Monte Carlo Singular Spectrum Analysis (SSA) Revisited: Detecting Oscillator Clusters in Multivariate Datasets

Andreas Groth*

Department of Atmospheric and Oceanic Sciences, and Institute of Geophysics & Planetary Physics, University of California at Los Angeles, Los Angeles, California

Michael Ghil

Department of Atmospheric and Oceanic Sciences, and Institute of Geophysics and Planetary Physics, University of California at Los Angeles, Los Angeles, California, and Geosciences Department and Laboratoire de Météorologie Dynamique (CNRS and IPSL), Ecole Normale Supérieure, Paris, France

ABSTRACT

Multichannel singular spectrum analysis (M-SSA) provides an efficient way to identify weak oscillatory behavior in high-dimensional data. To prevent the misinterpretation of stochastic fluctuations in short time series as oscillations, Monte Carlo (MC)–type hypothesis tests provide objective criteria for the statistical significance of the oscillatory behavior. Procrustes target rotation is introduced here as a key method for refining previously available MC tests. The proposed modification helps reduce the risk of type-I errors, and it is shown to improve the test's discriminating power. The reliability of the proposed methodology is examined in an idealized setting for a cluster of harmonic oscillators immersed in red noise. Furthermore, the common method of data compression into a few leading principal components, prior to M-SSA, is re-examined and its possibly negative effects are discussed. Finally, the generalized Procrustes test is applied to the analysis of interannual variability in the North Atlantic's sea surface temperature and sea level pressure fields. The results of this analysis provide further evidence for shared mechanisms of variability between the Gulf Stream and the North Atlantic Oscillation in the interannual frequency band.

1. Introduction

Over the last two decades, singular spectrum analysis (SSA) and its multivariate extension (M-SSA) have become widely used in the identification of intermittent or modulated oscillations in geophysical and climatic time series (Vautard et al. 1992). These methods are closely related to principal component analysis (PCA) or empirical orthogonal function (EOF) analysis; they are primarily designed for the reduction of the dimensionality of a given data set and the compression of a maximum of variance into a minimal number of robust components. In the identification of regularity, as well as the reduction to a simpler and easier to interpret picture of complex observations, SSA and M-SSA have demonstrated their usefulness in numerous applications; see Ghil et al. (2002b) for a comprehensive overview.

Both SSA and M-SSA decompose the time-delayed embedding of a given data set (Broomhead and King 1986a,b) into a set of data-adaptive orthogonal components, while M-SSA also takes cross-correlations into account. It turns out that these components can be classified essentially into trends, oscillatory patterns, and noise, and allow a reconstruction of a "skeleton" of the underlying dynamical system's structure (Vautard and Ghil 1989; Ghil and Vautard 1991; Vautard et al. 1992). Several practical aspects of SSA and its application to time series analysis are covered in Golyandina et al. (2001) and Golyandina and Zhigljavsky (2013).

In practice, we are usually confronted with the problem of the regular part of the behavior being rather weak and hidden by substantial noise. Without any a priori knowledge of the underlying dynamics—and

---
*Corresponding author address:
E-mail: andreasgroth@atmos.ucla.edu

by visual inspection, or by more common spectral methods alone—it may be difficult to find and formulate a proper model for the mechanisms that underlie the regularity, if any.

To prevent the misinterpretation of random fluctuations as oscillations in a univariate SSA analysis, Allen and Smith (1996, hereafter AS) formulated an objective statistical-significance test against a red-noise null hypothesis. In their Monte Carlo–type test, these authors simulate short-term temporal correlations of the time series by means of a first-order autoregressive process [AR(1) hereafter]; such a process does not support oscillatory behavior and it is, therefore, well adapted to the task.

In the case of multivariate data, it is not only temporal but spatial correlations, too, that have to be taken into account in the formulation of the null hypothesis. Allen and Robertson (1996) proposed a transformation of the data to pairwise uncorrelated principal components (PCs) by means of a conventional PCA analysis; their method then proceeds to fit independent AR(1) processes to each of the PCs.

Small mismatches in each of the null hypotheses are likely, however, to be amplified as the number of PCs increases, and this increases the risk of erroneously identifying oscillations where none are present. In an idealized experiment that uses harmonic oscillators hidden by irregular noise, we shall see that that such misidentification may occur even when the superimposed noise originates from an AR(1) process as well; that is, there is no formally erroneous specification in the definition of the null hypothesis.

In this paper, we introduce a modification of the Monte Carlo test of Allen and Robertson (1996) that helps reduce so-called type-I errors and improves the discriminating power of the test. We propose here to apply Procrustes target rotation (Green 1952; Hurley and Cattell 1962; Schönemann 1966) in matching M-SSA eigendecompositions of the null hypothesis' surrogate data with that of the observed data. In this setting, the Monte Carlo tests of AS and of Allen and Robertson (1996) emerge as special cases.

In our application to sea surface temperature (SST) data from the Simple Ocean Data Assimilation (SODA) reanalysis (Carton and Giese 2008; Giese and Ray 2011) and sea level pressure (SLP) data from the atmospheric 20CRv2 reanalysis (Compo et al. 2011), we rely furthermore on the varimax M-SSA analysis introduced recently by Groth and Ghil (2011). Feliks et al. (2013) applied such a varimax rotation to climatic indices, and showed that it helps reduce mixture problems in the EOFs and that it provides therewith much sharper spectral results.

In the present paper, we follow Feliks et al. (2011) and focus on SST data in the Gulf Stream region but combine it with SLP data in the entire North Atlantic region in a joint M-SSA analysis. The cleaner and sharper spectral results obtained herein support the previous authors' findings of very similar interannual peaks in the Gulf Stream SST data and the North Atlantic Oscillation (NAO). Given that the proposed Monte Carlo test improves M-SSA's discriminating power, our findings provide even stronger evidence for shared physical mechanisms between the Gulf Stream's meandering and the atmospheric NAO.

The remainder of the paper is organized as follows: In section 2, we first discuss the key features of M-SSA and the recently introduced varimax rotation of the M-SSA solution. We briefly review the Monte Carlo testing procedure as previously applied to M-SSA. In section 4, we introduce the concept of Procrustes target rotation as a generalization of such a test. In section 5, we apply the proposed testing methodology to an idealized statistical experiment. The results of this experiment are systematically evaluated in section 6. The proposed methodology is finally applied to observed SST and SLP data in section 7, and a summary of the results in section 8 concludes the paper.

2. Singular spectrum analysis (SSA)

Univariate SSA and its multivariate extension M-SSA rely on the classical Karhunen-Loève decomposition of a stochastic process. Broomhead and King (1986a,b) introduced them into dynamical system analysis, as a robust version of the Mañé-Takens idea to reconstruct dynamics from a time-delayed embedding of time series (Mañé 1981; Takens 1981). The method essentially diagonalizes the lag-covariance matrix with respect to a basis of orthogonal eigenvectors and computes the corresponding eigenvalues.

The M-SSA eigenvectors are often referred to as space-time empirical orthogonal functions (ST-EOFs, Plaut and Vautard 1994; Ghil et al. 2002b). The new data-adaptive eigenbasis is optimal in the sense that—for any truncation k with respect to the leading eigenelements—it minimizes the total mean squared error between the full data set and that k-truncated reconstruction. Hence, projecting the data set onto the data-adaptive eigenvectors simplifies the picture of a possibly high-dimensional complex system by viewing it in an optimal subspace.

Let $\mathbf{x} = \{x_d(n) : d = 1, \ldots, D; n = 1, \ldots, N\}$ be a multivariate time series with D channels of length N. Each channel d is embedded into an M-dimensional phase space by using lagged copies $\mathbf{X}_d(n) = [x_d(n), \ldots, x_d(n + M - 1)]$, with $n =$

$1, \ldots, N - M + 1$. The resulting matrix $\mathbf{X}_d$ is constant along the skew diagonals, and the full augmented trajectory matrix $\mathbf{X}$ is formed by concatenating all the channels according to $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_D)$. The matrix $\mathbf{X}$ has DM columns of reduced length $N' = N - M + 1$.

The first step of the M-SSA algorithm is to compute the covariance matrix $\mathbf{C}$. This matrix can be directly estimated from the trajectory matrix $\mathbf{X}$ (Broomhead and King 1986a,b; Allen and Robertson 1996) as

$$\mathbf{C} = \frac{1}{N'} \mathbf{X}' \mathbf{X}, \tag{1}$$

where $(\cdot)'$ indicates the transpose of a matrix, or via the Toeplitz approach of Vautard and Ghil (1989).

The latter explicitly imposes a Toeplitz structure—with constant sub- and superdiagonals—on the covariance matrix and the eigenvectors are then necessarily either symmetric or antisymmetric in the univariate case. This feature of the Toeplitz approach helps the detection of oscillations but it enhances the risk of spurious-oscillation identification as well. In the multichannel case (Keppenne and Ghil 1993; Plaut and Vautard 1994), each block has Toeplitz structure, while the "grand" block covariance matrix is symmetric. Hence all eigenvalues are real, but negative eigenvalues may appear as well. This negative bias has to be compensated by a positive bias in the positive eigenvalues, which may reduce the power of the statistical test.

The negative bias in the smallest eigenvalues of the Toeplitz approach originates from the fact that the covariance matrix is not directly estimated from the product of a trajectory matrix as in Eq. (1), but rather from the covariance function estimation, mapped into Toeplitz form. Vautard and Ghil (1989) suggested using the common unbiased covariance estimator,

$$c(\tau) = \frac{1}{N - \tau} \sum_{n=1}^{N-\tau} x(n) x(n + \tau), $$

as resulting from the scalar product of the longest-available matching segments. This estimate uses more information from the time series than the trajectory-matrix approach in Eq. (1).

A possible compromise, likewise suggested by Vautard et al. (1992), could be that of a biased estimator of the covariance function,

$$\tilde{c}(\tau) = \frac{1}{N} \sum_{n=1}^{N-\tau} x(n) x(n + \tau). $$

This scalar product can be written—by analogy with Eq. (1)—as a matrix product $\widetilde{\mathbf{C}} = \widetilde{\mathbf{X}}' \widetilde{\mathbf{X}} / N$ of the extended trajectory matrix, $\widetilde{\mathbf{X}}_d(n) = [x_d(n), \ldots, x_d(n + M - 1)]$, with $n = -M + 1, \ldots, N$, and all unknown values at either end, $n < 1$ and $n > N$, respectively, set to zero. Note that $\widetilde{\mathbf{X}}$ has thus columns of length $N + M - 1$. This way, the covariance matrix estimation $\widetilde{\mathbf{C}}$ has Toeplitz structure in each block and is positive semidefinite as well. A drawback of this approach is obviously the problem of the Gibbs phenomenon at either end of the time series due to the zero-padding.

For simplicity, we rely here on the trajectory approach of Eq. (1) to calculate $\mathbf{C}$, rather than on either version of the two versions above of the Toeplitz approach.

The next step of the algorithm is to diagonalize $\mathbf{C}$, which yields

$$\mathbf{\Lambda} = \mathbf{E}' \mathbf{C} \mathbf{E}, \tag{2}$$

where $\mathbf{\Lambda}$ is a diagonal matrix of DM real eigenvalues $\{\lambda_k : k = 1, \ldots DM\}$, and a matrix $\mathbf{E}$ whose columns are the associated eigenvectors $\mathbf{e}_k$. The ST-EOFs that enter $\mathbf{E}$ are composed of D consecutive segments $\mathbf{e}_{dk}$ of length M, each of which is associated with a channel in $\mathbf{X}_d$, $e_{dk}(m) \equiv e_k((d - 1)M + m)$.

Projecting the augmented trajectory matrix $\mathbf{X}$ onto the eigenbasis $\mathbf{E}$,

$$\mathbf{A} = \mathbf{X} \mathbf{E}, \tag{3}$$

yields the corresponding principal components (PCs) as columns $\mathbf{a}_k$ of $\mathbf{A}$. These PCs are pairwise uncorrelated at zero lag and have a reduced length of N'. Note that the trajectory approach ensures uncorrelated PCs, since $\mathbf{A}' \mathbf{A} = \mathbf{E}' \mathbf{X}' \mathbf{X} \mathbf{E} = N' \mathbf{\Lambda}$. This is not necessarily the case in the unbiased Toeplitz approach, in particular not for time series with strong nonstationarity over the observed time interval or for large values of $M \simeq N/2$.

In general, however, we are interested in the reconstruction of the dynamical behavior that generated the time series $\mathbf{X}$ by using only a subset $\mathcal{K} \subseteq \{1, \ldots, DM\}$ of the ST-EOFs. Such a reconstruction requires an inverse transformation of Eq. (3),

$$\mathbf{R}_{\mathcal{K}} = \mathbf{X} \mathbf{E} \mathbf{K} \mathbf{E}' \equiv \mathbf{A} \mathbf{K} \mathbf{E}', \tag{4}$$

with $\mathbf{K}$ a diagonal matrix of size $DM \times DM$, in which elements $K_{kk} = 1$ if $k \in \mathcal{K}$ and $K_{kk} = 0$ otherwise. Note that $\mathcal{K} = \{1, \ldots, DM\}$ yields a complete reconstruction of $\mathbf{X}$.

Given the PCs in $\mathbf{A}$ and the ST-EOFs in $\mathbf{E}$, the final M-SSA step is to average along the skew diagonals of $\mathbf{R}$,

$$r_{dk}(n) = \frac{1}{M_n} \sum_{m=L_n}^{U_n} a_k(n - m + 1) e_{dk}(m), \tag{5}$$

in order to determine a set of DM reconstructed components (RCs) for each of the original input channels $\mathbf{x}_d$ (Ghil and Vautard 1991; Vautard et al. 1992; Plaut and Vautard 1994). The normalization factor M_n and the summation bounds L_n and U_n for the central part of the time series, $M \leq n \leq N - M + 1$, are simply $(M_n, L_n, U_n) = (M, 1, M)$; for their values at either end, see Vautard et al. (1992) or Ghil et al. (2002b).

Still, PCA overall was primarily designed for dimensionality reduction and signal compression of multichannel data. It is thus not clear, a priori, how informative PCA in general or M-SSA in particular can be in the interpretation of the underlying system's dynamics or structure. Both PCA and M-SSA, in fact, suffer from degeneracy of eigenvectors when the corresponding eigenvalues are similar in size (North et al. 1982). Instead of clearly separating structural distinct dynamical phenomena—e.g., distinct oscillations in the case of M-SSA—one often observes a mixture of two or more eigenvectors.

A common approach to reduce such mixture effects and to improve the physical interpretation of the results is to perform a rotation of the eigenvectors, $\mathbf{E}^* = \mathbf{E}\mathbf{T}_V$ (Richman 1986; Jolliffe 2002). In Appendix A, we briefly recall the classical varimax rotation algorithm (Kaiser 1958) and motivate a necessary modification thereof for M-SSA eigenvectors (Groth and Ghil 2011).

In the following, we no longer explicitly distinguish between unrotated and rotated eigenelements and thus drop the superscript $(\cdot)^*$.

3. Monte Carlo SSA: a short review

Usually, the set of eigenvector-eigenvalue pairs are ranked in descending order of the eigenvalues. This informal ranking, however, should not be confused with the order of significance. It is only by testing against a specific null hypothesis that one can draw further conclusions about significant deterministic behavior, cf. AS.

In this context, AS proposed a framework of Monte-Carlo testing for SSA that compares the variance captured by the data eigenvalues with that of an ensemble of surrogate data. Since climate and other geophysical records tend to have larger power at lower frequencies, the authors discuss the null hypothesis of an AR(1) process; note, though that the class of null hypotheses is not restricted to such simple, linear stochastic processes. The AS approach also provides a low-bias estimator for the model parameters, given even a short time series. In the case of multivariate data sets, Allen and Robertson (1996) proposed first a rotation to uncorrelated principal components by means of a classical PCA, prior to the estimation of independent AR(1) processes.

Once an appropriate model for the null hypothesis has been formulated, an ensemble of surrogate data $\mathbf{x}_R$ of the same length N and dimension D as the original data set is generated and, for each realization, the covariance matrix $\mathbf{C}_R$ is determined. To compare the data eigendecomposition with the ensemble of surrogate covariance matrices, AS discussed several approaches.

A first possibility is to project each covariance matrix $\mathbf{C}_R$ onto the data eigenvectors $\mathbf{E}$ using Eq. (2),

$$\mathbf{\Lambda}_R^{(E)} = \mathbf{E}'\mathbf{C}_R\mathbf{E}. \tag{6}$$

Since Eq. (6) is not the eigendecomposition of $\mathbf{C}_R$, the matrix $\mathbf{\Lambda}_R^{(E)}$ is not necessarily diagonal, but its diagonal elements yield the resemblance between the null hypothesis' variance and the data variance in $\mathbf{\Lambda}$ in the eigen-directions of $\mathbf{E}$. From a sufficiently large ensemble of surrogate data, confidence intervals can be derived; outside of these intervals, the data eigenvalues can be considered to be significantly different from the null hypothesis.

However, as AS further pointed out, this test tends to be too lax. Since the eigendecomposition puts maximum variance into a minimal number of data adaptive components, artificial variance compression may occur: that is, SSA may account for too much variance in the largest eigenvalues and too little in the smallest. This increases the likelihood of the largest data eigenvalues being significant. Later on we shall see that this effect is amplified when the number of channels increases, e.g. when $DM \approx N'$.

Groth and Ghil (2011) have shown, furthermore, that this undesired effect of artificial variance compression can be at least partly reduced by a subsequent varimax rotation. The latter relaxes slightly the diagonal form of the eigenvalue matrix $\mathbf{\Lambda}^*$ in favor of a more simplified eigenvector structure that, in turn, flattens the spectrum of eigenvalues. In the presence of multiple oscillatory pairs, the rotated solution improves the pairing of nearly equal eigenvalues; see Fig. 1 in Groth and Ghil (2011). Even so, the problem of Eq. (6) not being truly optimal for constructing the surrogate covariance matrices is still present and, with it, the risk of the test not being stringent enough.

Alternatively, AS discussed the possibility of obtaining an eigendecomposition for each surrogate realization separately,

$$\mathbf{\Lambda}_R = \mathbf{E}'_R\mathbf{C}_R\mathbf{E}_R, \tag{7}$$

and to compare the shape of the ranked eigenspectra of $\mathbf{\Lambda}_R$ and $\mathbf{\Lambda}$; see also Elsner and Tsonis (1994) and

Elsner (1995). This way the effect of variance compression will be present in both the data and surrogate data eigenspectra. However, the rank order of the eigenvalues alone, without the information contained in the corresponding eigenvectors, has very little to say about the underlying dynamics. We will return to this idea to obtain an eigendecomposition for each surrogate realization separately in the definition of our hypothesis test in section 4.

Finally, AS discussed the following possibility: Instead of analyzing each surrogate realization separately, they average first over the entire set of all surrogate covariance matrices to estimate $\mathbf{C}_N$, the expected covariance matrix of the null hypothesis. Irregularities that are likely to occur in short realizations are smoothed out and the eigendecomposition $\mathbf{C}_N = \mathbf{E}_N \mathbf{\Lambda}_N \mathbf{E}'_N$ of this average covariance matrix is, in contrast to each single eigendecomposition in Eq. (7), much less susceptible to artificial variance compression. Both the data covariance matrix and all the surrogate covariance matrices are projected onto the eigenvectors of the null hypothesis,

$$\mathbf{\Lambda}^{(N)} = \mathbf{E}'_N \mathbf{C}\, \mathbf{E}_N, \tag{8a}$$

$$\mathbf{\Lambda}^{(N)}_R = \mathbf{E}'_N \mathbf{C}_R \mathbf{E}_N. \tag{8b}$$

Initially, this test gives only confidence intervals for $\mathbf{\Lambda}^{(N)}$ and not for $\mathbf{\Lambda}$. To link the two, AS proposed to associate the corresponding data eigenvectors $\mathbf{E}$ and null-hypothesis eigenvectors $\mathbf{E}_N$ via their dominant frequency. Paluš and Novotná (2004) took this idea further and paired the data eigenvectors $\mathbf{E}$ with the surrogate eigenvectors $\mathbf{E}_R$ in Eq. (7), based on their dominant frequency.

Such a pure frequency encoding of eigenvectors is quite helpful for single-channel SSA analysis (Ghil et al. 2002b) but it can be rather misleading in the multi-channel setting of M-SSA. In the latter, is is possible that two eigenvectors have similar frequency but different spatial patterns and thus could be linked to different types of dynamical behavior; e.g., two oscillators that are only slightly out of tune but not actually synchronized (Groth and Ghil 2011). The frequency pairing would then associate the same significance level to both eigenvectors, although their corresponding eigenvalues might be quite different.

In the end, the test in Eq. (8) may carry the opposite risk of being too conservative and thus not sufficiently sensitive. Since the null hypothesis approximates only certain aspects of the data set, its eigendecomposition may only be suitable for the description of the data set to a limited extent. In this context, AS noted that a weak signal may not align very well with the eigenvectors of the null hypothesis and that it could, therefore, be missed altogether.

Finally, this test on the EOFs of the null hypothesis may not take into account all the advantages of a varimax M-SSA.

Significance tests on variance can be complemented by using other statistical properties of the oscillatory modes to distinguish regular behavior from noise, e.g. Paluš and Novotná (2004) and Holmström and Launonen (2013).

4. Revisiting Monte Carlo SSA

In the following, we propose a Monte Carlo SSA algorithm that operates on the data eigenbasis in such a way as to keep its optimal data description properties, as well as its high sensitivity in the identification of weak signals. At the same time, we allow for small mismatches between the data generating process and the null hypothesis in order to reduce the risk of the test being either too lenient or overly conservative.

a. Procrustes target rotation

In the first step, we determine the eigendecomposition of the data set in Eq. (2) as well as the eigendecomposition of each surrogate realization in Eq. (7) separately. We recall that the projection in Eq. (6) can be rewritten as a similarity transformation,

$$\mathbf{\Lambda}^{(E)}_R = \mathbf{E}' \mathbf{C}_R \mathbf{E} \tag{9a}$$

$$= \mathbf{E}' \mathbf{E}_R \mathbf{\Lambda}_R \mathbf{E}'_R \mathbf{E} \tag{9b}$$

$$= \mathbf{T}' \mathbf{\Lambda}_R \mathbf{T}, \tag{9c}$$

with the transformation matrix $\mathbf{T} = \mathbf{E}'_R \mathbf{E}$. Insofar as $\mathbf{E}$ and $\mathbf{E}_R$ are of full rank, the transformation matrix $\mathbf{T}$ is orthogonal and the total variance is preserved; for the case of a rank-deficient covariance matrix, see the following section.

The linear maps in Eq. (9) are based on the structure of the eigenvectors and they disregard completely the eigenvalue spectrum. It is, however, the combination of both eigenvalues and eigenvectors that determines the spatio-temporal correlation structure of the underlying dynamical system. To improve the comparison of the data eigendecomposition with that of the surrogate data, we have to take both aspects into account.

In doing so, we first scale data and surrogate eigenvectors by their corresponding singular values $\mathbf{\Sigma} = \mathbf{\Lambda}^{1/2}$, and then look for an orthogonal rotation matrix $\mathbf{T}_E$ that provides maximal similarity between $\mathbf{E}\mathbf{\Sigma}$ and $\mathbf{E}_R \mathbf{\Sigma}_R \mathbf{T}_E$. In general, the projection $(\mathbf{E}_R \mathbf{\Sigma}_R)' \mathbf{E}\mathbf{\Sigma}$ is not orthogonal, but an optimal solution to this target rotation problem is given by the orthogonal Procrustes target rotation (Green 1952; Hurley and Cattell 1962; Schönemann 1966)

$$\mathbf{T}_E = \mathbf{U}\mathbf{V}' \tag{10a}$$

that yields the SVD

$$(\mathbf{E}_{\mathrm{R}}\mathbf{\Sigma}_{\mathrm{R}})' \, \mathbf{E}\mathbf{\Sigma} = \mathbf{U}\mathbf{S}\mathbf{V}'. \qquad (10b)$$

The solution $\mathbf{T}_E$ is optimal inasmuch as it minimizes the Frobenius norm of the difference $\mathbf{Y} = \mathbf{E}_{\mathrm{R}}\mathbf{\Sigma}_{\mathrm{R}}\mathbf{T}_E - \mathbf{E}\mathbf{\Sigma}$, i.e. $\{\mathrm{tr}\,(\mathbf{Y}'\mathbf{Y})\}^{1/2}$.[1]

Finally, we compare the data eigenvalues $\mathbf{\Lambda}$ with the diagonal elements of the similarity transformation,

$$\mathbf{\Lambda}_{\mathrm{R}}^{(E\Sigma)} = \mathbf{T}_E'\mathbf{\Lambda}_{\mathrm{R}}\mathbf{T}_E. \qquad (11)$$

Note that substituting for $\mathbf{\Lambda}_{\mathrm{R}}$ in Eq. (11) its eigen-decomposition (7) gives the equivalent formulation

$$\mathbf{\Lambda}_{\mathrm{R}}^{(E\Sigma)} = \mathbf{T}_E'\mathbf{E}_{\mathrm{R}}'\mathbf{C}_{\mathrm{R}}\mathbf{E}_{\mathrm{R}}\mathbf{T}_E. \qquad (12)$$

Hence, the surrogate covariance matrix is not directly projected onto the data eigenvectors, as in Eq. (6), thus avoiding the already discussed risk of a test that is too lenient. Instead, we project onto a close approximation of the former, i.e., onto $\mathbf{E}_{\mathrm{R}}\mathbf{T}_E \approx \mathbf{E}$, and we expect this modified test to be more conservative.

Cliff (1966) has suggested, as an alternative to the target rotation of one eigenbasis onto another, the possibility of comparing the two eigenbases with a common basis of maximum similarity. A solution to this problem is likewise given by Eq. (10), with the rotation of the data and surrogate data eigenvectors to $\mathbf{E}\mathbf{V}$ and $\mathbf{E}_{\mathrm{R}}\mathbf{U}$, respectively. As in the case of Eq. (8), though, this approach complicates the interpretation, since no confidence intervals for the data eigenvalues are available.

b. Rank-deficient covariance matrix

So far we have assumed covariance matrices $\mathbf{C}$ of full rank. For the trajectory-matrix–based approach in Eq. (1), this condition is, in general, fulfilled as long as $N' \geq DM$. This is a necessary condition for the Monte Carlo test to work correctly, and its violation yields to a loss of part of the total variance in the projection given by Eq. (6); such a loss will result in significance intervals that are too low.

To overcome this limiting factor in the M-SSA analysis of short time series, Allen and Robertson (1996) proposed the complementary eigendecomposition of a reduced covariance matrix $\mathbf{X}\mathbf{X}'$,

$$\frac{1}{DM}\mathbf{X}\mathbf{X}' = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'. \qquad (13)$$

Both, $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}\mathbf{X}'$ share the same non-vanishing eigenvalues and their eigenelements are linked via

$\mathbf{X} = \eta^{1/2}\mathbf{P}\mathbf{\Sigma}\mathbf{E}'$, the SVD of $\mathbf{X}$, where η is equal to the larger of N' and DM. In case of varimax rotation, $\mathbf{P}$ becomes rotated as well, $\mathbf{P}^* = \mathbf{P}\mathbf{T}_V$, and the SVD becomes $\mathbf{X} = \eta^{1/2}\mathbf{P}\mathbf{T}_V\mathbf{T}_V'\mathbf{\Sigma}\mathbf{T}_V\mathbf{T}_V'\mathbf{E}'$, with $\mathbf{T}_V\mathbf{T}_V' = \mathbf{I}_\eta$ the identity matrix. The left-singular vectors $\mathbf{P}$—also referred to as time EOFs (T-EOFs)—describe only temporal behavior; see also Appendix A1 in Ghil et al. (2002b).

The Allen and Robertson (1996) test continues by projecting the reduced covariance matrix onto $\mathbf{P}$,

$$\mathbf{\Lambda}_{\mathrm{R}}^{(P)} = \frac{1}{DM}\mathbf{P}'\mathbf{X}_{\mathrm{R}}\mathbf{X}_{\mathrm{R}}'\mathbf{P}; \qquad (14)$$

it derives the significance level for $\mathbf{\Lambda}$ from the statistics of the diagonal elements of $\mathbf{\Lambda}_{\mathrm{R}}^{(P)}$.

In the same way as for ST-EOFs, we could also construct a test based on the scaled T-EOFs, $\mathbf{P}\mathbf{\Sigma} \equiv \mathbf{A}$, i.e. on the PCs of $\mathbf{C}$. Similar to the scaled target rotation onto ST-EOFs, we first determine an orthogonal rotation matrix $\mathbf{T}_P = \mathbf{U}\mathbf{V}'$ from the SVD $(\mathbf{P}_{\mathrm{R}}\mathbf{\Sigma}_{\mathrm{R}})'\mathbf{P}\mathbf{\Sigma} \equiv \mathbf{A}_{\mathrm{R}}'\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}'$ and then compare the data eigenvalues $\mathbf{\Lambda}$ with the diagonal elements of a similarity transform,

$$\mathbf{\Lambda}_{\mathrm{R}}^{(P\Sigma)} = \mathbf{T}_P'\mathbf{\Lambda}_{\mathrm{R}}\mathbf{T}_P. \qquad (15)$$

The test based on the reduced covariance matrix is essentially a univariate test, since no cross-channel covariance information is used in $\mathbf{X}\mathbf{X}'$. The usefulness of the test thus strongly depends on the input channels being uncorrelated. This can be partly achieved by transforming the input channels into spatial PCs, although correlations at other time lags may not vanish.

A solution to this rank-deficiency problem is likewise given by the target rotation onto ST-EOFs, as described in the section before. In this algorithm, it is the multiplication of the eigenvectors by their corresponding singular values that automatically restricts the projection to non-vanishing eigenelements and the orthogonality constraints in $\mathbf{T}_E$ ensure the conservation of the total variance.

As a special case of this solution, we could also imagine to simply remove all vanishing eigenelements from $\mathbf{E}$ as well as $\mathbf{E}_{\mathrm{R}}$—without rescaling the elements retained by their corresponding singular values—and then determine an orthogonal transformation such as in Eq. (10). We will refer to this alternative as *unscaled* target rotation to distinguish it from the *scaled* target rotation in Eq. (11). It will be shown, however, that the rescaling of the EOFs by their singular value, in the latter, significantly improves the power of the modified Monte Carlo test proposed herein. Note that the unscaled target-rotation onto ST-EOFs includes the projection algorithm of AS as a special case for $DM \leq N'$, cf. Eq. (9).

---

[1] In case of varimax-rotated eigenelements, the target becomes $\mathbf{E}^*\mathbf{\Sigma}^* = \mathbf{E}\mathbf{T}_V\mathbf{T}_V'\mathbf{\Sigma}\mathbf{T}_V = \mathbf{E}\mathbf{\Sigma}\mathbf{T}_V$, and $\mathbf{T}_E$ simply includes a further rotation, given by $\mathbf{T}_E^* = \mathbf{T}_E\mathbf{T}_V$.

c. Composite null hypothesis

Once a certain part of the time series has been identified as signal, AS proposed a single-channel SSA composite test for the remainder of the time series against an AR(1) null hypothesis.

Following AS, we define $\mathbf{K}$ as a diagonal matrix of size $M \times M$ with $K_{kk} = 0$ if the corresponding EOF has been identified as signal and $K_{kk} = 1$ otherwise; i.e. the signal can be reconstructed from the filtered trajectory matrix $\mathbf{R} = \mathbf{XE}(\mathbf{I} - \mathbf{K})\mathbf{E}'$. The authors then proceed to fit the covariance matrix $\mathbf{C}_{\mathrm{N}}$ associated with the null hypothesis to the data covariance matrix $\mathbf{C}$ in a subspace spanned by the "noise" EOFs.

With the "noise projection matrix" $\mathbf{Q} = \mathbf{EKE}'$, the fitting conditions then are given by

$$\mathrm{tr}_0\left(\mathbf{QC}_{\mathrm{N}}\mathbf{Q}\right) = \mathrm{tr}_0\left(\mathbf{QCQ}\right), \qquad (16\mathrm{a})$$

$$\mathrm{tr}_1\left(\mathbf{QC}_{\mathrm{N}}\mathbf{Q}\right) = \mathrm{tr}_1\left(\mathbf{QCQ}\right), \qquad (16\mathrm{b})$$

where $\mathrm{tr}_j(\mathbf{C}) = (M - j)^{-1} \sum_{k=1}^{M-j} C_{k,k+j}$ is a generalized trace operator. Furthermore, AS provide an unbiased estimator of the AR(1) parameters given that the process mean underlying $\mathbf{x}$ is unknown, i.e., $\mathbf{x}$ is centered by subtracting its mean value instead.

In the test against a pure-noise null hypothesis, $\mathbf{K}$ and $\mathbf{Q}$ become identity matrices, and Eq. (16) equals the common fit of the variance and lag-1 correlation of the AR(1) process to that of the observed time series. This equivalence, however, is only correct when $\mathbf{C}$ has Toeplitz structure with constant sub- and superdiagonals. In the trajectory approach, this is in general not the case—i.e. for short time series with strong nonstationarity over the observed time interval, or for large M—and we estimate the AR(1) parameters from Eq. (16) instead, even when $\mathbf{Q}$ becomes an identity matrix. Doing so ensures, in particular, that the expectation value satisfy $\mathcal{E}\left\{\mathrm{tr}_0\left(\mathbf{K}\mathbf{\Lambda}_{\mathrm{R}}^{(E)}\right)\right\} = \mathrm{tr}_0\{\mathbf{K}\mathbf{\Lambda}\}$.

In our proposed target-rotation algorithm, the individual realizations of the surrogate covariance matrix $\mathbf{C}_{\mathrm{R}}$ are projected onto a Procrustes approximation of the data EOFs. To account for this modification of the original AS null-hypothesis test, the covariance matrix $\mathbf{C}_{\mathrm{N}}$ associated with the null hypothesis has to be projected onto a Procrustes approximation as well. Given the eigendecomposition of this matrix, $\mathbf{C}_{\mathrm{N}} = \mathbf{E}_{\mathrm{N}}\mathbf{\Lambda}_{\mathrm{N}}\mathbf{E}_{\mathrm{N}}'$, we derive an orthogonal rotation matrix $\mathbf{T} = \mathbf{UV}'$ that yields the SVD $\left(\mathbf{E}_{\mathrm{N}}\mathbf{\Lambda}_{\mathrm{N}}^{1/2}\right)' \mathbf{E}\mathbf{\Sigma} = \mathbf{USV}'$. The projection $\mathbf{Q}$ of the noise covariance matrix $\mathbf{C}_{\mathrm{N}}$ is modified thereafter to read $\widetilde{\mathbf{Q}} = \mathbf{E}_{\mathrm{N}}\mathbf{TKT}'\mathbf{E}_{\mathrm{N}}'$ and the conditions for the

AR(1) parameter estimation in case of the Procrustes target rotation become

$$\mathrm{tr}_j\left(\widetilde{\mathbf{Q}}\mathbf{C}_{\mathrm{N}}\widetilde{\mathbf{Q}}\right) = \mathrm{tr}_j\left(\mathbf{QCQ}\right), \quad \text{for } j = 0, 1. \qquad (17)$$

This modification of Eq. (16) is important inasmuch as more EOFs are identified as part of the signal in $\mathbf{K}$. In the test against a pure-noise null hypothesis, $\widetilde{\mathbf{Q}}$ becomes again an identity matrix, and Eq. (17) simplifies to (16).

The AS composite test we considered so far applies only to the single-channel case of SSA. In the multichannel case of M-SSA, the covariance matrix $\mathbf{C}$ includes cross-correlations as well. In the test against a pure-noise null hypothesis, this can be at least partly resolved by transforming the input channels into spatial PCs. The extent to which this is valid in the test against a composite null hypothesis, however, depends largely on the structure of $\mathbf{Q}$. Since individual channels become correlated in the filtered trajectory matrix $\mathbf{R}$, a test against independent AR(1) processes is no longer feasible.

This problem does not appear in the null hypothesis test on T-EOFs, since no cross-channel covariance information is used and the reduced covariance matrix,

$$\mathbf{XX}' = \sum_{d=1}^{D} \mathbf{X}_d\mathbf{X}_d', \qquad (18)$$

is simply the average of D individual $N' \times N'$ single-channel covariance matrices, each of rank M, as long as $M \leq N'$; see also Allen and Robertson (1996). This feature of the single-channel case has the advantage that independent AR(1) processes can be fitted to each of these single-channel covariances matrices, whether one tests against pure noise or against a composite null hypothesis. In the latter case, the noise projection matrix becomes $\mathbf{Q} = \mathbf{PKP}'$ and we proceed by fitting independently for each channel d a covariance matrix $\mathbf{C}_{\mathrm{N},d}$ of size $N' \times N'$ associated with a single-channel null hypothesis, according to

$$\mathrm{tr}_j\left(\mathbf{QC}_{\mathrm{N},d}\mathbf{Q}\right) = \mathrm{tr}_j\left(\mathbf{QX}_d\mathbf{X}_d'\mathbf{Q}\right), \quad \text{for } j = 0, 1. \quad (19)$$

In the multi-channel case, the covariance matrix $\mathbf{C}_{\mathrm{N}} = \sum_{d=1}^{D} \mathbf{C}_{\mathrm{N},d}$ given by Eq. (19) provides, in particular, a solution for

$$\mathrm{tr}_j\{\mathbf{QC}_{\mathrm{N}}\mathbf{Q}\} = \mathrm{tr}_j\{\mathbf{QXX}'\mathbf{Q}\}, \quad \text{for } j = 0, 1. \qquad (20)$$

Note that the problem in Eq. (20) is underdetermined, since its solutions actually involve two parameters for each of the D independent AR(1) processes, and the channel-wise solution from Eq. (19) is not unique.

This indeterminacy is a shortcoming of the test on T-EOFs, which are invariant with respect to an orthogonal rotation of the input channels. We will come back to this problem when comparing the null hypothesis test on T-EOFs with that on ST-EOFs in the presence of cross correlations. In the latter test, the ST-EOFs are not invariant and we proceed with the channel-wise solution of Eq. (19).

A drawback of this solution is the fact that only the projection onto T-EOFs, cf. Eq. (14), satisfies $\mathcal{E}\left\{\mathrm{tr}_0\left(\mathbf{K}\mathbf{\Lambda}_{\mathrm{R}}^{(P)}\right)\right\} = \mathrm{tr}_0\{\mathbf{K}\mathbf{\Lambda}\}$. We may always want, however, to verify the results of a significance test on T-EOFs with those obtained by projecting onto ST-EOFs, in case there is no a priori reason to assume uncorrelated input channels at all time lags $\{0, \ldots, M-1\}$.

Note that in the multichannel parameter estimation of Eq. (19), there is no equivalent formulation to the single-channel case of Eq. (17), i.e., no equivalent to a parameter estimation that would take the Procrustes target rotation into account.

5. Experimental design

To illustrate the limitations of the Monte Carlo tests in section 3, we consider the idealized case of a cluster of harmonic oscillators with red noise superimposed on the observations. Hence the model specification of the null hypothesis is correct.

We create $D = 5$ time series of length $N = 250$; each of these is a composite of harmonic oscillations with different fundamental periods $T = \{7.6, 5.0, 2.7, 2.3\}$. For each channel and each oscillator, the initial phases and amplitudes are randomly and independently drawn in the intervals $[0, 2\pi]$ and $[0, A]$, respectively. The upper limit $A = A(T)$ depends on the period length T and it has the same frequency dependence as the superimposed AR(1) process, i.e. $A(T) \sim \left|1 - \gamma\, e^{-2\pi i/T}\right|^{-1}$ with γ the damping parameter of the AR(1) process. We consider the optimal case of independent AR(1) processes for each of the D channels and set $\gamma = 0.65$. The signal-to-noise ratio here is 1:4.

The multichannel time series is first transformed to uncorrelated PCs by means of a classical PCA analysis and individual AR(1) processes are fitted to each of the PCs. Then, surrogate realizations of length N are created and transformed back.

Figure 1 shows the resulting eigenvalue spectrum for a typical realization of the simulation experiment. Both panels show the same spectrum but with different significance level.

In Fig. 1a, the error bars are derived from the projection algorithm of Eq. (6). For each surrogate re-
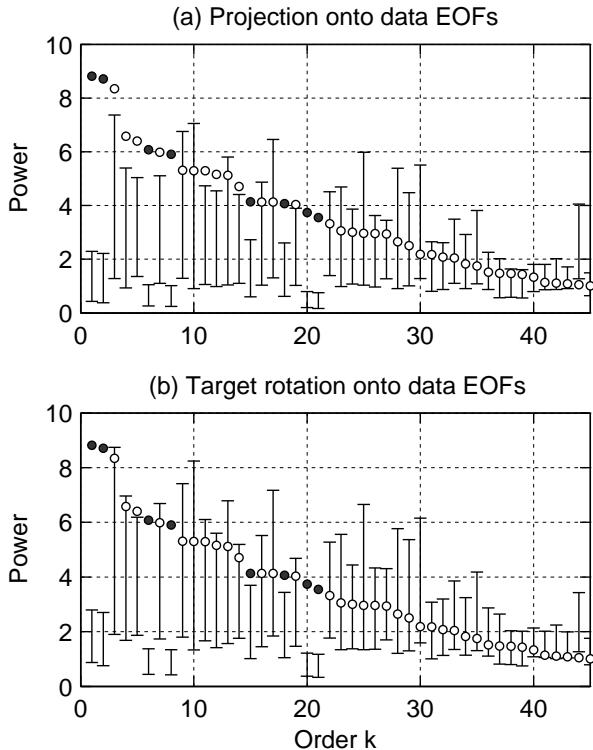

FIG. 1. Eigenvalue spectrum for a cluster of harmonic oscillators with observational red noise. (a, b) The eigenvalues that belong to the harmonic and the noise part—i.e., to the line spectrum and the continuous one—are marked as filled and open circles, respectively. The error bars indicate the 1% and 99% quantile from an ensemble of 500 surrogate data. The Monte Carlo test corresponds to (a) the classical projection algorithm of Eq. (6); and (b) the proposed scaled target-rotation algorithm of Eq. (11). The window length is $M = 40$ and the number of varimax rotated components is $S = 40$; see text for details.

alization, the covariance matrix is determined and projected onto the data ST-EOFs.

We clearly observe the limitation of the significance test, since more eigenvalues than expected are significant. Especially among the largest eigenvalues there are some that correspond to the noise (open circles) and that, nonetheless, appear as significant oscillations. The appearance of such *false positives* (FPs) reduces the precision and limits the explanatory power of the test. All true oscillations (filled circles), however, also reject the null hypothesis of random fluctuations. These oscillations are referred to as *true positives* (TPs).

In Fig. 1b, the error bars are derived from the scaled target-rotation algorithm of Eq. (11). We first determine the eigendecomposition for each of the former surrogate realizations and then look for an optimal orthogonal rotation towards the data eigende-
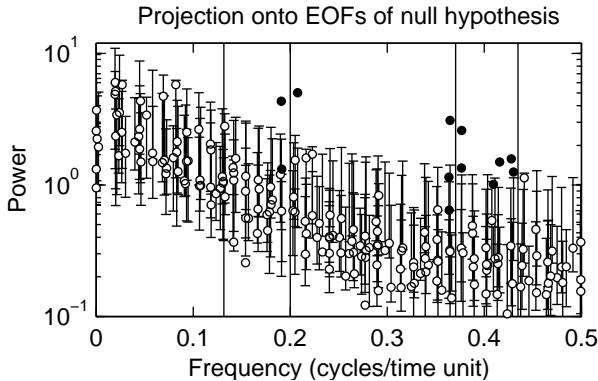
Projection onto EOFs of null hypothesis

FIG. 2. Same data and parameters as in Fig. 1. Both data and surrogates are projected onto the ST-EOFs of the null-hypothesis covariance matrix $\mathbf{C}_N$, cf. Eq. (8). Vertical lines indicate the frequencies of the harmonic oscillators.

composition, as described in section 4. We see that this algorithm reduces the number of FPs to none but keeps the number of TPs as high as in panel (a) and equal to the correct one.

In this experiment, we have chosen the parameters in the limiting case $DM = N'$, i.e. the covariance matrix $\mathbf{C}$ becomes a square matrix and we consider only significance test based on ST-EOFs. Allen and Robertson (1996) have already noted that, when DM reaches N', the hypothesis test becomes less reliable. In section 6 we will show that this undesirable effect is largely reduced by using scaled target-rotation algorithm.

Furthermore, we have performed a significance test on the null-hypothesis basis according to Eq. (8). It turns out that, with this test, only three out of four oscillations are detected. Figure 2 thus confirms the findings of AS that the test is less sensitive in the detection of weak signals. Moreover, we observe that each of the true oscillations are not necessarily captured by a single EOF pair of $\mathbf{C}_N$, rather than by three or even more significant eigenvalues. In contrast to single-channel SSA, it is quite likely in M-SSA to observe several EOFs with similar frequency but different spatial pattern, whose eigenvalues are found to be significant.

6. Sensitivity vs. specificity

a. Basic definitions

We consider now the identification of an oscillatory pair as a binary classification test. Such a test is characterized by its sensitivity and specificity.

The *sensitivity* is a statistical measure of the proportion of TPs that are correctly identified as such,

i.e.

$$\text{sensitivity} = \frac{\#\text{TP}}{\#\text{TP} + \#\text{FN}}, \qquad (21)$$

where $\#$ is the cardinality of a set and the FNs are the *false negatives*, i.e., the missed oscillations; sensitivity is also referred to as true positive rate or hit rate. *Specificity*, on the other hand, measures the true negative rate and is given by

$$\text{specificity} = \frac{\#\text{TN}}{\#\text{TN} + \#\text{FP}}, \qquad (22)$$

with the TNs being the *true negatives*.

Ideally, one wishes for a test that is 100 % sensitive, i.e., it identifies all actual oscillations, and also 100 % specific, i.e., it doesn't mistake any spurious oscillation as an actual one. But, in practice, there is always a trade-off between these two properties of a test, and any classification test has a minimum error, known as the Bayes error rate (Fukunaga 1990).

In the following, we proceed with the two tests based on the data EOFs that we first compared in Fig. 1 and evaluate more systematically their capability to identify true oscillatory components. Based on an ensemble of multiple repetitions of the experiment, we thus evaluate the sensitivity as well as the specificity of these two tests.

First, we determine the eigendecomposition of a noise-free realization of the set of harmonic oscillators of section 5, in order to get a reference set of eigenelements that describes the harmonic part of the full time series. Next, we add the desired amount of red noise and rerun the eigenanalysis.

To identify true oscillations in the eigendecomposition of the noise-contaminated data set, we (i) project the noise-free covariance matrix onto the noise-contaminated eigendecomposition, cf. Eq. (6), as well as (ii) apply the proposed target-rotation algorithm from the noise-free eigenelements to the noise-contaminated eigenelements, Eq. (12). This way, the noise-free realization is compared with the noise-contaminated eigendecomposition exactly in the same way as we compare the surrogate realizations in the two significance tests.

Corresponding to the four imposed oscillations in the noise-free realization, we consider the eight largest elements in $\mathbf{\Lambda}_R^{(E)}$ and $\mathbf{\Lambda}_R^{(E\Sigma)}$, respectively, as being true in the noise-contaminated eigendecomposition. We will see forthwith that it is not necessarily the largest eigenvalues that correspond to the true oscillations.

Typically, we expect to identify the same eigenelements as true oscillations—with or without Procrustes rotation—but small differences are possible. To avoid the identification algorithm being in favor of
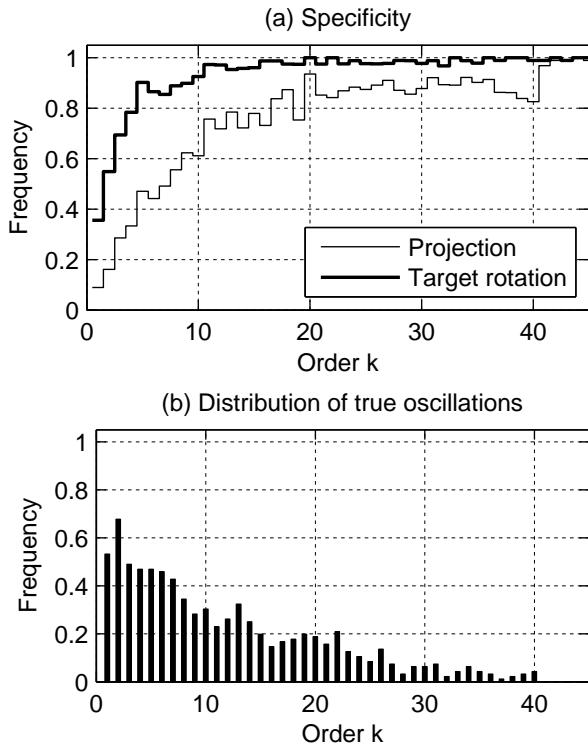
Fig. 3. Detection of oscillatory pairs in a cluster of harmonic oscillators with observational red noise. (a) Specificity when using the classical projection algorithm (light line) and the scaled target-rotation algorithm (heavy line), both based on ST-EOFs; and (b) distribution of true oscillations. In both panels, rank order k is plotted on the abscissa. A total of 100 noise-contaminated data sets have been tested; the parameters are the same as in Fig. 1.

one of the two significance tests, we keep only those noise-contaminated realizations that give the same identification results.

Once we have classified the eigenelements of the noise-contaminated data set into true and spurious oscillations, we proceed with the two significance tests as in the previous section, cf. Fig. 1, and determine the number of TPs and FPs, #TP and #FP, respectively, from an ensemble of several repetitions of the experiment.

b. Artificial variance compression

In the example of section 5 we have seen that the inclusion of eigenvalues, in addition to the EOFs, into the target rotation algorithm has reduced the number of FPs, in particular for higher-rank eigenvalues. For this near-singular case of the covariance matrix, i.e. $DM = N'$, the problem of artificial variance compression is exacerbated and M-SSA may yield, in the absence of suitable rotation, several FPs

among the higher-rank EOFs. At the same time, all true oscillations have been correctly identified; that is the sensitivity is maximal.

The experiment is repeated 100 times and the specificity is plotted as a function of the rank order k in Figure 3a. As expected, there is an enhanced rate of FPs, especially at higher-rank EOFs, and that, in turn, reduces the specificity in both cases. In the target-rotation algorithm, however, this undesirable effect has been largely reduced, and we show in the following subsections that the overall number of FPs does indeed not exceed the expected nominal level, according to the chosen significance level.

Figure 3b shows the distribution of true oscillations among the full spectrum of eigenvalues, as a fraction of the number of replicates of the test. As a consequence of the low signal-to-noise ratio, it turns out that the true oscillations are not necessarily attributed to the leading eigenvalues and that the eigenvalue rank order used originally by Broomhead and King (1986a,b) is, therefore, not a reliable method for separating signal from noise; this result provides additional motivation for using a Monte Carlo test against a red-noise process (cf. Allen and Smith 1996; Allen and Robertson 1996). To compare the success of the latter type of test with that of a simple rank-order test, we determine the number of TPs in the set of the eight largest eigenvalues as well. This criterion implies, of course, knowledge of the correct number of true oscillations, and it thus avoids estimating a break in the eigenvalue spectrum.

In the following, we first examine in greater detail the reliability of the two tests based on the data ST-EOFs—with and without scaling of the ST-EOFs—in different experimental settings. In particular, we consider the effect of modifying the number D of observed channels, the observation length N, and the window width M. Furthermore, we examine the effects of data compressions into PC. Finally, we compare the tests based on ST-EOFs with that based on T-EOFs.

c. Number D of observed channels

We first analyze the influence of the observed number of channels. In our experiment of a cluster of oscillators with uncorrelated observational red noise, we expect to improve the detection rate of shared oscillations as the number of channels—and hence the amount of information—increases, while at the same time the signal-to-noise ratio is enhanced.

Figure 4 shows #TP and #FP, i.e. the average number of TPs and FPs, as a function of the number D of channels. An increase in D does indeed help the extraction of the signal from the noise, and
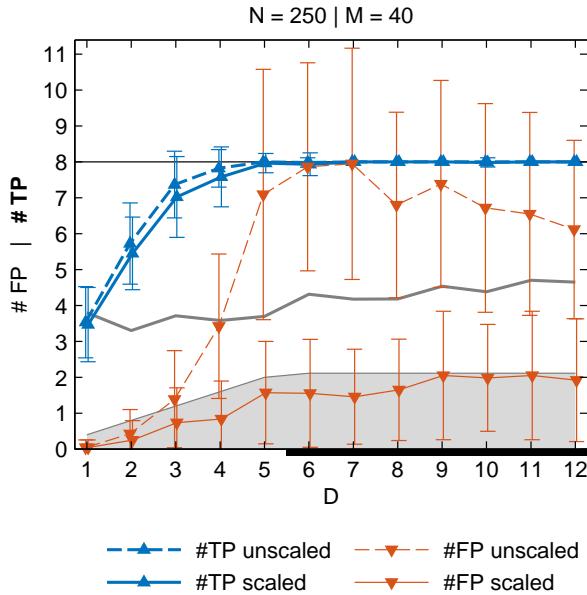
FIG. 4. Detection of oscillations in a cluster of harmonic oscillators with observational red noise. The average number of true positives (#TP) and false positives (#FP) is plotted, along with the corresponding standard deviation (error bars), based on an ensemble of 50 repetitions of the experiment, as the number D of observed channels increases. The parameters M and N are the same as in Fig. 1. The dashed lines correspond to the unscaled target-rotation algorithm onto ST-EOFs and the solid lines to the scaled one. The former includes the projection algorithm as a special case for $DM \leq N'$. A black bar along the horizontal axis indicates the D-interval within which the covariance matrix **C** becomes rank-deficient, $DM > N'$. In addition, the average number of TPs within the set of the eight largest eigenvalues (heavy gray line), and the expected number of FPs according to the significance level (gray shaded area) are shown.

the number of TPs converges toward its maximum value, which is $\max\{\#TP\} = 8$ in case of the four fundamental frequencies $T = \{7.6, 5.0, 2.7, 2.3\}$.

The convergence occurs already for $D = 5$, while in single-channel SSA, i.e. for $D = 1$ in the figure, only half of the oscillations have been identified. This result clearly illustrates the fact that M-SSA improves upon single-channel SSA by taking additional spatial information into account.

This improved detection rate, however, cannot be merely attributed to a concentration of oscillatory behavior in the largest eigenvalues alone. It turns out that no more than half of the eight largest eigenvalues are TPs (heavy gray line in Fig. 4).

While the sensitivity of the test does grow with D, it is indispensable to keep the specificity high as well. Figure 4 makes it clear that this is not the case for the projection algorithm, i.e. when $DM \leq N'$, in which the number of FPs (red dashed line) increases as well.

In particular, #FP greatly exceeds the number of expected FPs (gray shaded area), which equals the number of eigenvalues multiplied by the significance level. This excess of type-I error becomes more dramatic as one approaches the point at which the covariance matrix **C** reaches rank deficiency, $DM \lesssim N'$. For $DM > N'$, i.e. the unscaled target-rotation algorithm, the number of FPs remains likewise high, and reaches nearly the number of TPs. This renders the test useless, since only half of the significant eigenvalues can be attributed to true oscillations.

It is only the inclusion of eigenvalue information into the scaled target-rotation algorithm that finally helps control this type-I error, with the average number of FPs (solid red line in Fig. 4) now below the expected level of FPs over the entire range of D-values. We note that this algorithm has a slight tendency towards a more conservative behavior—a property that Procrustes methods have been often criticized for (e.g., Paunonen 1997)—but the detection rate remains comparable to that of the unscaled target-rotation algorithm. In particular, the scaled target-rotation algorithm would be the preferable choice with respect to the test's explanatory power.

To improve the detection of weak signals in the case of single-channel SSA, Paluš and Novotná (2004) proposed a test on the regularity of the oscillatory modes rather than on their variance. Their significance test has a demonstrably enhanced sensitivity, but it is not clear whether it remains sufficiently specific as well; see, for instance Figs. 2 and 4 in Paluš and Novotná (2004), with further "noise" EOFs becoming significant above the upper significance levels in the test on regularity. On the other hand, the frequency-pairing algorithm in their test is much less susceptible to the problem of artificial variance compression than the projection approach in Eq. (6); see again their Fig. 2, with the number of significant "noise" EOFs below the lower significance levels becoming largely reduced.

d. Length N of the observations

In the previous subsection we have seen that for the unscaled target-rotation algorithm, the specificity strongly depends on the ratio of the embedding dimension DM to the time series length N. It is the singular character of the covariance matrix **C** that enhances the artificial variance compression, and thus gives a high number of FPs in this algorithm. We examine next whether this undesirable property can be avoided when the observation length N considerably exceeds $(D + 1)M$ for fixed M.

For a fixed number D of channels, we have varied the observation length N and rerun the analysis as
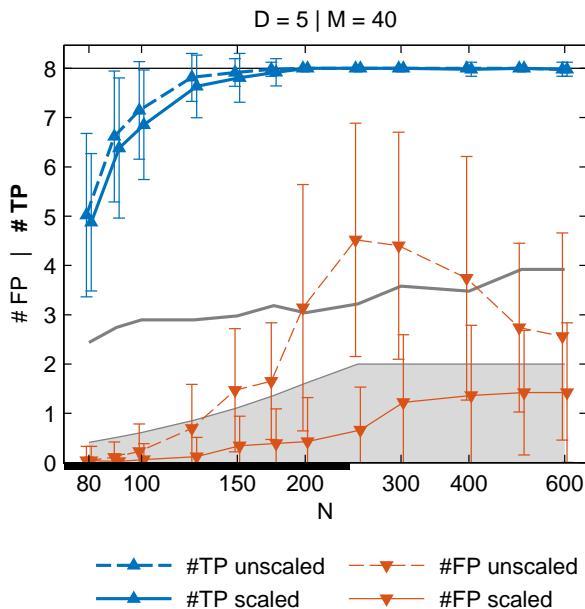
FIG. 5. Same as Fig. 4 but as a function of the length N of the observed time series.

FIG. 6. Same as Fig. 4 but as a function of the window length M.

before. Figure 5 shows the average number of TPs and FPs for both the unscaled and scaled target-rotation algorithms.

On the one hand, we see that the sensitivity is enhanced as the observation length N increases and that #TP tends towards the maximum value for sufficiently large N. It is only for very short observations that it drops to much lower values.

On the other hand, #FP in the unscaled target-rotation algorithm remains much higher than expected, even for large N. The scaled target-rotation algorithm, though, helps control the type I-error over the whole range of N-values in the figure. In particular, the latter algorithm remains superior to the projection algorithm even in full-rank case of very large N. This comparison demonstrates that the extent to which the artificial variance compression influences the hypothesis test is difficult to predict, as already indicated by Allen and Robertson (1996).

e. Window length M

The length N of the observations and the number D of channels is usually specified by the experimental setting, but the window length M is a flexible parameter to be judiciously chosen by the data analyst. In its choice, we are usually confronted with the general trade-off between a high spectral resolution, on the one hand, and a high temporal resolution, on the other; the decision may depend, therewith, on the specific problem. Increasing the window length
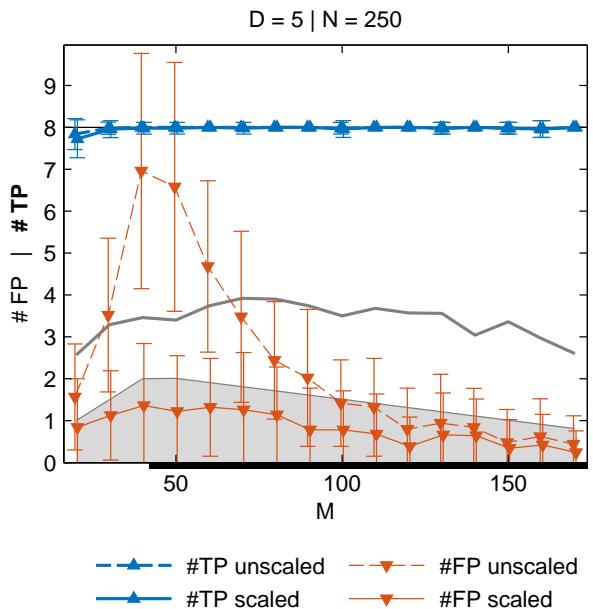
increases the number of RCs as well, and M-SSA will provide then a more detailed spectral decomposition of the data set, while incurring the risk of an excessive number of FPs.

Figure 6 shows the results of analyzing the same data set as in the previous subsections for various values of M. First of all, the figure shows a consistently high detection rate for both methods over a large range of M-values, with a slight decrease at small window length only. Note that a minimal window size of $M \simeq 16$ is necessary to separate the four oscillators even in the noise-free reference case.

At the same time, we observe that the scaled target-rotation algorithm controls well type-I errors over the entire range of M-values in the figure. The projection algorithm for full-rank covariance matrices, $M \lesssim 42$, runs increasingly the risk of an excessive number of FPs as these matrices get closer to being singular, and the risk remains high in the unscaled target-rotation algorithm for rank-deficient covariance matrices, $M \geq 42$. Only at much larger values of M does the number of FPs diminish to the nominal level, a feature that strongly limits the choice of M.

f. The effects of data compression

In the M-SSA analysis of high-dimensional data, it is common practice to perform first a conventional PCA analysis and to retain only a subset of leading

PCs for the subsequent M-SSA analysis. This pre-processing is meant to reduce the number of input channels and the computational cost, while retaining a large fraction of the total variance. Usually a small number L of channels is kept, no matter how large the data set (Dettinger et al. 1995; Allen and Robertson 1996; Robertson 1996; Ghil et al. 2002b); see also (Moron et al. 1998, Table 1). Since the resulting PCs are pairwise uncorrelated at zero lag, the M-SSA results can be simply tested against independent AR(1) processes, cf. section 3 herein.

Even though the transformation to conventional PCs turns out to be a helpful preprocessing step in the M-SSA analysis, its implications for the properties of the subsequent signal detection are rather complex. Given that the signal of interest involves typically only a small fraction of the total variance, we would expect it to show up only among the spatial EOFs (S-EOFs) with relatively small variance, while the leading S-EOFs might capture other large-scale effects. In this respect, the prior transformation to PCs could interfere with the detection of weak signals.

To study the implications of this type of preprocessing, we increase the number of observed channels in the previous example of a cluster of harmonic oscillators to $D = 250$ and reduce at the same time the observation length to $N = 130$. The parameter γ for the D superimposed AR(1) processes is randomly drawn from the interval $\gamma \in [0, 0.95]$, but we introduce at the same time correlations between all of them. That is, instead of adding independent noise realizations to each of the observed channels, we set the lag-0 covariance matrix $\mathbf{W}$ of the noise part to fit a Toeplitz structure, with $\mathbf{W}_{ij} = \kappa^{|i-j|}$ and $\kappa = 0.5$.

These correlations between channels are meant to simulate the effect of spatial correlations in a randomly perturbed spatio-temporal process and, given our choice of κ, we expect the noise part to dominate the prior PCA analysis.

Figure 7 illustrates the implications of a transformation of the input channels to conventional PCs. The variance that is captured by the L leading S-EOFs increases monotonically, as expected, as the number L of S-EOFs increases. Usually, one selects a good trade-off between a low number of components and a high fraction of the variance they capture, before proceeding to the M-SSA analysis.

Although the S-EOFs provide an efficient representation of the spatial aspects of the signal's variance, potentially important information about a possibly weak signal in the time domain may be missed. To illustrate this limitation, we project the reference signal—that is the cluster of harmonic oscillators wit-
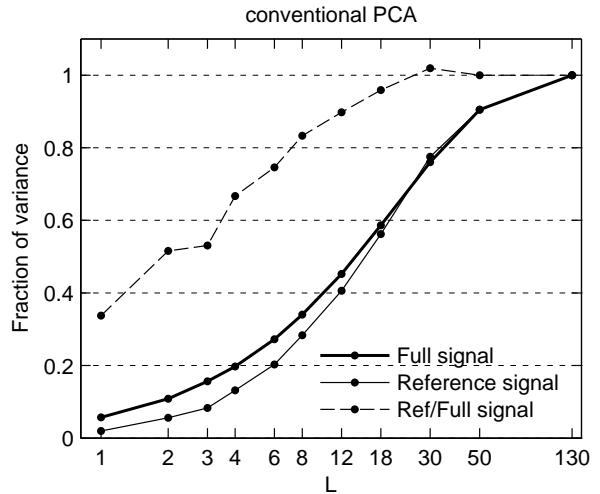

FIG. 7. Fraction of the total variance (heavy solid line) that is captured by the L leading S-EOFs of a conventional PCA analysis. The statistical experiment is similar to that in Fig. 4; it consists in a cluster of harmonic oscillators with spatial correlated observational red noise, but the number $D = 250$ of observed channels is much larger and the observation length is $N = 130$. Furthermore is shown the variance of the noise-free oscillators alone that is captured by the L leading S-EOFs (light solid line), and the variance ratio between the noise-free and the full signal (dashed line). All values are averaged over 50 realizations of the experiment.

hout noise—onto the same S-EOFs, and determine its variance as well (light solid line in the figure).

The variance of the reference signal increases monotonically with L, like that of the full signal, but the variance ratio between the reference and the full signal (dashed line) decreases markedly as L is reduced to values typically used in the type of preprocessing discussed herein; as a consequence, when $L \lesssim 4$, the signal-to-noise ratio is nearly half of its value at $L = N = 130$.

These effects have been further emphasized by the fact that the oscillations in our statistical experiment have random initial phases and that the PCA analysis is not taking this phase information into account. A similar negative effect is likewise possible in the detection of traveling oscillatory patterns, e.g. when having to rely on regionally averaged data, such as the SST data in section 7.

After the above preprocessing and retention of L leading conventional PCs, we next evaluate the detection rate of the subsequent M-SSA analysis. As before, we determine a set of reference ST-EOFs that correspond to oscillatory modes in order to identify true oscillatory behavior in the full signal's ST-EOFs. To account for the projection of the full signal onto S-EOFs, we project the noise-free reference signal
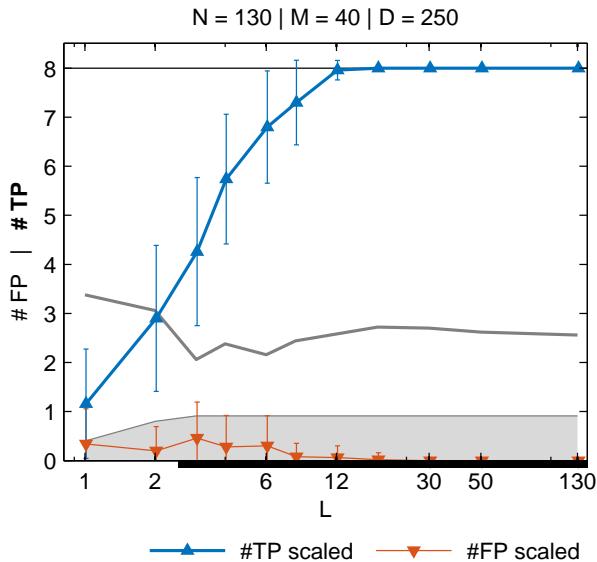
N = 130 | M = 40 | D = 250

Fig. 8. Same as Fig. 4 but as a function of the number L of retained PCs from a conventional PCA analysis in Fig. 7. Note that we only show the results for the scaled target rotation.

onto the same subspace as well and derive reference ST-EOFs from the M-SSA analysis of the corresponding PCs. Here we focus only on the scaled target-rotation algorithm and determine TPs and FPs in an ensemble of 50 repetitions of the experiment.

Figure 8 is similar to Figs. 4–6 and it shows the number of TPs and FPs as a function of L. It turns out that the detection rate, #TP (heavy solid line), is best for large L, and that it does reach its optimal value in this limit. On the other hand, as the data set is compressed into a decreasing number of PCs, the detection rate drops very markedly. This marked drop is comparable to that of Fig. 4, where D, the number of observed channels, affects the detection rate in a similar way.

In conclusion, we have seen that a compression of the data set into a few leading PCs can strongly influence the capability of M-SSA to extract weak but significant signals. In particular, in the presence of other high-variance components, such preprocessing may reduce substantially the signal-to-noise ratio.

When the number of channels D exceeds the length N of the data set, $D > N$, we rely here on PCA as a transformation of the data set into a set of $L = N$ pairwise uncorrelated PCs; see, for instance, Fig. A1 in Ghil et al. (2002b). Doing so preserves the total variance, but helps reduce computational costs.

Although it would be theoretically possible to perform M-SSA on the full set of D channels via the complementary eigendecomposition of the reduced covariance matrix (cf. section 2), it is computati-

onally more efficient and numerically stabler to perform the subsequent varimax rotation on ST-EOFs of length LM rather than DM, given that often $L = N \ll D$. To which extent, however, the superimposed spatial structure of the S-EOFs will affect the spatio-temporal structure of the varimax-rotated ST-EOFs so obtained is left for future studies.

g. Significance test on T-EOFs

So far, we have only focused on the two significance tests that are based on the unscaled and scaled target-rotation of ST-EOFs, respectively, in the two cases of a full-rank and a rank-deficient covariance matrix. In the latter case, we will compare those results now with that of the two tests that are based on the unscaled and scaled target-rotation of T-EOFs; cf. section 4.

Since no cross-channel covariance information is taken into account in the two tests on T-EOFs, we first transform the input channels into spatial PCs. As already discussed, this helps eliminate cross correlations at lag zero, but correlations at other time lags may remain and influence the tests on T-EOFs. To compare its reliability with that of the two tests on ST-EOFs, we consider correlated noise, as in the previous subsection, with the noise coupling strength set to $\kappa \geq 0$.

Figure 9 shows the number of TPs and FPs as a function of κ, for the two unscaled target rotation algorithms onto ST-EOFs and T-EOFs. Note that the latter case equals that of a projection onto T-EOFs as proposed by Allen and Robertson (1996), cf. Eq. (14) herein. The parameters N, M, and D are chosen to emphasize the severity of the problem of artificial variance compression: e.g., by taking $DM \lesssim N'$.

For uncorrelated noise, $\kappa = 0$, both tests are likewise susceptible to this problem and show an enhanced number of FPs. As the noise coupling strength increases, $\kappa > 0$, the FP rate of the test on ST-EOFs starts to decrease and reaches its nominal level at sufficiently strong coupling, $\kappa \gtrsim 0.4$. The FP rate of the test on T-EOFs, however, starts to increase even further as the coupling strength increases, and the test completely fails to distinguish true oscillations from spurious ones, i.e. the number of FPs largely exceeds that of the TPs, which renders the test meaningless.

The scaled target-rotation algorithm, on the other hand, clearly helps reduce the number of FPs not only in the test on ST-EOFs, but also in the test on T-EOFs, as shown in Fig. 10. When the coupling is weak, $\kappa \lesssim 0.1$, both tests show an FP rate that is close to the nominal level now. Still, only the test
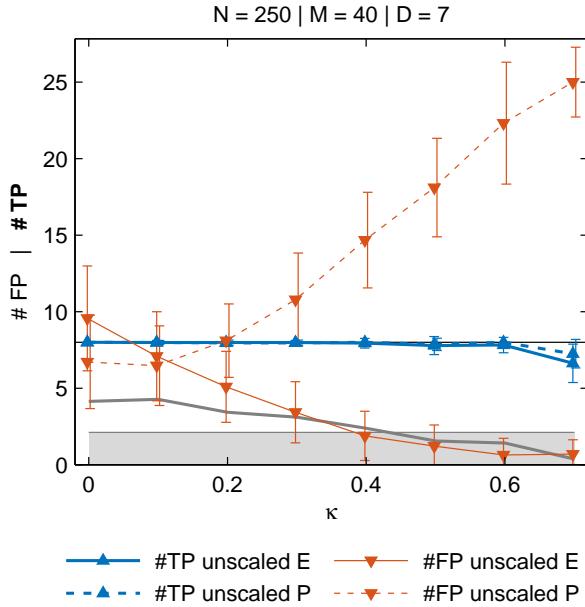
FIG. 9. Same as Fig. 4 but as a function of the coupling strength κ between the individual AR(1) processes. The solid lines correspond now to the unscaled target rotation onto ST-EOFs and the dashed lines to projecting onto T-EOFs, cf. Eq. (14).
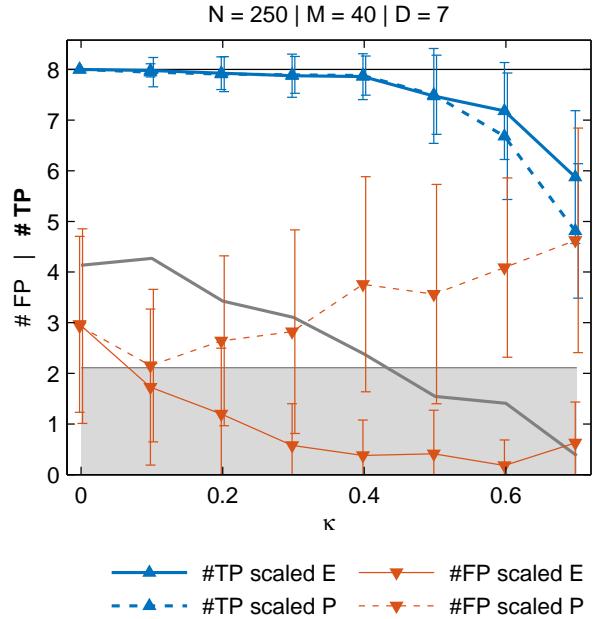
FIG. 10. Same as Fig. 9 but for the scaled target rotation. The solid lines correspond to the rotation onto ST-EOFs, cf. Eq. (11), and the dashed lines to the rotation onto T-EOFs, cf. Eq. (15).

on ST-EOFs is able to keep its FP rate below the nominal level as the coupling strength increases even further, $\kappa \gtrsim 0.2$, while the test on T-EOFs once again looses its discriminant power as κ increases. The TP rate at weak coupling is likewise high in both tests, although it does drop faster in the test on T-EOFs as the noise coupling increases.

In conclusion, the null hypothesis test that is based on the scaled target-rotation algorithm of ST-EOFs provides the most reliable results also for the case of a rank-deficient covariance matrix, $DM > N'$, and clearly outperforms the two tests that are based on T-EOFs in the presence of cross correlations. Since no cross-channel covariance information is retained in T-EOFs, cross correlations may reduce the discriminant power of the corresponding tests, even though the input channels have been transformed into spatial PC.

7. An application to North Atlantic data

The SODA reanalysis data set (Giese and Ray 2011, version 2.2.4) provides monthly SST fields over the 138-yr interval 1871–2008. The SST, following Feliks et al. (2011), is taken equal to the temperature in the upper 5 m of the ocean.

The analysis here is for the Gulf Stream region (30°–50°N, 76°–35°W) that includes the Cape Hat-

teras and the Grand Banks sub-regions. Feliks et al. (2011) identified in either one or both of these two regions, interannual spectral peaks of 8.5 yr, 4.2 yr and 2.8 yr. As discussed in the Introduction, these peaks are similar to those found in the NAO index (Gámiz-Fortis et al. 2002; Paluš and Novotná 2004; Feliks et al. 2013); hence the possibility of shared mechanisms between the ocean and atmosphere in the North Atlantic basin is worth examining further.

We include, therefore, in our analysis atmospheric SLP data that covers the North Atlantic region (25°–80°N, 80°W–33°E), taken from the 20CRv2 reanalysis project (Compo et al. 2011) in the same 138-yr interval. The SST and SLP data fields are first converted into anomalies, i.e. we remove at each grid point the average value over the 138-yr interval. To account for geographical variations in the grid-size resolution, we further multiply each grid point by the cosine of its latitude.

In the present analysis, we focus on interannual variability only and subsample the data in time with an annual sampling rate. The latter step implies a low-pass filtering of the monthly data with a Chebyshev type-I filter from which we take all the July values; see Feliks et al. (2013) for details.

To combine the annual SST and SLP anomalies into a joint M-SSA analysis, we further normalize each field to unit variance and concatenate all chan-
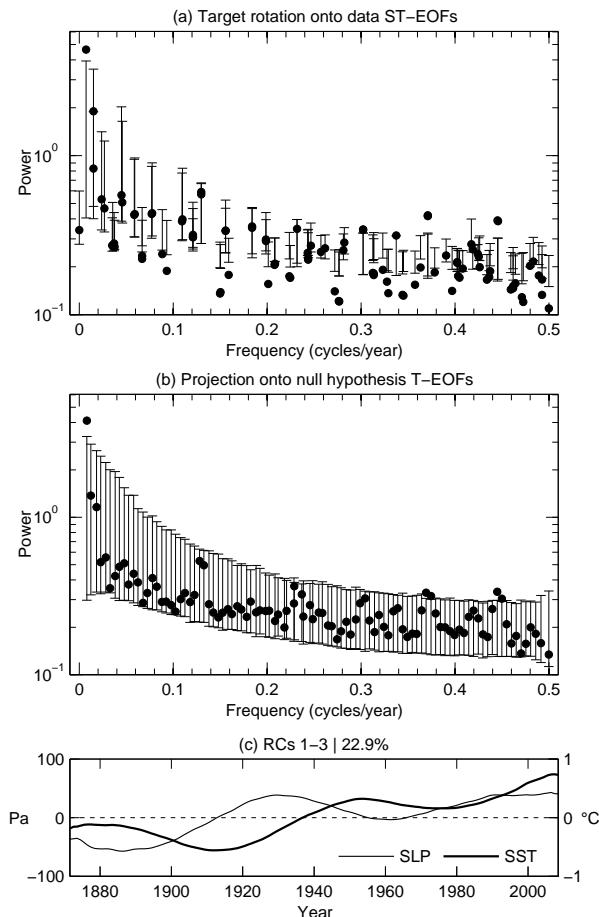
FIG. 11. Spectral properties of North Atlantic SST and SLP anomaly fields from a joint M-SSA analysis for the interval 1871–2008, with a window length of $M = 40$ yr; the subsequent varimax rotation uses ST-EOFs 4–30, while ST-EOFs 1–3 are excluded from the rotation since they are distinct from the others and not degenerate. (a) The estimated variance in each mode is shown as black dots, plotted as a function of their corresponding frequency. Lower and upper ticks on the error bars correspond to the 1% and 99% quantiles of a Monte Carlo test that uses the scaled target-rotation algorithm of ST-EOFs; the test ensemble has 2000 members. (b) Significance test on the T-EOFs of the reduced covariance matrix of the null hypothesis, using the same set of surrogate data; note that $LM \gg N'$. (c) Low-frequency variability of the spatially averaged SST and SLP anomalies from a reconstruction with ST-EOFs 1–3.

nels into a single large trajectory matrix. The resolution in time and space gives a total of $D = 24\,173$ channels of length $N = 138$ yr, from which we finally derive $L = 138$ PCs as input channels for our M-SSA analysis, as described in section 6f.

Figure 11a shows the M-SSA results—together with the significance level from a test against a null hypothesis of pure noise, as derived from the scaled target-rotation algorithm of ST-EOFs. It turns out

that five eigenvalues exceed the significance level of 99% in the interannual frequency band: the largest eigenvalue, which can be attributed to a trend component, and two oscillatory pairs at period lengths of 2.7 yr and 2.2 yr, respectively.

A significance test that is based on the T-EOFs of the null-hypothesis covariance matrix confirms these findings, cf. Fig. 11b. The probability to observe five or more excursions above the 99% quantile, as given by the binomial distribution, is approximately 1.3%. This probability means that, even without any prior knowledge of the underlying dynamics and the frequencies of interest, we can still reject the null hypothesis at a significance level that is only marginally lower than 99%.

It appears that the low-frequency variance in EOFs 1–3 dominates the entire spectrum and that the significance levels for the adjacent frequencies are likely, therefore, to be overestimated; i.e., EOFs 6-7 at a frequency of 0.05 year^{-1} appear below the 15% quantile. Removing the leading components from the significance test could thus help reduce the bias in the AR(1) parameter estimation toward the strong trend components. In a composite test, we hence exclude not only EOFs 1–3 from the parameter estimation of the AR(1) null hypothesis, but also the two oscillatory pairs that we found to be significant at the periods 2.7 yr and 2.2 yr, respectively.

Figure 12 shows the updated significance levels against a composite null hypothesis. In particular, at frequencies of $\lesssim 0.2$ cycles/yr, the AR(1) null hypothesis gives a considerably better fit to the remainder on the ST-EOFs. It turns out that we obtain an additional 7.7-yr mode in EOFs 4-5, which becomes significant in both tests now, apart from the seven EOFs that we had excluded from the significance test (target dots in the figure). This result is not surprising, as the probability to observe two or more such random excursion out of the remaining 129 EOFs is very likely, i.e. 37% at a 99% significance level, according to a binomial distribution; see also AS and the appendix therein. Without further evidence, the information from the two significance tests is therefore not sufficient to draw conclusions about the presence of such a 7.7-yr mode.

In the present case, however, we have prior reason to focus on a 7–8-yr frequency band. Feliks et al. (2011) identified significant oscillatory modes with periods of 8.5 yr and 10.5 yr in the SST field of the Cape Hatteras and Grand Banks region, respectively, and have shown that these modes spinup in an atmospheric model and become synchronized with a simulated NAO index.

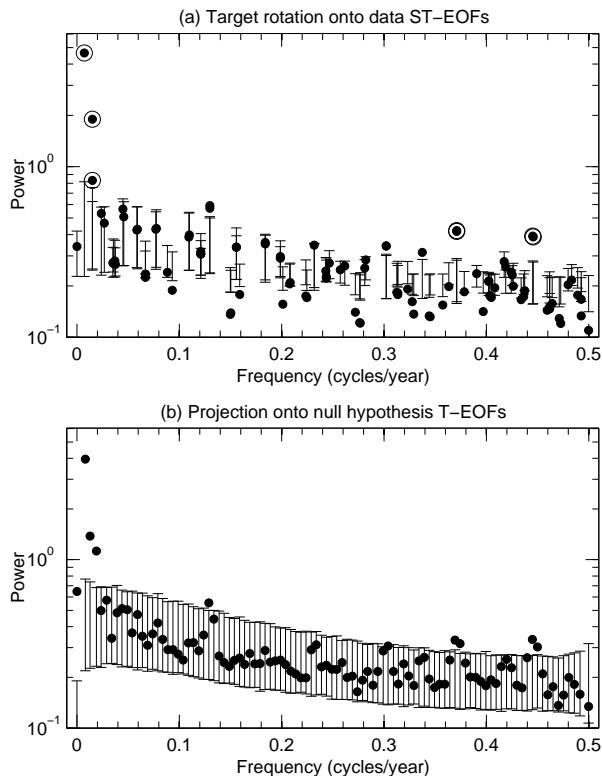These authors have also shown that the spatio-temporal pattern of the 8.5-yr mode shares certain

FIG. 12. Same as Fig. 11 but with the significance level derived from a test against a composite null hypothesis of seven ST-EOFs plus AR(1) noise. The seven ST-EOFs (target dots) correspond to the trend, ST-EOFs 1–3, and two oscillatory pairs at a period of 2.7 yr and 2.2 yr, respectively.

features with the so-called gyre mode (Jiang et al. 1995) that has a dominant 7–8-yr peak across a hierarchy of ocean models (Speich et al. 1995; Chang et al. 2001; Ghil et al. 2002a; Dijkstra and Ghil 2005). It is the modeling evidence in Feliks et al. (2011) combined with that obtained due to the markedly improved discriminant power of the present significance test that provides the requisite, stronger evidence for the existence of such a joint mode in the analyzed SST and SLP data.

Relatively small discrepancies between the earlier frequency results of Feliks et al. (2011) and the present ones might be due to the shorter duration of the SODA reanalysis used there, which was of only 50 yr in Carton and Giese (2008, version 2.0.2-4), as well as the absence of a subsequent varimax rotation toward unimodal ST-EOFs in the previous analysis. We have further analyzed the SST and SLP anomalies separately and the results support the existence of similar 7.7-yr, 2.7-yr, and 2.2-yr modes common to both of the fields (not shown).

Figure 13 shows the reconstruction of the SST and SLP anomaly fields from the three oscillatory modes found to be highly significant. In the SST anomalies (upper panels), we observe in all three modes a concentration of small areas of high variance and of alternating sign along the Gulf Stream front. This spatial structure yields a weaker vs. a stronger meandering of the eastward jet in the opposite phases of each mode, while the resulting deflection from the mean Gulf Stream position is largest in the 7.7-yr mode.

In the SLP anomalies (lower panels), we observe a clear dipole structure in all three modes. In both the 7.7-yr and 2.2-yr mode, this dipole has a meridional orientation, with the two extrema that appear near the Iceland low and the Azores high, respectively. The phase alternation in these two modes thus leads to a weaker vs. a stronger meridional SLP gradient. In the 2.7-yr mode, the dipole structure is tilted and it is oriented southwest-to-northeast. Hence, this oscillatory mode contributes to a a tilt of the total SLP pattern in that direction. In all three oscillatory modes, we are thus led to the conclusion that the variability in the Gulf Stream region is probably linked to variability in the NAO.

Finally, to compare our analysis of the full SST field in the Gulf Stream region with that of a simple univariate analysis of regionally averaged indicators, we have further analyzed the mean SST field, as well as the leading PC of the same region, in a single-channel SSA analysis. It turns out that in both cases, apart from a low-frequency component of high variance, no further oscillatory modes are found to be significant at the 99% level.

In the single-channel SSA analysis of the mean SLP anomalies, as well as the leading PC of the North Atlantic region, oscillatory modes similar to the three modes discussed are found to be significant, but only at a lower, 97.5% level. The single-channel SSA results thus confirm our findings of possibly negative effects of data compression on the detection of weak signals; these results clearly demonstrate, therewith, the advantages of a full multichannel spectral analysis vs. that of a simple scalar indicator.

8. Summary

In numerous applications, multichannel singular spectrum analysis (M-SSA) has proven an efficient tool for the identification of regular behavior in high-dimensional data (Ghil et al. 2002b, and references therein). Since M-SSA, like single-channel SSA (Vautard and Ghil 1989), can generate oscillatory-looking patterns from pure noise, Monte-Carlo type
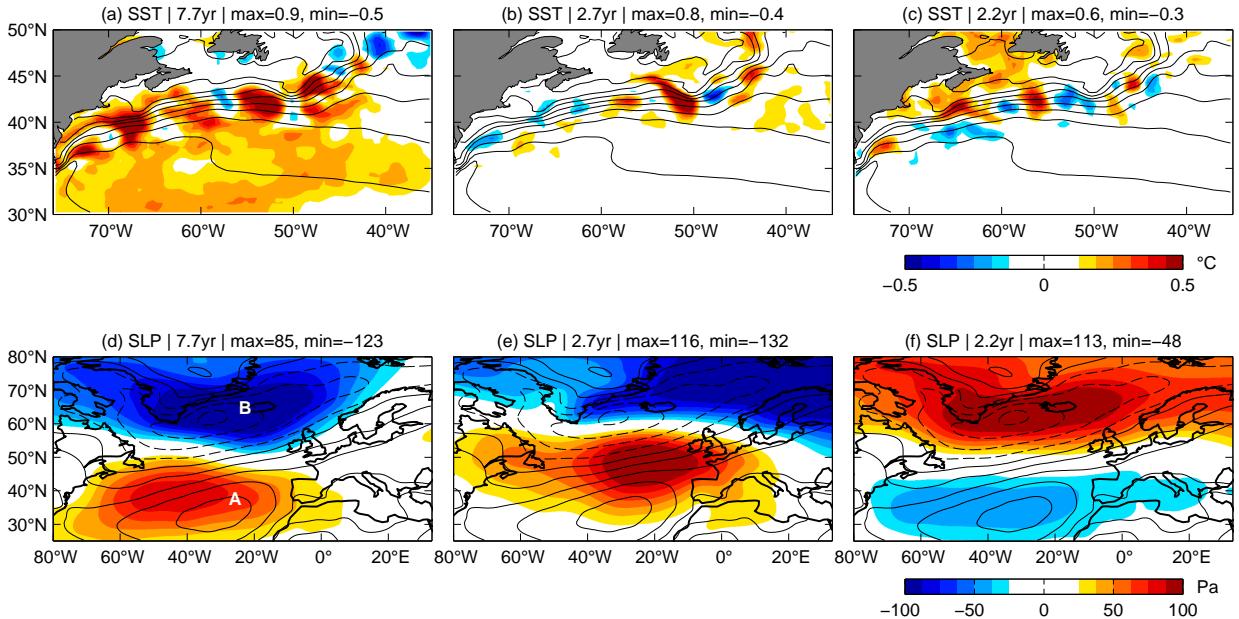
FIG. 13. Reconstruction of SST and SLP anomalies from the oscillatory modes with 7.7-yr, 2.7-yr, and 2.2-yr periods. In each mode, the leading spatial EOF from a classical PCA analysis is shown in color and transformed back into physical units. (a–c) SST field reconstruction in the Gulf Stream region, with contour lines that correspond to the mean SST field; contour interval is 2°C. (d–f) SLP field reconstruction in the North Atlantic region, with contour lines that correspond to the mean SLP field, i.e. solid and dashed lines are positive and negative differences from 1013.25 hPa, respectively; contour interval is 200 Pa. Extreme values (max and min) are given in the legend of each panel. Meteorological stations located in the Azores and Iceland are typically used to define an NAO index, and they are indicated as points A and B, respectively, in panel (d).

tests have been developed to provide objective criteria for its significance (Allen and Smith 1996; Allen and Robertson 1996). In the present paper, we have proposed several ways of improving such tests and studied their performance as a function of various parameters, such as the number D of observed channels, the length N of the time series, and the window parameter M.

We have shown that straightforward Monte Carlo tests for M-SSA are more likely to fail as the embedding dimension DM reaches the length N of the observed time series. We introduced here Procrustes target rotation into the M-SSA setting and showed that it markedly improves the discriminant power of Monte Carlo–type tests by reducing the risk of type-I errors, while maintaining their sensitivity.

Our M-SSA analysis relied on varimax-rotated ST-EOFs (Groth and Ghil 2011) and we have shown that, in particular, the scaled target-rotation algorithm of ST-EOFs provides a robust significance test for both full-rank and rank-deficient covariance matrices. In the latter case, it clearly outperforms the test based on T-EOFs of a reduced covariance matrix, as proposed by Allen and Robertson (1996), especially in the presence of cross correlations.

We have further shown the limitations of preprocessing large data sets via data compression onto a few leading S-EOFs by means of a conventional PCA analysis in the M-SSA setting, i.e. when the goal is the detection of weak but significant signals in the space-time domain by M-SSA. Once a certain part of the time series has already been identified as signal, we have further proposed a generalization of the single-channel SSA composite test of Allen and Smith (1996) to M-SSA.

The evaluation of the methods was carried out at first in an idealized experiment using a cluster of harmonic oscillators with observational red noise. The perturbing noise is generated by the same class of AR(1) processes as the one used in the null hypothesis, hence there is no formally erroneous specification of the latter.

The end-to-end testing algorithm that results from these various comparisons is summarized in Appendix B. We applied this algorithm—along with the new varimax rotation methods introduced and tested herein—to the analysis of interannual variability in the North Atlantic basin. This analysis combined the SST field in the Gulf Stream region that includes Cape Hatteras and the Grand Banks with the SLP

field over the entire North Atlantic basin. Given the more refined spectral results of varimax-rotated ST-EOFs and the improved discriminant power of our modified Monte Carlo test, we have been able to provide even stronger evidence for shared mechanisms between the Gulf Stream region and the North Atlantic Oscillation in the interannual frequency band.

APPENDIX

A. *Varimax rotation of ST-EOFs*

The common idea of so-called simple-structure rotations is to find a rotation that simplifies the interpretation of the eigenvectors and that reduces mixture effects. There are several ways to quantify the simplicity of an eigenvector's structure (Richman 1986). Varimax rotation attempts to find an orthogonal rotation $\mathbf{E}^* = \mathbf{E}\,\mathbf{T}_V$ that maximizes the variance of the squared elements (Kaiser 1958).

In PCA with $M = 1$, the functional of the eigenvectors one wishes to maximize is

$$V_1 = \sum_{k=1}^{S} \left\{ \frac{1}{D} \sum_{d=1}^{D} \left(\frac{e_{dk}^{*\,2}}{h_d^{*\,2}} \right)^2 - \left(\frac{1}{D} \sum_{d=1}^{D} \frac{e_{dk}^{*\,2}}{h_d^{*\,2}} \right)^2 \right\}; \tag{A1}$$

here S is the number of rotated eigenvectors $\mathbf{e}_k^*$ and $h_d^{*\,2} = \sum_{k=1}^{S} e_{dk}^{*\,2}$ is a normalization factor. Kaiser (1958) has given an explicit equation for the sequential rotation of pairs of eigenvectors.

Since the criterion V_1 maximizes the variance over all dimensions, a direct application to M-SSA eigenvectors would not only achieve the desirable effect of increasing the spatial variance between channels, but also increase the temporal variance of the data-adaptive sine-cosine–like characteristics in the eigenvectors. The latter effect can lead to an undesirable loss of correctly captured oscillatory pairs (Plaut and Vautard 1994).

For this reason, Groth and Ghil (2011) have proposed a modification of varimax that maximizes only the variance of the spatial component. Prior to the calculation of the varimax criterion in each rotation step, the authors first sum over the temporal part,

$$\bar{e}_{dk}^{*\,2} = \sum_{m=1}^{M} e_{dk}^{*\,2}(m), \tag{A2}$$

to obtain a participation index $\bar{e}_{dk}^{*\,2}$ of channel d to the kth ST-EOF, and then try to maximize the variance in the participation index instead. Thus, the criterion becomes

$$V_M = \sum_{k=1}^{S} \left\{ \frac{1}{D} \sum_{d=1}^{D} \left(\frac{\bar{e}_{dk}^{*\,2}}{\bar{h}_d^{*\,2}} \right)^2 - \left(\frac{1}{D} \sum_{d=1}^{D} \frac{\bar{e}_{dk}^{*\,2}}{\bar{h}_d^{*\,2}} \right)^2 \right\}, \tag{A3}$$

with the normalization $\bar{h}_d^{*\,2} = \sum_{k=1}^{S} \bar{e}_{dk}^{*\,2}$.

In this way, the criteria V_1 and V_M, respectively, attempt to bring the spatial component of the eigenvectors either close to one or to zero, and that can help the interpretation of their structure. Groth and Ghil (2011) have further shown that the varimax algorithm's original simplicity—namely a maximization of V_1 by pairwise rotations of eigenvectors—is maintained in the maximization of V_M as well.

As proposed by Groth and Ghil (2011), we scale here too each eigenvector by its singular value prior to rotation, in order to stabilize the results over a large range of the number S of rotated eigenvectors, and to minimize the risk of an overrotation (O'Lenic and Livezey 1988). That is, we first derive an orthogonal rotation matrix $\mathbf{T}_V$ that maximizes the criterion V_M in $\mathbf{B} = \mathbf{E}\mathbf{\Sigma}\mathbf{T}_V$, with $\mathbf{\Sigma} = \mathbf{\Lambda}^{1/2}$ the singular values. This yields a non-orthogonal $\mathbf{B}$, and Groth and Ghil (2011) propose to rotate the eigenvectors instead, $\mathbf{E}^* = \mathbf{E}\mathbf{T}_V$. Finally, the eigenvalues are updated by the diagonal elements of a similarity transform, $\mathbf{\Lambda}^* = \mathbf{T}_V'\mathbf{\Lambda}\mathbf{T}_V$.

As shown hereafter, the rotation $\mathbf{E}^* = \mathbf{E}\mathbf{T}_V$ gives indeed the best approximation, in the sense that it minimizes the distance between $\mathbf{E}^*$ and $\mathbf{B}$: Suppose we want to find an orthogonal rotation matrix $\widetilde{\mathbf{T}}$ that brings $\mathbf{E}\widetilde{\mathbf{T}}$ closest to $\mathbf{B}$. A solution to this problem is given by the orthogonal Procrustes algorithm, with the details given in section 4. In particular, $\widetilde{\mathbf{T}} = \mathbf{U}\mathbf{V}'$ yields the singular value decomposition (SVD) of $\mathbf{E}'\mathbf{B} \equiv \mathbf{E}'\mathbf{E}\mathbf{\Sigma}\mathbf{T}_V = \mathbf{U}\mathbf{S}\mathbf{V}'$. A simple, element-by-element inspection gives $\mathbf{U} = \mathbf{I}$ as the identity matrix, $\mathbf{S} = \mathbf{\Sigma}$, and hence $\widetilde{\mathbf{T}} = \mathbf{T}_V$.

B. *Summary of Monte-Carlo SSA algorithms*

Table 1 summarizes the different versions of the Monte-Carlo SSA algorithm that have been discussed in the present paper. This includes the origi-

TABLE 1. Original Monte-Carlo SSA algorithms of Allen and Smith (1996) and Allen and Robertson (1996) for full-rank and rank-deficient covariance matrices, respectively, in comparison with the proposed Procrustes target rotation algorithms.

$\mathbf{X} = \eta^{1/2}\mathbf{P\Sigma E'}$, $\mathbf{\Lambda} = \mathbf{\Sigma}^2$	full-rank	rank-deficient
$\mathbf{X}_R = \eta^{1/2}\mathbf{P}_R\mathbf{\Sigma}_R\mathbf{E}'_R$, $\mathbf{\Lambda}_R = \mathbf{\Sigma}^2_R$	$DM \leq N - M + 1 \quad (\eta = N - M + 1)$	$DM > N - M + 1 \quad (\eta = DM)$
1. Original Monte-Carlo SSA algorithms	$\begin{aligned}\mathbf{\Lambda}_R^{(E)} &= \eta^{-1}\mathbf{E'X}'_R\mathbf{X}_R\mathbf{E} \\ &= \mathbf{E'E}_R\mathbf{\Lambda}_R\mathbf{E}'_R\mathbf{E} \\ &= \mathbf{T'\Lambda}_R\mathbf{T} \\ \mathbf{T} &= \mathbf{E}'_R\mathbf{E}\end{aligned}$ Allen and Smith (1996)	$\begin{aligned}\mathbf{\Lambda}_R^{(P)} &= \eta^{-1}\mathbf{P'X}_R\mathbf{X}'_R\mathbf{P} \\ &= \mathbf{P'P}_R\mathbf{\Lambda}_R\mathbf{P}'_R\mathbf{P} \\ &= \mathbf{T'\Lambda}_R\mathbf{T} \\ \mathbf{T} &= \mathbf{P}'_R\mathbf{P}\end{aligned}$ Allen and Robertson (1996)
2. Extension via unscaled target rotation	$\begin{aligned}\mathbf{P}'_R\mathbf{P} &= \mathbf{USV'} \\ \mathbf{T} &= \mathbf{UV'} \\ \mathbf{\Lambda}_R^{(P)} &= \mathbf{T'\Lambda}_R\mathbf{T}\end{aligned}$	$\begin{aligned}\mathbf{E}'_R\mathbf{E} &= \mathbf{USV'} \\ \mathbf{T} &= \mathbf{UV'} \\ \mathbf{\Lambda}_R^{(E)} &= \mathbf{T'\Lambda}_R\mathbf{T}\end{aligned}$
3. Scaled target rotation of ST-EOFs	$\begin{aligned}(\mathbf{E}_R\mathbf{\Sigma}_R)'\mathbf{E\Sigma} &= \mathbf{USV'} \\ \mathbf{T}_E &= \mathbf{UV'} \\ \mathbf{\Lambda}_R^{(E\Sigma)} &= \mathbf{T}'_E\mathbf{\Lambda}_R\mathbf{T}_E\end{aligned}$	
4. Scaled target rotation of T-EOFs	$\begin{aligned}(\mathbf{P}_R\mathbf{\Sigma}_R)'\mathbf{P\Sigma} &= \mathbf{USV'} \\ \mathbf{T}_P &= \mathbf{UV'} \\ \mathbf{\Lambda}_R^{(P\Sigma)} &= \mathbf{T}'_P\mathbf{\Lambda}_R\mathbf{T}_P\end{aligned}$	

nal algorithms of Allen and Smith (1996) for single-channel SSA, and of Allen and Robertson (1996) for M-SSA (first row of the table), as well as their generalization to covariance matrices of arbitrary rank via the unscaled target-rotation algorithm (second row). The first column of the table deals with the case of a full-rank covariance matrix, while the second column presents the rank-deficient case.

All the algorithms in the upper half of the table (rows one and two) are based on the structure of the EOFs alone, i.e. they disregard completely the eigenvalue spectrum. To improve the comparison of the data eigendecomposition with that of the surrogate data, the scaled target-rotation algorithm is included in the lower half of the table (third and fourth row).

Note that all the steps in each of these algorithms remain exactly the same in the case of varimax-rotated data eigenelements: that is, the target eigenelements $\mathbf{E}$, $\mathbf{\Sigma}$, and $\mathbf{P}$ are just replaced by the rotated ones $\mathbf{E}^*$, $\mathbf{\Sigma}^*$, and $\mathbf{P}^*$, respectively, in all of the equations in Table 1.

References

Allen, M. R., and A. W. Robertson, 1996: Distinguishing modulated oscillations from coloured noise in multivariate datasets. *Climate Dyn.*, **12 (11)**, 775–784, doi:10.1007/s003820050142.

Allen, M. R., and L. A. Smith, 1996: Monte Carlo SSA: Detecting irregular oscillations in the presence of colored noise. *J. Climate*, **9**, 3373–3404, doi:10.1175/1520-0442(1996)009⟨3373:MCSDIO⟩2.0.CO;2.

Broomhead, D. S., and G. P. King, 1986a: Extracting qualitative dynamics from experimental data. *Physica D*, **20 (2-3)**, 217–236, doi:10.1016/0167-2789(86)90031-X.

Broomhead, D. S., and G. P. King, 1986b: On the qualitative analysis of experimental dynamical systems. *Nonlinear Phenomena and Chaos*, S. Sarkar, Ed., Adam Hilger, Bristol, England, 113–144.

Carton, J. A., and B. S. Giese, 2008: A reanalysis of ocean climate using Simple Ocean Data Assimilation (SODA). *Mon. Wea. Rev.*, **136 (8)**, 2999–3017, doi:10.1175/2007MWR1978.1.

Chang, K.-I., M. Ghil, K. Ide, and C.-C. A. Lai, 2001: Transition to aperiodic variability in a wind-driven double-gyre circulation model. *J. Phys. Oceanogr.*, **31 (5)**, 1260–1286, doi:10.1175/1520-0485(2001)031⟨1260:TTAVIA⟩2.0.CO;2.

Cliff, N., 1966: Orthogonal rotation to congruence. *Psychometrika*, **31 (1)**, 33–42, doi:10.1007/BF02289455.

Compo, G. P., and Coauthors, 2011: The twentieth century reanalysis project. *Q. J. R. Meteorol. Soc.*, **137 (654)**, 1–28, doi:10.1002/qj.776.

Dettinger, M. D., M. Ghil, and C. L. Keppenne, 1995: Interannual and interdecadal variability in United States surface-air temperatures, 1910–87. *Climatic Change*, **31 (1)**, 35–66, doi:10.1007/BF01092980.

Dijkstra, H. A., and M. Ghil, 2005: Low-frequency variability of the large-scale ocean circulation: A dynamical systems approach. *Rev. Geophys.*, **43 (3)**, doi:10.1029/2002RG000122/full.

Elsner, J., and A. A. Tsonis, 1994: Low-frequency oscillation. *Nature*, **372**, 507–508.

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

Feliks, Y., M. Ghil, and A. W. Robertson, 2011: The atmospheric circulation over the North Atlantic as induced by the SST field. *J. Climate*, **24**, 522–542, doi:10.1175/2010JCLI3859.1.

Feliks, Y., A. Groth, A. W. Robertson, and M. Ghil, 2013: Oscillatory climate modes in the Indian Monsoon, North Atlantic and Tropical Pacific. *J. Climate*, **26**, 9528–9544, doi:10.1175/JCLI-D-13-00105.1.

Fukunaga, K., 1990: *Introduction to statistical pattern recognition.* 2nd ed., Academic Press.

Gámiz-Fortis, S., D. Pozo-Vázquez, M. Esteban-Parra, and Y. Castro-Díez, 2002: Spectral characteristics and predictability of the NAO assessed through singular spectral analysis. *J. Geophys. Res.*, **107 (D23)**, ACL–11, doi:10.1029/2001JD001436.

Ghil, M., Y. Feliks, and L. Sushama, 2002a: Baroclinic and barotropic aspects of the wind-driven ocean circulation. *Physica D*, **167 (1)**, 1–35.

Ghil, M., and R. Vautard, 1991: Interdecadal oscillations and the warming trend in global temperature time series. *Nature*, **350 (6316)**, 324–327, doi:10.1038/350324a0.

Ghil, M., and Coauthors, 2002b: Advanced spectral methods for climatic time series. *Rev. Geophys.*, **40 (1)**, 1–41, doi:10.1029/2000RG000092.

Giese, B. S., and S. Ray, 2011: El Niño variability in Simple Ocean Data Assimilation (SODA), 1871–2008. *Journal of Geophysical Research: Oceans (1978–2012)*, **116 (C2)**, doi:10.1029/2010JC006695.

Golyandina, N., V. Nekrutkin, and A. Zhigliavsky, 2001: *Analysis of time series structure: SSA and related techniques.* Chapman & Hall/CRC.

Golyandina, N., and A. A. Zhigljavsky, 2013: *Singular Spectrum Analysis for Time Series.* Springer Berlin Heidelberg.

Green, B., 1952: The orthogonal approximation of an oblique structure in factor analysis. *Psychometrika*, **17 (4)**, 429–440, doi:10.1007/BF02288918.

Groth, A., and M. Ghil, 2011: Multivariate singular spectrum analysis and the road to phase synchronization. *Phys. Rev. E*, **84**, 036 206, doi:10.1103/PhysRevE.84.036206.

Holmström, L., and I. Launonen, 2013: Posterior singular spectrum analysis. *Stat. Anal. Data Min.*, **6 (5)**, 387–402, doi:10.1002/sam.11195.

Hurley, J. R., and R. B. Cattell, 1962: The Procrustes program: Producing direct rotation to test a hypothesized factor structure. *Behav. Sci.*, **7 (2)**, 258–262, doi:10.1002/bs.3830070216.

Jiang, S., F. Jin, and M. Ghil, 1995: Multiple equilibria, periodic, and aperiodic solutions in a wind-driven, double-gyre, shallow-water model. *J. Phys. Oceanogr.*, **25 (5)**, 764–786.

Jolliffe, I. T., 2002: *Principal Component Analysis.* 2nd ed., Springer.

Kaiser, H., 1958: The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, **23 (3)**, 187–200, doi:10.1007/BF02289233.

Keppenne, C. L., and M. Ghil, 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.

Mañé, R., 1981: On the dimension of the compact invariant sets of certain non-linear maps. *Dynamical Systems and Turbulence*, Lecture Notes in Mathematics, Vol. 898, Springer, Berlin, 230–242, doi:10.1007/BFb0091916.

Moron, V., R. Vautard, and M. Ghil, 1998: Trends, interdecadal and interannual oscillations in global sea-surface temperatures. *Climate Dyn.*, **14 (7)**, 545–569, doi:10.1007/s003820050241.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.*, **110 (7)**, 699–706, doi:10.1175/1520-0493(1982)110⟨0699:SEITEO⟩2.0.CO;2.

O'Lenic, E. A., and R. E. Livezey, 1988: Practical considerations in the use of rotated principal component analysis (RPCA) in diagnostic studies of upper-air height fields. *Mon. Wea. Rev.*, **116 (8)**, 1682–1689, doi:10.1175/1520-0493(1988)116⟨1682:PCITUO⟩2.0.CO;2.

Paluš, M., and D. Novotná, 2004: Enhanced Monte Carlo singular system analysis and detection of period 7.8 years oscillatory modes in the monthly NAO index and temperature records. *Nonlinear Processes Geophys.*, **11 (5/6)**, 721–729, URL http://www.nonlin-processes-geophys.net/11/721/2004/.

Paunonen, S. V., 1997: On chance and factor congruence following orthogonal Procrustes rotation. *Educ. Psychol. Meas.*, **57 (1)**, 33–59, doi:10.1177/0013164497057001003.

Plaut, G., and R. Vautard, 1994: Spells of low-frequency oscillations and weather regimes in the Northern Hemisphere. *J. Atmos. Sci.*, **51 (2)**, 210–236, doi:10.1175/1520-0469(1994)051⟨0210:SOLFOA⟩2.0.CO;2.

Richman, M. B., 1986: Rotation of principal components. *Int. J. Climatol.*, **6 (3)**, 293–335, doi:10.1002/joc.3370060305.

Robertson, A. W., 1996: Interdecadal variability over the North Pacific in a multi-century climate simulation. *Climate Dyn.*, **12 (4)**, 227–241, doi:10.1007/BF00219498.

Schönemann, P., 1966: A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, **31**, 1–10, doi:10.1007/BF02289451.

Speich, S., H. Dijkstra, and M. Ghil, 1995: Successive bifurcations in a shallow-water model applied to the wind-driven ocean circulation. *Nonlinear Processes Geophys.*, **2**, 241–268.

Takens, F., 1981: Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence*, Lecture Notes in Mathematics, Vol. 898, Springer, Berlin, 366–381.

Vautard, R., and M. Ghil, 1989: Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D*, **35 (3)**, 395–424, doi:10.1016/0167-2789(89)90077-8.

Vautard, R., P. Yiou, and M. Ghil, 1992: Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D*, **58 (1-4)**, 95–126, doi:10.1016/0167-2789(92)90103-T.