Enhanced Monte Carlo Singular System Analysis and detection of period 7.8 years oscillatory modes in the monthly NAO index and temperature records

M. Paluš[1] and D. Novotná[2]

[1]Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic
[2]Institute of Atmospheric Physics, Academy of Sciences of the Czech Republic, Boční II/1401, 141 31 Prague 4, Czech Republic

Abstract. An extension of the Monte Carlo Singular System Analysis (MC SSA) is described, based on evaluating and testing regularity of dynamics of the SSA modes against the colored noise null hypothesis, in addition to the test based on variance (eigenvalues). The application of the regularity index, computed from a coarse-grained estimation of mutual information, enhances the test sensitivity and reliability in detection of relatively more regular dynamical modes than those obtained by decomposition of colored noises, in particular, in detection of irregular oscillations embedded in red noise. This enhanced MC SSA is successfully applied in detection of period 7.8 years oscillatory modes in records of monthly average near-surface air temperature from several European locations, as well as in the monthly North Atlantic Oscillation index.

1 Introduction

Searching for dynamical mechanisms underlying experimental data in order to understand, model, and predict complex, possibly nonlinear processes, such as those studied in geophysics, in many cases starts with an attempt to identify trends, oscillatory processes and/or other potentially deterministic signals in a noisy environment. Singular system (or singular spectrum) analysis (SSA) in its original form (also known as principal component analysis, or Karhunen-Loève decomposition) is a method for identification and distinction from noise of important information in multivariate data. It is based on an orthogonal decomposition of a covariance matrix of multivariate data under study. The SSA provides an orthogonal basis onto which the data can be transformed, making thus individual data components ("modes") linearly in-

Correspondence to: M. Paluš
(mp@cs.cas.cz)

dependent. Each of the orthogonal modes (projections of the original data onto new orthogonal basis vectors) is characterized by its variance, which is given by the related eigenvalue of the covariance matrix. Here we will deal with a univariate version of SSA (which, however, can be generalised into a multivariate version, see, e.g. Allen and Robertson, 1996) in which the analyzed data is a univariate time series and the decomposed matrix is a time-lag covariance matrix, i.e. instead of several components of multivariate data, a time series and its time-lagged versions are considered. This type of the SSA application, which has frequently been used especially in the field of meteorology and climatology (Vautard and Ghil, 1989; Ghil and Vautard, 1991; Keppenne and Ghil, 1992; Yiou et al., 1994; Allen and Smith, 1994), can provide a decomposition of the studied time series into orthogonal components (modes) with different dynamical properties and thus "interesting" phenomena such as slow modes (trends) and regular or irregular oscillations (if present in the data) can be identified and retrieved from the background of noise and/or other "uninteresting" non-specified processes.

In the traditional SSA, the distinction of "interesting" components (signal) from noise is based on finding a threshold (jump-down) to a "noise floor" in a sequence of eigenvalues given in a descending order.

This approach might be problematic if the signal-to-noise ratio is not sufficiently large, or the noise present in the data is not white but "colored". For such cases statistical approaches utilizing the Monte Carlo simulation techniques have been proposed (Ghil and Vautard, 1991; Vautard et al., 1992) for reliable signal/noise separation. The particular case of the Monte Carlo SSA (MCSSA) which considers the "red" noise, usually present in geophysical data, has been introduced by Allen and Smith (1996). In this paper we present and develop an extension of the Monte Carlo Singular System Analysis based on evaluating and testing regularity of

dynamics of the SSA modes.

A brief introduction into the Monte Carlo singular system analysis is given in Sect. 2, its enhancement by testing the dynamics of modes is explained in Sect. 3. The practical implementation of the enhanced MCSSA, used here, as well as examples of its applications using numerically generated data are presented in Sect. 4. Section 5 summarizes the application of the enhanced MCSSA to monthly average near-surface temperature records from ten European locations. An enhanced MCSSA detection of an oscillatory mode in the NAO index series and its comparison with the related mode detected in the temperature data is described in Sect. 6. Discussion and conclusion are given in Sect. 7.

2   Monte Carlo Singular System Analysis

Let a univariate time series $\{y(i)\}$, $i=1, \ldots, N_0$, be a realization of a stochastic process $\{Y(i)\}$ which is stationary and ergodic. A map into a space of n-dimensional vectors $\mathbf{x}(i)$ with components $x^k(i)$, where $k=1, \ldots, n$, is given as

$$x^k(i)=y(i + k - 1). \tag{1}$$

The sequence of the vectors $\mathbf{x}(i)$, $i=1, \ldots, N=N_0 - (n-1)$, is usually referred to as the $n \times N$ trajectory matrix $\mathbf{X}=\{x_i^k\}$, the number n of the constructed components is called the embedding dimension, or the length of the (embedding) window. Suppose that the studied time series $\{y(i)\}$ results from a linear combination of m different dynamical modes, $m<n$. Then, in an ideal case, the rank of the trajectory matrix $\mathbf{X}$ is rank$(\mathbf{X})=m$, and, $\mathbf{X}$ can be transformed into a matrix with only m nontrivial linearly independent components. Instead of the $n \times N$ matrix $\mathbf{X}$ it is more convenient to decompose the symmetric $n \times n$ matrix $\mathbf{C} = \mathbf{X}^T\mathbf{X}$, since rank$(\mathbf{X})$ = rank$(\mathbf{C})$. The elements of the covariance matrix $\mathbf{C}$ are

$$c_{kl}=(1/N) \sum_{i=1}^{N} x^k(i)x^l(i), \tag{2}$$

where $1/N$ is the proper normalization and the components $x^k(i)$, $i=1, \ldots, N$, are supposed to have zero mean. The symmetric matrix $\mathbf{C}$ can be decomposed as

$$\mathbf{C}=\mathbf{V}\Sigma\mathbf{V}^{T}, \tag{3}$$

where $\mathbf{V}=\{v_{ij}\}$ is an $n \times n$ orthonormal matrix, $\Sigma= \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$, σ_i are non-negative eigenvalues of $\mathbf{C}$ by convention given in descending order $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$. If rank$(\mathbf{C})=m<n$, then

$$\sigma_1 \geq \ldots \geq \sigma_m > \sigma_{m+1}= \ldots =\sigma_n=0. \tag{4}$$

In the presence of noise, however, all eigenvalues are positive and the relation (4) takes the following form (Broomhead and King, 1986):

$$\sigma_1 \geq \ldots \geq \sigma_m >> \sigma_{m+1} \geq \ldots \geq \sigma_n > 0. \tag{5}$$

Then, the modes ξ_i^k

$$\xi_i^k=\sum_{l=1}^{n} v_{lk}x_i^l, \tag{6}$$

for $k=1, \ldots, m$ are considered as the "signal" part, and the modes ξ_i^k, $k=m + 1, \ldots, n$, are considered as the noise part of the original time series. The "signal" modes can be used to reconstruct the denoised signal $\tilde{x}_l^k$ as

$$\tilde{x}_i^k=\sum_{l=1}^{m} v_{kl}\xi_i^l. \tag{7}$$

Of course, the original time series x_i^k can be reconstructed back from the modes as

$$x_i^k=\sum_{l=1}^{n} v_{kl}\xi_i^l. \tag{8}$$

In the latter relation – decomposition (8), the modes ξ_i^k can also be interpreted as time-dependent coefficients and the orthogonal vectors $\mathbf{v}_k=\{v_{kl}\}$ as basis functions, usually called the empirical orthogonal functions (EOF's).

The clear eigenvalue-based signal/noise distinction (5) can be obtained only in particularly idealized situation when the signal/noise ratio is large enough and the background consists of a white noise. A kind of a colored noise, the "red" noise, which is particularly important for its presence in many geophysical processes (Allen and Smith, 1996), can be modeled by using an AR(1) model (autoregressive model of the first order):

$$u(i)-\hat{u}=\alpha(u(i - 1)-\hat{u})+\gamma z(i), \tag{9}$$

where $\hat{u}$ is the process mean, α and γ are process parameters, and $z(i)$ is a Gaussian white noise with a zero mean and a unit variance.

The red noises possess power spectra of the $1/f$ type, and their SSA eigenspectra have the same character, i.e. an eigenspectrum of a red noise is equivalent to a coarsely discretized power spectrum, where the number of frequency bins is given by the embedding dimension n. The eigenvalues related to the slow modes are much larger than the eigenvalues of the modes related to higher frequencies. Thus, in the classical SSA approach applied to a red noise, the eigenvalues of the slow modes might incorrectly be interpreted as a (nontrivial) signal, or, on the other hand, a nontrivial signal embedded in a red noise might be neglected, if its variance is smaller than the slow-mode eigenvalues of the background red noise. Therefore the mutual comparison of eigenvalues inside an eigenspectrum cannot lead to a reliable detection of a nontrivial signal, if a red noise is present in studied data. In order to correctly detect a signal in a red noise, the following approach has been proposed (Allen and Smith, 1996):

First, the eigenvalues are plotted not according to their values, but according to a frequency associated with a particular mode (EOF), i.e. the eigenspectrum in this form becomes a sort of a (coarsely) discretized power spectrum in general,

not only in the cases of red noises (when the eigenspectra have naturally this form, as mentioned above).

Second, an eigenspectrum obtained from a studied data is compared, in a frequency-by-frequency way, with eigenspectra obtained from a set of realizations of an appropriate noise model (such as the AR(1) model, Eq. 9), i.e. an eigenvalue related to a particular frequency bin obtained from the data is compared with a range of eigenvalues related to the same frequency bin, obtained from the set of so-called surrogate data, i.e. the data artificially generated according to the chosen noise model (null hypothesis) (Allen and Smith, 1996; Smith, 1992; Theiler et al., 1992; Paluš, 1995).

The detection of a nontrivial signal in an experimental time series becomes a statistical test in which the null hypothesis that the experimental data were generated by a chosen noise model is tested. When (an) eigenvalue(s) associated with some frequency bin(s) differ(s) with a statistical significance from the range(s) of related noise model eigenvalues, then one can infer that the studied data cannot be fully explained by the considered null hypothesis (noise model) and could contain an additional (nontrivial) signal. This is a rough sketch of the approach, for which we will use the term Monte Carlo SSA (MCSSA), as coined by Allen and Smith (see (Allen and Smith, 1996) where also a detailed account of the MCSSA approach with analyses of various levels of null hypotheses is given) although the same term was earlier used for other SSA methods, which considered white noise background (Ghil and Vautard, 1991; Vautard et al., 1992).

3 Enhancement of MCSSA by testing dynamics of modes

The above MCSSA is a sophisticated technique, however, it still assumes a very simple model that the signal of interest has been linearly added to a specified background noise and therefore the variance in the frequency band, characteristic for the searched signal, is significantly greater than the typical variance in this frequency band obtained from the considered noise model. If the studied signal has a more complicated origin, e.g. when an oscillatory mode is embedded into a background process without significantly increasing variance in a particular frequency band, the standard MCSSA can fail. In order to be able to detect any interesting dynamical mode independently of its (relative) variance, Paluš and Novotná (1998) have proposed to test also dynamical properties of the SSA modes against the modes obtained from the surrogate data. How can we characterize dynamics in a simple, computationally effective way?

Consider a complex, dynamic process evolving in time. A series of measurements done on such a system in consecutive instants of time $t=1, 2, \ldots$ is usually called a time series $\{y(t)\}$. Consider further that the temporal evolution of the studied system is not completely random, i.e. that the state of the system in time t in some way depends on the state in which the system was in time $t-\tau$. The strength of such a dependence per a unit time delay τ, or, inversely, a rate at which

the system "forgets" information about its previous states, can be an important quantitative characterization of temporal complexity in the system's evolution. The time series $\{y(t)\}$, which is a recording of (a part of) the system temporal evolution, can be considered as a realization of a stochastic process, i.e. a sequence of stochastic variables. Uncertainty in a stochastic variable is measured by its entropy. The rate in which the stochastic process "produces" uncertainty is measured by its entropy rate.

The concept of entropy rates is common to the theory of stochastic processes as well as to the information theory where the entropy rates are used to characterize information production by information sources (Cover and Thomas, 1991).

Alternatively, the time series $\{y(t)\}$ can be considered as a projection of a trajectory of a dynamical system, evolving in some measurable state space. A. N. Kolmogorov, who introduced the theoretical concept of classification of dynamical systems by information rates, was inspired by the information theory and generalized the notion of the entropy of an information source (Sinai, 1976). The Kolmogorov-Sinai entropy (KSE) (Cornfeld et al., 1982; Petersen, 1983; Sinai, 1976) is a topological invariant, suitable for classification of dynamical systems or their states, and is related to the sum of the system's positive Lyapunov exponents (LE) according to the theorem of Pesin (1977).

Thus, the concept of entropy rates is common to theories based on philosophically opposite assumptions (randomness vs. determinism) and is ideally applicable for characterization of complex geophysical processes, where possible deterministic rules are always accompanied by random influences.

However, possibilities to compute the exact entropy rates from experimental data are limited to a few exceptional cases. Therefore Paluš (1996) has proposed "coarse-grained entropy rates" (CER's) instead. The CER's are relative measures of regularity and predictability of analyzed time series and are based on coarse-grained estimates of information-theoretic functionals. In the simplest case, applied here, we use so called mutual information. The mutual information $I(X; Y)$ of two random variables X and Y is given by $I(X; Y)=H(X)+H(Y)-H(X, Y)$, where the entropies $H(X)$, $H(Y)$, $H(X, Y)$ are defined in the usual Shannonian sense (Cover and Thomas, 1991). A detailed account on relations between the entropy rates and the information-theoretic functionals is given in (Paluš, 1996, 1997a). For a time series $\{x(t)\}$, considered as a realization of a stationary and ergodic stochastic process $\{X(t)\}$, $t=1, 2, 3, \ldots$, we compute the mutual information $I(x; x_\tau)$ as a function of time lag τ. In the following we will mark $x(t)$ as x and $x(t+\tau)$ as x_τ. For defining the simplest form of CER let us find such $\tau_{\max}$ that for $\tau' \geq \tau_{\max}$: $I(x; x_{\tau'}) \approx 0$ for the analysed datasets. Then we define the norm of the mutual information

$$||I(x; x_\tau)|| = \frac{\Delta\tau}{\tau_{\max}-\tau_{\min}+\Delta\tau} \sum_{\tau=\tau_{\min}}^{\tau_{\max}} I(x; x_\tau) \qquad (10)$$

with $\tau_{min}=\Delta\tau=1$ sample as a usual choice. The CER h^1 is then defined as

$$h^1 = I(x, x_{\tau_0}) - ||I(x; x_\tau)||. \qquad (11)$$

It has been shown that the CER h^1 provides the same classification of states of chaotic systems as the exact KSE (Paluš, 1996). Since usually $\tau_0=0$ and $I(x; x)=H(X)$ which is given by the marginal probability distribution $p(x)$, the sole quantitative descriptor of the underlying dynamics is the mutual information norm (Eq. 10) which we will call the regularity index. Since the mutual information $I(x; x_\tau)$ measures the average amount of information contained in the process $\{X\}$ about its future τ time units ahead, the regularity index $||I(x; x_\tau)||$ gives an average measure of predictability of the studied signal and is inversely related to the signal's entropy rate, i.e. to the rate at which the system, or process, producing the studied signal, "forgets" information about its previous states.

4   Implementation of the enhanced MCSSA and numerical examples

We realize the enhanced version of MCSSA as follows.

1. The studied time series undergoes SSA as described in Sect. 2, i.e. using an embedding window of length n, the $n \times n$ lag-correlation matrix $\mathbf{C}$ is decomposed using the SVDCMP routine (Press et al., 1986). In the eigenspectrum, the position of each eigenvalue on the abscissa is given by the dominant frequency associated with the related EOF, i.e. detected in the related mode. That is, the studied time series is projected onto the particular EOF, the power spectrum of the projection (mode) is estimated, and the frequency bin with the highest power is identified. This spectral coordinate is mapped onto one of the n frequency bins, which equidistantly divide the abscissa of the eigenspectrum.

2. An AR(1) model is fitted on the series under study, the residuals are computed.

3. The surrogate data are generated using the above AR(1) model, "scrambled" (randomly permutated in temporal order) residuals are used as innovations.

4. Each realization of the surrogates undergoes SSA as described in item 1. Then, the eigenvalues for the whole surrogate set, in each frequency bin, are sorted and the values for the 2.5th and 97.5th percentiles are found. In eigenspectra, the 95% range of the surrogates eigenvalue distribution is illustrated by a horizontal bar between the above percentile values.

5. For each frequency bin, the eigenvalue obtained from the studied data is compared with the range of the surrogate eigenvalues. If an eigenvalue lies outside the range given by the above percentiles, the null hypothesis of the AR(1) process is rejected, i.e. there is a probability $p<0.05$ that the data can be explained by the null noise model.

6. For each SSA mode (a projection of the data onto a particular EOF) its regularity index is computed, as well as for each SSA mode for all realizations of the surrogate data. The regularity indices are processed and statistically tested in the same way as the eigenvalues. The regularity index is based on mutual information obtained by a simple box-counting approach with marginal equiquantization (Paluš, 1995, 1996, 1997a). In general, this testing approach is not limited to the particular regularity index used, but can be based on a suitable information/entropy measure obtained by a different algorithm, employing novel methods such as that of recurrence plots (Romano et al., 2004); or even different complexity measures (Wackerbauer et. al., 1994).

Performing MCSSA using the embedding window of the length n, there are n eigenvalues in the eigenspectrum, and n statistical tests are done. Therefore the problem of the simultaneous statistical inference should be considered in applications (see (Paluš, 1995) and references therein). However, since here we are interested in a detection of a signal in a specific frequency band (and not in rejecting the null hypothesis by a digression from the surrogate range by an eigenvalue or a regularity index in any frequency band), we will not discuss this topic here.

Rejecting the null hypothesis of the AR(1) (or other appropriate) noise model, one can infer that there is something more in the data than a realization of the null hypothesis (noise) model. The rejection based on the eigenvalues indicates a different covariance structure than the noise model used. The rejection based on the regularity index indicates that the studied data contains a dynamically interesting signal with higher regularity and predictability than a mode obtained by linear filtration of the considered noise model.

The presented approach is demonstrated here by using numerically generated data.

1.   A periodic signal with randomly variable amplitude (Fig. 1a) was mixed with a realization of an AR(1) process with a strong slow component (Fig. 1b). The used noise model is defined as $x_i=0.933x_{i-1}+\xi_i$, where ξ_i are Gaussian deviates with a zero mean and a unit variance. The signal to noise ratios obtained by mixing the signals were 1:2 (Fig. 1c), and 1:4 (Fig. 1d). (The given signal/noise ratios are the ratios of the standard deviations.) The latter two series are analyzed by the presented method.

The eigenspectrum of the time series consisting of the signal (Fig. 1a) and the AR(1) noise (Fig. 1b) in the ratio 1:2 (Fig. 1c) is presented in Fig. 2a, where logarithms of the eigenvalues are plotted as the bursts ("LOG POWER"). The series is considered as unknown experimental data, so that an AR(1) model is fitted on the data and the surrogates are generated as described above. The vertical bars in the eigenspectrum represent the surrogate eigenvalue ranges from 2.5th to 97.5th percentiles, which were obtained from 1500 sur-
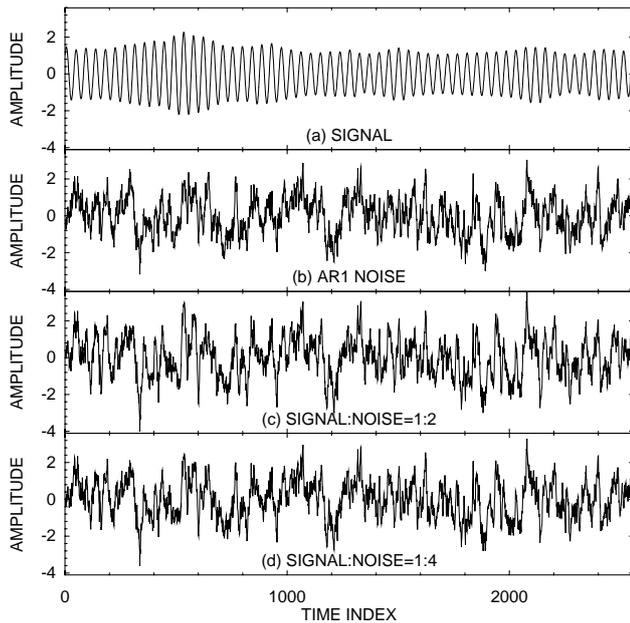
Fig. 1. Numerically generated test data: **(a)** A periodic signal with randomly variable amplitude was mixed with **(b)** a realization of an AR(1) process with a strong slow component, obtaining the signal to noise ratio 1:2 **(c)**, and 1:4 **(d)**.

Fig. 2. The standard – eigenvalue based **(a)–(c)** and the enhanced – regularity index based **(d)–(f)** MCSSA analysis of the numerical data, presented in Fig. 1. (a) The full eigenspectrum and (b) the low-frequency part of the eigenspectrum – logarithms of eigenvalues ("LOG POWER") plotted according to the dominant frequency associated with particular modes, for the signal to noise ratio 1:2. (c) Low frequency part of the eigenspectrum for the signal to noise ratio 1:4. (d) The regularity spectrum and (e) its low frequency part for the signal to noise ratio 1:2. (f) Low frequency part of the regularity spectrum for the signal to noise ratio 1:4. Bursts – eigenvalues or regularity indices for the analysed data; bars – 95% of the surrogate eigenvalues or regularity index distribution, i.e. the bar is drawn from the 2.5th to the 97.5th percentiles of the surrogate eigenvalues/regularity indices distribution.

rogate realizations (here, as well as in the following examples). The eigenvalues of the AR(1) surrogates uniformly fill all the n frequency bins (here, as well as in the following example $n=100$), while in the case of the test data, some bins are empty, others contain one, two or more eigenvalues. We plot the surrogate bars only in those positions, in which (an) eigenvalue(s) of the analyzed data exist(s). Note the $1/f$ character of the surrogate eigenspectrum, i.e. the eigenvalues plotted according to increasing dominant frequency associated with the related modes are monotonously decreasing in a $1/f^\alpha$ way. The low-frequency part of the eigenspectrum from Fig. 2a is enlarged in Fig. 2b. The two data eigenvalues related to the frequency 0.02 (cycles per time unit) are clearly above the surrogate bar, i.e. they are significant on the 95% level and the null hypothesis is rejected. Further study of the significant modes shows that they are related to the embedded in noise signal from Fig. 1a, in particular, one of the modes contains the signal together with some noise of similar frequencies, and the other include an oscillatory mode shifted by $\pi/2$ relatively to the former. Note that the simple SSA based on the mutual comparison of the data eigenvalues could be misleading, since the AR(1) noise itself "produces" two or three eigenvalues which are larger than the two eigenvalues related to the signal embedded in the noise.

The same analysis applied to the series possessing the signal/noise ratio 1:4 (Figs. 2c), however, fails to detect the embedded signal – all eigenvalues obtained from the test data are well confined between the 2.5th and 97.5th percentiles of the surrogate eigenvalues distributions. Applying the test based on the regularity index on the mixture with the signal
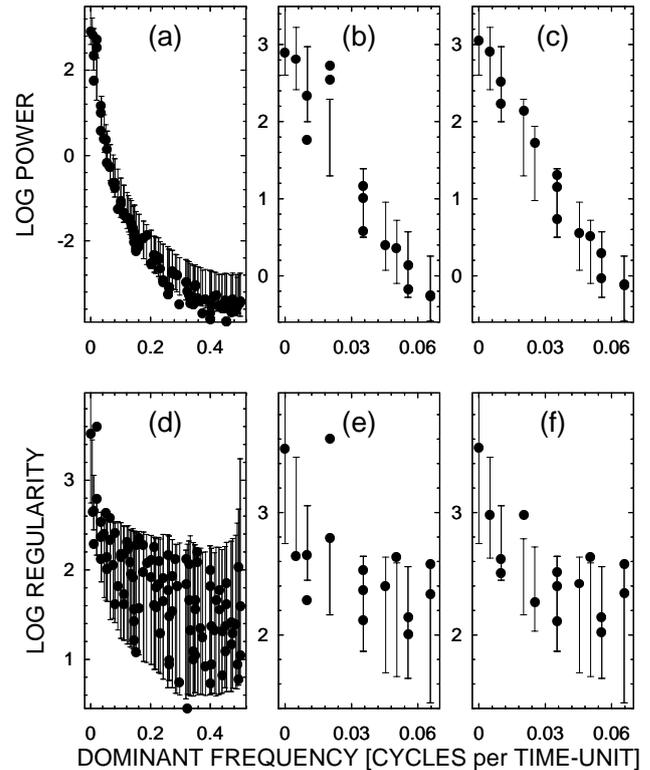
to noise ratio 1:2 (Figs. 2d and 2e) one data regularity index has been found significantly higher than the related surrogate indices. It was obtained from the mode related to the frequency bin 0.02, as in the case of the significant eigenvalues in Figs. 2a and 2b. This is the mode which contains the embedded signal (Fig. 1a) together with some noise of similar frequencies. The orthogonal mode, related to the same frequency bin, which has the variance comparable to the former (Figs. 2a and 2b), has its regularity index close to the 97.5th percentile of the surrogate regularity indices distribution. In other words, if a (nearly) periodic signal is embedded in a (colored) noise, the SSA approach, in principle, is able to extract this signal together with some noise of close frequencies, and produces an orthogonal "ghost" mode which has a comparable variance, however, its dynamical properties are closer to those of the modes obtained from the pure noise
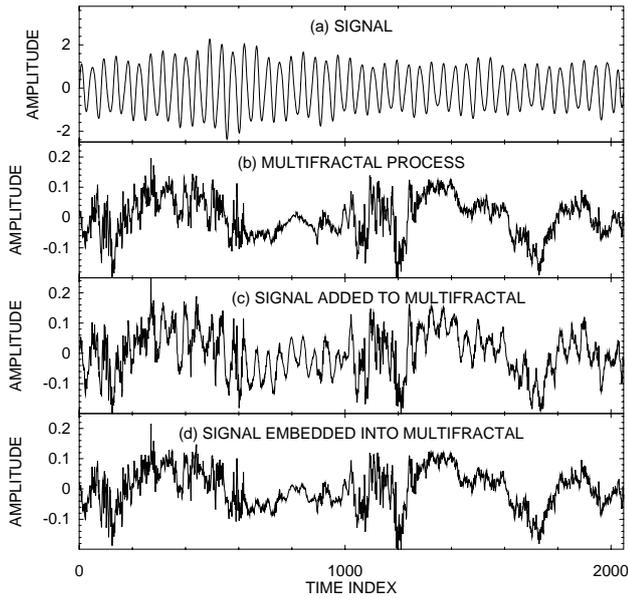
Fig. 3. Numerically generated test data: **(a)** The wavelet filtered signal from Fig. 1a was embedded into **(b)** a realization of a multifractal process, obtaining the ratio of related wavelet coefficients 1:2 **(c)**, and 0.5:0.5 **(d)**.

Fig. 4. The low frequency parts of the MCSSA eigenspectra **(a)**–**(c)** and regularity spectra **(d)**–**(f)** for the signal embedded into a multifractal process with wavelet coefficient ratio 1:2 (a), (d) and 0.5:0.5 (b), (c), (e), (f). Bursts – eigenvalues or regularity indices for the analysed data; bars – 95% of the surrogate eigenvalues or regularity index distribution obtained from the AR(1) (a), (b), (d), (e) and the multifractal (c), (f) surrogate data.

(null model), as measured by the regularity index (Eq. 10). Nevertheless, the regularity index used as a test statistic in the MCSSA manner is able to detect the embedded signal with a high statistical significance in this case (signal:noise = 1:2), as well as in the case of the signal to noise ratio 1:4 (Fig. 2f), when the standard (variance-based MCSSA) failed (Fig. 2c). In the latter case the orthogonal "ghost" mode did not appear, and the regularity index of the signal mode is lower than in the previous case, since the mode contains larger portion of the isospectral noise, however, the signal mode regularity index is still safely above the surrogate bar, i.e. significant with $p < 0.05$ (Fig. 2f).

2. As a more complex example we "embed" the test signal (Fig. 1a) into a realization of a multifractal process (Fig. 3b) generated as a log-normal random cascade on a wavelet dyadic tree (Arneodo et al., 1998) using the discrete wavelet transform (Press et al., 1986). Using the wavelet decomposition, we embedd the most significant part of the signal (Fig. 1a) related to a particular wavelet scale – this wavelet-filtered signal is illustrated in Fig. 3a. The mixing is done in the space of the wavelet coefficients, in the first case (in Fig. 3 referred to as "signal added to multifractal") the standard deviation (SD) of the signal wavelet coefficients is the double of the SD of the wavelet coefficients of the multifractal signal in the related scale (Fig. 3c), i.e. the added signal deviates from the covariance structure of the "noise" (multifractal) process, while in the second case we adjusted the SD of both sets of the wavelet coefficients to 50% of the SD of the wavelet coefficients of the original multifractal signal in the related scale (Fig. 3d), so that the total variance in
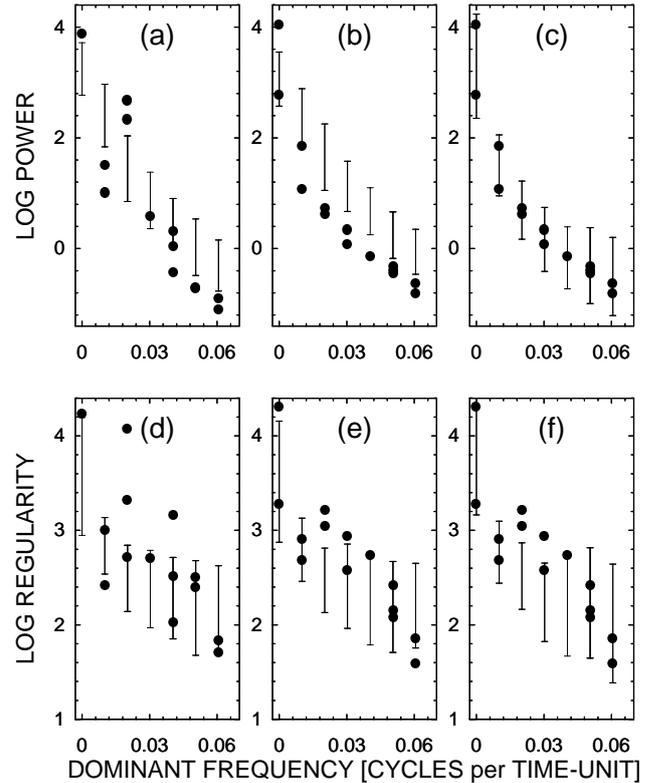
this scale (frequency band) does not excess the related variance of the "clean" multifractal. Then, it is not surprising, that the variance-(eigenvalues)-based MCSSA test, using the AR(1) surrogate data (Figs. 4a and 4b) clearly distinguishes the signal from the multifractal background in the first case (Fig. 4a) including its orthogonal "ghosts", while in the second case no eigenvalue is over the AR(1) surrogate range, but the slow trend mode (Fig. 4b). The AR(1) process is unable to correctly mimic the multifractal process – the slow mode (the zero frequency bin) scores as a significant trend over the AR(1) surrogate range, while the variance on subsequent frequencies is overestimated (Figs. 4a and 4b). On the other hand, even the AR(1) surrogate model is able to detect the added signal in the first case (Fig. 4a). If we use realizations of the same multifractal process as the surrogate data, the signal is detected in the first case (not presented, just compare the bursts on frequency 0.02 in Fig. 4a and the related surrogate bar in Fig. 4c), while in the second case, the eigenvalues-based MCSSA neglects the signal embedded into the multifractal "noise" – all the data mixture eigenvalues (bursts) are inside the multifractal surrogate bars (Fig. 4c). In the MCSSA tests using the regularity index, the embedded signal
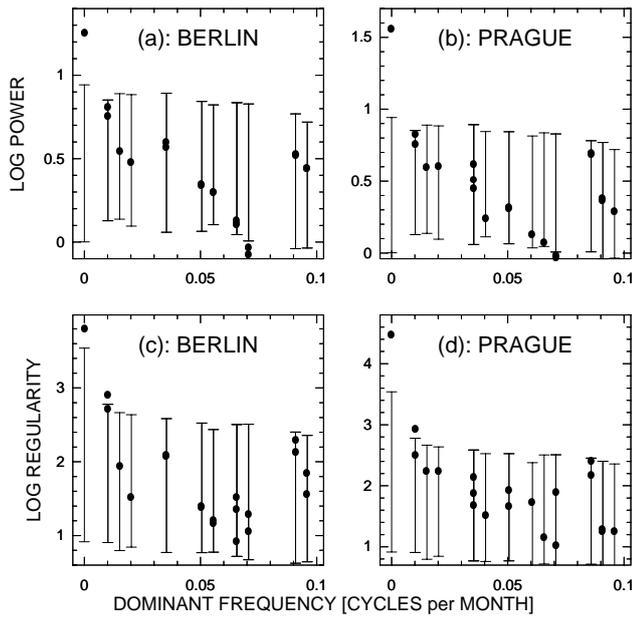
Fig. 5. Enhanced MCSSA analysis of the Berlin **(a), (c)** and Prague **(b), (d)** near-surface air temperature series. Low-frequency parts of eigenspectra (a), (b) and regularity index spectra (c), (d). For the burst/bars key see the caption of Fig. 2.

Fig. 6. Enhanced MCSSA analysis of the Prague near-surface air temperature series **(a), (c)** and the NAO index **(b), (d)**. Low-frequency parts of eigenspectra (a), (b) and regularity index spectra (c), (d). For the burst/bars key see the caption of Fig. 2. Both datasets span the period 1824–2002, the embedding dimension n=480 months was used.

is safely detected together with its orthogonal "ghosts" and higher harmonics not only in the first case (Fig. 4d), but also in the second case, either using AR(1) (Fig. 4e) or the multifractal surrogate data (Fig. 4f), when it is, from the point of view of the covariance structure, indistinguishably embedded into the multifractal process.

5   Application of the enhanced MCSSA to temperature records

The above numerical examples demonstrated the power of the enhancement of the MCSSA in which we test also the dynamical properties of the SSA modes, namely its regularity, against the dynamical properties of the surrogate SSA modes. In the following we apply this approach to the monthly average near-surface air temperature series from ten European stations (Stockholm, De Bilt, Paris – Le Bourget, Geneve – Cointrin, Berlin – Tempelhof, Munich – Riem, Vienna – Hohe Warte, Budaors, Wroclaw II, obtained from the Carbon Dioxide Information Analysis Center Internet server (ftp://cdiac.esd.ornl.gov/pub/ndp041) and to a series from Prague – Klementinum station from the period 1781–1988. The long-term monthly averages were subtracted from the data, so that the annual cycle was effectively filtered-out.

The enhanced MCSSA analyses of the Berlin and Prague temperature series, using the embedding window of length n=100 (months), are presented in Fig. 5. In the classical MCSSA test based on eigenvalues (Figs. 5a and 5b) the only significance has been found for the zero frequency mode, i.e. there is a significant long-term trend present, inconsistent
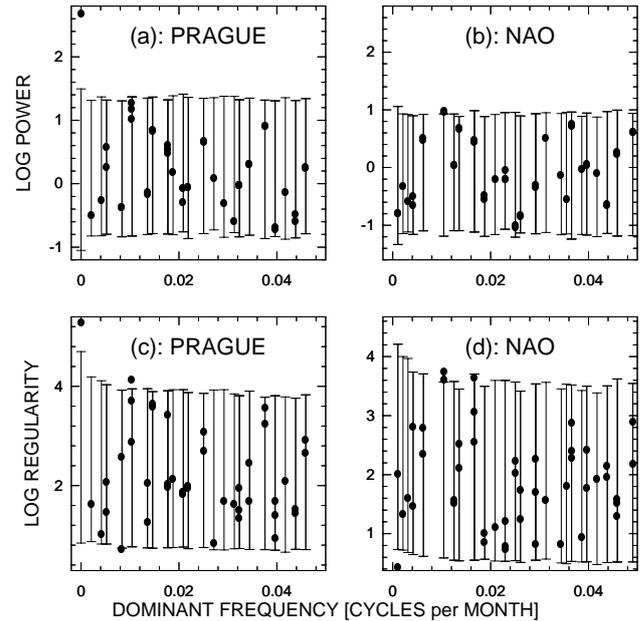
with the hypothesis of the AR(1) noise, however, no oscillations or other dynamical phenomena exceeding the AR(1) model, have been detected. The situation is different using the test based on the regularity index (Figs. 5c and 5d), when, in addition to the significant long-term trend, also another mode, related to oscillatory dynamics with a period of 7.8 years (approx. 0.01 cycles per month, Figs. 5c and 5d), has been found significantly different from the AR(1) null hypothesis.

Similar result has been found in the analysis of the series from Wroclaw and De Bilt. In the data from the other six stations only the long-term trend has been found significant, but no oscillations. This result could lead to the question of simultaneous statistical inference, namely to the probability of randomly occurring significances in a part of the data set. Considering geographical locations of the stations, however, we can see a nonrandom pattern in the occurrence of the significant results, since the period 7.8 year cycle has been found in the stations located slightly over 50° of northern latitude.

6   Period 7.8 years cycle detected in the NAO index

The North Atlantic Oscillation is a dominant pattern of atmospheric circulation variability in the extratropical Northern Hemisphere and it is a major controlling factor of basic meteorological variables including the temperature (Hurrell et al., 2001). The NAO index is traditionally defined as the
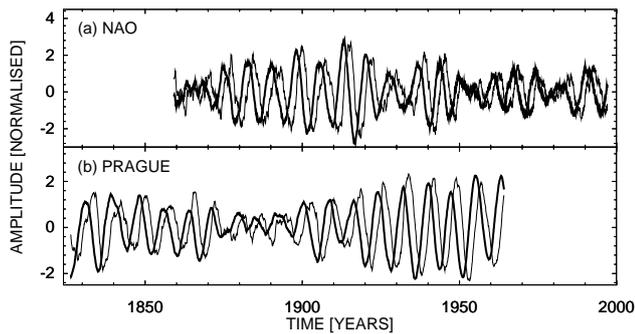
Fig. 7. The period 7.8 years oscillatory mode extracted from the monthly NAO index **(a)** and the monthly Prague near-surface air temperature series **(b)**.

normalized pressure difference between the Azores and Iceland. The NAO data, used here, and their description are available at http://www.cru.uea.ac.uk/cru/data/nao.htm.

Gámiz-Fortis et al. (2002) have applied the standard MCSSA method on the winter NAO index, i.e. yearly sampled values obtained by averaging December, January and February index values, and, using an embedding window of length $n=40$ years were able to identify an oscillatory mode with a period about 7.7 years. According to other authors (Fernandez et al., 2003) the NAO index can be considered just as a pink noise with a very little possible predictability. Recalling our results with the temperature records, we ask if we could detect an oscillatory mode with the period around 7.8 years in the monthly NAO index. As the first step we update the Prague temperature data and remove a portion of their historical part in order to have the data set covering the same period as the available NAO data, i.e. the period of 1824–2002. Since Gámiz-Fortis et al. (2002) used the embedding window of 40 years, we will use the same time window, but in monthly sampling we have $n=480$. Repeating the analysis for the Prague temperature data we obtain the same result as above: In the eigenvalue-based MCSSA the only significant mode is related to slow trends (Fig. 6a), while in the test using the regularity index also the oscillatory mode with the period of 7.8 years is detected (Fig. 6c). Analysing the NAO index, the same oscillatory mode is already apparent in the eigevalue-based MCSSA test, however, its eigenvalues lie on the edge of significance (Fig. 6b). Using the regularity index (Fig. 6d) the period 7.8 years mode is reliably detected, i.e. its regularity index lies clearly above the surrogate bar (while the regularity index of its orthogonal "ghost" is again on the edge of significance). In Fig. 7 we illustrate the detected modes together with their orthogonal twins. Due to the embedding dimension $n=480$ months, there is an uncertainty of the exact timing of the modes equal to the embedding window of 40 years. We adjusted the temporal coordinate of the modes by maximizing their correlation with the original data.

7   Conclusions

An extension of the Monte Carlo SSA method has been described, based on evaluating and testing regularity of dynamics of the SSA modes against the colored noise null hypothesis in addition to the test based on variance (eigenvalues). It has been demonstrated that such an approach could enhance the test sensitivity and reliability in detection of relatively more regular dynamical modes than those obtained by decomposition of colored noises. The standard MCSSA can detect only those signals whose variance significantly exceed the variance of background noise in the frequency range of the signal to be detected. The proposed enhanced MCSSA version can detect signals with relatively small variance, or even signals which are "embedded" into the variance/frequency structure of the background noise, if the signals have more regular dynamics than related SSA modes obtained by linear filtering of the model background noise.

The enhanced MCSSA has been applied to records of monthly average near-surface air temperature from ten European locations. In the part of the latter, located over 50° of northern latitude, an oscillatory mode with a period of 7.8 years has been detected. Then the same oscillatory mode has been detected in the monthly NAO index. Can the existence of the same oscillatory mode in the NAO index and in the temperature records be regarded as an evidence that the NAO influences the European temperature (also) on this temporal scale? Before answering this question, possible relations between these oscillatory modes should be carefully studied. Analyses of possible phase synchronization (Rosenblum et al., 1996; Paluš, 1997b) and causality relations (Rosenblum and Pikovsky, 2001; Paluš and Stefanovska, 2003) are the next step planned in this project. The present result, however, is already important, since the discovered period 7.8 years oscillatory modes in the NAO index and in the temperature records can play an important role in predictions and evaluation of climate changes on near-decadal scales at the mid- and higher latitudes in European regions.

References

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

Allen, M. R. and Robertson, A. W.: Distinguishing modulated oscillations from coloured noise in multivariate datasets, Climate Dynamics, 12, 775–784, 1996.

Allen, M. R. and Smith, L. A.: Monte Carlo SSA: Detesting irregular oscillation in the presence of colored noise, J. Climate, 9, 12, 3373–3404, 1996.

Arneodo, A., Bacry, E., and Muzy, J. F.: Random cascades on wavelet dyadic trees, J. Math. Phys., 39, 8, 4142–4164, 1998.

Broomhead, D. S. and King, G. P.: Extracting qualitative dynamics from experimental data, Physica D, 20, 217–236, 1986.

Cornfeld, I. P., Fomin, S. V., and Sinai, Ya. G.: Ergodic Theory, Springer, New York, 1982.

Cover, T. M. and Thomas, J. A.: Elements of Information Theory, J. Wiley & Sons, New York, 1991.

Fernandez, I., Hernandez, C. N., and Pacheco, J. M.: Is the North Atlantic Oscillation just a pink noise?, Physica A, 323, 705–714, 2003.

Gámiz-Fortis, S. R., Pozo-Vázques, D., Esteban-Parra, M. J., and Castro-Díez, Y.: Spectral characteristics and predictability of the NAO assessed through Singular Spectral Analysis, J. Geophys. Res., 107, D23, 4685, 2002.

Ghil, M. and Vautard, R.: Interdecadal oscillations and the warming trend in global temperature time series, Nature, 350, 6316, 324–327, 1991.

Hurrell, J. W., Kushnir, Y., and Visbeck, M.: Climate – The North Atlantic oscillation, Science, 291, 5504, 603–605, 2001.

Keppenne, C. L. and Ghil, M.: Adaptive filtering and the Southern Oscillation Index, J. Geophys. Res., 97, 20 449–20 454, 1992.

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

Paluš, M.: Testing for nonlinearity using redundancies: Quantitative and qualitative aspects, Physica D, 80, 186–205, 1995.

Paluš, M.: Coarse-grained entropy rates for characterization of complex time series, Physica D, 93, 64–77, 1996.

Paluš, M.: Kolmogorov entropy from time series using information-theoretic functionals. Neural Network World, http://www.uivt.cas.cz/~mp/papers/rd1a.ps, 3/97, 269–292, 1997.

Paluš, M.: Detecting phase synchronization in noisy systems, Phys. Lett. A, 235, 341–351, 1997.

Paluš, M. and Novotná, D.: Detecting modes with nontrivial dynamics embedded in colored noise: Enhanced Monte Carlo SSA and the case of climate oscillations, Phys. Lett. A, 248, 191–202, 1998.

Paluš, M. and Stefanovska, A.: Direction of coupling from phases of interacting oscillators: An information-theoretic approach, Phys. Rev. E, 67, 055201(R), 2003.

Pesin, Ya. B.: Characteristic Lyapunov exponents and smooth ergodic theory, Russian Math. Surveys, 32, 55–114, 1977.

Petersen, K.: Ergodic Theory, Cambridge University Press, Cambridge, 1983.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T.: Numerical Recipes: The Art of Scientific Computing, Cambridge Univ. Press, Cambridge, 1986.

Romano, M. C., Thiel, M., Kurths, J., von Bloh, W.: Multivariate recurrence plots, Phys. Lett. A, 330, 3–4, 214–223, 2004.

Rosenblum, M. G., Pikovsky, A. S., and Kurths, J.: Phase synchronization of chaotic oscillators, Phys. Rev. Lett., 76, 1804–1807, 1996.

Rosenblum, M. G. and Pikovsky, A. S.: Detecting direction of coupling in interacting oscillators, Phys. Rev. E, 64, 045202(R), 2001.

Sinai, Ya. G.: Introduction to Ergodic Theory, Princeton University Press, Princeton, 1976.

Smith, L. A.: Identification a prediction of low-dimensional dynamics, Physica D, 58, 50–76, 1992.

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

Vautard, R. and Ghil, M.: Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, Physica D, 35, 395–424, 1989.

Vautard, R., Yiou, P., and Ghil, M.: Singular spectrum analysis: a toolkit for short noisy chaotic signals, Physica D, 58, 95–126, 1992.

Wackerbauer, R., Witt, A., Altmanspacher, H., Kurths, J., and Scheingraber, H.: A comparative classification of complexity-measures. Chaos Solitons and Fractals, 4, 1, 133–173, 1994.

Yiou, P., Ghil, M., Jouyel, J., Paillard, D., and Vautard, R.: Nonlinear variability of the climatic system, from singular and power spectra of Late Quarternary records, Clim. Dyn., 9, 371–389, 1994.