Detecting Oscillations Hidden in Noise: Common Cycles in Atmospheric, Geomagnetic and Solar Data

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 `mp@cs.cas.cz`
[2] Institute of Atmospheric Physics, Academy of Sciences of the Czech Republic, Boční II/1401, 141 31 Prague 4, Czech Republic `nov@ufa.cas.cz`

Summary

In this chapter we present a nonlinear enhancement of a linear method, the singular system analysis (SSA), which can identify potentially predictable or relatively regular processes, such as cycles and oscillations, in a background of colored noise. The first step in the distinction of a signal from noise is a linear transformation of the data provided by the SSA. In the second step, the dynamics of the SSA modes is quantified in a general, nonlinear way, so that dynamical modes are identified which are more regular, or better predictable than linearly filtered noise. A number of oscillatory modes are identified in data reflecting solar and geomagnetic activity and climate variability, some of them sharing common periods.

Keywords

signal detection, statistical testing, Monte Carlo SSA, sunspots, geomagnetic activity, NAO, air temperature, solar-terrestrial relations

1 Introduction

The quest for uncovering physical 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. The distinction of a relatively regular part of the total variability of a complex natural process can be a key for understanding not only such a process itself, but also interactions with other

processes or phenomena, if they posses, for instance, oscillations on a similar temporal scale.

Singular system (or singular spectrum) analysis (SSA) [1, 2, 3] in its original form (closely related to the principal component analysis or Karhunen-Loève decomposition) is a method for identification and distinction of important information in multivariate data from noise. 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 independent. Each of the orthogonal modes (projections of the original data onto the 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. [4]) 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 SSA application, which has frequently been used especially in the field of meteorology and climatology [5, 6, 7, 8, 9], can provide a decomposition of the studied time series into orthogonal components (modes) with different dynamical properties. 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 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 [6, 10] for reliable signal/noise separation. The particular case of Monte Carlo SSA (MCSSA) which considers the "red" noise, usually present in geophysical data, has been introduced by Allen & Smith [11]. In this paper, we present and apply an extension of the Monte Carlo singular system analysis based on evaluating and testing the regularity of dynamics of the SSA modes. In our approach, we retain the decomposition exploiting the linear covariance structure of the data, however, in the testing (detection) part of the method, we evaluate the regularity of dynamics of the SSA modes using a measure of general, i.e., nonlinear dependence. The latter gives a clue in inferring whether the studied data contain a component which is more regular and predictable, in a general, nonlinear sense, than linearly filtered noise. Attempts to generalize SSA-like approach to accounting for nonlinear dependence structures are also known [12, 13, 14, 15], however, are not considered in this chapter.

2 Monte Carlo singular system analysis and its enhancement

2.1 The basic univariate 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 n-dimensional time series (the trajectory matrix $\mathbf{X}$) 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. In the univariate SSA, it is supposed that this procedure decomposes the original series $\{y(i)\}$ into a sum of several components and noise. Exceptional care must be taken when the trajectory matrix $\mathbf{X}$ is constructed from a time series possibly containing short-range correlated or nonlinear signals such as chaotic signals. The emergence of additional, linearly independent modes when the lags used in construction of the trajectory matrix are larger than the correlation length of such a signal has been discussed in [16].

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

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

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 [17]:

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

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).

2.2 Monte Carlo singular system analysis

The clear signal/noise distinction based on the eigenvalues $\sigma_1, \sigma_2, \ldots, \sigma_n$ can only be obtained in particularly idealized situation when the signal/noise ratio is large enough and the background consists of white noise. In many geophysical processes, however, so-called "red" noise with power spectrum of the $1/f^\alpha$ (power-law) type is present [11]. Its SSA eigenspectrum also has the $1/f^\alpha$ character [18], i.e., the eigenspectrum of 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 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 red noise might be neglected if its variance is smaller than the slow-mode eigenvalues of the background red noise. Therefore, Allen & Smith [11] proposed comparing the SSA spectrum of the analyzed signal with the SSA spectrum of a red-noise model fitted to the studied data. Such a red-noise process 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 Gaussian white noise with zero mean and unit variance.

In order to correctly detect a signal in red noise, we will apply the following approach, inspired by Allen & Smith [11]:

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 case of red noise (when the eigenspectra have naturally this form, as mentioned above).

Second, the eigenspectrum obtained from the studied data set is compared, in a frequency-by-frequency way, with the eigenspectra obtained from a set of realizations of an appropriate noise model (such as the AR(1) model (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) [11, 19, 20, 21]. Allen & Smith [11] also discuss other relevant approaches how to compare the eigenvalues from the tested data and the surrogates.

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 (MC-SSA), as coined by Allen & Smith [11] (see [11] 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 a white noise background [6, 10].

2.3 Enhanced MCSSA: testing dynamics of the SSA modes

The MCSSA described above is a sophisticated technique. However, it still assumes a very simple model, i.e., that the signal of interest has been linearly added to a specified background noise. Therefore the variance in the frequency band, characteristic for the searched signal, is significantly larger 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 independent of its (relative) variance, Paluš & Novotná [22] have proposed to test also dynamical properties of the SSA modes against the modes obtained from surrogate data. From this idea, the question arises how we can 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 at time t in some way depends on the state in which the system was at time $t - \tau$. The strength of such a dependence per 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 record of (a part of) the system's 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 with 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 information theory where the entropy rates are used to characterize information production by information sources [23].

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 information theory and generalized the notion of the entropy of an information source [24]. The Kolmogorov-Sinai entropy (KSE) [24, 25, 26] 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 [27].

Thus, the concept of entropy rates is common to theories based on philosophically opposite assumptions (randomness vs. determinism) and is ideally applicable for a 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 [28] has proposed "coarse-grained entropy rates" (CERs) instead. The CERs 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 the 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 [23]:

Let X and Y be random variables with sets of values Ξ and Υ, respectively, probability distribution functions (PDF) $p(x)$, $p(y)$, and a joint PDF $p(x, y)$. The entropy $H(X)$ of a single variable, say X, is defined as

$$H(X) = -\sum_{x \in \Xi} p(x) \log p(x), \tag{10}$$

and the joint entropy $H(X, Y)$ of X and Y is

$$H(X, Y) = -\sum_{x \in \Xi} \sum_{y \in \Upsilon} p(x, y) \log p(x, y). \tag{11}$$

The mutual information $I(X; Y)$ then can be expressed as

$$I(X; Y) = \sum_{x \in \Xi} \sum_{y \in \Upsilon} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}. \tag{12}$$

A detailed account on relations between entropy rates and information-theoretic functionals is given in [28, 29]. 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. We mark $x(t)$ as x and $x(t + \tau)$ as x_τ. For defining the simplest form of CER let us find τ_{max} such 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

$$||I(x; x_\tau)|| = \frac{\Delta\tau}{\tau_{max} - \tau_{min} + \Delta\tau} \sum_{\tau=\tau_{min}}^{\tau_{max}} I(x; x_\tau) \tag{13}$$

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

$$h^1 = I(x, x_{\tau_0}) - ||I(x; x_\tau)||. \tag{14}$$

It has been shown that the CER h^1 provides the same classification of states of chaotic systems as the exact KSE [28]. 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 (13) 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.

There are plenty of approaches to estimate the mutual information $I(x; x_\tau)$ [30]. If we are not interested in an exact value, but rather in a relative comparison of values obtained from the tested data and from the surrogate set, a simple box-counting approach based on marginal equiquantization [21, 28, 29] is satisfactory. The latter means that the marginal boxes (bins) are not defined equidistantly, but in a such a way that there is approximately the same number of data points in each marginal bin.

2.4 Implementation of the enhanced MCSSA

We realize the enhanced MC SSA as follows:

1. The studied time series undergoes SSA as briefly described above or in [31], i.e., using an embedding window of length n, the $n \times n$ lag-correlation matrix $\mathbf{C}$ is decomposed using the SVDCMP routine [32]. In the eigen-spectrum, 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 partic-ular EOF, the power spectrum of the projection (mode) is estimated, and the frequency bin with the highest power is identified. This spectral co-ordinate is mapped onto one of the n frequency bins, which equidistantly divide the abscissa of the eigenspectrum.

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

3. The surrogate data are generated using the above AR(1) model, where "scrambled" (randomly permutated in temporal order) residuals are used as innovations, i.e., the noise term $\gamma z(i)$ in (9) .

4. Each realization of the surrogates undergoes SSA as described in step 1. Then, the eigenvalues for the whole surrogate set are sorted in each frequency bin, and the values for the 2.5th and 97.5th percentiles are found. In the 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 above 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 such an eigenvalue as observed can emerge from the background of the null noise model.

6. For each SSA mode (a projection of the data onto a particular EOF), the regularity index is computed, as well as for each SSA mode for all the realizations of 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 [21, 28, 29].

Performing MCSSA using the embedding window of the length n, there are n eigenvalues in the eigenspectrum, and n statistical tests are done. Therefore, in general, the problem of simultaneous statistical inference should be consid-ered (see [21] and references therein). However, in many relevant applications 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), therefore we will not discuss this topic here.

Rejecting the null hypothesis of the AR(1) (or another appropriate) noise model, one can infer that there is "something more" in the data than a re-alization 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 pre- dictability than a mode obtained by linear filtration of the considered noise model.

3 Numerical examples

3.1 A signal in AR(1) background
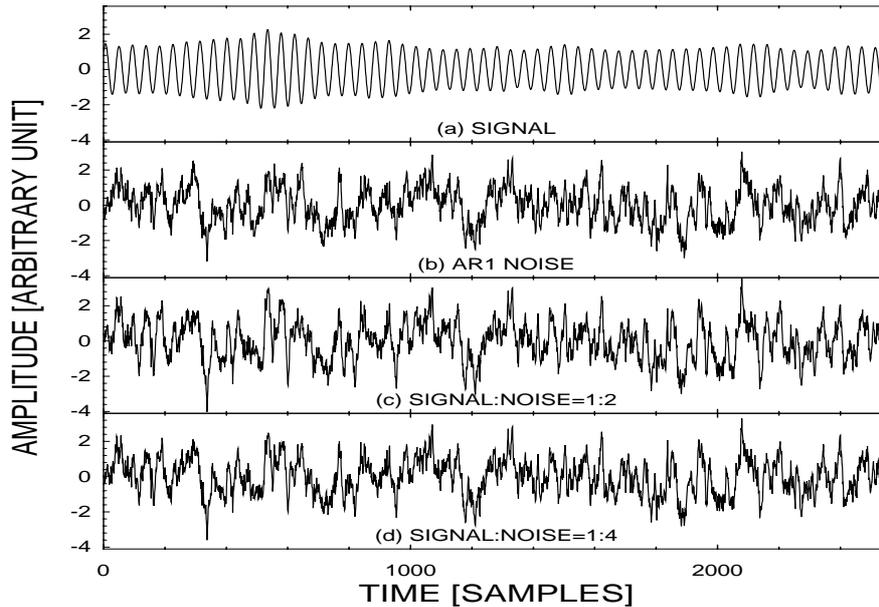
Fig. 1. Numerically generated test data: (a) A periodic signal with randomly vari- able amplitude was mixed with (b) a realization of an AR(1) process with a strong slow component, obtaining the signal to noise ratio 1:2 (c), and 1:4 (d). Adapted from Paluš & Novotná [31].

For an example of the application of the presented approach, let us consider numerically generated data data – a periodic signal with randomly variable amplitude (Fig. 1a) 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 zero mean and unit variance. The signal to noise ratios (i.e., the ratios of the respective standard deviations of signal and noise component) obtained by mixing the signals were 1:2 (Fig. 1c), and 1:4 (Fig. 1d). The latter two series are analyzed by the presented method.
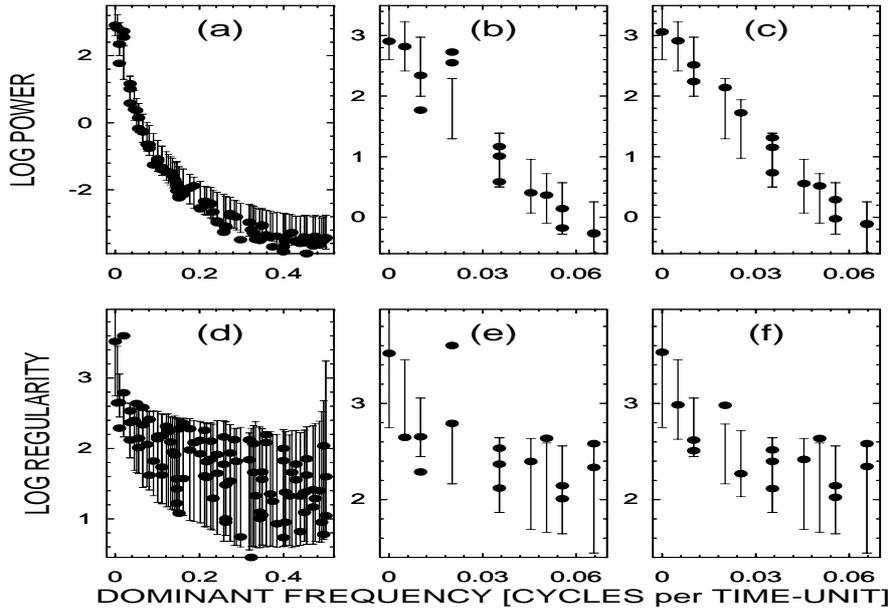
Fig. 2. The standard – eigenvalue based (a–c) and the enhanced – regularity index based (d–f) MCSSA analysis of the numerical data, presented in Fig. 1. (a) The full eigenspectrum and (b) the low-frequency part of the eigenspectrum – logarithms of eigenvalues ("LOG POWER") plotted according to the dominant 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. Adapted from Paluš & Novotná [31].

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 surrogate realizations (here, as well as in the following example). 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 $1/f^\alpha$ character of the surrogate eigenspectrum, i.e., the eigenvalues plotted against the 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 outside the range of those from the surrogates, i.e., they are statistically significant, the null hypothesis is rejected, and a signal not consistent with the null hypothesis is detected. A close look to the significant modes shows that they are related to the embedded 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 one. 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 to the mixture with the signal to noise ratio 1:2 (Figs. 2d,e), for one data eigenvalue, the 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, b. 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 a variance comparable to the former one (Figs. 2a,b), has its regularity index close to the 97.5th percentile of the surrogate regularity indices distribution. With other words, if a (nearly) periodic signal is embedded in a (colored) noise background, the SSA approach, in principle, is able to extract this signal together with some noise of neighboring 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 (null model), as measured by the regularity index (13). 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 was 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).
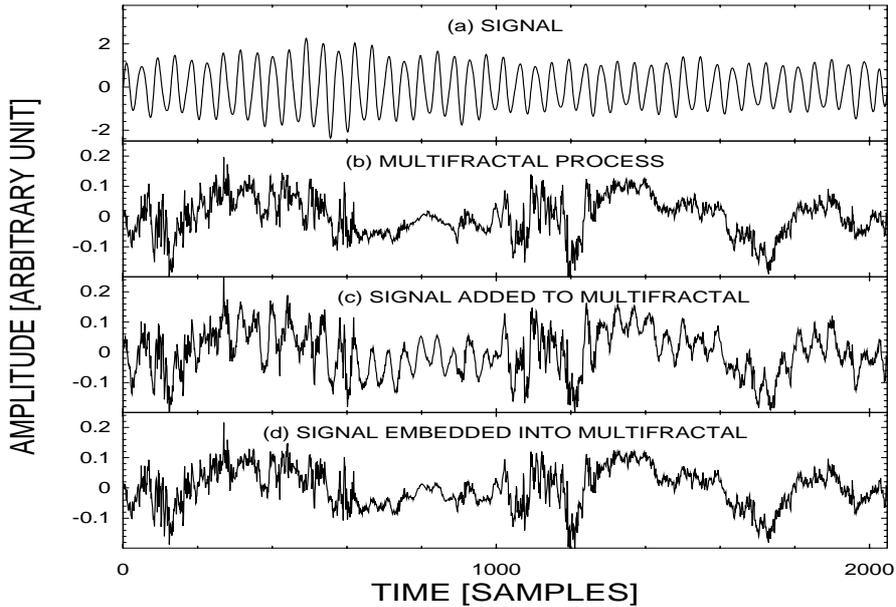
Fig. 3. Numerically generated test data: (a) The wavelet filtered signal from Fig. 1a was embedded into (b) a realization of a multifractal process, obtaining the ratio of related wavelet coefficients 1:2 (c), and 0.5:0.5 (d). Adapted from Paluš & Novotná [31].

3.2 A signal in multifractal background

As a more complex example we "embed" the test signal (Fig. 1a) into a realization of a multifractal process (Fig. 3b) generated by a log-normal random cascade on a wavelet dyadic tree [33] using the discrete wavelet transform [32]. Using wavelet decomposition, we embed 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 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 twice 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. In the second case, we adjusted the SD of both sets of wavelet coefficients to 50% of the SD of the wavelet coefficients of the original multifractal signal in the associated scale (Fig. 3d), so that the total variance in this scale (frequency band) does not exceed the corresponding variance of the "clean" multifractal. Then, it is not surprising, that the variance-(eigenvalues)-based MCSSA test, using the AR(1) surrogate data (Fig. 4a,b), 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 outside 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 (Fig. 4a,b). 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 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.
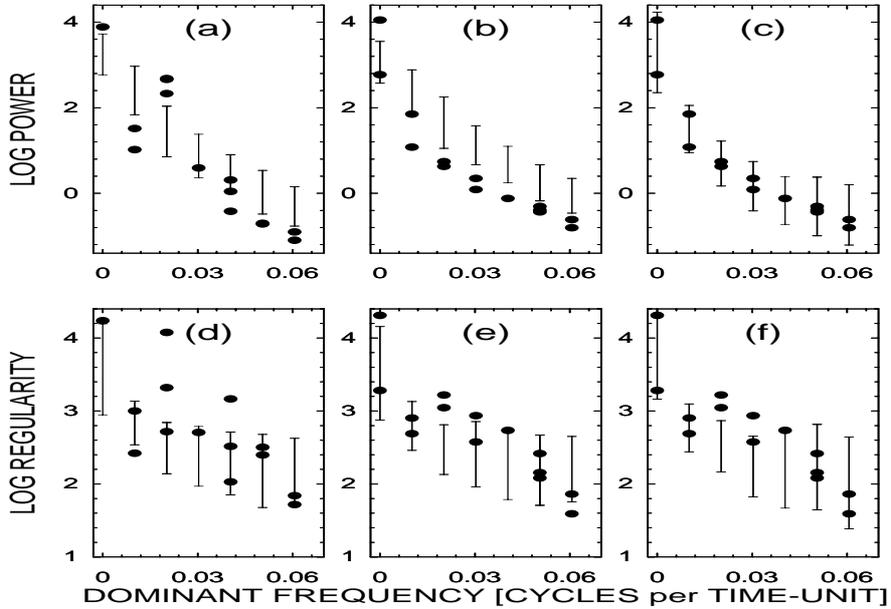
Fig. 4. The low frequency parts of the MCSSA eigenspectra (a–c) and regularity spectra (d–f) for the signal embedded into a multifractal process with wavelet coefficient ratio 1:2 (a,d) and 0.5:0.5 (b,c,e,f). Bursts – eigenvalues or regularity indices for the analysed data; bars – 95% of the surrogate eigenvalues or regularity index distribution obtained from the AR(1) (a,b,d,e) and the multifractal (c,f) surrogate data. Adapted from Paluš & Novotná [31].

4 Detection of irregular oscillations in geophysical data

Temperature measurements are among the longest available instrumental data characterizing the long term evolution of the atmosphere and climate in a particular location. For instance, the data from the Prague–Klementinum station are available since 1775. On the other hand, large-scale circulation patterns reflect a more global view on the atmospheric dynamics. The North Atlantic Oscillation (NAO) is a dominant pattern of atmospheric circulation variability in the extratropical Northern Hemisphere, accounting for about 60% of the total sea-level pressure variance. The NAO has a strong effect on European weather conditions, influencing meteorological variables including the temperature [34]. The NAO – temperature relationship, however, is not straightforward and its mechanism is not yet fully understood.

The possible influence of the solar variability on the climate change has been a subject of research for many years, however, there are still open questions and unsolved problems (for reviews, see e.g. [35, 36, 37]). Probably the longest historical record of the solar variability are the well-known sunspot numbers. After the sunspot numbers, aa index, the time series of the geomagnetic activity provides the longest data set of solar proxies which goes back to 1868 [38]. Since there are no direct measurements of solar irradiance available until the beginning of the 1980's, the data of geomagnetic variations are used for an additional study of solar activity, especially of irradiance.

It might be interesting if the atmospheric data, both the local and global, and the geomagnetic and solar data possess any common, repeating variability pattern such as cycles or oscillatory modes. The enhanced MCSSA can give an answer to such a question.

4.1 The data

The NAO index is traditionally defined as the 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`.

Monthly average near-surface air temperature time series from ten European stations were used (see [31] for details), obtained from the Carbon Dioxide Information Analysis Center Internet server (`ftp://cdiac.esd.ornl.gov/pub/ndp041`) as well as a time series from the Prague–Klementinum station from the period 1781 – 2002. The long-term monthly averages were subtracted from the data, so that the annual cycle was effectively filtered out.

The aa-index is defined by the average, for each 3-hour period, of the maximum of magnetic elements from two near-antipodal mid-latitude stations in Australia (Melbourne) and England (Greenwich). The data spanning the period 1868–2005 were obtained from World Data Centre for Solar-Terrestrial Physics, Chilton, `http://www.ukssdc.ac.uk/data/wdcc1/wdc_menu.html`.

The monthly sunspot data, spanning the period 1749–2006, has been obtained from the SIDC-team, Royal Observatory of Belgium, Ringlaan 4, 1180 Brussels, Belgium, `http://sidc.oma.be/DATA/monthssn.dat`.

4.2 The results
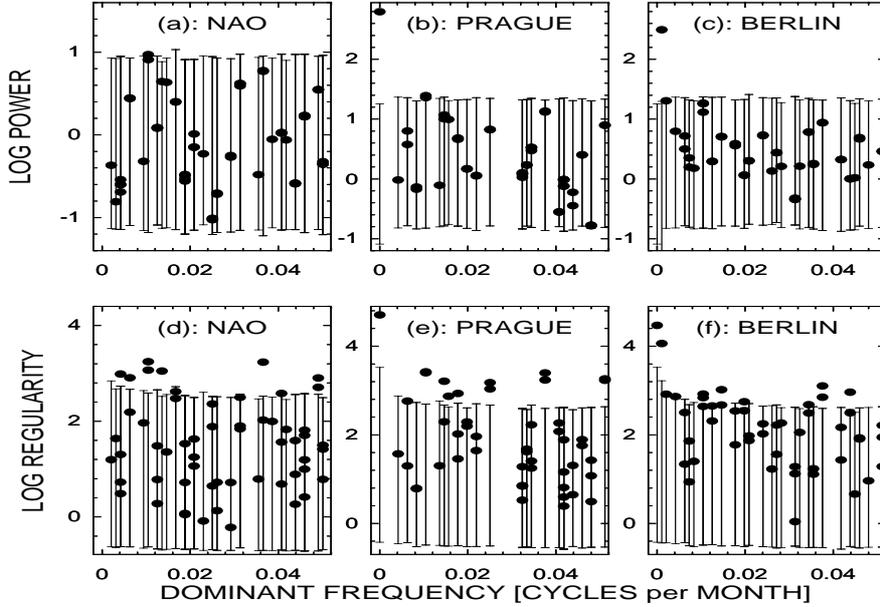
Fig. 5. Enhanced MCSSA analysis of the monthly NAO index (a,d) and monthly average near-surface air temperature series from Prague–Klementinum (b,e) and Berlin (c,f). Low-frequency parts of eigenspectra – logarithms of eigenvalues ("LOG POWER") (a,b,c) and regularity index spectra (d,e,f). Bursts – eigenvalues 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. The datasets span the period 1824–2002, the embedding dimension $n = 480$ months was used.

Figure 5 presents the results from the enhanced MCSSA for the considered monthly NAO index and the monthly average near-surface air temperature time series from Prague (Prague–Klementinum station) and Berlin, obtained using the embedding dimension $n = 480$ months. In the standard MCSSA, the only eigenvalue undoubtedly distinct from the surrogate range is the trend (zero frequency) mode in the temperature (Figs. 5b,c). Further, there are two modes at the frequency 0.0104 just above the surrogate bar in the Prague temperature and NAO test (Figs. 5a,b). These results, however, are still "on the edge" of significance and are not very convincing. In the case of Berlin,

the eigenvalues of the modes at the frequency 0.0104 are confined within the surrogate range (Fig. 5c).

A quite different picture is obtained from the analyses based on the regularity index (Figs. 5d,e,f). Several oscillatory modes have been detected with a high statistical significance. The distinction of the regularity indices of these modes from the related surrogate ranges is clear and even the simultaneous statistical inference cannot jeopardize the significance of the results. The significant modes in the NAO are located at the frequencies (in cycles per month) 0.004, 0.006, 0.0104, 0.014, 0.037 and 0.049, corresponding to the periods of 240, 160, 96, 73, 27 and 20 months, respectively. Besides the zero frequency (trend) mode, the significant modes in the Prague temperature are located at the frequencies 0.0104, 0.014, 0.016, 0.018, 0.025, 0.037 and 0.051, corresponding to the periods of 96, 68, 64, 56, 40, 27 and 20 months, respectively. In the case of Berlin, there are some differences, namely the modes with the periods 20, 40, 56 and 64 months are missing, while modes with periods 23, 29 and 58 months, as well as a slow mode next to the zero frequency mode appeared. The significant modes with the periods 27, 68 and 96 months were detected in both the records.

The modes with a period of 8 years were extracted and analysed in [31], their mean frequency was estimated with higher precision as 7.8 years. Besides the latter modes (and the trend mode in the temperature), the highest regularity index was obtained for the modes with a period of 27 months (frequency 0.037). This frequency lies within the range of the quasi-biennial oscillations (QBO). The behavior of these modes was studied in some detail in [39].
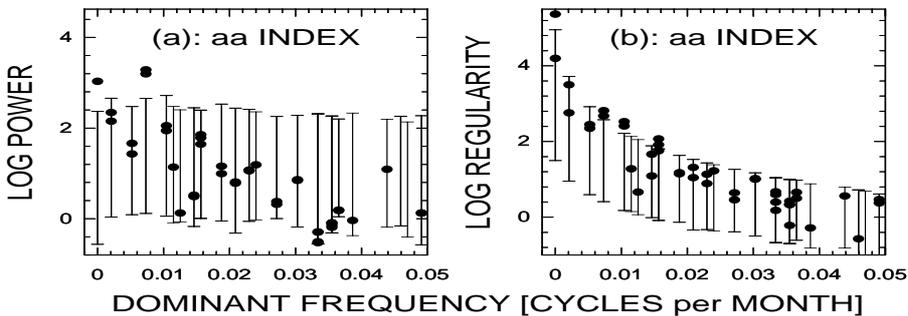
Fig. 6. Enhanced MCSSA analysis of the monthly aa index. The low-frequency part of the eigenspectrum – logarithms of eigenvalues ("LOG POWER") (a) and the regularity index spectrum (b). Bursts – eigenvalues or regularity indices for the analysed data; bars – 95% of the surrogate eigenvalues or regularity index distribution, i.e., the bar is drawn from the 2.5th to the 97.5th percentiles of the surrogate eigenvalues/regularity indices distribution. The dataset spans the period 1868–2005, the embedding dimension $n = 480$ months was used.

The results of the enhanced MCSSA analysis of the aa index are presented in Fig. 6. In the standard (eigenvalue) analysis (Fig. 6a), we can see significant modes representing the trend, i.e., the zero frequency mode, and a mode with a frequency of 0.0073 which corresponds to the period of 136 months, i.e. to the 11-year solar activity cycle. The analysis based on the regularity index (Fig. 6b) confirms the previous two modes and adds two more ones on frequencies of 0.0104 and 0.016, corresponding to periods of 96 and 64 months.
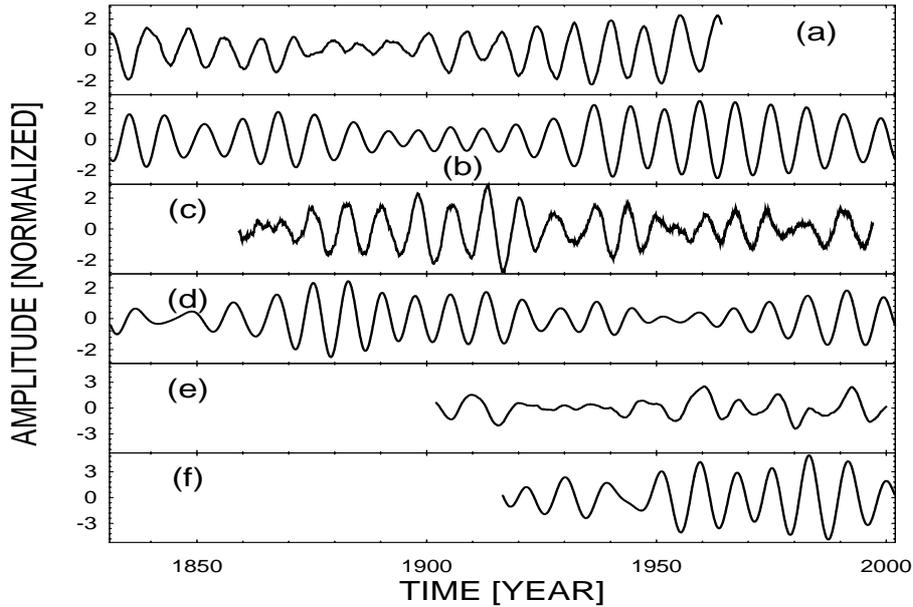
Fig. 7. The oscillatory modes with the approximately 8-year period extracted by using SSA (a,c,e) and CCWT (b,d,f) from the near-surface air temperature (a,b), the NAO index (c,d), and the aa index (e,f).

The mode with the period of 96 months or 8 years has been detected in all the above analyzed data sets, i.e., in the near-surface air temperature, the NAO index, and the geomagnetic aa index. The time series of the modes extracted using the SSA, i.e., by projecting the input data on the particular EOF, are presented in Figs. 7a,c,e. When the modes are extracted using SSA, there is an uncertainty of timing of the modes given by the embedding window, and a part of the data equal to the embedding window is lost. We positioned the SSA modes on the time axis by maximizing the cross-correlation between the mode and the original data. This approach, however, does not always give unambiguous results. Therefore, Paluš & Novotná [39] studied the possible relationships of the QBO modes from the temperature and NAO index not only using the SSA-extracted modes, but also using modes extracted

from the data by means of complex continuous wavelet transform (CCWT) [40]. Here we compare the modes with the period 96 months extracted by SSA (Figs. 7a,c,e) with the modes obtained by using CCWT with the central wavelet frequency set to the period of 96 months (Figs. 7b,d,f). The SSA mode and the wavelet mode, obtained from the Prague temperature data (Fig. 7a and Fig. 7b, respectively) are shifted by π (a half of the period), otherwise their agreement is very good. The timing of the SSA and CCWT modes from the NAO index (Fig. 7c and Fig. 7d, respectively) is consistent, however, the wavelet transform performs stronger smoothing. In the aa index, the CCWT mode is smoother and slightly shifted in time in comparison with the SSA mode (Fig. 7f and Fig. 7e, respectively).
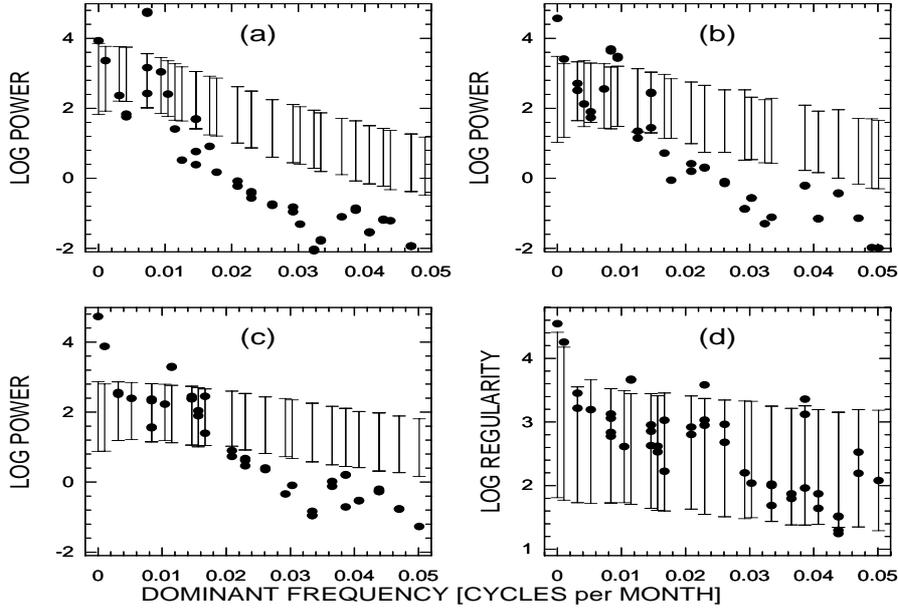
Fig. 8. Enhanced MCSSA analysis of the monthly sunspot data. Low-frequency parts of eigenspectra – logarithms of eigenvalues ("LOG POWER") for the raw sunspot data (a), the sunspot data after removal of the mode with the period 136 months (b), and for the sunspot data after removal of the modes with the periods 136, 120 and 106 months (c). (d): Low-frequency part of the regularity index spectrum for the sunspot data after removal of the modes with the periods 136, 120 and 106 months. 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. The dataset spans the period 1749–2006, the embedding dimension $n = 480$ months was used.

Analyzing the monthly sunspot data, the only clear significance in both the eigenspectrum (Fig. 8a) and the regularity index spectrum is the mode with a period of 136 months. The long-term trend at the zero-frequency mode lies at the edge of significance (Fig. 8a). After removal of the 136 month mode and subsequent analysis of the data residuals, the zero frequency mode becomes highly significant and another slow mode, with a period about 80 years emerges. Two new significant modes, related to the 11-year solar cycle appear on the frequency bins following the frequency bin of the previously defined mode with the period 136 months. Their periods are 120 and 106 months (Fig. 8b). After removal of all three modes (i.e., the modes with the periods 136, 120 and 106 months) which can be considered as a decomposition of the 11-year cycle, the standard MCSSA analysis of the sunspot data residuals uncovers another interesting oscillatory mode in the frequency bin corresponding to a period of 7.4 years (Fig. 8b). The enhanced MCSSA analysis of the sunspot data residuals confirms all the modes from the standard MCSSA (zero frequency and period 80 and 7.4 years) and adds two new significant modes with the periods of 43.5 and 26 months (Fig. 8d).
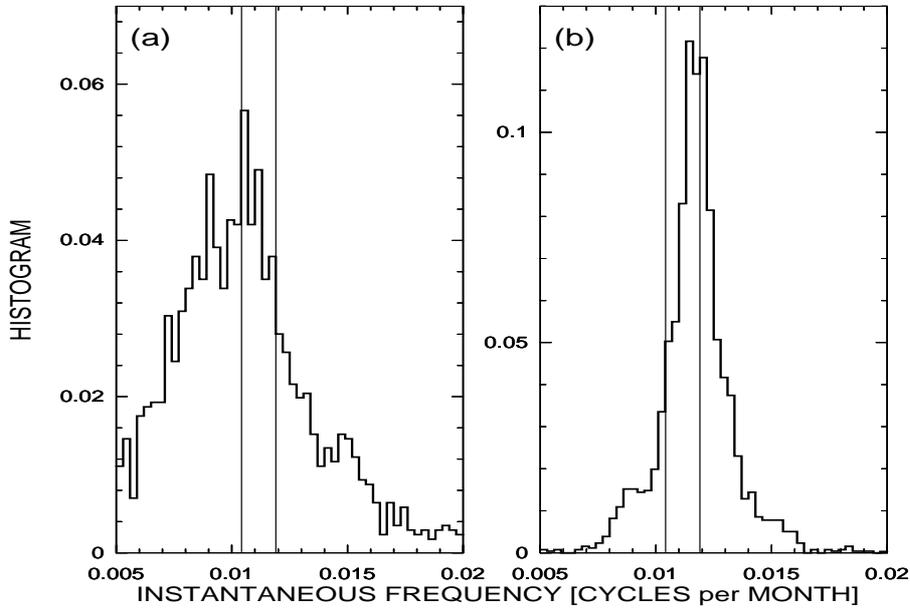
Fig. 9. Histograms of the instantaneous frequencies of the 7.8 yr temperature mode (a) and the 7.4 yr sunspot mode (b). The thin vertical lines mark the frequencies corresponding to the period of 8 and 7 years, reading from the left to the right side.

It is important to note that the frequency or period accuracy of the SSA approach is limited by the number of frequency bins given by the embedding

dimension. The accuracy of the frequency (or the period) of a particular mode can be increased after the extraction of this mode from the original data and its subsequent spectral or autocorrelation analysis, as Paluš & Novotná [22, 31] have done for the temperature mode. On the other hand, oscillatory modes from natural processes are never strictly periodic and their frequency is variable. We illustrate this variability by presenting histograms of instantaneous frequencies of the two close modes – the mode with the period 7.8 yr from the Prague temperature (Fig. 9a), and the period 7.4 yr mode obtained from the sunspot data residuals after modes related to the 11 yr cycle have been previously removed (Fig. 9b). The instantaneous frequencies were obtained by differentiation of the instantaneous phases [41, 42]. The latter can easily be computed by applying the analytic signal approach to the two orthogonal (shifted by $\pi/2$) components of each oscillatory mode, see Refs. [39, 43] for details. Thus the presented histograms are not necessarily equivalent to the power (Fourier) spectra, but they better reflect possibly nonstationary fluctuations of the frequencies of the modes. We can see that the most probable period of the sunspot mode is 7.4 years, with the slight tendency to higher frequencies (Fig. 9b), while in the case of the temperature mode, the most probable period is 7.8 years, with considerable weight on slower frequencies (Fig. 9a). There is, however, a great deal of common frequencies of the two modes, giving thus the possibility of interactions during some time intervals.

Considering both the available accuracy and the natural variability of the frequency of the detected oscillatory modes, the periods given here should be understood as limited accuracy estimates of average periods of particular modes.

The common occurrence of the oscillatory modes with the periods of approximately 11, 5.5, and 2.2 years and in the range $7 - 8$ years in the sunspot numbers, the aa index, the near-surface air temperature and the NAO index is summarized in Tab. 1.

Source data	Period [years]			
	≈ 11	$7 - 8$	≈ 5.5	≈ 2.2
sunspots	+	+	−	+
aa	+	+	+	−
T	−	+	+	+
NAO	−	+	−	+

Table 1. Occurrence of the most significant oscillatory modes with periods of approximately 11, $7 - 8$, 5.5 and 2.2 years in the sunspot numbers, the aa index, the average near-surface air temperature and the NAO index.
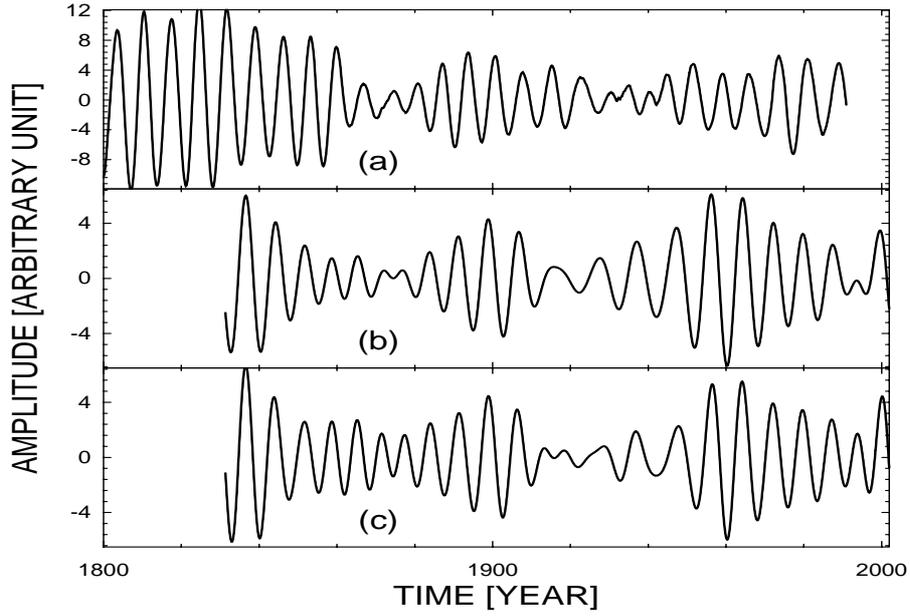
Fig. 10. The oscillatory mode with the approximately 7.4 yr period obtained from the sunspot data residuals after previously removed modes related to the 11 yr cycle, extracted by using SSA (a) and CCWT with the central wavelet frequency corresponding to periods 8 yr (b) and 7.4 yr (c).

We can see that the modes with a period in the range $7-8$ years have been detected in all the analysed datasets. These modes, obtained from the near-surface air temperature, from the NAO index and the geomagnetic aa index have already been presented in Fig. 7, the related modes from the sunspot data are illustrated in Fig. 10. Again, we can compare the mode extracted by SSA in the natural EOF base (Fig. 10a) with the modes obtained by CCWT with the Morlet basis [40], using two close central wavelet frequencies corresponding to the periods 8 yr (Fig. 10b) and 7.4 yr (Fig. 10c). We can see that the wavelet extracted modes have a more limited frequency range and the wavelets with different central frequency are able to better fit the mode shapes in different temporal segments dominated by different frequencies.

5 Discussion and Conclusion

The Monte Carlo Singular System Analysis has been extended by evaluating and testing the regularity of the dynamics of the SSA modes against the colored noise null hypothesis in addition to the test based on variance (eigenvalues). The nonlinear approach to the measurement of regularity and predictability of the dynamics, based on a coarse-grained estimate of the mu-

tual information, gives a possibility to detect dynamical modes which are more regular than those obtained by decomposition of colored noise. Using numerical examples, we have demonstrated that such an enhanced MCSSA test is more sensitive in detection of oscillatory modes hidden in a noisy background. There are, however, some facts about accuracy and consistency of the results which should not be neglected. Already in the previous section, we have discussed the accuracy of the estimation of the period of detected oscillatory modes. We have stated that we are only able to provide a limited accuracy estimate of an average period or frequency, since the frequency of oscillatory modes in the studied natural phenomena is variable. One should keep this fact in mind in comparisons of results found in the literature. Not only frequency, but also the relative variance and the regularity of the oscillatory modes is variable. Due to this nonstationary behaviour, any conclusion about the existence and significance of a mode is dependent on the temporal range of analysed data. Obtained eigenvalues and regularity indices give an average quantification of the relative variance and regularity, respectively, for the analysed time span of the data. It is possible that in some data segments, the results can change. Thus it is reasonable to combine the MCSSA analysis with a wavelet analysis, using the latter one as an exploratory tool and the former one as a hypothesis testing tool.

Another important question is that of the relevance of the used null hypothesis. While in many cases the simple AR(1) process seems to work satisfactorily, for instance, in the case of the sunspot numbers, it is not generally appropriate. In this case, the AR(1) process does not fit the long-range dependence in the data, but the short-range correlation inside the 11yr cycle. As a consequence, the covariance structures of the data and the null noise model are not consistent (see Figs. 8a, b, c where the surrogate bars overestimate the data eigenvalues). The situation is improved after removal of the modes related to the 11yr cycle, and especially, in the case of the regularity test, the null hypothesis seems to be consistent with the noise part of the data (Fig. 8d). In the further development of the MCSSA, it is desirable to consider also more sophisticated null hypotheses including long-range correlated, fractal and multifractal models, since such properties have been observed in geophysical data, especially in the long-term air temperature records [44, 45].

The enhanced MCSSA has been applied to records of monthly average near-surface air temperature from several European locations, to the monthly NAO index, as well as to the monthly aa index and the sunspot numbers. A number of significant oscillatory modes have been detected in all the different source data, some of them with common periods (Tab. 1). While the 11yr solar cycle is shared by the solar and geomagnetic data, the quasi-biennial mode is present in the atmospheric data and also in the solar data. The mode with the period in the range 7 – 8 years is present in all the analysed data, i.e., in the atmospheric temperatures, in the NAO index, in the aa index and in the sunspot numbers.

It is interesting to note that the oscillatory mode with a period of 7.8 years has been detected in the NAO, in the Arctic Oscillation (AO), in the Uppsala winter near-surface air temperature, as well as in the Baltic Sea ice annual maximum extent by Jevrejeva and Moore [46]. Applying MCSAA on the winter NAO index, Gámiz-Fortis *et al.* [47] detected oscillations with the period 7.7 years. Moron *et al.* [48] have observed oscillatory modes with the period about 7.5 years in the global sea surface temperatures. Da Costa and de Verdiere [49] have detected oscillations with the period 7.7 years in interactions of the sea surface temperature and the sea level pressure. Unal and Ghil [50] and Jevrejeva *et al.* [51] observed oscillations with periods 7 – 8.5 years in a number of sea level records. Feliks and Ghil [52] report the significant oscillatory mode with the 7.8 year period in the Nile River record, Jerusalem precipitation, tree rings and in the NAO index. Our first application of the enhanced MCSSA [22] yielded the observation of the mode with the period 7.8 years in near-surface air temperature from several European locations. Recently, the enhanced MCSSA analyses of the temperature data were refined and the analysis of the NAO index was added [31]. In the present work the number of processes containing the oscillatory mode with the period in the range 7 – 8 years was extended by the geomagnetic activity aa index and the sunspot numbers.

These findings give a solid basis for further research of relations among the dynamics reflected in the analysed data and thus between the solar and geomagnetic activity and the climate variability. The existence of oscillatory modes open the possibility to apply the recently developed synchronization analysis [53, 54] which already has found successful applications in studies of relations between atmospheric phenomena. Maraun & Kurths [55] discovered epochs of phase coherence between El Niño/Southern Oscillation and Indian monsoon, while Paluš & Novotná [39] demonstrated phase synchronization or phase coherence between the above mentioned QBO modes extracted from the temperature and the NAO index. The analysis of instantaneous phases of oscillatory processes allows to detect very weak interactions [53] and also causality relations if one oscillatory process drives the other one [56, 57]. In such analysis, Mokhov & Smirnov [58] have demonstrated that the NAO interacts with (or is influenced by) the other main global atmospheric oscillatory process – the El Niño Southern Oscillation. We believe that the synchronization analysis will help uncovering the mechanisms of the tropospheric responses to the solar and geomagnetic activity and contribute to a better understanding of the solar-terrestrial relations and their role in climate change.

Acknowledgements

References

1. J.B. Elsner, A.A. Tsonis, *Singular Spectrum Analysis. A New Tool in Time Series Analysis.* Springer, Berlin (1996)
2. N. Golyandina, V. Nekrutkin, A. Zhigljavsky, *Analysis of Time Series Structure. SSA and Related Techniques*, Chapman & Hall/CRC, Boca Raton (2001)
3. M.R. Ghil, M. Allen, M.D. Dettinger, K. Ide, D. Kondrashov, M.E. Mann, A.W. Robertson, A. Saunders, Y. Tian, F. Varadi, P. Yiou, Advanced spectral methods for climatic time series. *Rev. Geophys.*, 40, 3-11–3-13 (2002)
4. M.R. Allen, A.W. Robertson, Distinguishing modulated oscillations from coloured noise in multivariate datasets. *Climate Dynamics*, 12, 775–784 (1996)
5. R. Vautard, M. Ghil, Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, *Physica D*, 35, 395–424 (1989)
6. M. Ghil, R. Vautard, Interdecadal oscillations and the warming trend in global temperature time series. *Nature*, 350(6316), 324–327 (1991)
7. C.L. Keppenne, M. Ghil, Adaptive filtering and the Southern Oscillation Index. *J. Geophys. Res.*, 97, 20449–20454 (1992)
8. P. Yiou, M. Ghil, J. Jouyel, D. Paillard, R. Vautard, Nonlinear variability of the climatic system, from singular and power spectra of Late Quarternary records. *Clim. Dyn.*, 9, 371–389 (1994)
9. M.R. Allen, L.A. Smith, Investigatings the origins and significance of low-frequency modes of climate variability. *Geophys. Res. Lett.*, 21, 883–886 (1994)
10. R. Vautard, P. Yiou, M. Ghil, Singular spectrum analysis: a toolkit for short noisy chaotic signals. *Physica D*, 58, 95–126 (1992)
11. M.R. Allen, L.A. Smith, Monte Carlo SSA: Detecting irregular oscillation in the presence of colored noise. *J. Climate*, 9(12), 3373–3404 (1996)
12. W. W. Hsieh, Nonlinear multivariate and time series analysis by neural network methods. *Rev. Geophys.*, 42, RG1003 (2004)
13. S. S. P. Rattan, W.W. Hsieh, Nonlinear complex principal component analysis of the tropical Pacific interannual wind variability. *Geophys. Res. Lett.*, 31(21), L21201 (2004)
14. W. W. Hsieh, Nonlinear principal component analysis of noisy data. *Neural Networks*, 20, 434-443 (2007)
15. W. W. Hsieh, A. J. Cannon, Towards robust nonlinear multivariate analysis by neural network methods. This volume.
16. M. Paluš, I. Dvořák, Singular-value decomposition in attractor reconstruction: pitfalls and precautions. *Physica D* 55, 221–234 (1992).
17. D.S. Broomhead, G.P. King, Extracting qualitative dynamics from experimental data. *Physica D*, 20, 217–236 (1986)
18. J.B. Gao, Y. Cao, J.-M. Lee, Principal component analysis of $1/f^{\alpha}$ noise. *Phys. Lett. A*, 314, 392–400 (2003)
19. L.A. Smith, Identification and prediction of low-dimensional dynamics. *Physica D*, 58, 50–76 (1992)
20. J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, J.D. Farmer, Testing for nonlinearity in time series: the method of surrogate data. *Physica D* 58, 77–94 (1992)

21. M. Paluš, Testing for nonlinearity using redundancies: Quantitative and qualitative aspects. *Physica D*, 80, 186–205 (1995)
22. M. Paluš, D. Novotná, 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)
23. T.M. Cover, J.A. Thomas, *Elements of Information Theory.* J. Wiley & Sons, New York (1991)
24. Ya.G. Sinai, *Introduction to Ergodic Theory.* Princeton University Press, Princeton (1976)
25. I.P. Cornfeld, S.V. Fomin, Ya.G. Sinai, *Ergodic Theory.* Springer, New York (1982)
26. K. Petersen, *Ergodic Theory*, Cambridge University Press, Cambridge (1983)
27. Ya.B. Pesin, Characteristic Lyapunov exponents and smooth ergodic theory. *Russian Math. Surveys*, 32, 55–114 (1977)
28. M. Paluš, Coarse-grained entropy rates for characterization of complex time series. *Physica D*, 93, 64–77 (1996)
29. M. Paluš, Kolmogorov entropy from time series using information-theoretic functionals. *Neural Network World*, 7(3), 269–292 (1997) (`http://www.cs.cas.cz/~mp/papers/rd1a.pdf`)
30. K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, J. Bhattacharya, Causality detection based on information-theoretic approaches in time series analysis. *Phys. Rep.*, 441, 1–46 (2007).
31. M. Paluš, D. Novotná, Enhanced Monte Carlo Singular System Analysis and detection of period 7.8 years oscillatory modes in the monthly NAO index and temperature records. *Nonlin. Proc. Geophys.*, 11, 721–729 (2004)
32. W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes: The Art of Scientific Computing.* Cambridge Univ. Press, Cambridge (1986)
33. A. Arneodo, E. Bacry, J.F. Muzy, Random cascades on wavelet dyadic trees. *J. Math. Phys.*, 39(8), 4142–4164 (1998)
34. J.W. Hurrell, Y. Kushnir, M. Visbeck, Climate - The North Atlantic oscillation. *Science*, 291(5504), 603 (2001)
35. D. Rind, The Sun's Role in Climate Variations. *Science*, 296, 673–677 (2002)
36. E. Bard, M. Frank, Climate change and Solar variability: What's new under the sun? *Earth Planet. Sci. Lett.*, 248(1-2), 1–14 (2006)
37. R.P. Kane, Sun-Earth relation: Historical development and present status- A brief review. *Advances in Space Research*, 35(5), 866–881 (2005)
38. P.N. Mayaud, The aa indices: a 100year series characterizing the magnetic activity. *J. Geophys. Res.*, 77(34), 6870–6874 (1972)
39. M. Paluš, D. Novotná, Quasi-Biennial Oscillations extracted from the monthly NAO index and temperature records are phase-synchronized. *Nonlin. Proc. Geophys.* 13, 287–296 (2006)
40. C. Torrence, G.P. Compo, A practical guide to wavelet analysis. *Bull. Amer. Meteorological Society*, 79(1), 61–78 (1998)
41. M. Paluš, D. Novotná, Sunspot cycle: a driven nonlinear oscillator? *Phys. Rev. Lett.*, 83, 3406-3409 (1999)
42. M. Paluš, J. Kurths, U. Schwarz, N. Seehafer, D. Novotná, I. Charvátová, The solar activity cycle is weakly synchronized with the solar inertial motion. *Phys. Lett. A*, 365, 412–428 (2007)

43. M. Paluš, D. Novotná, P. Tichavský, Shifts of seasons at the European mid-latitudes: Natural fluctuations correlated with the North Atlantic Oscillation. *Geophys. Res. Lett.*, 32, L12805 (2005)

44. R. B. Govindan, D. Vjushin, S. Brenner, A. Bunde, S. Havlin, H.-J. Schellnhuber, Long-range correlations and trends in global climate models: Comparison with real data. *Physica A*, 294(1-2), 239–248 (2001)

45. R. B. Govindan, A. Bunde, S. Havlin, Volatility in atmospheric temperature variability. *Physica A*, 318(3-4) 529–536 (2003)

46. S. Jevrejeva, J. C. Moore, Singular Spectrum Analysis of Baltic Sea ice conditions and large-scale atmospheric patterns since 1708. *Geophys. Res. Lett*, 28(23), 4503–4506 (2001)

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

48. V. Moron, R. Vautard, M. Ghil, Trends, interdecadal and interannual oscillations in global sea surface temperatures. *Clim. Dyn.*, 14, 545 – 569 (1998)

49. E. D. da Costa, A. C. de Verdiere, The 7.7 year North Atlantic Oscillation. *Q. J. R. Meteorol. Soc.*, 128, 797 – 817 (2002)

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

51. S. Jevrejeva, A. Grinsted, J. C. Moore, S. Holgate, Nonlinear trends and multiyear cycles in sea level records. *J. Geophys. Res.* 111, C09012 (2006)

52. Y. Feliks, M. Ghil, Interannual, synchronized oscillations over the Norh Atlantic, Eastern Mediterranean and Ethiopian Plateau. *Geophysical Research Abstracts* 9, 05600 (2007)

53. A. Pikovsky, M. Rosenblum, J. Kurths, *Synchronization. A Universal Concept in Nonlinear Sciences.* Cambridge University Press, Cambridge (2001)

54. M. Paluš, Detecting phase synchronization in noisy systems. *Phys. Lett. A*, 235, 341–351 (1997)

55. D. Maraun, J. Kurths, Epochs of phase coherence between El Niño/Southern Oscillation and Indian monsoon. *Geophys. Res. Lett.*, 32(15), L15709 (2005)

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

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

58. I.I. Mokhov, D.A. Smirnov, El Niño-Southern Oscillation drives North Atlantic Oscillation as revealed with nonlinear techniques from climatic indices. *Geophys. Res. Lett.*, 33, L03708 (2006)