Advances in Adaptive Data Analysis
Vol. 6, No. 4 (2014) 1450011 (34 pages)

DOI: 10.1142/S1793536914500113

World Scientific
www.worldscientific.com

SINGULAR SPECTRUM DECOMPOSITION: A NEW METHOD FOR TIME SERIES DECOMPOSITION

PIETRO BONIZZI* and JOËL M. H. KAREL†

Department of Knowledge Engineering
Maastricht University, P.O. Box 616
Maastricht, 6200 MD, The Netherlands
**pietro.bonizzi@maastrichtuniversity.nl*
†*joel.karel@maastrichtuniversity.nl*

OLIVIER MESTE

Lab. I3S, UNS CNRS, UMR 7271
06900 Sophia Antipolis, France
meste@i3s.unice.fr

RALF L. M. PEETERS

Department of Knowledge Engineering
Maastricht University, P.O. Box 616
Maastricht, 6200 MD, The Netherlands
ralf.peeters@maastrichtuniversity.nl

Received 3 February 2014
Revised 17 June 2014
Accepted 21 June 2014
Published 24 September 2014

This study introduces singular spectrum decomposition (SSD), a new adaptive method for decomposing nonlinear and nonstationary time series in narrow-banded components. The method takes its origin from singular spectrum analysis (SSA), a nonparametric spectral estimation method used for analysis and prediction of time series. Unlike SSA, SSD is a decomposition method in which the choice of fundamental parameters has been completely automated. This is achieved by focusing on the frequency content of the signal. In particular, this holds for the choice of the window length used to generate the trajectory matrix of the data and for the selection of its principal components for the reconstruction of a specific component series. Moreover, a new definition of the trajectory matrix with respect to the standard SSA allows the oscillatory content in the data to be enhanced and guarantees decrease of energy of the residual. Through the numerical examples and simulations, the SSD method is shown to be able to accurately retrieve different components concealed in the data, minimizing at the same time the generation of spurious components. Applications on time series from both the biological and the

*Corresponding author.

physical domain are also presented highlighting the capability of SSD to yield physically meaningful components.

Keywords: Singular spectrum analysis; principal component analysis; nonstationary time series; empirical mode decomposition.

1. Introduction

The application of decomposition methods to time series is an important analysis step allowing patterns and behaviors to be extracted as components providing insight into the mechanisms producing the time series. Of foremost importance is that the decomposition methods provide components which are physically meaningful. Unfortunately, actual data may be most likely too short, nonstationary (the time–frequency properties of the time series change over time), and they may represent nonlinear processes (processes for which the superposition principle does not hold, and the output cannot be represented by a linear combination of the inputs) [Huang *et al.* (1998)], and in many cases standard techniques like Fourier analysis provide for these data results with little physical sense. During the last decade, empirical mode decomposition (EMD) has been proposed as a new adaptive time–frequency data analysis methods for noisy nonlinear and nonstationary processes [Huang *et al.* (1998)]. However, components retrieved by EMD may be affected by mode mixing. That means, either a single component consisting of distinct scales (representing different behaviors), or similar scales residing in different components, making physical meaning of individual components unclear. Mode mixing is often a consequence of signal intermittency, a frequent ingredient in real time series. A different method which has been shown suitable to analyze short and noisy time series is the singular spectrum analysis (SSA or also "Caterpillar" SSA). SSA is a principal component analysis-based nonparametric spectral estimation method which uses no prior knowledge about the underlying physics or biology of the system [Vautard and Ghil (1989); Vautard *et al.* (1992); Ghil *et al.* (2002)]. The advantage of SSA-derived components over sines and cosines from Fourier analysis is that SSA-components are not necessarily harmonic functions and, being data adaptive, can capture highly nonharmonic oscillatory shapes, making it suitable for the analysis of nonlinear and nonstationary time series. In standard SSA, a so called trajectory matrix is generated by windowing the time series and storing all windows in the matrix, to enhance the periodic content of the original data. One main disadvantage is that the window length, or embedding dimension, has to be properly chosen, as the reconstructed components (and their physical meaningfulness) strongly depend on it. Moreover, the principal components to be selected for the reconstruction of a specific component series have to be carefully chosen to represent physically meaningful mechanisms and avoid mode mixing. Therefore, their frequency content has to be narrow-banded. Even though guidelines have been provided on the choice of the window length and component series reconstruction [Vautard *et al.* (1992); Golyandina *et al.* (2001); Golyandina (2010)], no standard approach exist to

make these processes automated. Consequently, SSA has been used for reconstruction of specific component series carefully identified out of the principal component analysis of the trajectory matrix, rather than for time series decomposition.

In this paper, we present the singular spectrum decomposition (SSD), a novel iterative SSA-based time series decomposition method. In the SSD method, the choice of the embedding dimension and the selection of the principal components for reconstruction of a specific component series have been made fully data-driven. This makes SSD an adaptive decomposition method. Similar to EMD, the decomposition is based on the extraction of the energy associated with various intrinsic time scales. As will be shown, SSD avoids mode mixing and provides accurate separation between intermittent components at the transition points. To achieve this, a new definition of the trajectory matrix is proposed and fed into the SSD method, which also guarantees decrease of energy of the residual at each iteration. A preliminary version of the SSD method was presented in Bonizzi *et al.* [2012]. In this paper, time–frequency spectra are presented as energy–time–frequency distribution obtained by the Hilbert spectrum. Qualitative comparisons with ensemble EMD (EEMD) are also provided for some of the numerical examples and applications. EEMD is a noise-assisted data analysis method recently proposed by Wu and Huang [2009] in order to overcome the scale separation problem (mode mixing) typical of standard EMD. Indeed, as discussed in Huang *et al.* [1998, 1999], intermittence could not only cause serious aliasing in the time–frequency distribution, but also make the physical meaning of individual components unclear. EEMD defines the exact component series as the mean of the corresponding component series obtained via EMD over an ensemble of trials, generated by adding different realizations of white Gaussian noise of finite variance (defined by the user) to the original signal (Matlab code for EMD and EEMD is available at http://perso.ens-lyon.fr/patrick.flandrin/emd). The paper is organized as follows. Section 2 reviews the SSA algorithm, its properties, and its capabilities as a data analysis method. Section 3 introduces the SSD method, by illustrating the fundamental choices adopted for the automated selection of the window length and of the principal components of the trajectory matrix for reconstruction of a specific component series. Section 4 shows the performance of SSD through numerical simulations. Section 5 illustrates the usefulness and capability of SSD through real applications. A discussion will be then presented in Sec. 6. Finally, conclusions are given in Sec. 7. Appendix provides the theoretical details about the convergence of the SSD method.

2. Singular Spectrum Analysis

SSA is a linear analysis and prediction method for time series, incorporating elements of classical time series analysis, multivariate statistics, multivariate geometry, dynamical systems, and signal processing [Golyandina *et al.* (2001); Zhigljavsky (2010)]. Despite its linear nature, the data-adaptive character of the eigenelements which SSA is based on makes it a suitable method for the spectral analysis of short

and noisy time series, without any *a priori* knowledge of the process generating the data [Vautard *et al.* (1992)]. Indeed, it has been shown that SSA gives adaptive spectral filters associated with the dominant oscillations in the time series. This allows to localize in time intermittent oscillation spells to a certain extent, making SSA suitable for the analysis of nonlinear and nonstationary time series [Vautard and Ghil (1989); Vautard *et al.* (1992)]. Application areas of SSA are diverse, such as mathematics and physics, economics and financial mathematics, meteorology and oceanology, geophysics, and more recent as engineering, medicine, and econometrics [Golyandina *et al.* (2001); Zhigljavsky (2010); Ghil *et al.* (2002)]. Different platforms and toolkit for SSA-based spectral analysis can be found at http://www.atmos.ucla.edu/tcd/ssa/ or at http://www.gistatgroup.com/cat/.

Given a zero-mean (centered, i.e. the average is subtracted before processing, [Vautard *et al.* (1992)]) time series $x(n)$, $n = 1, \ldots, N$, "Caterpillar" SSA consists of four stages: embedding, decomposition, grouping, and reconstruction [Golyandina *et al.* (2001); Golyandina and Osipov (2007)].

Embedding. The starting point of SSA is to embed the time series $x(n)$ in a vector space of dimension M. Let M be a strictly positive integer (named embedding dimension or window length, $M \in \mathbb{Z}^+$), $1 < M < N$. The embedding procedure forms $K = N - M + 1$ lagged vectors $\mathbf{x}_i = (x(i), \ldots, x(i + M - 1))^{\mathrm{T}}$, with $i = 1, \ldots, N - M + 1$. The trajectory matrix of the series $x(n)$ is then given by $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K]$, such that this is a Hankel matrix (constant cross-diagonals) of size $(M \times K)$. Hence, the embedding procedure constructs a sequence of M-dimensional vectors from the original time series $x(n)$, by using lagged copies of the scalar data.

Decomposition. The singular value decomposition (SVD) of the trajectory matrix $\mathbf{X}$ is then computed [Golyandina *et al.* (2001)], providing $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}$, with $\mathbf{U} = (M \times M)$ and $\mathbf{V} = (K \times K)$ being orthogonal matrices containing the left and right singular vectors, respectively, and $\mathbf{D} = (M \times K)$ being a matrix containing the singular values σ_i on the main diagonal and zero elsewhere (where $\sigma_i = \sqrt{K\lambda_i}$, with λ_i being the corresponding eigenvalues of $\mathbf{C}$, which in turn is the covariance matrix of $\mathbf{X}$, i.e. $\mathbf{C} = \frac{1}{K}\mathbf{X}\mathbf{X}^{\mathrm{T}}$). Matrix $\mathbf{X}$ is thus decomposed in a sum of rank-one matrices $\mathbf{X}_i$ such that $\mathbf{X} = \sum_{i=1}^{M} \mathbf{X}_i = \sum_{i=1}^{M} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathrm{T}}$, where the ith eigentriple $(\sigma_i, \mathbf{u}_i, \mathbf{v}_i)$ is composed of the ith right ($\mathbf{u}_i$; ith column of matrix $\mathbf{U}$) and left eigenvectors ($\mathbf{v}_i$; ith column of matrix $\mathbf{V}$), and the ith singular value σ_i.

Grouping. Hence, the grouping procedure partitions the set of indices $\{1, \ldots, M\}$ into m disjoint subsets $\{I_1, \ldots, I_m\}$ (eigentriple grouping; partitions carried out manually by the user), with each group I_c containing a set of principal components $\{i_1, \ldots, i_p\}$ representing specific components of the signal (trend, oscillations, harmonic components, etc.). Let $I_c = \{i_1, \ldots, i_p\}$. Then the resulting matrix $\mathbf{X}_{I_c}$ corresponding to the group I_c is defined as $\mathbf{X}_{I_c} = \mathbf{X}_{i_1} + \cdots + \mathbf{X}_{i_p}$. These matrices are computed for $I_c = I_1, \ldots, I_m$ and the SVD expansion leads to the decomposition

$\mathbf{X} = \mathbf{X}_{I_1} + \cdots + \mathbf{X}_{I_m}$. Each $\mathbf{X}_{I_c}$ represents a set of eigentriples which describes a specific component series $g_c(n)$ in the original time series.

Reconstruction. The last step in basic SSA is to reconstruct the components of the original series. This is achieved by *diagonal averaging* of each $\mathbf{X}_{I_c}$ to provide the cth component of the series $x(n)$, $g_c(n)$, $n = 1, \ldots, N$, where the nth sample is obtained by averaging over the cross-diagonal $i + j = \text{const} = n + 1$ of $\mathbf{X}_{I_c}$. This is because each $\mathbf{X}_{I_c}$ can be seen as the Hankel matrix for the corresponding embedded component series $g_c(n)$.

However, SSA may still fail in localizing the spells of intermittent components due to inaccurate manual estimation of the embedding dimension M and eigentriple selection for the reconstruction of a specific component series. This is not trivial and the final result is highly dependent on M. For instance, the number of significant eigenvalues over the noise floor, and thus the way the singular spectrum splits into noise and signal components can change dramatically when M is varied [Vautard and Ghil (1989)], thus providing different principal component sets for different values of M.

2.1. *Choice of the embedding dimension*

As pointed out by Vautard and Ghil [1989], there is a compromise between the amount of information one hopes to retain, the larger the M the better, and the statistical confidence one needs to achieve, the smaller the M the better. Given a peak in the power spectrum of the time series $x(n)$, with maximal spectral density at $f_{\max}$ and width $2 \cdot \delta f$, SSA can isolate correctly intermittent oscillation if $\frac{1}{f_{\max}} \leq M \leq \frac{1}{2 \cdot \delta f}$. Hence, M has to be chosen within the period of the oscillation and the average life span of its spells [Vautard *et al.* (1992)]. In practice, this latter quantity is not known *a priori*. Vautard and Ghil [1992] found that SSA is typically successful at analyzing periods in the range $(\frac{1}{5}M, M)$. Golyandina [2010] found a choice of M close to one-half of the time series length to be appropriate in most of the cases. This value is similar to the value $M = \frac{N}{3}$ provided by Vautard and Ghil [Vautard *et al.* (1992)] in case of presence of a sizable trend (a remarkable trend) in the time series. However, in order to obtain meaningful component series, one has to evaluate the stable features of the eigenset by varying M over a reasonable range.

2.2. *Selection of components for reconstruction*

In the absence of any knowledge about the components representing a time series, the problem of decomposition into several components can be reduced to the sequential extraction of the components one by one, and assuming only two main components series in the original time series: signal and noise [Golyandina *et al.* (2001)]. A proper SSA decomposition can identify three main sets: trends of the original series, oscillatory components (for example, seasonality), and noise. Hence, the main problem in the reconstruction of specific component series becomes the identification of

the leading SVD components of the trajectory matrix, with respect to the noise floor. Vautard and Ghil [Vautard and Ghil (1989); Vautard *et al.* (1992)] showed that vigorous albeit irregular oscillations are associated with a pair of nearly equal eigenvalues which stand out from the spectrum of the covariance matrix (after detrending). Oscillating pairs of eigenelements must then be spectrally localized around the same frequency. The condition for separability is then that the deterministic part of the time series can be approximated by a time series of finite rank r [Golyandina (2010)]. This means that all except the r leading eigenvalues are close to zero. However, when analyzing actual series with the help of SSA, one can hardly hope to obtain the exact harmonics or linear trend, even if these harmonics or linear trend are indeed present in the series. This is a consequence of the influence of noise and of the nonparametric nature of the method [Golyandina *et al.* (2001)]. Possible difficulties in estimating the rank r of the trajectory matrix and in identifying the leading components out of the noise floor make it arduous to achieve a reliable identification of eigenelements describing different oscillations in the data, especially when the noise becomes large.

3. The Singular Spectrum Decomposition Algorithm

In the proposed SSD method, the decomposition of the original time series into meaningful component series is achieved through sequential extraction of the components one by one, by means of an iterative approach. Based on the considerations made in Secs. 2.1 and 2.2, the automated choice of the embedding dimension in the SSD algorithm has to be accurately set at each iteration such as to describe a well defined frequency band in the spectrum of the original time series. Moreover, the automated identification of the SVD components for the reconstruction of the component series $g^{(j)}(n)$ at iteration j has to be set such as to identify all and only the SSA-derived components in the defined frequency band and minimize generation of spurious components.

3.1. *Modified trajectory matrix for SSD algorithm*

With respect to the standard way of building the trajectory matrix $\mathbf{X}$ (as introduced in Sec. 2), a new approach is proposed for the SSD algorithm. Given a time series $x(n)$ of length N, and an embedding dimension M, an $(M \times N)$ matrix $\mathbf{X}$ is generated such that its ith row is obtained as: $\mathbf{x}_i = (x(i), \ldots, x(N), x(1), \ldots, x(i-1))$, with $i = 1, \ldots, M$. Hence, $\mathbf{X} = [\mathbf{x}_1^{\mathrm{T}}, \mathbf{x}_2^{\mathrm{T}}, \ldots, \mathbf{x}_M^{\mathrm{T}}]^{\mathrm{T}}$. For instance, given the time series $x(n) = \{1, 2, 3, 4, 5\}$, and an embedding dimension $M = 3$, the corresponding trajectory matrix will be

$$\mathbf{X} = \left[\begin{array}{ccc|cc} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 1 \\ 3 & 4 & 5 & 1 & 2 \end{array}\right].$$

Note that the left-hand side block corresponds to the trajectory matrix $\mathbf{X}$ exploited in the standard SSA algorithm. This new formulation "wraps around" the time series $x(n)$ in the trajectory matrix $\mathbf{X}$ with the advantage of enhancing its oscillatory content, and providing useful properties for the decrease of energy of the residual, as shown in detail in Appendix A.1.

The diagonal averaging procedure introduced in Sec. 2 needs to be adapted to this new definition of the trajectory matrix. In order to carry out the average along the ith cross-diagonal of $\mathbf{X}$, the wrapped part of the right-hand block must be correctly appended to the top right of the left-hand block. For the example above:

$$\mathbf{X} = \left[\begin{array}{ccc|cc} & & 1 & & \\ & 1 & 2 & & \\ 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & * \\ 3 & 4 & 5 & * & * \end{array}\right],$$

where the asterisks mark the former locations of the elements composing the wrapped part moved to the right top of the left-hand block. With this approach, each cross-diagonal contains the same number of elements M (with $M = 3$ in this example, the empty locations should be ignored).

Care must be taken in case of a sizable trend in the time series. In this particular situation, the standard definition of the trajectory matrix has to be preferred, to avoid undesirable border effects due to generation of spurious oscillations at the edges of the reconstructed trend component. To prevent this, a test for sizable trend is performed at the first iteration. If the test is positive, a standard trajectory matrix is generated to guarantee reliable estimation of the trend. The new definition of $\mathbf{X}$ is then used in all the following iterations.

3.2. *Choice of the embedding dimension*

Taking into account the influence of the window length M on the decomposition provided by SSA and the results presented in Bonizzi *et al.* [2012], the following criterion has been adopted for the automated choice of M at iteration j:

- The power spectral density (PSD) of the residual time series at iteration j, $v_j(n) = x(n) - \sum_{k=1}^{j-1} v_k(n)$ ($v_0(n) = x(n)$), is computed. The frequency in its PSD associated with the dominant peak, $f_{\max}$, is then estimated.
- At the first iteration, if the normalized frequency $f_{\max}/F_s$ (with F_s being the sampling frequency), is less than a given threshold (set at 10^{-3} in our implementation), a sizable trend is assumed to characterize the residual. M is then set to $\frac{N}{3}$, as suggested by Vautard and Ghil [Vautard *et al.* (1992)] in case of a sizable trend present in the time series.
- Otherwise, and for iterations $j > 1$, the window length is defined as: $M = 1.2\frac{F_s}{f_{\max}}$.

The factor 1.2 allows the window length to cover a time span 20% larger than the average period of the wanted component series, to improve its identification by SSA.

The adaptive choice for M guarantees SSD to provide adaptive spectral filters associated with the dominant oscillations in the residual series, exploiting the property of SSA of concentrating the transfer functions related to the dominant eigenfunction within regions where sharp peaks in the power spectrum occur, facilitating the detection of such oscillations.

3.3. *Reconstruction of the jth component series*

Reconstruction of the jth component series $g^{(j)}(n)$ is carried out as follows: At the first iteration, if a sizable trend has been detected, only the first left and right eigenvectors are used to obtain $g^{(1)}(n)$, such that $\mathbf{X}_1 = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^{\mathrm{T}}$, and $g^{(1)}(n)$ is obtained from diagonal averaging of $\mathbf{X}_1$. Indeed, as a sizable trend is the component with the highest energy in the analyzed signal, it should be mainly reflected in the first eigentriple provided by SSA.

Otherwise, and for iterations $j > 1$, a component series $g^{(j)}(n)$ has to be retrieved such as to describe a well defined time scale. In this sense, its frequency content is to be concentrated in the frequency band $[f_{\max} - \delta f, f_{\max} + \delta f]$, with δf representing half the width of the dominant peak in the PSD of the residual time series $v_j(n)$. Hence, a subset I_j ($I_j = \{i_1, \ldots, i_p\}$) is created from all eigentriples whose left eigenvectors show a dominant frequency in their spectra in the range $[f_{\max} - \delta f, f_{\max} + \delta f]$ and the one eigentriple which contributes the most to the energy of the dominant peak of the selected mode. The corresponding component series is then reconstructed by diagonal averaging of the matrix $\mathbf{X}_{I_j} = \mathbf{X}_{i_1} + \cdots + \mathbf{X}_{i_p}$ along the cross-diagonals (see Sec. 2). As mentioned in Sec. 2.1, the width δf of the dominant peak in the PSD is related to the average life span of the wanted oscillation spells. As this quantity is not known *a priori*, it has to be estimated from the PSD of $v_j(n)$.

To achieve a good estimate of the dominant peak width δf, a spectral model defined by a superposition of Gaussian functions is used for describing the PSD profile, thereby accounting for the presence of the dominant peak with respect to the presence of additional peaks. The model is defined as a sum of three Gaussian functions, each function representing a spectral peak

$$\gamma(f, \theta) = \sum_{i=1}^{3} A_i e^{-\frac{(f-\mu_i)^2}{2\sigma_i^2}}, \tag{1}$$

where A_i is the magnitude of the ith Gaussian, σ_i is its width, μ_i is its location, and $\theta = [\mathbf{A}\ \sigma]^T$ is the parameter vector, with $\mathbf{A} = [A_1, A_2, A_3]$ and $\sigma = [\sigma_1, \sigma_2, \sigma_3]$.

In a biomedical context, a similar algorithm was used in Corino *et al.* [2008] to improve the update of the spectral profile for use in time–frequency analysis of atrial fibrillation. Unlike in Corino *et al.* [2008], the model in Eq. (1) is characterized by three Gaussian functions not in harmonic relation with each other. In particular,

the first Gaussian function accounts for the dominant spectral peak, the second function for the second highest spectral peak, while the third functions accounts for everything in between the two main spectral peaks. The purpose of the second and third Gaussian function is to prevent artificial widening of the first Gaussian function over frequency bands describing different scales than the one corresponding to the dominant spectral peak. With such a model, information about δf is then retained in σ_1. Since the first and second Gaussian functions are expected to be located at the frequencies corresponding to the two dominant spectral peaks ($f_{\max}$ and f_2, respectively), their locations are constrained to be centered at these frequencies. The third function is then located halfway between $f_{\max}$ and f_2. Hence

$$\mu_1 = f_{\max}, \quad \mu_2 = f_2, \quad \mu_3 = f_3 = \frac{f_{\max} + f_2}{2}.$$

The model parameters are obtained from weighted least squares fitting of the model to the observed PSD. The optimization procedure fits the model to the entire PSD using the following initial parameter values (indexed as $(\cdot)^{(0)}$):

$$A_1^{(0)} = \frac{1}{2}\mathrm{PSD}(f_{\max}) \quad \sigma_1^{(0)} = f : \mathrm{PSD}(f) = \frac{2}{3}\mathrm{PSD}(f_{\max}),$$

$$A_2^{(0)} = \frac{1}{2}\mathrm{PSD}(f_2) \qquad \sigma_2^{(0)} = f : \mathrm{PSD}(f) = \frac{2}{3}\mathrm{PSD}(f_2),$$

$$A_3^{(0)} = \frac{1}{4}\mathrm{PSD}(f_3) \qquad \sigma_3^{(0)} = 4|f_{\max} - f_2|.$$

Similar to Corino *et al.* [2008], the optimal values are determined using a nonlinear minimization algorithm, such as the Levenberg–Marquardt algorithm [Lourakis and Argyros (2005)]. Given the estimated value for σ_1, a value for δf is then obtained as $\delta f = 2.5\sigma_1$. The factor 2.5 is chosen such that almost 99% of the area under the dominant peak is accounted for, by excluding at the same time baseline noise.

This choice for $\mathbf{X}_{I_j}$ avoids the need of identifying the leading eigenvalues with respect to the noise floor (or similarly, the estimation of the rank r of the trajectory matrix). Hence, components describing different scales are automatically discarded at iteration j, and left to be identified at future iterations (as their frequency content is outside $[f_{\max} - \delta f, f_{\max} + \delta f]$). The underlying hypothesis is that at each iteration SSD picks up the oscillation describing most of the energy in the residual series, under the assumption that SSA tries to find spectral bands with a high percentage of explained energy.

Moreover, a second run of the algorithm is carried out on the jth component series just retrieved, $g_{[1]}^{(j)}(n)$ (for $j = 1$, this second run is performed in case of no sizable trend detected). This allows to achieve a better estimate of $g^{(j)}(n)$, by polishing up the result of the first run. To guarantee convergence of the algorithm, the second estimate has to fulfill the condition $\sum_{i=1}^{N}[g_{[2]}^{(j)}(i)(v^{(j)}(i) - g_{[2]}^{(j)}(i))] > 0$ to be accepted. Otherwise, the estimate from the first run, $g_{[1]}^{(j)}(n)$, is taken as estimate of $g^{(j)}(n)$. This condition ensures that the method is not generating a

residual $v^{(j+1)}(n)$ with higher energy than $v^{(j)}(n)$. A theoretical justification for this condition can be found in Appendix A.2.

Finally, a scaling factor $\hat{a}$ is applied to $g^{(j)}(n)$ such as to adjust its variance to the residual time series $v^{(j)}(n)$, s.t.

$$\hat{a} = \min_a \|v^{(j)}(n) - a g^{(j)}(n)\|_2^2, \quad \text{giving } \hat{a} = \frac{g^{\mathrm{T}} v}{g^{\mathrm{T}} g} \text{ and } \tilde{g}^{(j)}(n) = \hat{a} g^{(j)}(n).$$

3.4. *Stopping criterion*

Each time a new component series $\tilde{g}^{(j)}(n)$ is estimated, a new residual is calculated: $v^{(j+1)}(n) = v^{(j)}(n) - \tilde{g}^{(j)}(n)$, which represents the input to the next iteration $j+1$. The normalized mean square error (NMSE) between the residual and the original signal is then computed

$$\mathrm{NMSE}^{(j)} = \frac{\sum_{i=1}^{N} (v^{(j+1)}(i))^2}{\sum_{i=1}^{N} (x(i))^2}.$$

The decomposition process is stopped when the NMSE is less than a given threshold (default th = 1%). The final decomposition gives

$$x(n) = \sum_{k=1}^{m} \tilde{g}^{(k)}(n) + v^{(m+1)}(n),$$

where m is the number of component series identified. It can be proven that the energy of the residual decreases at each iteration. A proof is provided in Appendix A.1, together with additional properties of the SSD method.

4. Method Validation

The SSD algorithm achieves extraction of band limited intermittent oscillatory components minimizing both reduction of the amplitude of the fit within the spells and generation of artificial periodicities off the spells. This may be hard to achieve with basic SSA or its extensions (as sequential SSA) as a suitable window length has to be set at each stage and information on the amount and type of component series forming the observed time series may be missing.

For instance, consider the harmonic time series in Fig. 1, sampled at 1 kHz.

This shows a low-frequency sinusoidal wave with dominant frequency 5 Hz and unit amplitude carrying a sinusoidal wave with dominant frequency 15 Hz and amplitude of 0.1. An additional high-frequency (75 Hz) intermittent oscillation with unit amplitude is added to the second half of the signal. Using basic SSA, an accurate grid search of the optimal window length to separate the three components gives the results shown in Fig. 2(a), for a window length $M = 106$. The reconstruction of each harmonic component was achieved by means of all eigentriples whose right singular vectors showed a dominant frequency in a frequency range $2 \cdot \delta f = 1$ Hz centered at 5, 15 and 75 Hz, respectively. SSA achieves separation of

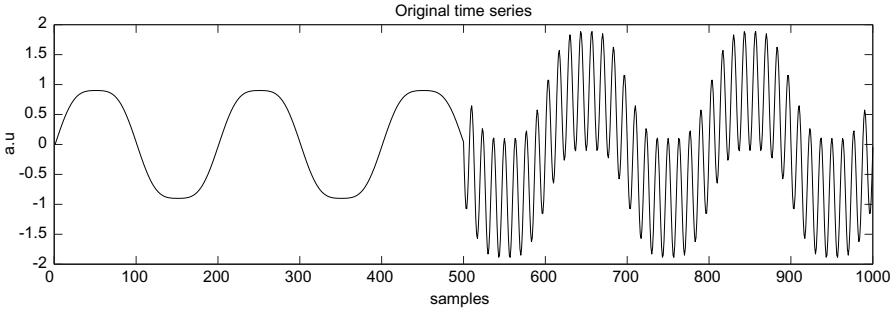


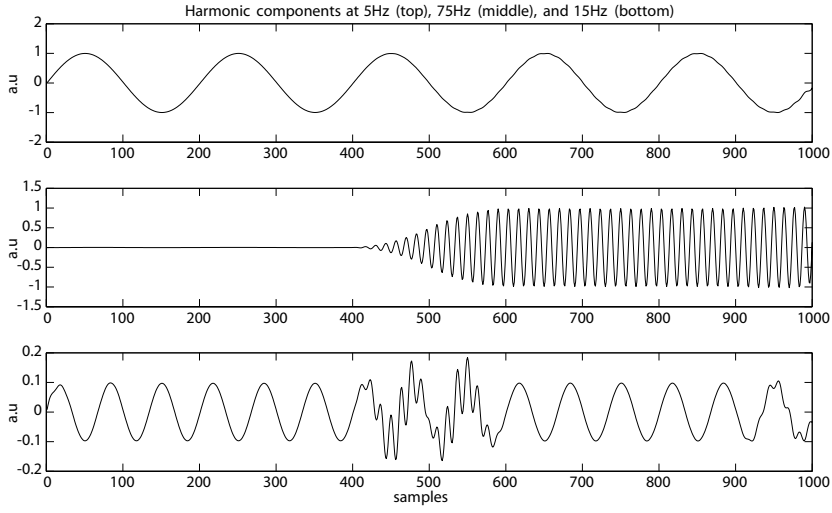
Fig. 1. Low-frequency sinusoidal wave with dominant frequency 5 Hz and unit amplitude carrying a sinusoidal wave with dominant frequency 15 Hz and amplitude of 0.1. An additional high-frequency (75 Hz) intermittent oscillation with unit amplitude is added to the second half of the signal; a.u.: arbitrary units.

the three harmonic components when a suitable window length is chosen, and with prior knowledge of their (frequency domain) properties. However, artifacts affects the 15 Hz component at the transition with the 75 Hz component, and negligible spurious components have been generated off the spells for the 75 Hz component.

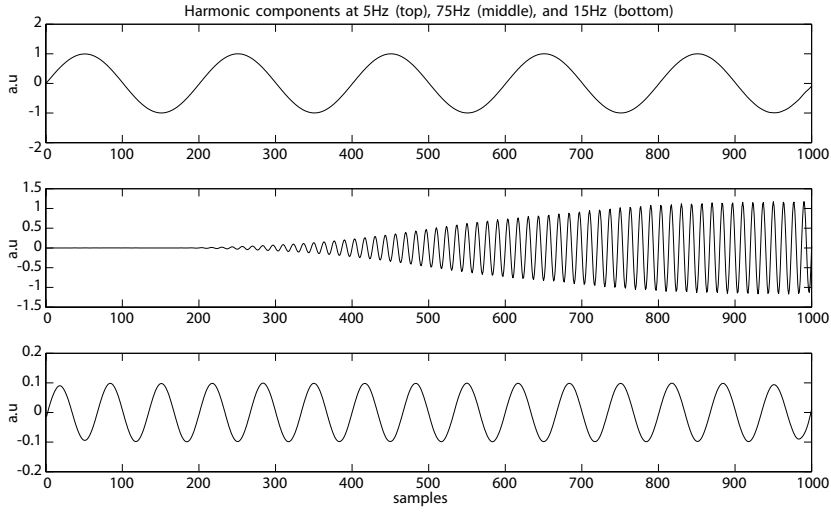
As discussed in Sec. 2.1, with no prior knowledge about the amount or type of components constituting the signal under analysis (as in most real applications), one-third of the time series length is considered an appropriate choice for the window length in most cases. Figure 2(b) shows the results of the separation provided by SSA when $M = \frac{N}{3}$. Even though mode separation is properly achieved, for such a choice of the window length important spurious components have been generated off the spell of the 75 Hz component.

The output of SSD is given in Fig. 3 (th = 0.5%). This is achieved in a completely automated way, by means of the algorithm presented in Sec. 3. It can be noted how SSD separates the two harmonic waves at 5 and 15 Hz, the second one with very low amplitude, without altering the identification of the intermittent component at 75 Hz. In this example, a threshold of 0.5% is important to be able to pick up the low amplitude 15 Hz component. The default threshold of 1% would provide only the first two components in Fig. 3. Hence, care must be taken on the choice of the threshold.

Performance of SSD is compared with band-pass filtering. This is carried on simulated low field potentials (LFPs), which represent the relatively slow varying temporal components of the neural signal (the synchronized dendritic input) picked up from within a few hundreds of microns of a recording electrode. The study of the processes underlying LFPs is considered important in understanding information transmission in the brain [Maier *et al.* (2010)]. Similar to Liang *et al.* [2005], we created a composite time series consisting of four components, each amplitude modulated and frequency-modulated in the gamma (30–90 Hz), beta (13–30 Hz), alpha (8–12 Hz), and theta (4–7 Hz) frequency bands, to represent typical frequency components in LFP data. Each component consisted of a noisy, narrow-band oscillation



(a)



(b)

Fig. 2. (a) Three component series provided by basic SSA of the signal in Fig. 1 with $M = 106$. (b) Three component series provided by basic SSA with $M = \frac{N}{3}$ of the signal in Fig. 1. a.u.: arbitrary units.

around a central frequency at 40, 20, 10 and 5 Hz, respectively, modulated in both amplitude and frequency (interested readers are referred to Liang *et al.* [2005] for details about the creation of these simulated data). An example of the composite time series is given in Fig. 4(a). Figure 4(b) shows the six component series obtained by applying SSD to the simulated LFP data (th = 1%), together with their PSDs.

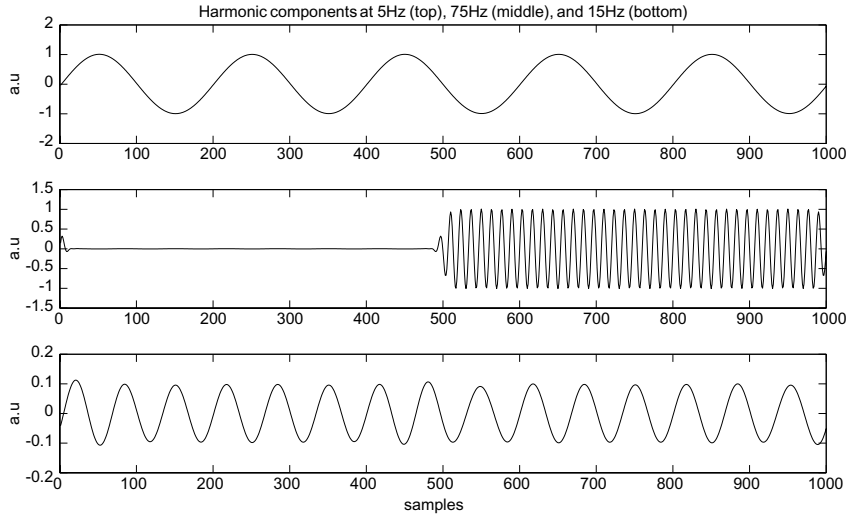


Fig. 3. Three component series provided by the SSD algorithm applied to the signal in Fig. 1. a.u.: arbitrary units.

It can be noted that SSD is able to retrieve the four fundamental waves composing the original time series, at the same time minimizing the number of spurious components (a low frequency trend and a high frequency component in this example) and their amplitudes.

To avoid pre-determined sub-band filtering and to attain a rigorous comparison, band-pass filtering is applied as an alternative to SSA into the SSD algorithm. That means, at each iteration a band-pass finite impulse response (FIR) filter with pass-band $[f_{\max} - \delta f, f_{\max} + \delta f]$ is applied to the residual signal. δf is estimated as introduced in Sec. 3.3. The filter order is optimally estimated at each iteration (Parks–McClellan optimal equiripple FIR order estimator). The algorithm runs until the same stopping criterion of SSD is reached. Results are shown in Fig. 4(b). It is clear how band-pass filtering fails to attain separation of the four waves in different components.

Finally, with respect to a chirp (a signal in which the frequency increases or decreases with time), similar to EEMD, SSD decomposes the signal in several narrow-band components (results not shown).

4.1. *Numerical results from a classic nonlinear system*

To explore the capability of SSD to handle nonlinear time series, the method was applied to a classic nonlinear system also analyzed in Huang *et al.* [1998]: the Rössler equation. The Rössler equation is given by the system

$$\dot{x} = -(x + y), \quad \dot{y} = x + \frac{1}{5}y, \quad \dot{z} = \frac{1}{5} + z(x - \mu),$$

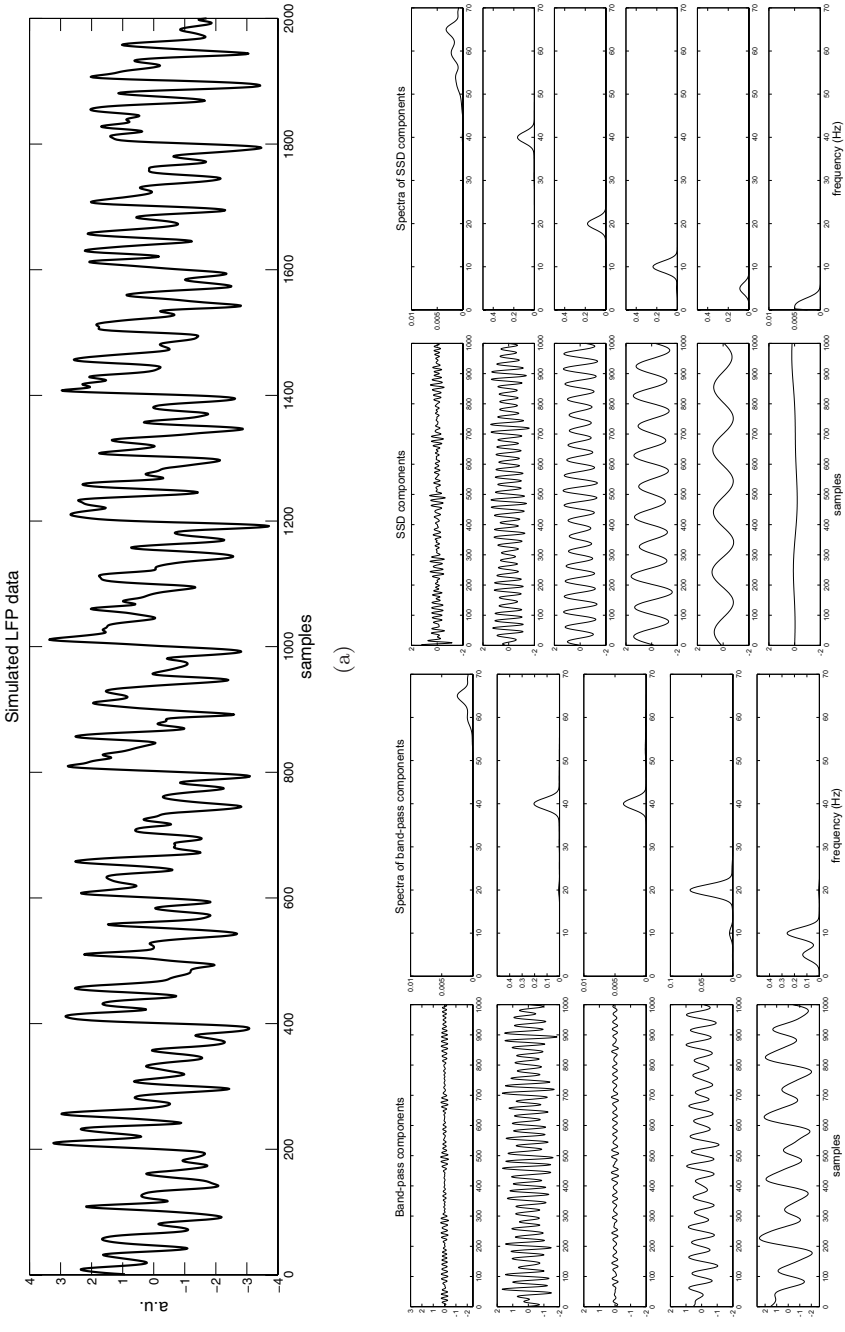
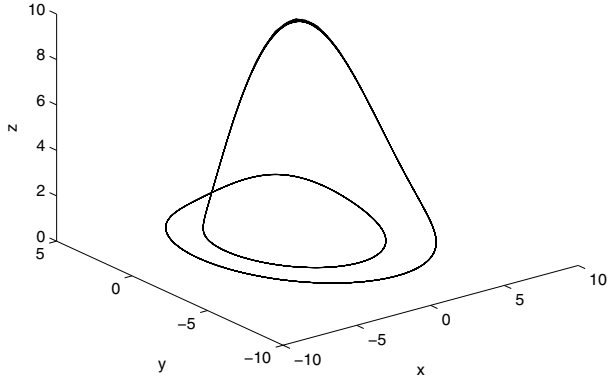
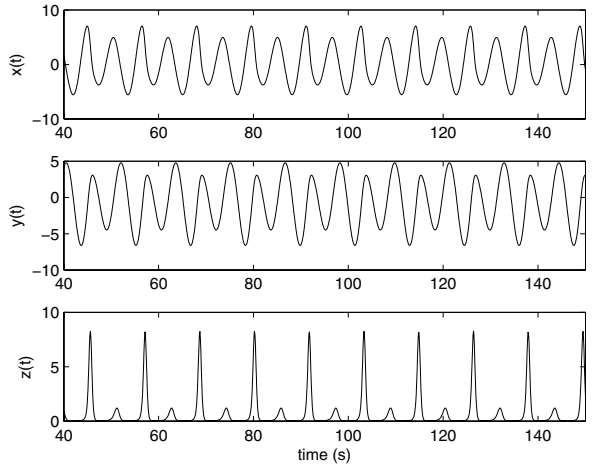


Fig. 4. (a) Simulated LFP data. (b) Five component series provided by band-pass filtering together with their corresponding PSDs (left two columns). Six component series provided by SSD together with their corresponding PSDs (right two columns). a.u.: arbitrary units.



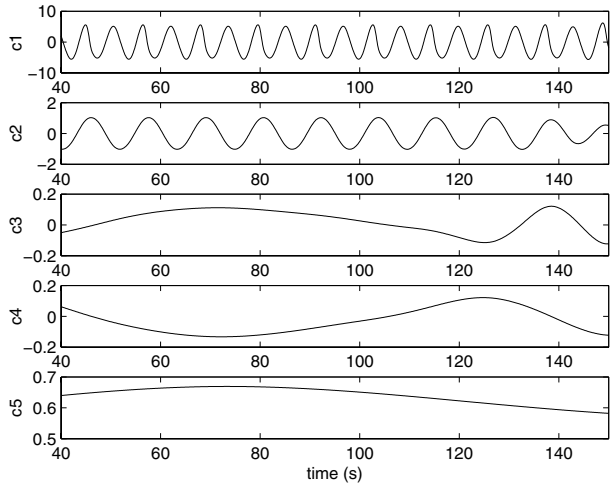
(a)



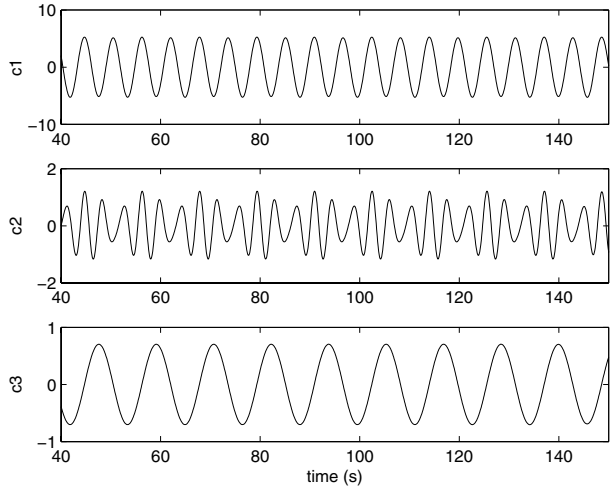
(b)

Fig. 5. (a) Three-dimensional representation of the numerical solution of the Rössler equation given in Eq. (2), for the period-doubling case. (b) The time series for the numerical solution of the Rössler equation given in Eq. (2).

with $\mu = 3.5$ representing the well known period-doubling event. Figure 5(a) presents the phase diagram. Figure 5(b) shows the time series for the numerical solution of the Rössler equation given in Eq. (2). The wave form of the x-component is regular with an exact period of twice the high-frequency oscillation. EMD of the x-component gives the five IMF components as shown in Fig. 6(a). Similar to what is presented in Huang *et al.* [1998]: $c1$ represents a nonuniform high-frequency component, $c2$ a uniform intermediate component, and the other components are low-frequency components with low energy likely generated by numerical error.



(a)



(b)

Fig. 6. (a) The IMF components of the x-component of the Rössler equation given in Eq. (2): Component $c1$ is nonuniform but the most energetic, $c2$ is uniform and represents steady low-frequency band at about 0.1 Hz, the other components are likely due to numerical error. (b) The SSD components of the x-component of the Rössler equation given in Eq. (2): Component $c1$ is uniform and the most energetic, $c2$ is nonuniform, and $c3$ is uniform.

SSD of the x-component gives the three component series as shown in Fig. 6(b): $c1$ represents a uniform high-frequency component (the most energetic one), $c2$ a nonuniform intermediate component, and $c3$ a uniform low-frequency component. Unlike EMD, SSD does not provide any spurious component. Hence, the SSD seems less affected by numerical error. The Hilbert spectrum of the EMD-components

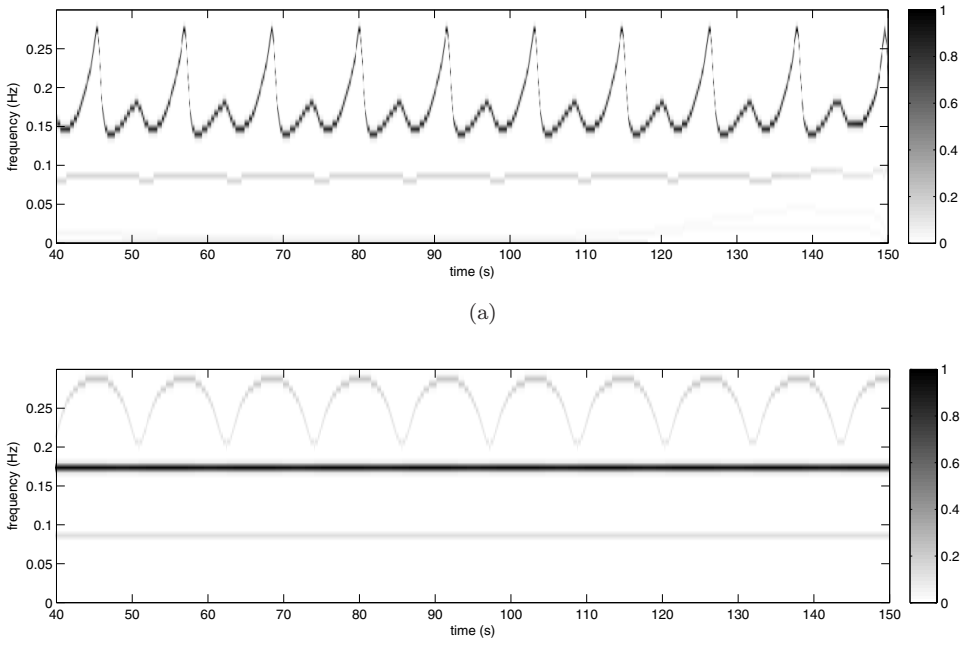


Fig. 7. (a) The EMD derived Hilbert spectrum for the x-component. The most energetic component shows regular intra- and interwave frequency modulations. (b) The SSD derived Hilbert spectrum for the x-component. The two steady components representing the basic wave periods are well separated from the intrawave frequency modulations.

(Fig. 7(a)) contains only two prominent components: The high-frequency component fluctuating widely between 0.15 and 0.26 Hz, and the steady low-frequency band at slightly less than 0.1 Hz. The fluctuating high-frequency component is strictly periodic with two basic wave periods as the period doubling requires. Moreover, within the two wave periods, the frequency undergoes intrawave frequency modulation that corresponds to the alternative long and more rounded waves interlaced with the short and much sharper ones. The first EMD-component shows both those intrawave and interwave frequency modulations. The Hilbert spectrum of the SSD-components (Fig. 7(b)) contains: The steady high-frequency component at about 0.17 Hz, the steady low-frequency component at slightly less than 0.1 Hz (similar to the one retrieved by EMD). These two components are supposed to represent the two basic wave periods as the period doubling requires. Moreover, the remaining component represents the intrawave frequency modulations that correspond to the alternative long and more rounded waves interlaced with the short and much sharper ones. Interestingly, SSD separated the contributions of the basic waves from the intrawave fluctuations, while EMD incorporates in the same IMF the steady high-frequency component at about 0.17 Hz and the intrawave modulations.

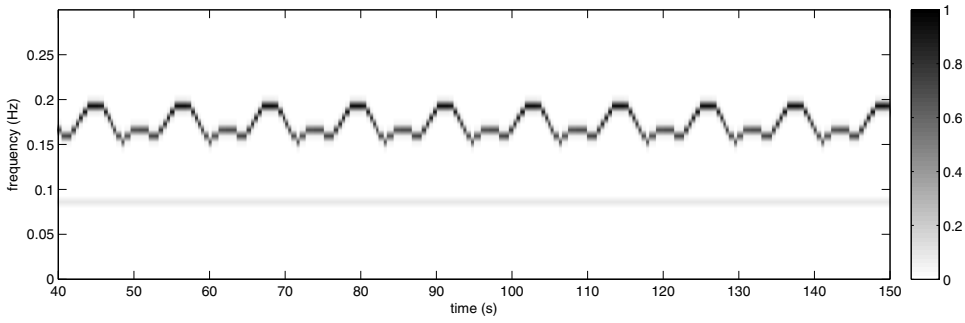


Fig. 8. The Hilbert spectrum for the x-component, with components $c1$ and $c2$ added together.

To show this, the SSD components $c1$ and $c2$ were added together and a new Hilbert spectrum was computed (Fig. 8). It can be noted that the component $c1 + c2$ behaves very similarly to the first component provided by EMD, a high-frequency component fluctuating widely between 0.15 and 0.2 Hz, strictly periodic with two basic wave periods. EEMD showed performance similar to EMD (results not shown).

5. Applications on Real Time Series

In this section, we will describe applications of the SSD method through a number of examples. In a similar way as stated by Huang *et al.* [1998], this will allow to understand "whether SSD indeed helps in reaching the ultimate goal of data analysis: To isolate and extract the physically meaningful components in data, and thereby to understand the properties of data and its underlying physics". The easiest way to demonstrate the power of SSD and its usefulness is to apply it to data of natural and biological phenomena. In this section, SSD is applied to three real cases: The first case is neurological LFP signals acquired from animals performing specific visual tasks. The second case is about equatorial region ocean surface elevation measured by altimeter during a tsunami. The third case is the length-of-day (LOD) data (previously analyzed in Huang *et al.* [2003], Wu and Huang [2009]). A threshold of 1% was used for all applications but the LOD data, for which 0.1% was used. The heterogeneity of these applications will also allow to underline the robustness of the SSD algorithm to diverse types of noise.

5.1. *Low field potential data*

In the previous section, SSD performance was evaluated on simulated low field potentials (LFPs). In vision, it is well-established that stimulation causes a relative enhancement of power within a spectral band centered in the gamma-band frequency region of 30 Hz to 100 Hz [Fries *et al.* (2001); Fries (2005)]. Specific theoretical proposals have pointed out the prominent role of coherence in the gamma-band

range as a vehicle for communication among different neuronal populations, which has been shown to not require a stable frequency but to be maintained despite a shift in peak frequency governed by stimulus contrast [Roberts *et al.* (2010)]. A detailed time–frequency analysis of LFPs requires not only LFPs from distinct populations to be studied separately from population spiking but also to focus on specific frequency bands, asking for these bands to be isolated. As gamma oscillations have been discovered to be nonstationary in the visual cortex [Roberts *et al.* (2010)], and EEG signal is known to be nonlinear [Liang *et al.* (2005)], tools able to capture nonstationary and nonlinear dynamics are required. In Bonizzi *et al.* [2012], a prototype version of the current SSD algorithm was applied to real LFP recordings from macaque V1 cortex undergoing a visual fixation task. The method generated gamma-band components characterized by higher energy than those given by EMD, hence more suitable in tracking the changes in the gamma-band energy content before and after stimulus presentation.

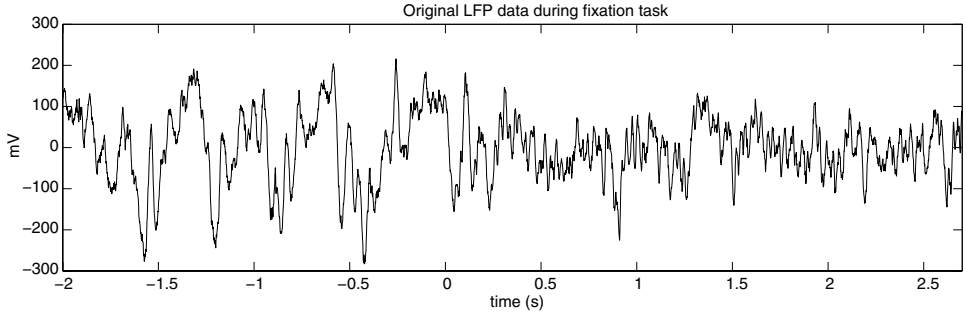
An example of a single-trial LFP recording is shown in Fig. 9(a) (time $t = 0$ represents stimulus onset). SSD was applied to the LFP recording, giving the component series shown in Fig. 9(b). SSD-components have been ordered by decreasing dominant frequency. Note how the second component, the main gamma-related SSD component (components with a dominant frequency in the gamma-range [30, 90] Hz), properly captures the dynamic enhancement of this activity after stimulus presentation.

The Hilbert spectrum of the SSD gamma-components is shown in Fig. 10 (top: contribution of the first and second components of Fig. 9(b)). It can be noted how SSD can track over time the nonstationary behavior in the gamma-band, and highlight the activation of the main gamma-related component after stimulus onset. Results for EMD and EEMD are also shown for completeness. These methods seem to have more difficulties in following the dynamical changes in the gamma-band after stimulus onset.

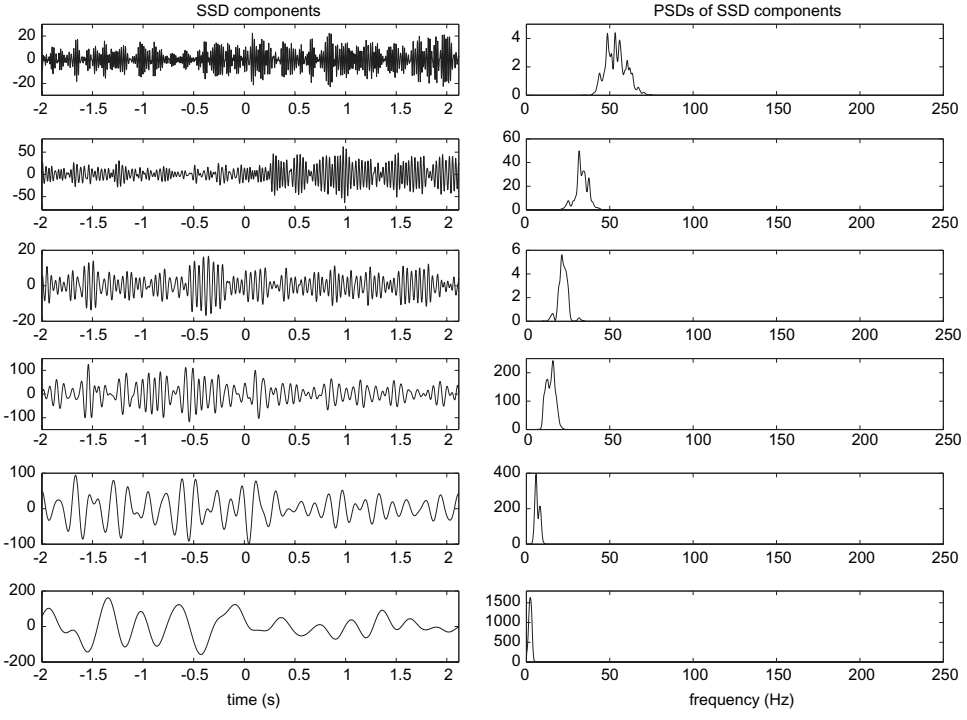
As put forward in Bonizzi *et al.* [2012], the SSD method may be then useful to analyze single-trial LFP recordings to generate physiologically meaningful measurements of oscillatory neural behavior, and to improve the analysis of functional connectivity among different brain areas.

5.2. *Tide and tsunami data*

In a similar way as in Huang *et al.* [1998], we analyzed the raw data from the NOAA tidal gauge located inside Kahului Harbour, Maui for approximately four days from March 10, 2011 (unlike Huang *et al.* [1998], in which data for five days from October 4, 1994, were analyzed). This gauge is capable of recording data at 1 min steps. On March 11, 2011 an undersea earthquake off the coast of Japan occurred. On March 11, tsunami-induced waves arrived at the site and created water level changes of a magnitude larger than that of the tidal signal and persisted for several tidal cycles. The tidal data are shown is Fig. 11.



(a)



(b)

Fig. 9. (a) LFP data during fixation task. (b) Six component series provided by SSD (left column) together with their corresponding PSDs (right column).

The combination of the transient tsunami waves with the pseudo-stationary tidal wave makes the whole time series nonstationary. Since the transient tsunami data and the tide have many harmonic components in the same frequency range, filtering cannot separate the two signals cleanly. The SSD method provides

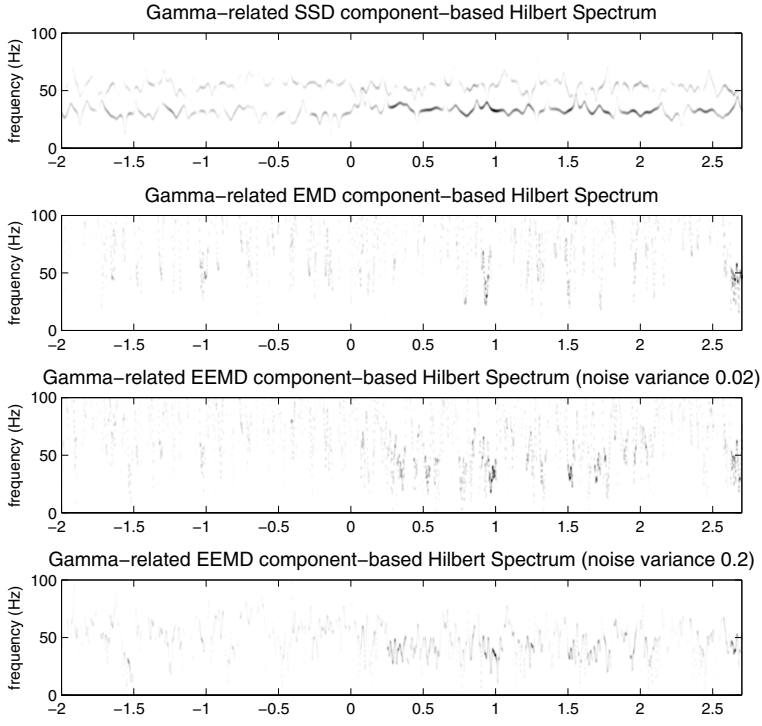


Fig. 10. The Hilbert spectra of the gamma-related components of LFP data during fixation task extracted with SSD, EMD, and EEMD, respectively. For EEMD, both noise variances 0.02 and 0.2 have been analyzed.

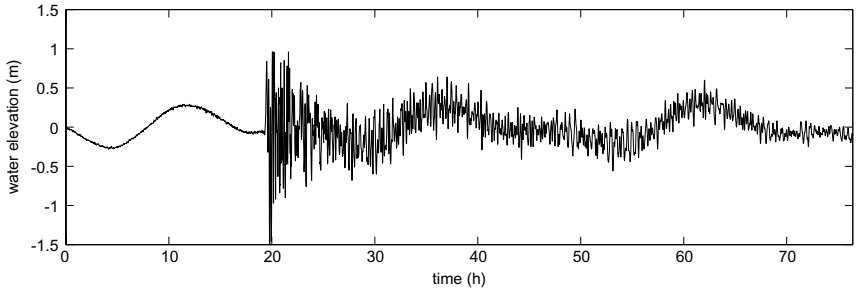


Fig. 11. The tidal data collected inside the Kahului Harbour, Maui, from March 10–13, 2011. On March 11, tsunami waves reached the tidal station with a signal of larger strength than that of the tides.

thirteen components from the data as shown in Fig. 12. It is easy to recognize the high-frequency signal representing the tsunami-induced waves, and the last two low-frequency components representing the semi-diurnal and diurnal tide, respectively.

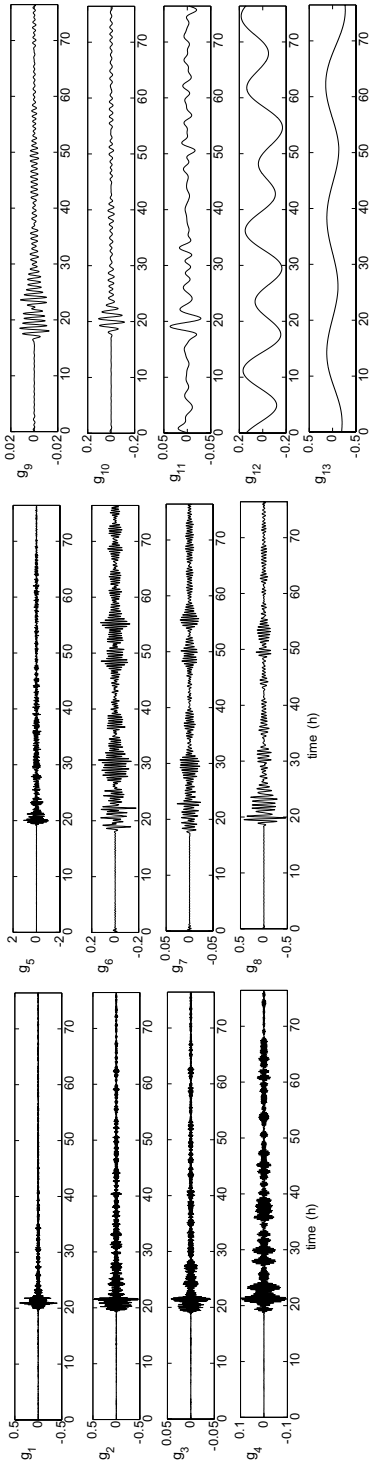


Fig. 12. The thirteen SSD-components obtained for the data in Fig. 11.

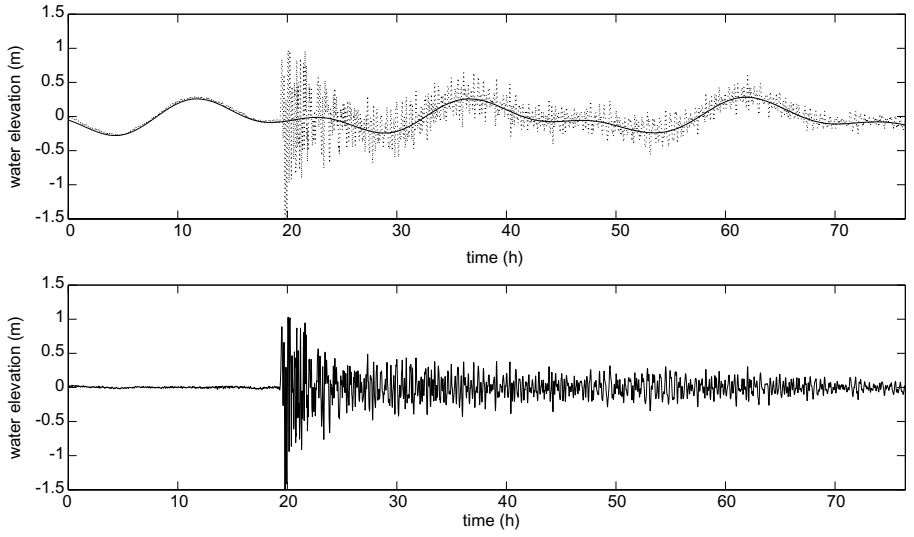
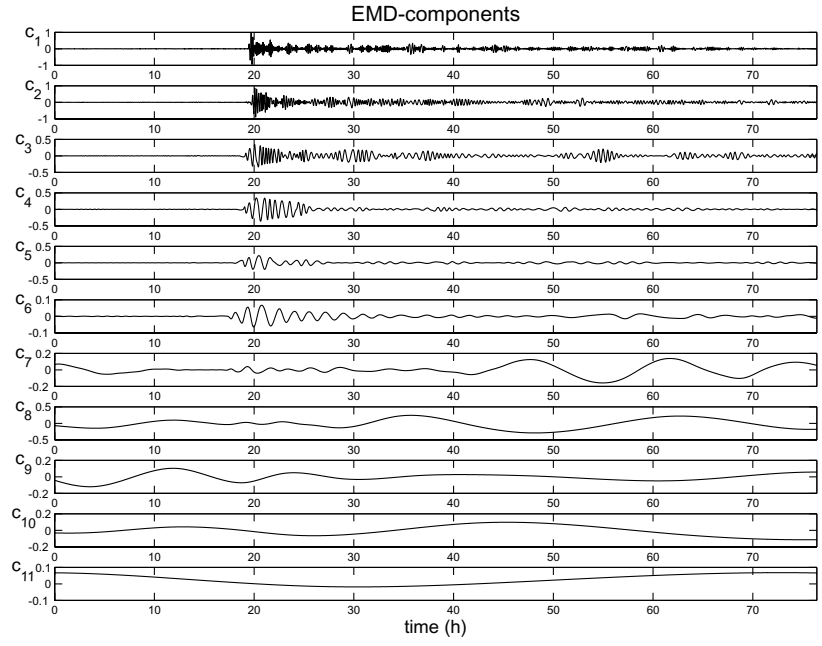


Fig. 13. Separation of the tides and the tsunami waves by the SSD-components. Top panel: The tides obtained by means of SSD-component twelve and thirteen (thick solid line), overlapped to the raw data (thin dotted line). Bottom: The tsunami waves represented by the sum of the SSD components one to eleven. Note how separability of the two motion waves is achieved.

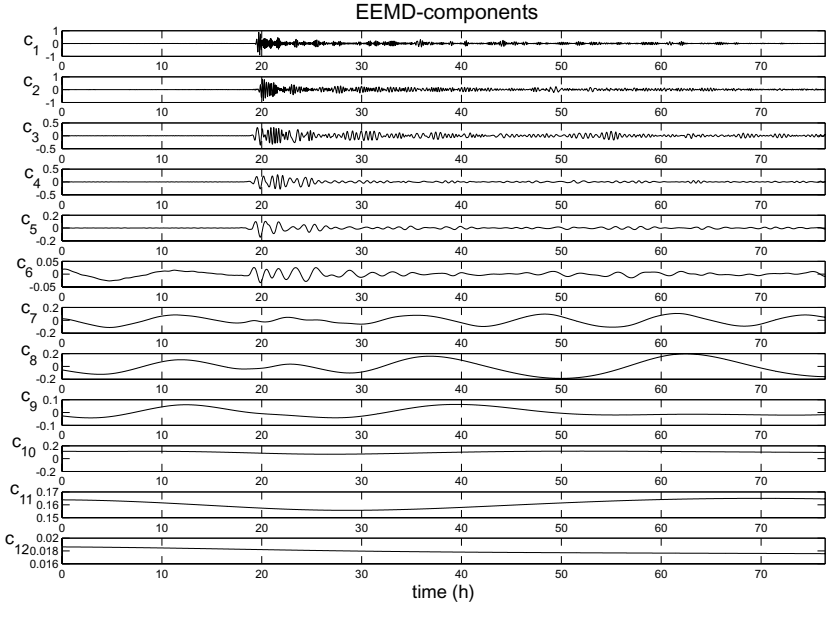
Figure 13 shows the tidal component overlapped to the raw data (top), reconstructed by means of components twelve and thirteen, and the tsunami-induced waves, reconstructed by using all the remaining components (bottom). Note how SSD separates the tide and the transient tsunami without any ambiguity.

A qualitative comparison with EMD and EEMD is also provided. The eleven EMD-components and the twelve EEMD-components for the data in Fig. 11 are shown in Figs. 14(a) and 14(b), respectively. In this particular case, both the EMD and the EEMD output seem to be affected by mode mixing. High-frequency tsunami-induced waves are mixed with tidal waves in EMD-components c_7 and c_8, while the semi-diurnal and diurnal tides are mixed in c_9. Similarly, tsunami-induced waves are mixed with tidal waves in EEMD-components c_6 and c_7. Hence, ambiguity is present in the separation of tides and the transient tsunami waves.

The larger number of tsunami components generated by SSD w.r.t. EMD and EEMD can be explained in light of the results presented for the Rössler equation in Sec. 4. Considering the tsunami waves likely characterized by inter- and intrawave frequency modulation, SSD separates the contributions of the basic waves from the intrawave fluctuations, while EMD and EEMD include these contributions in the same components. Moreover, as in Huang *et al.* [1998], the tsunami wave frequencies are phase locked with the tidal cycle. Finally, SSD separates without ambiguity tides and the transient tsunami waves. This may help to analyze and compare different tsunamis and their dependences on tide levels.



(a)



(b)

Fig. 14. (a) The eleven EMD-components and (b) the twelve EEMD-components for the data in Fig. 11 (noise variance for EEMD equal to 0.02).

5.3. *Length-of-day data*

Universal time and length of day are subject to variations due to the zonal tides (smaller than 2.5 ms in absolute value), to oceanic tides (smaller than 0.03 ms in absolute value), to atmospheric circulation, and to transfer of angular momentum to the Moon orbital motion (see information available at http://www.iers.org/nn 10398/IERS/EN/Science/EarthRotation/UT1LOD.html). The variations in LOD can be split into several components, according to their causes. The dynamical influence of the liquid core of the earth and climatic variations in the atmosphere account for slow variations while the rest of the atmospheric excitation can be split into a seasonal oscillation and a residual oscillation. The LOD data being decomposed here cover the period from Jan 1962 to July 2012. These data are taken from the comb2011_noon.eop series and are available at ftp://euler.jpl.nasa.gov/keof/combinations/2011. A detailed interpretation of the different oscillations composing the LOD data was given in Huang *et al.* [2003], were LOD data were studied by means of EMD while analyzing a confidence limit for the method and Hilbert spectral analysis, and in Wu and Huang [2009], where LOD data were analyzed by means of EEMD (the LOD series analyzed in Huang *et al.* [2003] and Wu and Huang [2009] were both shorter than the one analyzed in this study). The LOD data and the 11 components retrieved by SSD are displayed in Figs. 15(a) and 15(b), respectively.

The first component has quasi-regular spikes with an average period around 9 days, similar to the 7 day average period component retrieved by Huang *et al.* [Huang *et al.* (2003); Wu and Huang (2009)]. The second component has an average period about 14 days, linked to semi-monthly tides (similar to component two in Huang *et al.* [2003] and Wu and Huang [2009]). The combined components three and four have an average period of 28 days, linked to monthly tides (similar to component three in Huang *et al.* [2003] and Wu and Huang [2009]). Component five has a period of about three months of unknown causes, as also mentioned in Huang *et al.* [2003]. Component six is a semi-annual component (similar to component five in Huang *et al.* [2003] and Wu and Huang [2009]). Component seven has a period slightly shorter than 8 months. Component eight is an annual component (similar to component six in Huang *et al.* [2003] and Wu and Huang [2009]).

Component nine has an average period of about two years and a half. Such a component has not been retrieved in Huang *et al.* [2003] and Wu and Huang [2009]. Silva *et al.* have recently described associations between LOD and geomagnetic secular acceleration for 6-year periodic oscillation [Silva *et al.* (2012)]. In the same study they also identified a 2.5-year periodic signal in the CHAOS-3 geomagnetic field model with unknown origin. This unknown periodicity may be linked to the one which has been identified from LOD data in this study.

Moreover, careful examination of component ten suggests that this component seems to contain the 6-year periodic oscillation aforementioned [Silva *et al.* (2012)].

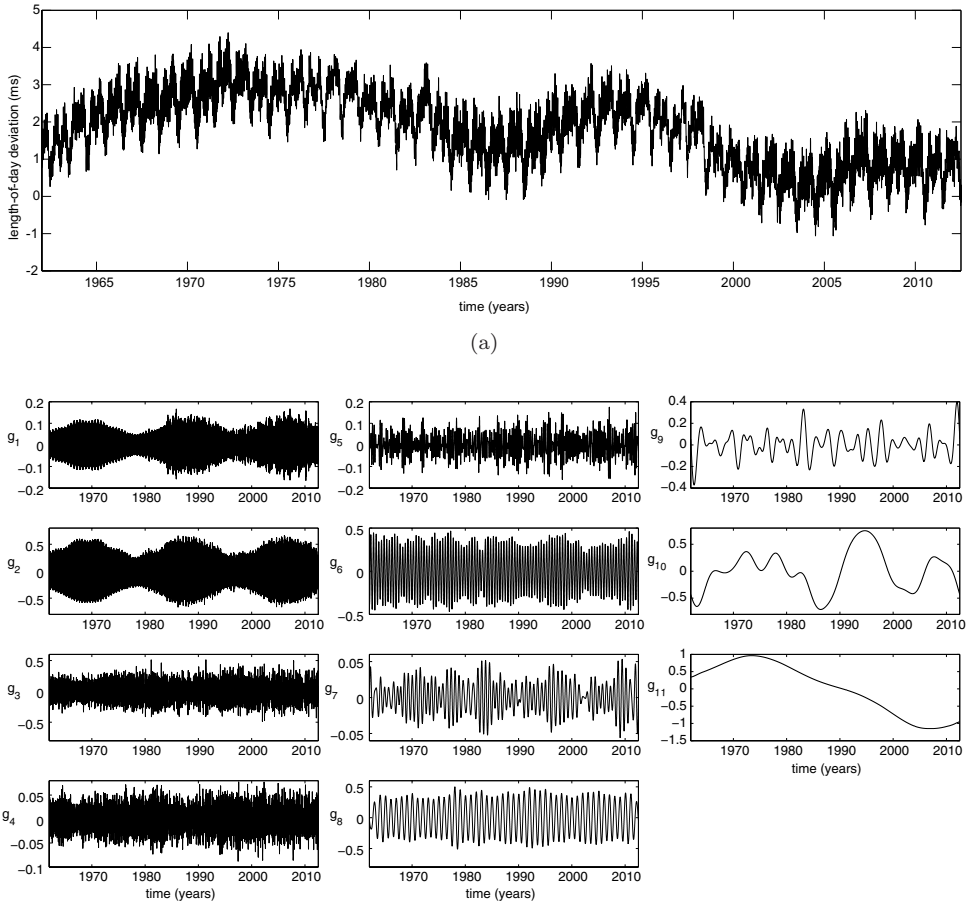


Fig. 15. (a) The daily LOD data, extracted from the comb2011_noon.eop series. (b) The eleven SSD-components for the LOD data.

The most likely origin of the oscillation is in association with fluid core motions and inner-core coupling [Holme and de Viron (2013)].

Figure 16 (left) depicts components nine and ten with superimposed a sine wave at a frequency of 2.5 years and about 6 years, respectively, to visually emphasize the presence of these two oscillations in the LOD data. The lower frequency content characterizing the second half of component ten and hiding the 6-year oscillations may be due to trend residuals. These components were also properly retrieved by EEMD (noise standard deviation of 0.2, Fig 16 right). This was obtained at the expense of a manual search for a suitable noise standard deviation, showing that SSD is able to achieve a decomposition as physically meaningful as EEMD, at the same time being free from optimal manual parameter setting for the problem at hand.

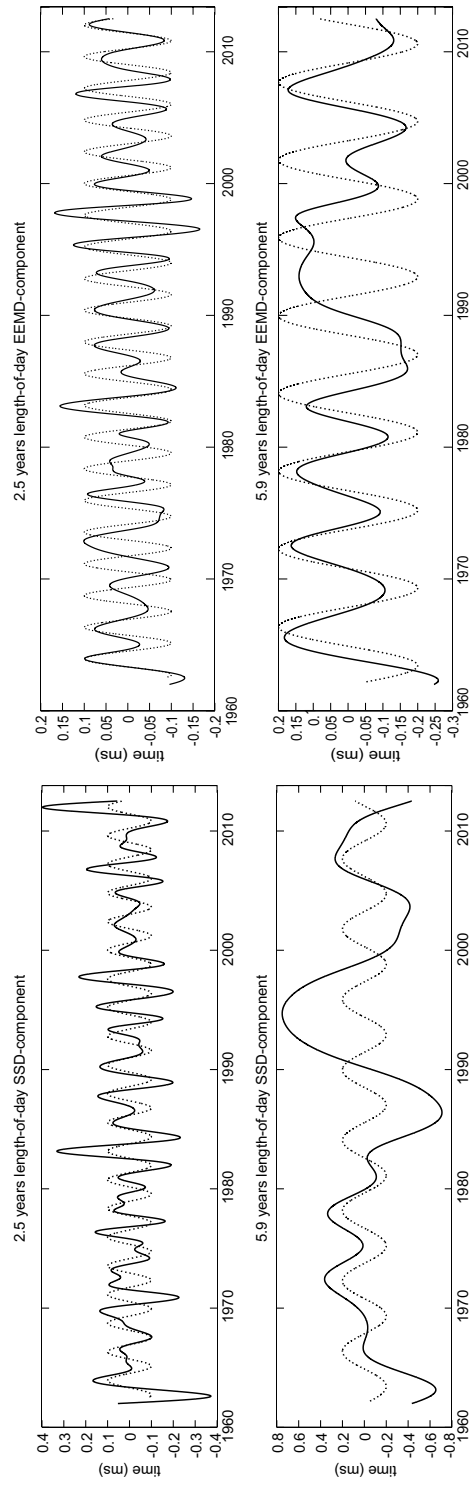


Fig. 16. Top: LOD SSD-components nine (top-left) and LOD EEMD-component with superimposed a sine wave at a period of 2.5 years. Bottom: LOD SSD-components ten and LOD EEMD-component with superimposed a sine wave at a period of about 6 years.

6. Discussion

In this paper, we have presented the SSD, a novel time series decomposition method for the identification of the various components (scales) concealed in the data. The component series retrieved by the SSD method do carry physical significance in general, as proven by the several numerical simulations and applications. The method is based on SSA, a nonparametric spectral estimation method proven to be successful in the analysis of noisy, nonlinear, and nonstationary short time series. With respect to the standard SSA algorithm, the main results are:

(i) The choice of the window length, or embedding dimension, for the construction of the trajectory matrix has been automated by setting it at the inverse of the dominant frequency of the data at a given iteration. With respect to the parameter ranges suggested in Vautard *et al.* [1992], this choice allows the window length to be sufficiently small to avoid the generation of spurious components off the spells where this oscillation resides, potentially given by larger windows.
(ii) The selection of the principal components of the trajectory matrix for the reconstruction of a specific scale has been automated, by selecting all the principal components with a dominant frequency within the estimated width of the dominant spectral peak ($2 \cdot \delta f$) of the residual at a given iteration. The estimate of the width $2 \cdot \delta f$ has been achieved by means of a Gaussian mixture model fitted to the power spectral density of the data.
(iii) A new definition for the trajectory matrix has been proposed which allows an effective use of all data and emphasizes the oscillations contained in the data. Moreover, the properties of this new matrix guarantee decrease of energy of the residual at each iteration.
(iv) A new set of parameters is derived from the data for each component series to extract, making the method totally adaptive.

The use of different window lengths to iteratively separate the components of data characterized by weak separability has been suggested in the literature under the name of "sequential" SSA [Golyandina *et al.* (2001); Varadi *et al.* (1999); Yiou *et al.* (2000)]. However, compared with the SSD method proposed in this paper, sequential SSA does not provide specific criteria (but few suggestions) for the choice of the window length, nor for the selection of the principal components of interest, but it employs the same main principles suggested for the caterpillar SSA, which are finally left to the user's subjectivity, rather than being embedded in the method. For oscillatory components, the identification of the associated oscillatory pair of eigentriples with the same period is implicitly performed by SSD [Vautard and Ghil (1989); Vautard *et al.* (1992); Ghil *et al.* (2002); Alexandrov and Golyandina (2005)]. Alexandrov proposed auto-SSA as a method which allow to use caterpillar-SSA approach in batch processing for solution of oscillatory components extraction [Alexandrov (2007)]. The method focuses on the selection of the pairs of eigentriples

representing specific periodic components. Even though this selection has been automated, the method strongly depends on the manual choice of two thresholds: One to check if the periodograms of SVD components from a pair have maxima at near frequency points, and the second to check whether the selected maxima values are large enough, to avoid false positive. Moreover, the method does not address the choice of the window length, leaving it to the user.

The SSD method has proven to be versatile, robust, and physically sound for a diverse selection of applications. However, there are aspects of it which need further attention. First, the trend identification and separation needs improvement. The current trend test, based on evaluating the contribution of the DC component in the original data, even though empirically reliable requires statistical strengthening. One alternative could be the Mann–Kendall trend test (a special case of the Kendall tau test) [Vautard *et al.* (1992)]. But our trials of the Mann–Kendall trend test at the first iteration of the SSD method showed that in some cases too many principal components may be selected for trend reconstruction, giving a trend component containing different scales. Therefore, we decided to achieve trend reconstruction by the first principal component only.

Second, in few occasions, when the energy of the residual signal falls below the noise level of the data, convergence slows as the Gaussian fitting procedure starts describing the numerous peaks characterizing the PSD. In this respect, a related open question is the choice of an appropriate stopping threshold for each application at hand. Care must be taken on the appropriateness of the default value of 1% of the original variance.

Third, the end effects in terms of spurious oscillations which may characterize the edges of the reconstructed SSD-components need further investigation. However, the numerical examples show that these end effects seem generally of small amplitude and they cannot be visually detected in the applications.

Finally, unlike Vautard *et al.* [1992], no amplitude threshold is used for the dominant PSD peak of the residual signal. We assume picking up the frequency describing most of the energy in the data, leaving everything not described for the following iteration(s).

The different results obtained for EMD and SSD should suggest that SSD offers a different point of view in the analysis of nonlinear time series with respect to EMD, rather than incorrectly suggesting that SSD performs better than EMD for those signals. A more detailed comparison of the two methods is left to future studies.

7. Conclusion

The SSD method has been presented which provides us with an adaptive energy-frequency based decomposition of the data in narrow-banded components. Its data-driven and adaptive nature and its effective use of the data make SSD a valuable alternative for decomposition and analysis of noisy, nonlinear, and nonstationary time series. The several examples and applications have testified the capability of

the method to yield physically significant components. Unlike standard SSA, in the SSD method the choice of the window length and selection of the principal components of the corresponding trajectory matrix for the reconstruction of a specific component series are automated and adaptive. Moreover, the new definition of the trajectory matrix allows for the oscillatory content in the data to be enhanced and better identified, and for the energy of the residual to decrease after each iteration.

Appendix A. About Convergence of the SSD Algorithm and Other Properties

A.1. *Convergence*

We consider the finite, discrete time series $x(n) = [x_1, x_2, \ldots, x_N]$. Given a window length equal M, the corresponding trajectory matrix, with the new definition given in Sec. 3.1, is a matrix of delayed versions of $x(n)$ with "wrap around" ($M \leq N$)

$$\mathbf{A} = \left[\begin{array}{cccc|ccc} x_1 & x_2 & \ldots & x_{N-M+1} & x_{N-M+2} & \ldots & x_N \\ x_2 & x_3 & \ldots & x_{N-M+2} & x_{N-M+3} & \ldots & x_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_M & x_{M+1} & \ldots & x_N & x_1 & \ldots & x_{M-1} \end{array}\right],$$

which is an $M \times N$ rectangular Hankel matrix. The Frobenius norm of $\mathbf{A}$ is

$$\|\mathbf{A}\|_F^2 = \sum_{i=1}^{M}\sum_{j=1}^{N} a_{ij}^2 = M\left(\sum_{j=1}^{N} x_j^2\right) = M\|x\|^2,$$

which is then equivalent to M times the energy of the signal $x(n)$ (where $\|x\|^2$ can be interpreted as the energy of the signal $x(n)$). The SVD of $\mathbf{A}$ gives $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}$, with $\mathbf{U}$ $M \times M$ orthogonal, $\mathbf{V}$ $N \times N$ orthogonal, and $\mathbf{D}$ $M \times N$, such that $D = \mathrm{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_M\}_{M\times N}$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_M \geq 0$. Given a subset $J \subset \{1, \ldots, M\}$ of selected indices for which the frequency content of the signal associated with $\sigma_j \mathbf{u}_j \mathbf{v}_j^{\mathrm{T}}$ ($j \in J$) overlaps well with the selected dominant mode of the residual signal v, we define

$$\mathbf{B} = \sum_{j\in J} \sigma_j \mathbf{u}_j \mathbf{v}_j^{\mathrm{T}},$$

where $\mathbf{u}_j$ denotes the jth column of U, and $\mathbf{v}_j$ the jth column of V. This gives a rank-k approximation $\mathbf{B}$ of $\mathbf{A}$, with k the number of elements in the set J. The remainder $\mathbf{C}$ is then

$$\mathbf{C} = \mathbf{A} - \mathbf{B} = \sum_{j\notin J}^{M} \sigma_j \mathbf{u}_j \mathbf{v}_j^{\mathrm{T}}.$$

Therefore,

$$\|\mathbf{A}\|_F^2 = \sum_{i=1}^{M} \sigma_i^2, \quad \|\mathbf{B}\|_F^2 = \sum_{j\in J} \sigma_j^2, \quad \|\mathbf{C}\|_F^2 = \sum_{j\notin J} \sigma_j^2$$

and

$$\|\mathbf{A}\|_F^2 = \|\mathbf{B}\|_F^2 + \|\mathbf{C}\|_F^2, \tag{A.1}$$

due to the "orthogonality" of the SVD. The averaging procedure along the cross-diagonals (with the procedure explained in Sec. 3.1) gives the average matrices $\overline{\mathbf{B}}$ and $\overline{\mathbf{C}}$, with each element replaced by the mean of its cross-diagonal. Since averaging is a linear operation

$$\mathbf{A} = \mathbf{B} + \mathbf{C} \rightarrow \overline{\mathbf{A}} = \overline{\mathbf{B}} + \overline{\mathbf{C}}.$$

Moreover, $\mathbf{A} = \overline{\mathbf{A}}$ by construction, hence $\mathbf{A} = \overline{\mathbf{B}} + \overline{\mathbf{C}}$. Given g as the signal associated with the selected mode and v the remainder

$$g \Leftrightarrow \overline{\mathbf{B}}, \quad v \Leftrightarrow \overline{\mathbf{C}},$$

with $g + v = x$, we have

$$\|\overline{\mathbf{B}}\|_F^2 = M\|g\|^2, \quad \|\overline{\mathbf{C}}\|_F^2 = M\|v\|^2.$$

Now, the averaging procedure decreases the total energy. Indeed, for all $s_1, \ldots, s_m \in \mathbb{R}$:

$$s_1^2 + \cdots + s_m^2 \geq m\left(\frac{s_1 + \cdots + s_m}{m}\right)^2.$$

The proof is straightforward and can be organized in a way that is quite common in statistics:

$$\frac{s_1^2 + \cdots + s_m^2}{m} = \mathbb{E}(s^2), \quad \left(\frac{s_1 + \cdots + s_m}{m}\right)^2 = \mathbb{E}(s)^2.$$

Denote $\mathbb{E}(s) = \mu$ and let $s = \mu + y$, with $\mathbb{E}(y) = 0$. Then:

$$\begin{aligned}\mathbb{E}(s^2) &= \mathbb{E}((\mu + y)^2) = \mu^2 + 2\mu\mathbb{E}(y) + \mathbb{E}(y^2) = \mu^2 + \mathbb{E}(y^2) \\ &= \mathbb{E}(s)^2 + \mathbb{E}(y^2) \geq \mathbb{E}(s)^2 \quad \text{because } \mathbb{E}(y^2) \geq 0.\end{aligned}$$

Consequently

$$\|\overline{\mathbf{B}}\|_F^2 \leq \|\mathbf{B}\|_F^2 \quad \text{and} \quad \|\overline{\mathbf{C}}\|_F^2 \leq \|\mathbf{C}\|_F^2.$$

Therefore, the monotonic decrease of energy follows:

$$\begin{aligned}M\|x\|^2 = \|\mathbf{A}\|_F^2 \geq \|\mathbf{C}\|_F^2 \geq \|\overline{\mathbf{C}}\|_F^2 = M\|v\|^2, \\ M\|x\|^2 = \|\mathbf{A}\|_F^2 \geq \|\mathbf{B}\|_F^2 \geq \|\overline{\mathbf{B}}\|_F^2 = M\|g\|^2.\end{aligned}$$

This result together with Eq. (A.1) gives

$$\|v\|^2 + \|g\|^2 \leq \|x\|^2, \tag{A.2}$$

showing that a certain amount of energy is captured at each iteration. As the set J varies for each iteration, there is no guaranteed rate of decrease of energy per iteration, and no guarantee for the algorithm to avoid local minima. However,

the inclusion in J of the principal component which contributes the most to the energy of the dominant peak of the selected mode should help avoid this unwanted circumstance. The additional final scaling applied to the derived SSD-component g (described in Secs. 3.3 and 3.4) is such that

$$\tilde{g} = \hat{a}g \text{ with } \hat{a} = \frac{g^T x}{g^T g} = \frac{g^T(g+v)}{g^T g} = 1 + \frac{g^T v}{g^T g} \geq 1.$$

Given $x = \tilde{g} + \tilde{v}$

$$\|x\|^2 = \|\tilde{g}\|^2 + \|\tilde{v}\|^2 = |a|^2\|g\|^2 + \|\tilde{v}\|^2 \quad \text{and} \quad \|x\|^2 \geq \|g\|^2 + \|\tilde{v}\|^2,$$

such that $\|\tilde{v}\|^2 \leq \|v\|^2$. Therefore, scaling has the effect of increasing the energy of g, thus it decreases by construction the energy of the residual v, improving convergence.

A.2. *Condition for the second run*

From Eq. (A.2) and the results of the previous section, we know that

$$x = g + v \rightarrow \|x\|^2 = \|g\|^2 + \|v\|^2 + 2g^{\mathrm{T}}v \geq 0 \quad \text{and} \quad \|x\|^2 \geq \|g\|^2 + \|v\|^2,$$

for which we can conclude that

$$g^{\mathrm{T}}v \geq 0.$$

This condition is satisfied by construction for $g_{[1]}$. However, it needs to be tested on $g_{[2]}$. In particular, $g_{[2]}^{\mathrm{T}}v > 0$, from which it follows:

$$\|x\|^2 \geq \|g_{[2]}\|^2 + \|v\|^2 \quad \text{and} \quad \|v\|^2 \leq \|x\|^2.$$

A.3. *Zero-mean nature of the residual*

Another important effect of the wrapped around definition of the trajectory matrix is that starting with a zero-mean signal x such that $x = g + v$, the signal g associated with the selected mode and the residual v will also have zero-mean. Given a zero-mean signal x and the definition for $\mathbf{A}$ in Sec. A.1 we have that

$$(1, 1, \ldots, 1)\mathbf{A}^{\mathrm{T}} = (0, 0, \ldots, 0).$$

Combining this with the SVD: $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}$, it becomes

$$(1, 1, \ldots, 1)\mathbf{V}\mathbf{D}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}} = (0, 0, \ldots, 0) \rightarrow (1, 1, \ldots, 1)\mathbf{V}\mathbf{D}^{\mathrm{T}} = (0, 0, \ldots, 0),$$

$$(1, 1, \ldots, 1)[\mathbf{v}_1\mathbf{v}_2 \ldots \mathbf{v}_N] \begin{bmatrix} \sigma_1 & 0 & \ldots & 0 \\ 0 & \sigma_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_M \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix} = (0, 0, \ldots, 0)$$

$$\rightarrow (1, 1, \ldots, 1)[\sigma_1\mathbf{v}_1\sigma_2\mathbf{v}_2 \ldots \sigma_M\mathbf{v}_M] = (0, 0, \ldots, 0),$$

from which it follows that for any column $j \leq M$ of $\mathbf{V}$

$$(1, 1, \ldots, 1)\sigma_j \mathbf{v}_j = 0 \quad \text{and} \quad (1, 1, \ldots, 1)\mathbf{v}_j = N\overline{\mathbf{v}}_j = 0,$$

with $\overline{\mathbf{v}}_j$ being the average of vector $\mathbf{v}_j$. Given any selection of indices $J \subset \{1, \ldots, M\}$ (with $\sigma_j > 0$) we have

$$\mathbf{B} = \sum_{j \in J} \sigma_j \mathbf{u}_j \mathbf{v}_j^{\mathrm{T}} \quad \text{and}$$

$$(1, 1, \ldots, 1)\mathbf{B}^{\mathrm{T}} = (1, 1, \ldots, 1) \sum_{j \in J} \sigma_j \mathbf{v}_j \mathbf{u}_j^{\mathrm{T}} = N \sum_{j \in J} \sigma_j \overline{\mathbf{v}}_j \mathbf{u}_j^{\mathrm{T}} = (0, 0, \ldots, 0).$$

Therefore, the sum of the elements of any row of $\mathbf{B}$ gives 0, and so must do the sum of all its elements: $\sum_{i=1}^{M} \sum_{j=1}^{N} b_{ij} = 0$. Now, if each element in $\mathbf{B}$ is replaced by the average along the corresponding cross-diagonal (with the wrap-around procedure explained in Sec. 3.1), such as to obtain $\overline{\mathbf{B}}$, then

$$\sum_{i=1}^{M} \sum_{j=1}^{N} \bar{b}_{ij} = 0,$$

and because the matrix $\overline{\mathbf{B}}$ is again a Hankel matrix (with wrap around), having the same entries in each of its rows (but in shifted positions) we have for every row i

$$(1, 1, \ldots, 1)\mathbf{b}_i^{\mathrm{T}} = \bar{g} = 0.$$

This proves the selected mode to be zero-mean. This also holds for the residual v associated with $\mathbf{C} = \mathbf{A} - \mathbf{B}$. Moreover, it holds when $\tilde{g} = \hat{a}g$ is computed.

References

Alexandrov, T. and Golyandina, N. (2005). Automatic extraction and forecast of time series cyclic components within the framework of SSA. *Proc. 5th St. Petersburg Workshop on Simulation*, **5**: 45–50.

Alexandrov, T. (2007). A method of extraction of quasi-periodic time series components using singular spectrum analysis. *EC2 Conf.*

Bonizzi, P. *et al.* (2012). Singular spectrum analysis improves analysis of local field potentials from macaque V1 in active fixation task. *Proc. IEEE EMBS*, **34**: 2945–2948.

Corino, V. D. A. *et al.* (2008). Improved time-frequency analysis of atrial fibrillation signals using spectral modeling. *IEEE Trans. Biomed. Eng.*, **55**: 2723–2730.

Fries, P. *et al.* (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, **291**: 1560–1563.

Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends Cogn Sci.*, **9**: 474–480.

Ghil, M. *et al.* (2002). Advanced spectral methods for climatic time series. *Rev. Geophys.*, **40**: 3.1–3.41.

Golyandina, N., Nekrutkin, V. and Zhigljavsky, A. A. (2001). *Analysis of Time Series Structure: SSA and Related Techniques*. CRS Press, Boca Raton.

Golyandina, N. and Osipov, E. (2007). The caterpillar-SSA method for analysis of time series with missing values. *J. Stat. Plan. Inference*, **137**: 2642–2653.

Golyandina, N. (2010). On the choice of parameters in singular spectrum analysis and related subspace-based methods. *Stat. Interface*, **3**: 259–279.

Holme, R. and de Viron, O. (2013). Characterization and implications of intradecadal variations in length of day. *Nature*, **499**: 202–204.

Huang, N. E. *et al.* (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. A*, **454**: 903–995.

Huang, N. E., Shen, Z. and Long, S. R. (1999). A new view of nonlinear water waves: The Hilbert Spectrum. *Annu. Rev. Fluid Mech.*, **31**: 417–457.

Huang, N. E. *et al.* (2003). A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proc. R. Soc. A*, **459**: 2317–2345.

Liang, H. *et al.* (2005). Empirical mode decomposition of field potentials from macaque V4 in visual spatial attention. *Biol. Cybern.*, **92**: 380–392.

Lourakis, M. I. A. and Argyros, A. A. (2005). Is Levenberg–Marquardt the most efficient optimization algorithm for implementing bundle adjustment? *Proc. Tenth IEEE Int. Conf. Comput. Vis.*, **2**: 1526–1531.

Maier, A. *et al.* (2010). Distinct superficial and deep laminar domains of activity in the visual cortex during rest and stimulation. *Front. Syst. Neurosci.*, **4**: 1–11.

Polukoshko, S. and Hofmanis, J. (2009). Use of Caterpillar-SSA for the analysis and forecasting of industrial and economic indicators. *Proc. 7th Int. Scientific and Practical Conf.*, **II**: 241–248.

Roberts, M., Fries, P. and de Weerd, P. (2010). The peak frequency of the gamma band shifts with stimulus contrast. *FENS Forum.*

Silva, L., Jackson, L. and Mound, J. (2012). Assessing the importance and expression of the 6 year geomagnetic oscillation. *J. Geophys. Res.*, **117**, B10101.

Vautard, R. and Ghil, M. (1989). Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D*, **35**, 3: 395–424.

Vautard, R., Yiou, P. and Ghil, M. (1992). Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Phys. D, Nonlinear Phenom.*, **58**: 95–126.

Varadi, F. *et al.* (1999). Searching for signal in noise by random-lag singular spectrum analysis. *Astrophys. J.*, **526**: 1052–1061.

Wu, Z. and Huang, N. E. (2009). Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.*, **1**: 1–41.

Yiou, P., Sornette, D. and Ghil, M. (2000). Data-adaptive wavelets and multi-scale singular spectrum analysis. *Physica D*, **142**: 254–290.

Zhigljavsky, A. (2010). Singular spectrum analysis for time series: Introduction to this special issue. *Stat. Interface*, **3**: 255–258.