

Sliding Window Empirical Mode Decomposition -its performance and quality

Pawel Stepien

Correspondence:
pstepien@ibib.waw.pl
Nalecz Institute of Biocybernetics
and Biomedical Engineering PAS,
Warsaw, Poland

Abstract

Background: In analysis of nonstationary nonlinear signals the classical notion of frequency is meaningless. Instead one may use *Instantaneous Frequency (IF)* that can be interpreted as the frequency of a sine wave which locally fits the signal. *IF* is meaningful for monocomponent nonstationary signals and may be calculated by *Hilbert transform (HT)*.

Methods: A multicomponent signal may be decomposed into its monocomponents. *Empirical Mode Decomposition (EMD)*, developed by Norden E. Huang, is a new method of such breaking down of a signal into its monocomponents. *EMD* combined with *HT* (called *Hilbert-Huang Transform*) is a good tool for analyzing nonstationary signals, but unfortunately the traditional *EMD* algorithm consumes a lot of time and computer resources. I propose a modified *EMD* algorithm - *Sliding Window EMD, SWEMD*.

Results: Proposed algorithm speeds up (about 10 times) the computation with acceptable quality of decomposition.

Conclusions: *Sliding Window EMD* algorithm is suitable for decomposition of long signals with high sampling frequency.

Keywords: Empirical Mode Decomposition; EMD; Fast EMD; Sliding Window Empirical Mode Decomposition; SWEMD; Signal analysis

Background

Empirical Mode Decomposition (EMD) is a new method of analyzing nonlinear nonstationary signals. It was proposed by Norden E. Huang in 1996 [1]. *EMD* is an entirely data-driven algorithm and it does not depend on any predefined basis functions. *EMD* breaks down nonstationary, multicomponent signal into its monocomponents. Such monocomponents are called *Intrinsic Mode Functions (IMFs)* (see subsection 'Intrinsic mode function'). Analyzing long signals with *EMD* is time consuming or even impossible in reasonable time due to the fact, that spline interpolation of large number of points takes a lot of computer resources. For such signals 'classical algorithm' needs some modifications. There are several modifications of *EMD* improving its performance. C. D. Blakely developed *FastEMD* method [2]; instead of cubic spline interpolation, matrix-free moving least-squares approximation is used. Q. Chen et al. proposed representing of *IMFs* as B-splines [3]. This method was tested on earthquake records. It decomposes signal into more modes than 'classical' *EMD* so it gives better

frequency resolution. Unfortunately authors did not give information about time performance of this method. S. Meignen and V. Perrier [4] eliminate sifting process (subsection 'EMD algorithm') by calculating *IMFs* through the resolution of a quadratic programming problem with equality and inequality constraints. This method should greatly accelerate decomposition of short signals (no sifting iterations), but for long signals it can take longer than 'classical' *EMD* (quadratic optimization problem is solved in polynomial time, spline interpolation is solved in linear time). A. Roy and J. F. Doherty [5] use the raised cosine filter based interpolation instead splines, *RCEMD*. They tested it on signal being a superposition of two harmonics with different frequencies. Depending on harmonic separation complexity (the ratio of the frequencies), calculations were performed from 0 to 10 times faster than when using the 'classical' method. Their method also needs lower sampling rates of signal than the 'classical' *EMD* [6]. Further investigation of this method [7] shows that *RCEMD* algorithm is generally faster than *EMD* for short signal, so the authors developed windowed version of *RCEMD* which speeds up calculation about 10 times.

For long data sets, I suggest to calculate *EMD* in a small window and to slide this window along the time axis (*Sliding Window EMD*, *SWEMD*) [8]. In this paper I describe proposed algorithm (subsection 'Sliding Window Empirical Mode Decomposition'), test its performance (subsection 'Performance of *SWEMD*') and quality (subsection 'Quality of *SWEMD*'). Then I compare time-frequency characteristics of specimen signals obtained by my algorithm, with those obtained by 'classical' *EMD* and those obtained by typical calculation of *EMD* in time windows (subsection 'Quality of *SWEMD*').

Methods

Empirical Mode Decomposition

Intrinsic mode function

EMD decomposes signal into, so called, *intrinsic mode functions (IMFs)*. Each *IMF* is a signal that must meet the following criteria [1]:

- the number of extrema and zeros are equal or their difference is not greater than 1,
- the signal has "zero mean" - the mean value of the envelope determined by maxima and the envelope determined by minima is equal 0 at every point.

These conditions show the essence of the *EMD* decomposition: non-stationary signal is decomposed into symmetrical oscillations around zero amplitude, which are easy for further analysis.

EMD algorithm

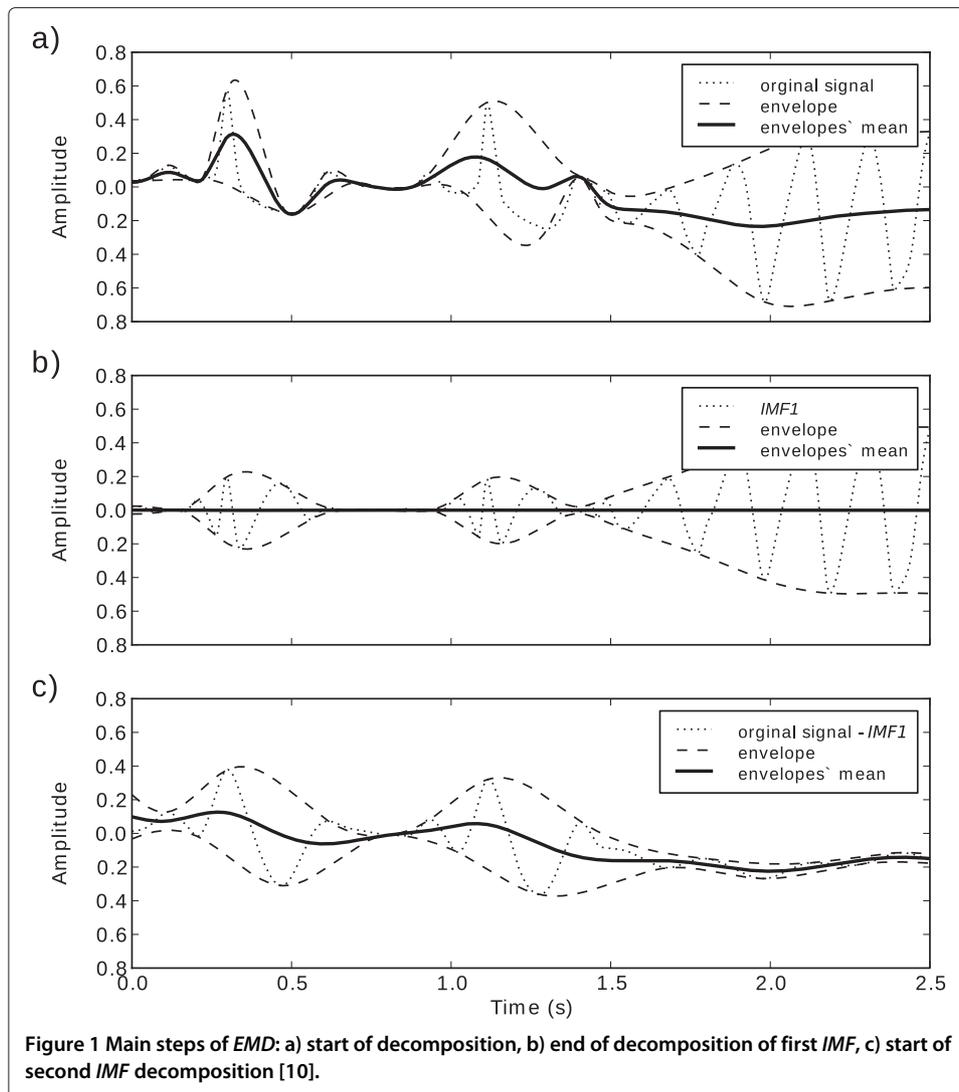
The first step of the *EMD* algorithm [1,9] is extraction of extrema from original signal $x(t)$ and creation of the upper envelope e_{max} by cubic spline interpolation of maxima and of the lower envelope e_{min} by interpolation of minima (Figure 1a). Then mean value of envelopes is calculated:

$$m(t) = \frac{e_{max} + e_{min}}{2}. \quad (1)$$

The mean value from Eq. 1 is subtracted from original data:

$$imf_1(t) = x(t) - m(t). \quad (2)$$

Above procedure is called *sifting process*.



Then $imf_1(t)$ is treated as input data for next *sifting process*. Mean value $m(t)$ of envelopes of $imf_1(t)$ is calculated and this value is subtracted from $imf_1(t)$:

$$imf_1(t) := imf_1(t) - m(t), \quad (3)$$

where ' $:=$ ' means *it becomes*.

Sifting process is iterated till $imf_1(t)$ fulfills conditions of IMF signal ((Figure 1b), see subsection 'Intrinsic mode function'). When the first *sifting process* is finished (see subsection 'Sifting process stopping criteria') then the original signal is reduced by the first mode (Figure 1c):

$$IMF_1(t) := imf_1(t) \quad (4)$$

$$r_1(t) = x(t) - IMF(t). \quad (5)$$

The residue $r_1(t)$ of subtraction (5) is treated as input data for extraction of second IMF (next *sifting loop*). Procedure is looped to extract all IMFs:

$$r_i(t) = r_{i-1}(t) - imf_i(t), \quad (6)$$

where i is index of current mode. Decomposition is finished when residue $r_i(t)$ has less than three extrema (there is no possibility to construct envelopes) or all its points are nearly equal zero.

The sum of all *IMF* components and the residue gives the original signal:

$$r_n + \sum_{i=1}^n imf_i(t) = x(t), \quad (7)$$

where n is number of all modes.

Algorithm of *EMD* is shown, as a block diagram, in Figure 2.

Sifting process stopping criteria

According to the second criteria of *IMF*, the mean of its envelope is equal to 0 at every point of *IMF*. In fact, it is very difficult to get numerically this value equal to zero at every point. What is needed is an approximation. In my work I use the stop conditions of the sifting procedure proposed by Rilling et al. [9]. In each iteration ratio of the mean value of the envelope of iterated mode and the amplitude of this envelope is checked:

$$\sigma(t) = \left| \frac{m(t)}{a(t)} \right|, \quad (8)$$

where:

$$m(t) = \frac{e_{max}(t) + e_{min}(t)}{2}; \quad a(t) = \frac{e_{max}(t) - e_{min}(t)}{2} \quad (9)$$

and e_{max} is the upper envelope, e_{min} - the lower envelope.

There are two thresholds:

- θ_1 - guaranteeing globally small fluctuations of the envelopes' mean around zero,
- θ_2 - locally allowing higher fluctuations.

Sifting process is stopped if $\sigma(t) < \theta_1$ is true for $(1 - \alpha)$ part of the number of signal's points, and if $\sigma(t) < \theta_2$ is true for the rest of points. Typical values for these parameters are as follows:

$$\alpha = 0,05; \quad \theta_1 = 0,05; \quad \theta_2 = 10 \cdot \theta_1. \quad (10)$$

These values of parameters allow a compromise between quality and speed of decomposition process.

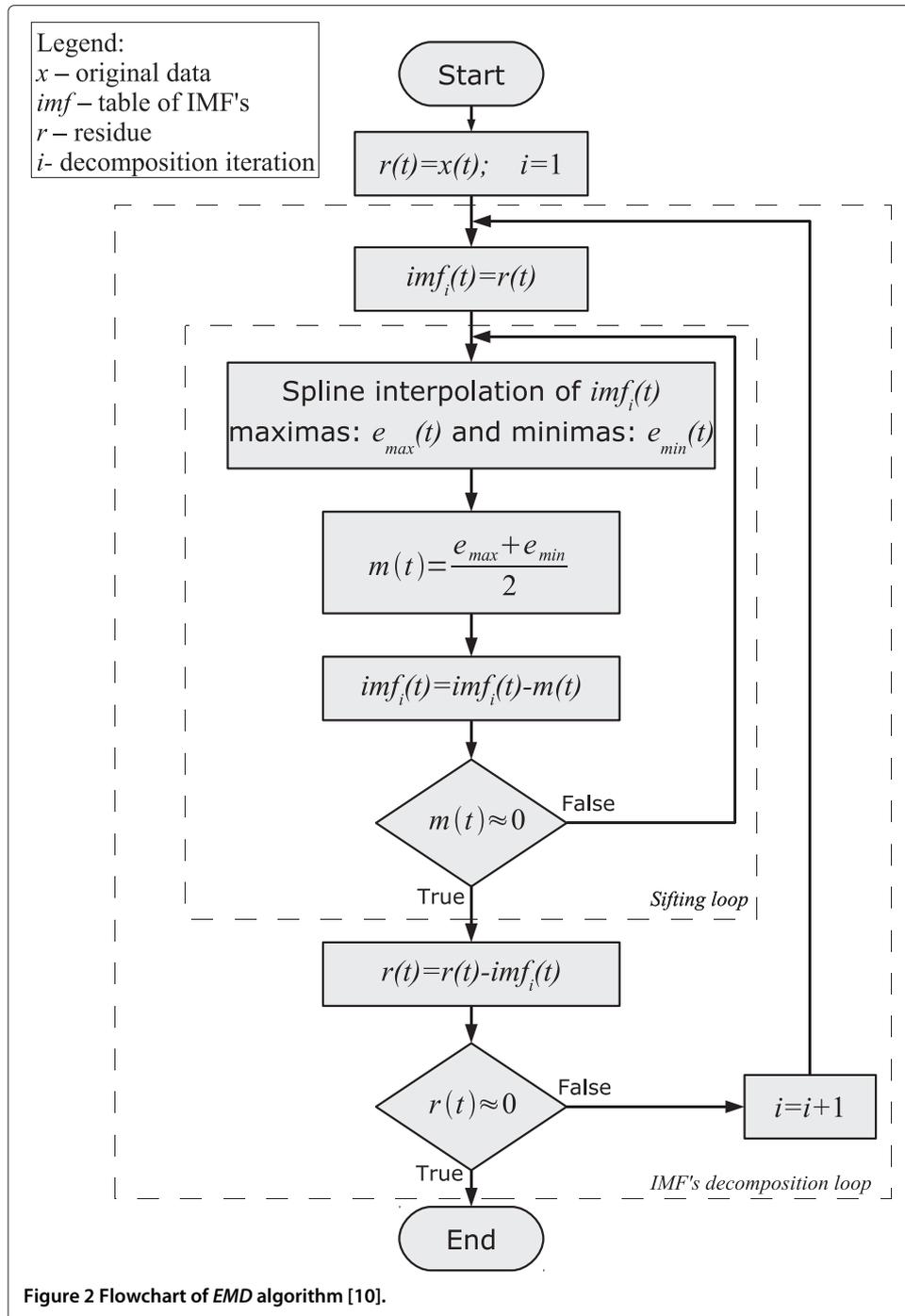
Hilbert-Huang spectrum

Signal decomposed into *IMFs* can be further analyzed with time-frequency characteristic by obtaining *Hilbert-Huang spectrum*. For this purpose, the analytic signal is created from each *EMD* mode:

$$imf_{a_k}(t) = imf_k(t) + jH[imf_k(t)] \quad (11)$$

where $H(imf_k(t))$ is *Hilbert transform* of k -th *IMF* and j means imaginary part. *Hilbert transform* is defined as:

$$H[x(t)] = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x(t')}{t - t'} dt', \quad (12)$$



where P denotes Cauchy principal value:

$$P \int_{-\infty}^{\infty} f(x) dx = \lim_{\rho \rightarrow \infty} \int_{-\rho}^{\rho} f(x) dx. \quad (13)$$

From analytic signal the instantaneous amplitude can be obtained as a module of this signal:

$$a_k(t) = |imf_{a_k}(t)| \quad (14)$$

and the instantaneous frequency as a differential of argument of this signal:

$$f_k(t) = \frac{1}{2\pi} \frac{d \arg(\text{imf}_{a_k}(t))}{dt}. \quad (15)$$

The sum of all modes' instantaneous frequencies and amplitudes gives the *Hilbert-Huang Spectrum (HSS)*:

$$HHS(f_k, t) = \text{Re} \left[\sum_{k=1}^N a_k(t)^j \int f_k(t) dt \right]. \quad (16)$$

With *HHS*, the *marginal Hilbert-Huang Spectrum* can be calculated. It gives information about the contribution of a given frequency to the total power and is defined as follows:

$$hhs(f_k) = \int_{-\infty}^{\infty} HHS(f_k, t) dt. \quad (17)$$

Sliding Window Empirical Mode Decomposition

Sliding Window Empirical Mode Decomposition, SWEMD, is based on calculation of *EMD* in a relatively small window and sliding this window along the time axis [8,10]. Size of the window depends on the signal's frequency band - there must be at least 5 maxima and 5 minima in the window for correct spline interpolation. To prevent possible discontinuities in *IMFs* between windows, the total number of modes and also the number of sifting iterations must be set *a priori* [9]. The number of sifting steps should be tailored for each module. These parameter depends on the sampling frequency and on analyzed signal, its complexity and spectrum. The empirical determination of these values can be difficult. Thus, I propose a simple algorithm for determining the number of sifting iterations: decomposition of several signal's windows by classic algorithm and counting the average number of sifting steps for each mod.

The block diagram of *SWEMD* algorithm is displayed in Figure 3. To eliminate possible windows' boundary effects, surroundings points (*S* in Figure 3) are added to the beginning and end of the window. I suggest to use the number of surroundings points equal to the size of the window or its multiple. *EMD* sifting process is done on the window and surroundings part of the data. Calculation of all the modes in the window at time *t* is called *SWEMD* step.

During calculation performed on the subsequent modes in a *SWEMD* step, it may occur (especially for modes containing low frequencies) that the window size is too small for spline interpolation (not enough extrema). In such a case, the calculation for this mode (and the next modes) is canceled in this *SWEMD* step. In the next step these modes will be calculated in a longer window. Therefore, in the implementation of the algorithm, the table for storing indices of the windows' start point is used (*Ind* in Figure 3). If extraction of mode *M* in *SWEMD* step at time *t* is successful then the array element corresponding to the given mode (*Ind[M]*) is assigned the value of the last window point: *t + win*. In a case of too small number of extrema, *Ind[M]* remains unchanged, therefore, the size of the window in the next step will be greater by *win* points from the basic size: *win*.

The calculation of the mode *M* in the *SWEMD* step begins with extracting from the analyzed signal *y* points corresponding to the window in time *t* and its surroundings (in Figure 3 these data are assigned to the variable *R*). Then, in order to prepare input data for

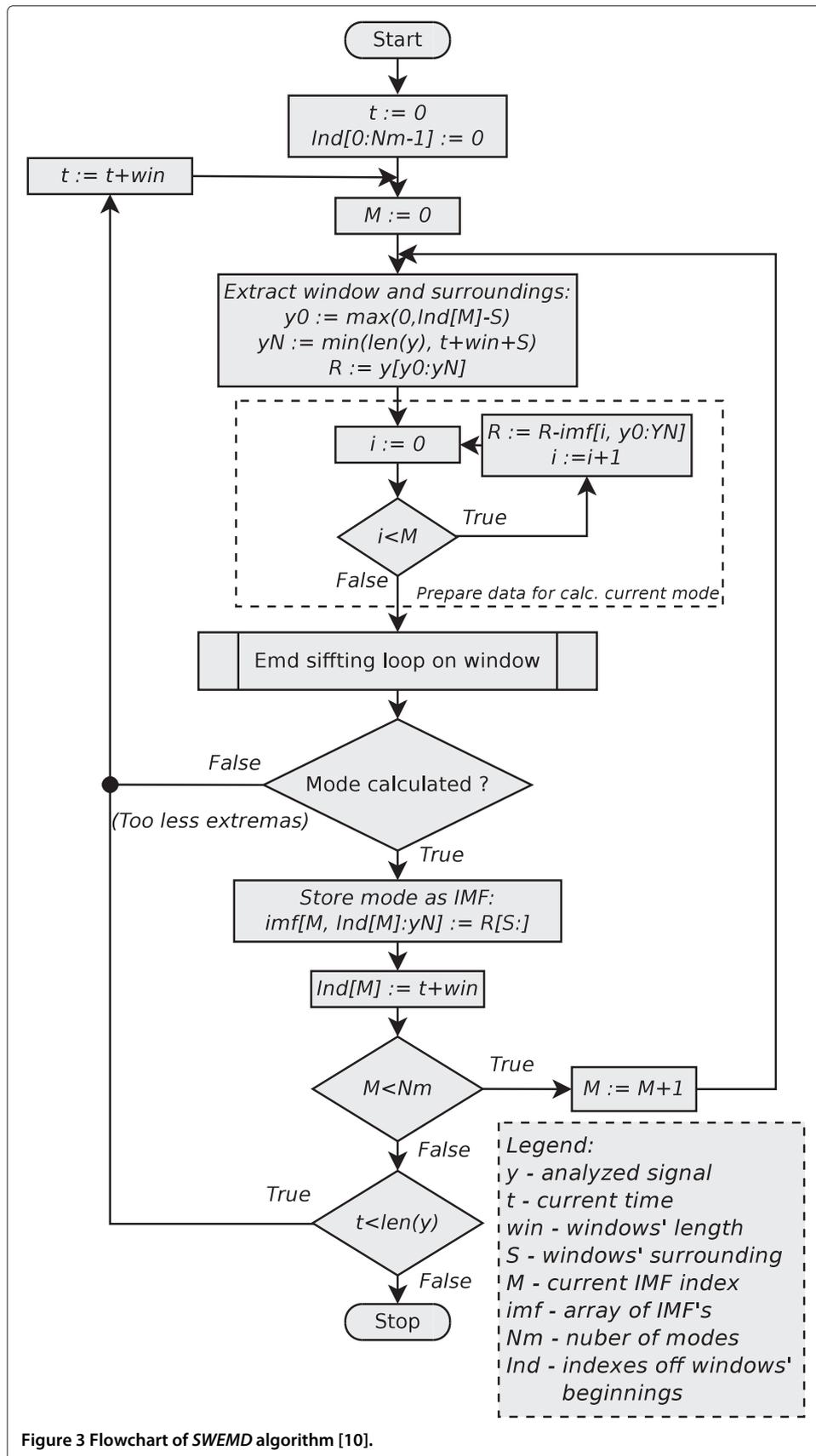
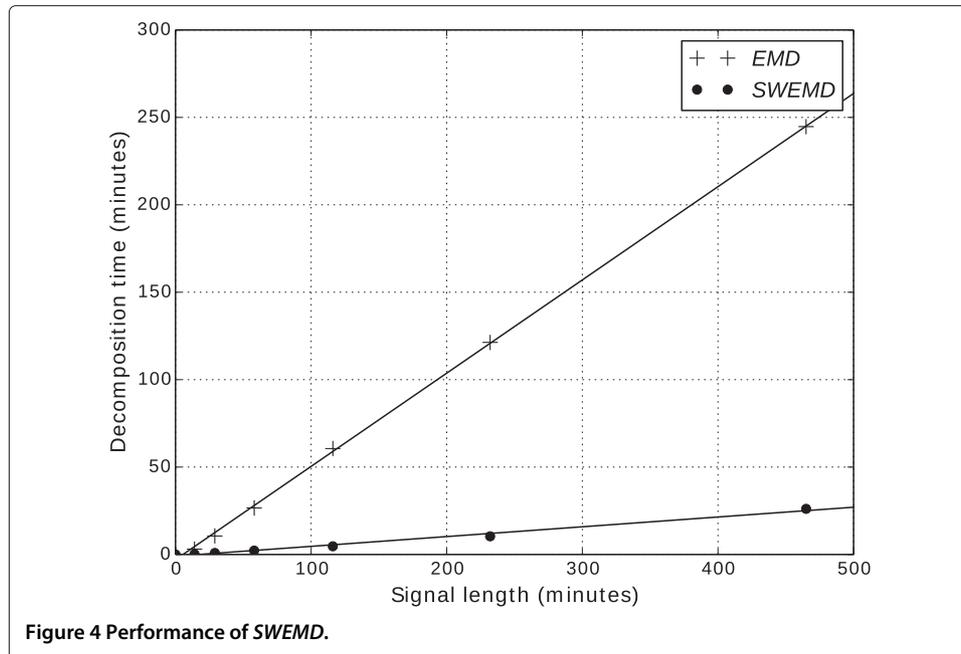
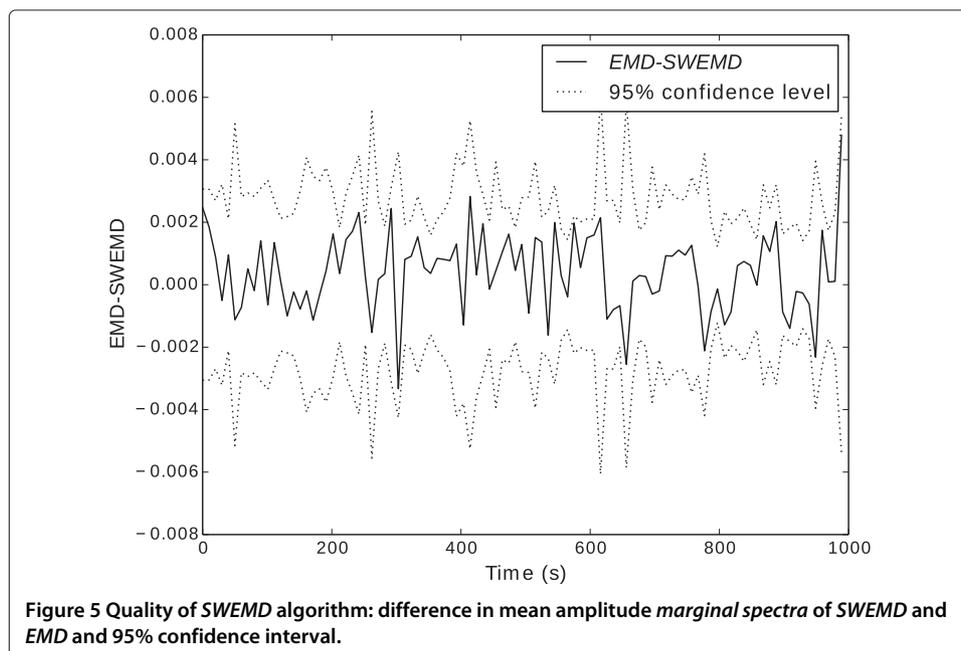


Figure 3 Flowchart of SWEMD algorithm [10].



calculation of the mode M , the signal R is reduced by the $IMFs$ of orders lower than M (in Figure 3 block labeled “Prepare data for calc. Current mode”). Prepared data are subjected to the *sifting process*. If there is not enough extrema, calculations in this $SWEMD$ step are interrupted and the counting of the first IMF in the next step ($t + win$) is started. If the *sifting process* is completed correctly, the resulting mode (after discarding surroundings) is placed in the array $imf[.,.]$ (Figure 3) that stores $IMFs$ at appropriate indices. Then, the value of the start point of the window for the mode M in the next $SWEMD$ step is changed and a process of calculating the module $M + 1$ is started or current step is completed (if mode M was the last to achieve).



Results and discussion

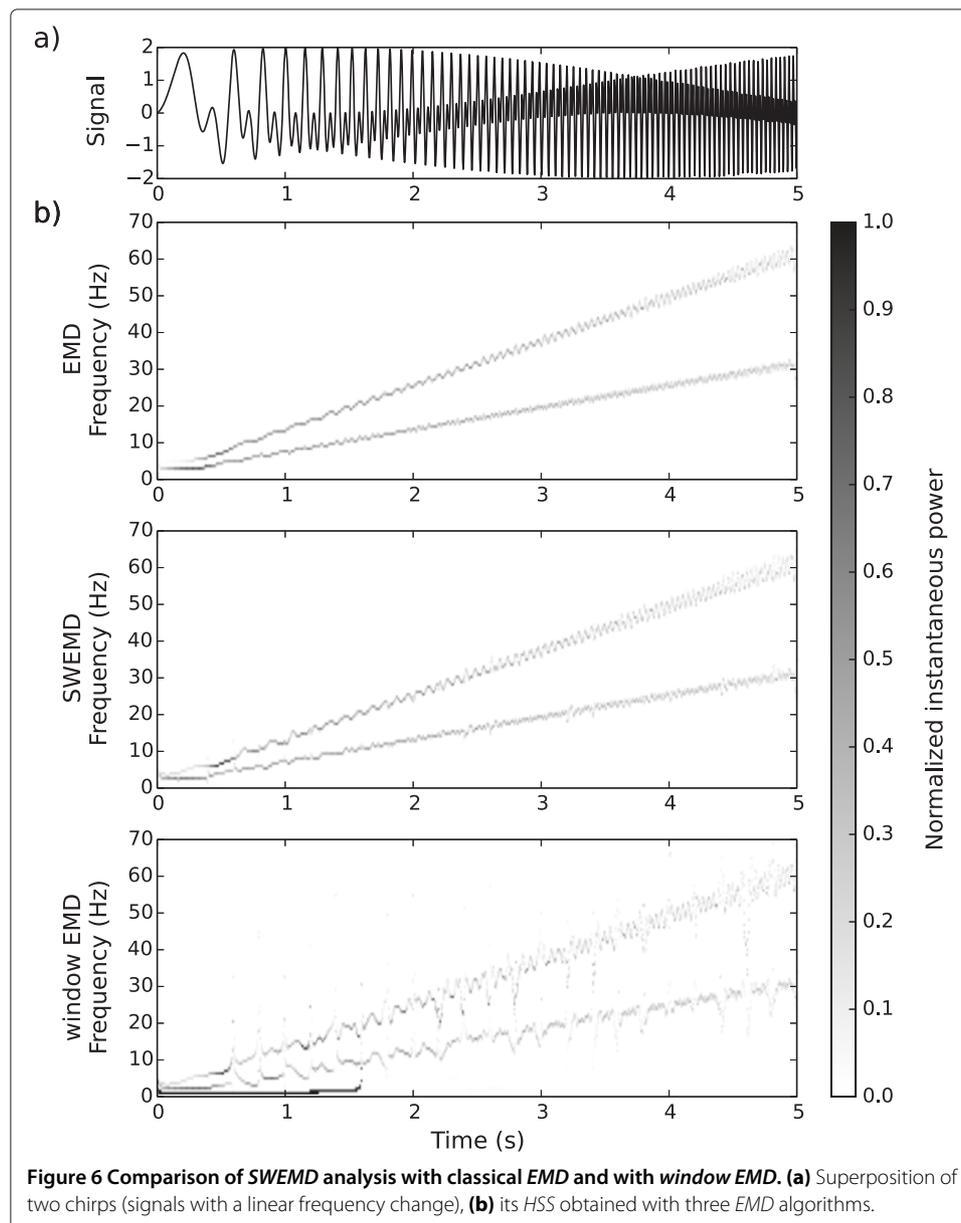
Performance of SWEMD

Performance was tested on several *EEG* signals with different length. Figure 4 shows the comparison of the time of calculation *SWEMD* algorithm to 'classical' *EMD* depending on the length of the analyzed signal. Proposed algorithm performs calculations about 10 times faster.

Quality of SWEMD

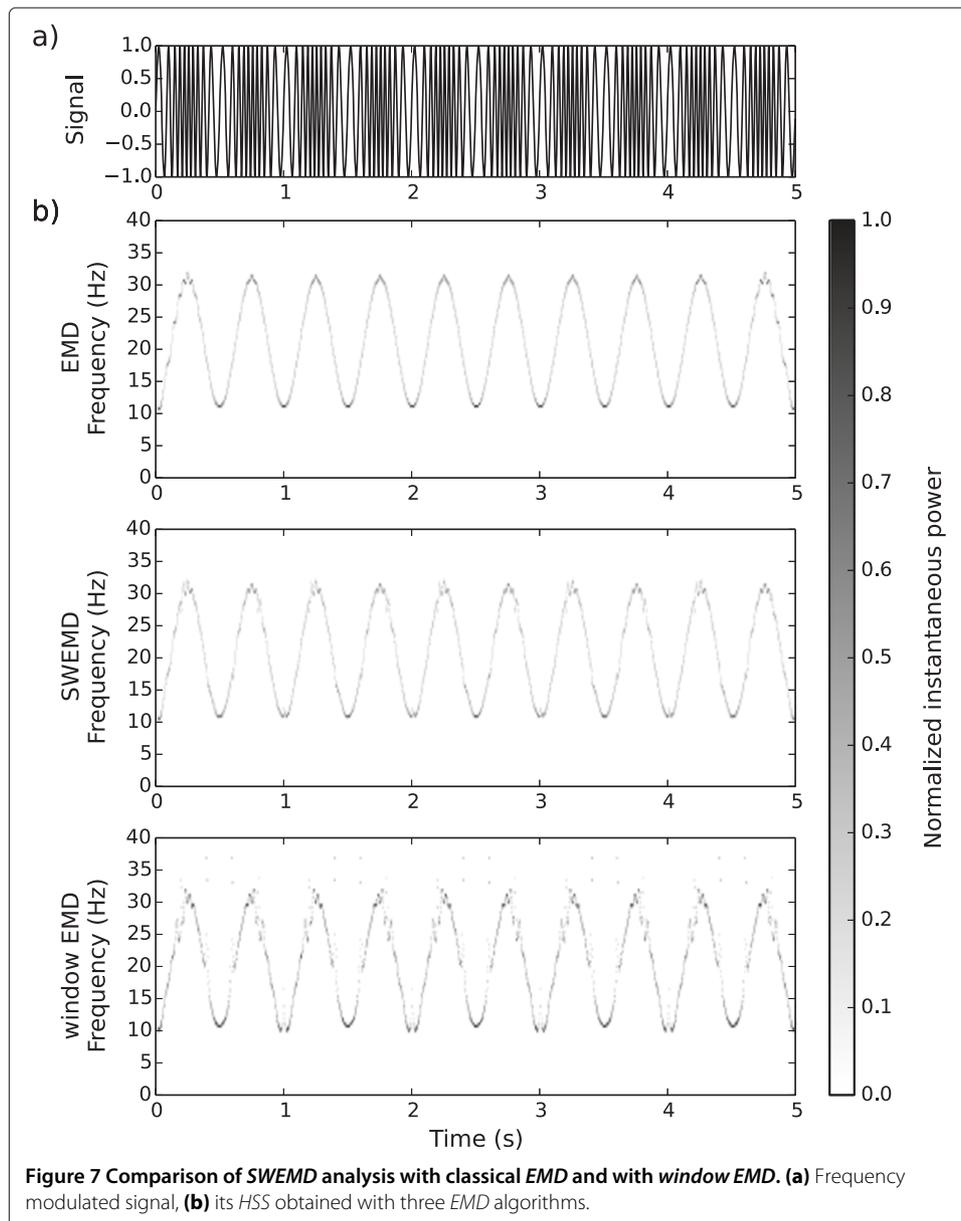
In order to verify the quality of the proposed algorithm I compared the results *SWEMD* and 'classical' *EMD*. The test consisted of the following steps:

- decomposition of the test *EEG* signal using *EMD* and *SWEMD*;



- calculation of the *marginal Hilbert-Huang spectrum* for both *EMD* and *SWEMD* in time windows length of 10 seconds (the same window was used in the calculations of *SWEMD*);
- calculation of the difference of the *EMD* and *SWEMD* marginal spectra's mean amplitudes and the standard deviation of this difference;
- designate 95% confidence interval [11] for this difference

The test was performed on several *EEG* signals. The specimen results of the quality test of the algorithm *SWEMD* is shown in Figure 5. As it is seen in Figure 5 the difference in the amplitudes of marginal spectra of *EMD* and *SWEMD* at any point is approximately zero (solid line in Figure 5), and lies in the 95% confidence interval (dotted line in Figure 5). Therefore the results of both algorithms are statistically consistent.



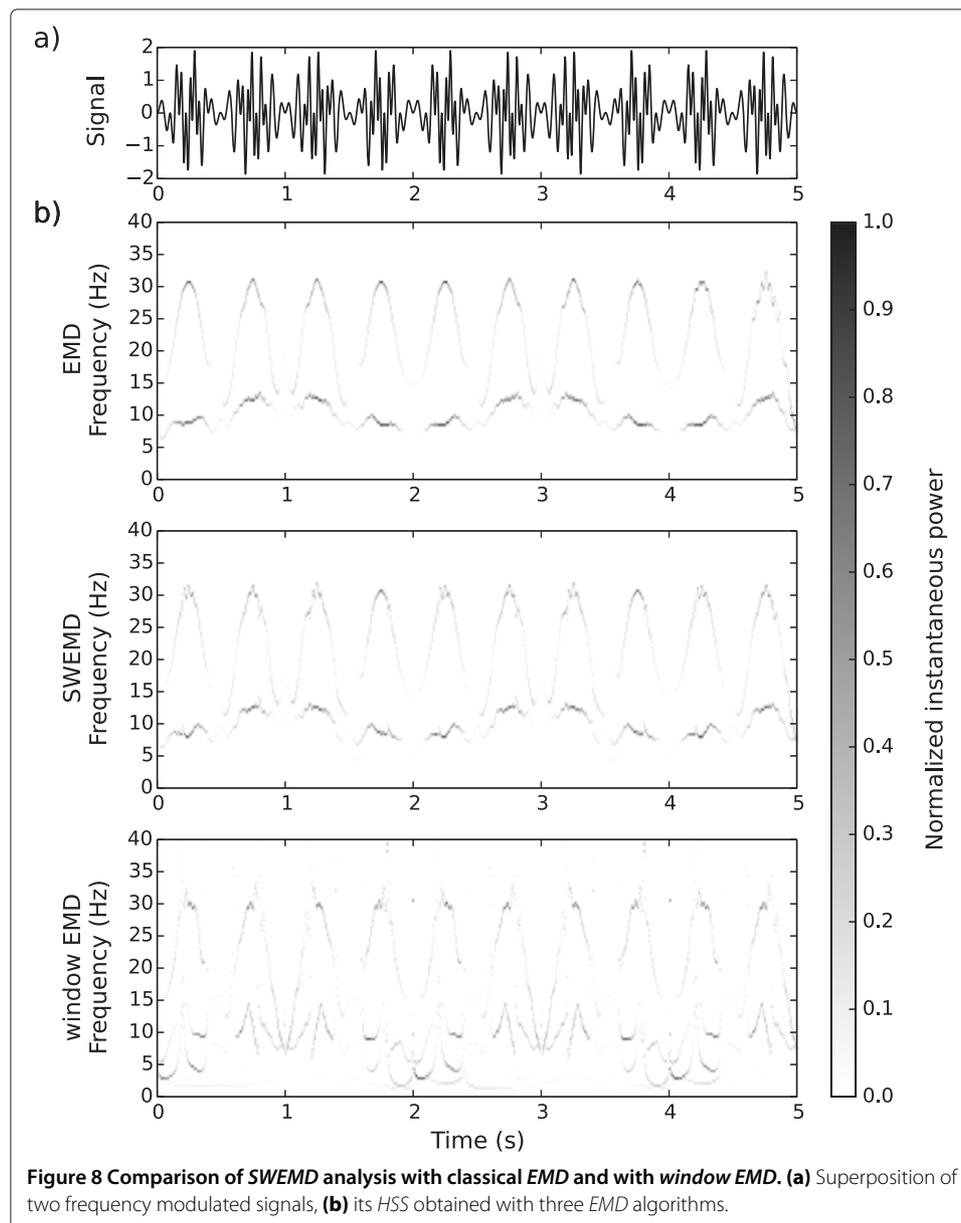
I also compared results of *SWEMD* analysis with classical *EMD* and with concatenated results of calculation of *EMD* in windows (*window EMD*) on simple signals.

Superposition of two chirp signals

Superposition of two chirps (signals with a linear frequency change) is shown in Figure 6a. *Hilbert-Huang spectra* calculated with three *EMD* algorithms: classical, *SWEMD*, and calculation of *EMD* in windows are compared in Figure 6b. Results of *SWEMD* are the same as *EMD*. Simple calculation in windows shows very strong boundary effects.

Frequency modulated signal

Frequency-modulated signal is shown in Figure 7a. The carrier frequency is 20 Hz, signal is modulated with 2 Hz sine. Figure 7b compares results of *Hilbert-Huang spectra*



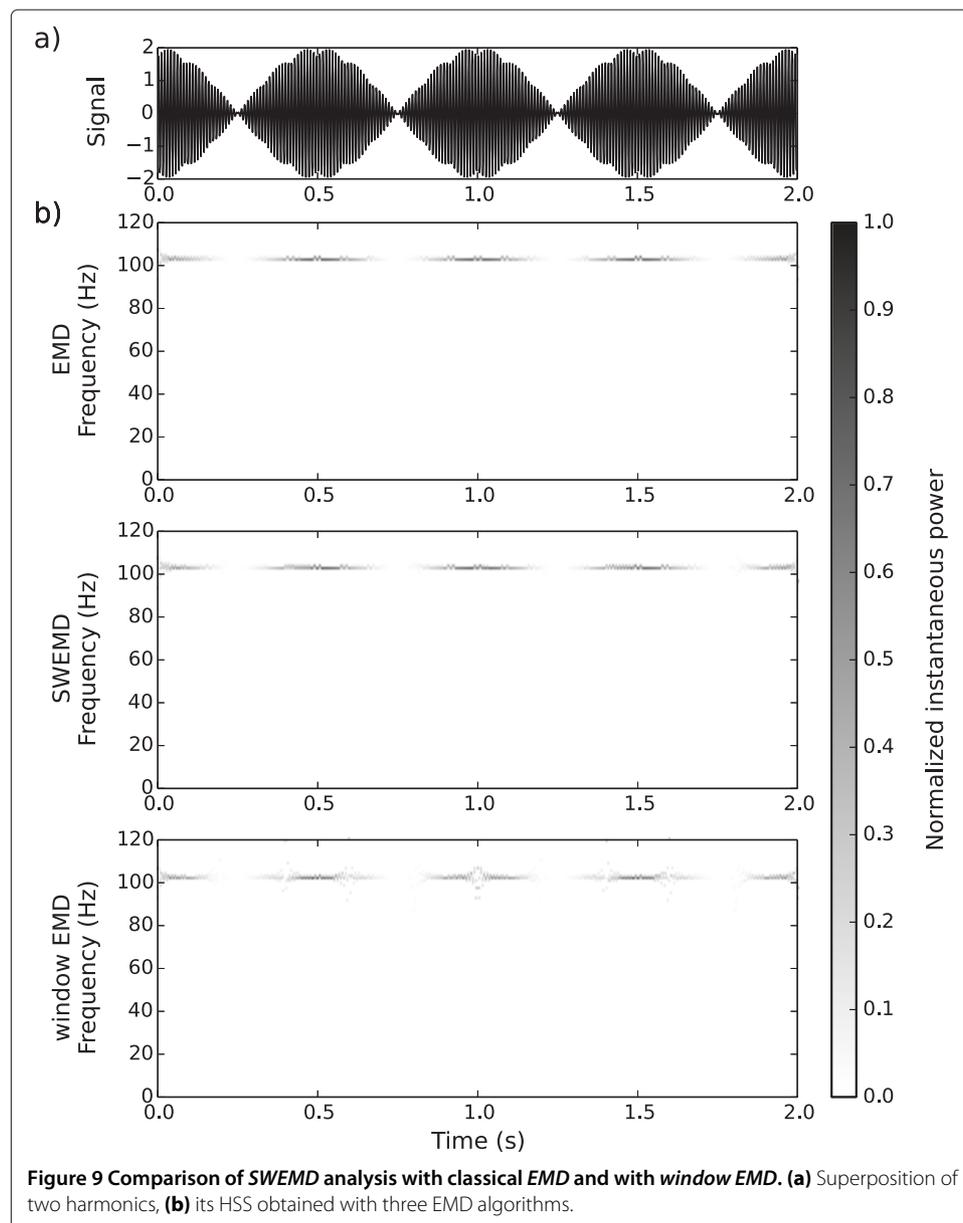
obtained using three different *EMD* algorithms. *SWEMD* results show very small window boundary effects compared to typical windowing of *EMD*.

Superposition of two FM signals

Figure 8a shows superposition of signal from Figure 7a with another *FM* signal (the carrier frequency is 10 Hz modulated with 0.5 Hz). Results in Figure 8b are similar for *SWEMD* and *EMD* but *window EMD* gives greater errors, especially in decomposition of signal with 10 Hz carrier frequency.

Superposition of two harmonics

Superposition of two sines 101 Hz and 103 Hz is shown in Figure 9a. Such superposition gives 102 Hz signal amplitude modulated with 1 Hz. Also for this signal *SWEMD*, unlike typical calculation in windows, gives the same results as 'classical' *EMD* (Figure 9b).



Conclusions

Before using the classical *EMD* one needs to consider whether multiple spline interpolation does not take too much time. In particular this applies to analysis of long signals or acquisition systems with high sampling frequency or a large number of channels. Proposed algorithm, *Sliding Window EMD*, can speed up computation about 10 times. Examples of using *SWEMD* presented in this article show that it has similar quality as classical *EMD* - boundary effects in typical windowing are nearly completely reduced in *SWEMD*. Proposed algorithm is suitable also in real time analysis of data.

To compare the performance of 'classical' *EMD* method and its modifications [2-5] including *SWEMD* it would be necessary to apply all these methods to the same set of signals.

Competing interests

The author declares that he has no competing interests.

Acknowledgements

This work was supported by Nalecz Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, Warsaw, through statutory project.

Received: 24 July 2014 Accepted: 11 September 2014

Published online: 22 November 2014

References

1. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen N-C, Tung CC, Liu HH: **The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis.** *Proc Roy Soc Lond Math Phys Sci* 1971, **454**:903–995.
2. Blakely CD: **A fast Empirical Mode Decomposition technique for nonstationary nonlinear time series.** In *Center for Scientific Computation and Mathematical Modeling*. College Park, USA: University of Maryland. Preprint CSCAMM-05-10; 2005. Online: [http://www.cscamm.umd.edu/publications/fastEMDcBlakelyFinal_CS-05-10.pdf]
3. Chen Q, Huang N, Riemenschneider S, Xu Y: **A B-spline approach for empirical mode decompositions.** *Adv Comput Math* 2006, **24**:171–195.
4. Meignen S, Perrier V: **A new formulation for empirical mode decomposition based on constrained optimization.** *IEEE Signal Process Lett* 2007, **14**(12):932–935.
5. Roy A, Doherty JF: **Raised cosine interpolation for empirical mode decomposition.** In *Information Sciences and Systems, 2009. CISS 2009. 43rd Annual Conference on*. Baltimore: IEEE; 2009.
6. Roy A, Doherty JF: **Improved signal analysis performance at low sampling rates using raised cosine empirical mode decomposition.** *Electron Lett* 2010, **46**(2):176–177.
7. Roy A, Doherty JF: **Raised cosine filter-based empirical mode decomposition.** *IET Signal Process* 2011, **5**(2):121–129.
8. Stepien P: **Application of empirical mode decomposition method in biomedical signal analysis (in polish).** *PhD thesis*. Warsaw University of Technology; 2011. Supervisor: W. Klonowski.
9. Rilling G, Flandrin P, Goncalves P: **On empirical mode decomposition and its algorithms.** In *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP-03*. Grado, Italy; 2003. Online: [http://perso.ens-lyon.fr/paulo.goncalves/pub/NSIP03_GRPFPG.pdf]
10. Pierzchalski MP, Stepien RA, Stepien P: **New nonlinear methods of heart rate variability analysis in diagnostics of atrial fibrillation.** *Int J Biol Biomed Eng* 2011, **5**(4):201–208.
11. Huang NE, Wu MLC, Long SR, Shen SSP, Qu W, Gloersen P, Fan KL: **A confidence limit for the empirical mode decomposition and Hilbert spectral analysis.** *Proc R Soc Lond Math Phys Sci* 2003, **459**(2037):2317.

doi:10.1140/epjnbp/s40366-014-0014-9

Cite this article as: Stepien: Sliding Window Empirical Mode Decomposition -its performance and quality. *EPJ Nonlinear Biomedical Physics* 2014 **2**:14.