We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Biomedical Time Series Processing and Analysis Methods: The Case of Empirical Mode Decomposition

Alexandros Karagiannis[1],
Philip Constantinou[1] and Demosthenes Vouyioukas[2]
*[1]National Technical University of Athens, School of Electrical and Computer Engineering,
Mobile RadioCommunication Laboratory
[2]University of the Aegean,
Department of Information and Communication Systems Engineering
Greece*

1. Introduction

1.1 Typical measurement systems chain

Computational processing and analysis of biomedical signals applied on the time series follow a chain of finite number of processes. Typical schemes front process is the acquisition of signal via the sensory subsystem. Next steps in the acquisition processing and analysis chain include buffers and preamplifiers, the filtering stage, the analog-digital conversion part, the removal of possible artifacts, the event detection and the analysis and feature extraction. Figure 1 depicts this process.
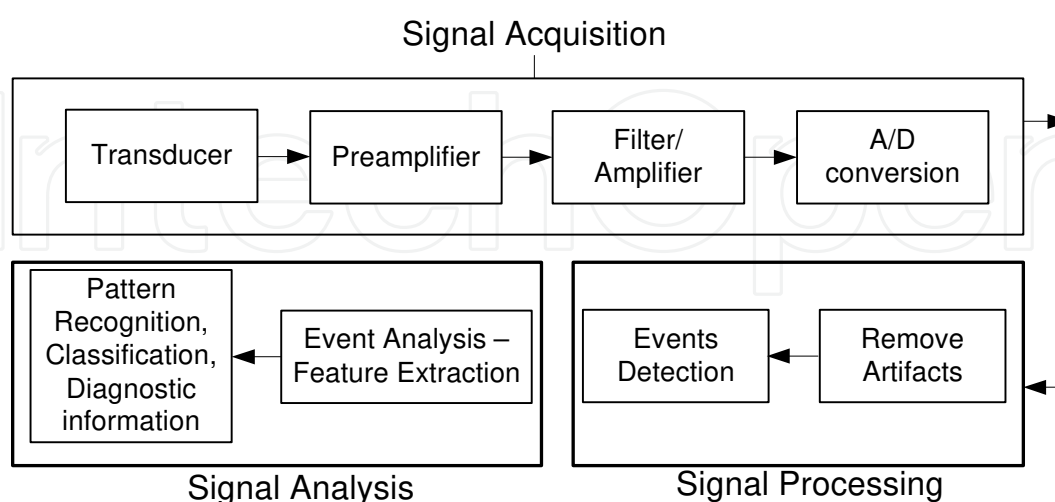

Fig. 1. Chain of processes from the acquisition of a biomedical signal to the analysis stage

Biomedical signal measurement, parameter identification and characterization initiate by the acquisition of diagnostic data in the form of image or time series that carry valuable

information related to underlying physical processes. The analog signal usually requires to be amplified and bandpass or lowpass filtered. Since most signal processing is easier to implement using digital methods, the analog signal is converted to digital format using an analog-to-digital converter. Once converted, the signal is often stored, or buffered, in memory. Digital signal processing algorithms applied on the digitized signal are mainly categorized as artifact removal processing methods and events detection methods. The last stage of this a typical measurement system refers to digital signal analysis with a higher level of sophistication techniques that extract features out of the digital signal or make a pattern recognition and classification in order to deliver useful diagnostic information.

A transducer is a device that converts energy from one form to another. In signal processing applications, the purpose of energy conversion is to gather information, not to transform energy. Usually, the output of a biomedical transducer is a voltage (or current) whose amplitude is proportional to the measured energy. The energy that is converted by the transducer may be generated by the physical process itself or produced by an external source. Many physiological processes produce energy that can be detected directly. For example, cardiac internal pressures are usually measured using a pressure transducer placed on the tip of catheter introduced into the appropriate chamber of the heart.

Whilst the most extensive signal processing is usually performed on digital data using software algorithms, some analog signal processing is usually necessary. Noise is inherent in most measurement systems and it is considered a limiting factor in the performance of a medical instrument. Many signal processing techniques target at the minimization of the variability in the measurement. In biomedical measurements, variability has four different origins: physiological variability; environmental noise or interference; transducer artifact; and electronic noise. The physiological variability is due to the fact that the biomedical signal acquired is affected by biological factors other than those of interest. Environmental noise originates from sources external or internal to the body. A classic example is the measurement of fetal ECG where the desired signal is corrupted by the mother's ECG. Since it is not known a priori the sources of environmental noise, typical noise reduction techniques have partially successful results compared to adaptive techniques which present better behavior in filtering.

Source	Cause
Physiological	Other variables present in the measured variable of interest
Environmental	Other sources of similar energy form
Electronic	Thermal or shot noise

Table 1. Sources of Measurement Variability

Transducer artifact is produced when the transducer responds to energy modalities other than that desired. For example, recordings of electrical potentials using electrodes placed on the skin are sensitive to motion artifact, where the electrodes respond to mechanical movement as well as the desired electrical signal. They are usually compensated by transducer design modifications.

Johnson or thermal noise is produced by resistance sources, and the amount of noise generated is related to the resistance and to the temperature:

$$V_{el} = \sqrt{4kTRB} \tag{1}$$

where R is the resistance in Ohms, T is the temperature in degrees Kelvin, k is Boltzman's constant (k = 1.38*10^{-23} J/^{o}K) and B is the bandwidth, or range of frequencies, that is allowed to pass through the measurement system.

It is a common assumption that electronic noise is spread evenly over the entire frequency range of interest. However it is common to describe relative noise as the noise that would occur if the bandwidth were 1.0 Hz. Such relative noise specification can be identified by the unusual units required: volts/$\sqrt{Hz}$ or amps/$\sqrt{Hz}$.

When multiple noise sources are present, as is often the case, their voltage or current contributions to the total noise add as the square root of the sum of the squares, assuming that the individual noise sources are independent. For voltages

$$V_T = (\sqrt{V_1^2 + V_2^2 + ... + V_N^2})^{1/2} \tag{2}$$

where V_1, V_2, ..., V_N are the voltages caused by any source of noise.

The relative amount of signal and noise present in the time series acquired by means of measurement systems is quantified by signal to noise ratio, SNR. Both signal and noise are measured in RMS values (root mean squared). SNR is expressed in dB (decidels) where

$$SNR = 20\log(\frac{Signal}{Noise}) \tag{3}$$

Various types of filters are incorporated according to the frequency range of interest in measurement systems. Lowpass filters allow low frequencies to pass with minimum attenuation whilst higher frequencies are attenuated. Conversely, highpass filters pass high frequencies, but attenuate low frequencies. Bandpass filters reject frequencies above and below a passband region. Bandstop filter passes frequencies on either side of a range of attenuated frequencies. The bandwidth of a filter is defined by the range of frequencies that are not attenuated.

The last analog element in a typical measurement system is the analog-to-digital converter (ADC). In the process of analog-to-digital conversion an analog or continuous waveform, x(t), is converted into a discrete waveform, x(n), a function of real numbers that are defined only at discrete integers, n. Slicing the signal into discrete points in time is termed time sampling or simply sampling. Time slicing samples the continuous waveform, x(t), at discrete prints in time, nTs, where Ts is the sample interval. Since the binary output of the ADC is a discrete integer whilst the analog signal has a continuous range of values, analog-to-digital conversion also requires the analog signal to be sliced into discrete levels, a process termed quantization.

The speed of analog to digital conversion is specified in terms of samples per second, or conversion time. For example, an ADC with a conversion time of 10 μsec should, logically, be able to operate at up to 100000 samples per second (or simply 100 kHz). Typical conversion rates run up to 500 kHz for moderate cost converters, but off-the-shelf converters can be obtained with rates up to several MHz. Lower conversion rates are usually acceptable for biological signals.

Most of biomedical signals are low energy signals and their acquisition takes place in the presence of noise and other signals originating from underlying systems that interfere with

the original one. Noise is characterized by certain statistical properties that facilitate the estimation of Signal to Noise ratio.

Biomedical data analysis aims at the determination of parameters required for models development of the underlying system and its validation. Problems usually encountered at the processing stage are related to the small length of sampled time series or the lack of stationarity and non linearity of the process that produces the signals.

1.2 Difficulties in acquisition and biomedical signal analysis

The proximity of the sensory subsystem to the physical phenomenon, biomedical signal's dynamic nature as well as the interconnections and interactions of multiple physical systems are set difficulties in acquisition and biomedical signal processing and analysis. The impact of measurement equipment and different sources of artifacts and noise in biomedical signals such as electrocardiogram are considered in the determination of properties that affect the processing stage.

1.3 Sensor proximity

Most of physiological systems are located deep inside the human body and this sets a difficulty in biosignal acquisition and measurement. A typical case is electrocardiogram which is acquired by means of electrodes in the level of chest. The measured signal is a projection of a moving 3D cardiac electric vector at a level defined by the electrodes. If the purpose of the electrocardiogram acquisition is related to the monitoring of cardiac rhythm then this signal provides sufficient information. However, if the purpose is the atrium electric activity monitoring the processing and analysis of this signal is difficult.

Proximity to the physiological system that produces biosignals is usually accomplished by means of invasive methods which require certain conditions for the patients and the available equipment.

1.4 Signal variability

Physiological systems are dynamic systems controlled by numerous variables. Biomedical signals represent the dynamic nature of the underlying physiological systems. These processes as well as the variables have a deterministic or random (stochastic) nature and in some cases they are periodic.

A normal electrocardiogram may present a normal cardiac rhythm with easily identifiable and detectable complexes. A normal electrocardiogram could be characterized as deterministic and periodic signal; however a patient's circulatory system may have significant time variability both in the form of the complexes and the cardiac rhythm.

The dynamic nature of biological systems results in the stochastic and non stationary nature of biomedical signals. Statistical parameters such as average value and variance as well as the spectral density are time variant. In this case, a common approach is the signal analysis in wide time windows in order to include all the possible conditions of the underlying biological systems.

1.5 Interconnections and Interactions between physiological systems

Various physiological systems of the human body are not independent; on the contrary they are interconnected and interact. Some of the interactions cause physiological variable compensation, feedback loops or even affect other physiological systems. These operations

in the level of physiological systems interactions should be considered as well in monitoring, processing and biomedical signal analysis.

1.6 Measurement equipment and measurement procedures

The front end of a measurement system which is the transducer subsystem and the connection with the rest of the measurement equipment affects the performance of the measurement system and may cause significant changes in signal's characteristics.

1.7 Artifacts and interference

When electrocardiogram is acquired it is required the immobility of the body in order to minimize the interference from other signals such as electromyogram. Even the respiratory signal can cause interference to the electrocardiogram.

Artifacts in acquired biomedical signal and interference from other physiological systems raise the need for biomedical signal processing techniques in order to deal with these phenomena.

1.8 Measurement equipment sensitivity

Monitoring of biomedical signals in the range of a few microvolts or millivolits which are produced by physiological systems demands the use of equipment with increased levels of sensitivity as well as low levels of noise. Shielded cables are used in order to minimize the electromagnetic interference from other medical equipment or any other sources of electromagnetic fields.

2. Spectral and statistical properties of biomedical signals

In scientific study, noise can come in many ways: it could be part of the natural processes generated by local and intermittent instabilities and sub-grid phenomena; it could be part of the concurrent phenomena in the environment where the investigations were conducted; and it could also be part of the sensors and recording systems. A generic model for the acquired signal is described by formula 4:

$$x(t) = s(t) + n(t) \tag{4}$$

where x(t) represents the acquired data, s(t) is the true signal and n(t) is noise. Once noise contaminates data, data processing techniques are employed to remove it.

For the obvious cases, when the processes are linear and noise have distinct time or frequency scales different from those of the true signal, Fourier filters can be employed to separate noise from the signal. Historically, Fourier based techniques are the most widely used.

The problem of separating noise and signal is complicated and difficult when there is no knowledge of the noise level in the data. Knowing the characteristics of the noise is an essential first step.

Most of the biosignals are characterized by the small levels of their energy as well as the existence of various types of noise during the acquisition. Any signal of no interest rather than the true signal is characterized as artifact, interference or noise. The existence of noise deteriorates the performance of a measurement system and the processing and analysis stages.

The amplitude of a deterministic signal can be calculated by a closed form mathematical formula or predicted if the amplitude of previous samples is considered. All the other signals are characterized as random signals. Kendal and Challis [1] , [2] proposed a test for the determination of the randomness of a signal which is based on the number of signal's extremas.

2.1 Noise

The term random noise refers to the interference of a biosignal caused by a random process. Considering a random variable η with probability density function $p_\eta(\eta)$, the average value μ_η of the random process η is defined as

$$\mu_\eta = E[\eta] = \int_{-\infty}^{+\infty} \eta p_\eta(\eta) d\eta \tag{5}$$

where E[.] is the expected value of random variable η.
Mean square value of random process is defined as

$$E[\eta^2] = \int_{-\infty}^{+\infty} \eta^2 p_\eta(\eta) d\eta \tag{6}$$

and the variance of the process is defined as

$$\sigma_\eta^2 = E[(\eta - \mu_\eta)^2] = \int_{-\infty}^{+\infty} (\eta - \mu_\eta)^2 p_\eta(\eta) d\eta \tag{7}$$

The square root of the variance provides the standard deviation σ_η of the process.

$$\sigma_\eta^2 = E[\eta^2] - \mu_\eta^2 \tag{8}$$

The average value of a stochastic process η(t) represents the DC component of the signal; the mean square value represents the mean energy of the signal and the mean square root of the variance represents the RMS value. These statistical parameters are the essential components in the SNR estimation.

2.2 Ensemble averages

When the probability density function of a random process is not known then it is common practice to estimate the statistical expected value of the process via the averages computed at sample sets of the process.
The estimation of average defined in t_1 is

$$\mu_x(t) = \lim_{M \to \infty} \frac{1}{M} \sum_{k=1}^{M} x_k(t_1) \tag{9}$$

The autocorrelation function $\varphi_{xx}(t_1, t_1+\tau)$ of a random process is defined

$$\varphi_{xx}(t_1, t_1 + \tau) = E[x(t_1)x(t_1 + \tau)] = \int_{-\infty}^{+\infty} x(t_1)x(t_1 + \tau) p_x(x) dx \tag{10}$$

2.3 Non stationary biomedical time series

Biomedical data analysis aims at the determination of parameters which are required for the development of models for the underlying physiological processes and the validation of those models. The problems encountered in the analysis of biomedical time series are due to the total data length, the non stationarity of the time series and the non linearity of the underlying physiological processes. The first two problems are related. Biomedical time series which are short in terms of time duration could be shorter than the longer time scale of a stationary process and to be characterized in this way as a non stationary process.

Fourier spectral analysis is a general method for the energy distribution of signal's frequency components. It has dominated in the data analysis and has been applied in almost all the biomedical time series acquired. However Fourier transform is applicable under certain conditions that set limitations. Linearity and strict periodicity as well as the strict stationary process are some of the conditions that should be satisfied in order to apply Fourier transform and interpret in a correct way the physical meaning of the results.

The stationarity requirement is not particular to the Fourier spectral analysis; it is a general one for most of the available data analysis methods. According to the traditional definition, a time series, x(t), is stationary in the wide sense, if, for all t

$$E[\,|\,x(t)^2\,|\,] < \infty$$
$$E[x(t)] = \mu_x \qquad\qquad\qquad (11)$$
$$Cov(x(t_1), x(t_2)) = Cov(x(t_1 + \tau), x(t_2 + \tau)) = Cov(t_1 - t_2)$$

in which E(.) is the expected value defined as the ensemble average of the quantity, and C(.) is the covariance function. Stationarity in the wide sense is also known as weak stationarity, covariance stationarity or second-order stationarity.

Few of the biomedical data sets, from either natural phenomena or artificial sources, can satisfy the definition of stationarity. Other than stationarity, Fourier spectral analysis also requires linearity. Although many natural phenomena can be approximated by linear systems, they also have the tendency to be nonlinear. For the above reasons, the available data are usually of finite duration, non-stationary and from systems that are frequently nonlinear, either intrinsically or through interactions with the imperfect probes or numerical schemes. Under these conditions, Fourier spectral analysis is of limited use [3]. The uncritical use of Fourier spectral analysis and the adoption of the stationary and linear assumptions may give misleading results.

3. Biomedical signal processing and analysis methods

Many waveforms—particularly those of biological origin–are not stationary, and change substantially in their properties over time. For example, the EEG signal changes considerably depending on various internal states of the subject. A wide range of approaches have been developed in order to extract both time and frequency information from a waveform. Basically they can be divided into two groups: time–frequency methods and time–scale methods. The latter are better known as Wavelet analysis.

3.1 The spectrogram

The first time–frequency methods were based on the straightforward approach of slicing the waveform of interest into a number of short segments and performing the analysis on each

of these segments, usually using the standard Fourier transform [4]. A window function is applied to a segment of data, effectively isolating that segment from the overall waveform, and the Fourier transform is applied to that segment. This is termed the spectrogram or "short-term Fourier transform" (STFT).

Since it relies on the traditional Fourier spectral analysis, one has to assume the data to be piecewise stationary. This assumption is not always justified in non-stationary data. Furthermore, there are also practical difficulties in applying the method: in order to localize an event in time, the window width must be narrow, but, on the other hand, the frequency resolution requires longer time series.

3.2 Wigner-Ville distribution

A number of approaches have been developed to overcome some of the shortcomings of the spectrogram. The first of these was the Wigner-Ville distribution. It is a special case of a wide variety of similar transformations known under the heading of Cohen's class of distributions.

The Wigner-Ville, and in fact all of Cohen's class of distributions, use a variation of the autocorrelation function where time remains in the result. This is achieved by comparing the waveform with itself for all possible lags, but instead of integrating over time.

The Wigner-Ville distribution is sometimes also referred to as the Heisenberg wavelet. By definition, it is the Fourier transform of the central covariance function. For any time series, x(t), we can define the central variance as

$$C_c(\tau, t) = x(t - \frac{\tau}{2})x^*(t + \frac{\tau}{2})$$ (12)

Then the Wigner-Ville distribution is

$$V(\omega, t) = \int_{-\infty}^{+\infty} C_c(\tau, t)e^{-i\omega\tau} d\tau$$ (13)

The classic method of computing the power spectrum was to take the Fourier transform of the standard autocorrelation function. The Wigner-Ville distribution echoes this approach by taking the Fourier transform of the instantaneous autocorrelation function, but only along the τ (i.e., lag) dimension. The result is a function of both frequency and time.

3.3 Evolutionary spectrum

The evolutionary spectrum was proposed by Priestley [5]. The basic idea is to extend the classic Fourier spectral analysis to a more generalized basis: from sine or cosine to a family of orthogonal functions $\varphi(\omega, t)$ indexed by time, t, and defined for all real ω, the frequency. Any real random variable, x(t), can be expressed as

$$x(t) = \int_{-\infty}^{+\infty} \varphi(\omega, t) dA(\omega, t)$$ (14)

in which dA(ω,t), the Stieltjes function for the amplitude, is related to the spectrum as

$$E(|dA(\omega, t)|^2) = d\mu(\omega, t) = S(\omega, t)d\omega$$ (15)

where $\mu(\omega,t)$ is the spectrum, and $S(\omega,t)$ is the spectral density at a specific time t, also designated as the evolutionary spectrum.

4. Empirical mode decomposition

A recently proposed method, the Hilbert-Huang Transform (HHT) [3], satisfies the condition of adaptation employed in nonlinear - nonstationary time series processing. HHT consists of EMD and Hilbert Spectral Analysis (HSA) [6]. The lack of mathematical foundation and analytical expressions sets difficulty in the theoretical study of the method. Nevertheless there has been an exhaustive validation in an empirical fashion especially in the time-frequency representations [7].

Empirical Mode Decomposition (EMD) lies in the core of HHT method decomposing nonstationary time series originating from nonlinear systems in an adaptive fashion without predefined basis function. An intrinsic mode function (IMF) set is produced through an iterative process which is related to the underlying physical process.

Unlike wavelet processing, Hilbert-Huang transform decomposes a signal by direct extraction of the local energy associated with the time scales of the signal. This feature reveals the applicability of HHT in both nonstationary time series and signals originating from nonlinear biological systems.

Literature references' variety reveals the extensive range of EMD applications in several areas of the biomedical engineering field. Particularly there are publications concerning the application of EMD in the study of Heart Rate Variability (HRV) [8], analysis of respiratory mechanomyographic signals [9], ECG enhancement artifact and baseline wander correction [10], R-peak detection [11], Crackle sound analysis in lung sounds [12] and enhancement of cardiotocograph signals [13]. The method is employed for filtering electromyographic (EMG) signals in order to perform attenuation of the incorporated background activity [14]. Numerous research papers have been published concerning applications of EMD in biomedical signals and especially towards the direction of optimizing traditional techniques of acquisition and processing of signals such as Doppler ultrasound for the removal of artifacts [15], the analysis of complex time series such as human heartbeat interval [16], the identification of noise components in ECG time series [17] and the denoising of respiratory signals [18].

Lack of solid theoretical foundation concerning empirical mode decomposition constitutes the basis for a series of problems regarding the adaptive nature of the method as well as the selection of an efficient interpolation technique. Identification of nonlinear characteristics of the physical process and optimum threshold selection for the implementation of the algorithm set challenges for further research on EMD method.

The empirical mode decomposition does not require any known basis function and is considered a fully data driven mechanism suited for nonlinear processes and nonstationary signals.

Each component extracted (IMF) is defined as a function with

- Equal number of extrema and zero crossings (or at most differed by one)
- The envelopes (defined by all the local maxima and minima) are symmetric with respect to zero. This implies that the mean value of each IMF is zero.

Given a signal x(t), the algorithm of the EMD can be summarized as follows :

1. Locate local maxima and minima of $d_0(t)=x(t)$.

2. Interpolate between the maxima and connect them by a cubic spline curve. The same applies for the minima in order to obtain the upper and lower envelopes $e_u(t)$ and $e_l(t)$, respectively.
3. Compute the mean of the envelopes:

$$m(t) = \frac{e_u(t) + e_l(t)}{2} \tag{16}$$

4. Extract the detail $d_1(t) = d_0(t) - m(t)$ (sifting process)
5. Iterate steps 1-4 on the residual until the detail signal $d_k(t)$ can be considered an IMF (satisfy the two conditions): $c_1(t) = d_k(t)$
6. Iterate steps 1-5 on the residual $r_n(t) = x(t) - c_n(t)$ in order to obtain all the IMFs $c_1(t), .., c_N(t)$ of the signal. The result of the EMD process produces N IMFs ($c_1(t), c_2(t), ... c_N(t)$) and a residual signal ($r_N(t)$) :

$$x(t) = \sum_{n=1}^{N} c_n(t) + r_N(t) \tag{17}$$

In step 5, in order to terminate the sifting process it is commonly used a criterion which is the sum of difference

$$SD = \sum_{t=0}^{T} \frac{|d_{k-1}(t) - d_k(t)|^2}{d_{k-1}^2(t)} \tag{18}$$

When SD is smaller than a threshold, the first IMF is obtained and this procedure iterates till all the IMFs are obtained. In this case, the residual is either a constant, or a monotonic slope or a function with only one extremum.

Implementation of the aforementioned sifting process termination criterion along with the conditions that should be satisfied in order to acquire an IMF result in a set of check points in the algorithm (Eq. 19, 20, 21).

$$\sum \left\{ \frac{MA}{EA} > Threshold_1 \right\}_{boolean} < TOLERANCE \tag{19}$$

$$\frac{MA}{EA} < Threshold_2 \tag{20}$$

$$\left| \sum zeros - \sum extrema \right| <= 1 \tag{21}$$

where MA is the absolute value of m(t) and EA is given by the equation 22.

$$EA = \left| \frac{e_u(t) - e_l(t)}{2} \right| \tag{22}$$

Control of the progress of the algorithm and the IMF extraction process is determined by equations 19-22 and termination as well as the number of IMFs are related to the selection of threshold values. Different values result in different set of IMFs and significant computation

effect on the whole process of the EMD algorithm especially at the level of the number of iterations required. Optimum threshold values are still under investigation in the research field concerning the method as well as in the effect on the set of IMFs and the relation of certain IMFs with the underlying physical process [19].

Each component extracted (IMF) is defined as a function with equal number of extrema and zero crossings (or at most differed by one) with its envelopes (defined by all the local maxima and minima) being symmetric with respect to zero.

The application of the EMD method results in the production of N IMFs and a residue signal. The first IMFs extracted are the lower order IMFs which captures the fast oscillation modes while the last IMFs produced are the higher order IMFs which represent the slow oscillation modes. The residue reveals the general trend of the time series.
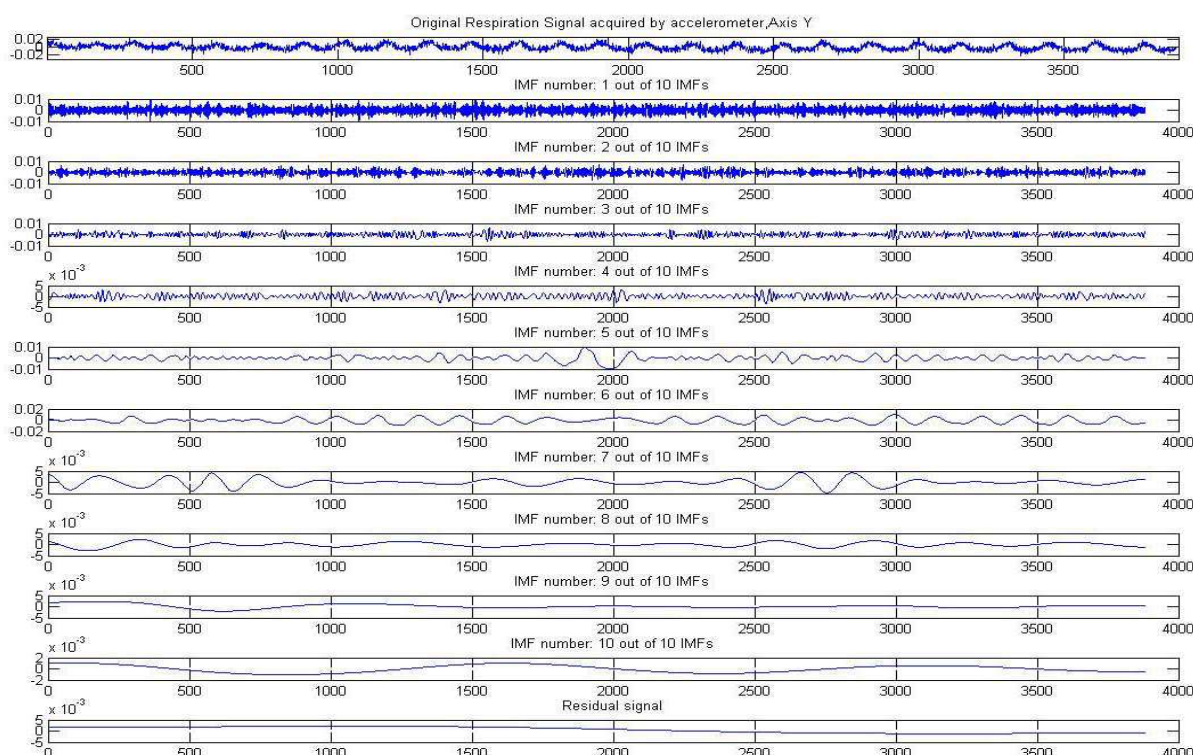
Fig. 2. Experimental respiratory signal processed with Empirical Mode Decomposition.
At the upper plot is depicted the original signal Axis Y of a dual axis accelerometer
which is sampled in both axes by a mote of a Wireless Sensor Network [18].

5. Statistical significance of IMFs

Intuitively, a subset of IMF set produced after the application of EMD on biomedical ECG time series is related to the signal originating from the physical process. Although high correlation values between the noise corrupted time series with specific IMFs may occur, there is a difficulty in defining a physical meaning and identifying those IMFs that carry information related to the underlying process.

The lack of EMD mathematical formulation and theoretical basis complicates the process of selecting the IMFs that may confidently be separated from the ones that are mainly attributed to noise. Flandrin et al [21] studied fractional Gaussian noise and suggested that

EMD acts as a dyadic filter. Wu and Huang [20] confirmed Flandrin's findings by studying White Gaussian Noise in time series processed with EMD. Wu and Huang empirically discovered a linear relationship between mean period and time series energy density expressed in log-log scale.

Study of noise statistical characteristics initiates the computation of the IMF's energy distribution function. The establishment of energy distribution spread function for various percentiles according to literature conclusions mentioned in this section constitutes an indirect way to quantify IMFs with strong noise components thus defines their statistical significance.

Each IMF probability function is approximately normally distributed, which is expected from the central limit theorem. This finding implies that energy density of IMFs should have a chi-square distribution (x^2).

Determination of the IMF mean period is accomplished by counting the number of extrema (local maxima-minima) or the number of zero crossings. The application results on typical 6000 samples MIT-BIH record 100 [23] for both unfiltered and Savitzky-Golay filtered time series are summarized in tables 2 and 3 respectively. Mean period is expressed in time units (sec) by taking into consideration the number of local maxima and the frequency sampling of the time series [22].

Energy Density of the nth IMF is calculated by mathematical expression 23.

$$E_n = \frac{1}{N} \sum_{j=1}^{N} [c_n(j)]^2 \tag{23}$$

Energy distribution and spread function constitute the basis for the development of a test in order to determine the IMFs statistical significance. The algorithm implemented is described below assuming that biomedical ECG time series are corrupted by White Gaussian Noise:

1. Decompose the noisy time series into IMFs via EMD.
2. Utilize the statistical characteristics of White Gaussian Noise in the time series to calculate energy spread function of various percentiles.
3. Select the confidence interval (95%, 99%) to determine upper and lower spread lines.
4. Compare the energy density of the IMFs with the spread function.

IMF energies that lie outside the area defined by the spread lines, determine the statistical significance of each one. The application results are depicted in figure 3 for a MIT-BIH ECG record 100 time series of 6000 samples length processed with Savitzky-Golay method. As far as step 2 of the algorithm concerns, a detailed approach is described in [20] with analytical formula expression for the determination of spread lines at various percentiles.

Statistical significance test indicates a way to separate information from noise in noise corrupted time series. Nevertheless, partial time series reconstruction by proper selection of the IMFs outside the spread lines area reveals that noisy components still exist in reconstructed time series. The interpretation of an IMF subset physical meaning by means of instantaneous frequencies, a typical characteristic of IMFs revealed when treated with Hilbert Transform, is based on the assumption that instantaneous frequencies related to the underlying process are spread in the whole IMF set. Combining this observation with the addition of white Gaussian noise and the application of the algorithm that takes into advantage the statistical characteristics of WGN one draws the conclusion that the algorithm proposed is lossy in terms of physical meaning in the reconstructed time series. A loss of information related to the underlying process is caused due to exclusion of an IMF subset.

This observation reveals a trade off situation in the level of partial signal reconstruction between the amount of information related to the physical process in the reconstructed time series and the noise level. Inclusion of wider IMF subset in the reconstruction process also increases noise levels and deteriorates SNR in the reconstructed time series.

Reconstruction process results of the proposed algorithm are presented in [17] for a MIT-BIH ECG record time series of 6000 samples length which is EMD processed and the algorithm of IMFs statistical significance is applied. Cross correlation value of 0.7 is achieved only by including the statistically significant IMFs.

IMF	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
# extrema	1764	943	691	537	430	351	265	211	212	112	85	44	22	8	5	1
Mean Period (sec)	0.003	0.006	0.009	0.011	0.014	0.017	0.023	0.028	0.028	0.054	0.071	0.136	0.273	0.750	1.200	6.001

Table 2. IMFs mean period of 6000 samples unfiltered MIT-BIH ECG record 100

IMF	1	2	3	4	5	6	7	8	9	10	11	12	13
# extrema	1123	735	456	349	283	245	125	94	64	38	21	7	1
Mean Period (sec)	0.005	0.008	0.013	0.017	0.021	0.025	0.048	0.064	0.094	0.158	0.286	0.857	6.001

Table 3. IMFs mean period of 6000 samples Savitzky-Golay filtered MIT-BIH ECG record 100
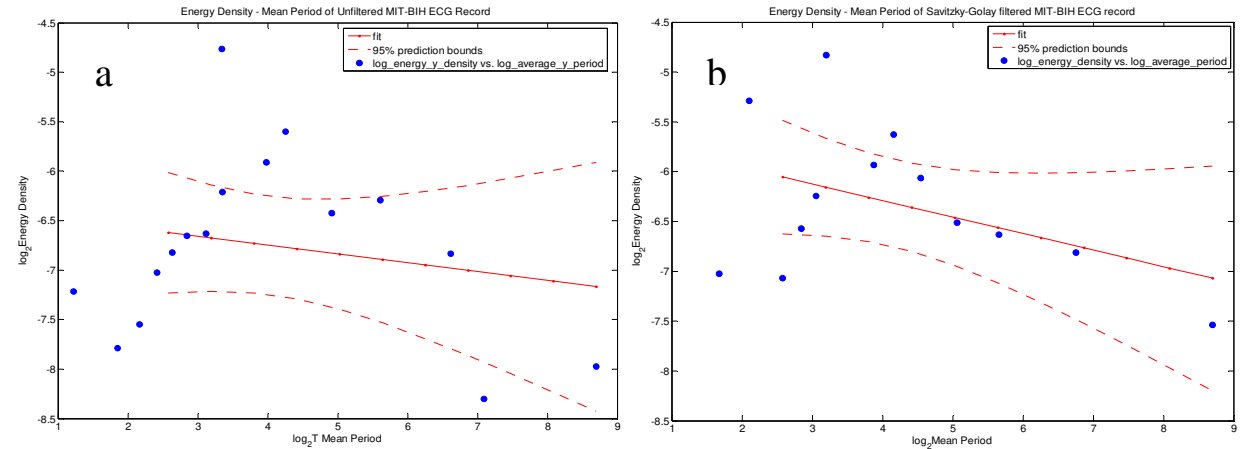

Fig. 3. IMF Energy Density of MIT-BIH ECG record 100 of 6000 samples as a function of the Mean Period. Fitting of the experimental results exhibits a linear relationship for log-log scale of IMF's Energy Density and Mean Period at 95% confidence interval.

6. Noise assisted data processing with empirical mode decomposition

Time series are considered to be IMFs if they satisfy two conditions concerning the number of zero crossings and extrema (equal or at most differ by one) and the required symmetry of the envelopes with respect to zero. A complete description of the EMD algorithm is included in [3].

The majority of data analysis techniques aim at the removal of noise in order to facilitate the following stages of the processing-analysis chain. However, in certain cases, noise is added to the time series to assist the detection of weak signals and delineate the underlying process. A common technique in the category of Noise Assisted Data Analysis (NADA) methods is pre-whitening. Adding noise to time series is an assistive way for the investigation of analysis method sensitivity. Furthermore the superimposition of noise samples following specific distribution functions in time series facilitates the study of EMD performance in processing of typical noise corrupted biomedical signals.

In the framework of NADA applications on biomedical signals, the addition of White Gaussian Noise (WGN) boosts the tendency of time series to develop extrema. EMD sensitivity in extrema detection is related to the interpolation technique. In the current implementation, cubic spline curve is selected as the interpolation technique; still there are multiple arguments in literature for different interpolation schemes.

The proposed methodology is depicted in figure 4. Simulated biomedical signals, in this case electrocardiogram (ECG), are contaminated with WGN in a controlled way. The study of EMD performance is accomplished by comparative evaluation of the method results in respect of three aspects. First, EMD performance is studied by investigating the statistical significance of an IMF set. Secondly, computation time of the method's application on biomedical signals is measured in both possible routes depicted in methodology diagram and thirdly the size of the IMF set is monitored.

The preprocessing stage is carefully selected after an exhaustive literature review and represents three different filtering techniques in order to tackle with various artifacts present in ECG time series. Namely, it constitutes a preparative stage, which changes the spectral characteristics of the time series in a predefined way.

Mainly there are two modes of operation in electrocardiography, the monitor mode and diagnostic mode. Highpass and lowpass filters are incorporated in monitor mode with cutoff frequencies in the range of 0.5-1Hz and 40Hz respectively. The selection of the aforementioned cutoff frequencies is justified by the accomplishment of artifact limitation in routine cardiac rhythm monitoring (Baseline Wander reduction, power line suppression). In diagnostic mode, lowpass filter cutoff frequency range is wider from 40Hz to 150Hz whereas for highpass filter cutoff frequency is usually set at 0.05Hz (for accurate ST segment recording).

Apparently noise assisted data analysis methods coexistence with noise reduction techniques set two antagonistic factors. The target for the addition of white Gaussian noise is threefold. It simulates a typical real world biomedical signal case whereas the superimposition of noise samples increase the number of extrema developed in the time series in order to evaluate EMD application results due to the high sensitivity of the method in extrema detection. Finally, the study of the IMF set statistical significance is facilitated taking under consideration the noise samples distribution function as well as the statistical properties of the noisy time series.

Preprocessing stage implemented as various filtering techniques is commonly incorporated in typical biomedical signal processing chain. Apart from the trivial case of taking into account these techniques to process ECG time series, preprocessing stage is introduced prior to the application of EMD method in order to comparatively evaluate the performance of the mixed scheme in terms of size of IMF set and its statistical significance as well as the total computation time. Each technique deals with specific types of artifacts in ECG time series and a significant part of initial noise level is still present in the time series processed via EMD.

The flowchart of the proposed methodology is applied on both simulated and real record ECG time series and the branch outputs are compared in order to evaluate the pre-processing stage and the effect in EMD performance.
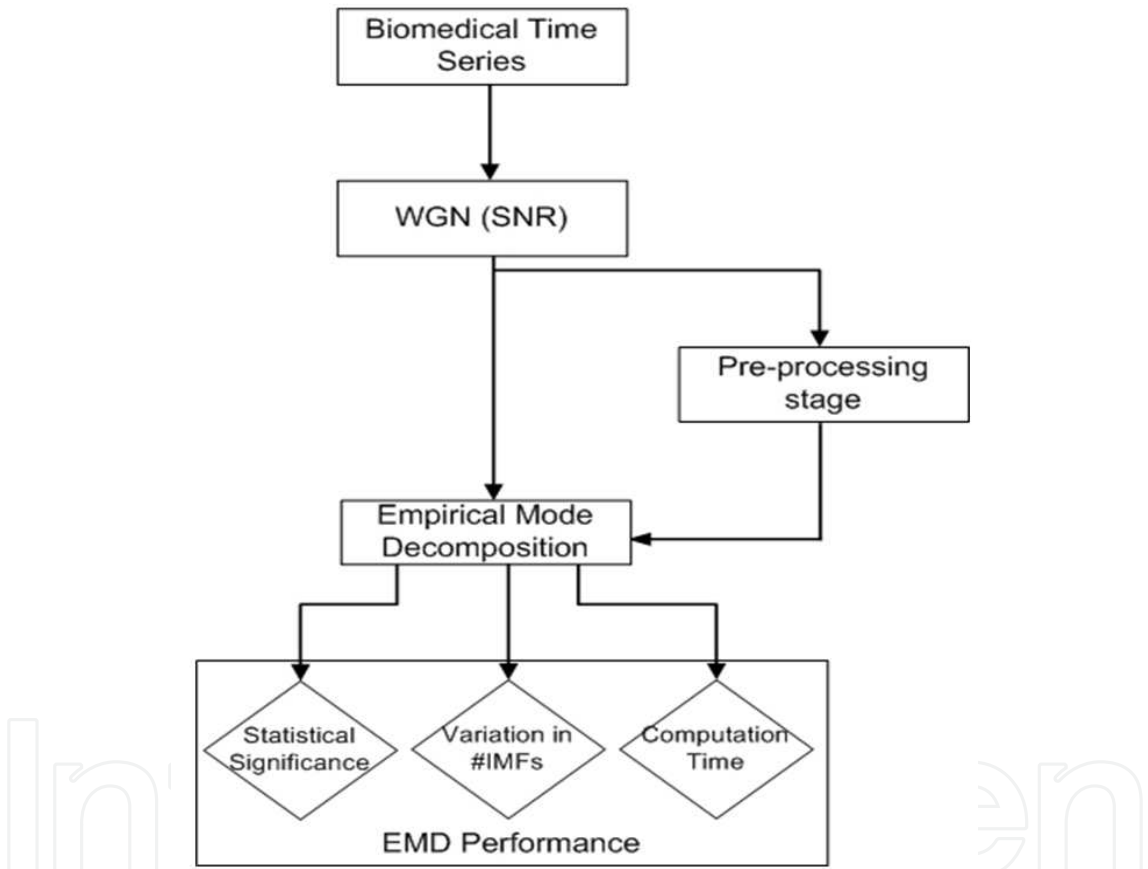
Fig. 4. Methodology process for the performance study of EMD applied on ECG time series

Results of the proposed methodology are provided in [22] and [17] with more details concerning the pre-processing stage which is implemented as typical filters and the way this stage affects the output of the empirical mode decomposition application on the simulated and real biomedical time series. Empirical mode decomposition performance is checked in terms of statistical significance of the IMF set produced, the variation of the IMF set length as a function of time series length and SNR and the computation time.

Some results are included in this chapter and depicted in figure 5.
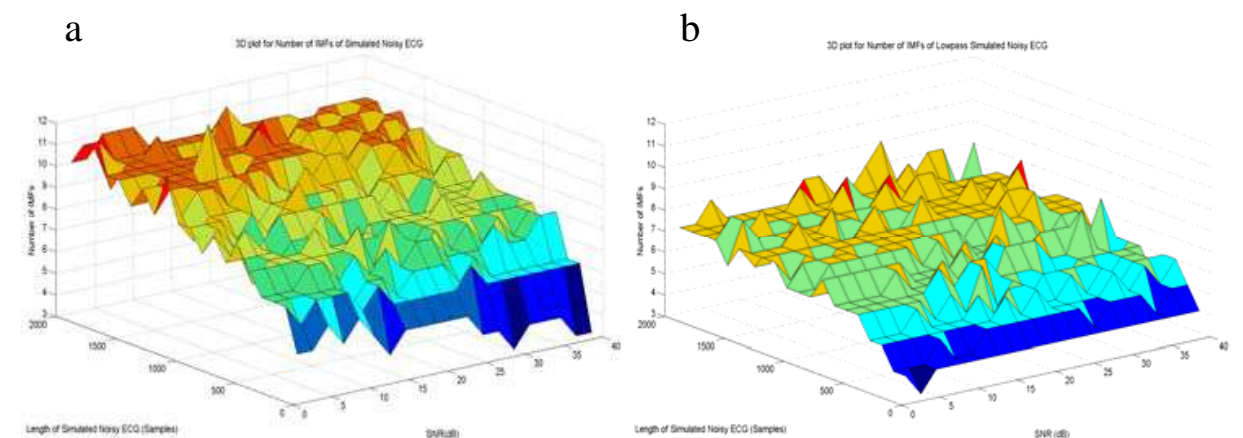
a b

Fig. 5. a. 3D plots of the number of IMFs as a function of the SNR and the length of
a simulated White Gaussian Noise corrupted ECG time series without the application
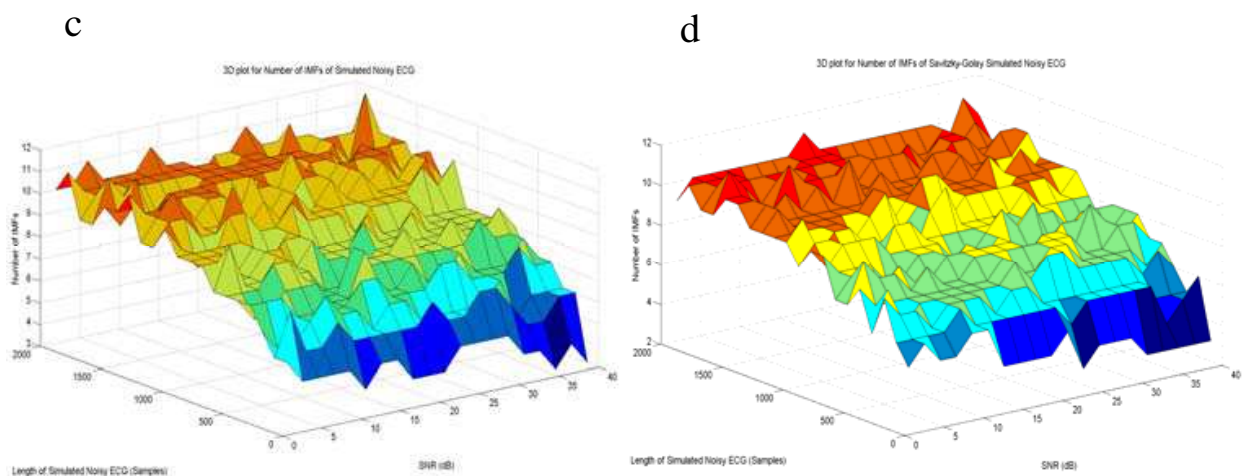of preprocessing stage (a) and with application of a lowpass filter (b),
See [22].

c d

Fig. 5. b. 3D plots of the number of IMFs as a function of the SNR and the length of
a simulated White Gaussian Noise corrupted ECG time series without the application
of preprocessing stage (c) and with application of the Savitzky-Golay filter (d),
See [22].

Savitzky-Golay method is considered mainly for its wide acceptance in ECG processing and
especially for the ability of the filter to preserve the peaks with minimal distortion. Minor
effects are expected on the peaky nature of the noise corrupted ECG time series. As a result,
the variation in the number of extracted IMFs after the application of EMD on Savitzky-
Golay filtered ECG time series is relatively small.
The effect on the peaky nature of time series processed with lowpass filters results in the
reduction of the IMF set size. Various cut-off frequencies attenuate in a different way high
frequency content. Number of extrema is decreased in the lowpass filtered time series

however distribution of peaks in the time series is dependent on the frequency components distorted by the different cut-off frequencies.

7. Computation time considerations for empirical mode decomposition

Considering the characteristics of EMD algorithm a straight forward way for computation time estimation takes into account the size of IMF set as well as the number of iterations required in order to produce this set. This goes down to implementation issues concerning the EMD algorithm and the thresholds used in termination criterion as well as the maximum number of iterations allowed.

Multiple lengths of noise corrupted simulated ECG time series of various SNR levels are studied. For demonstration reasons the minimum and maximum number of samples (1000, 8000) are depicted in figure 6 along with the computation time of unfiltered EMD processed time series.

Computation time of EMD processed ECG time series is depicted in figure 6 for comparison reasons. In both graphs EMD performance in terms of computation time is worst compared to the corresponding performance of ECG time series preprocessed with the suitable filter. Overall, EMD performance of LP_1 highlights the important role of suitable preprocessing stage selection [22].
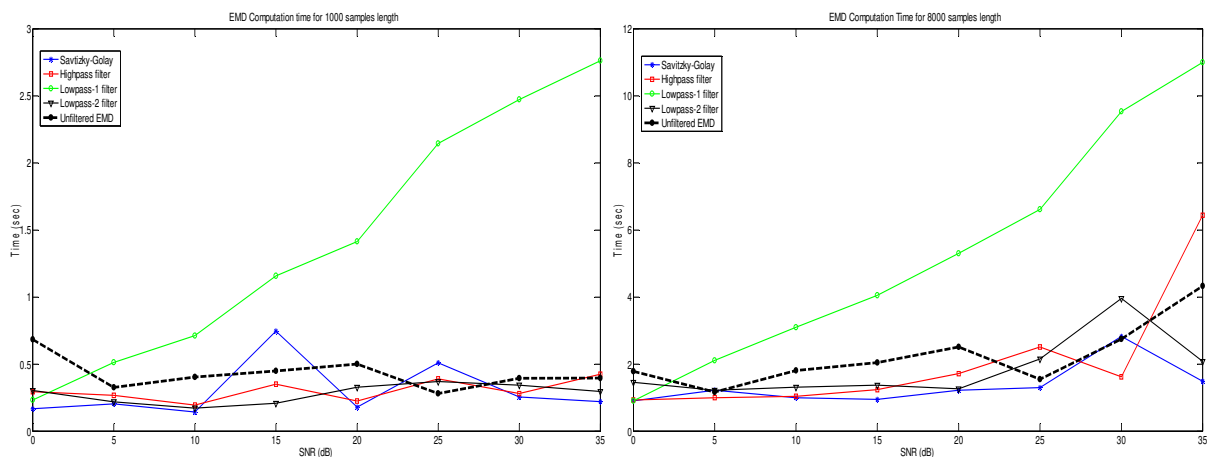
Fig. 6. Comparison results of EMD Computation Time for 1000 and 8000 samples of Simulated ECG time series

8. Conclusions - discussion

In practice, in noisy time series it is difficult to separate confidently information from noise. The implemented algorithm deduces a 95% bound for the white Gaussian noise in ECG time series. The core idea is based on the assumption that energy density of an IMF exceeds a noise bound if it represents statistically significant information.

Preprocessing stage affects the spectral characteristics of the input signal and any distortions of the time series' statistical and spectral contents have an effect in EMD performance. Based on the inherent properties of the time series to be processed, one may

select an appropriate preprocessing stage in order to achieve smaller number of IMFs and minimization of computation time without changing in a significant degree the physical content of IMFs.

Total computation time is an essential aspect that should be taken under consideration when implementing EMD algorithm on resource constrained systems. It is concluded that time series length, number of extrema and total number of iterations are significant parameter determining total computation time.

Simulation campaigns remain the only way to study EMD performance and various issues related to the method due to the lack of analytical expression and solid theoretical ground.

EMD implementation takes into account the termination criterion, a significant parameter to be optimized in order to avoid numerous iterations for the extraction of IMFs. Research effort is still to be undertaken to investigate in what degree tight restrictions in number of iterations drain the physical content of IMFs. An optimization procedure for both termination criterion and number of iterations is an open issue in this field.

Considering ECG time series of low SNR levels, noise is prevalent resulting in smoother spline curves and generally faster extraction due to smaller number of iterations. In high SNR, a tendency is observed towards the increase of computation time raising the issue of the optimum magnitude of noise to be added in the signal in NADA methods.

Empirical mode decomposition is a widely used method which has been applied on multiple biomedical signals for the processing and analysis. Focus is given on both application issues as well as the properties of the method and the formulation of a mathematical basis. Since this issue is addressed the only option remains the simulation and numerical experiments. It has been proved that empirical mode decomposition has various advantages compared to other methods which are employed in biomedical signal processing such as wavelets, Fourier analysis, etc. Research interest about the method is rapidly growing as it is represented by the number of related publications.

9. References

[1] Kendall M. Time-Series. Charles Griffin, London,UK,2nd edition,1976

[2] Papoulis A. Probability, Random Variables and Stochastic Processes. McGraw-Hill, New York, NY, 1965

[3] Huang, N. E. , Z. Shen, and S. R. Long, M. C. Wu, E. H. Shih, Q. Zheng, C. C. Tung, and H. H. Liu, 1998: The empirical mode decomposition method and the Hilbert spectrum for non-stationary time series analysis, Proc. Roy. Soc. London, 454A, 903-995.

[4] Semmlow J.L., Biosignal and Biomedical Image Processing, Signal Processing and Communications Series, Mercel Dekker, NY, 2004

[5] Priestley, M. B. 1965 Evolutionary spectra and non-stationary processes. J. R. Statist. Soc. B27, 204{237.

[6] S. Hahn: Hilbert Transforms in Signal Processing. Artech House, 442pp, 1995

[7] N.E Huang, M.C Wu, S.R Long, S.S.P Shen, W. Qu, P. Gloersen, K.L Fan, A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. Proc. R. Soc. A 459, 2317–2345 pp. doi:10.1098/rspa.2003.1123, 2003

[8] J. C. Echeverría, J. A. Crowe, M. S. Woolfson and B. R. Hayes-Gill, Application of empirical mode decomposition to heart rate variability analysis. Med. Biol. Eng. Comput. Volume 39, Number 4, 471-479pp, DOI: 10.1007/BF02345370, 2001

[9] Abel Torres, José A. Fiz, Raimon Jané, Juan B. Galdiz, Joaquim Gea, Josep Morera, Application of the Empirical Mode Decomposition method to the Analysis of Respiratory Mechanomyographic Signals, Proceedings of the 29th Annual International Conference of the IEEE EMBS Cité Internationale, Lyon, France

[10] M. Blanco-Velasco, B. Weng, KE Barner, ECG signal denoising and baseline wander correction based on the empirical mode decomposition. Comput. Biol Med; 38(1):1-13pp 2008 Jan

[11] AJ Nimunkar, WJ Tompkins. R-peak detection and signal averaging for simulated stress ECG using EMD. Conf Proc IEEE Eng Med Biol Soc. 2007; 1261-1264pp, 2007

[12] S. Charleston-Villalobos, R. Gonzalez-Camarena, G. Chi-Lem,; T. Aljama-Corrales, Crackle Sounds Analysis by Empirical Mode Decomposition. Engineering in Medicine and Biology Magazine, IEEE Vol. 26, Issue 1, Page(s):40 – 47pp, Jan.-Feb. 2007

[13] B.N. Krupa, M.A. Mohd Ali, E.Zahedi. The application of empirical mode decomposition for the enhancement of cardiotocograph signals. Physiol. Meas. 30, 729-743pp, 2009

[14] A. O. Andrade, V. Nasuto, P. Kyberd, C. M. Sweeney-Reed, F.R. V. Kanijn, EMG signal filtering based on Empirical Mode Decomposition, Biomedical Signal Processing and Control, Volume 1, Issue 1, 44-55 pp, DOI: 10.1016/j.bspc.2006.03.003, January 2006

[15] Y. Zhang, Y Gao, L Wang, J Chen, X Shi. The removal of wall components in Doppler ultrasound signals by using the empirical mode decomposition algorithm, IEEE Trans Biomed Eng. Sep; 54(9):1631-1642 pp, 2007

[16] Yeh JR, Sun WZ, Shieh JS, Huang NE Intrinsic mode analysis of human heartbeat time series, Ann Biomed Eng. 2010 Apr;38(4):1337-1344 pp. Epub 2010 Jan 30

[17] Karagiannis A., Constantinou, P., Noise components identification in biomedical signals based on Empirical Mode Decomposition, 9th International Conference on Information Technology and Applications in Biomedicine, ITAB 2009. 10.1109/ITAB.2009.5394300, 2009

[18] Karagiannis A., Loizou L., Constantinou, P., Experimental respiratory signal analysis based on Empirical Mode Decomposition, First International Symposium on Applied Sciences on Biomedical and Communication Technologies,. ISABEL 2008. 10.1109/ISABEL.2008.4712581, 2008

[19] Karagiannis, A.; Constantinou, P.; , "Investigating performance of Empirical Mode Decomposition application on electrocardiogam," Biomedical Engineering Conference (CIBEC), 2010 5th Cairo International , vol., no., pp.1-4, 16-18 Dec. 2010 doi: 10.1109/CIBEC.2010.5716048

[20] Z. Wu, N.E. Huang: A study of the characteristics of white noise using the empirical mode decomposition method. Proc. R. Soc. London, Ser. A, 460, 1597-1611 pp, 2004

[21] P.Flandrin, G. Rilling, P. Goncalves, Empirical Mode Decomposition as a filter bank. IEEE Signal Process Letter, 11, 112-114 pp, 2004.

[22] Karagiannis, A.; Constantinou, P.; "Noise-Assisted Data Processing With Empirical Mode Decomposition in Biomedical Signals," Information Technology in Biomedicine, IEEE Transactions on, vol.15, no.1, pp.11-18, Jan. 2011 doi: 10.1109/TITB.2010.2091648

[23] http://www.physionet.org/physiobank/database/mitdb

This book presents a collection of recent and extended academic works in selected topics of biomedical signal processing, bio-imaging and biomedical ethics and legislation. This wide range of topics provide a valuable update to researchers in the multidisciplinary area of biomedical engineering and an interesting introduction for engineers new to the area. The techniques covered include modelling, experimentation and discussion with the application areas ranging from acoustics to oncology, health education and cardiovascular disease.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

INTECH
open science | open minds