Universidade do Porto

FEUP

Master in Biomedical Engineering

HEART RATE VARIABILITY: A FRACTAL ANALYSIS

AUTHOR: ÓSCAR BARQUERO PÉREZ
TUTOR: JOAQUIM PONTES MARQUES DE SÁ
CO-TUTOR: JOSÉ LUIS ROJO ÁLVAREZ

June 1, 2008

Dissertation
Submitted to the
FEUP, Universidade do Porto
in Partial Fulfillment
of the Requirements for the Degree of
Master of Science in Biomedical Engineering.

HEART RATE VARIABILITY: A FRACTAL ANALYSIS

Óscar Barquero Pérez

Universidade do Porto
Faculdade de Engenharia
June 1, 2008

A mi familia y a Rebeca

Resumo

A Variabilidade da Frequencia Cardaca (VFC) é definida como a variação no intervalo entre batimentos cardíacos consecutivos, o as variações entre ritmos cardíacos instantâneos consecutivos que ocorrem no coração como consequência de um complexo equilíbrio dinâmico interno. Dado que o estado do sistema nervoso autônomo, e várias doenças relacionadas, podem ser investigadas de forma não invasiva por meio da VFC, existe um grande conjunto de índices utilizados para avaliar a condição do sistema cardíaco através do sinal de VFC.

O objetivo deste trabalho é definir o conceito de VFC e os principais métodos utilizados para avaliá-la, assim como descrever em detalhe os métodos de quantificação da VFC basseados na teoría fractal. Estes métodos tencionam quantificar as propriedades de correlação fractal dos sinais auto-semelhantes, como o movimento Browniano fraccionário (mBf), e têm ganhado recentemente um grande interesse na avaliação da VFC debido ao facto desta exibir algumas propriedades características de sinais fractais, nomeadamente, comportamento espectral $1/f$ e propriedades de dependência de longo-prazo. No entanto, os significados fisiológicos destes índices fractais e a sua relação com a idade não estão completamente estabelecidos.

Neste trabalho, primeiro apresentamos um estudo detalhado sobre a escholha dos parâmetros envolvidos nos algoritmos utilizados na caracterização das propriedades de correlação fractal, assim como a comparação dos seus desempenhos, utilizando o mBf como sinal sintético. A seguir, estudamos os distintos métodos utilizando bases de dados reais con dois objectivos: primeiro, analisar as capacidades de dicriminação das características fractais do ritmo cardíaco en termos de distinção entre sujeitos saudáveis e sujeitos com Insuficiência Cardíaca Congestiva (ICC); e segundo, estudar a dependência destas características fractais com a idade, quer nos sujeitos saudáveis quer nos sujeitos com ICC, a fim de perceber melhor a relação entre os índices fractais e seus significados fisiológicos.

Face aos resultados, encontramos que, entre os diferentes índices estudados o expoente de

escala α_1 é o melhor en termos de discriminação entre sujeitos saudáveis e sujeitos com ICC. Encontramos também que o expoente de escala α_2 e o expoente de Hurst possuem um crescimento estacionário em função da idade para sujeitos saudáveis, enquanto o expoente de escala α_1 não muda significativamente entre sujeitos jovens e sujeitos idosos.

Desta forma é possível concluir que a perda de complexidade debida a ideda reflecte-se como uma ruptura nas propriedades de correlação fractais do sinal de VFC quantificadas pelo expoente de Hurst e α_2, enquanto as propriedades fractais quantificadas pelo α_1 são preservadas com a idade. A perda de complexidade debida à ICC reflecte-se como uma completa ruptura das propriedades fractais do sinal de VFC.

Abstract

Heart Rate Variability (HRV), is defined as the variation in the interval between consecutive heart beats, or the variations between consecutive instantaneous heart rates that occur in a person as a consequence of a complex internal dynamic balance. Since the state of the autonomic nervous system, and of several related diseases, can be investigated non-invasively by the HRV, there exist a large set of indices used to assess the condition of the cardiac system via the HRV signal.

The aim of this work is to define the HRV concept and the main methods used to assess it, as well as to describe in detail the HRV quantification methods based on the fractal theory. These methods try to quantify the fractal correlation properties of self-similar signals, as the fractional Brownian motion (fBm), and they have recently gained interest in the HRV assessment due to the fact that HRV exhibits some properties characteristic of fractal signals, namely, $1/f$ spectral behavior and long-range dependence properties. However, the physiological meaning of these fractal indices and its relation with aging have not yet been completely established.

In this work, we first present a detailed study on the choice of the parameters for the algorithms that characterize the fractal correlation properties, as well as the comparison of their performance, using the fBm as a synthetic signal. Afterwards, we study the different methods using real datasets with two main objectives: first, the analysis of the discrimination capabilities of the heart rate fractal features allowing to distinguish between healthy and Congestive Heart Failure (CHF) subjects; and second, to study the dependence of these fractal features on aging for both healthy and CHF subjects, in order to get a better understanding on the relationship between the fractal indices and their physiological meaning.

We find that the scaling exponent α_1 is the best index to distinguish between healthy and CHF subjects. We find also that the scaling exponent α_2 and the Hurst exponent have a steady increase with aging for healthy subjects, whereas the DFA scaling exponent α_1 does not change

significantly between young and elderly subjects.

Therefore, we can conclude that complexity loss due to aging is reflected as a breakdown in the fractal correlation properties of the HRV signal quantified by Hurst exponent and α_2, whereas the fractal properties quantified by α_1 are preserved. The complexity loss due to CHF is reflected as a complete breakdown of the fractal properties of the HRV signal.

ACKNOWLEDGEMENTS

I would like to thank professors J. P. Marques de Sá and J. L. Rojo Álvarez (my personals inspectors Hopkins and Lestrade) for support, guidance, and advice (for present to me and discuss with me this complicated and stimulating *case*).

I especially thank Rebeca, she is my particular *Watson*, and also my particular *Irene Adler* (as well as I am hers) in this intricate *case*, *let's go to catch Moriarty*!.

I would also like to thank JL and Doctor Arcadi Garcia Alberola (my particular Mycroft Holmes in this amazing *case*) for the opportunity of working in an amazing area and making so much easy to reconcile study and work.

Also, I would like to thank the rest of the team from URJC ($\star$). Also, the people from INEB Signal Processing Group.

To my friends, and to my family.

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

El español siempre lo sabe todo. Y si de algo no sabe, dice: de esto ya hablaremos más adelante.
(J.L. López Aranguren, Spanish Philosopher)

Over the last three decades the study of the variations in the beat-to-beat timing of the heart, usually known as Heart Rate Variability (HRV), has been of great interest. HRV has been recognised as a marker of the autonomic activity which is related to the cardiovascular mortality [Malik 96]. The apparently easy derivation of this measure has popularized its use, but there still exists some discrepancies on its meaning and its significance as well as on its estimate on clinical practice. Therefore, its measurement from the Holter recordings has not yet become a routine clinical tool, but only a *research toy* [Huikuri 99].

The analysis techniques from the fractal theory applied to assess the HRV have recently gained wide interest. It was shown that HRV exhibits a $1/f$ spectral behavior in a wide range of frequencies, which is a property of the self-similar, fractal signals. The Hurst exponent is a quantitative measure of the fractal correlation properties of self-similar signals. In this work, we present a set of algorithms proposed for finding the Hurst exponent of fractal signals. In order to study in depth these algorithms we use a well-known process to model self-similar signals, namely, the fractional Brownian motion (fBm), whose Hurst exponent is known.

The objective of this work is, firstly, to study the performance and robustness of the different algorithms for estimate the Hurst exponent using the fBm as a synthetic signal. Secondly, once the usefulness of the estimation methods had been established, we study the estimation methods applied to real datasets from healthy and pathological subjects, in order to analyse the

capabilities of characterizing the different fractal properties from different cardiovascular states, as well as to study the dependence of these fractal properties on aging.

The structure of this work is as follows:

Chapter 1: In Chapter 1, the concept of Heart Rate Variability as well as its origins and its main applications are briefly introduced. Also, the methods of assessing Heart Rate Variability, the classical ones and the more advanced ones are outlined.

Chapter 2: In Chapter 2, the fBm as a suitable model to represent physical phenomena which involve long-term dependence and $1/f$ spectra is presented, as well as six different methods to estimate the fractal properties, namely, the Hurst exponent of fBm realizations.

Chapter 3: In Chapter 3, a set of synthetic experiments over fBm simulations are run in order to tune the free parameters of the estimation methods and to evaluate the performance of the methods.

Chapter 4: In Chapter 4, the estimation methods are applied on real datasets from healthy subjects and subjects suffering from Congestive Heart Failure. The aim of the experiments are twofold, on the one hand, to characterize the fractal correlation properties of the HRV from healthy and pathological subjects in order to discriminate both conditions, and on the other hand to study the variation of these fractal properties with aging.

Chapter 5: In Chapter 5, the main results obtained along the work, the limitations of the work, as well as further studies are presented.

Chapter 2

HEART RATE VARIABILITY

It is an old maxim of mine that when you have excluded the impossible, whatever remains, however improbable, must be the truth.
(Sherlock Holmes, Sir Arthur Conan Doyle)

The Heart Rate Variability (HRV) has been shown to be a powerful risk stratification criterion to reveal the relationship between the autonomic nervous system and cardiovascular mortality. In this chapter, the concept of Heart Rate Variability (HRV) is introduced, and the standard methods for measurement, and quantification of HRV are briefly presented.

2.1 What is Heart Rate Variability

HRV is the term used to describe the variations in the time intervals between consecutive heart beats. HRV is usually studied in any electrocardiogram (ECG) of sufficient duration, by analyzing a time series of beat-to-beat intervals. The extraction of the beat-to-beat intervals from the ECG can be achieved by measuring the time intervals between QRS complexes, which are the electrical marks registered in an ECG when a cardiac beat occurred, see Figure 2.1. Since HRV is related to Sinoatrial activity, the onset of P wave is an appropriate fiducial point. However, this approach is not very accurate since the P wave has a low amplitude; sometimes it is completely missing. Therefore, the fiducial point is commonly related to the R wave of the QRS complex [Sörnmo 05], see Figure 2.2.

HRV is due to the activity of the Sinoatrial Node (SAN), as the source of the repetitive impulses that generate the normal beats [Mateo 03]. The SAN is in turn influenced by the

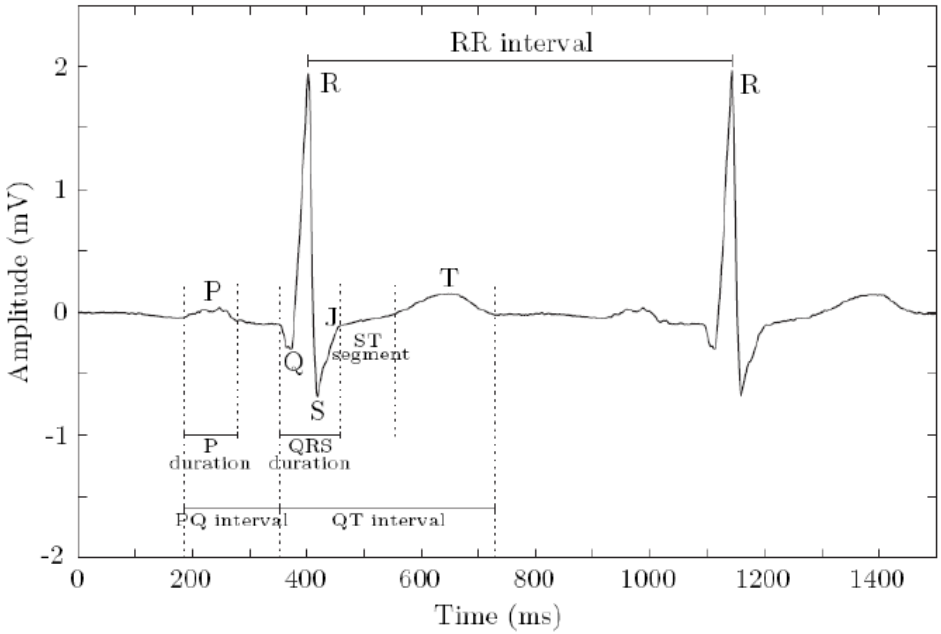


Figure 2.1: *Wave definitions of a cardiac cycle. The* ***P wave*** *reflects the sequential depolarization of the right and left atria, the* ***QRS complex*** *reflects depolarization of the right and left ventricles, and the* ***T wave*** *reflects ventricular repolarization. Taken from [Laguna 07]*

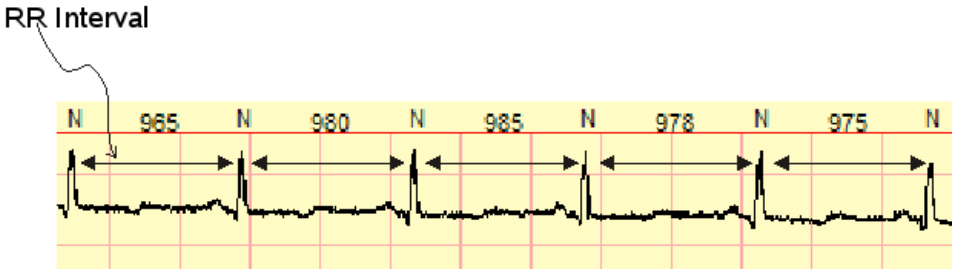


Figure 2.2: *RR intervals time series extracted from an ECG. The N label means normal beat. The time intervals between successive beats is in milliseconds.*

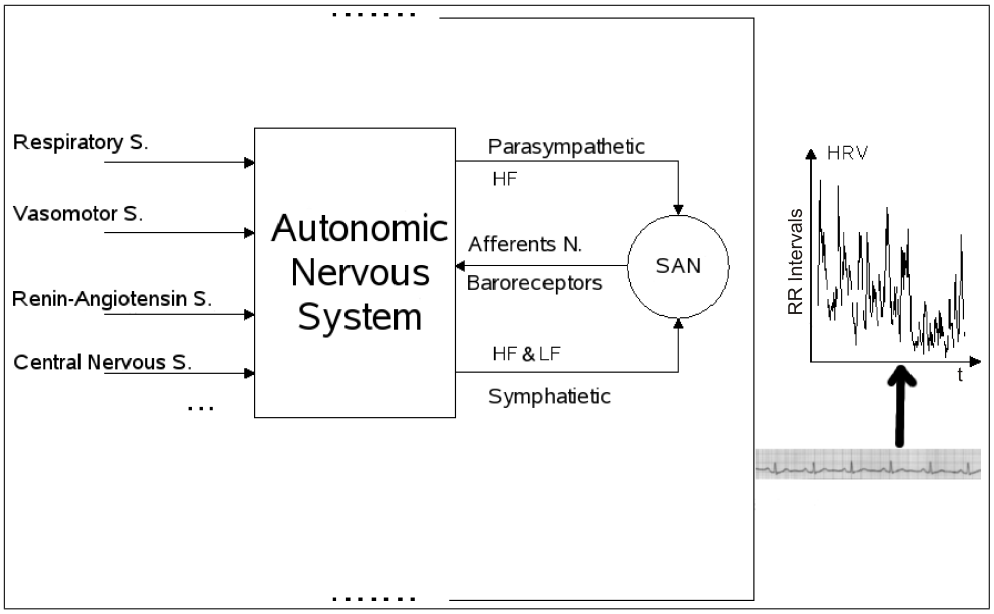


Figure 2.3: *HRV is due to the SAN activity which is modulated by the complex interactions between various systems.*

Autonomic Nervous System (ANS) activity, by the parasympathetic and sympathetic branches, that interacts in a complex way with a variety of reflexes and systems [Hainsworth 04]. The general behavior is outlined in Fig 2.3. HRV has been suggested as a noninvasively tool to assess the state of the system that controls the heart rhythm and its relationship with cardiovascular mortality.

The clinical relevance of the HRV has been established in numerous studies. First, Hon and Lee, in 1965, reported that fetal distress was preceded by alterations in the interbeat intervals. Wolf *et al.* in 1977 established an association of higher risk of post-infarction mortality in patients with reduced HRV [Malik 96, Malik 89b]. Indeed, it has been shown that low HRV is associated with some cardiac illness: myocardial infarction, atherosclerosis, heart failure and even with aging [Huikuri 99].

The underlying assumption when studying HRV is that short-term and long-term variations in heart rate have different physiological origins and the magnitude of these variations has been shown to be indicative of the autonomic state of the subject [Clifford 02], e.g., after a myocardial infarction, the innervation level of the heart decreases, and part of the nervous control of this organ can be lost. The HRV reflects this control loss and it makes possible the classification of cardiac sudden death risk groups [Malik 89b]. The degeneration of the ANS due to the aging

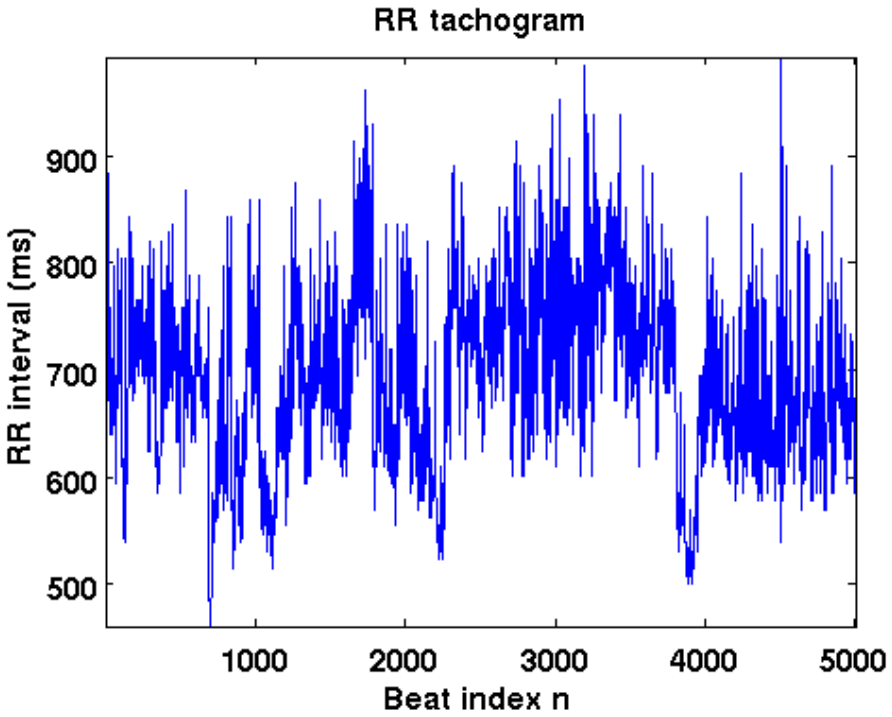


Figure 2.4: *RR-tachogram from a healthy subject.*

can also be inferred by the analysis of the HRV. Therefore, ideally, it would be possible to characterize different cardiovascular states, by just measuring the HRV, which would represent a noninvasive alternative to other methods.

2.2 Standard Descriptors of HRV

Once the meaning of the HRV has been outlined, in this section the standard methods which have been found to be the most useful for assessing the HRV will be listed and briefly described.

HRV analysis is usually done on RR-tachograms, also known as interval tachograms, that are time series of the RR-interval durations as a function of the interval number [Rompelman 77]. If t_n is the occurrence time of the n-th R-wave, the interval tachogram $IT(n)$ is given by:

$$IT(n) = (t_n - t_{n-1}) \tag{2.1}$$

In Fig 2.4 a RR-tachogram is plotted.

The methods used in HRV analysis can be very roughly divided into three main groups: time-domain methods, frequency-domain methods, and nonlinear methods.

Descriptor	Units	Description
AVNN	ms	Mean of NN intervals.
SDNN	ms	Standard deviation of all NN intervals.
SDANN	ms	Standard deviation of the averages of NN intervals in all 5 minute segments of the entire recording.
SDNNindex	ms	Mean of the standard deviations of NN intervals for all 5 minute segments.
RMSSD	ms	The Square root of the mean of the sum of the squares of differences between adjacent NN intervals.
NN50		Number of pairs of adjacent NN intervals differing by more than 50 ms in the entire recording.
pNN50	%	NN50 divided by the total number of NN intervals.

Table 2.1: *Statistical descriptors of the HRV.*

2.2.1 Time-domain methods

Time-domain methods are the simplest ones on computational terms. They treat the RR-interval sequence as an unordered set of intervals and employ different techniques to express the variance of such data. They can be split into two categories [Malik 96]: *statistical descriptors*, and *geometrical descriptors*.

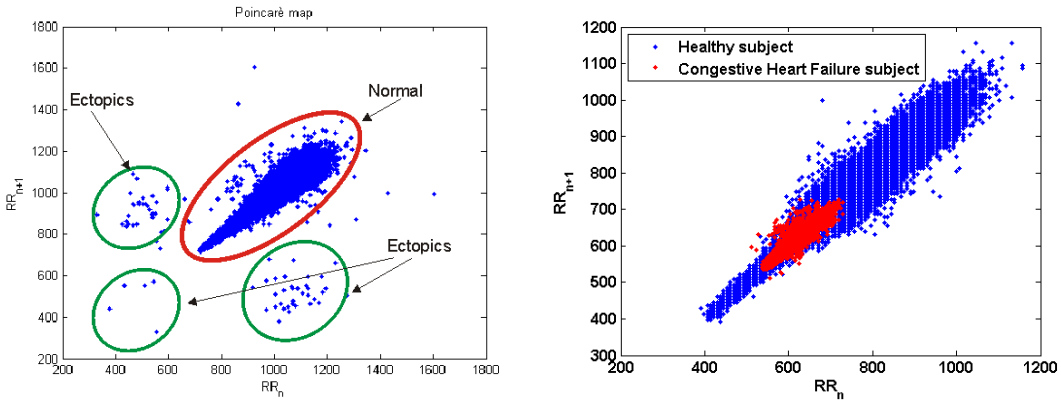
Statistical descriptors

Statistical descriptors generally involve the calculation of the standard deviation or variance of the RR-interval time series. These may be divide in two classes [Malik 96]:

- Those derive directly from the RR-intervals.
- Those derive from the differences between adjacent RR-intervals.

Table 2.1 summarizes the most common time-domain descriptors for characterization of HRV.

The main advantages are that these indices are computationally simples. However, statistical descriptors are highly depend on the data quality. This quality may be affected by outliers, and artifacts (i.e. noise generated by the equipment). Note also that these indices based on differences of adjacent RR-intervals reflects the high frequency variations (short-term variations) due to the derivative high-pass filter nature.



(a) (b)

Figure 2.5: *Poincaré map example (a). Comparison of Poincarè maps from a healthy subject and a congestive heart failure subject (b).*

Geometrical Descriptors

The geometrical descriptors come up from the search for more robust descriptors to face the lack of quality. They use the RR-intervals to construct a geometrical pattern and asses the HRV using a parameter of the pattern. There are three basic approaches in geometrical methods [Malik 95]:

1. The HRV is assessed by some geometrical measurements from the used pattern, e.g. the width histogram width.

2. The geometrical pattern is interpolated with a mathematically defined shape, and the parameters of the shape are used as HRV measures. e.g. the distribution of the histogram is approximated by a triangle.

3. The general pattern of the geometrical form is classified into one of several predefined categories, which represents different classes of HRV. e.g. Lorentz plots.

The most common geometrical patterns are: sample histogram from RR-intervals, sample histogram from the differences of adjacent RR-intervals, and Lorentz plots, that is a graph in which each RR interval is plotted against next RR interval, also known as Poincaré plots [Yang 06, Piskorski 07, Brennan 01], see Figure 2.5.

Table 2.2 summarizes the most common geometrical descriptors.

Descriptor	Units	Description
Triangular index	ms	Total number of all NN intervals divided by the maximum of the density function (height of the histogram of all NN intervals).
TINN	ms	Base width of the minimum square difference triangular interpolation of the highest peak of the histogram of all NN intervals
Lorenz plot scattering	ms	Representation of each NN-interval duration versus the duration of the previous interval.
Differential index	ms	Difference between the widths of the histogram of differences between adjacent NN intervals measured at selected heights.
Logarithmic index		Coefficient φ of the negative exponential curve $K \exp^{-\varphi t}$ which is the best approximation of the histogram of absolute differences between adjacent intervals.

Table 2.2: *Geometric descriptors of HRV.*

The main advantage using geometrical methods is their relative robustness to inclusion of outliers or artifacts. However, the application of these methods is limited to Holter recordings (preferably 24 hours duration) in order to get a sufficient number of intervals to construct a reliable histogram.

2.2.2 Frequency-domain methods

Frequency-domain methods are based on Power Spectral Density (PSD) estimation, that provides the basic information on how the power (i.e. the variance) is distributed as a function of frequency [Kay 81].

The HRV spectrum found in healthy subjects usually exhibits harmonic components. The different systems that modulate the heart rate (i.e. that modulate the behavior of the SAN) oscillate spontaneously with specific frequencies. Thus, when the PSD is taken from a HRV signal, it is expected to extract information on those systems related to cardiac autonomic function, that is, to identify the harmonic frequency components that correspond to each system. It is possible, for example, to quantify the power of the different spectral components in PSD as a measurement of the contribution of each system to the global variability [Cerutti 95, Sörnmo 05, Kamath 93].

In HRV analysis in the frequency domain, the PSD is generally estimated from RR-tachogram. Since it is a representation of the beat-to-beat variability, it is inherently a discrete and uneven time series (that is the reason for the *variability*). However, almost all of the PSD estimation methods require evenly sampled data. The two usual approaches are:

- Pretend that the data are, in fact, evenly sampled, so apply the PSD techniques directly to the RR-tachogram. In this approach, it is necessary to pay attention to the fact that the end up is not frequency domain but the so called *beatquency domain* [Lisenby 77]

- Interpolate and re-sampling the RR-tachogram in order to obtain an evenly sampled data and then apply PSD techniques on evenly data [Sörnmo 05]. In this approach the units are Hz.

The spectral analysis of HRV may be applied to short-term recordings, usually 5 minutes segments, or to long-term recordings, usually 24 hours recordings, which is the standard for Holter recordings in clinical environment. In short-term recordings, the PSD estimation shows three main spectral components [Sayers 73, Malik 96, Akselrod 81]:

1. Very Low Frequency (VLF): dc–0.04 Hz.

2. Low Frequency (LF): 0.04–0.15 Hz.

3. High Frequency (HF): 0.15–0.4 Hz.

The PSD measures in the LF and HF bands are closely associated with autonomic balance; an increase in parasympathetic activity is primarily related to an increase of the HF power, whereas an increase in sympathetic activity is related to an increase of the LF power. However, it is also accepted that the LF band has influences from the parasympathetic activity [Cerutti 95, Sörnmo 05], see Figure 2.6. The ratio between the power corresponding to the two components serves as index of autonomic balance [Malik 96].

The physiological origin of the VLF is not well known, a specific physiological process ascribable to this component could be inexistent [Malik 96].

Power spectral analysis of the HRV in long-term recordings (24 hours) includes an ultra-low frequency component (ULF), in addition to VLF, LF, and HF components. The oscillations in the ULF and VLF bands are thought to be due to long-term regulatory mechanisms such as the thermoregulatory system, the renin-angiontensin system and other factors, but the exact relation is not clear [Cerutti 95].

The methods for power spectral estimations are divided into two main groups: nonparametric an parametric. Both methods are briefly introduced next.

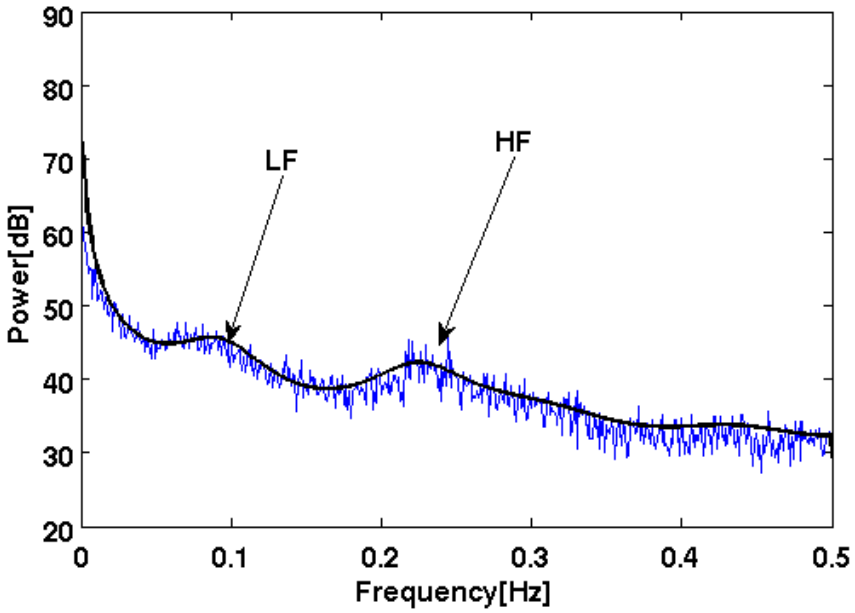


Figure 2.6: *PSD estimation of the same RR-tachogram. Comparison between nonparametric methods (fine line) and parametric methods (bold line), Welch periodogram and AR model respectively.*

Nonparametric methods

The nonparametric methods of spectral analysis are based on the Fourier Transform (FT). The PSD is usually estimated by the *Periodogram*, whose expression is:

$$P_{xx}(f) = \frac{1}{N\Delta t}\,|X(f)|^2 \tag{2.2}$$

where $X(f)$ is the FT of the time series that represents the HRV, N is the length of the time series, and Δt is the sampling period. The raw periodogram is an estimator with some limitations, namely, it is a biased estimator due to sharp truncation of the sequence (i.e. due to the *window*) and it is an inconsistent estimator, i.e. the variance at a given frequency does not decrease as the number of samples used increases. *Welch periodogram* is a modified periodogram that aims to avoid some of the previous limitations. The Welch method is based on splitting the time series in P overlapped segments multiplied by a time window. Then, the pooled average of the periodograms in each windowed segment is computed. The advantage of doing so is that the obtained estimate is asymptotically consistent [Oppenheim 00].

The *Lomb Periodogram* is another approach on the PSD estimation. It is based on fitting

non-evenly sampled sine waves of different frequency to the data using least-squares methods; hence, obtaining one component on the spectrum for each sine wave. This method does not need evenly sampled data and constitutes an alternative to the classical periodogram combined with interpolation and resampling [Laguna 95, Moody 93].

The advantages of the nonparametric methods are the simplicity of the algorithms involved and the speed of processing due to the use of the Fast Fourier Transform (FFT) as a base of the methods.

Parametric methods

The parametric methods assume that the HRV signal is the output of a given mathematical model. The output signal is generated by feeding noise to a filter. Then, the PSD can be estimated by the parameters of the model built from the time series [Cerutti 95, Malik 04]. It is necessary to verify the suitability of the chosen model and its complexity (i.e. the order of the model) [Moody 06]. The Auto-Regressive model (AR model) has been the most common parametric model used in the HRV spectral analysis [Baselli 86]. It is based on the assumption that each value of the series depends only on a weighted sum of the previous values of the same series plus noise, the expression is:

$$y(k) = \sum_{i=1}^{N} a_i y(k-i) + n(k) \quad , \tag{2.3}$$

where a_i are the coefficients of the model, $y(k)$ is the k-th value of the time series and $n(k)$ is the noise. There are different methods to obtain the coefficients of the model, e.g. *the autocorrelation method*, *the modified covariance method*, *Burg's method*. Once the AR model is estimated, the PSD can be estimated straightforwardly from the coefficients:

$$P_{xx}(f) = \frac{\Delta t \sigma_n^2}{\left| 1 + \sum_{n=1}^{N} a_n e^{-j2\pi f n \Delta t} \right|^2} \quad , \tag{2.4}$$

where Δt is the sampling period, and σ^2 is the noise variance.

The advantages of the parametric methods are smoother spectral components, and accurate PSD estimation even on a small number of samples.

Figure 2.6 shows a comparison between parametric and nonparametric PSD estimation from a healthy subject RR-tachogram. Note the similarity and the differences between both PSD estimations.

2.2.3 Nonlinear Methods

It has been suggested that the irregular and complex heart rate fluctuations observed in during normal sinus rhythm in healthy subjects, even at rest, are due in part to deterministic chaos, and that a variety of diseases may involve a (paradoxical) decrease in this type of nonlinear variability [Huikuri 03, Goldberger 99].

The linear methods do not provide information on nonlinear structures, since the linear methods assume that the time series under study is the output of a linear system. Therefore, the methods based on chaos theory and nonlinear theory have gained recent interest, in order to extract all the information that the HRV signal conveys.

The most common nonlinear measures used in HRV analysis can be divided into three categories, namely, chaos theory descriptors, fractal descriptors and information descriptors.

Chaos Theory Descriptors

A nonlinear system that exhibits highly sensitive to initial conditions and it is difficult to predict, it is considered to have a chaotic behavior. The descriptors from chaos theory attempt to characterize the dynamical behavior of the system [Lombardi 00]. The most common descriptors applied to HRV signal are:

- *The correlation dimension*, D_2. This descriptor assess the fractal dimension of the system attractor in the state space, i.e. measures the complexity of the dynamical system [Lombardi 96].

- *The Lyapunov exponent.* This descriptor measures the sensitive dependence on initial conditions of the chaotic systems [Signorini 94].

The nonlinear hypothesis implies that different pathologies involve a decrease in such a *nonlinear variability*, i.e. a loss of complexity in the systems. It should be noted that the application of this kind of nonlinear indices to HRV analysis has met some controversy. It is necessary to establish the chaotic behavior of the system prior to applying them. The use of these indices is usually justified by considering them simply as numerical indices [Glass 99].

Fractal Descriptors

It has been proposed that the heart beat fluctuations reveal statistical properties of fractals, i.e. do not have a characteristic scale (so called scale-invariance), presenting statistical self-similarity, long-range dependence, and power law correlations. Thus, it seems a good approach to apply the quantitative tools from fractal mathematics. Under this assumption the pathologies are associated with a breakdown of the fractal structure and can lead to structures with a dominating scale [Goldberger 99].

The most common indices applied in the HRV studies are:

- Power-law correlation exponents ($1/f$ spectrum exponent).
- The Hurst exponent.
- Scaling exponents from the Detrended Fluctuations Analysis (DFA) [Huikuri 00]

Information descriptors

The information descriptors quantify the irregularity of temporal series. This approach has to do with the information amount that contains a time series. The motivation to apply information descriptors to HRV signal is that many diseases are marked by less complex dynamical behavior. When a physiological systems become less complex its *information content* is degraded. Then, the information descriptors can assess the complexity of the physiological systems measuring the information content of the HRV signal. The descriptors most commonly used in the literature are:

- The Approximate Entropy (ApEn) [Pincus 90].
- The Sample Entropy (SampEn) [Richman 00].
- The Multiscale Entropy[Costa 02].

2.3 Conclusions

The aim of this chapter was to build a framework on HRV, and show "the sate of art" of the HRV topic. In doing so, firstly the heart rate variability concept was presented as well as a brief description of its physiological origins. It was outlined its relevance as a noninvasive tool.

Finally, a description of the standard measures of HRV was presented, namely, the classical ones: time-domain and frequency-domain methods; and the more advances, nonlinear methods.

Within all the standard measurements, the more challenging are the nonlinear ones, because allow to extract different, and maybe more complete, information about the state of the cardiovascular system. The idea underlying all nonlinear methods is that in a healthy subject the systems must to be capable to cope with the exigencies of a constantly changing environment, therefore the variability is viewed as a mark of the complex capability of behavior of the system (health system). In contrast, a reduced variability is viewed as a mark of a system that can not cope with changes, i.e. a system that has a highly regular behavior, a system that has highly predictable outputs. It must be noted that it is not yet clear how the variations in nonlinear indices values are related with the physiological components of the systems. This is a big handicap for the clinical applications.

Chapter 3

FRACTAL CHARACTERIZATION OF TIME SERIES

You do not really understand something unless you can explain it to your grand mother.
(Albert Einstein)

Early work on HRV analysis used linear systems approaches, as mentioned in the previous chapter. The fluctuations were considered as the sum of the outputs of a series of systems, which allows the application of the superposition principle. The HRV was quantified [Malik 96] with simple statistics or by splitting the spectra in frequency bands and obtaining the power in the bands [Berger 01]. The complex fluctuations of the heart rate often reveal power-law correlations, long-term dependencies, and scale invariance, hence suggesting a *fractal structure*, which means that the HRV signal is statistically self-similar [Goldberger 02, Goldberger 99]. Application of fractal techniques seems to be a promising approach to assess the cardiovascular system.

In this chapter, we present some common fractal indices that aim to characterize the statistical self-similarity properties of time series. First, the Fractional Brownian Motion (fBm) is introduced as a useful and usual model of self-similar process, and then, fractal indices are explained and applied to this model in order to yield an appropriate background of the fractal characterization of the time series.

3.1 Fractional Brownian Motion

Intuitively, a fBm is a *random walk* process, which means that it basically consists of steps in a random direction and with a step-length that has some characteristic value. Because it presents a deep connections with the concepts of self-similarity, fractal, long-range dependence or $1/f$ processes, it provides useful models for various field where such a concepts are relevant.

In order to get a good understanding of the fBm, the Brownian motion is next presented.

The **Brownian motion**, named after the botanist Robert Brown, who first observed this kind of motion in the random movement of particles suspended in a fluid. The mathematical model describing the Brownian motion is called Wiener process (after Norbert Wiener) [Dugnol 00].

Consider the movement of a particle on a line with the following constrains:

1. The particle starts at the origin in time $t = 0$.
2. In each discrete time step, the particle moves randomly either L or $-L$ displacement, with probability 0.5 each one.

The random variable $B(t)$ represents the particle position each time. In a more formal way, the function $B(t)$ is a random process with independent Gaussian increments with these properties [Dugnol 00, Mandelbrot 68]:

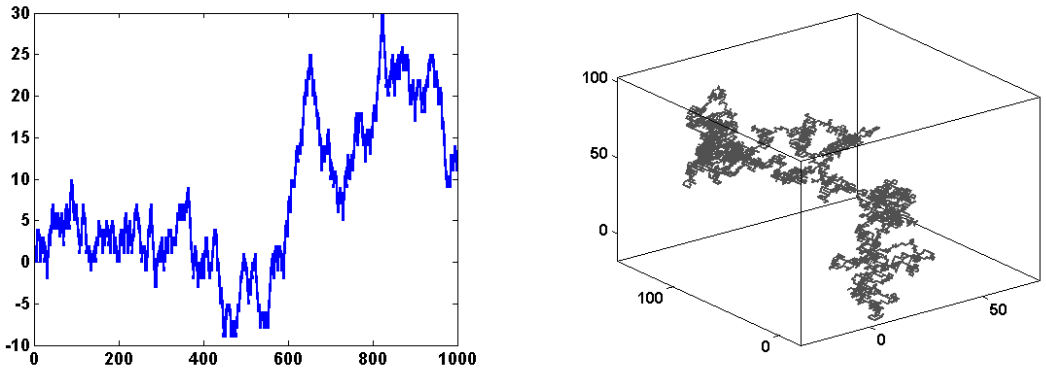
- $B(0) = 0$ with probability 1.
- The increments follow a Gaussian distribution with mean zero and variance $|t_2 - t_1|$

$$B(t_2) - B(t_1) \sim \mathcal{N}(0, |t_2 - t_1|) \tag{3.1}$$

  in such a way that, if the intervals (t_1, t_2) and (t_3, t_4) do not overlap, $B(t_2) - B(t_1)$ is independent of $B(t_4) - B(t_3)$.

Note also that the standard deviation of the increment $B(t+\tau) - B(t)$ is equal to $\tau^{1/2}$, commonly known as $\tau^{1/2}$ *law*. Therefore, the total variance is proportional to the time elapsed. In Fig 3.1 a Brownian motion simulation for one-dimension and for three-dimensions are presented.

fBm is a natural extension of ordinary Brownian motion [Flandrin 92, Mallat 01]. It is a Gaussian zero mean nonstationary stochastic process, denoted by $B_H(t)$. It is indexed by scalar parameter H, called Hurst index, Hurst exponent or Hurst parameter, which is a real number that satisfies $0 < H < 1$.



(a) (b)

Figure 3.1: *Brownian motions simulations. (a) 1-D Brownian motion simulation. (b) 3-D Brownian motion simulation*

A fBm process $B_H(t)$ with parameter H and starting value $B_H(0) = b_0$ at time $t = 0$, is defined for $t > 0$ as

$$B_H(t) - B_H(0) = \frac{1}{\Gamma(H+1/2)} \left\{ \int_{-\infty}^{0} \left[(t-s)^{H-1/2} - (-s)^{H-1/2} \right] dB(s) + \right.$$
$$\left. + \int_{0}^{t} (t-s)^{H-1/2} dB(s) \right\} \quad (3.2)$$

where $B(t)$ is the ordinary Brownian motion, and Γ represents the Gamma function

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} exp(-x) dx$$

Therefore, the fBm is a moving average of $dB(t)$, in which past increments of $B(t)$ are weighted by $(t-s)^{H-1/2}$ [Mandelbrot 68].

The most important properties of the fBm are next presented:

1. The covariance of the fBm is

$$\mathrm{E}\left\{B_H(t) B_H(s)\right\} = \frac{\sigma_H^2}{2} \left(|t|^{2H} + |s|^{2H} - |t-s|^{2H} \right) \quad (3.3)$$

where $\sigma_H^2 = \Gamma(1-2H)\frac{cos(\pi H)}{\pi H}$. Since covariance does not depend only on $t-s$, this proves that the fBm is nonstationary. The variance is obtained straightforwardly

$$\mathrm{Var}\left\{B_h(t)\right\} = \frac{\sigma_H^2}{2} \left(|t|^{2H} + |t|^{2H} - 0 \right) = \sigma_H^2 |t|^{2H} \quad (3.4)$$

2. The fBm is *statistically self-similar*, that is

$$B_H(ct) \stackrel{d}{=} c^H B_H(t) \tag{3.5}$$

where c is a constant, and $\stackrel{d}{=}$ means that the two members have the same distribution functions. This property means that the fBm can not be statistically distinguished from its re-scaled version (see Fig 3.2).

3. The increments of fBm are a Gaussian zero-mean *stationary* process with the following variance:

$$\mathrm{E}\left\{|B_H(t) - B_H(s)|^2\right\} = \sigma_H^2 |t-s|^{2H} \tag{3.6}$$

The value of the H parameter determines what kind of process the fBm is, allowing to categorize it as follows

- If $H = 1/2$, the process is in fact the Brownian motion, which means that the increments are uncorrelated.
- If $H < 1/2$, the successive increments of the process are negatively correlated, which results in a rough behavior of the signal.
- Ff $H > 1/2$, the successive increments of the process are positively correlated, which results in a relative smooth graph.

In Fig 3.3 two fBm simulations are presented, with values of the H equal to 0.1 for (a) and 0.9 for (b). Note that for $H = 0.1$ the negative correlation accounts for the high variability, which results in a rough graph, whereas for $H = 0.9$ there are clearly periods in which the sample path increases and periods in which it decreases, which result in smooth graph.

3.2 Power-Law Spectral Processes and fBm

Let be $X(t)$ a random stationary process with an autocorrelation function $R_X(\tau)$. The PSD for this process is

$$S(\omega) = \int_{\infty}^{\infty} R_x(\tau) e^{-j\omega\tau} d\tau \tag{3.7}$$

If $X(t)$ is statistically self-similar, this property is expressed in the frequency domain as a power-law:

$$S_X(\omega) \propto f^{-\beta} \tag{3.8}$$

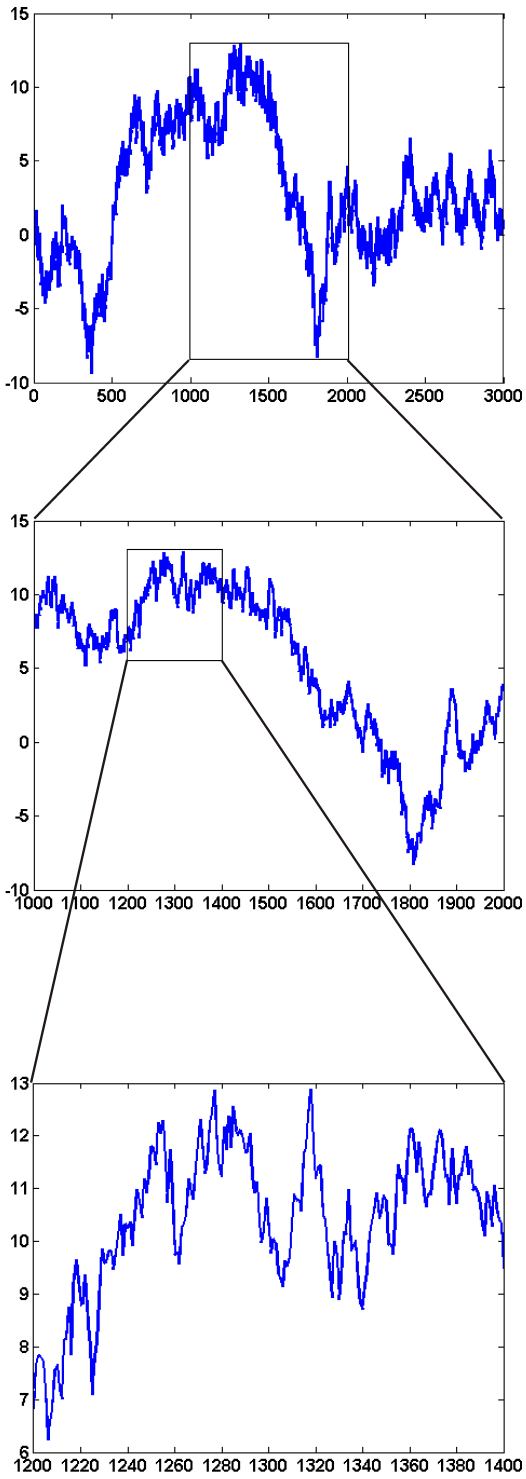
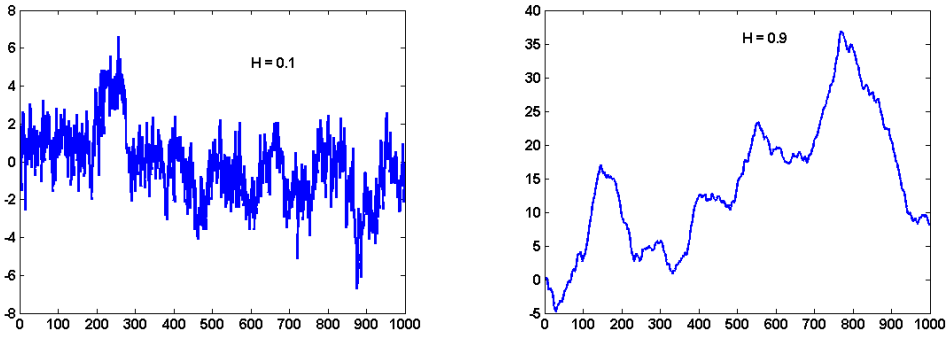


Figure 3.2: *A fBm process is statistically self-similar, which means that a re–scaled version can not be distinguish from the original version, in a statistic sense*



(a) (b)

Figure 3.3: *Two fBm simulations. Left (a) fBm with* $H = 0.1$ *results in a rough graph and right (b) fBm with* $H = 0.9$ *results in a smooth graph.*

It is possible to establish this power-law as an original criteria of self-similarity, i.e., their statistics are invariant to dilations and contractions of the time axis, to within an amplitude factor. As such, $1/f$ processes have no characteristic scale, i.e. exhibits *scale invariance*. In sum, the self-similarity inherent to $1/f$ processes typically gives rise to fractal structure in the associated time waveforms [Dugnol 00, Wornell 93]. Then, the power-law observed in the ULF and VLF components of the HRV spectrum, firstly reported in [Kobayashi 82], can be thought as an indication of a self-similar (fractal) structure.

In a more formal way, $X(t)$ is called an $1/f$ *process* if its PSD has the following expression:

$$S_x(\omega) \propto |\omega|^{-\beta} \tag{3.9}$$

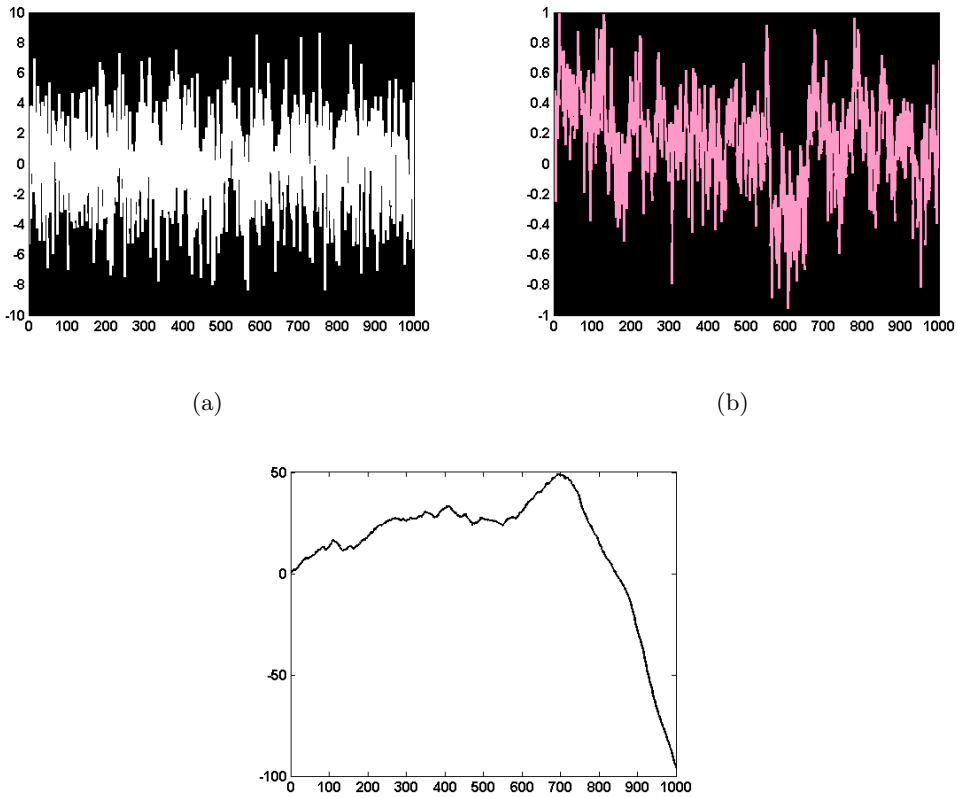
with $-\infty < f < \infty$ and $-1 < \beta < 3$.

Since fBm process is nonstationary, it is not possible to compute a classical PSD for this kind of process. However, it is possible to deduce an averaged spectrum with the following expression [Flandrin 89]

$$S_{B_H}(\omega) = \frac{\sigma_H^2}{|\omega|^{\beta}} = \frac{\sigma_H^2}{|\omega|^{2H+1}} \tag{3.10}$$

The fBm correspond to $1/f$ processes with $\beta = 2H + 1$ taking values between 1 and 3, because the H parameter is defined only for the range $(0, 1)$. But there are other interesting $1/f$ processes with different values for β parameter:

- For $\beta = 0$, it results in the so called *white noise*. Each value of the process is independent from the passed value, then results in a *surprise*, see Fig 3.4(a).



(a) (b)

(c)

Figure 3.4: *Colored noises with values of the β parameter: $\beta = 0$ white noise (a), $\beta = 1$ pink noise (b) and $\beta = 2$ black noise (c) respectively.*

- For $\beta = 1$, it results in the so called *pink noise*. This noise seems to be ubiquitous in nature, that is, there are many natural phenomena exhibiting scale invariance. Its structure displays alternate behavior from surprisingly changes to predictable changes. Note that this value of the β parameter is the usually attributed to healthy subjects in HRV analysis [Goldberger 99, Huikuri 00], see Fig 3.4(b). The existence of $1/f$ processes has been attributed to the so-called multiplicative (or cascade) effect.

- For $\beta > 2$, it results in the so called *black noise*. This noise is associated with financial crashes, and down markets trends, see Fig 3.4(c).

3.3 Hurst Exponent Estimation Methods

In this section, six different methods for the estimation of the Hurst exponent are presented. The H exponent is a quantitative index of the properties of a self-similar signal, e.g. the fBm. This self-similarity has the consequence that one individual realization of a self-similar process is a fractal curve, which means that it has a *fractal dimension*, i.e. a noninteger dimension. There is a straightforward relationship between the H exponent and the fractal dimension [Flandrin 92, Schepers 92], given by

$$D = E + 1 - H \tag{3.11}$$

where D is the fractal dimension, E is the Euclidean dimension. The Euclidean dimension is 1 for time series, and then

$$D = 2 - H \tag{3.12}$$

According to the possible range of the H parameters, it follows that D for the fBm process take values between 1 and 2 [Falconer 86]. There is an intuitive meaning for these values. If the value of H parameter is near 1, then the fBm realization is very smooth and it is almost a common curve. Therefore, the dimension must be the dimension of a line, i.e., $D \approx 1$. If the value of H is near 0, then the fBm realization is very rough and is therefore close to filling up a certain plane region, and the dimension must be the dimension of a plane, i.e., $D \approx 2$.

The H exponent characterizes the fractal structure of the self-similar signals. Accordingly, it is a candidate for the characterization of the fractal structure of HRV signal. Next, six methods for the H index estimation from a time series are presented.

3.3.1 Estimation Method Based on PSD

The self-similar characteristics of the fBm process in the time domain is expressed as a power law in the frequency domain [Flandrin 89]. It is possible to associate it with an average spectrum given by

$$S_{B_H}(\omega) = \frac{\sigma_H^2}{|\omega|^\beta} \tag{3.13}$$

with $\beta = 2H + 1$. This power law corresponds to a linear relationship in the log-log graph of the spectrum, this is,

$$\log(S_{B_H}(\omega)) \propto -\beta \log(\omega) \tag{3.14}$$

The method presented here is based on the determination of H using the *periodogram*, which is an estimate of the PSD.

Periodogram

The periodogram is an estimate of the PSD for random stationary signals. Assume that the infinite length random signal $x[n]$ is windowed by a window $w[n]$ of length N, resulting in the sequence $v[n] = w[n]x[n]$. The Fourier Transform (FT) of $v[n]$ is:

$$V(\omega) = \sum_{n=0}^{N-1} w[n]x[n]\mathrm{e}^{-j\omega n} \tag{3.15}$$

The periodogram is the estimate of the PSD given by:

$$P_{xx}(\omega) = \frac{1}{LU} |V(\omega)|^2 \tag{3.16}$$

where constant U is a normalization factor that eliminates the bias due to the window. Estimate $P_{xx}(\omega)$ is called *periodogram* when the window is rectangular and, it is called *modified periodogram* for other types of windows [Mitra 01, Oppenheim 00].

If the original time series has a non-zero mean, the PSD will have an impulse at zero frequency, and if the mean is relatively large, this component will dominate the PSD estimate. It is usual to subtract the mean value of the time series before computing the periodogram.

The variance of the periodogram for large realization sizes becomes proportional to the square of the original spectrum:

$$\mathrm{Var}\left\{P_{xx}(\omega)\right\} \propto S_x(\omega)^2 \tag{3.17}$$

where $S_x(\omega)$ is the original spectrum. This means that the periodogram strongly fluctuates around the true spectrum. A smoother PSD estimate can be obtained by the *periodogram averaging* method, which is discussed next.

Periodogram Averaging

The periodogram averaging PSD estimation method was first proposed by Bartlett and later modify by Welch. The method is based on the computation of the modified periodogram of R overlapping portions (sequences) of length N and then averaging these R periodograms. The r-th segment is:

$$x_r[n] = x[n + Kr]w[n], \qquad 0 \leq n \leq N - 1. \tag{3.18}$$

If $K < N$ there will be overlapping between adjacent segments. The periodogram of this r-th segment is

$$P_{xx}^{r}(\omega) = \frac{1}{LU} \left| X_r(\omega) \right|^2 \tag{3.19}$$

The Welch estimate is then given by the average of all R periodograms,

$$P_{xx}(\omega) = \frac{1}{R} \sum_{r=0}^{R-1} P_{xx}^{r}(\omega) \tag{3.20}$$

The variance of the periodogram, when estimated with the Welch method is

$$\mathrm{Var}\left\{ P_{xx}(\omega) \right\} \approx \frac{1}{R} S_x(\omega) \tag{3.21}$$

which means that the variance is inversely proportional to the number of periodograms R and tends to zero as R increases. Overlaps of more than 50% does not improve the results. Another advantage of segmenting the signal is that it partly corrects for the error of using the periodogram on a non-stationary signal like the fractal signal (e.g. fBm and HRV signals).

Hurst Parameter Estimation from Welch Periodograms

The Welch periodogram method estimate of the power spectrum of a fBm process can be used to determine the Hurst parameter. Once the PSD is estimated it is possible to determine β parameter as the slope of the straight line fitted using least squares.

$$\log(P_{xx}(f)) = -\beta \log(f) + C \tag{3.22}$$

Then, the computation of H index is straightforwardly given by

$$H = \frac{\beta - 1}{2} \tag{3.23}$$

Figure 3.5 shows an example of H estimation with Welch periodogram of a fBm realization with an actual H index $H = 0.8$. The value estimated from the slope of the regression line $H = 0.7696$ closely agrees with the real value.

From now on this estimation method will be called *Periodogram method*.

3.3.2 Continuous Wavelet Transform and Periodogram

The second method uses the Continuous Wavelet Transform (CWT) and the periodogram in order to estimate the PSD in a way that suits the characteristics of the fBm, namely, the nonstationarity and the self-similarity. First, the CWT will shortly be introduced, then, the estimation of the H parameter with the CWT and the periodogram will be presented.

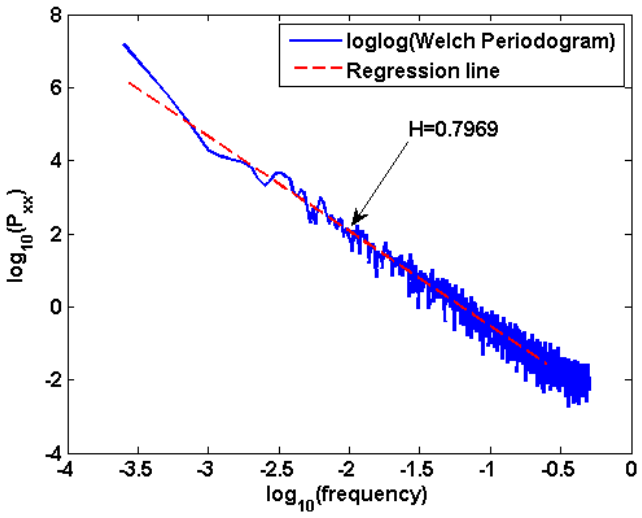


Figure 3.5: *Example of H estimation with Welch periodogram. The fBm realization has a real H index* $H = 0.8$. *The H estimated from the slope of the regression line (dotted line) is* $H = 0.7969$.

The CWT

The Wavelet Transform (WT) is a time-scale analysis. While the FT compares the similarity of the signal with a set of infinite sine waves of different frequencies, the WT compares the signal with finite waves and their scaled versions, the so-called wavelets [Flandrin 89, Sörnmo 05].

The CWT of a continuous signal $x(t)$ is defined by the correlation between $x(t)$ and a scaled and translated version of a function $\psi(t)$ called *mother wavelet*

$$W_x(s,\tau) = \int_{-\infty}^{+\infty} x(t) \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \, dt \tag{3.24}$$

where $s > 0$ is the scale parameter, τ is the translation parameter, and $\psi(t)$ is an arbitrary (but localized) function called *wavelet*, normalized so that its FT $\Psi(\omega)$ satisfies $\Psi(0) = 0$ and

$$\int_0^{+\infty} |\Psi(\omega)|^2 \frac{d\omega}{\omega} = 1 \tag{3.25}$$

$$|\Psi(\omega)| \underset{|\omega|\to\infty}{\longrightarrow} 0 \tag{3.26}$$

These requirements imply that wavelet function $\psi(t)$ must have bandpass characteristics, and hence, the CWT can be viewed as a kind of a bandpass analysis.

The CWT is the sum along time of the signal multiplied by scaled, shifted versions of a wavelet, which results in wavelets coefficients that are functions of time and scale [Misiti 07].

The computation algorithm of the CWT can be summarized in some simple steps, see Fig 3.6:

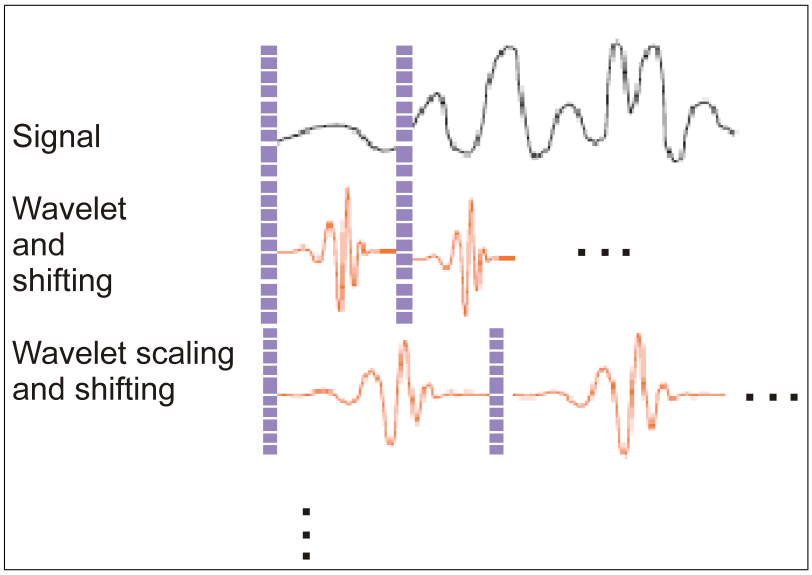


Figure 3.6: *CWT algorithm computation. The wavelet is compared with the whole signal, and then it is scaled and compared again with the whole signal.*

1. Choose a wavelet function and compare it to the starting segment of the signal. The comparison represents how much correlated are the wavelet and this signal segment, which results in a coefficient C.

2. Shift the wavelet and repeat step 1 until covering the whole signal.

3. Scale the wavelet and repeat steps 1 and 2.

Hurst exponent estimation with the CWT

Two important features of the fBm are to be taken into account when it is analyzed [Flandrin 89]:

- *Nonstationarity*, which needs *time-dependent* analysis.

- *Self-similarity*, which needs *scale-dependent* analysis.

The wavelet analysis fulfills both conditions, because it is by nature a time-scale method and appears as a natural tool analysis of fBm. The CWT of a fBm process is also itself a zero-mean random process

$$W_{B_H}(s,\tau) = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} B_H(\tau)\psi\left(\frac{t-\tau}{s}\right)\, dt \tag{3.27}$$

The autocorrelation function of the CWT of a fBm with respect a fixed scale s is

$$R_{B_H}(t_1, t_2; s) = \mathrm{E}\left\{W_{B_H}(t_1, s) W_{B_H}(t_2, s)\right\} \tag{3.28}$$

and using $\Psi(0) = 0$, and the Eq(3.3) we have

$$R_{B_H}(t_1, t_2; s) = -\frac{\sigma_H^2}{2}\, s^{2H+1} \int_{-\infty}^{+\infty} \gamma_\psi\left(u - \frac{t_1 - t_2}{s}\right) |u|^{2H} du \tag{3.29}$$

with

$$\gamma_\psi(u) = \int_{-\infty}^{+\infty} \psi(\theta)\psi(\theta - u) d\theta \tag{3.30}$$

The autocorrelation function $R_{B_H}(t_1, t_2; s)$ only depends on the difference $t_1 - t_2$, and then it is possible to obtain the PSD by taking the FT. In a practical implementation, the PSD can be estimated with the periodogram as explained in Section 3.5, which results in

$$S_{B_H}(\omega; s) = s\,|\Psi(s\omega)|^2\, \frac{1}{|\omega|^{2H+1}} \tag{3.31}$$

which can be interpreted as the fBm spectrum, as seen through the filter of nominal scale s.

The global spectrum can be obtain by adding up the contributions of each scale, this is

$$S_{B_H}(\omega) = \int_0^{+\infty} S_{B_H}(\omega; s) \frac{ds}{s} = \frac{1}{|\omega|^{2H+1}} \tag{3.32}$$

From this expression, the estimation of the H exponent is straightforward by taking the logarithm of the global spectrum

$$\log(S_{B_H})(\omega) = K + (2H + 1)\log(\omega) \tag{3.33}$$

which shows a linear trend with slope $2H + 1 = \beta$.

Figure 3.7 presents an example of estimation of the H exponent with CWT and periodogram. The fBm realization has a parameter $H = 0.8$. The estimation is obtained from the slope of the regression line fitted by least squares method to the global spectrum. The regression is only done in the frequencies that present a clear $1/f^\beta$ behavior. The result estimate is $H = 0.8329$.

From now on this estimation method will be called *CWT-and-periodogram method*.

3.3.3 Discrete Wavelet Transform and Periodogram

This method is the discrete version of the previous one. PSD estimate is obtained with the Discrete Wavelet Transform (DWT) and with the help of the periodogram. The DWT is next briefly introduced, and then the method of H parameter estimation is outlined.

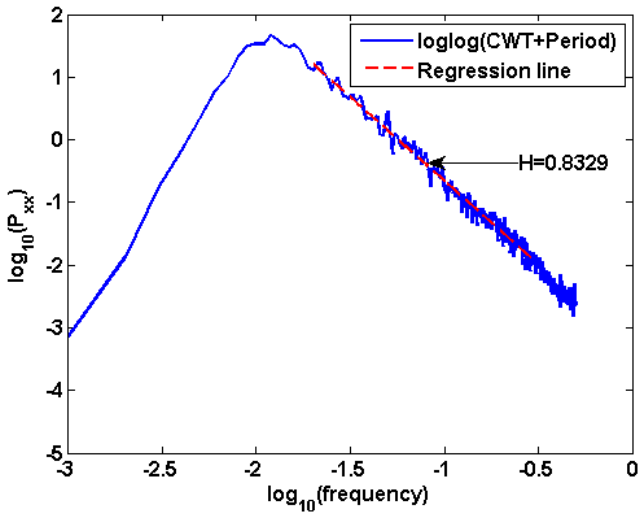


Figure 3.7: *Example of H estimation with CWT and Periodogram. The fBm realization has a real H index* $H = 0.8$. *The H estimated from the slope of the regression line (dotted line) is* $H = 0.8329$.

The DWT

The CWT produced an huge amount of data that is highly redundant. Therefore, it is necessary a less demanding method. The DWT uses discretized versions of the scaling and translation parameters s and τ. The method used to discretize is the *dyadic sampling* of the two parameters

$$s = 2^{-j}, \ \tau = k2^{-j} \tag{3.34}$$

where j and k are both integers. Accordingly, the wavelet is

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k) \tag{3.35}$$

The coefficients are now calculated according to

$$d(j,k) = \int_{-\infty}^{+\infty} x(t)\psi_{j,k}(t)dt \tag{3.36}$$

One approach in DWT is the *Multiresolution Analysis*. In this approach, the signal is viewed as the sum of a smooth (*coarse*) part and a detailed (*fine*) part. The smooth part is called the *approximation signal*, and it reflects the main features of the signal, whereas the detailed part contains the faster oscillations. In this DWT computation approach, the signal is divided into

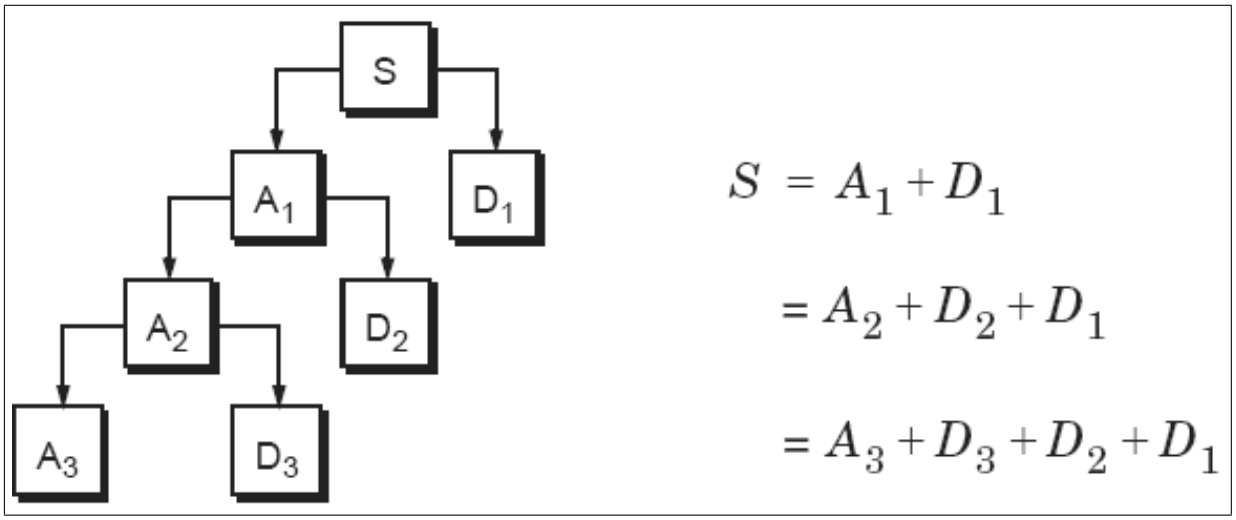


Figure 3.8: *Multiresolution method scheme. Decomposition of the original signal in approximations and details.*

approximations and *details* at different levels of resolution. Detail signals y_j at each level j can be calculated by using

$$y_j(t) = \sum_{k=-\infty}^{+\infty} d_j(k)\psi_{j,k}(t) \tag{3.37}$$

where $d_j(k)$ are the coefficients of the DWT applied to signal $x(t)$. The signal at level J can be recovered by:

$$x(t) = a_j(t) + \sum_{j \leq J} y_j(t) \tag{3.38}$$

Figure 3.8 presents the scheme of the multiresolution analysis.

Hurst Exponent Estimation with DWT and Periodogram

In this approach, the estimation is made from the global power spectrum. First, the signal is decomposed into its approximations. Then, the power spectrum for each approximation is obtained with the periodogram method, and finally, the power spectrum of the signal is reconstructed by summing up all the approximation spectrums. This method emphasises the lower frequencies due to the contribution of the approximations

$$S_{B_H}(\omega) = \sum_{j=1}^{N} \frac{S_{A_H}(\omega; j)}{(2^j)^2} = \frac{1}{|\omega|^{2H+1}} \tag{3.39}$$

Again, the power spectrum shows a linear trend in a log-log graph. The estimation of the H exponent is obtained from the slope of the regression line fitted with least squares method to

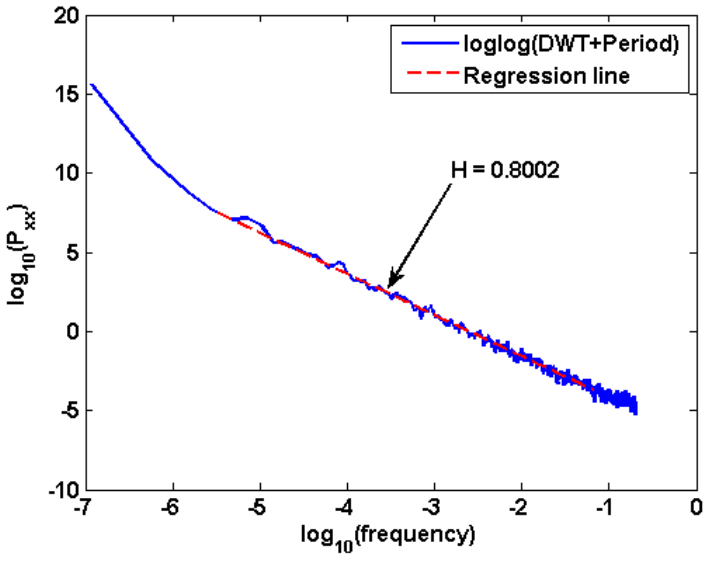


Figure 3.9: *Example of H estimation with DWT approximations and Periodogram. The fBm realization has a real H index* $H = 0.8$. *The H estimated from the slope of the regression line (dotted line) is* $H = 0.8002$.

the power spectrum.

Figure 3.9 presents an example of the estimation of the Hurst exponent with the DWT approximations and periodogram method. The fBm realization has an index value $H = 0.8$. The method estimation results in $H = 0.8170$.

From now on this estimation method will be called *DWT-and-periodogram method*

3.3.4 Variance of the DWT Coefficients Method

The DWT coefficients of a fBm process can be calculated as follows,

$$d_j[n] = 2^{j/2} \int_{-\infty}^{+\infty} B_H(t)\psi(2^{-j}t - n)dt \tag{3.40}$$

where $\psi(t)$ is the wavelet function. For each scale 2^j, the wavelet coefficients $d_j[n]$ form a discrete sequence for random coefficients [Flandrin 92].

The method here analyzed is based on a property that satisfies the DWT coefficients of fBm process. The coefficients of a scale j are time sequences which are self-similar and stationary

$$\mathrm{E}\left\{d_j[n]d_j[n]\right\} = -\frac{\sigma_H^2}{2} \int_{-\infty}^{+\infty} \gamma_\psi(\tau - n + m)\,|\tau|^{2H}\,d\tau \tag{3.41}$$

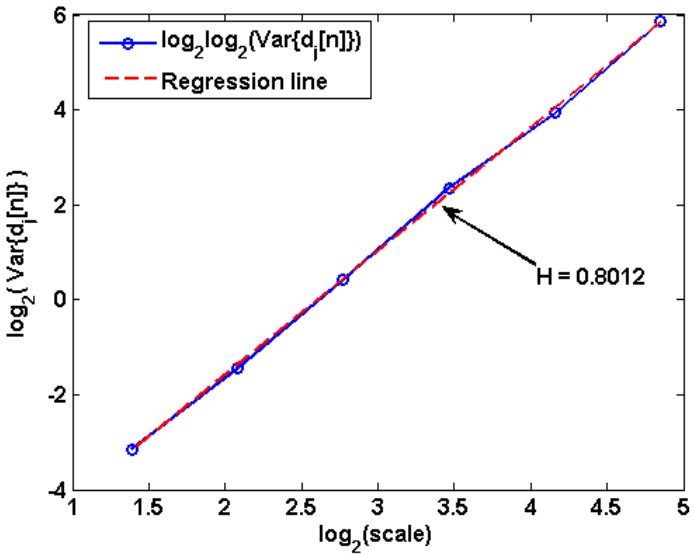


Figure 3.10: *Example of H estimation with Variance of DWT coefficients method. The fBm realization has a real H index* $H = 0.8$. *The H estimated from the slope of the regression line (dotted line) is* $H = 0.8012$.

with $\gamma_\psi(t) = \int \psi(t)\psi(t-\tau)dt$. Note that $\mathrm{E}\{d_j[n]d_j[n]\}$ depend only on $(n-m)$.

The variance of the wavelets coefficients yields a power-law behavior with the scale

$$\mathrm{Var}\{d_j[n]\} = \frac{\sigma_H^2}{2} K (2^j)^{2H+1} \tag{3.42}$$

where K is a constant which depends on both the chosen wavelet and the Hurst exponent.

The estimation of the Hurst index can be achieved by taking the logarithm of the previous expression:

$$\log_2(\mathrm{Var}\{d_j[n]\}) = (2H+1)j + K_2 \tag{3.43}$$

where $K_2 = \log_2(K)$.

Therefore, the Hurst exponent can be obtained from the slope of the variance as a function of scale in a log-log graph [Flandrin 92].

Figure 3.10 presents an example of the estimation of the Hurst exponent with the Variance of DWT coefficients method. The fBm realization has a real value index $H = 0.8$, the method estimate is $H = 0.8025$.

From now on this estimation method will be called *Variance-of-DWT-details method*

3.3.5 Rescaled Range Analysis

Rescaled range analysis is a way to characterise a time series that provides simultaneously a measure of variance and of the long-term correlation (or *memory*). The trend-corrected method is based on the statistical self-similarity in the signal. In the standard approach, one measures the ratio R/S of the range, R of the sum of the deviations from the local mean divided by the standard deviation, S from the mean [Bassingthwaighte 94]. For fractional Gaussian noise (fGn) processes, R/S is a power law function of the length τ of each segment of the set of segments into which the data set was divided [Taqqu 95], this is,

$$\frac{R(\tau)}{\tau} \propto \tau^H \tag{3.44}$$

fGn processes are long-range dependence random processes. They are related with fBm by differentiation. The fGn is the consequence of associating a derivative to the fBm [Mandelbrot 68].

Be fGn realization $x[n]$, with partial sum:

$$y[n] = \sum_{i=1}^{n} x[i] \tag{3.45}$$

and sample variance $S^2[n]$. The R/S statistic is given by

$$\frac{R[n]}{S[n]} = \frac{1}{S[n]} \left[\max_{0 \le t \le n} \left(y[t] - \frac{t}{n} y[n] \right) - \min_{0 \le t \le n} \left(y[t] - \frac{t}{n} y[n] \right) \right] \tag{3.46}$$

This statistic follows the rule

$$\mathrm{E} \left\{ \frac{R[n]}{S[n]} \right\} \approx n^H \tag{3.47}$$

The estimation of the H parameter from the R/S statistic can be achieved as follows:

1. Select a window with a fixed length.
2. The time series of length N is first subdivided into K blocks of size the length of the window.
3. R and S are computed for each block.
4. Compute the average of the R/S value, $R/S(\tau)_{avg}$ for all the blocks.
5. Select other length for the window a go to steps 2–5, (see Fig 3.11).

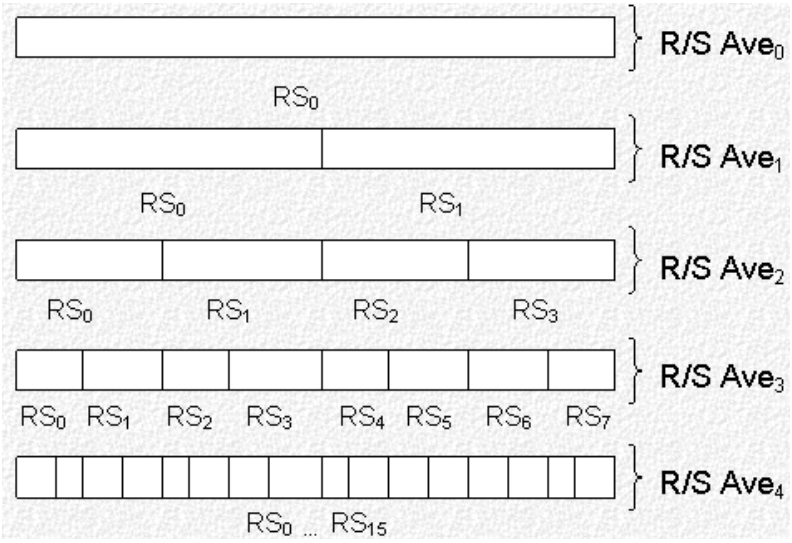


Figure 3.11: *Implementation of the R/S analysis.*

The values of $R/S(\tau)_{avg}$ are plotted as a function of the window length τ in a log-log graph. It is possible to find a power-law. The H parameter can be estimated as the slope of the regression line fitted by least-square method to the plot (see Fig 3.12). Note that when using R/S to estimate the H index of a fBm, it is first necessary to obtain the finite difference signal of the fBm realization.

3.3.6 Detrended Fluctuations Analysis

The Detrended Fluctuation Analysis (DFA) is a modified root mean square analysis of a random walk (i.e Brownian Motion).

Let be $x[n]$ a time series of length N, the DFA algorithm is as follows [Peng 95]:

1. First, the time series is integrated:

$$y(k) = \sum_{i=1}^{k} B(i) - \bar{B} \tag{3.48}$$

   where $\bar{B}$ is the mean of the time series.

2. Next, the integrated time series is divided into segments of equal length n.

3. In each segment, a least squares line is fitted to the data. This line represents the trend in that segment, see Fig 3.13. The straight line is denoted by $y_n(k)$ in each segment.

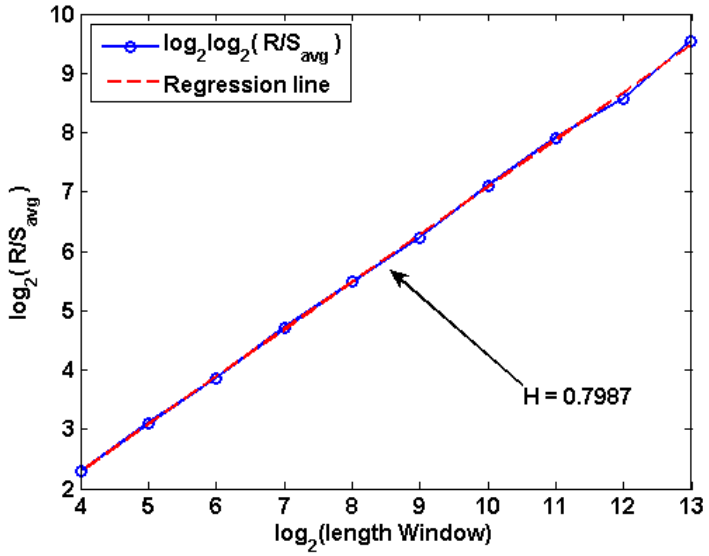


Figure 3.12: *Estimation of the H index with R/S analysis. The real index is $H = 0.8$ and the estimation obtained is $H = 0.7987$.*

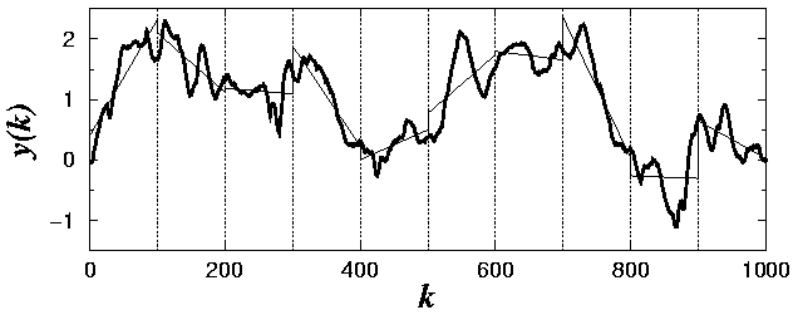


Figure 3.13: *The solid straight line segments represent the trend estimates in each segment.*

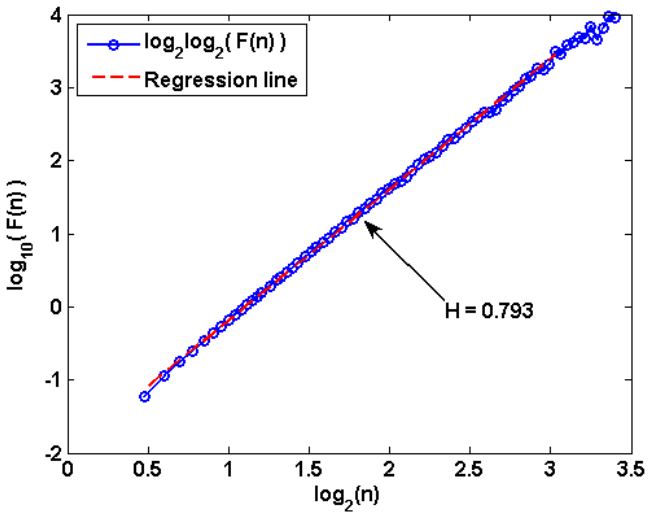


Figure 3.14: Estimation of the H index with DFA analysis. The real index is $H = 0.8$ and the estimation obtained is $H = 0.793$.

4. Finally, subtracting this trend from $y(k)$, the root-square fluctuation is calculated by:

$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(y(k) - y_n(k)\right)^2} \tag{3.49}$$

The computation is repeated over all scales, i.e. segments size to provide a relationship between $F(n)$ and the segment size n

$$F(n) \propto n^{\alpha} \tag{3.50}$$

The α exponent is related to the H parameter. Since integration raises scaling exponents by one, the Hurst parameter and the α parameter are related by [Borg 05]:

$$\alpha = H + 1 \tag{3.51}$$

Scaling exponent α can be estimated as the slope of a straight line fitted to the log-log graph using least-squares. In Figure 3.14, an example of the estimation of the H parameter is presented. The fBm realization has an index $H = 0.8$, and the scaling exponent α estimated is $\alpha = 1.793$, therefore, $H = 0.793$.

The advantages of DFA over conventional methods (periodogram and R/S method) are that it permits the detection of long-range correlations in time series with non-stationarities, and also avoids the spurious detection of apparent long-range correlations that are an artifact of

non-stationarity [Peng 95]. Therefore, is a suit method for HRV signal. In this method the HRV signal is thought as a composition of two classes of fluctuations, one arising from the dynamics of the complex systems which shows long-range correlations and the other type of fluctuation that have a characteristic time scale (frequencies related to a some stimuli), although highly non-stationary. The latter type of fluctuation can be considered as *noise*, an treated as *trend*, this is the reason for removed the trends in the algorithm. This trend can be distinguished from the more subtle fluctuations that may reveal intrinsic correlation properties of the dynamics.

Chapter 4

EXPERIMENTS USING SYNTHETIC SIGNALS

El 99% de la genialidad es sudor.
(Pedro Miguel Etxenike, Basque Physicist)

In this chapter, several experiments are performed in order to test the performance of the methods presented in the previous chapter. The fBm is the synthetic signal used in the experiments.

The aim of the tests is to compare the accuracy of the methods over signals with known Hurst parameter. The aim is also tuning the free parameters of the different methods (e.g. window choice in Periodogram method, wavelet choice in wavelet-based methods).

4.1 Synthetic Signal

The fBm signal model is the synthetic signal model used in the tests presented in this chapter. The fBm simulations are generated with the Matlab® function `wfbm`. This function implements the algorithm proposed by Abry and Sellan [Abry 96]; it uses a wavelet based approach to synthesize the fBm signals. The Matlab® function has two input parameters: `H`, the value of the Hurst parameter ($0 < H < 1$), and `L`, number of signal samples. The output of the signal is the variable `fbm`, that is, the fBm simulation with `H` parameter and `L` length. The following code is an example of fBm simulation:

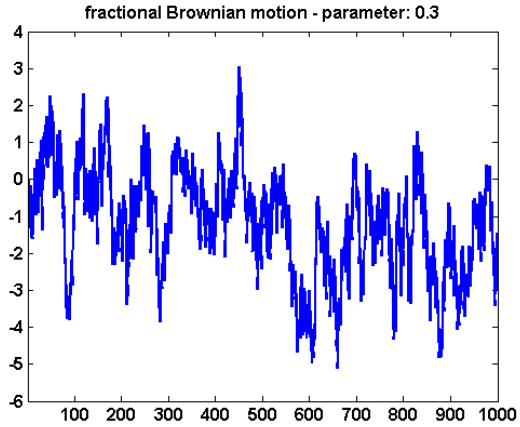


Figure 4.1: *Example of fBm simulation.*

```
% Generate fBm for H = 0.3 and H = 0.7
% Initialize the randn generator
randn('state',1)
% Set the parameter H and the sample length
H = 0.3; lg = 1000;
% Generate and plot wavelet-based fBm for H = 0.3
fBm03 = wfbm(H,lg);
plot(fBm03)
```

The result of one instance of the simulation is plotted in Figure 4.1.

4.2 Free Parameter Tuning Experiments

The estimation methods presented in the previous section have some parameters that are not fixed at any particular value by the theory. These parameters, called *free parameters*, are usually set (tuned) by carrying out controlled experiments where the outputs are known (in general). The methodology followed in this section is to compare the output of the synthetic experiment with the output of the estimation method as a function of the free parameter. When there exist more than one free parameter, a one-by-one approach is followed, that is, while one free parameter is being tuned the remaining ones are either set to some *reasonable* value, or set to the value previously obtained in the tuning process.

In this section, a set of experiments based on the fBm signal, which has a known H exponent, are developed in order to tune the free parameters of the estimation methods.

4.2.1 Free Parameters of the Periodogram method

The Periodogram method has two free parameters that need to be tuned, namely, window choice, and the sequence length, N, in which the signal is divided.

Window Choice

A window with good properties is the *Kaiser window*. It is nearly optimal in the sense of its peak's concentration around $w = 0$ frequency [Oppenheim 00]. For discrete time, the Kaiser window is expressed by,

$$w[n] = \frac{I_0\left(K_\beta\sqrt{1-\frac{4n^2}{(N-1)^2}}\right)}{I_0(K_\beta)} \quad with \quad -\frac{N-1}{2} \leq n \leq \frac{N-1}{2} \tag{4.1}$$

where N, the window length (i.e. the segment length), controls the main lobe width, K_β controls the amplitude for the sidelobes, and $I_0(x)$ is the modified zeroth-order Bessel function. The discrete-time Fourier transform of $w[n]$ is given by

$$W_K(\omega) = \frac{(N+1)\sinh\left(\sqrt{K_\beta^2 - \left(\frac{(N+1)\omega}{2}\right)^2}\right)}{I_0(K_\beta)\sqrt{K_\beta^2 - \left(\frac{(N+1)\omega}{2}\right)^2}}. \tag{4.2}$$

The larger the value of K_β, the narrower the window becomes, $K_\beta = 0$ corresponds to a rectangular window; conversely, for larger K_β the main lobe of $W_K(\omega)$ increases in width, while the sidelobes decrease in amplitude. Thus, this parameter controls the tradeoff between main lobe width and sidelobes area, as is illustrated in Figure 4.2. For large K_β, the shape of the Kaiser window (in both time and frequency domain) tends to a Gaussian curve.

The K_β parameter of the Kaiser window is a free parameter that needs to be tuned. To test the Periodogram method dependence on the variable K_β the following experiment was carried out. One hundred realizations of the fBm process were obtained for five different values of the H parameter ($H = 0.1, 0.3, 0.5, 0.7, 0.9$). Estimates of H parameter were obtained for each realization by Periodogram method varying the K_β parameter in the range $[2, 9]$. The length of the sequences, N (i.e. the other free parameter) is set to 4000. Figure 4.3(a) shows the mean and standard deviation of the H parameter estimate vs. the K_β parameter.

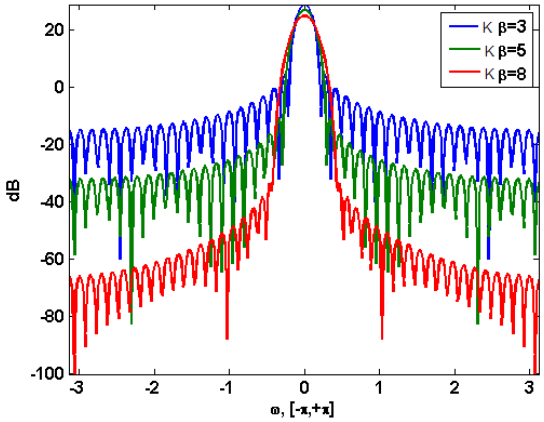


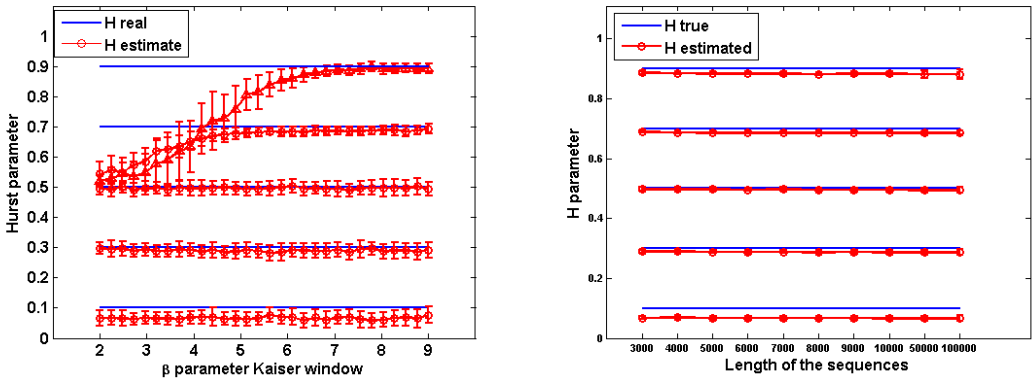
Figure 4.2: *Spectral representation of Kaiser window with three different values,* $K_\beta = 3$, $K_\beta = 5$, *and* $K_\beta = 8$.

The performance of the estimate has two different behaviors, which depend on the value of the true H parameter. For values between 0 and 0.5 the estimate is relatively independent of the K_β parameter choice, whereas for values between 0.5 and 1 it strongly depends on the K_β choice. When H is in the range $(0.5, 1)$, and for small values of the K_β parameter (between 2 and 6) the Periodogram method underestimates and it has large error. Conversely, for larger K_β (between 6 and 9) the Periodogram method estimates are more accurate and with lower error.

Therefore, the value for K_β parameter must be chosen from values between 6 and 9. We chose $K_\beta = 6.8$, which is a value that gives good estimates with small error for the whole range of the true H parameter and is one of the smallest possible values for K_β, which allows to obtain main lobe widths as smalls as possible.

Segment Length Choice

Welch's periodogram method is based on the computation of the modified periodogram of R overlapping segments of length N, and then, the average of these R periodograms is obtained. Therefore, the length of the segments N can be considered as a free parameter. Segmenting the signal has the advantage that it partially corrects for the error of using the periodogram on a non-stationary signal like fractal signals. In this experiment, the periodogram method dependence on the length of the segments is tested. The tests were performed for fBm signals with five different values of the H parameter, $H = 0.1, 0.3, 0.5, 0.7, 0.9$ and a total number of samples $L = 100000$.



(a) (b)

Figure 4.3: *Kaiser window K_β tuning experiment (a). Segment length tuning experiment (b).*

The H parameter estimates are obtained for each realization of the fBm by periodogram method with different values of N: $N = 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 50000, 100000$. The parameter K_β of the Kaiser window was set to 6.8.

Figure 4.3 shows the mean and standard deviation of the H parameter estimates vs the N free parameter. The estimates are relatively constant for all segments length and for all true H values, the error has small values and also a constant behavior. We chose $N = 10000$, because this choice allows to segment the signal in almost stationary sequences, and with enough samples to obtain reliable linear regressions in the frequency domain.

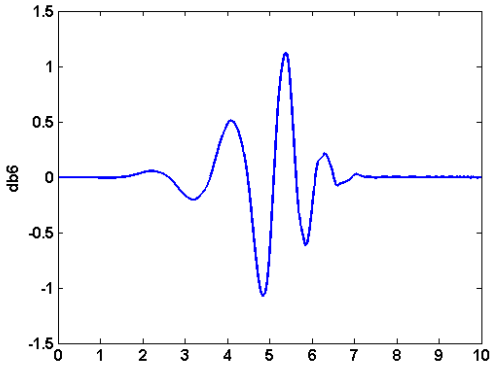
4.2.2 CWT-and-Periodogram Method Free Parameters

CWT-and-periodogram method has three free parameters that need to be tuned. Two of them are the same as for the Periodogram method, namely, window choice, and segment length. For these parameters, we used the values derived in the previous section for the Periodogram method, i.e. Kaiser window with $K_\beta = 6.8$ and segment length $N = 10000$. The remaining free parameter is the wavelet function for the CWT.

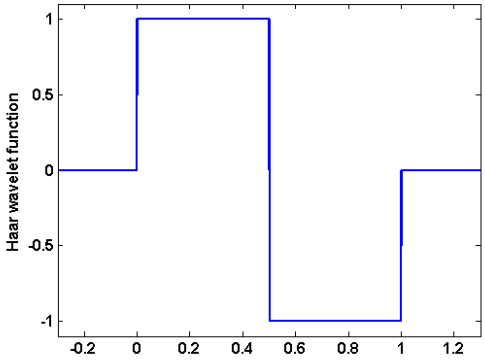
Wavelet choice

The wavelet will be chosen from the *Daubechies wavelets family*, which is a family of orthogonal wavelets. Figure 4.4 (a) shows the Daubechies wavelet *db*6.

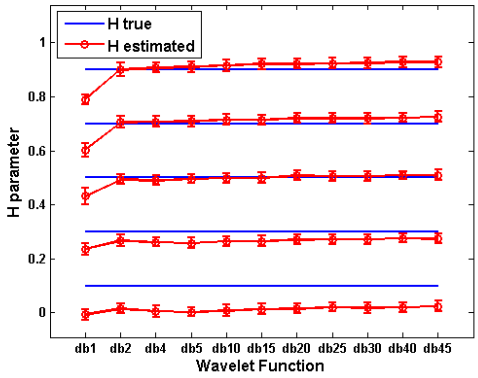
In order to choose the wavelet function from the Daubechies family one hundred fBm realiza-



(a)



(b)



(c)

Figure 4.4: *Daubechies-6 wavelet function (a). Haar wavelet function (b). Wavelet choice experiment for CWT-and-periodogram method (c).*

tions were generated for each different value of the H parameter from the set $\{0.1, 0.2, 0.3, 0.5, 0.7, 0.9\}$. The H estimate was obtained for each realization by CWT-and-periodogram method varying the wavelet function from the following set:

$$\text{Wavelet Function} = \{db1, db2, db4, db5, db10, db15, db20, db25, db30, db40, db45\}$$

Figure 4.4 (c) shows the mean and standard deviation for the H estimates as a function of the *Wavelet function*.

The estimate has two main different behaviors, which depend on the value of the true H parameter. For values between 0 and 0.5 the CWT-and-periodogram method tends to underestimate, whereas for values between 0.5 and 1 the method tends to overestimate. The general behavior of the mean H estimates is to grow when the Daubechies wavelet function grows (i.e.

$db1, db2, ..., db45$). The wavelet functions $db1$ is an exceptional case, which gives the worst H estimates. The wavelet function $db1$ correspond to the so called *Haar wavelet*, which is the simples wavelet possible (see Figure 4.4 (b)). The Haar Wavelet's mother wavelet function $\psi(t)$ can be described as

$$\psi(t) = \begin{cases} 1 & 0 \leq t < 1/2, \\ -1 & 1/2 \leq t < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

The reason for the bad estimates is the shape of the Haar wavelet; this shape does not match very well the shape of the fBm signal, and therefore gives bad H estimates.

The wavelet function that gives tighter estimates for all true H values is $db5$, therefore we chose $db5$ as the wavelet function for the following experiments with the CWT-and-periodogram method.

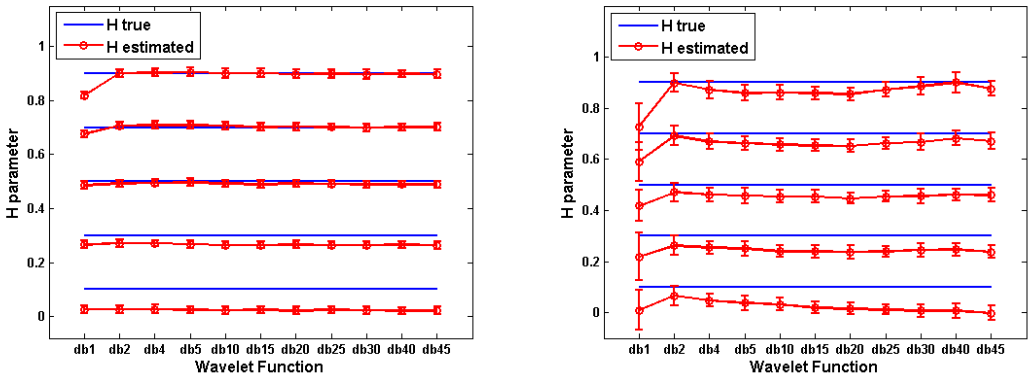
4.2.3 DWT-and-Periodogram Method Experiments

DWT-and-periodogram method is the discrete version of the CWT-and-periodogram method. Therefore, the free parameters that need to be tuned are the same, namely, the wavelet function and the free parameters of the Periodogram method. As in the CWT-and-periodogram method, the value of the free parameters for the periodogram are the previously obtained, namely, Kaiser window with $K_\beta = 6.8$, and segment length $N = 10000$.

Wavelet Function Choice

The experiment for the wavelet function has the same scheme as the experiment for CWT-and-peridogram method, in the previous section. Figure 4.5 (a) shows the mean and standard deviation of the H estimates for the DWT-and-periodogram method.

We rule out the first wavelet function because it corresponds to the Haar wavelet function (see Section 4.2.2, page 43). The general behavior is approximately steady, for both mean and standard deviation, and for all wavelet functions. It seems, for this method, that the H exponent estimates are independent of wavelet function. We chose *db5* as the wavelet function for the DWT-and-periodogram method, just for reasons of symmetry with the CWT-and-periodogram method.



(a) (b)

Figure 4.5: *Wavelet function choice experiment for DWT-and-periodogram method (a), and for the Variance-of-DWT-details method (b).*

4.2.4 Free Parameter of the Variance-of-DWT-Details Method

This is also a wavelet-based method, but it does not use the periodogram; thus, it only has one free parameter, namely, the wavelet function.

Wavelet Function Choice

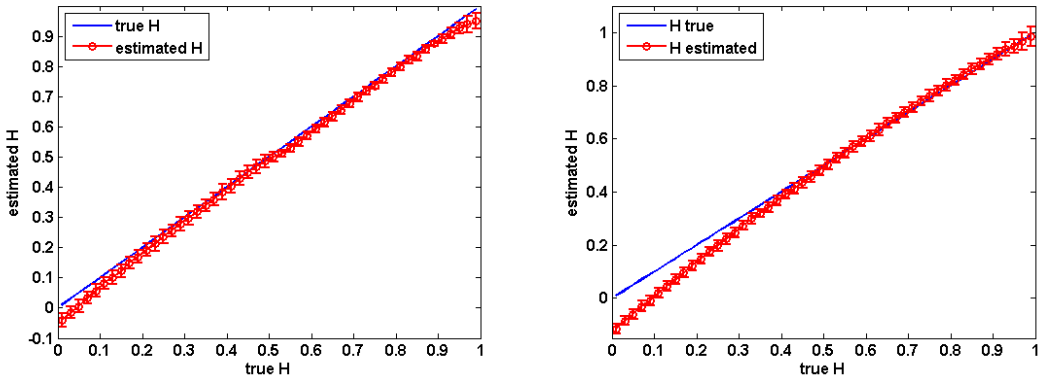
Following, the same scheme as in the CWT-and-periodogram method (see *Wavelet choice* in Subsection 4.2.2), the mean and standard deviation of the H estimates were computed.

Figure 4.5 (b) shows the mean and standard deviation of the H estimates as a function of the wavelet. The estimates for the first wavelet function, *db1*, is, once more, ruled out, because it corresponds to the Haar wavelet function (see Section 4.2.2, in page 43). For the remaining wavelet functions the behavior is quite irregular. However, the tighter estimates are obtained with the wavelet function *db2*; thus, we chose *db2* as wavelet function for the Variance-of-DWT-details method.

4.3 Performance and Robustness Tests

4.3.1 Performance Tests

To test the performance of each estimation method, fBm signals were used with H running from 0.01 to 0.99 with 0.02 steps. For each step, one hundred realizations of the fBm signals



(a) (b)

Figure 4.6: *H estimates (mean and standard deviation) results with Periodogram method (a), and with CWT-and-periodogram method (b).*

were generated, and the H parameter was estimated for each realization with each estimation method. Then, the mean and the standard deviation of the estimations were plotted together with the true H value. This test structure is used in all the following experiments. We consider an estimate as being *tight* if it does not deviate from its true value more than 2%.

Periodogram Method Performance

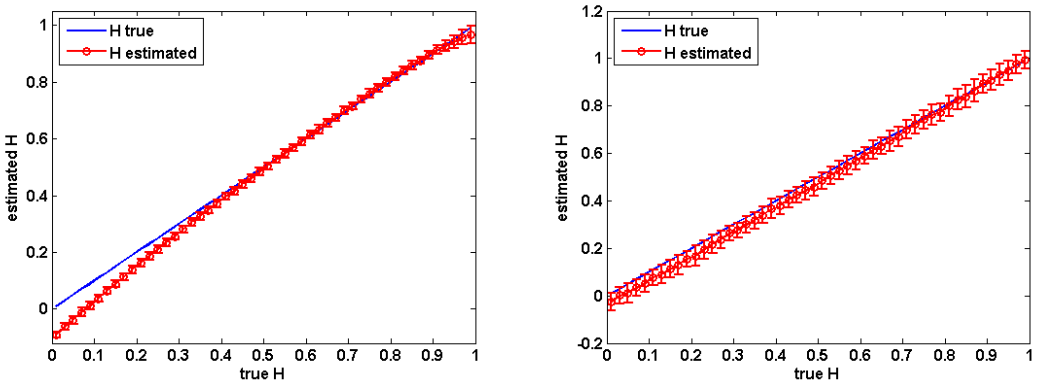
Figure 4.6 (a) shows the mean and standard deviation of the H parameter estimated by Periodogram method.

This method gives tight mean estimates of H for fBm signals with true H between 0.25 and 0.9. The standard deviation is greater for true H values between 0.1 and 0.5 than for the range between 0.5 and 0.9, but can be considered small for the whole range of H values.

CWT-and-Periodogram Method Performance

Figure 4.6 (b) shows the mean and standard deviation of the H parameter estimated with the CWT-and-periodogram method.

This method has two differentiated behaviors with respect to the dependence on the H range. For H values between 0 and 0.5, the global behavior is to underestimate the H parameter, and the smaller the H value the worst the estimates are. For H values between 0.5 and 1, the estimates are very tight for all the values, although with large values for the standard deviation.



(a) (b)

Figure 4.7: *H estimates (mean and standard deviation) results with DWT-and-periodogram method (a), and with Variance-of-DWT-details method (b).*

DWT-and-Periodogram Method Performance

Figure 4.7 (a) shows the mean and standard deviation of the H parameter estimated with the DWT-and-Periodogram method.

The behavior is similar to CWT-and-periodogram method, that is, the method underestimates for true H values below 0.5, whereas it gives very tight estimates with small error for values between 0.5 and 0.9. The method gives slightly worse estimates for true H values near 1.

Variance-of-DWT-Details Method Performance

Figure 4.7 (b) shows the mean and standard deviation of the H parameter estimated by Variance-of-DWT-details method.

It gives systematically good estimates for all true H values, and it is the only method that does not give bad estimates for the extreme values of true H. The method seems to exhibit a small bias, that apparently, is a linear bias that depends on the H value. Following, we try to estimate this bias computationally.

For this purpose, we hypothesize that there is a linear bias in the Variance-of-DWT-details method, described by

$$b_H = H - \hat{H}, \tag{4.4}$$

$$b_H = \gamma H + \delta \tag{4.5}$$

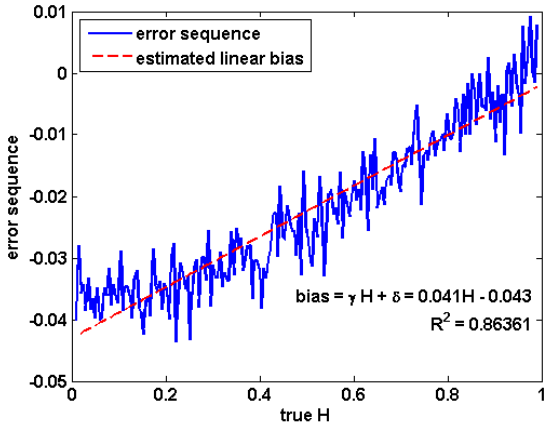


Figure 4.8: *Error sequence and regression line.*

where b_H means that the bias is a function of the H parameter, and $\hat{H}$ is the estimator. In order to obtain a reliable estimation of the bias, first, we repeat the performance test, but this time with H running from 0.01 to 0.99 with 0.0049 steps, which means two hundred H values. Again, we generated one hundred fBm realizations for each H. We applied the Variance-of-DWT-details method for each realization and obtained one hundred estimates for each true H. Next, we obtained the error sequence

$$e(n) = \langle \hat{H}(n) \rangle - H(n), \text{ for } n = 1, \ldots, 200 \tag{4.6}$$

where H is the true H value, and $\langle \hat{H} \rangle$ is the mean of the one hundred estimates for each true H. Finally, we fitted a regression line to the plot of this error sequence vs. true H; from this regression line we obtained the γ and δ parameters of the linear bias (see Figure 4.8). The estimated linear bias is

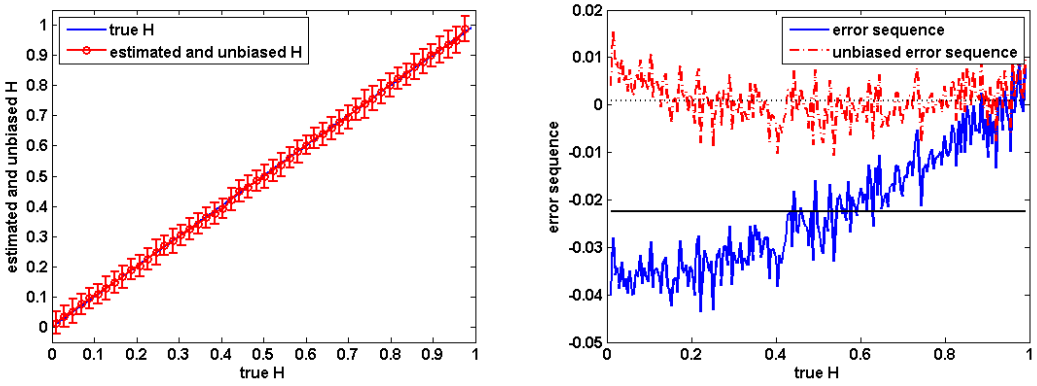
$$b_H = \gamma H + \delta = 0.041H - 0.043 \tag{4.7}$$

The procedure to obtain unbiased H estimates is then the following: First, we obtain the H estimates, $\hat{H}$, using the Variance-of-DWT-details method as was explained previously. Second, we compute the bias for this estimate,

$$b_{\hat{H}} = \gamma \hat{H} + \delta = 0.041\hat{H} - 0.043 \tag{4.8}$$

and finally, we subtract this bias to the original estimate, which gives the unbiased estimate

$$\hat{H}_{unbiased} = \hat{H} - b_{\hat{H}} \tag{4.9}$$



(a) (b)

Figure 4.9: *Unbiased H estimates with Variance-of-DWT-details method (a), and error sequences of biased and unbiased H estimates with their mean values (b).*

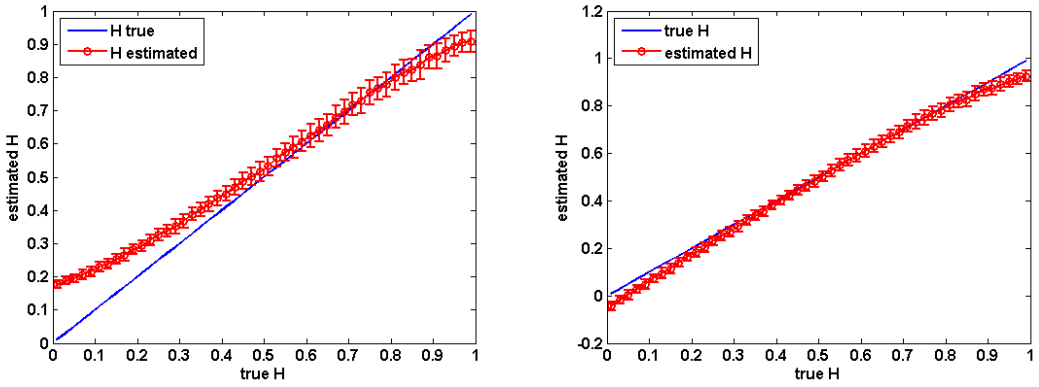
Figure 4.9 (a) shows the mean and standard deviation of the unbiased estimated H, where each estimate was unbiased following the procedure explained previously. Figure 4.9 (b) shows the comparison between the error sequence for biased H estimates (solid line) and the error sequence for unbiased H estimates (dotted line); it also shows the mean of the error sequences for biased and unbiased estimates with solid and dotted lines respectively.

The mean of the error sequence for the unbiased estimate is 9.2415×10^{-4}, whereas the mean error for biased estimates is -0.0225. Therefore, the estimates of the bias allows one to obtain accurate estimates of the H exponent. Moreover, the unbiased estimates of H are very tight for the whole range of true H values, even for extreme values.

R/S Method Performance

Figure 4.10 (a) shows the mean and standard deviation of the H parameter estimates with the R/S method.

This method gives the worst estimates of all methods, and seems to have a non-linear bias. For small values of the true H, the method overestimates, and evolves in a way that underestimates for large true H values. The standard deviation is small for low true H values, but increases as true H does. There are some proposals to obtain a bias-corrected version of the R/S estimator. In [Mielniczuk 07], the authors propose a scheme to obtain an unbiased R/S estimator based on a linear bias, which is estimated empirically. They show that if the bias were



(a) (b)

Figure 4.10: *Performance test for RS method (a), and DFA method (b).*

exactly linear with slope -0.618, a twofold bias correction of the biased estimate $\hat{H}_{R/S}$ would produce an unbiased estimator.

DFA Method Performance

Figure 4.10 (b) shows the mean and standard deviation of the H estimates obtained with the DFA method.

The DFA method gives very tight estimates for true H values between 0.3 and 0.8. For values below 0.3 the method slightly underestimates, and for values above 0.8 the methods also underestimates, but in this case with higher error. Even so, this method gives better results in extreme values of the true H parameter than the methods based on spectral estimates such as the Periodogram method, the CWT-and-periodogram method, and DWT-and-periodogram method.

4.3.2 Robustness against Data Length

To test the robustness of the methods against the data length one hundred realizations of fBm signals were generated for each different values of the H parameter, $H = 0.1, 0.3, 0.5, 0.7, 0.9$, and for ten different values of the data length,

$$\text{Data Length} = 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 50000, 100000$$

The data lengths were chosen in order to match the length of the real data in HRV analysis. that is performed in the next chapter. The recordings used to assess the HRV are obtained from a

24 hour monitorization. Then, if it is considered one beat per second, which means 60 beats per minute, a recording of 24 hours will have, approximately, 24 hours $\times$ 3600 samples/hour = 86400 samples.

Figures 4.11 (a), (b), and (c) show the mean and standard deviation for H estimated by the Periodogram method, the CWT-and-periodogram method, and the DWT-and-periodogram method, respectively. The behavior of the estimates is very similar for the three methods. The mean can be considered constant for all the data length values and for all true H values, whereas, the standard deviation has high values for the two first data lengths, 4000 and 5000, but becomes smaller for high data lengths, from 10000 to 100000. Therefore, these three methods give accurate H estimates for signal length over 10000 samples, although, they give similar estimates, in mean, for all data lengths.

Figure 4.11 (d) shows the mean and standard deviation for H (unbiased) estimated by the Variance-of-DWT-details method. The behavior of the standard deviation is similar to the behavior of the standard deviation for the previous methods, that is, it is high for small data lengths, and becomes smaller for higher data lengths. However, the behavior of the mean is slightly dependent on the data length. The method gives bad results (underestimated) for small data lengths, but as the data length grows, the difference between the H mean estimates and the true H becomes smaller. For data lengths higher than 15000 the H mean estimates become more accurate and more independent on the data length. Therefore, the method needs sequences larger than 15000 samples to obtain tight estimates with small error.

Figure 4.11 (e) shows the mean and standard deviation for H estimated by RS method. This method has a strong dependence on the data length. Particularly, for true H near 1. In general, the method tends to gives small H estimates as the data length increases.

Figure 4.11 (f) shows the mean and the standard deviation for H estimated by the DFA method. This method has opposed behaviors in the range of true H values from 0 to 0.5 and from 0.5 to 1. For the first range the method tends to give higher H estimates as the data length increases, which cause that the estimates to come closer to the true H value for large data lengths. Whereas, for the second range the method tends to gives smaller H estimates as the data length increases, which causes that the estimates diverge from the true H value. The behavior for true H near 0.5 is constant, thus independent of the data length.

The strong dependence on data length of the two last methods, RS and DFA method, can be due to the algorithm characteristics. Both methods estimate the H exponent by splitting the

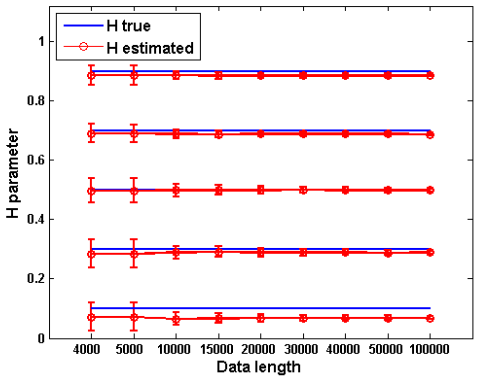
sequence into equal segment lengths for different scales and obtaining some quantity (the R/S rate in RS method, and the root mean square fluctuation in DFA method). Ideally this value increases as a function of the scale by following a power-law which exponent is related to the H exponent. In the practice, that is, with finite length sequences, the power law is revealed only in some scales, and the number of the scales is dependent on the data length.

4.4 Conclusions

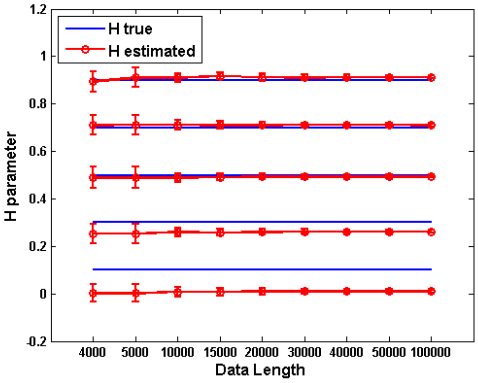
In this chapter, we developed some experiments in order to, first, tune the free parameters of the estimating methods of H parameter, and second, to study the performance and the robustness against data length of the methods.

From the performed experiments, we found that the spectral-based methods, namely Periodogram method, CWT-and-periodogram method, and DWT-and-periodogram method, gave bad estimates for extreme values of true H, i.e. near 0 and near 1, whereas they gave tight estimates for central values. The Variance-of-DWT-details method gives good estimates for all tested range of values, but it has a linear bias which depends on the true H exponent. In order to avoid this shortcoming, we estimated the linear bias experimentally. The unbiased estimates with this method were the most accurate ones. The RS method was the worst of all methods; indeed, it has non-linear bias. There exist in the literature some attempts to obtain an unbiased RS method, but we prefer to discard this unbiased versions (Ockham's razor) due to the fact that the other studied methods outperform the R/S method, and moreover, we do not know exactly how this unbiased estimators will behave in real, noisy, situations. Finally, the DFA method gave tight estimates for almost all true H values, and with small error.

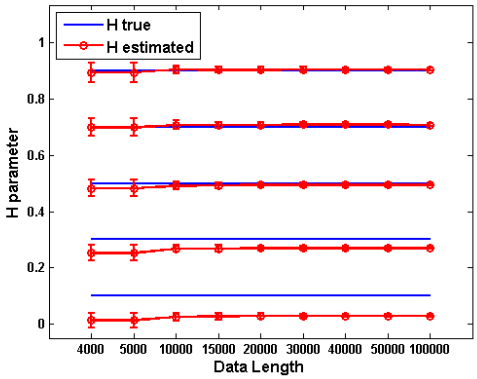
All spectral-based methods are very robust against data length, giving tight estimates for very short sequences, over 5000 samples. The Variance-of-DWT-details method is most dependent on data length, and needs more than 15000 samples to obtain tight estimates. The RS method is very dependent on the data length, whereas the DFA method is slightly dependent on the data length.



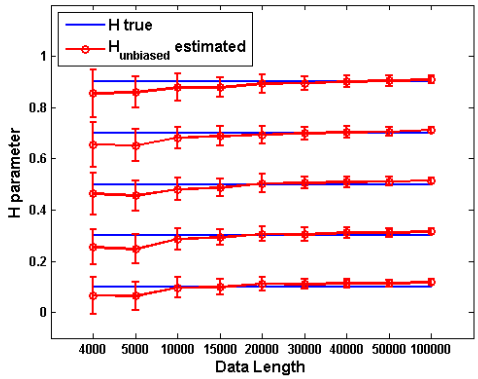
(a) *Periodogram method.*



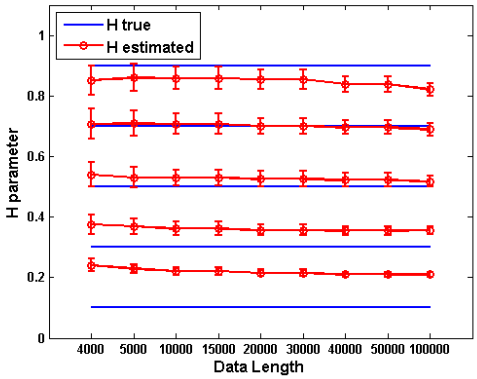
(b) *CWT-and-periodogram method.*



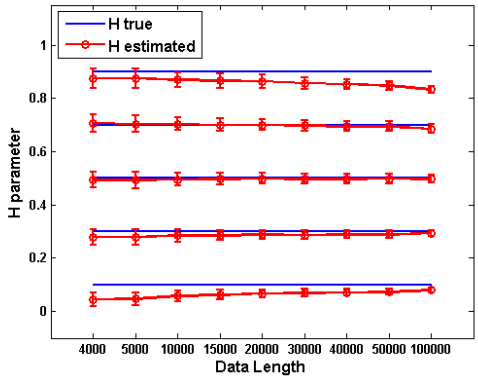
(c) *DWT-and-Periodogram method.*



(d) *Variance-of-DWT-details method.*



(e) *RS method.*



(f) *DFA method.*

Figure 4.11: *Robustness against data length tests for all estimating methods.*

Chapter 5

EXPERIMENTS USING REAL DATA

If you are out to describe the truth, leave elegance to the tailor.
(Albert Einstein)

The aim of this chapter is to present some experiments on real HRV signal applying the estimation methods introduced in previous chapters. The main objectives are two: first, characterize the fractal correlation properties for healthy and pathological subjects (with Congestive Heart Failure, CHF), in order to study the properties of H estimates in discriminating both states. Second, to analyse the dependence of the fractal correlations properties on aging, in other words, to asses the loss of HRV due to aging, using the H parameter.

5.1 Data Sets

We apply the methods of fractal characterization to real RR-intervals time series, which are derived from 24-hour Holter recordings, named after its inventor Norman Holter. The Holter recording is an ambulatory ECG recording technique in which, during a day of normal daily activities, the patient carries a solid state recording device that stores the ECG [Sörnmo 05]. The Holter recordings were obtained from healthy subjects and subjects with CHF, which is a condition that can result from any structural or functional cardiac disorder that impairs the ability of the heart to fill with or pump a sufficient amount of blood through the body [Wikipedia 08]. Both data sets were obtained from Physionet [Goldberger 00].

The data of the healthy group were obtained from 24-hour Holter monitor recordings of

72 healthy subjects, 35 men and 37 women aged from 20 to 76 years old. The data of the CHF group were obtained from 24-hour Holter monitor recordings of 44 subjects, from 22 to 79 years old, including 19 men and 6 women (the gender information was not available for all the recordings). Some of the recordings from both groups have time information, which was used to study the variation of HRV during different periods of the day (18 healthy subjects aged from 20 to 50 years old and 15 subjects of the CHF group aged from 22 to 71 years old).

All datasets were preprocessed in order to exclude artifacts, missed detections and isolated ectopic beats, i.e. non-normal beats. Furthermore, RR intervals lower than 200 ms and greater than 2000 ms were excluded, and also those which differed more than 20% from the previous and the subsequent RR intervals [Malik 89a].

5.2 Real HRV Signals and Hurst Exponent Estimation Methods

In the application of the estimation methods, presented in the Chapter 3, to real HRV signals, it is necessary to take into account some peculiarities of these signals.

First, the $1/f$ spectral behavior of the HRV signal is only exhibited at low frequencies, below 0.02 Hz. Therefore, the methods based in spectral analysis (i.e., the Periodogram method, the CWT-and-periodogram method and the DWT-and-periodogram method) must produce the H estimates only for a low frequency band. We use the frequency band most widely used in the literature [Bigger 96], that is, between 10^{-4} and 10^{-2} Hz. Figure 5.1 shows an example of the H exponent estimated by the Periodogram method in a real HRV signal from a healthy subject.

This frequency band constraint makes impossible the use of the CWT-and-periodogram method, because, this method does not produce a reliable spectral reconstruction in this frequency band (see Figure 5.2); therefore, we discard this method for the following experiments with real HRV signals.

Second, there are claims in the literature stating that the HRV is, at least, bi-fractal, that is, one needs two different scaling exponents in order to adequately characterize the fractal properties of the HRV. In [Peng 95], the authors propose using of two scaling exponents for the DFA method to characterize the fractal correlation properties of the HRV, one for the short-term, α_1, and the second one for long-term, α_2. Practically, α_1 is estimated by fitting a regression line to log $F(s)$ vs. log n for $4 \leq \mathrm{n} \leq 16$ (i.e less than 16 beats), whereas, α_2 is similarly obtained

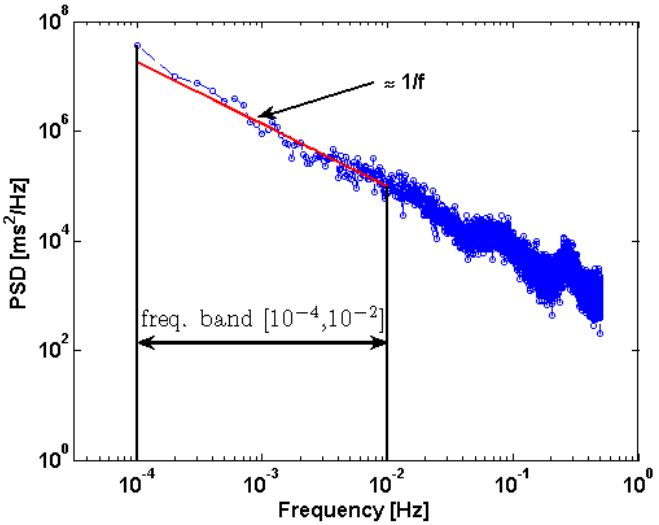


Figure 5.1: $1/f$ *spectral behavior in a real HRV signal, this spectral behavior only holds for frequencies between* 10^{-4} *and* 10^{-2} *Hz.*

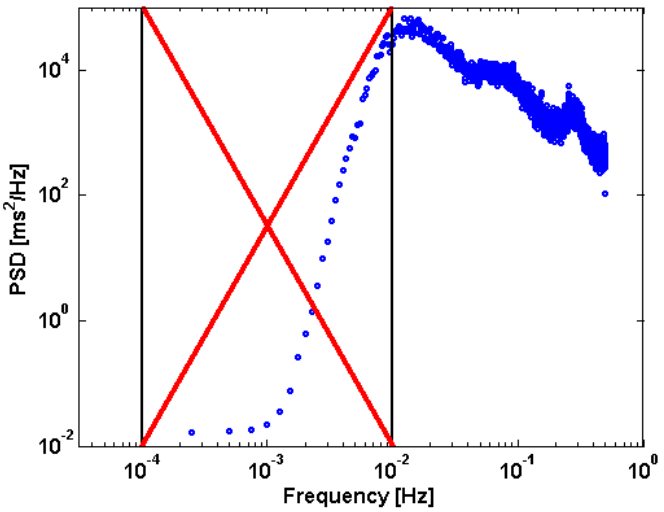


Figure 5.2: *The CWT-and-periodogram method does not produce a reliable spectral reconstruction in the frequency band where the HRV signal shows scaling behavior.*

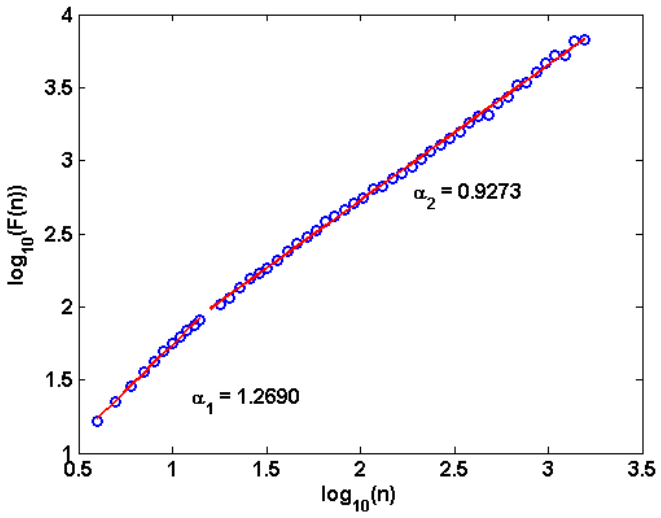


Figure 5.3: *Bi-scaling behavior of HRV in a healthy subject. Two different scaling exponents are necessary to characterize the fractal correlation properties.*

for $16 \leq \mathrm{n} \leq \mathrm{N}/4$. For very large scales, $s > N/4$, $F(s)$ becomes statistically unreliable because the number of segments N_s for the averaging procedure becomes very small [Kantelhardt 02]. Figure 5.3 shows this bi-scaling behavior in a HRV signal example from a healthy subject. There are also some works that show that the HRV signal is, not only bi-fractal, but multi-fractal, that is, one needs a set of scaling exponents (more than two) to characterize the fractal properties of the HRV signal [Ivanov 99].

5.3 Discriminating Healthy and CHF subjects

In this section, the capabilities of the H exponent to distinguish between healthy and CHF subjects are analyzed. The H exponent is estimated by the four methods selected in the previous section, namely: The Periodogram method, the DWT-and-peridogram method, the Variance-of-DWT-details method, and the DFA method. Also, the DFA scaling exponents, α_1 and α_2, are studied.

5.3.1 Healthy vs. CHF Subjects in 24 Hours

Firstly, the H estimates are obtained over the entire HRV signal, i.e. over 24 hours, and the values obtained in healthy and CHF subjects are compared. Table 5.1 shows the mean and

standard deviation of the H parameter estimated by all the methods, and the scaling exponents α_1 and α_2 from DFA, for both healthy and CHF subjects. Significant difference between healthy and CHF groups ($p < 0.05$) has been highlighted. The Student's t-test is the statistic test used to compare both groups, and this is the test used for all the experiments in this work except otherwise stated

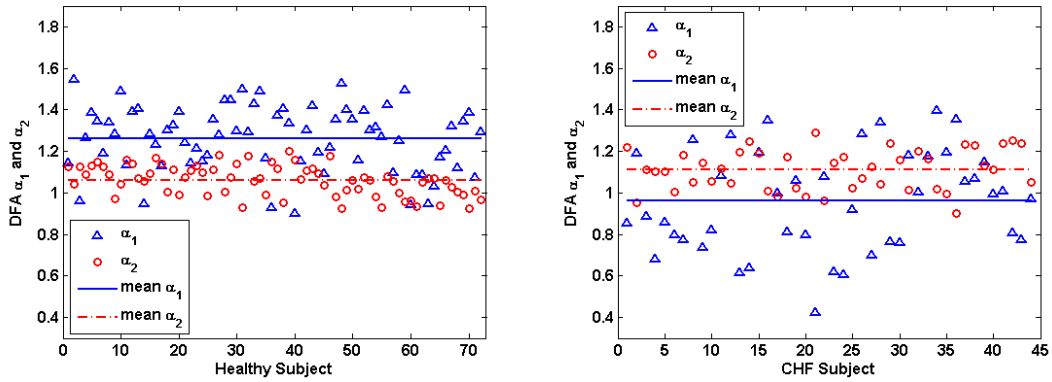
Figure 5.5 shows the notched boxplots for all the estimation methods. The boxes have lines at the lower quartile, median, and upper quartile values. Whiskers extend from each end of the box to 1.5 times the interquartile range. Outliers are displayed with a + sign. Notches display the variability of the median between samples.

The H parameter estimated by all the methods, except for the Periodogram method, shows significant difference ($p < 0.05$) between healthy and CHF subjects. The change in the H exponent for CHF subjects, whose values are higher than for healthy subjects, agrees with the hypothesis of that complexity loss due to CHF, reflected as a breakdown in the fractal correlation properties [Berger 01, Goldberger 99], that is, a departure from $1/f$ structure $H \approx 0$ and an approach to a Brownian motion $H \underset{\text{CHF}}{\longrightarrow} 0.5$, which means that consecutive increments of the signal are uncorrelated. The behavior of DFA scaling exponent α_1 is different; indeed, it gives higher values for healthy subjects than for CHF subjects, moreover, it gives the best results to distinguish between healthy and CFH subjects.

From the results, we can also conclude that the methods based on a spectral estimation produce worse results than the other methods. This can be due to the fact that they are less suitable for the analysis of scaling properties in highly non-stationary signals, as the HRV signal, than the wavelet-based methods or the DFA-based methods.

	Healthy	CHF	p-value
H_Period	0.13 ± 0.08	0.15 ± 0.12	0.2504
H_DWTPeriod	$\mathbf{0.12 \pm 0.08}$	$\mathbf{0.17 \pm 0.13}$	**0.0282**
H_VarDWT	$\mathbf{0.13 \pm 0.06}$	$\mathbf{0.18 \pm 0.07}$	$\mathbf{1.54 \times 10^{-4}}$
H_DFA	$\mathbf{0.09 \pm 0.06}$	$\mathbf{0.13 \pm 0.06}$	**0.0058**
α_1 DFA	$\mathbf{1.26 \pm 0.16}$	$\mathbf{0.96 \pm 0.24}$	$\mathbf{7.27 \times 10^{-13}}$
α_2 DFA	$\mathbf{1.06 \pm 0.07}$	$\mathbf{1.11 \pm 0.1}$	**0.0014**

Table 5.1: *Mean ± standard deviation of H parameter estimated by all methods for healthy and CHF subjects. Significant difference between healthy and CHF groups ($p < 0.05$) has been highlighted.*



(a) (b)

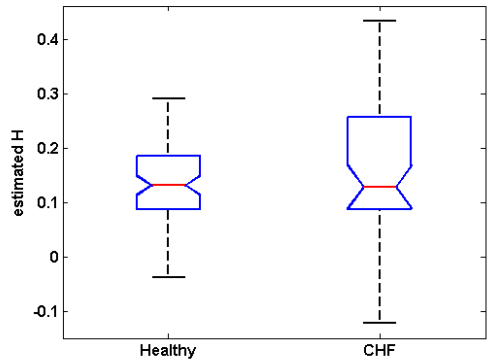
Figure 5.4: *DFA scaling exponents α_1 ($\triangle$) and α_2 ($\circ$) for each healthy (a) and CHF (b) subject. The means are also plotted. These plots show the cross-over phenomenon due to CHF.*

An important fact to be pointed out, is that there is a cross-over phenomenon in DFA α_1 and α_2 values for healthy and CHF subjects. The scaling exponent α_1 shows higher values than α_2 for healthy subjects, whereas the scaling exponent α_2 shows higher values than α_1 for CHF subjects. This result agrees with the results reported in the literature [Peng 95]. Figure 5.4 (a) shows the DFA α_1 scaling exponent, represented as $\triangle$ symbol, and DFA α_2 scaling exponent, represented as $\circ$, for each healthy subject. Also the mean for α_1, solid line, and for α_2, dash-dot line, are plotted. Whereas Figure 5.4 (b) is the equivalent representation but CHF subjects.

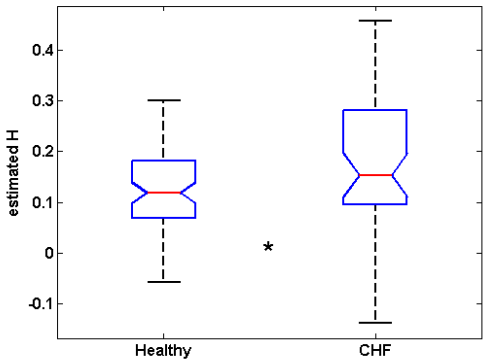
The cross-over phenomenon is a general behavior, i.e., it holds in average, but it does not hold for each and every subject; in fact, for healthy subjects α_1 is higher than α_2 in 63 cases of 72, whereas for CHF subjects α_2 is higher than α_1 in 28 cases of 44. As the results suggest, the cross-over phenomenon is mainly due to the change in the α_1 exponent, whereas the change in α_2 is more subtle. Therefore, the fractal correlations properties characterized by α_1 are the most affected by the CHF. The main drawback is that there is yet no clear explanation for the physiological behavior underlying this phenomenon, so, it is difficult to extract deeper implications.

5.3.2 Healthy vs. CHF Subjects in Different Time Periods

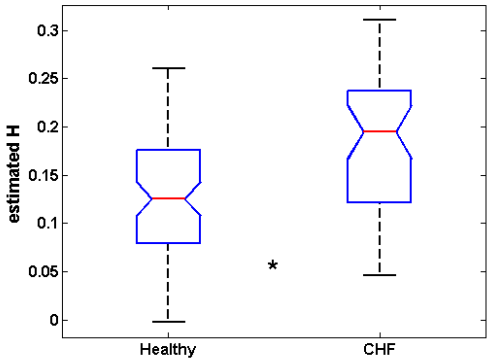
The behavior of the different H estimates with regard for the time period of the day are studied in this subsection. Only the Holter recordings with available time information are used



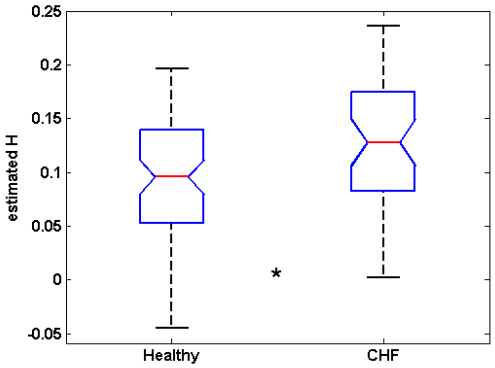
(a) Periodogram method



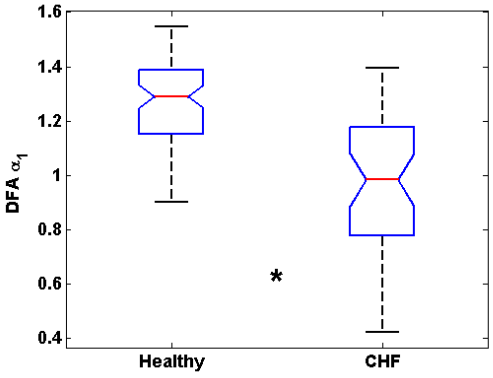
(b) DWT-and-periodogram method



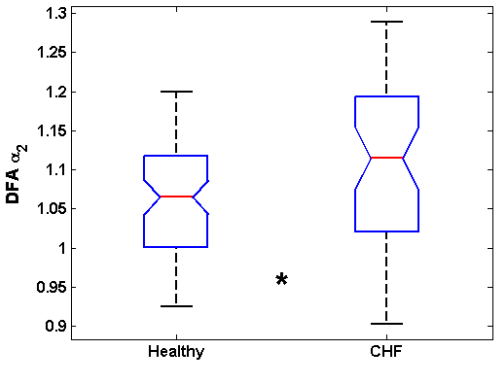
(c) Variance-of-DWT-details method



(d) Hurst by DFA method



(e) DFA α_1



(f) DFA α_2

Figure 5.5: *Notched box plots for the estimation methods, Periodogram method (a), DWT-and-periodogram method (b), Variance-of-DWT-details method (c), DFA method (d), and for the scaling exponents from DFA, α_1 (e), and α_2 (f). Symbol $*$ means significant difference between healthy and CFH subjects ($p < 0.05$).*

		Day	Night	p-value
Healthy	**H_Period**	$\mathbf{0.13 \pm 0.08}$	$\mathbf{-0.11 \pm 0.13}$	$\mathbf{1.67 \times 10^{-7}}$
	H_DWTPeriod	$\mathbf{0.15 \pm 0.09}$	$\mathbf{-0.09 \pm 0.13}$	$\mathbf{2.29 \times 10^{-7}}$
	H_VarDWT	$\mathbf{0.09 \pm 0.05}$	$\mathbf{-0.01 \pm 0.05}$	$\mathbf{1.08 \times 10^{-6}}$
	H_DFA	$\mathbf{0.06 \pm 0.04}$	$\mathbf{-0.03 \pm 0.07}$	$\mathbf{4.97 \times 10^{-5}}$
	α_1 DFA	$\mathbf{1.25 \pm 0.12}$	$\mathbf{1.12 \pm 0.24}$	**0.0472**
	α_2 DFA	$\mathbf{1.03 \pm 0.06}$	$\mathbf{0.94 \pm 0.08}$	$\mathbf{7.05 \times 10^{-4}}$
CHF	**H_Period**	$\mathbf{0.19 \pm 0.13}$	$\mathbf{0.00 \pm 0.16}$	**0.0016**
	H_DWTPeriod	$\mathbf{0.21 \pm 0.14}$	$\mathbf{0.03 \pm 0.20}$	**0.0057**
	H_VarDWT	$\mathbf{0.20 \pm 0.07}$	$\mathbf{0.07 \pm 0.09}$	$\mathbf{1.39 \times 10^{-4}}$
	H_DFA	$\mathbf{0.15 \pm 0.07}$	$\mathbf{0.02 \pm 0.07}$	$\mathbf{4.72 \times 10^{-5}}$
	α_1 DFA	0.88 ± 0.26	0.97 ± 0.20	0.3378
	α_2 DFA	$\mathbf{1.15 \pm 0.10}$	$\mathbf{1.00 \pm 0.08}$	$\mathbf{5.63 \times 10^{-5}}$

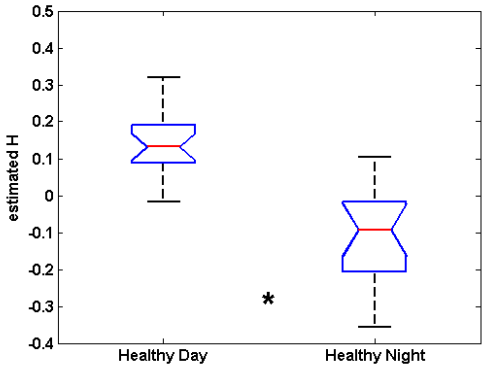
Table 5.2: *Mean ± standard deviation of H parameter estimated by all methods for day and night periods, and for healthy and CHF subjects. Significant difference between healthy and CHF groups ($p < 0.05$) has been highlighted.*

here.

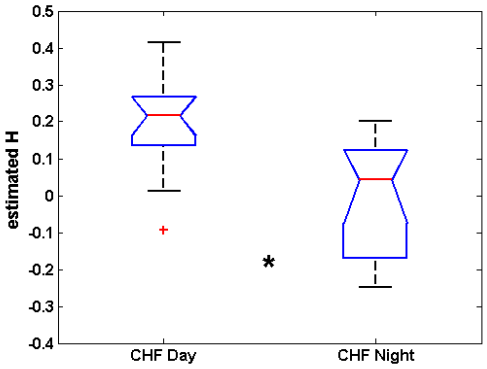
First, day and night time periods are compared, where the day period comprises the interval time between 06 : 00 and 23 : 59, whereas the night period comprises the interval time between 00 : 00 to 05 : 59 for healthy and CHF groups separately. Table 5.2 shows the mean and standard deviation of the H estimates for healthy subjects, and also shows the mean and standard deviation of the H estimates for CHF subjects, for the day and night periods. Figure 5.6 and Figure 5.7 show the distribution of the H estimates for each estimation method and for DFA scaling exponents α_1 and α_2, comparing day and night periods for both healthy and CHF subjects. The Lilliefors test is previously used to assess the normality of H estimate distribution, with a level of significance $\alpha = 0.05$, since the number of the recordings is not large enough to assume normal distribution.

The results suggest that the H parameter changes significantly between day and night for both healthy and CHF subjects. The only exception is the DFA scaling exponent α_1 which for CHF subjects does not present a significant variation. Also, α_1 in healthy subjects has a p-value ($p = 0.0472$) which is close to the chosen level of the significance ($\alpha = 0.05$).

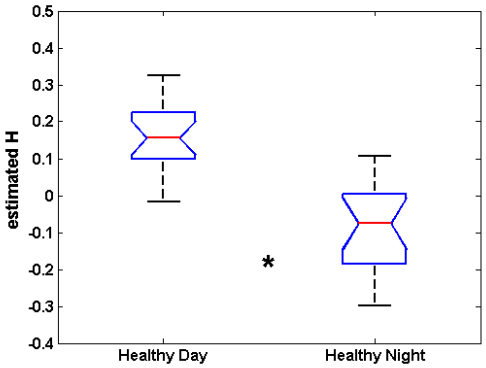
The H parameter and scaling exponents show higher values in day period than in night period for all the estimation methods that give significant differences. The reason can be explained as



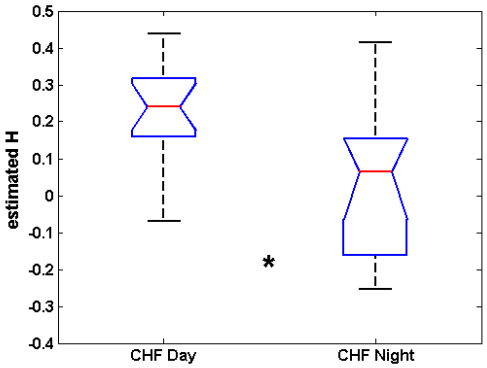
(a) Periodogram method Healthy



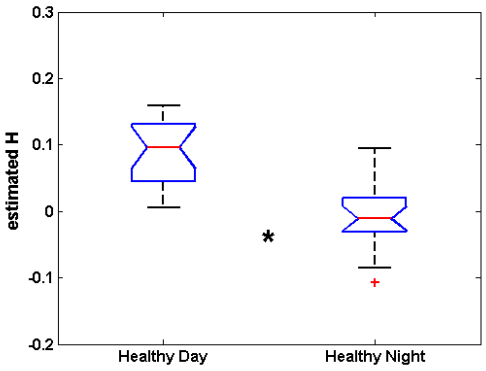
(b) Periodogram method CHF method



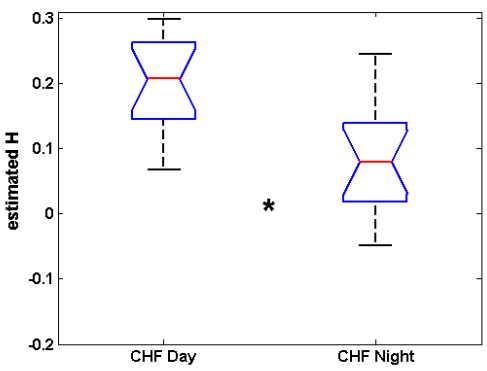
(c) DWT-and-periodogram method Healthy



(d) DWT-and-periodogram method CHF

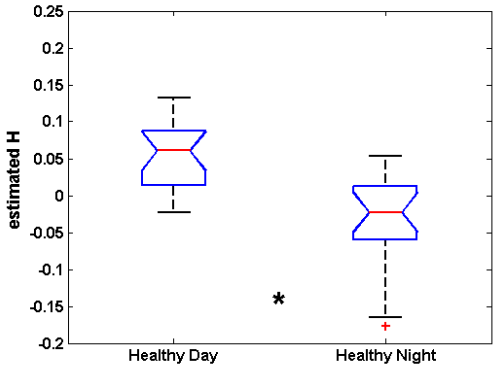


(e) Variance-of-DWT-details method Healthy

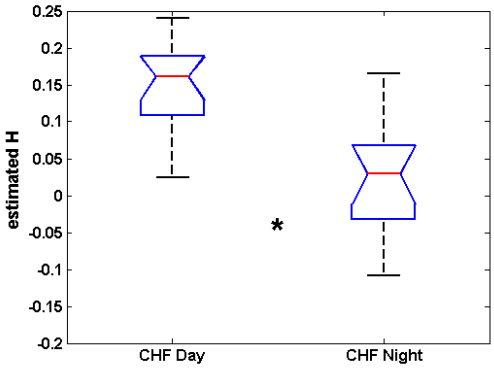


(f) Variance-of-DWT-details method CHF

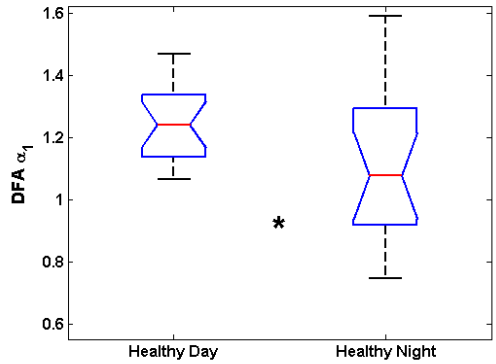
Figure 5.6: *Notched box plots for the H parameter estimated in a day and night periods by the following methods, Periodogram method healthy subjects (a), Periodogram method CHF subjects (b), DWT-and-periodogram method for healthy subjects (c), DWT-and-periodogram method for CHF subjects (d), Variance-of-DWT-details for healthy subjects (e), and Variance-of-DWT-details for CHF subjects (f). The $*$ symbol means significant difference between day and night period ($p < 0.05$).*



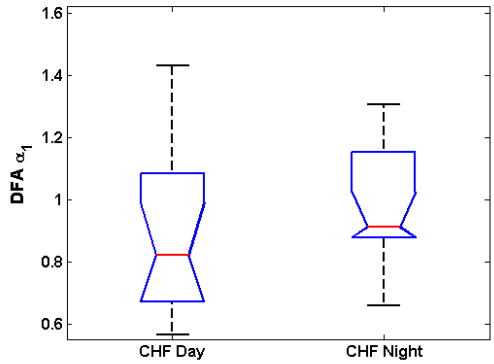
(a) Hurst by DFA method Healthy



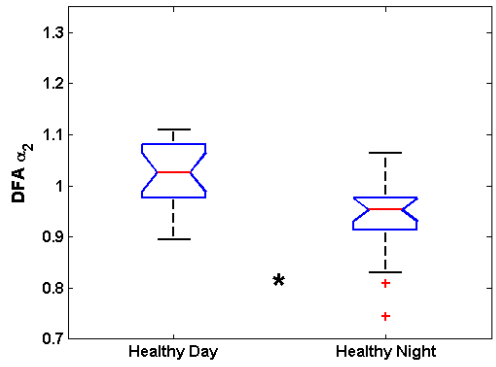
(b) Hurst by DFA method method



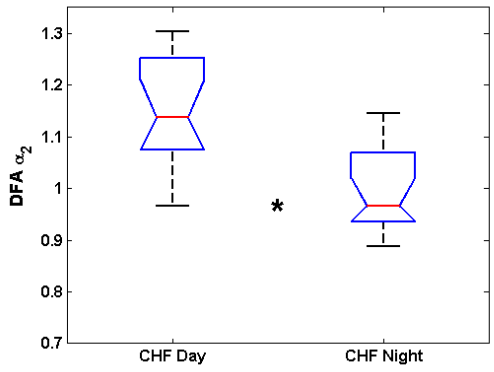
(c) DFA α_1 Healthy



(d) DFA α_1 CHF



(e) DFA α_2 Healthy



(f) DFA α_2 CHF

Figure 5.7: *Notched boxplots for H parameter estimated in a day and night periods by the following methods, Hurst by DFA method healthy subjects (a), Hurst by DFA method CHF subjects (b), DFA α_1 healthy subjects (c), DFA α_1 method for CHF subjects (d), DFA α_2 healthy subjects (e), and DFA α_2 for CHF subjects (f). Symbol $*$ means significant difference between day and night period ($p < 0.05$).*

follows: it is possible to think that the HRV is composed by fluctuations that arise from a complex nonlinear dynamical system, and other fluctuations which are the response to changes in the environmental conditions superimpose and uncorrelated with the former ones. These two classes of fluctuations may be shown to have very different correlation properties. The fluctuations arising from the dynamics of the complex, multiple-component system should exhibit long-term correlations, whereas the other fluctuations can have characteristic time scales, i.e. frequencies related to the stimuli [Peng 95]. Obviously the fluctuations due to the environmental stimuli are more present in the day period than in the night period, when the fluctuations due only to the complex dynamical behavior are outstanding.

The DFA method takes advantage of this hypothesis, and interpret the fluctuations relating to the environmental stimuli as a *noise* that can be treated as a *trend* and distinguished from the more subtle fluctuations (i.e. **D**etrended) that may reveal intrinsic correlation properties of the dynamics (i.e. **F**luctuation **A**nalysis).

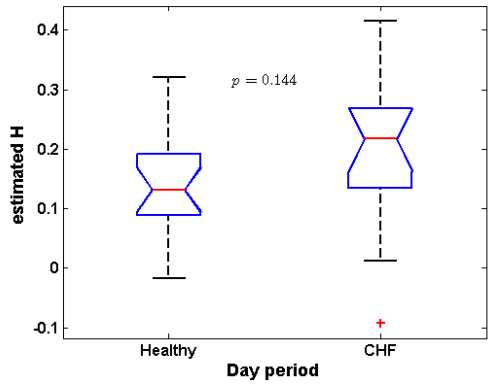
The next step to study the behavior of the H parameter with regard to the period of the day is to compare the H estimates for healthy and CHF subjects, for both day and night periods. Figure 5.8 and Figure 5.9 show the distribution from each estimation method by means of notched boxplots accompanied by its p-values.

The comparison in day periods between healthy and CHF subjects shows that Periodogram and DWT-and-peridogram estimation methods do not give significant differences between healthy and CHF subjects, whereas the other methods, namely, Variance-of-DWT-details, Hurst by DFA, and DFA scaling exponents α_1 and α_2, show significant differences for an $\alpha = 0.05$ level of significance.

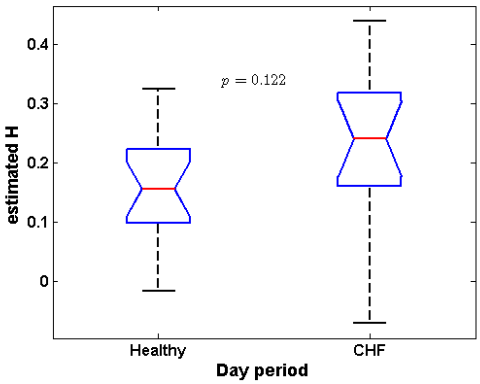
For the night period, the DWT-and-periodogram method and the DFA scaling exponent α_1 are the only ones that do not show significant difference between healthy and CHF subjects for an $\alpha = 0.05$ level of significance. As a general conclusion, the H estimates and scaling exponents that show significant difference, give the larger separation between healthy and CHF subjects for the day period. For the night period this separation is more subtle.

5.4 Aging Curve

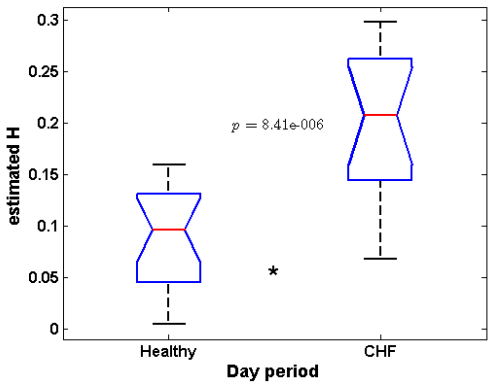
The aim of this section is to study the changes on the fractal correlation properties, quantified by the H parameter, with aging. A loss of complexity in the physiological regulation of elderly



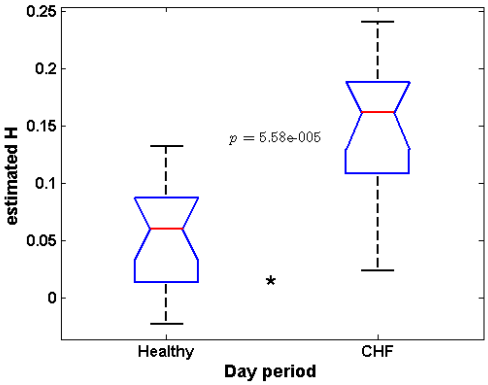
(a) Periodogram method Day



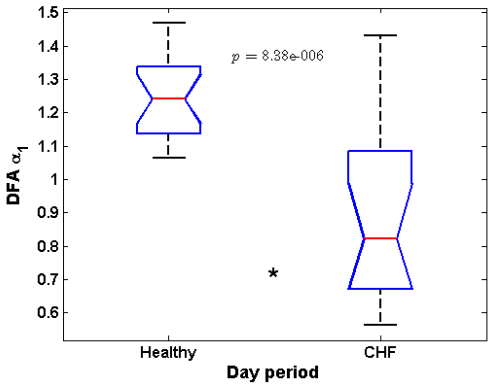
(b) DWT-and-Periodogram method Day



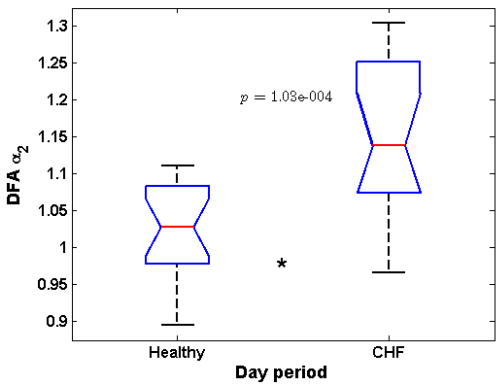
(c) Variance-of-DWT-details method Day



(d) Hurst by DFA method Day

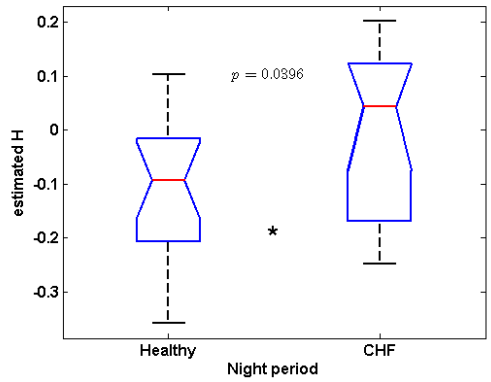


(e) DFA α_1 Day

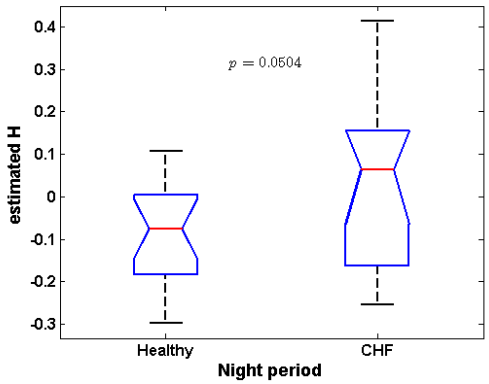


(f) DFA α_2 Day

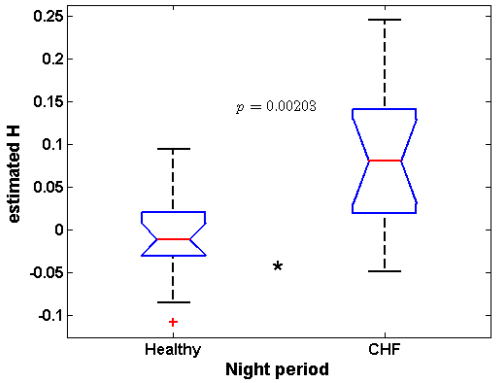
Figure 5.8: *Notched box plots for the H parameter estimated in the day period, Periodogram method (a), DWT-and-periodogram method (b), Variance-of-DWT-details method (c), Hurst by DFA method (d), DFA α_1 (e), and DFA α_2 (f). The $*$ symbol means significant difference between healthy and CHF subjects in ($p < 0.05$).*



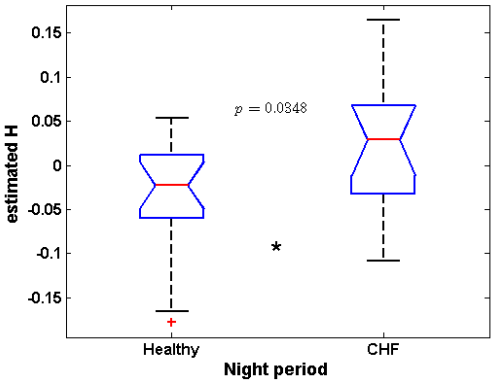
(a) Periodogram method Night



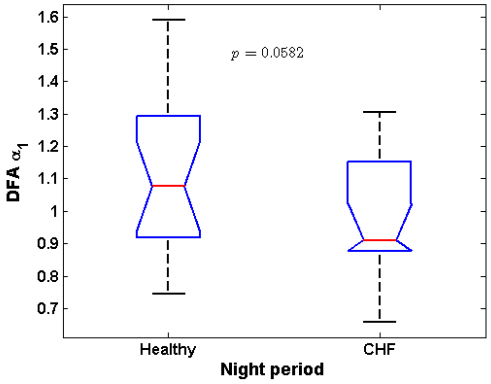
(b) DWT-and-Periodogram method Night



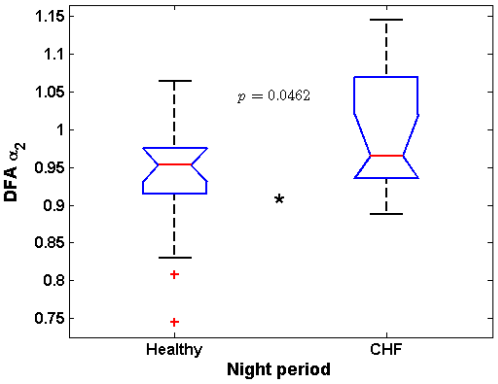
(c) Variance-of-DWT-details method Night



(d) Hurst by DFA method Night



(e) DFA α_1 Night



(f) DFA α_2 Night

Figure 5.9: *Notched box plots for the H parameter estimated in the night period, Periodogram method (a), DWT-and-periodogram method (b), Variance-of-DWT-details method (c), Hurst by DFA method (d), DFA α_1 (e), and DFA α_2 (f). The $*$ symbol means significant difference between healthy and CHF subjects in ($p < 0.05$).*

subjects has been reported in the literature [Kaplan 91, Goldberger 99]. Under the hypothesis that this complexity loss has an influence on the H parameter, the relationship between the changes on this parameter and the aging is studied.

First, the discrimination capabilities of the H exponent to distinguish between young and elderly subjects are studied. In order to do so, the data bases are split into two groups: young group, which is made up of subjects between 20 and 50 years old, and elderly group, which is made up of subjects between 51 and 80 years old. The aim of this first analysis is to elucidate whether the change in the H exponent with aging is significant or not, i.e. whether the loss of complexity due to aging is reflected as significant variation of the fractal correlation properties quantified by H parameter.

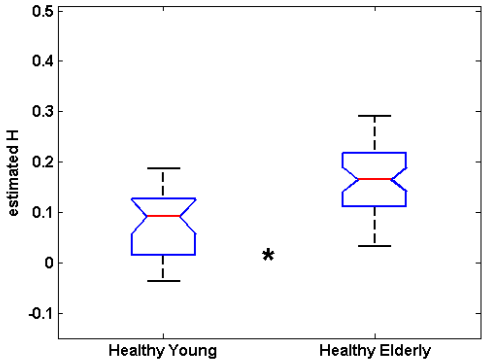
Additionally, we study the dependence of H parameter on age for each healthy subject and CHF subject by using linear regression of the indices vs. age, and then obtaining the slope (variation per year index). The aim of the second analysis is to discover how is this change in the correlations properties of the HRV signal, i.e. to determine how the change in the H parameter (if it occurs) evolves with aging, for both healthy and CHF subjects.

5.4.1 Young Group vs. Elderly Group

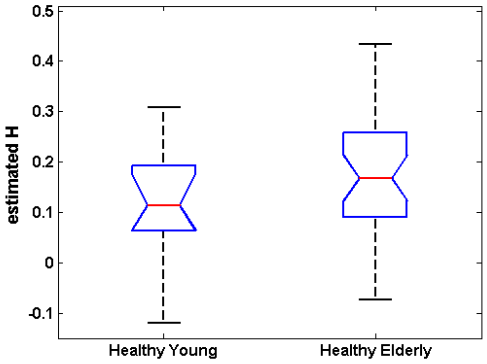
Table 5.3 shows the mean and standard deviation of the H estimates for young and elderly groups for both healthy and CHF subjects.

Graphical summaries of the distributions from each estimation method comparing young and elderly groups for both healthy and CHF subjects, are depicted in Figure 5.10 and Figure 5.11, which shows the corresponding notched boxplots. Once again, Lillefors test is used to assess normality of distributions with level of significance $\alpha = 0.05$.

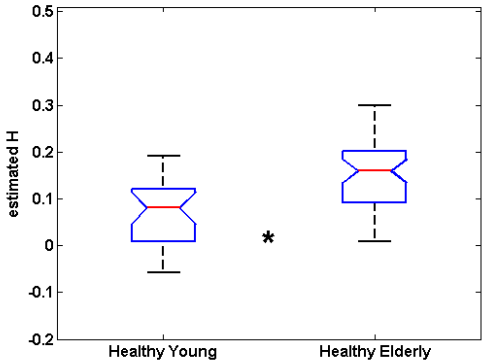
From the results, we can conclude that, for healthy subjects, the H parameter estimated by all methods, and the DFA scaling exponent α_2, show significant differences between young and elderly groups, whereas the DFA scaling exponent α_1 does not show it. Therefore, the fractal correlation properties quantified by α_1 does not change significantly with aging. For the CHF subjects, nor the H parameter, estimated by any method, nor the α_1 and α_2 scaling exponents, show significant differences between young and elderly groups. Therefore, the damage caused by the CHF induce a breakdown of the fractal correlation properties of the HRV for both young and the elderly subjects.



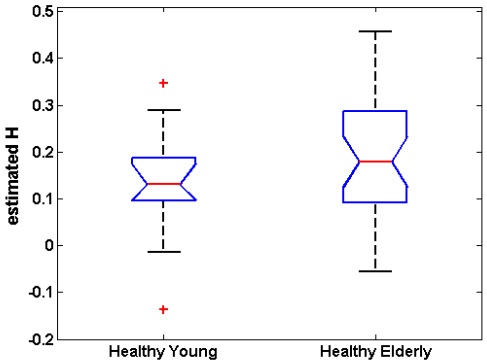
(a) Periodogram method Healthy



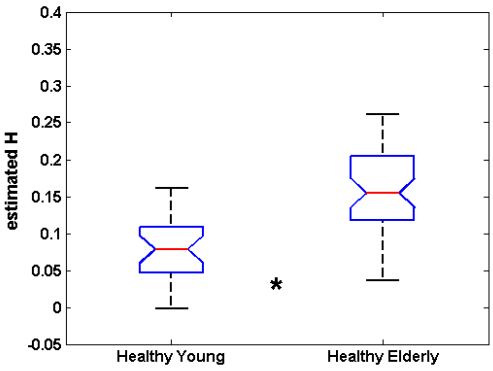
(b) Periodogram method CHF method



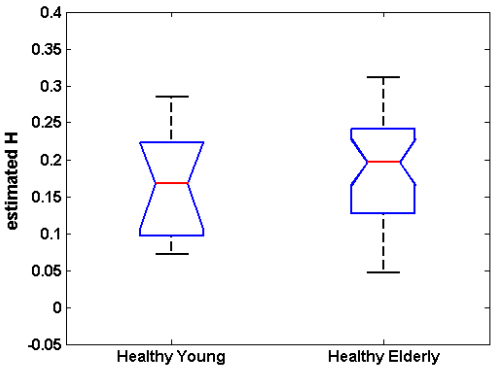
(c) DWT and periodogram method Healthy



(d) DWT and periodogram method CHF

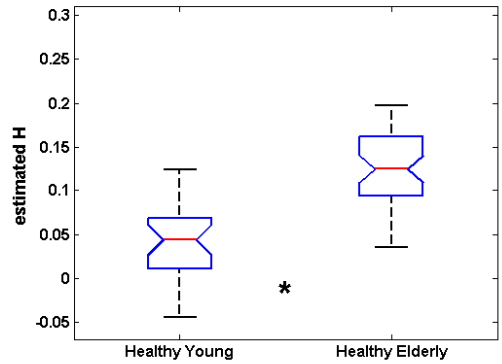


(e) Variance-of-DWT-details method Healthy

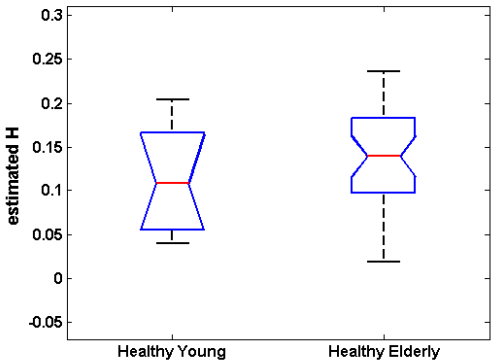


(f) Variance-of-DWT-details method CHF

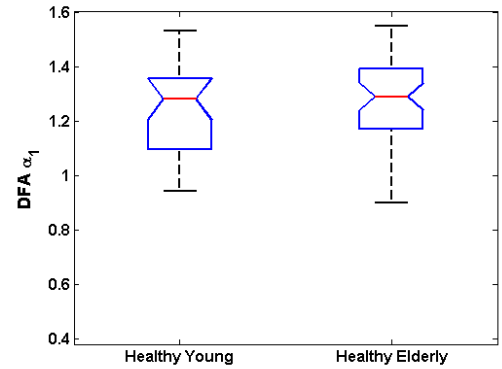
Figure 5.10: *Notched box plots for H parameter estimated for young and elderly groups, Periodogram method healthy subjects (a), Periodogram method CHF subjects (b), DWT and periodogram method for healthy subjects (c), DWT and periodogram method for CHF subjects (d), Variance-of-DWT-details for healthy subjects (e), and Variance-of-DWT-details for CHF subjects (f). Symbol $*$ means significant difference between day and night period ($p < 0.05$).*



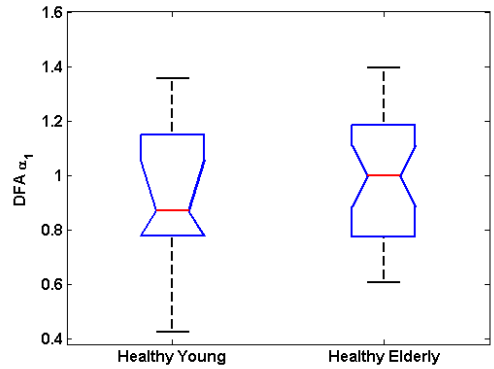
(a) Hurst by DFA method Healthy



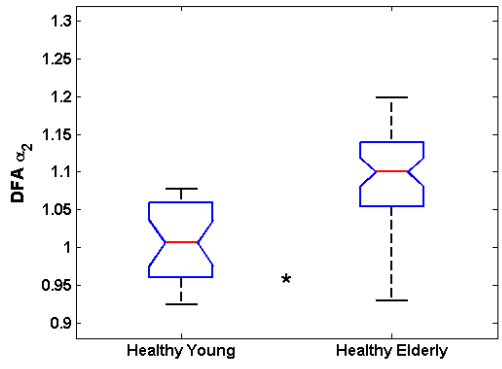
(b) Hurst by DFA method method



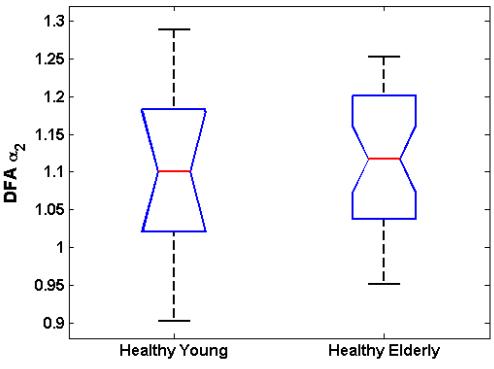
(c) DFA α_1 Healthy



(d) DFA α_1 CHF



(e) DFA α_2 Healthy



(f) DFA α_2 CHF

Figure 5.11: *Notched box plots for H parameter estimated for young and elderly groups, Hurst by DFA method healthy subjects (a), Hurst by DFA method CHF subjects (b), DFA α_1 healthy subjects (c), DFA α_1 method for CHF subjects (d), DFA α_2 healthy subjects (e), and DFA α_2 for CHF subjects (f). Symbol $*$ means significant difference between day and night period ($p < 0.05$).*

		Young	Elderly	p-value
Healthy	**H_Period**	$\mathbf{0.08 \pm 0.06}$	$\mathbf{0.16 \pm 0.07}$	$\mathbf{2.37 \times 10^{-6}}$
	H_DWTPeriod	$\mathbf{0.07 \pm 0.07}$	$\mathbf{0.15 \pm 0.07}$	$\mathbf{1.40 \times 10^{-5}}$
	H_VarDWT	$\mathbf{0.08 \pm 0.04}$	$\mathbf{0.16 \pm 0.06}$	$\mathbf{4.20 \times 10^{-8}}$
	H_DFA	$\mathbf{0.04 \pm 0.04}$	$\mathbf{0.12 \pm 0.05}$	$\mathbf{1.79 \times 10^{-10}}$
	α_1 DFA	1.25 ± 0.16	1.27 ± 0.15	0.4987
	α_2 DFA	$\mathbf{1.01 \pm 0.05}$	$\mathbf{1.09 \pm 0.07}$	$\mathbf{3.31 \times 10^{-7}}$
CHF	H_Period	0.12 ± 0.13	0.16 ± 0.13	0.3638
	H_DWTPeriod	0.13 ± 0.14	0.18 ± 0.13	0.3280
	H_VarDWT	0.16 ± 0.08	0.18 ± 0.07	0.4244
	H_DFA	0.11 ± 0.06	0.13 ± 0.06	0.3121
	α_1 DFA	0.93 ± 0.28	0.98 ± 0.24	0.5908
	α_2 DFA	1.10 ± 0.12	1.12 ± 0.09	0.6065

Table 5.3: *Mean ± standard deviation for young and elderly groups for both healthy and CHF subjects. Significant difference between young and elderly groups ($p < 0.05$) has been highlighted.*

5.4.2 Aging Curve

Table 5.4 shows the variation/year index and the determination coefficient of the H estimates and DFA scaling exponents in terms of the dependence on age, for healthy and CHF subjects.

For the healthy subset, the H parameter estimated by all the estimation methods and DFA scaling exponent α_2 reveal a linear relationship (steady increase) with aging, whereas DFA scaling exponent α_1 does not show this linear relationship, which means that it is not possible to assign it a variation/year index. The slope (variation/year index) of the linear relationship between H estimates and age shows a very similar value, including for α_2, which is approximately 0.0023, namely: 0.0023 for H parameters estimated by Periodogram method, DFA method and for scaling exponent α_2; 0.0024 for H estimated by DWT-and-periodogram method; 0.0022 for H estimated by Variance-of-DWT-details method. The DFA scaling exponent α_1 does not show significant linear relationship with age, which agrees with the fact of that this parameter did not show significant difference between young and elderly group in the previous analysis.

For the CHF subjects, there is no indication of a significant linear relationship between the H parameter and age for any estimation method. This is in agreement with the results shown in the previous section, where the estimation methods did not show any significant differences between young and elderly groups for CHF subjects. For this reason, the variation per year

		Var./year Indx.	Determ. coeff.
Healthy	**H_Period**	**0.0023**	**0.2462**
	H_DWTPeriod	**0.0024**	**0.2171**
	H_VarDWT	**0.0022**	**0.3073**
	H_DFA	**0.0023**	**0.3980**
	α_1 DFA	*xxxx*	0.0096
	α_2 DFA	**0.0023**	**0.2642**
CHF	H_Period	*xxxx*	0.0431
	H_DWTPeriod	*xxxx*	0.0575
	H_VarDWT	*xxxx*	0.0442
	H_DFA	*xxxx*	0.0413
	α_1 DFA	*xxxx*	0.0028
	α_2 DFA	*xxxx*	0.0407

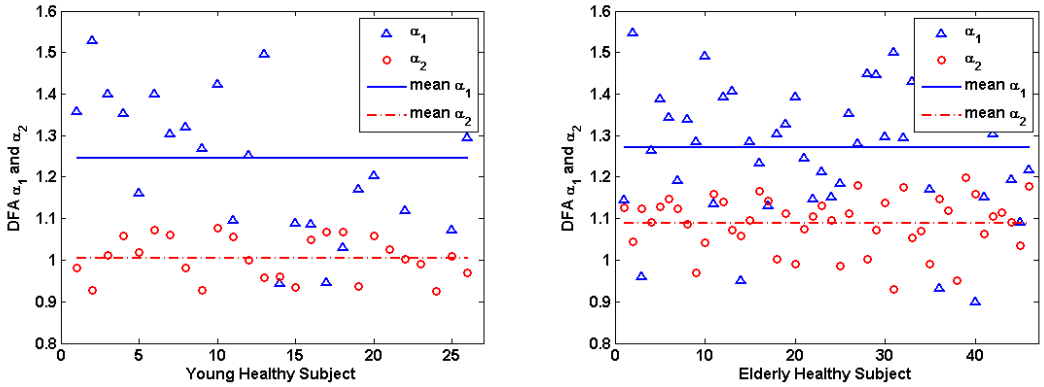
Table 5.4: *Variation/year index and determination coefficient for linear regression in healthy and CHF subjects. Significant relationship ($p < 0.05$) has been highlighted, when there is no significant relationship the variation/year index is replaced by xxxx.*

indices for all the estimation methods are not shown, and are filled with the *xxxx* symbol.

Figure 5.13 shows the dependence of the estimate H exponent on age for healthy subjects and for all estimation methods, namely: Periodogram method Fig 5.13 (a), DWT-and-periodogram method Fig 5.13 (b), Variance-of-DWT-details method Fig 5.13 (c), H parameter estimated by DFA method Fig 5.13 (d), and scaling exponents from DFA α_1 Fig 5.13 (e), and α_2 Fig 5.13 (f). When there exists no evidence for linear relationship, the regression line is plotted as a dotted-line. This only happens for DFA α_1, which agrees with the previous results when the discrimination capabilities between young and elderly were studied.

Figure 5.14 shows the dependence of the estimate H exponent on age for CHF subjects and for all estimation methods, namely: Periodogram method Fig 5.14 (a), DWT-and-periodogram method Fig 5.14 (b), Variance-of-DWT-details method Fig 5.14 (c), H parameter estimated by DFA method Fig 5.14 (d), and scaling exponents from DFA α_1 Fig 5.14 (e), and α_2 Fig 5.13 (f). When there exists no evidence of linear relationship, the regression line is plotted as a dotted-line. All the estimation methods show no linear relationship, which agrees with the previous results when the discrimination capabilities between young and elderly were studied.

The steady increase found in the H parameter for all estimation methods and for healthy subjects suggest a departure from the so called $1/f$ structure and therefore the loss of complexity



(a) (b)

Figure 5.12: *Scaling exponents α_1 ($\triangle$) and α_2 ($\circ$) from DFA for each young healthy (a) and elderly healthy(b) subject, also the means are plotted. Figures shows that the cross-over phenomenon does not hold for healthy subjects when comparing young and elderly.*

due to aging is reflected as a breakdown of the fractal correlation properties quantified by H, whereas there exists no reason to say that the scaling exponent α_1 change with aging, and therefore, it is possible to conclude that the fractal properties quantified by α_1 are preserved by aging. However, when a patient is suffering from a CHF, there seems to exist a breakdown of the fractal correlation properties quantified for H exponent, α_1, and α_2, and for both young and elderly subjects.

It should be noted that the cross-over phenomenon that occurs when comparing healthy and CHF subjects, does not hold for the case of comparing young and elderly subjects. Figure 5.12 (a) shows the DFA α_1 scaling exponent, represented as $\triangle$, and DFA α_2 scaling exponent, represented as $\circ$, for each young and elderly subject (Fig 5.12 (b)). Also the mean for α_1, as solid line, and for α_2, as dash-dot line, are plotted. This is the equivalent representation to the Figure 5.4 which shows the same scheme for healthy and CHF subjects.

This representation shows that the cross-over phenomenon does not occur for elderly subjects, which means that the loss of complexity due to aging does not affect in the same way that the CHF does. The CHF causes greater change in the fractal correlation properties quantified by α_1, in a way that the mean value of α_1 for CHF is lower than the mean value of α_2 (cross-over phenomenon), whereas aging causes greater change in the fractal correlation properties quantified by α_2, but preservers those quantified by α_1, hence it does not originate the cross-

over phenomenon.

5.5 Conclusions

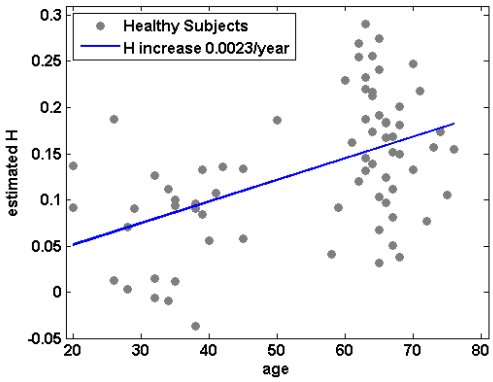
In this chapter, the different methods of estimating the H parameter are applied to real HRV signals, which are derived from 24-hours ECG monitoring (Holter) in both healthy and CHF subjects. Some features of the H parameter estimates from real HRV signals have been pointed out which have taken into account, namely the bi-fractal behavior and the frequency range where the $1/f$ spectral behavior holds.

Afterwards, the capabilities of the fractal correlation properties quantified by the H parameter and the DFA scaling exponents, to distinguish between the states of health and CHF were studied. All the methods gave H estimates that were significantly different for healthy and for CHF subjects, except for the Periodogram method. Moreover, the α_1 scaling exponent was the index that shown the best discrimination capabilities. The cross-over phenomenon was illustrated, in which the α_1 and α_2 mean values for healthy and CHF subjects were crossed-over,

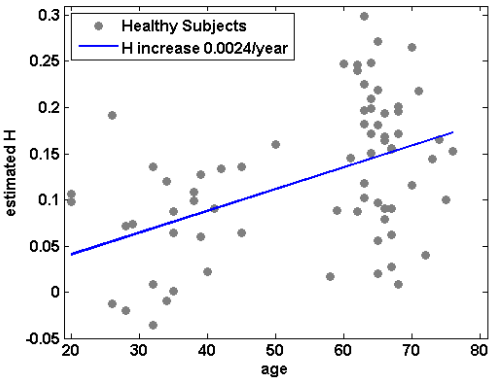
$$\begin{aligned} \bar{\alpha}_1 &> \bar{\alpha}_2, \quad \text{for Healthy subjects,} \\ \bar{\alpha}_2 &> \bar{\alpha}_1, \quad \text{for CHF subjects.} \end{aligned}$$

It was also showed that there exists a significant difference between the H parameter between day period and night period, for both healthy and CHF subjects, with the exception of α_1 for CHF subjects (even the difference for healthy subjects was very subtle). This difference can be due to the presence of environmental stimuli in the day period.

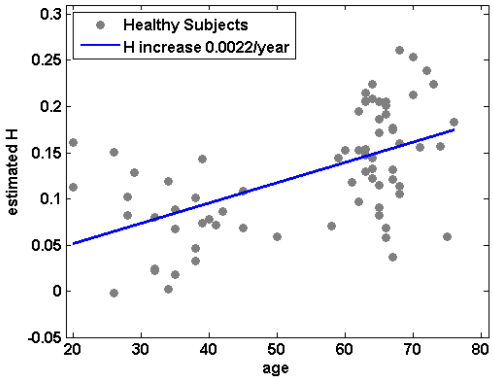
In the last section, the change in the fractal correlation properties, quantified by the H parameter and α_1 and α_2 scaling exponents, with age were studied. It was shown that the complexity loss due to aging causes, in healthy subjects, is reflected on a significant change in the H estimates and in α_2, but not in α_1. Also the CHF causes a breakdown of the fractal correlation properties of the same magnitude in both young and elderly subjects. Finally, we tried to estimate the variation per year of the H parameter and α_2, and it was found that from our data, the variation is approximately 0.0023 units per year, for healthy subjects. We can conclude that the fractal correlation properties quantified by the H parameter and α_2 are degraded with age, whereas aging preserves those properties quantified by α_1.



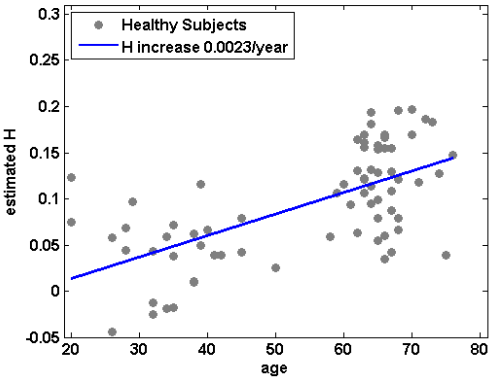
(a) Periodogram method Healthy



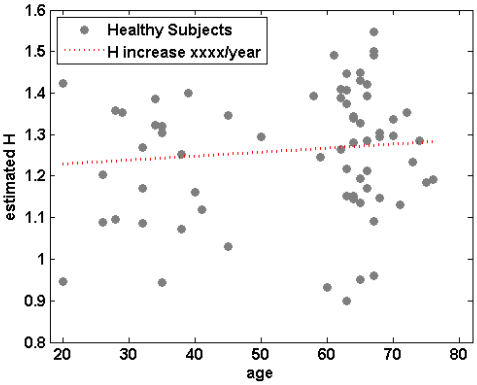
(b) DWT-and-periodogram method Healthy



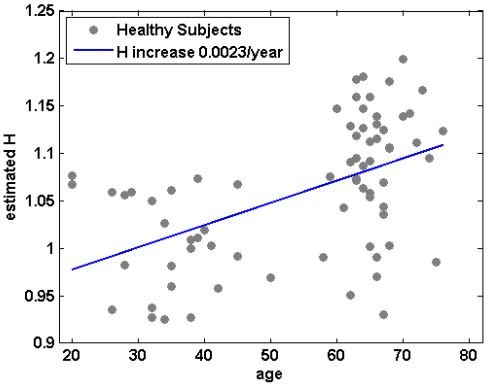
(c) Variance-of-DWT-details method Healthy



(d) Hurst by DFA method Healthy

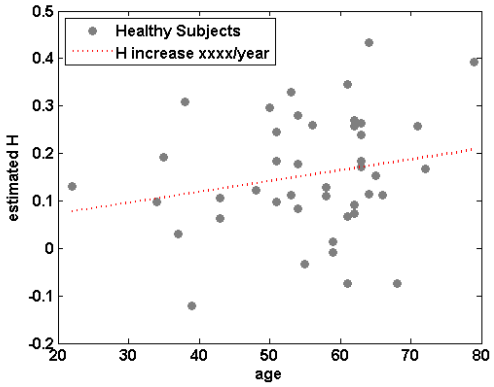


(e) DFA α_1 method Healthy

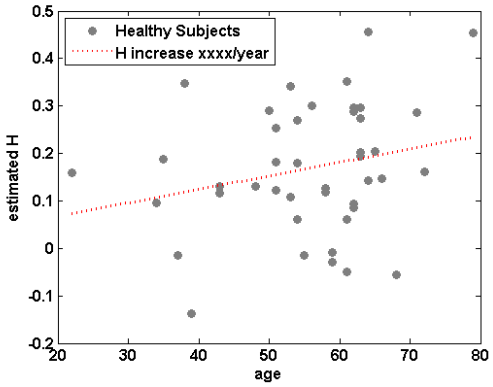


(f) DFA α_2 method Healthy

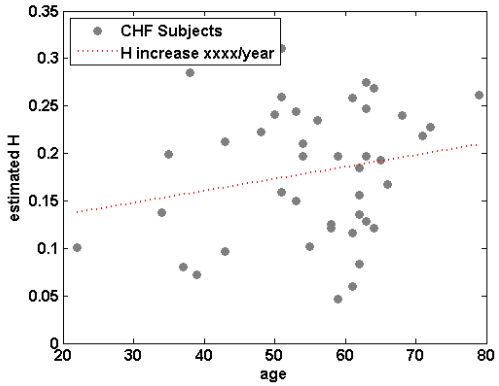
Figure 5.13: *Dependence of the estimate H exponent on age for healthy subjects estimated by Periodogram method (a), DWT-and-periodogram method (b), Variance-of-DWT-details method (c), H parameter estimated by DFA method (d), and scaling exponents from DFA α_1 (e), and α_2 (f). When there exists no evidence for linear relationship, the regression line is plotted as a dotted-line.*



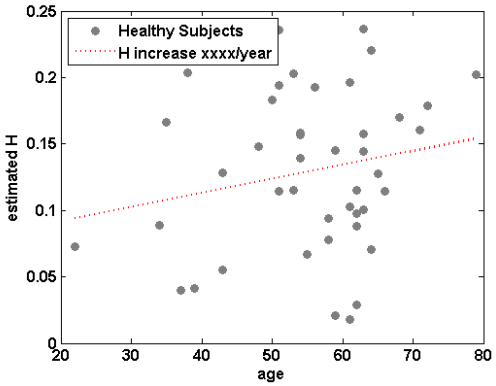
(a) Periodogram method Healthy



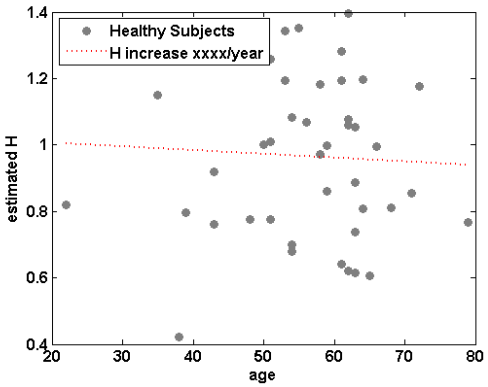
(b) DWT-and-periodogram method CHF



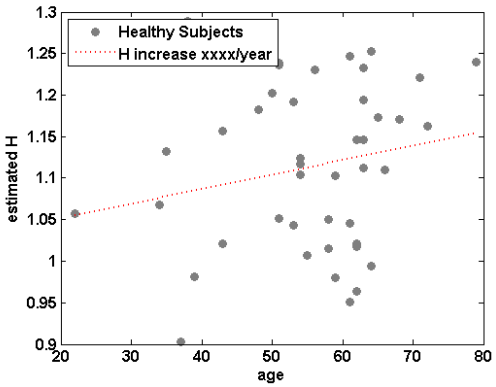
(c) Variance-of-DWT-details method CHF



(d) Hurst by DFA method CHF



(e) DFA α_1 method CHF



(f) DFA α_2 method CHF

Figure 5.14: *Dependence of the estimate H exponent on age for healthy subjects estimated by Periodogram method (a), DWT-and-periodogram method (b), Variance-of-DWT-details method (c), H parameter estimated by DFA method (d), and scaling exponents from DFA α_1 (e), and α_2 (f). When there exists no evidence for linear relationship, the regression line is plotted as a dotted-line.*

Chapter 6

DISCUSSIONS AND CONCLUSSIONS

Discovery consists of seeing what everybody has seen, and thinking what nobody has thought.
(A. Szent Győrgi, Hungarian Physiologist)

In this final chapter, we want to summarize and discuss the main topics pursued during the work and propose further studies that are a natural continuation of this work.

The work was mainly focused on the study of HRV. The RR signal reflects, partially, the characteristics of the cardiovascular system by which is modulated; at the same time, it is indirectly influenced by several other physiological subsystems, acting on different time scales. The focus of the work was to characterize the fractal correlation properties of the HRV signal.

In chapter 1, the background of the HRV analysis was outlined, whereas in the chapter 2, the fractal characterization of the fractal, self-similar, time series was presented by means of the study of the fBm properties, which is a well-known self-similar signal. Some different methods to characterize the fractal correlation properties, by the Hurst exponent estimate of time series were also presented.

In chapter 3, we studied in detail the performance of the different estimation methods, as well as the tuning of the free parameters in the algorithms. This analysis allowed us to compare the different methods, and we found, that in general, all the methods have a significant bad behavior in the extreme values of the valid range for H parameter. We also found, that the method we called Variance-of-DWT-details shows a linear bias, and due to the fact that the method has a very good performance in general, we spent some time trying to compensate experimentally this bias, i.e. obtaining an unbiased estimator. Finally we found, that this unbiased estimator

was the best one. Another estimation method, the so-called R/S method, showed a non-linear bias and in general a bad performance, hence we chose to discard this method from further considerations.

In chapter 4 we applied the estimation methods to real datasets, from healthy and pathological subjects (suffering from CHF). The main objectives were two, namely, discriminate between healthy and CHF subjects by its fractal correlation properties, and study the variation of this fractal properties due to aging.

For the first objective, we found that the DFA scaling exponent α_1 was the best index discriminating both cardiovascular states, whereas, the other estimators in spite of the fact that they gave significant differences, have shows bad results when trying to discriminate healthy from CHF. We could confirm the fact that there exists a cross-over phenomenon in CHF subjects, which means that, while for healthy subjects the DFA scaling exponent α_1 is greater than α_2, for CHF subjects the scaling exponent α_1 is smaller than α_2. This phenomenon remains as yet unclear, that is, there still is no clear physiological-based explanation for this finding. All the results suggested that the CHF causes a breakdown of the fractal correlation properties of the HRV signals.

For the second objective, we found that there exists hard evidence for a relationship between the variation of the H exponent and aging, such that it can be quantified by a variation per year index. The results suggested a variation of approximately 0.0023 units per year for the H estimates and scaling exponent α_2, which confirms the hypothesis that aging causes a departure from the $1/f$ structure, that is $H \approx 0$ and an approach to Brownian motion structure $H \underset{\text{aging}}{\longrightarrow} 0.5$. Moreover, the results for α_1 showed that this scaling exponent does not have a significant variation between young and elderly subjects; therefore, the fractal correlation properties quantified by α_1 are preserved with aging. A possible hypothesis could be that the mechanism involved in the cardiovascular response for this time scale —we remember here that α_1 is computed for time scales comprising between 4 to 16 beats, i.e., between a pair of seconds and twenty seconds more or less— has to maintain its complex response in order to cope with the environmental changes even for elderly subjects, whereas those mechanisms involved in large scales could be less complex without serious injury.

From a technical point of view, there exists some further improvements in the algorithms that can be achieved, namely:

- A similar signal detrended strategy as devised for DFA can be applied to other algorithms. This could improve the behavior of such estimators, because, this detrended procedure treats the fluctuations that are responses to the environmental stimuli as a trend that must be removed.

- All the algorithms are based on fitting a regression line to some function that exhibits a scaling behavior, but in real applications this only occurs in some ranges; therefore an automatic process to find the scaling region where the function is actually linear in a log-log plot, may allow to achieve accurate estimates in real situations.

- The application of non-linear detrended process in the DFA method, can be more suitable to real situations.

- Extension of the proposed algorithms to a multifractal (multi-scaling) analysis. There exits some previous works that seem to be promissory for the DFA method [Kantelhardt 02] and for the Variance-of-DWT-details method [Lashermes 04].

- A more detailed robustness analysis, for example, using bootstrap techniques, for both free parameter selection and for preprocessing signal analysis. This latter issue has not been taken into account in this work.

Whereas from a more experimental point o view we can make some remarks:

- More comprehensive datasets, for both pathological and healthy subjects. It is obvious that for more accurate results, and in order to establish more reliable conclusions it is necessary to deal with larger datasets.

- To extend the study to different pathologies in order to get a better understanding of the fractal properties damage originated by the different pathologies.

- An important challenge for biomedical engineers is to fill the gap between the research investigation and the clinical applicability of the results, one of multiple ways to do this is by means of providing a clear physiological meaning to the different indices. This could be achieved using mathematical models of the cardiovascular systems and studying the changes in the different indices [Ursino 03, Ursino 98, Rojo-Álvarez 07].

Appendices

Appendix A

MATLAB FUNCTIONS

In this appendix, the main functions developed in Matlab® for this work are presented.

A.1 Matlab Function for the Periodogram Method

```
function [H,H2,Psd,f] = periodH(x,kB,kN,hrv,flag)
%"""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
% Estimate the Hurst exponente using the slope of the power spectrum in a
% log-log plot. The psd is estimated by Welch's periodogram and with Kaiser
% window.
% USE:
% [H,Psd,f] = periodH(x,kB,kN,hrv,flag)
% INPUT:
%       .-x ---> Time series
%       .-kB ---> Parameter for the Kaiser Window.
%       .-kN ---> Segments lenght.
%       .-hrv ---> Indicate (hrv=1) wheter x is a RR interval time series.
%       .-flag ---> Plot (flag=1) the estimate procedure.
% OUTPUT:
%       .-H ---> Estimated Hurst parameter.
%       .-H2 ---> Estimated scaling exponente for HRV signals.
%       .-Psd ---> Estimated PSD.
```

```
%       .-f ---> Vector of frequencies
%"""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
% AUTHORS:
% OSCAR BARQUERO PEREZ
%
% DATE: 27-09-2007


%Default
H2 = [];
if isempty(flag)
    flag= 0;
end
fs = 1; %Fs is 1 by default
n2 = nextpow2(x);
if isempty(hrv)
    hrv = 0;
end
if nargin<2
    kB = 6.8;
end
[Psd,f] = pwelch(x-mean(x),kaiser(kN,kB),[],kN,fs);
%Linear Regression
%Regression for HRV signals
if hrv
ini=min(find(f>=1e-4));
fin=min(find(f>=1e-2));
end
if hrv
    ini2 = min(find(f>=1e-2))+10;
    fin2 = min(find(f>=0.4));
end
%Regresion for non-HRV signals
```

```
if ~hrv
    ini = min(find(log10(f)>=-3));
    fin = min(find(log10(f)>=-0.6));

end
p = polyfit(log10(f(ini:fin)),log10(Psd(ini:fin)),1);
H = (-p(1)-1)/2; %Estimate

%Mirror effect correction for non-HRV signals
if H<0.5
    if ~hrv
        ini = min(find(log10(f)>=-3));

        fin = min(find(log10(f)>=-1));
    end
end

p = polyfit(log10(f(ini:fin)),log10(Psd(ini:fin)),1);
H = (-p(1)-1)/2;
%Scaling exponente for high frequencies
if hrv
p2 = polyfit(log10(f(ini2:fin2-100)),log10(Psd(ini2:fin2-100)),1);
H2 = (-p2(1)-1)/2;
y2 = polyval(p2,log10(f(ini2:fin2-100)));
end
y = polyval(p,log10(f(ini:fin)));

%Plot
if flag
 close all
 plot(log10(f),log10(Psd),'.-')
 hold on
```

```
 plot(log10(f(ini:fin)),y,'r-','LineWidth',1.3)
 if hrv
 plot(log10(f(ini2:fin2-100)),y2,'r-','LineWidth',1.3)
 end
end
```

A.2 Matlab Function for the CWT-and-periodogram Method

```
function H = cwtPeriodH(x,wave,kB,kN,hrv,flag)
%"""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
% Estimate the Hurst exponente using the slope of the power spectrum in a
% log-log plot. The psd is estimated by CWT coeffients and Welch's
% periodogram with Kaiser window
% USE:
% H = cwtPeriodH(x,wave,kN,kB,flag)
% INPUT:
%       .-x ---> Time series.
%       .-wave ---> Wavelet function.
%       .-kB ---> Parameter for the Kaiser Window.
%       .-kN ---> Segments lenght.
%       .-hrv ---> Indicate (hrv=1) wheter x is a RR interval time series.
%       .-flag ---> Plot (flag=1) the estimate procedure.
%       principal.
% OUTPUT:
%       .-H ---> Estimated Hurst exponent.
%"""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""

if isempty(kN)
kN = 10000;
end
if isempty(kB)
kB = 6.8;
end
scaleleng = 63;
scale = 2:scaleleng+1;

coefs = cwt(x,scale,wave);
%Welch Periodogram for each scale
```

```
for m = 1:scaleleng
    [Ps(:,m),f] = pwelch(coefs(m,:),kaiser(kN,kB),[],kN,1);
    Ps(:,m) = Ps(:,m) / (scale(m)^2);
end
%Sum all contributions
Pint = sum(Ps');
%HRV signal
if hrv
ini=min(find(f>=1e-4));
[maxP,ini] = max(Pint);
fin=min(find(f>=1.5e-2));
end
%Non-HRV signal
if ~hrv
    [maxP,ini] = max(Pint);
    ini = ini+7;
    fin = length(Pint)-200;
    fin = min(find(log10(f)>=-0.6));

end
p = polyfit(log10(f(ini:fin)),log10(Pint(ini:fin)'),1);
H = (-p(1)-1)/2;
y = polyval(p,log10(f(ini:fin)));
%Plot
if flag
    figure
    plot(log10(f),log10(Pint))
    hold on
    plot(log10(f(ini:fin)),y,'r:','MarkerSize',3)
end
```

A.3 Matlab Function for DWT-and-periodogram Method

```
function H = dwtApproxPeriodH(x,wave,kN,hrv,flag)
%"""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
%  Hurst estimatee by the DWT-and-periodogram method
% USE:
% H = dwtApproxPeriodH(x,wave,kN,hrv,flag)
% INPUT:
%       .-x ---> Time series.
%       .-wave ---> Wavelet function.
%       .-kB ---> Parameter for the Kaiser Window.
%       .-kN ---> Segments lenght.
%       .-hrv ---> Indicate (hrv=1) wheter x is a RR interval time series.
%       .-flag ---> Plot (flag=1) the estimate procedure.
%       principal.
% OUTPUT:
%       .-H ---> Estimated Hurst exponent.
%"""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""

kB = 6.8;
lev = wmaxlev(length(x),wave);
[c,l] = wavedec(x,lev,wave);
Px = pwelch(x,kaiser(kN,kB),[],kN,1);
for m = 1:lev
    app = wrcoef('a',c,l,wave,m);
    [Ps(:,m),f] = pwelch(app,kaiser(kN,kB),[],kN,1);
    Ps(:,m) = Ps(:,m)/ (2^m)^2;
end
Psum = sum(Ps');
Psum = Psum' + Px;
if hrv
ini=min(find(f>=1e-4));
```

```
fin=min(find(f>=1e-2));
end
if ~hrv
    ini = 10;
    fin = length(Psum)-200;
    fin = min(find(log10(f)>=-0.6));

end
p = polyfit(log(f(ini+1:fin)),log(Psum(ini+1:fin)),1);
H = (-p(1)-1)/2;
y = polyval(p,log(f(ini+1:fin)));
if flag
    close all
    plot(log(f(1:end)),log(Psum(1:end)))
    hold on
    plot(log(f(ini+1:fin)),y,'r:')
end
```

A.4 Matlab Function for Variance-of-DWT-Details Method

```
function [H] = dwtVarDetH(x,wtype,flag)

%"""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
% Hurst exponent estimated by the Variance-of-DWT-details method.
% USE:
% [H] = dwtVarDetH(x,wtype,flag)
% INPUT:
%       .-x ---> Time series
%       .-wtype ---> Wavelet function
%       .-flag ---> Plot (flag=1) the estimate procedure.
% OUTPUT:
%       .-H ---> Parmetro de Hurst estimado.
%
%"""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""

if nargin<3
    flag = 0;
end
lev = wmaxlev(length(x),wtype)+2; %Max level
[c,l] = wavedec(x,lev,wtype); %Wavelet decomposition
%Variance compute.

%1) Robust method
stdc = wnoisest(c,l,1:lev);
%2) Traditional method
dcell = detcoef(c,l,'cells');
varD = ones(1,length(dcell));
for m = 1:length(dcell)
    varD(m) = var(dcell{m});
end
```

```
p = polyfit(2:lev-2,log2(stdc(2:lev-2).^2),1);
Hro = (p(1)-1)/2;


scal = 2.^(1:lev);
fin = min(find(log(scal)>=8));
p2 = polyfit(log(scal(2:fin)),log(varD(2:fin)),1);
H = (p2(1)-1)/2;
y = polyval(p2,log(scal(2:fin)));
%Plot
if flag
    close all
    plot(log(scal),log(varD))
    hold on
    plot(log(scal(2:fin)),y,'r:o')
end
```

A.5 Matlab Function for the RS Method

```
function [H] = rsHurst(datos,flag)
%"""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
% Hurst exponent estimated by the RS method.
% USE:
% [H] = rsHurst(datos,flag)
% INPUT:
%       .-datos ---> Time series
%       .-flag ---> Plot (flag=1) the estimate procedure.
% OUTPUT:
%       .-H ---> Parmetro de Hurst estimado.
%"""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""

datos = diff(datos); %First compute the fGn = diff(fBm).
maxPot2 = 2.^floor(log2(length(datos)));
datosAux = datos(1:maxPot2);
tamCaja = maxPot2;
numIter = 1;
RSave = [];
cajas = 1;
tC = [tamCaja];
while (tamCaja >= 8)
    inds = [];
    m = 1;
    for l = 1:cajas %Index for this segment lenght
        indsaux = (m : m+tamCaja-1) ;
        m = m + tamCaja;
        inds = [inds; indsaux];
    end
    RSiter = [];
    for g = 1:size(inds,1) %RS compute for this segment length
```

```
        RSaux = RScalc(datosAux(inds(g,:)));
        RSiter = [RSiter; RSaux];
    end
    RSave = [RSave;mean(RSiter)];
    numIter = numIter + 1;
    tamCaja = tamCaja / 2;
    cajas = 2.^(numIter-1);
    tC = [tC;tamCaja];
end
tC = tC(1:end-1);

[H, rectaReg, barrasError] = myregress(tC(end-1:-1:1),RSave(end-1:-1:1),flag);

%%Auxiliar Functions I%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [H, rectaReg, barrasError] = myregress(tC,RSave,plotFlag)
x = [ones(length(tC),1) log2(tC)];
y = log2(RSave);
[c,cint,r,rint,stats] = regress(y,x,0.05);
rectaReg = x*c;
barrasError = abs(rint(:,1)-r);
H = abs(c(2));

%Plot
if plotFlag
    plot(log2(tC),y,'o-')
    hold on
    plot(log2(tC),rectaReg,'r--');
end

%%Auxiliar Functions II%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function RS = RScalc(datos)
%RS compute
```

```
N = length(datos);
mediaDatos = sum(datos)/N; %Data mean
X = cumsum(datos-mediaDatos);
R = max(X) - min(X); %Range
S = sqrt(sum((datos-mediaDatos).^2)/N); %STD
RS = R/S; %RS
```

A.6 Matlab Function for the DFA method

```
function [alpha] = dfa(serie,flag)
%"""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
% Hurst exponent estimated by the DFA method.
% USE:
% [alpha] = dfa(serie,flag)
% INPUT:
%       .-serie ---> Time series.
%       .-flag ---> Plot (flag=1) the estimate procedure.
% OUTPUT:
%       .-alpha  ---> Sacling exponent, realted with the Hurst exponent by
%        H = alpha + 1.
%"""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""

if nargin<2
    flag = 0;
end
serie = serie(:);
serie = serie';
serie = serie - mean(serie);
y = cumsum(serie); %Integration
L = length(y);
%Segementation
npuntos = 80;
n = round(logspace(log10(3),log10(length(serie)/4),npuntos) );
f = zeros(1,npuntos);
a = 0;
for m = n
    a = a+1;
    seg = (1:m:L);%segments
    yn = zeros(1,L);
```

```
    for k = 1:length(seg)-1
        yaux = y(seg(k):seg(k+1)-1);
        x = 1:length(yaux);
        %Detrended procedure.
        [pen,ord] = myregress(x,yaux);
        ytrend = pen*x + ord; %Trend
        yn(seg(k):seg(k+1)-1) = ytrend';
    end
    %Fluctuation
    f(a) = sqrt((1/seg(end)).*sum((y(1:seg(end)-1)-yn(1:seg(end)-1)).^2));
end
p = polyfit(log10(n(2:end-10)),log10(f(2:end-10)),1);
alpha = p(1);
y = polyval(p,log10(n(2:end-10)));
%plot
if flag
    plot(log10(n),log10(f),'o-')
    hold on
    y = polyval(p,log10(n(2:end-10)));
    plot(log10(n(2:end-10)),y,'r:')
end
%%Auxiliar Function I%%%%%%%%%%%%%%%%%%%%%%
function [a,b]=myregress(x,y)
A=[sum(x.^2),sum(x); sum(x),length(x)];
C=[sum(x.*y);sum(y)];
v=inv(A)*C;
a=v(1); b=v(2);
```

A.7 Matlab Function for DFA Scaling Exponent

```
function [alpha1,alpha2] = dfa2(serie,flag)
%"""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
% DFA scaling exponents alpha1 and alpha2 estimate.
% USE:
% [alpha1,alpha2] = dfa(serie,flag)
% USE:
% [alpha] = dfa(serie,flag)
% INPUT:
%       .-serie ---> Time series.
%       .-flag ---> Plot (flag=1) the estimate procedure.
% OUTPUT:
%       .-alpha1 ---> Scaling exponent short-term.
%       .-alpha2 ---> Scaling exponent long-term.
%"""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""


if nargin<2
    flag = 0;
end
serie = serie - mean(serie);
y = cumsum(serie);
L = length(y);

%Alpha1%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
npuntos = 10;
f1=zeros(1,npuntos);
for n1=4:16
   %Segmentation
   seg =(1:n1:L);
   yn = zeros(1,L);
   for k = 1:length(seg)-1
```

```
        yaux = y(seg(k):seg(k+1)-1);
        x = 1:length(yaux);
        [pen,ord] = myregress(x,yaux);
        [a,b]=myregress(x,yaux);
        ytrend = pen*x + ord;
        yn(seg(k):seg(k+1)-1) = ytrend';
    end
    f1(n1) = sqrt((1/seg(end)).*sum((y(1:seg(end)-1)-yn(1:seg(end)-1)).^2));
end
n1=4:16;
f1=f1(4:end);
p1 = polyfit(log10(n1(1:end-2)),log10(f1(1:end-2)),1);
alpha1 = p1(1);

%Alpha2%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

npuntos = 40;
n = round(logspace(log10(16),log10(length(serie)/64),npuntos) );
f = zeros(1,npuntos);
a = 0;
%Segmentation
for m = n
    a = a+1;
    seg = (1:m:L);
    yn = zeros(1,L);
    for k = 1:length(seg)-1
        yaux = y(seg(k):seg(k+1)-1);
        x = 1:length(yaux);
        [pen,ord] = myregress(x,yaux);
        ytrend = pen*x + ord;
        yn(seg(k):seg(k+1)-1) = ytrend';
    end
```

```
    f(a) = sqrt((1/seg(end)).*sum((y(1:seg(end)-1)-yn(1:seg(end)-1)).^2));
end
p = polyfit(log10(n(1:end)),log10(f(1:end)),1);
alpha2 = p(1);
y = polyval(p,log10(n(1:end)));
y1 = polyval(p1,log10(n1(1:end-2)));

%Plot
if flag
    plot(log10(n(2:end)),log10(f(2:end)),'^')
    hold on
    plot(log10(n1(1:end-2)),log10(f1(1:end-2)),'^')
    plot(log10(n(1:end)),y,'r-')
    plot(log10(n1(1:end-2)),y1,'r-')
end

%%Auxiliar FunctionI%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [a,b]=myregress(x,y)
A=[sum(x.^2),sum(x); sum(x),length(x)];
C=[sum(x.*y);sum(y)];
v=inv(A)*C;
a=v(1); b=v(2);
```

Bibliography

[Abry 96] P. Abry & F. Sellan. *The wavelet-based synthesis for the fractional Brownian motion proposed by F. Sellan and Y. Meyer: Remarks and fast implementation.* Applied and Computational Harmonic Analysis, vol. 3, no. 4, pages 377–383, 1996.

[Akselrod 81] S. Akselrod. *Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control.* Science, vol. 213, pages 220–222, 1981.

[Baselli 86] G. Baselli, S. Cerutti, S. Civardi, D. Liberati, F. Lombardi, A. Malliani & M. Pagani. *Spectral and cross-spectral analysis of heart rate and arterial blood pressure variability signals.* Computers and Biomedical Research, vol. 19, no. 6, pages 520–534, 1986.

[Bassingthwaighte 94] J. B. Bassingthwaighte & G. M. Raymond. *Evaluating rescaled range analysis for time series.* Annals of Biomedical Engineering, vol. 22, no. 4, pages 432–444, 1994.

[Berger 01] R. D. Berger. *Broken Fractals: Where's the Break?* Journal of Cardiovascular Electrophysiology, vol. 12, no. 1, pages 33–35, 2001.

[Bigger 96] J. T. Bigger, R. C. Steinman, L. M. Rolnitzky, J. L. Fleiss, P. Albrecht & R. J. Cohen. *Power Law Behavior of RR-Interval Variability in Healthy Middle-Aged Persons, Patients With Recent Acute Myocardial Infarction,*

and Patients With Heart Transplants. Circulation, vol. 93, pages 2142–2151, 1996.

[Borg 05] F. G. Borg. *Review of Nonlinear Methods and Modelling.* ArXiv Physics, 2005.

[Brennan 01] M. Brennan, M. Palaniswami & P. Kamen. *Do Existing Measures of Poincar Plot Geometry Reflect Nonlinear Features of Heart Rate Variability?* IEEE Transactions on Biomedical Engineering, vol. 48, no. 11, pages 1342–1347, 2001.

[Cerutti 95] S. Cerutti, A. M. Bianchi & L. T. Mainardi. *Spectral Analysis of the Heart Rate Variability.* In Marek Malik & A. J. Camm, editors, Heart Rate Variability. Futura Publishing Company, New York, 1995.

[Clifford 02] G. D. Clifford. *Signal Processing Methods for Heart Rate Variability.* PhD thesis, University of Oxford, 2002.

[Costa 02] M. Costa, A. L. Goldberger & C.-K. Peng. *Multiscale Entropy Analysis of Complex Physiologic Time Series.* Physical Review Letters, vol. 89, no. 6, pages 068102–1, 2002.

[Dugnol 00] B. Dugnol, A. M. Fernández, D. A. Cernea & R. Sariego. *Geometría fractal.* website: `http://coco.ccu.uniovi.es/`, 2000.

[Falconer 86] K. J. Falconer. The geometry of fractal sets. Cambridge University Press, New York, NY, USA, 1986.

[Flandrin 89] P. Flandrin. *On the spectrum of fractional Brownian motions.* IEEE Transactions on Information Theory, vol. 35, no. 1, pages 197–199, 1989.

[Flandrin 92] P. Flandrin. *Wavelet analysis and synthesis of fractional Brownian motion.* IEEE Transactions on Information Theory, vol. 38, no. 2, pages 910–917, 1992.

[Glass 99] L. Glass. *Chaos and Heart Rate Variability.* Journal of Cardiovascular Electrophysiology, vol. 10, no. 10, pages 1358–1360, 1999.

[Goldberger 99] A. L. Goldberger. *Nonlinear Dynamics, Fractals, and Chaos Theory: Implications for Neuroautonomic Heart Rate Control in Health and Disease.* In C.L. Bolis & J. Licinnio, editors, The Autonomic Nervous System. World Health Organization, Geneva, 1999.

[Goldberger 00] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng & H. E. Stanley. *PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals.* Circulation, vol. 101, no. 23, pages e215–e220, 2000.

[Goldberger 02] A. L. Goldberger, L. A. N. Amaral, J. M. Hausdorff, P. Ch. Ivanov, C.-K. Peng & H. E. Stanley. *Fractal dynamics in physiology: Alterations with disease and aging.* Proceedings of the National Academy of Sciences, vol. 99, page 119, 2002.

[Hainsworth 04] R. Hainsworth. *Physiological background of heart rate variability.* In Marek Malik & A. John Camm, editors, Dynamic Electrocardiography. Blackwell Futura, 2004.

[Huikuri 99] H. Huikuri, T. Makikallio, J. Airaksinen, R. Mitrani, A. Castellanos & R.J. Myerburg. *Measurement of Heart Rate Variability: A Clinical Tool or a Research Toy?.* Journal of the American College of Cardiology, vol. 34, no. 7, pages 1878–1883, 1999.

[Huikuri 00] H. V. Huikuri, T. H. Mäkikallio, C.-K. Peng, A. L. Goldberger, U. Hintze & M. Moller. *Fractal Correlation Properties of R-R Interval Dynamics and Mortality in Patients With Depressed Left Ventricular Function After an Acute Myocardial Infarction.* Circulation, vol. 101, pages 47–53, 2000.

[Huikuri 03] H. V. Huikuri, T. H. Mäkikallio & J. Perkiömäki. *Measurement of Heart Rate Variability by Methods Based on Nonlinear Dynamics.* Journal of Electrocardiology, vol. 36, page 95, 2003.

[Ivanov 99] P.Ch. Ivanov, L. A. N. Amaral, A. L. Goldberger, S. Havlin, M. B. Rosenblum, Z. Struzik & H. E. Stanley. *Multifractality in healthy heartbeat dynamics*. Nature, vol. 399, pages 461–465, 1999.

[Kamath 93] M: V. Kamath & E. L. Fallen. *Power Spectral Analysis of Heart Rate Variability: A Noninvasive Signature of Cardiac Autonomic Function*. Critical Reviews in Biomedical Engineering, vol. 21, no. 3, pages 245–311, 1993.

[Kantelhardt 02] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde & H. E. Stanley. *Multifractal detrended fluctuations analysis of nonstationary time series*. Physica A, vol. 316, pages 87–114, 2002.

[Kaplan 91] D. T. Kaplan, M. I. Furman, S. M. Pincus, S. M. Ryan & L. A. Lipsitz. *Aging and the complexity of cardiovascular dynamics*. Biophysical Journal, vol. 59, pages 945–949, 1991.

[Kay 81] S. M. Kay & S. L. Marple. *Spectrum analysis-A modern perspective*. Proceedings of the IEEE, vol. 69, no. 11, pages 1380–1419, 1981.

[Kobayashi 82] M. Kobayashi & T. Musha. *1/f Fluctuation of Heartbeat Period*. IEEE Transactions on Biomedical Engineering, vol. 29, no. 6, pages 456–457, 1982.

[Laguna 95] P. Laguna, G. B. Moody & R. G. Mark. *Power spectral density of unevenly sampled heart rate data*. Engineering in Medicine and Biology Society, 1995., IEEE 17th Annual Conference, vol. 1, pages 157–158, 20-25 Sep 1995.

[Laguna 07] P. Laguna. *Tratamiento de señal aplicado al ECG. Course slides*. `http://www.tsc.uc3m.es/~antonio/trat_senal_ecg/Seminario.html`, 2007.

[Lashermes 04] B. Lashermes, R. Abry & P. Chainais. *Scaling exponents estimation for multiscaling processes*. Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on, vol. 2, pages ii–509–12 vol.2, 17-21 May 2004.

[Lisenby 77] M. J. Lisenby & P. C. Richardson. *The Beatquency Domain: An Unusual Application of the Fast Fourier Transform.* IEEE Transactions on Biomedical Engineering, vol. 24, no. 4, pages 405–408, July 1977.

[Lombardi 96] F. Lombardi, G. Sandrone, A. Mortara, D. Torzillo, M. T. La Rovere, M. G. Signorini, S. Cerutti & A. Malliani. *Linear and Nonlinear Dynamics of Heart Rate Variability After Acute Myocardial Infarction With Normal and Reduced Left Ventricular Ejection Fraction.* American Journal of Cardiology, vol. 77, pages 1283–1288, 1996.

[Lombardi 00] F. Lombardi. *Chaos Theory, Heart Rate Variability, and Arrhythmic Mortality.* Circulation, vol. 101, pages 8–10, 2000.

[Malik 89a] M. Malik, T. Cripps, T. Farrell & A.J. Camm. *Prognostic value of heart rate variability after myocardial infarction. A comparison of different data-processing methods.* Medical and Biological Engineering and Computing, vol. 27, pages 603–621, 1989.

[Malik 89b] M. Malik, T. Farrell, T. Cripps & A. J. Camm. *Heart rate variability in relation to prognosis after myocardial infarction: Selection of optimal processing techniques.* European Heart Journal, vol. 10, pages 1060–1074, 1989.

[Malik 95] M. Malik. *Geometrical Methods for Heart Rate Variability Assesment.* In Marek Malik & A. J. Camm, editors, Heart Rate Variability. Futura Publishing Company, New York, 1995.

[Malik 96] M. Malik. *Heart Rate Variability. Standards of measurement, physiological interpretation, and clinical use.* European Heart Journal, vol. 17, pages 354–381, 1996.

[Malik 04] M. Malik. *Standard Measurement of Heart Rate Variability.* In M. Malik & A. J. Camm, editors, Dynamic Electrocardiography. Blackwell Futura, 2004.

[Mallat 01] S. Mallat. A wavelet tour of signal processing. Academic Press, 2nd edition, 2001.

[Mandelbrot 68] B. B. Mandelbrot & J. W. van Ness. *Fractional Brownian motions, fractional noises and applications.* SIAM Review, vol. 10, pages 422–437, 1968.

[Mateo 03] J. Mateo & P. Laguna. *Analysis of Heart Rate Variability in the Presence of Ectopic Beats Using Heart Timing Signal.* IEEE Transactions on Biomedical Engineering, vol. 50, no. 3, page 334, 2003.

[Mielniczuk 07] J. Mielniczuk & P. Wojdyllo. *Estimation of Hurst exponent revisited.* Computational Statistics And Data Analysis, vol. 51, pages 4510–4525, 2007.

[Misiti 07] M. Misiti, Y. Misiti, G. Oppenheim & J. M. Poggi. *Wavelet Toolbox 4 User's Guide*, 2007.

[Mitra 01] S. K. Mitra. Digital signal processing: A computer-based approach. McGraw-Hill School Education Group, 2001.

[Moody 93] G. B. Moody. *Spectral analysis of heart rate without resampling.* Computers in Cardiology 1993. Proceedings., pages 715–718, 5-8 Sep 1993.

[Moody 06] G. B. Moody. *Frequency Domain Measures: The Fourier Transform, the Lomb Periodogram, and Other Methods*, 2006. HRV 2006 course, Physionet.

[Oppenheim 00] A. V. Oppenheim, R. W. Schafer & J. R. Buck. Tratamiento de señales en tiempo discreto. Prentice Hall, 2nd edition, 2000.

[Peng 95] C.-K. Peng, S. Havlin, H. E. Stanley & A. L. Goldberger. *Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series.* Chaos, vol. 5, pages 82–87, 1995.

[Pincus 90] S. M. Pincus. *Approximate entropy as a measure of system complexity.* Proceedings of the National Academy of Sciences, vol. 88, pages 2297–2301, 1990.

[Piskorski 07] J. Piskorski & P. Guzik. *Geometry of the Poincar plot of RR intervals and its asymmetry in healthy adults.* Physiological measurement, vol. 28, pages 287–300, 2007.

[Richman 00] J. S. Richman & J. R. Moorman. *Physiological time-series analysis using approximate entropy and sample entropy.* American Journal of Physiology - Heart and Circulatory Physiology, vol. 278, pages 2039–2049, 2000.

[Rojo-Álvarez 07] JL. Rojo-Álvarez, A. Sánchez-Sánchez, O. Barquero-Pérez, R. Goya-Esteban, E. Everss, I. Mora-Jiménez & A. García-Alberola. *Analysis of Physiological Meaning of Detrended Fluctuation Analysis in Heart Rate Variability Using a Lumped Parameter Model.* IEEE Computers in Cardiology 2007., 2007.

[Rompelman 77] O. Rompelman, A. J. R. M. Coenen & R. I. Kitney. *Measurement of heart-rate variability: Part 1—Comparative study of heart-rate variability analysis methods.* Medical and Biological Engineering and Computing, vol. 15, pages 233–239, 1977.

[Sayers 73] B. McA. Sayers. *Analysis of Heart Rate Variability.* Ergonomics, vol. 16, no. 1, pages 17–32, 1973.

[Schepers 92] H. E. Schepers, J. H. G. M. van Beek & J. B. Bassingthwaighte. *Four Methods to Estimate the Fractal Dimension from Self-Affine Signals.* IEEE Engineering in Medicine and Biology, vol. 11, no. 2, pages 57–64,71, 1992.

[Signorini 94] M. F. Signorini & S. Cerutti. *Lyapunov Exponents Calulated from Heart Rate Variability Time Series.* IEEE Computers in Cardiology, page 119, 1994.

[Sörnmo 05] L. Sörnmo & P. Laguna. Bioelectrical signal processing in cardiac and neurological applications. Academic Press, June 2005.

[Taqqu 95] M. S. Taqqu, V. Teverovsky & W. Willinger. *Estimators for long-range dependence: an empirical study.* Fractals, vol. 3, no. 4, pages 785–788, 1995.

[Ursino 98] M. Ursino. *Interaction between carotid baroregulation and the pulsating heart: a mathematical model.* American Journal of Physiology - Heart and Circulatory Physiology, vol. 275, pages H1733–H1747, 1998.

[Ursino 03] M. Ursino & E. Magosso. *Role of short-term cardiovascular regulation in heart period variability: a modeling study.* American Journal of Physiology - Heart and Circulatory Physiology, vol. 284, pages H1479–H1493, 2003.

[Wikipedia 08] Wikipedia. *Heart Failure.* `http://en.wikipedia.org/wiki/Heart_failure`, 2008.

[Wornell 93] G. W. Wornell. *Wavelet-Based Representations for the* $1/f$ *Family of Fractal Processes.* Proceedings of the IEEE, vol. 81, no. 10, pages 1428–1450, 1993.

[Yang 06] A. C.-C. Yang. *Poincar Plots: A Mini-Review*, 2006. HRV 2006 course, Physionet.