Performance analysis of Hurst exponent estimators using surrogate-data and fractional lognormal noise models: Application to breathing signals from preterm infants

Xavier Navarro, Fabienne Porée, Alain Beuchée, Guy Carrault

Performance analysis of Hurst exponent estimators using surrogate-data and fractional lognormal noise models: Application to breathing signals from preterm infants

X Navarro[1,2], F Porée[1,2], A Beuchée[1,2,3], G Carrault[1,2]

[1]Université de Rennes 1, LTSI, Rennes, F-35000, France;

[2]INSERM, U1099, Rennes, F-35000, France;

[3]CHU Rennes, Pôle Médico-Chirurgical de Pédiatrie et de Génétique Clinique, Rennes, F-35000, France

January 23, 2013

Address for correspondence:

Xavier Navarro

LTSI – INSERM U1099

Université de Rennes 1

Bat. 22, Campus de Beaulieu, 35042 Rennes Cedex (France)

e-mail: javier.navarro@univ-rennes1.fr

Phone: +33 2 23 23 62 20

Fax: +33 2 23 23 69 17

Abstract

The use of the Hurst exponent (H) to quantify the fractal characteristics of biological signals and its potential to detect abnormalities has aroused, recently, the interest of many researchers. Numerous techniques to estimate H are described in the literature, yet the choice of the most performing one is not straightforward. In this paper, we proposed some tests using artificial signals from experimental data and stochastic models to evaluate the robustness of three estimation techniques. Different surrogate-data tests, including a novel method to parametrize the degree of correlation in experimental signals with H (Hurst-adjusted surrogates), were first carried out. Then, simulated signals with prescribed H were obtained from fractional Gaussian noise modified properly to follow the lognormal laws observed in empirical data. The tests were applied to examine detrended fluctuation analysis (DFA), discrete wavelet transform and least squares based on standard deviation (LSSD) methods in the particular case of inter-breath interval signals from preterm infants. Simulations showed that none of the estimators were robust for every breathing pattern (regular, erratic and periodic) and should not be applied blindly without performing the preliminary tests proposed here. The LSSD technique was the most precise in general, but DFA was more robust with highly spiked patterns.

1 Introduction

The immaturity of the cardiorespiratory control system in premature newborns is known to be the origin of bradycardia, a decay of the heart beat rate, and apnea of prematurity (AOP), the cessation of breathing for 10-20 seconds [1]. Both events, frequently inter-related, provoke a diminution of the oxygen blood concentration increasing the risk of the newborns morbidity and mortality [2]. Other causes of apnea include infection, abnormal body temperature, or neurological problems. Regardless of their origin, sighs and respiratory

pauses are the responsible of the variable manner in which the infants breathe during sleep. Typically, three different patterns can be identified: 1) Regular breathing, a quiet, low variable breathing in amplitude and frequency, 2) erratic breathing, high variable breathing in amplitude and frequency including several episodes of AOP and 3) periodic breathing, the alternation of pauses lasting a few seconds followed by several rapid and shallow breaths [3]. The inter-breath interval (IBI) signals, formed by the succession of the respiratory cycle times, appear consequently as complex signals with spikes (apneas) of variable number and duration according to the pattern. Despite their random-like aspect, IBI signals exhibit long-range dependence (LRD) or fractal properties [4] as other physiological signals from preterm infants, such as the heartbeat rate [5]. Since changes in the fractal structure are related to aging and disease [6], the study of LRD for clinical purposes has become an issue of growing interest.

A well-known measure of the fractal properties is the Hurst exponent (H). It can be estimated by a large number of time or frequency domain methods, but their performances can differ substantially depending on the way they are defined or the context where they are applied. The numerous comparative analyses available in the literature basically prescribe H in simulated data, then it is estimated by different methods to evaluate the error. The works of Taqqu et al [7], and more recently Rea et al [8], presented an extensive review of this kind of empirical study, analyzing several estimators with different lengths of time-series generated by fractional Gaussian noise (fGn) [9] and fractional autoregressive-integrated moving average (FARIMA) models [10]. Other works focusing on non-Gaussian conditions [11, 12] tested the estimators with infinite variance FARIMA series and concluded that robustness decreases when short-range dependence (SRD) structures are present. The effect of non-stationarities on the estimated value of H ($\widehat{H}$) has also aroused numerous studies in view of the fact that real data often contains local trends and shifts in the mean, which are known to produce a false detection of LRD and bias [11]. In such conditions, the detrended

fluctuation analysis (DFA) [13] and a wavelet-based (DWT) estimator [14] result in good performances [15, 16].

However, the models to generate simulated time-series in the mentioned works might not be realistic for some experimental data, and more specifically, IBI signals from preterm infants. In a preliminary study [17], we analyzed five estimators through a basic model which approximated apneic patterns adding spikes ad hoc to fGn, finding out that DFA, DWT and the least squares based on standard deviation method (LSSD)[18] showed the best performances. The present work not only analyzes accurately the structure of IBIs, but also formalizes more rigorous models to generate artificial signals, with the purpose of evaluating the three aforementioned methods under realistic conditions.

This paper is organized as follows: In Section 2, we first introduce long-range dependent processes, the Hurst exponent, and the three estimators under evaluation. In Section 3, we explain how the real IBI signals are obtained and analyzed with surrogate-data tests, as well as how artificial IBIs are generated. Next, Section 4 describes the characteristics of IBI signals and the error of estimations through the different tests. A discussion of the main implications of these findings follows in Section 5 and some general remarks conclude this paper.

2 Long-range dependence and the Hurst exponent

2.1 Definitions

Long-range dependence or long-range memory is a property naturally present in many physical phenomena. It is characterized by self-similar (or fractal) behavior, meaning that similar statistical properties are preserved at different scale levels, which are related by a constant known as the Hurst exponent (H) [19]. By contrast, short-range dependence exhibits statistical similarity only at short scales. LRD was firstly reported in hydrology and subsequently

used in diverse areas of research such as geophysics, econometrics, network traffic and biology. Among the several definitions of LRD, the approach based on the second-order properties of a stochastic process is the most commonly accepted in the literature.

Let X_i be a stationary stochastic process with $i = 1, 2, ..$ the discrete sequence of observations. Let μ, $\gamma(k)$, $\rho(k)$ ($k \geq 0$) denote the mean, the auto-covariance and the autocorrelation, and $\sigma^2 = \gamma(0)$ the variance of the process. X_i is then considered self-similar if it has an autocorrelation of the form:

$$\rho(k) \sim k^{-\beta} L(i), \quad \text{as } k \to \infty, \tag{1}$$

where $\beta \in [0, 1]$, and L is a slowly varying function at infinity, i.e

$$\lim_{i \to \infty} \frac{L(ix)}{L(i)} = 1 \quad \text{for any } x > 0.$$

Given m a positive scalar representing a scale greater than 1 ($X_k^{(1)} = X$), a new data-series, the mean aggregated stochastic process, is obtained for every scale:

$$X_k^{(m)} = \frac{1}{m} \sum_{i=km}^{(k+1)m-1} X_i, \tag{2}$$

with an autocorrelation denoted $\rho^{(m)}(k)$ at each scale m.

The process X is called second-order (exactly) self-similar with a characteristic Hurst exponent ($H = 1 - \beta/2$) if the variance and autocorrelation are the same at each scale, so for all $m = 1, 2, 3, ...$,

$$\text{var}[X^{(m)}] = \sigma^2 m^{2H-2}, \tag{3}$$

$$\rho^{(m)}(k) = \rho(k) = \frac{1}{2}[(k+1)^{2H} + (k-1)^{2H}] - k^{2H}. \tag{4}$$

In finite processes (with k large enough), X is defined asymptotically self-similar if $\rho^{(m)}(k) \to$

$\rho(k)$ as $m \to \infty$. Self-similar processes are then scale-invariant, i.e. autocorrelations at low scale levels (aggregate processes $X_k^{(m)}$) are similar at higher scale (X).

The long-range dependence property is given by the fact that autocorrelation is non-summable, $\sum_k \rho(k) = \infty$, and described by a slow, positive decaying function. The Hurst exponent ranges from 0 to 1 but LRD, which describes natural phenomena, implies $H > 0.5$. A value equal to 0.5 means that the process is completely uncorrelated or random (white noise) and values under 0.5 describe processes with negative correlations or with an anti-persistent behavior.

2.2 H estimators under analysis

A considerable number of methods to estimate H, providing different approaches to quantify the self-similar behavior, have been proposed and analyzed in the literature. In this work, we considered the detrended fluctuation analysis and the discrete wavelet transform-based, already employed in respiratory signals, and the least squares based on standard deviation, a method used in hydrology yet unexplored in biomedical data.

2.2.1 Detrended fluctuation analysis

DFA, introduced by Peng et al [13] to estimate long-range dependence in non-stationary signals, has already been used to quantify the fractal content in IBI series from adults [20] and infants [21]. The data-series X_n of length N are first integrated and then divided into blocks of equal size m. A least squares line, representing the local trend, is fit to the data in each block. The y coordinate of the fitted line, $y_m(k)$, is then subtracted to the integrated series $y(k)$ to remove the trend in each box. Next, $F(m)$, the root-mean-square fluctuation

of this integrated and detrended time series is calculated:

$$F(m) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} [y(k) - y_m(k)]^2}. \tag{5}$$

This function is repeated successively over all time scales (box sizes) to characterize its relationship with the box size m. A power law, given by $F(m) \sim m^\alpha$ as $m \to \infty$, indicates the presence of LRD. α is the scaling exponent, a generalization of the Hurst exponent, and it is obtained by finding the slope of the line relating $F(m)$ to m in a log plot.

2.2.2 Discrete wavelet transform-based estimation method

Abry et Weitch [14] proposed a semi-parametric joint estimator of H based on the DWT, probed to be robust with non-stationarities even when signals contains SRD. It takes advantage of the scaling properties of the wavelet basis, which captures optimally the scaling self-similar nature of LRD processes.

Briefly, the method performs first a wavelet decomposition of a given discrete time series X_n, providing $d_x(j, k)$, the wavelet coefficients or details. Next, at each fixed octave j the details are squared and then averaged across k to produce an estimate of the variance of the wavelet coefficients, called μ_j. A plot of $\log_2(\mu_j)$ against j is done to identify the range of octaves where scaling occurs. Finally, H is computed by performing a weighted linear regression over those scales. The algorithm employed in the present work utilizes Daubechies wavelets to perform DWT and identifies automatically the octaves with scaling [22].

2.2.3 Least squares based on standard deviation method

Classical statistics assume that a sample is constituted by independent random variables, but this is not consistent with finite processes exhibiting long range dependence. In particular, the variances of the sample mean decay slower than those from the sample size, hence

important bias can be introduced if the properties of self-similarity are not taken into account [23]. It has been shown in [24] that a robust estimator of variance $\widehat{S}^2$ with known H is given by:

$$\widehat{S}^2 = \frac{n-1}{n - n^{2H-1}}S^2 = \frac{1}{n - n^{2H-1}} \sum_{i=1}^{n}(X_i - \overline{X})^2, \tag{6}$$

with $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Note that when $H = 0.5$ (absence of long-term memory), the above expression becomes the classic variance estimator.

Koutsoyiannis proposed a method to estimate both H and standard deviation, σ, using modified statistics for long-range dependent processes [18]. Combining expressions (2) and (6), assuming that $\mathrm{E}(\widehat{S}) = \sigma$, the standard deviation of $X^{(m)}$ at each scale m could be approximated by $\widehat{S}^{(m)} \approx C_m(H)m^{H+1}\sigma$, where:

$$C_m(H) = \sqrt{\frac{n/m - (n/m)^{2H-1}}{n/m - 1/2}}. \tag{7}$$

Then, the next step was to minimize the fitting error e^2 to estimate both H and σ:

$$e^2(H, \sigma) = \sum_{m=1}^{m'} \frac{[\ln \sigma^{(m)} - \ln S^{(m)}]^2}{m^p}, \tag{8}$$

where m' is the maximum scale level so that $m' = n/10$ and p is a weight factor. After a Monte-Carlo study, it was found that $p = 2$ gave the best H estimation. A complete analytical and iterative procedure is detailed in the original paper [18] and a performance analysis in [25].

3   Methodology

In this section, we describe the methodology employed to evaluate the performance of the H estimators under test in IBI signals. After the selection of real signals (3.1) and the

extraction of IBIs (3.2), we designed two surrogate-data tests (3.3) with two goals in mind: 1) to verify the existence of LRD and SRD, and 2) to examine the performance of the estimations regarding the dispersion of $\widehat{H}$. Moreover, a novel surrogate technique, that will be referred as Hurst-adjusted surrogates (HAS), is introduced here to generate artificial data from real IBIs governed by H. Furthermore, to reinforce the results from the surrogate analysis, a second strategy evaluates the error of the estimators using simulated signals with prescribed H (3.4).

3.1 Selection of real breathing signals

Twenty-four breathing signals from a database of 43 preterm infants were selected in order to characterize the three respiratory patterns. Patients were born between 29 and 33 weeks of gestation at the University Hospital of Rennes (France) and were three to 10 days of postnatal age. Exclusion criteria were: ongoing inflammatory response, medication known to influence the autonomic nervous system except caffeine, intratracheal respiratory support, intracerebral lesion or malformation. This study was under parental consent and approved by the local ethics committee (03/05-445).

All recordings (Powerlab system; ADInstruments, Oxfordshire, UK) were performed in the neonatal intensive care unit (NICU) and consisted of a 1-hour recording at a 400-Hz sampling rate of abdominal strain gauges. Signals were low-pass filtered (20 Hz cutoff frequency) and subsampled by a factor of 10. Artifacts, mostly due to gross body movements or clinical intervention, were rejected by visual inspection. After preprocessing raw signals with Matlab®, 20-minute excerpts were selected by clinicians according to their predominant breathing pattern so that they were classified in three groups of eight signals.

3.2 Extraction of IBI signals

Respiratory rate or IBI signals are formed by collecting the time durations in all cycles of a breathing trace (see Figure 1-a). Since the fluctuations in the abdominal trace are proportional to the variations of the tidal volume, a breathing cycle is the elapsed time between successive minima, called total time (t_{tot}). Inspiration (t_i) and expiration (t_e) times correspond to the first and second half-cycles delimited by the maximum (see Figure 1-c). A peak detector was employed to find local extrema in the respiratory signal, but cycles were considered valid only if the difference in amplitude between consecutive maxima et minima exceeded a threshold, defined as 0.2 times the interquartile range of the trace.

IBI signals are constituted by a discrete sequence of time intervals, $X_i = t_{tot}(i)$, where index i is not proportional to time. Data-series coming from irregularly observed experiments have already been studied by the Hurst exponent in other research areas as ecology or clinical trials [26]. In this paper, H is also estimated directly on our experimental signals, assuming that it should not be interpreted in strictly temporal terms, but rather as a parameter describing the self-affinity in a signal.

3.3 Design of surrogate-data tests

Surrogate-data tests are utilized to statistically infer the nature of the process generating the observed data [27]. The realization of a surrogate set, artificial signals preserving certain statistical properties of the original data, determines the formulation of the null hypothesis, H_0, which is to be tested with a statistic capable to discern original and surrogates properties. If the statistics from the surrogate set and from the empirical data are significantly different, H_0 can be rejected.

In this work, we designed two surrogate tests, each posing different null hypotheses and modifying differently the original data:

- Random-shuffled surrogates (RSS) test: Temporal correlations are destroyed (theoretically, $H = 0.5$) and spectrum is whitened in surrogate-data. The addressed null hypothesis is that LRD in empirical data is generated by uncorrelated noise. H can be used directly as discriminant statistic thanks to its ability to detect LRD.

- Small-shuffled surrogates (SSS) test: Surrogate-data have their samples shuffled in a small scale, consequently local correlations (SRD) are lost and long-term structures preserved (H is theoretically unchanged). The null hypothesis addressed with SSS is that empirical data do not contain short term dependent processes. The average mutual information (AMI) computed in short lags to capture short-term dynamics can be used as discriminant statistic [28].

Although the main goal of RSS and SSS tests is to demonstrate the existence of LRD and SRD, the set of artificial data generated by both methods will be used for a second purpose: to evaluate the performance of estimators measuring the bias from the theoretical values of H.

However, RSS and SSS confine H to very specific values, and hence the tests of robustness remains incomplete. The Hurst-adjusted surrogates overcome this limitation by generating artificial signals with variable H. The principle consists in replicating the degree of correlation within an ordinary realization of fGn (governed by H) to an empirical record X. This procedure is done as follows:

1. Generate Y, a realization of fGn with the desired H and the same standard deviation of X and form iy, the vector of indexes of Y.

2. Obtain Y' by sorting in ascending order Y, and get iy', a vector containing the original positions of the fGn samples.

3. Form a new series X' by sorting X, also in ascending order.

4. Reorder X' using the indexes of iy'.

These artificial signals, equally distributed than the original ones, are characterized by an H approximately equal to the value prescribed in fGn.

3.4   Generation of simulated signals with prescribed H

The utilization of surrogate-data to evaluate the estimators has the advantage of reproducing realistic conditions, but at the expenses of imposing approximated values of H on artificial signals. Therefore, to conduce more complete performance studies, simulated self-similar signals with an exact prescription of H are necessary.

In order to generate processes exhibiting LRD, several methods can be utilized. One of the most employed is the fractional Gaussian noise, a stochastic model able to synthesize exact self-similar realizations [9]. The FARIMA model is another alternative if both short-range and long-range dependence need to be introduced but, for simplicity, and knowing that the original series does not contain important SRD processes (see section 4.2), we utilized the fGn model.

fGn is originated from the family of fractional Brownian motions (fBm) $B_H = \{B_H(t), t \in \mathbb{R}\}$, that are Gaussian self-similar stochastic processes with stationary increments indexed by the Hurst parameter H. So, a process $X = \{X_k, k \in \mathbb{Z}\}$, with an autocorrelation function $\rho(k)$ as equation (1), is fractional Gaussian noise if $X_k = B_H(k+1) - B_H(k)$.

We employed the *circulant embedding* technique [29] to synthesize stationary, normally-distributed realizations of fGn with $H \in \{0.5, 0.55, ...0.9\}$. However empirical data are described by lognormal distributions (see next section), hence fGn need to be transformed to follow the same marginal law to test the estimators under similar conditions.

A Gaussian discrete series, $X_G(n)$ can be transformed into a log-normally distributed

series $X_L(n)$ using the relation [30]:

$$X_L(n) = e^{\mu_G + \sigma_G X_G(n)}, \tag{9}$$

where μ_G and σ_G,the mean and standard deviation of the normal distribution, are given by:

$$\mu_G = \ln \frac{\mu_L}{\sqrt{1 + cv^2}} \quad ; \quad \sigma_G = \sqrt{\ln(1 + cv^2)}, \tag{10}$$

where μ_L is the mean of the lognormal distribution and $cv = \sigma_G/\mu_G$ its coefficient of variation, a measure of the relative dispersion of the series. The transformation of normally distributed series to lognormal processes can be applied to fGn, obtaining fractional Lognormal noise (fLn) with approximately the same characteristic Hurst exponent [31]. Values of cv in close proximity to zero produce a highly peaked distribution with low variance, whereas greater values increase the tail. A few examples of fLn with several values of cv and H can be seen in Figure 2.

4 Results

In this section the characteristics of real IBI signals are first described, and then the hypotheses of long- and short-range dependence are tested. Finally, the results from the test of robustness using artificial and simulated data are presented.

4.1 Characteristics of real IBI signals

Subsequent to the application of the peak detector on the 24 respiratory records, we computed the mean duration of cycles, $\overline{t_{tot}}$, and the standard deviation, $std(t_{tot})$. Apnea were defined as cycles $\geq 3 * \overline{t_{tot}}$, hence the number of apneas per hour, N_{Ap}, and the main duration of apneas, D_{Ap} were computed for each IBI signal. Regarding the three groups of

13

infants, these parameters showed statistically significant differences in a Mann-Whitney U-test, whereas the gestational age (GA) and the weight did not (see Table 1).

With regard to the marginal laws, the presence of apneas (spikes), variable with the individual breathing pattern, and the fact that there are no negative times, determines right skewed distributions in IBI signals, which were best fit with lognormal laws [32]. In our IBI signals, the fitted lognormal parameters between the three patterns were also significantly different. Later on, the mean and standard deviation of the distribution parameters (Table 1) will be employed to simulate the fLn series.

On the other hand, the estimation of H on real data (see Figure 3) revealed that, as expected for biological signals, LRD ($H > 0.5$) is present in the great majority of the cases with the exception of some periodic IBI. Likewise, the different sensitivities of DFA, DWT and LSSD is evidenced by the meaningful differences of $\widehat{H}$ on the same groups.

4.2 Long and short-range dependence in real IBI signals

Given that LRD is detected on the real IBIs, it should be verified if 1) LRD is not caused by uncorrelated processes in IBIs (apneas and local trends), and if 2) IBI signals do not contain strong short-term dependent dynamics, known to perturb the estimation of the Hurst exponent [11].

Hence, long- and short-range dependence need to be examined by a RSS and a SSS test. In next lines, the following notation concerning H will be utilized:

- H_r and $\widehat{H_r}$: The real value of H of an empirical signal and its estimation, respectively.

- H_s and $\widehat{H_s}$: The theoretical value of H imposed by a surrogate technique and its estimated value.

In both tests, a set of 100 surrogates was obtained from each original signal. Then, $\widehat{H_s}$ and $\widehat{H_r}$ were compared by a Mann-Whitney U-test to find possible statistically significant

differences.

4.2.1 Results of RSS tests

The RSS provided data-series with $H_s = 0.5$, meaning that LRD is eliminated by un-sorting randomly the empirical data. Statistically significant differences were found between $\widehat{H_s}$ and $\widehat{H_r}$ using the three estimators (p_{DFA}, p_{DWT} and $p_{LSSD} < 0.0001$) in regular and erratic group. However, these differences were not found in the periodic group ($p_{DFA} = 0.585; p_{DWT} = 0.919; p_{LSSD} = 0.176$), presumably due to the nature of this breathing pattern, which contains less LRD and hence closer values of H_r to 0.5.

Therefore, the null hypothesis, that long-range dependence is caused by an uncorrelated process, can be rejected by regular and erratic groups, but not by the periodic group.

4.2.2 Results of SSS tests

The SSS produced a set of signals without SRD containing approximately the same LRD ($H_s \approx H_r$). Comparing $\widehat{H_s}$ and $\widehat{H_r}$ it could be stated that SRD does not affect the estimation of LRD because no statistically significant differences were found, but the use of the statistic given by AMI, as introduced previously, is a more rigorous criterion. AMI was computed with lags from 1 to 10 to quantify the dependence between close cycles in each original signal and in a set of 100 SSS. The analysis of the 24 IBI signals (see Figure 4) revealed that there are statistically independent differences regarding the AMI of original and surrogate populations in the closest cycles: breaths one to three in regular patterns, one and two in erratic and only the first breath ahead in periodic breathing. Therefore, the null hypothesis, that there is no short range dependence in IBI signals, can be rejected only in the mentioned lags.

4.3 Robustness of estimators

As can be observed in Figure 3, the meaningful differences regarding $\widehat{H_r}$ poses a dilemma to choose the most appropriate estimator in each pattern. The following tests elucidate and compare the robustness of DFA, DWT and LSSD employing artificial data generated by surrogate techniques and simulated data generated by the fLn model.

4.3.1 Results in surrogate-data

The knowledge of H_s in surrogate-data allowed to carry out a first evaluation of the estimators with signals identically distributed than the real ones. The robustness was measured by the error of estimations, ϵ_{sur}, defined as:

$$\epsilon_{sur} = H_s - \widehat{H_s}. \tag{11}$$

We used the set of 100 RSS ($H_s = 0.5$) and SSS ($H_s \approx H_r$) generated in previous tests to compute ϵ_{sur} on the 3 groups of IBIs (see first and second columns in Figure 5). The sample size of surrogate-data, as real IBIs, is comprised approximately between 500 and 1000 samples due to the variations on breathing rates and apneas. In view of the boxplots dispersion and bias from the middle lines ($\epsilon_{sur} = 0$), DWT is the least accurate and least precise technique, whereas LSSD exhibits very small bias and the least scattered interquartile ranges (IQRs).

The HAS method was employed to obtain surrogates with different levels of LRD, setting H_s from 0.6 to 0.9 in steps of 0.05. The evaluation of the robustness in this case was measured by the mean surrogate error, $\overline{\epsilon_{sur}}$, computed as the mean of the different ϵ_{sur} obtained by 100 realizations of HAS for each value of H_s (see third column in Figure 5). The boxplots show, in general, more dispersed and positive-biased IQRs, revealing that the presence of a stronger LRD perturbs the estimations. Of note, only LSSD has its IQRs below the simulation step

$(\Delta H_t = 0.05)$ in all patterns, despite having the greatest bias.

4.3.2 Tests with fLn

Prior to evaluating the estimation errors, we first verified if the apneas generated by the fLn model are realistic. One hundred fLn series of 1024 samples were realized for cv from 0.2 to 1 in steps of 0.05, then the mean duration (D_{Ap}) and number of apneas per hour (N_{Ap}) were obtained. Both parameters are plotted in Figure 6 as a function of cv, together with the corresponding points from empirical IBIs. As it can be observed, fLn signals can reproduce the frequency of real apneas, with the 87.5% of empirical N_{Ap} into the 95% confidence intervals (CI) of simulated values. However, the duration of apneas shows some limitations based on the fact that the 47.5% of empirical D_{Ap} (specially long apneas) are not included into the 95% CI obtained by the simulated fLn.

Simulations were performed by generating 100 fLn signals using the average parameters σ_L and μ_L for the three breathing patterns (see Table 1), and H was prescribed from 0.5 to 0.9 in steps of 0.05. To evaluate the robustness of DWT, DFA and LSSD, we utilize the estimation error from simulated values, ϵ_{sim}, the homologous of expression (11):

$$\epsilon_{sim} = H_r - \widehat{H_r}, \tag{12}$$

where H_r is in this case the prescribed value of H by fLn and $\widehat{H_r}$ its estimation. In Figure 7 ϵ_{sim} is plotted as a function of H_r for 1024-sample data, comparing the curves obtained for each simulated pattern and estimator.

It can be noticed that, in general, LSSD is the most precise technique, showing the narrowest 95% CI. However, its accuracy diminishes slightly as the simulated pattern becomes more spiky and the strength of LRD increases. On the other hand, the DWT-based method has the worst precision yet enhances the accuracy with irregular and periodic patterns, and

DFA shows the best performance in periodic-like patterns. Observing the sense of bias, DFA and DWT are prone to underestimate the real value of H whereas LSSD behaves conversely.

Simulations using 512-sample signals were also undergone, and the confidence intervals of the estimation errors and bias (not shown here) exceeded 20 to 40% the values obtained in simulations for 1024 samples. Therefore, it can be stated that robustness estimating H decreases as the length of data diminishes, as reported previously [8].

5   Discussion

In this paper, we investigated the usage of the Hurst exponent in the particular case of IBI signals from preterm infants. Since the way these newborns breathe can be highly variable, we selected representative signals of the typical immature breathing patterns, regular, erratic and periodic, to bring up two main issues: Do all patterns can be analyzed reliably by the Hurst exponent? How commonly used estimators perform on these signals? To uncover these questions, we generated both artificial data, using surrogate-data techniques, and simulated data, using the fLn model.

5.1   Main findings

The application of the surrogate-data tests allowed first to evaluate the properties of IBI signals. The random-shuffled surrogates found significant long-range dependence, as reported by earlier works [21, 32], and the small-shuffled surrogates proved the existence of short-term memory regarding the three first consecutive breaths in regular breathing, as it had been found in adults [33]. Not surprisingly, the SRD is weaker as the randomness and apneas increase, i.e. typically immature respiration, where it decreased to two and one consecutive breaths in erratic and periodic patterns, respectively. As randomness is associated with uncorrelated and irregular processes (weak LRD), this result is also consistent with the

lowest values of $\widehat{H_r}$ characterizing periodic patterns. On the contrary, regular patterns show the highest $\widehat{H_r}$, suggesting a more predictable long-term behavior within breaths.

The Surrogate-techniques served as well to test the performance of the estimators using artificial data generated from experimental signals. Interestingly, the HAS method provides identically distributed signals than real IBIs approximating H, permitting to assess the estimation errors in a realistic scenario including spikes. Evaluating the results of this test, it can be stated that H tends to be overestimated in situations in which LRD becomes stronger in artificial signals, particularly if the pattern analyzed contains more spikes, as erratic and periodic breathing. A possible explanation could be the inefficacy of HAS to replicate the exact correlations of fGn on the basis of apneic patterns, but there is still an uncertainty if the presence of spikes itself introduces the before-mentioned bias.

This is the reason why our stochastic model to simulate IBIs, initially Gaussian, was modified to generate apneas with approximately the same probability as the real signals. In this way, H can be prescribed reliably by the fLn while apneas can be regulated by the distribution parameters of empirical data. Nevertheless, the fLn approach has some limitations. Severe apneas, which in some patterns exceed frequently 10 seconds, cannot be reproduced within this model because the tail of the fitted distributions restricts greater values. Other marginal laws could be more suitable for highly spiked patterns, such as generalized extreme distributions, but they will be the object of further research.

However, it is noteworthy that the behavior of the estimators with simulated fLn is similar to the trends observed in the surrogates analysis: LSSD appears to be the most precise but least accurate, DWT the most imprecise and DFA seems to be a compromise between precision and accuracy. In spite of the limitations of the fLn model, it can be stated that the use of DWT on IBI signals, regardless of the pattern, should be avoided due to the unacceptable bias in estimations (the 75% CI exceed the simulation step). LSSD could be a good choice in case-control studies where two populations need to be characterized by H in

19

regular-erratic patterns, because its good precision gives the best discriminant power despite overestimating the real value of H. On the other hand, the robustness of DFA facing spiky signals makes it a more reliable choice in periodic breathing or mixed patterns.

In relation to the length of the breathing signals, it has to be considered that long and clean records are difficult to obtain at NICU, so the number of samples in IBI series could be in some cases insufficient to compute H reliably. Therefore, the optimal length of analysis should be a trade-off between a realistic sample size and a tolerable level of bias in estimators, but keeping in mind that the interest of the sample size is not only technical. Indeed, the choice of long IBIs can result in capturing different breathing dynamics, associated with changes in sleep [34]. As sleep stages are quite variable in duration, we consider appropriate to set the range of the analysis between 512 and 1024 samples, equivalent to time intervals from 10 to 25 minutes.

5.2   Clinical implications

Several breathing disorders are challenging problems in the premature infant. These disorders include frequent and severe apneas, erratic and periodic breathing, exaggerated or insufficient response to laryngeal chemoreflex [35, 36]. These events, when they repeat, are usually the reason why the hospitalization in intensive care unit is required or prolonged [37] and are associated with a worse neurodevelopmental outcome [38]. They are related to immaturity of the respiratory control and the inability of the premature infant to cope with a challenging environment. The analysis of the respiratory pattern variability, proposed in the present study, add comprehensive information in the documentation of these adverse events which may help to improve their management, if introduced in a bed side monitoring system.

Another challenging problem in neonatal medicine is infection [39]. Premature infants are at high risk of hospitalized acquired infection due to immaturity of their immune system and frequent need for intensive care. Early and correct diagnosis of infection is often difficult in

these patients due to non-specific and subtle presenting symptoms [40]. In this perspective, the analysis of cardiorespiratory signals appears to be a promising tool. Indeed, heart rate variability [41, 42] and cardiorespiratory interactions [43] appear to be affected by infection and these characteristics are already implemented in neonatal monitoring systems to help the physician in its decisions [44]. Since respiratory control is affected by infection, these methods of exploring the complexity of the respiratory signal, alone or as part of a multiparametric model, may help in the recognition and early diagnosis of infection, including in the context of a highly immature respiratory control.

Apart of these two areas of application for H calculation, other potential areas are numerous in the neonatal medicine looking at the development of these techniques in the adult intensive care medicine [45, 26, 46], H could be used as well in neonatal monitoring systems as an additional non-invasive indicator of both pathology and maturation.

6 Conclusion

The Hurst exponent is gaining interest in many fields thanks to its ability to quantify the long-range dependence. However, its application on some non-Gaussian signals with spiky and changing patterns like the variability signals analyzed here, should be examined carefully. As previous works demonstrated, the estimation of H by traditionally-employed methods may introduce significant bias, even to mislead the interpretation of LRD in the observed data. We believe that the tests proposed here to test experimentally the robustness of H estimators using artificial -but realistic- data, could be useful as a preliminary analysis to predict the behavior of any long-range dependence estimator on any specific data. Choosing the most appropriate method will indeed minimize the bias and imprecision of such a sensitive parameter, allowing its confident use as an index of interest in many research areas.

References

[1] H. C. Miller, F. C. Behrle, and N. W. Smull. Severe apnea and irregular respiratory rhythms among premature infants: A clinical and laboratory study. *Pediatrics*, 23(4):676–685, 1959.

[2] S. Saigal and L. W. Doyle. An overview of mortality and sequelae of preterm birth from infancy to adulthood. *The Lancet*, 371(9608):261–269, 2004.

[3] H. Rigatto and J. P. Brady. Periodic breathing and apnea in preterm infants. i. evidence for hypoventilation possibly due to central respiratory depression. *Pediatrics*, 50(2):202–218, 1972.

[4] P. D. Larsen, D. E. Elder, Y. C. Tzeng, A. J. Campbell, and D. C. Galletly. Fractal characteristics of breath to breath timing in sleeping infants. *Respiratory Physiology & Neurobiology*, 139(3), 2004.

[5] T. Nakamura, H. Horio, S. Miyashita, Y. Chiba, and S. Sato. Identification of development and autonomic nerve activity from heart rate variability in preterm infants. *Biosystems*, 79(1–3):117–124, 2005.

[6] J. B. Bassingthwaighte, L. S. Liebovitch, and B. J. West. *Fractal Physiology*. American Physiological Society, 1994.

[7] M. S. Taqqu, V. Teverovsky, and W. Willinger. Estimators for Long-Range dependence: An empirical study. *Fractals*, 3:785–798, 1995.

[8] W. Rea, L. Oxley, M. Reale, and J. Brown. Estimators for long range dependence: An empirical study. *arXiv:0901.0762*, 2009.

[9] B. Mandelbrot and J.W. Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968.

[10] G. Box, G. Jenkins, and G. Reinsel. *Time Series Analysis: Forecasting & Control (3rd Edition)*. Prentice Hall, 1994.

[11] M. S. Taqqu and V. Teverovsky. On estimating the intensity of Long-Range dependence in finite and infinite variance time series. In *A Practical Guide To Heavy Tails: Statistical Techniques and Applications*, pages 177–217, 1996.

[12] J. Barunik and L. Kristoufek. On hurst exponent estimation under heavy-tailed distributions. *Physica A: Statistical Mechanics and its Applications*, 389(18):3844–3855, 2010.

[13] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger. Mosaic organization of DNA nucleotides. *Physical Review E*, 49(2):1685–1689, 1994.

[14] P. Abry and D. Veitch. Wavelet analysis of long-range-dependent traffic. *Information Theory, IEEE Transactions on*, 44(1):2 –15, 1998.

[15] F. Esposti, M. Ferrario, and M. G. Signorini. A blind method for the estimation of the hurst exponent in time series: Theory and application. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 18(3):033126–033126–8, 2008.

[16] S. Stoev, M. S. Taqqu, C. Park, and J. S. Marron. On the wavelet spectrum diagnostic for hurst parameter estimation in the analysis of internet traffic. *Comput. Netw.*, 48(3):423–445, 2005.

[17] X. Navarro, A. Beuchée, F. Porée, and G. Carrault. Performance analysis of hurst's exponent estimators in higly immature breathing patterns of preterm infants. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 701–704. IEEE, 2011.

[18] D. Koutsoyiannis. Climate change, the hurst phenomenon, and hydrological statistics. *Hydrological Sciences Journal*, 48(1):3–24, 2003.

[19] B. Mandelbrot. *The Fractal Geometry of Nature*. W.H. Freeman, 1983.

[20] C. K. Peng, J. E. Mietus, Y. Liu, C. Lee, J. M. Hausdorff, H. E. Stanley, A. L. Goldberger, and L. A. Lipsitz. Quantifying fractal dynamics of human respiration: age and gender effects. *Annals of Biomedical Engineering*, 30(5):683–692, 2002.

[21] M. Cernelc, B. Suki, B. Reinmann, G. L. Hall, and U. Frey. Correlation properties of tidal volume and end-tidal O2 and CO2 concentrations in healthy infants. *Journal of Applied Physiology*, 92(5):1817–1827, 2002.

[22] D. Veitch, P. Abry, and M. S. Taqqu. On the automatic selection of the onset of scaling. *Fractals*, 11(2):377–390, 2003.

[23] H. Kunsch. Contrasts under Long-Range correlations. *The Annals of Statistics*, 21(2):943–964, 1993.

[24] J. Beran. *Statistics for Long-Memory Processes*. Chapman & Hall, 1994.

[25] H. Tyralis and D. Koutsoyiannis. Simultaneous estimation of the parameters of the Hurst–Kolmogorov stochastic process. *Stochastic Environmental Research and Risk Assessment*, 25(1):21–33, 2011.

[26] D. Lai. Estimating the hurst effect and its application in monitoring clinical trials. *Computational Statistics & Data Analysis*, 45(3):549–562, 2004.

[27] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. Doyne Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1-4):77–94, 1992.

[28] T. Nakamura and M. Small. Small-shuffle surrogate data: Testing for dynamics in fluctuating data with trends. *Physical Review E*, 72(5):056216, 2005.

[29] C. R. Dietrich and G. N. Newsam. Fast and exact simulation of stationary gaussian processes through circulant embedding of the covariance matrix. *SIAM J. Sci. Comput.*, 18(4):1088–1107, 1997.

[30] J. Aitchison and J. A. C. Brown. *The Lognormal Distribution.* Cambridge University Press, 1969.

[31] B. Mandelbrot. Possible refinement of the lognormal hypothesis concerning the distribution of energy dissipation in intermittent turbulence. In *Statistical Models and Turbulence*, volume 12 of *Lecture Notes in Physics*, pages 333–351. 1972.

[32] U. Frey, M. Silverman, A. L. Barabási, and B. Suki. Irregularities and power law distributions in the breathing pattern in preterm and term infants. *Journal of Applied Physiology (Bethesda, Md.: 1985)*, 85(3):789–797, 1998.

[33] M. J. Tobin, K. L. Yang, A. Jubran, and R. F. Lodato. Interrelationship of breath components in neighboring breaths of normal eupneic subjects. *American journal of respiratory and critical care medicine*, 152(6 Pt 1):1967–1976, 1995.

[34] L. Curzi-Dascalova, C. Gaudebout, and C. Dreyfus-Brisac. Respiratory frequencies of sleeping infants during the first months of life: correlations between values in different sleep states. *Early Human Development*, 5(1):39–54, 1981.

[35] A. Beuchée, E. Nsegbe, M. St Hilaire, G. Carrault, B. Branger, P. Pladys, and J-P. Praud. Prolonged dynamic changes in autonomic heart rate modulation induced by acid laryngeal stimulation in non-sedated lambs. *Neonatology*, 91(2):83–91, 2007.

[36] A. Dragomir, Y. Akay, A. K. Curran, and M. Akay. Investigating the complexity of respiratory patterns during the laryngeal chemoreflex. *Journal of NeuroEngineering and Rehabilitation*, 5:17, 2008.

[37] E. C. Eichenwald, M. Blackwell, J. S. Lloyd, T. Tran, R. E. Wilker, and D. K. Richardson. Inter-neonatal intensive care unit variation in discharge timing: influence of apnea and feeding management. *Pediatrics*, 108(4):928–933, 2001.

[38] A. Janvier, M. Khairy, A. Kokkotis, C. Cormier, D. Messmer, and K. J. Barrington. Apnea is associated with neurodevelopmental impairment in very low birth weight infants. *Journal of perinatology: official journal of the California Perinatal Association*, 24(12):763–768, 2004.

[39] N. S. Boghossian, G. P. Page, E. F. Bell, B. J. Stoll, J. C. Murray, C. M. Cotten, S. Shankaran, M. C. Walsh, A. R. Laptook, N. S. Newman, E. C. Hale, S. A. McDonald, A. Das, and R. D. Higgins. Late-onset sepsis in very low birth weight infants from singleton and multiple-gestation births. *The Journal of pediatrics*, 2013. PMID: 23324523.

[40] J. Bekhof, J. B. Reitsma, J. H. Kok, and I. H. Van Straaten. Clinical signs to identify late-onset sepsis in preterm infants. *European journal of pediatrics*, 2012.

[41] M. P. Griffin, D. E. Lake, and J. R. Moorman. Heart rate characteristics and laboratory tests in neonatal sepsis. *Pediatrics*, 115(4):937–941, 2005.

[42] A. Beuchée, G. Carrault, J. Y. Bansard, E. Boutaric, P. Bétrémieux, and P. Pladys. Uncorrelated randomness of the heart rate is associated with sepsis in sick premature infants. *Neonatology*, 96(2):109–114, 2009.

[43] G. Carrault, A. Beuchée, P. Pladys, L. Senhadji, and A. Hernandez. Time-frequency relationships between heart rate and respiration: A diagnosis tool for late onset sepsis

in sick premature infants. In *Computers in Cardiology, 2009*, pages 369–372. IEEE, 2009.

[44] J. R. Moorman, J. B. Delos, A. A. Flower, H. Cao, B. P. Kovatchev, J. S. Richman, and D. E. Lake. Cardiovascular oscillations at the bedside: early diagnosis of neonatal sepsis using heart rate characteristics monitoring. *Physiological measurement*, 32(11):1821–1832, 2011.

[45] V. E. Papaioannou, I. Chouvarda, N. Maglaveras, C. Dragoumanis, and I. Pneumatikos. Changes of heart and respiratory rate dynamics during weaning from mechanical ventilation: A study of physiologic complexity in surgical critically ill patients. *Journal of Critical Care*, 26(3):262–272, 2011.

[46] P. B. DePetrillo, d'A. Speers, and U. E. Ruttimann. Determining the hurst exponent of fractal time series and its application to electrocardiographic analysis. *Computers in Biology and Medicine*, 29(6):393–406, 1999.

7 Figure captions

Figure 1: a) Example of a pre-processed respiratory signal corresponding to a patient born after 32 weeks of gestation. b) One-minute excerpt showing apnea and irregular breaths, with detected minima (o) and maxima (*). c) Definition of inspiration, expiration and total times within a cycle. d) Resulting IBI signal with the horizontal line marking the apnea threshold $(3 * \overline{t_{tot}})$. e) Empirical distribution of the IBI signal, whose fitted lognormal parameters are $\mu = 0.29$ and $\sigma = 0.43$.

Figure 2: Examples of fLn realizations with several coefficients of variation $(cv = \sigma_G/\mu_G)$ at $H = 0.6, 0.7$ and 0.8. a) $cv = 0.3$ b) $cv = 0.7$ c) $cv = 1$.

Figure 3: Illustration of the discrepancy within the estimations of the Hurst exponent $(\widehat{H_r})$. Regular IBIs (a) showed in general the highest LRD, followed by erratic (b) and periodic (c) patterns. In all cases, $\widehat{H_r}$ in periodic IBIs presented statistically significant differences $(p < 0.05)$ with regard to regular and erratic breathing.

Figure 4: Average Mutual Information (AMI) computed for the first 10 lags in original IBI series (upper panels) and small-shuffled surrogates (lower panels). There were significantly significant differences in the first three breaths for regular breathing (a), in the first two breaths for erratic (b) and in the first breath for erratic patterns (c) after performing a Mann-Whitney U-test $(*, p < 0.001; **, p < 0.005; ***, p < 0.05)$.

Figure 5: First and second columns: Estimation error from surrogates (ϵ_{sur}) in RSS $(H_s = 0.5)$ and SSS $(H_s \approx H_r)$. Third column: Averaged estimation error $(\overline{\epsilon_{sur}})$ for HAS prescribing H_s $(0.6, ...0.9)$. Results for regular (a), erratic (b) and periodic (c) IBIs.

Figure 6: Comparison of the mean duration (D_{Ap}) and the number of apneas per hour (N_{Ap}) obtained by simulated IBIs (grey lines) and from real signals (regular IBIs in triangles, erratic in circles and periodic in diamonds). These plots were obtained by simulated IBIs at $H = 0.6$, but similar curves were obtained with values between 0.55 and 0.9.

Figure 7: Performance of estimators in 100 realizations of fLn. The median of the simulation error (ϵ_{sim}) in solid lines and 75% and 95% confidence intervals in semi-dotted (-.-) and discontinuous (- -) lines, respectively. Curves for regular breathing in (a), erratic in (b) and periodic in (c).

Table 1

Table 1: Description and some characteristics of the selected signals of each pattern. μ_L and σ_L correspond to the mean and standard deviation of the lognormal law, respectively.

	Regular	Erratic	Periodic
GA (weeks)	31.9 ± 1.5	31.2 ± 0.9	31.1 ± 0.5
Weight (g)	1338 ± 353	983 ± 249	1061 ± 149
$\overline{t_{tot}}$ (s)	$0.93 \pm 0.17^{*}$	$1.29 \pm 0.22^{\S}$	$1.39 \pm 0.18^{*\S}$
$std(t_{tot})$	$0.29 \pm 0.08^{*}$	$0.93 \pm 0.25^{*\S}$	$1.58 \pm 0.56^{*\S}$
N_{Ap} (num/h)	$10.6 \pm 7.70^{*}$	$68.2 \pm 19.5^{*\S}$	$122.9 \pm 29.7^{*\S}$
D_{Ap} (s)	$2.26 \pm 1.42^{*}$	$5.58 \pm 1.31^{*\dagger}$	$7.20 \pm 1.65^{*\dagger}$
μ_L	$-0.12 \pm 0.17^{\dagger\ddagger}$	$0.15 \pm 0.22^{\dagger}$	$0.02 \pm 0.14^{\ddagger}$
σ_L	$0.25 \pm 0.06^{*}$	$0.46 \pm 0.02^{*\S}$	$0.58 \pm 0.08^{*\S}$

$^{*,\,\S}$ paired groups with $p < 0.01$
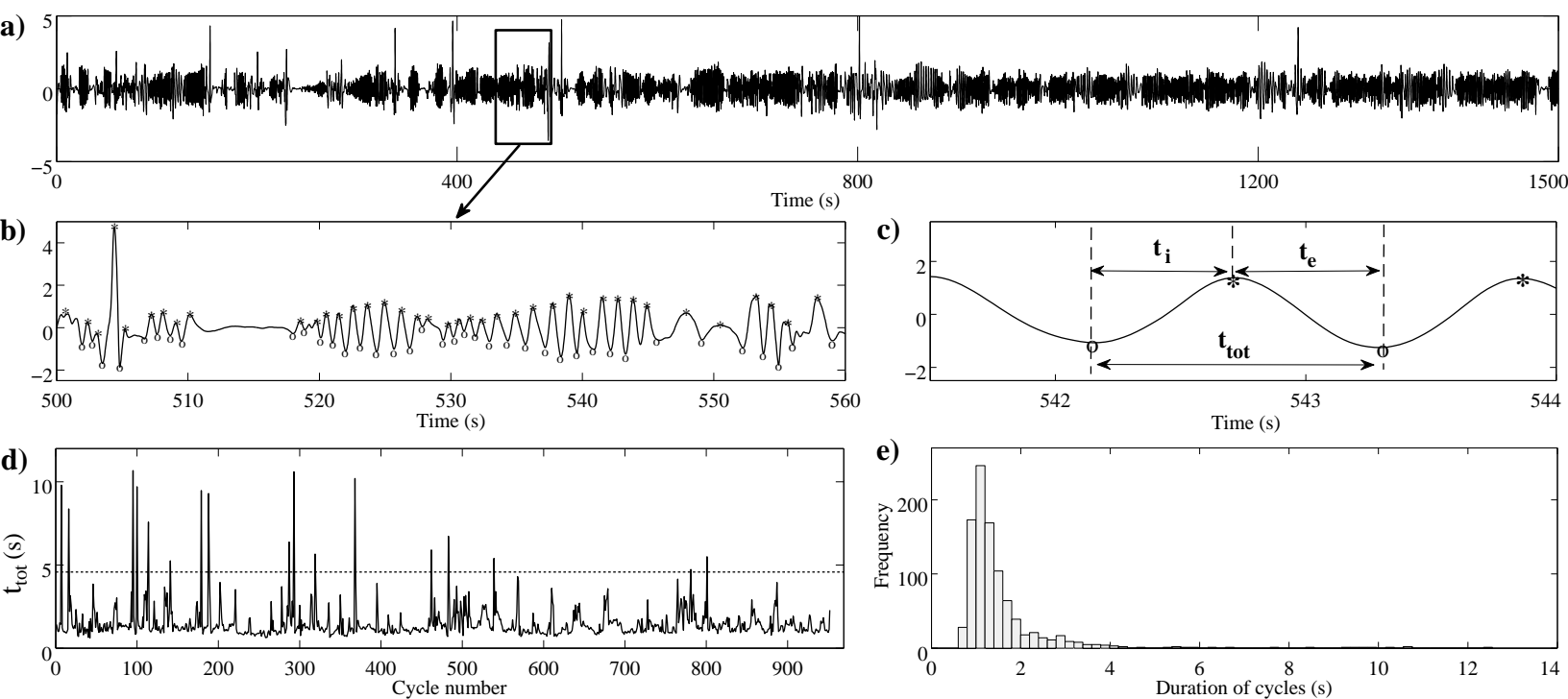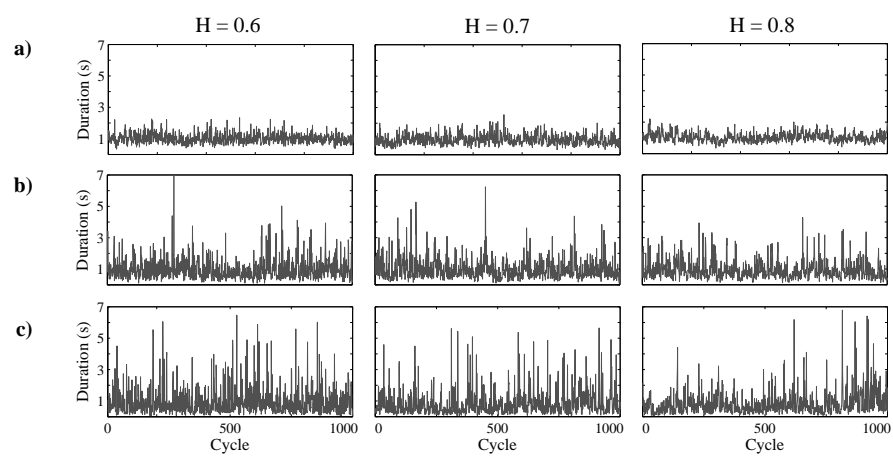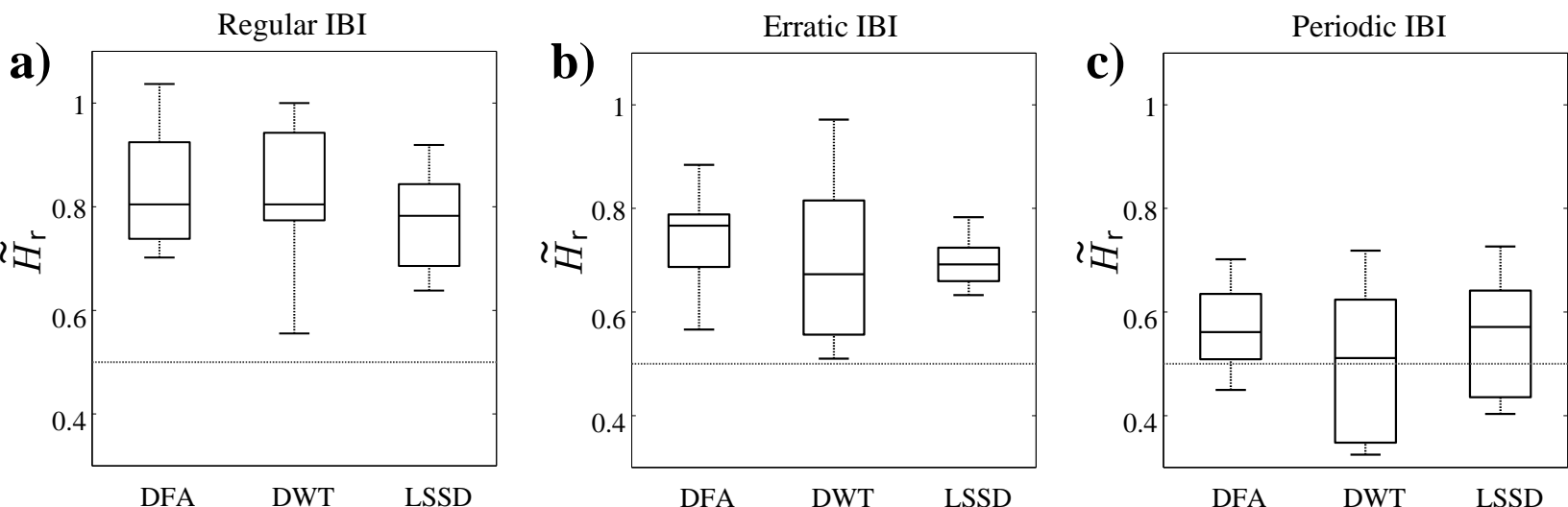$^{\dagger,\,\ddagger}$ paired groups with $p < 0.05$

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7

DFA   DWT   LSSD

a)

ϵ_{sim}

b)

ϵ_{sim}

c)

ϵ_{sim}

H   H   H