

Research Article

Methodological Framework for Estimating the Correlation Dimension in HRV Signals

Juan Bolea,^{1,2} Pablo Laguna,^{1,2} José María Remartínez,^{3,4} Eva Rovira,^{3,4}
Augusto Navarro,^{3,4,5} and Raquel Bailón^{1,2}

¹ Communications Technology Group (GTC), Aragón Institute for Engineering Research (I3A), IIS Aragón, University of Zaragoza, 50018 Zaragoza, Spain

² CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), 50018 Zaragoza, Spain

³ Anaesthesiology Service, Miguel Servet University Hospital, 50009 Zaragoza, Spain

⁴ Medicine School, University of Zaragoza, 50009 Zaragoza, Spain

⁵ Aragón Health Sciences Institute (IACS), 50009 Zaragoza, Spain

Correspondence should be addressed to Juan Bolea; jbolea@unizar.es

Received 28 August 2013; Revised 17 December 2013; Accepted 20 December 2013; Published 30 January 2014

Academic Editor: Mika P. Tarvainen

Copyright © 2014 Juan Bolea et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a methodological framework for robust estimation of the correlation dimension in HRV signals. It includes (i) a fast algorithm for on-line computation of correlation sums; (ii) log-log curves fitting to a sigmoidal function for robust maximum slope estimation discarding the estimation according to fitting requirements; (iii) three different approaches for linear region slope estimation based on latter point; and (iv) exponential fitting for robust estimation of saturation level of slope series with increasing embedded dimension to finally obtain the correlation dimension estimate. Each approach for slope estimation leads to a correlation dimension estimate, called \hat{D}_2 , $\hat{D}_{2(\perp)}$, and $\hat{D}_{2(\max)}$. \hat{D}_2 and $\hat{D}_{2(\max)}$ estimate the theoretical value of correlation dimension for the Lorenz attractor with relative error of 4%, and $\hat{D}_{2(\perp)}$ with 1%. The three approaches are applied to HRV signals of pregnant women before spinal anesthesia for cesarean delivery in order to identify patients at risk for hypotension. \hat{D}_2 keeps the 81% of accuracy previously described in the literature while $\hat{D}_{2(\perp)}$ and $\hat{D}_{2(\max)}$ approaches reach 91% of accuracy in the same database.

1. Introduction

Heart rate variability (HRV) has been widely used as a marker of the autonomic nervous system (ANS) regulation of the heart. Classical HRV indices include global descriptive statistics which characterize HRV distribution in the time domain (mean heart rate and standard deviation of the normal-to-normal beat interval, among others) and in the frequency domain (power in the very low frequency, and low frequency (LF) and high frequency (HF) bands). The activity of the two main branches of the ANS, sympathetic and parasympathetic systems, has been related with the power in the LF and HF bands, respectively [1].

HRV data often present nonlinear characteristics, possibly reflecting intrinsic physiological nonlinearities, such as changes in the gain of baroreflex feedback loops or delays

in conduction time, which are not properly described by classical HRV indices.

The most widespread methods used to characterize nonlinear system dynamics are based on chaos theory. The question of whether HRV arises from a low-dimensional attractor associated with a deterministic nonlinear dynamical system or whether it has a stochastic origin is still under debate.

Of great interest is the concept of system complexity, which refers to the richness of process dynamics. Complexity measures are based on the theory of nonlinear systems but may be applied to both linear and nonlinear systems. Several techniques attempting to assess complexity have been developed such as detrended fluctuation analysis [2], Lempel-Ziv complexity [3], Lyapunov exponents [4], the correlation dimension (D_2) [5], and approximate and sample entropies [6].

The reduction of HRV complexity has been associated with age, disease, and unbalanced cardiovascular regulation [7]. Complexity measures have been proven to characterize HRV signals more successfully than linear approaches in certain applications [8]. In [9–11] point correlation dimension of HRV signals predicted hypotension events in pregnant women during spinal anesthesia for cesarean section, something which time and frequency domain indices were unable to do.

While these measurements are of considerable interest, their application to HRV has some pitfalls that could mislead their interpretation. One of such limitations arises from their application to limited time series. Correlation dimension estimation is highly dependent on the length of the time series [12]. Several studies have reported the effect of data length on D_2 estimation, as well as proposals to alleviate this effect [13, 14]. Stationarity is another requirement that a time series has to fulfil to obtain reliable results. However, satisfying the constraint of finite series and stationarity at the same time is usually difficult [15]. Yet another limitation of these measurements is the long computational time required. Data length exponentially increases the computational time cost in a classical sequential approach. In the case of D_2 , several attempts to try to reduce this factor have been reported by Widman et al. [16] and Zurek et al. [17], the latter proposing parallel computing using MPI (Message Passing Interface).

The main goal of this study is to propose a methodological framework for robust and fast estimation of D_2 and its application in HRV signals. Section 2 starts with a definition of the correlation dimension and its classical estimation. An algorithm for its fast computation is proposed. Robustness is addressed by fitting the log-log curve to a sigmoid function after which three alternative approaches for D_2 estimation are presented. Section 3 introduces synthetic and real (HRV signals) data where the proposed estimates are evaluated and interpreted. Section 4 presents the results while Section 5 sets out the discussion and conclusions of the study.

2. Methods

2.1. Correlation Dimension. Let $x(n)$, $n = 1, \dots, N$ be the time series of interest, which in HRV analysis will be the RR interval series normalized to unit amplitude, with N being the total number of beats. A set of m -dimensional vectors, $\mathbf{y}_m(i)$, called reconstructed vectors, are generated [18]:

$$\mathbf{y}_i^m = [x(i), x(i + \tau), x(i + 2\tau), \dots, x(i + (m - 1)\tau)]^T, \quad (1)$$

where τ represents the delay between consecutive samples in the reconstructed space. Then, the amount of reconstructed vectors is $N_m = N - \tau(m - 1)$ for each m -dimension. The distance between each pair of reconstructed vectors, $\mathbf{y}_i^m, \mathbf{y}_j^m$, is denoted as

$$d_{i,j}^m = d(\mathbf{y}_i^m, \mathbf{y}_j^m) \quad (2)$$

and this can be computed as the norm of the difference vector $\Delta \mathbf{y}_{i,j}^m = \mathbf{y}_i^m - \mathbf{y}_j^m$. (In Appendix A different norms and their effect on correlation dimension estimates from finite time

series are discussed.) The correlation sum which represents the probability of the reconstructed vector pair distance being smaller than a certain threshold r is computed as

$$\begin{aligned} C_m(r) &= \frac{1}{N_m(N_m - 1)} \sum_{i,j=1}^{N_m} H(r - d_{i,j}^m) \\ &= \frac{1}{N_m(N_m - 1)} \sum_{i=1}^{N_m} c_i^m(r), \end{aligned} \quad (3)$$

where $H(\cdot)$ is the Heaviside function defined as:

$$H(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0, \end{cases} \quad (4)$$

and $c_i^m(r) = \sum_{j=1}^{N_m} H(r - d_{i,j}^m)$.

For deterministic systems, $C_m(r)$ decreases monotonically to 0 as r approaches 0, and it is expected that $C_m(r)$ is well approximated by $C_m(r) \approx r^{D_2^m}$. Thus, D_2^m can be defined as

$$D_2^m = \lim_{r \rightarrow 0} \frac{\log C_m(r)}{\log(r)}. \quad (5)$$

For increasing m , D_2^m values tend to saturate to a value D_2 which constitutes the final correlation dimension estimate.

2.2. Fast Computation of Correlation Sums. One important limitation of D_2 estimation is the long computational time required mainly due to the sequential estimation of correlation sums. This section describes an algorithm for the fast computation of correlation sums based on matrix operations (MO). A matrix \mathbf{S} which contains the differences between all pairs of samples of $x(n)$ is computed as

$$\mathbf{S} = \mathbf{X} - \mathbf{X}^T, \quad (6)$$

where \mathbf{X} is the $N \times N$ matrix:

$$\mathbf{X} = \begin{pmatrix} x(1) & x(2) & \cdots & x(N) \\ x(1) & x(2) & \cdots & x(N) \\ x(1) & x(2) & \cdots & x(N) \\ \vdots & \vdots & \ddots & \vdots \\ x(1) & x(2) & \cdots & x(N) \end{pmatrix}, \quad (7)$$

$$\mathbf{S} = \begin{pmatrix} x_{1,1} & \cdots & x_{i,1} & x_{i+1,1} & \cdots & x_{i+m-1,1} & \cdots & x_{N,1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1,j} & \cdots & x_{i,j} & x_{i+1,j} & \cdots & x_{i+m-1,j} & \cdots & x_{N,j} \\ x_{1,j+1} & \cdots & x_{i,j+1} & x_{i+1,j+1} & \cdots & x_{i+m-1,j+1} & \cdots & x_{N,j+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1,j+m-1} & \cdots & x_{i,j+m-1} & x_{i+1,j+m-1} & \cdots & x_{i+m-1,j+m-1} & \cdots & x_{N,j+m-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1,N} & \cdots & x_{i,N} & x_{i+1,N} & \cdots & x_{i+m-1,N} & \cdots & x_{N,N} \end{pmatrix}, \quad (8)$$

where $x_{i,j}$ symbolizes $x(i) - x(j)$. For instance, the dashed box contains the elements of the difference vector $\Delta y_{i,j}^m$ for $\tau = 1$. For each embedded dimension m and the reconstructed vector i , the difference vectors $\Delta y_{i,j}^m$ generates a S_i^m matrix:

$$S_i^m = \begin{pmatrix} x_{i,1} & x_{i+1,2} & \cdots & x_{i+m-1,m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{i,j} & x_{i+1,j+1} & \cdots & x_{i+m-1,j+m-1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{i,N_m} & x_{i+1,N_m+1} & \cdots & x_{i+m-1,N_m} \end{pmatrix} = \begin{pmatrix} \Delta y_{i,1}^m \\ \vdots \\ \Delta y_{i,j}^m \\ \vdots \\ \Delta y_{i,N_m}^m \end{pmatrix}^T. \quad (9)$$

The selected norm is applied to the matrix S_i^m , generating the norm vector \mathbf{d}_i^m , whose elements are distances $\mathbf{d}_{i,j}^m$. To compute the limit in (5), distances should be compared with a set of thresholds, which implies the repetition of the whole process as many times as the number of thresholds. This repetition is avoided since distances in \mathbf{d}_i^m are compared with a whole set of thresholds $\mathbf{r} = [r_1, r_2, \dots, r_{N_r}]$:

$$\Gamma_i^m = \begin{pmatrix} H(r_1 - d_{i,1}^m) & H(r_2 - d_{i,1}^m) & \cdots & H(r_{N_r} - d_{i,1}^m) \\ H(r_1 - d_{i,2}^m) & H(r_2 - d_{i,2}^m) & \cdots & H(r_{N_r} - d_{i,2}^m) \\ \vdots & \vdots & \vdots & \vdots \\ H(r_1 - d_{i,N_m}^m) & H(r_2 - d_{i,N_m}^m) & \cdots & H(r_{N_r} - d_{i,N_m}^m) \end{pmatrix}, \quad (10)$$

where Γ_i^m is a $N_m \times N_r$ matrix, which contains ones and zeros. The accumulative addition of each column represents the partial correlation sum of the i th reconstructed vector for a set of thresholds \mathbf{r} :

$$\mathbf{c}_i^m = \Gamma_i^{mT} \mathbf{1} = \begin{pmatrix} c_i^m(r_1) \\ c_i^m(r_2) \\ \vdots \\ c_i^m(r_{N_r}) \end{pmatrix}, \quad (11)$$

where $\mathbf{1}$ is a $N_m \times 1$ vector whose elements are equal to one.

Finally, the procedure has to be repeated varying the i index N_m times to compute $C_m(r)$. This technique saves computational time due to the usage of a set of thresholds in one step.

2.3. New Approaches for D_2 Assessment

2.3.1. Sigmoid Curves as Surrogates of Log-Log Curves. D_2 has to be estimated from (5) whose numerator and denominator both tend to $-\infty$ as r tends to 0. Therefore, applying L'Hôpital's rule the equation can be rewritten as [19]

$$D_2^m = \lim_{r \rightarrow 0} \frac{\partial \log C_m(r)}{\partial \log(r)}. \quad (12)$$

Since the size of the time series is finite, choosing small values of r to evaluate this limit is problematic. For values of r close to 0, very few distances contribute to the correlation sum, making the estimation unreliable. The evaluation of this

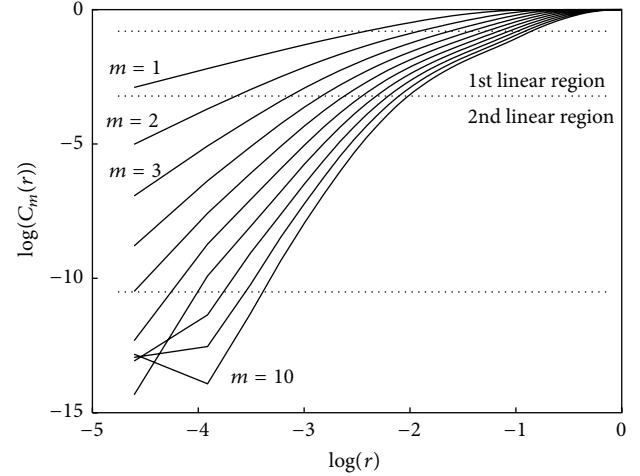


FIGURE 1: Log-log curves for a dynamic system. Data correspond to an RR interval series extracted from 30 mins of ECG recording.

expression is usually done in a linear region in the $\log(C_m(r))$ versus $\log(r)$ representation, called the log-log curve. The slope of this linear region is considered an estimate of D_2^m . There are different approaches for estimating this slope. Maximum slope searching can be done by directly computing the increments in the log-log curve. Another approach is to estimate numerically the maximum of the first derivative of the log-log curve. Nevertheless these approaches encounter some limitations due to the usual nonequidistant sampling of r values in the logarithmic scale. Yet another limitation arises in the presence of dynamic systems whose log-log curves display several linear regions, as can be seen in Figure 1 where the data corresponds to an RR interval series extracted from a 30 minute ECG recording. In order to estimate the slope of the linear region of the log-log curve, an attempt to artificially extend the linear region is made by excluding the self-comparisons ($d_{i,i}$) from the correlation sums.

However, the basis of the approach proposed in this work to improve D_2 estimation lies in considering self-comparisons. Figure 2 illustrates how log-log curves behave in both situations, considering or not considering self-comparisons. As it is shown, both share part of the linear region. Our proposal is to use sigmoidal curve fitting (SCF) over the log-log curves to obtain an analytic function whose maximum slope in the linear region is well defined. These log-log curves are reminiscent of the biasymptotic fractals studied by Rigaut [20] and Dollinger et al. in [21] in which exponential fittings were proposed. The sigmoidal fitting is applied to the interpolated log-log curves computed with evenly spaced r values.

A modified Boltzmann sigmoid curve was used by Navarro-Verdugo et al. [22] as a model for the phase transition of smart gels:

$$f(x) = A_2 - \frac{(A_2 - A_1)}{B + e^{(-x+x_0)/\alpha}}, \quad (13)$$

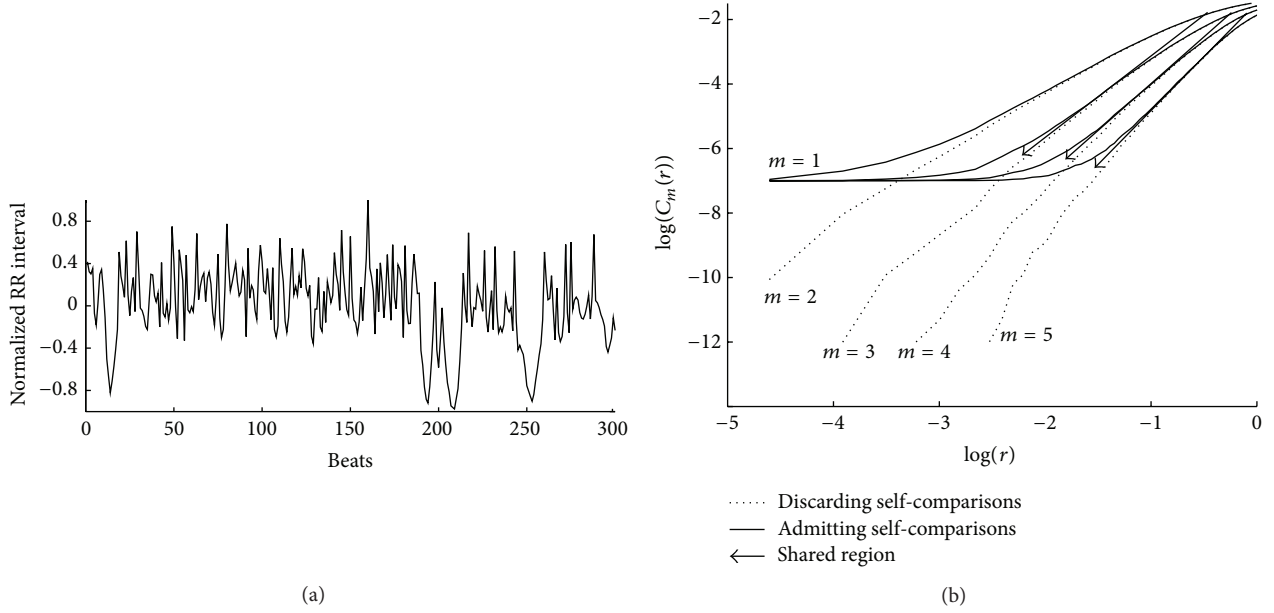


FIGURE 2: Log-log curves discarding and accepting self-comparisons. Arrows show the slope of the scaling range. Data correspond to an RR interval series of 300 beats.

where A_1, A_2, x_o , and B are the parameter designed. The first derivative of $f(x)$ is

$$\frac{df(x)}{dx} = -\frac{(A_2 - A_1)e^{-(x+x_o)/\alpha}}{[\alpha(B + e^{-(x+x_o)/\alpha})^2]}. \quad (14)$$

In our study the sigmoid curve $f(x)$ is fitted to log-log curve. The first derivative, (14), is determined analytically and its maximum constitutes the slope of the linear range, that is, \widehat{D}_2^m . Note that hat notation refers the use of SCF.

In order to achieve a good fitting, the thresholds, r , have to guarantee that both asymptotes are reached. In this work $r \in [0.01 \ 3]$ with a step of 0.01. The upper asymptote is reached when all comparisons are above the threshold, $C_m(r) \approx 1$, and the lower when only the self-comparison is below, $C_m(r) = 1/N_m(N_m - 1)$.

The SCF approach is robust in the presence of dynamic systems which exhibit log-log curves with more than one linear region since when the fitting is not good enough, no estimate is given. In this work, the requirement for a good fitting is to achieve a regression factor greater than 0.8.

As the embedding dimension m increases, the linear regions of the log-log curves tend to be parallel to each other. Thus, \widehat{D}_2^m estimates tend to saturate to a certain value, which is considered the correlation dimension \widehat{D}_2 . The correlation dimension is estimated fitting the \widehat{D}_2^m versus m curve following a modified version of that used by Carvajal et al. [23]:

$$D_2^m = D_2(1 - Ae^{-km}), \quad (15)$$

where A , introduced in this study in order to reach the saturation level more quickly than previously proposed, and k are exponential growth factors.

2.3.2. New Approaches for D_2 Estimation. As mentioned previously, we chose \widehat{D}_2^m as the maximum slope on each fitted sigmoid curve. Nevertheless the linear range is composed of more than one point. Instead of considering only one point per curve, in this study we propose a new approach for D_2 estimation considering a set of points extracted from these linear ranges.

The proposed strategy is based on selecting one point of the linear range in the SCF log-log curve of the lowest embedded dimension m and moving forward to the next embedded dimension $m + 1$, selecting the point of the corresponding SCF log-log curve with minimum distance to the former curve (i.e. where the perpendicular to the m th log-log curve intersects the $(m + 1)$ th log-log curve, as in the gradient descent technique); see Figure 3. The procedure is repeated up to the maximum embedded dimension analyzed. Then, several sets of slopes are computed (one for each point in the linear region around the maximum slope of the SCF log-log curve of the lowest embedded dimension) providing a set of correlation dimension estimates per embedded dimension ($\widehat{D}_{2(\perp),r}^m$). The dependence on r in the notation indicates that each set of correlation dimension estimates is linked to an r value, corresponding to the first value of each set.

Finally, (15) is used to estimate the final correlation dimension ($\widehat{D}_{2(\perp),r}$) for each set of points. These ($\widehat{D}_{2(\perp),r}$) estimates are linked to the $\log(r)$ value of the lowest embedded dimension. Finally, the maximum of the $\widehat{D}_{2(\perp),r}$ is selected as the new D_2 estimate, called $\widehat{D}_{2(\perp)}$.

Another new approach for D_2 estimation based on sample entropy (SampEn) is now presented. SampEn was defined by Zurek et al. [17] as

$$\text{SampEn}(m, r, N) = \log(C_m(r)) - \log(C_{m+1}(r)), \quad (16)$$

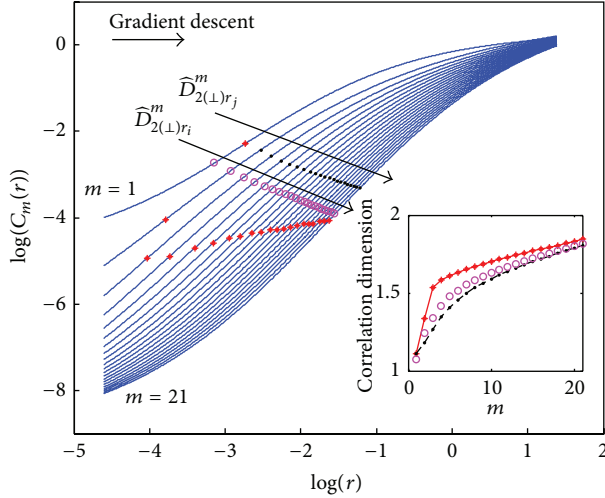


FIGURE 3: Maximum slope points are marked with crosses over fitted sigmoid curves. Points calculated using gradient descent criteria from two starting points are shown in dots and circles. r_j is the point which corresponds to the maximum slope in the lowest embedded dimension. The inset illustrates the correlation dimension estimation of the three sets of points.

where, in this case, $C_m(r)$ is computed as in (3), but without considering self-comparisons. Let us define $\text{SampEn}_{i=j}(m, r)$ as the sample entropy considering self-pairs, which is easily computed for all embedded dimensions m and a huge set of thresholds r using the fast algorithm described in Section 2.2. We can generate a $\text{SampEn}_{i=j}(m, r)$ surface from the fitted sigmoid curves, as can be seen in Figure 4, an example of a 300-beat RR interval series extracted from one recording of the database used in [10]. For each embedded dimension, the value of r which maximizes $\text{SampEn}_{i=j}(m, r)$ is used to evaluate the slope of the linear region of the SCF log-log curves, $\widehat{D}_{2(\max)}^m$, yielding another D_2 estimate, called in this paper $\widehat{D}_{2(\max)}^m$.

3. Materials

The selected time series chosen to validate the approaches proposed in this paper to estimate D_2 are the Lorenz attractor, the $\text{MIX}(P)$ process, and real HRV signals, respectively.

Lorenz Attractor. The Lorenz system is described by three coupled first order differential equations whose solution exhibits a chaotic behaviour system for certain parameter values and initial conditions. This is called the Lorenz attractor:

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x), \\ \frac{dy}{dt} &= \rho x - y - xz, \\ \frac{dz}{dt} &= -\beta z + xy. \end{aligned} \quad (17)$$

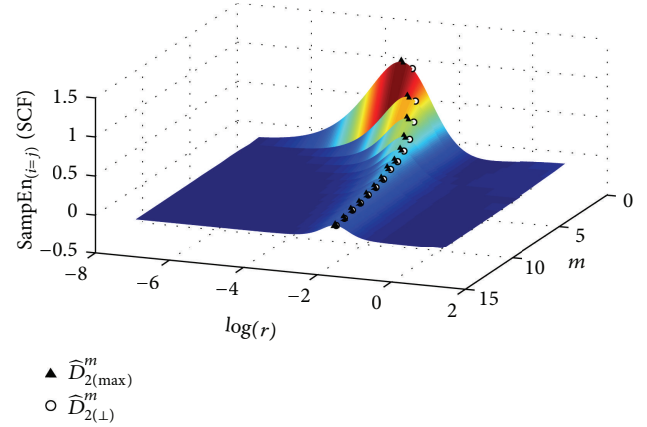


FIGURE 4: $\text{SampEn}_{i=j}(m, r)$ surface for a 300-beat RR interval series. For each embedded dimension maximum point is marked with solid triangle. Circles correspond to the r values which define $\widehat{D}_{2(L)}^m$.

For parameter values $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$ the theoretical D_2 value is 2.02 [24]. In this study, the system equations are discretized with a time step of 0.01.

MIX(P) Signals. $\text{MIX}(P)$ is a family of stochastic processes that samples a sine for $P = 0$ and becomes more random as P increases ($P = 1$ completely random) [5] following the expression

$$\text{MIX}(P)_j = (1 - Z_j) X_j + Z_j Y_j, \quad (18)$$

where $X_j = \sqrt{2} \sin(2\pi j/12)$, $Y_j \equiv i.i.d$ uniform random variables on $[-\sqrt{3}, \sqrt{3}]$, and $Z_j \equiv i.i.d$ random variables, with $Z_j = 1$ with probability P , and $Z_j = 0$ with probability $1 - P$. MIX indicates a mixture of deterministic and stochastic components.

HRV Signals. The HRV representation used in this study is the time difference between the occurrences of consecutive normal heart beats, the so-called RR interval. Ectopic beats as well as missed and false detections introduce some extra variability in the RR interval series which is not representative of the ANS activity. Thus, they were detected and corrected [25]. The RR interval series analysed in this study belongs to a database recorded at the Miguel Servet University Hospital in Zaragoza (Spain). That database was used to predict hypotension events during spinal anesthesia in elective cesarean delivery by HRV analysis [10]. It consists of ECG signals from 11 women with programmed cesarean section recorded at a 1000 Hz sampling frequency immediately before the cesarean surgery. Five of them suffered a hypotension event during the surgery (Hyp) and 6 did not (NoHyp). The series analysed correspond to 5 minutes in a lateral decubitus position. See [10] for further database details.

4. Results

All the results presented in this section are computed using the ℓ_∞ -norm. The effect of different norms in D_2 estimation is discussed in Appendix A.

TABLE 1: Computational time of correlation sums estimated for Lorenz attractor series of different sample lengths. S_p is the speed-up achieved and defined as $S_p = T_{\text{Seq}}/T_{\text{MO}}$, where T_{Seq} is the time demand for a sequential algorithm and T_{MO} the time demand for the proposed technique based on matrix operations.

N (samples)	T_{Seq} (s)	T_{MO} (s)	S_p
300	1086	0.9	≈ 1200
5000	$8.69e5$	50	≈ 16000
10000	$3.63e6$	300	≈ 12000

The computational time cost of the correlation sums depends on the length of the data, the maximum embedded dimension considered, and the amount of thresholds used. The results shown in Table 1 correspond to computational time required for different data lengths of Lorenz attractor series, a maximum embedded dimension of 16, and a set of 300 threshold values r evenly spaced from 0.01 to 3. The computational time required for a sequential approach is denoted by T_{Seq} whereas the time required for the proposed technique, based on matrix operations, is denoted by T_{MO} . This allows defining the speed up achieved by the novel approach as the ratio between both measurements, $S_p = T_{\text{Seq}}/T_{\text{MO}}$. As it is shown in Table 1, S_p increases with data length. For a 300-sample data (usual length for a 5 minute RR interval series) correlation sums are estimated in approximately 1 s.

The Lorenz attractor series was used to validate the new proposed methodologies. Figure 5(a) displays the SCF log-log curves for embedded dimensions m from 1 to 10. The sets of points where the slope is evaluated according to (15) are displayed for different starting points. For each starting point, the corresponding set of points is selected following a gradient descent technique. Figure 5(b) shows the slope estimate ($\widehat{D}_{2(\perp),r}^m$) versus m for each starting point. Figure 5(c) displays the correlation dimension estimate ($\widehat{D}_{2(\perp),r}$) versus $\log(r)$ for each starting point. The maximum ($\widehat{D}_{2(\perp),r}$) constitutes the novel D_2 estimate ($\widehat{D}_{2(\perp)}$).

Table 2 displays correlation dimension estimates using the different approaches presented in this study. Note that although the three approaches give results close to the theoretical value of Lorenz attractor correlation dimension, the $\widehat{D}_{2(\perp)}$ approach is the closest one. Relative errors for approaches \widehat{D}_2 and $\widehat{D}_{2(\text{max})}$ are above 4%, while for $\widehat{D}_{2(\perp)}$ it is just 1%; D_2 estimated as described in [10] is also included for comparison purposes.

\widehat{D}_2 was applied to a set of MIX series with different P values (0.1, 0.4, and 0.8). These estimates can be considered as measures of the randomness of the signals when these signals are finite stochastic processes; see Figure 6.

The same database for HRV analysis as in [10] was used. The results shown in Table 2 are divided into hypotension and nonhypotension groups. The approaches proposed in this paper were applied as well as the classical correlation dimension estimate described in [10] included for comparison purposes. The distribution of the data was found to be not normally distributed by the Kolmogorov-Smirnov test,

TABLE 2: D_2 estimated by different approaches for Lorenz attractor series (5000 samples) and HRV signals (300 samples). Data expressed as median | interquartile range.

	Lorenz		HRV	
		Hyp	NoHyp	P value
D_2 [10]	1.93	5.92 0.57	4.39 1.26	0.028
\widehat{D}_2	1.93	5.94 0.60	4.81 0.79	0.028
$\widehat{D}_{2(\perp)}$	2.01	6.41 0.65	5.11 0.77	0.028
$\widehat{D}_{2(\text{max})}$	1.93	5.88 0.49	4.83 0.70	0.010

TABLE 3: ROC area for the analysis of all studied correlation dimension estimates for the database used. Accuracy, sensibility, and specificity estimated with the correspondent cut points are expressed in percentage.

	ROC area	Acc.	Sen.	Spe.
D_2 [10]	0.900	81.8	71.4	100
\widehat{D}_2	0.900	81.8	71.4	100
$\widehat{D}_{2(\perp)}$	0.900	90.9	100	85.7
$\widehat{D}_{2(\text{max})}$	0.967	90.9	83.3	100

and therefore the Mann-Whitney U test was applied to evaluate their statistical differences in medians. The differences between both groups for all estimates were found to be statistically significant with a P value lower than 0.03. In order to evaluate the discriminant power of the proposed measures, ROC analysis was performed. Area of the ROC curve, accuracy, sensitivity, and specificity for all the proposed approaches and the classical D_2 estimate used in [10] were displayed in Table 3. The proposed \widehat{D}_2 estimates maintain the accuracy achieved in [10] while the techniques based on the SampEn surface and the gradient descent actually increase it.

5. Discussion and Conclusion

In this paper a methodological framework has been proposed to compute the correlation dimension (D_2) of a limited time series such as HRV signals which includes fast computation of the correlation sums, sigmoidal curve fitting of log-log curves, three approaches for estimating the slope of the linear region, and exponential fitting of the \widehat{D}_2^m versus m curves.

One important limitation for the application of D_2 to HRV analysis is the long computational time required for the correlation sums. In an attempt to solve this problem, an algorithm has been proposed based on matrix operations. In [17] another approach was described based on parallel computing which decreased the time demand. Nevertheless, the computational times achieved in the present work were obtained with a regular computer (Windows 7 based PC, Intel Core i7 3.5 GHz, 16 Gb RAM with Matlab R2011a). As an example, for a signal of 300 sample length (a usual length in typical 5 min HRV analysis, ≈ 300 beats), the time demand was reduced with respect to the sequential approach from 18 minutes to 1 second, which allows the online computation of D_2 in clinical practice. Computational time required for the proposed approaches is discussed in Appendix B.

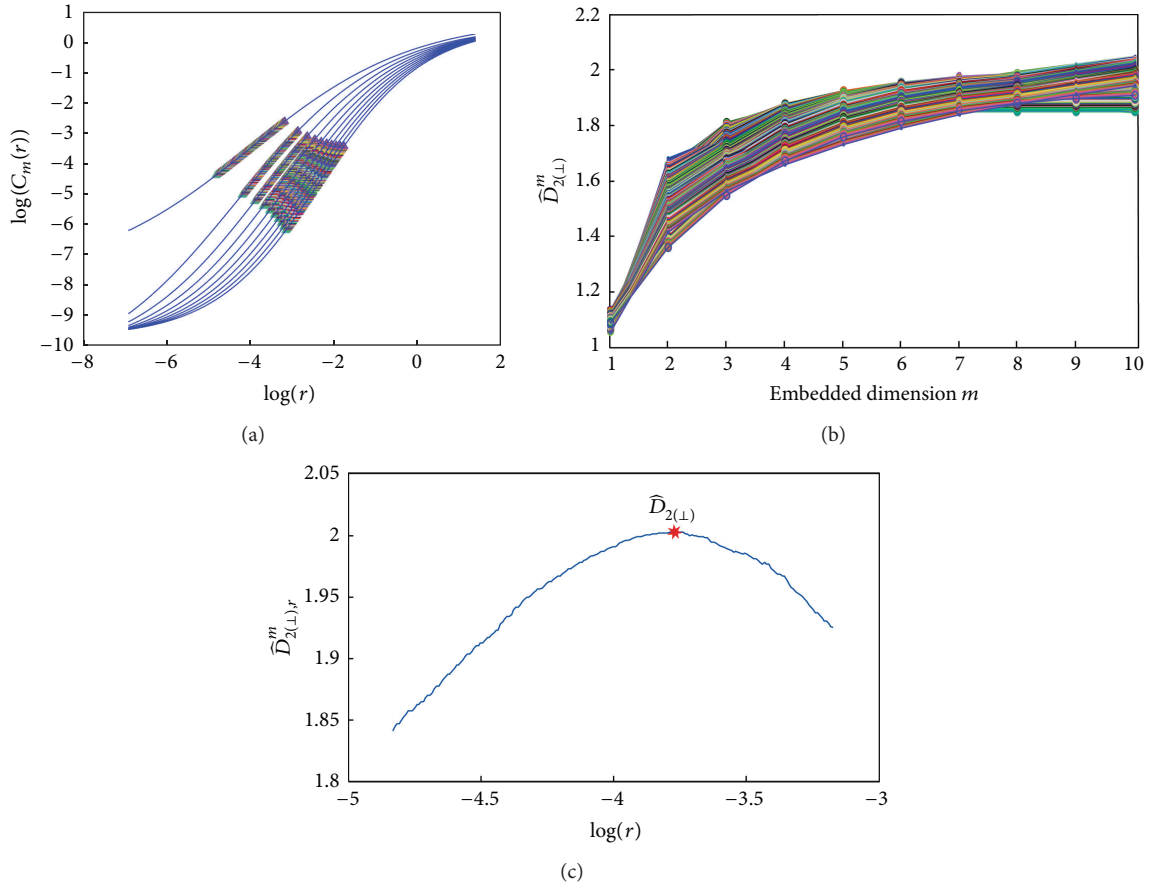


FIGURE 5: (a) Set of points where slope is estimated from the fitted sigmoid curves in the approach proposed in Section 2.3. (b) Set of $\widehat{D}_{2(L),r}^m$ estimates for different starting points versus embedded dimensions are fitted by the exponential equation (15). (c) Correlation dimension estimate for each set corresponding to different starting points. Data extracted from Lorenz attractor of 5000-sample length.

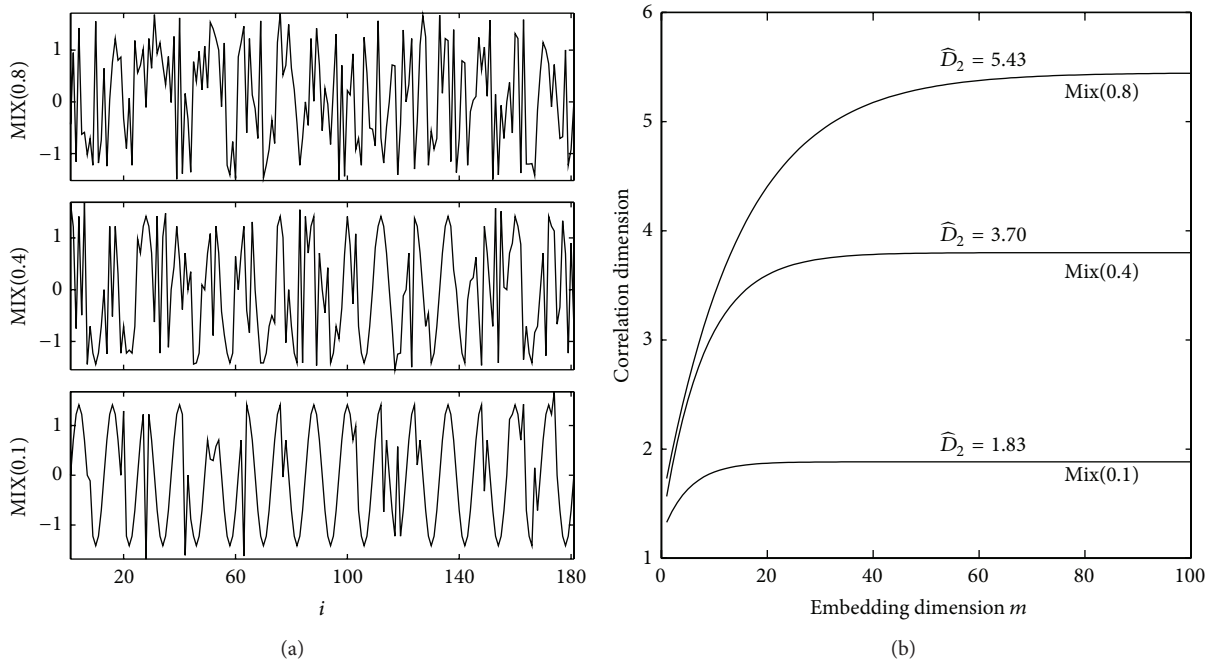


FIGURE 6: MIX signals with different degrees of randomness and the effects on the estimation of the \widehat{D}_2 .

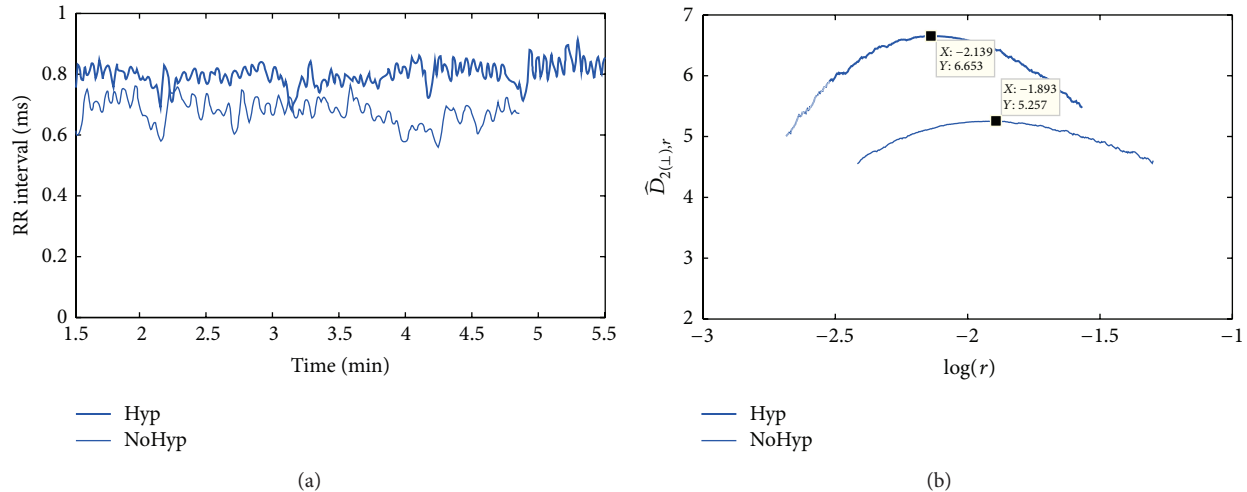


FIGURE 7: The left panel shows two RR intervals, one corresponding to a patient who developed a hypotension event (Hyp) and the other to one who did not (NoHyp); the right panel shows the $\widehat{D}_{2(L),r}$ estimation using the perpendicular points in the log-log curves.

Another limitation of the D_2 estimate is its reliability. One of the system characteristics that can lead to an unreliable measurement of D_2 is the nonstationarity of the data. Several techniques attempting to characterize these dynamic systems have been reported, mainly focused on changing the τ parameter or even taking into account the time between the vectors [13, 26, 27]. Searching the linear region of the log-log curves becomes a difficult challenge when the system is nonstationary since more than one linear region can appear and classical D_2 estimate is unreliable in those cases. The SCF approach is more robust since it does not give any estimate if the fitting is not good enough.

The novel approaches proposed in this study for the estimation of D_2 use the SCF approach. $\widehat{D}_{2(L)}$ exploits the fact that the linear region of the log-log curves is almost parallel for high embedded dimensions. This allows a set of points surrounding the maximum slope point to be considered, and therefore several correlation dimension estimates are obtained for these starting points. $\widehat{D}_{2(\max)}$ is based on the differences between two consecutive log-log curves that define the $\text{SampEn}_{i=j}$ surface. This surface showed maximum values for each embedded dimension, m , and a specific threshold, r , providing another estimation of the correlation dimension. $\widehat{D}_{2(L)}$ was found to be the closest to the theoretical correlation dimension value for the Lorenz attractor series with 5000 size points and for $h = 0.01$, with a relative error of 1%, while \widehat{D}_2 and $\widehat{D}_{2(\max)}$ obtained a relative error of 4% with the same data.

The correlation dimension is known to be a surrogate of the fractal dimension of a chaotic attractor [12]. However, when applied to limited time series, nonzero finite correlation dimension values do not imply the existence of an underlying chaotic attractor. For example, when applied to MIX processes, nonzero finite D_2 values were obtained, higher for more random processes. Thus, although D_2 cannot be interpreted as the fractal dimension of an underlying chaotic attractor, it still gives a measure of the complexity of the process at least regarding its unpredictability.

Thus, the D_2 estimate in HRV signals may shed light on the degree of complexity of the ANS or how many degrees of freedom it has. The group of women (Hyp) suffering hypotension events occurring during the surgery of a programmed cesarean section under spinal anesthesia showed higher D_2 values than the group who did not (NoHyp), in the lateral decubitus position. As an example Figure 7 shows one patient of each group and the $\widehat{D}_{2(L)}$ estimate. All the proposed correlation dimension estimates not only maintain the accuracy obtained in [10], they also increase it. Predicting hypotension is a challenge since it occurs in the 60% of the cases producing fetal stress [28]. If the goal is to predict those who are going to suffer hypotension, then the estimates that performed 100% of specificity will be selected, being D_2 [10], \widehat{D}_2 , and $\widehat{D}_{2(\max)}$. Otherwise, if the goal is to use prophylaxis in the less number of patients to prevent hypotension, then the estimates that performed 100% of sensitivity will be chosen, and in this case it is $\widehat{D}_{2(L)}$. The effect of prophylaxis on patients who finally are not going to suffer a hypotension event and the relation with fetal stress needs further studies.

The contribution of this paper to the field is the proposal of a methodological framework for a reliable estimation of the correlation dimension from a limited time series, such as HRV signals, avoiding or at least alleviating the misleading interpretations that can be made from classical correlation dimension estimates. The computational speed-up achieved may allow this framework to be considered for monitoring in clinical practice. Nevertheless, the main limitation for the application of these methodologies to HRV analysis lies in its relation with the underlying physiology, which is still unclear and needs further studies. In spite of the fact that the framework proposed in this paper is focused on the characterization of HRV signals, its applicability could be extended to a wide range of fields. However, an evaluation would be needed to ascertain whether the proposed approaches are appropriate in each particular case.

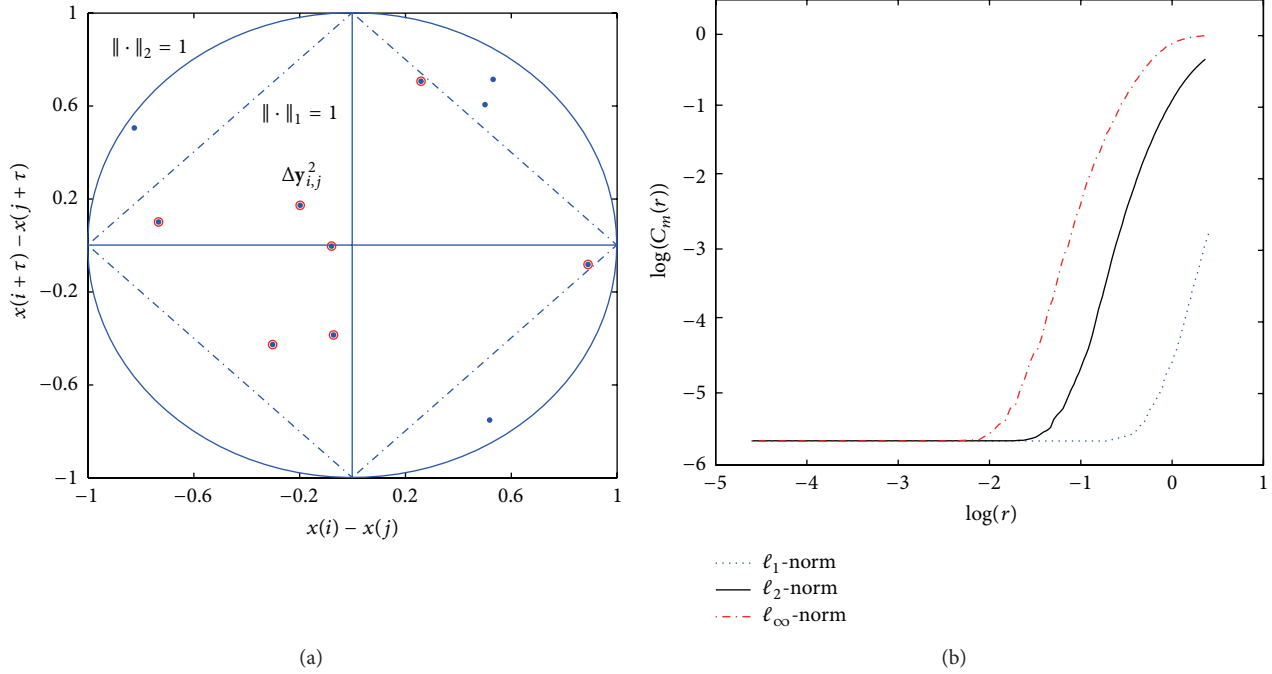


FIGURE 8: In the left panel, vector differences of any two reconstructed vectors are shown (i.e., for $m = 2$, $\Delta y_{i,j}^2$), where solid circles and dashed lines represent the points whose ℓ_2 -norm and ℓ_1 -norm are equal to 1, respectively. The dots are the differences below ℓ_2 -norm unity and the dots with circles are below ℓ_1 -norm unity. In the right panel log-log curves of one HRV signal (300 samples) used in the study are shown for a $m = 10$ and ℓ_1 -, ℓ_2 - and ℓ_∞ -norms.

Appendices

A. Use of Norms and Thresholds in Correlation Dimension Estimates

The correlation dimension is considered norm invariant [14]. However, the effect of selecting the norm in correlation dimension estimates deserves further attention when applied to a finite data set. The norm of the difference vector $\Delta y_{i,j}^m$ defines the distance $d_{i,j}^m$ in (2). Norms can be defined from ℓ_1 ($\|\cdot\|_1$) to ℓ_∞ ($\|\cdot\|_\infty$). Left panel in Figure 8 shows norm unity for ℓ_1 and for ℓ_2 . Moreover, it is illustrated how a distance $d_{i,j}^2$ can be lower than the norm unity or not depending on which norm is used. The norm unity is chosen as an example of any threshold r used in the correlation dimension algorithm. Therefore, by fixing the set of thresholds, the appearance of the linear region of the log-log curve can be compromised.

In Figure 8 the right panel shows how the application of different norms shifts the log-log curves losing the entire linear region in some cases due to the fixed range of thresholds. The range of these thresholds should be long enough to ensure that the linear regions are contained therein; thus, the election of the norm compromises the set of thresholds used.

In the SCF approach it is particularly important that the two asymptotic regions should be represented in the log-log curve. Therefore, the correct selection of the norm and the range of the set of thresholds are critical to assure the goodness of the SCF approach. Table 4 shows the correlation dimension estimates for 5000 data length of Lorenz attractor

TABLE 4: Correlation dimension estimates for the different proposed approaches, using different norms for Lorenz attractor series (5000 samples) using different norms.

	Lorenz attractor		
	ℓ_1	ℓ_2	ℓ_∞
D_2 [10]	1.95	1.94	1.93
\bar{D}_2	1.69	1.70	1.93
$\bar{D}_{2(\perp)}$	1.84	1.74	2.01
$\bar{D}_{2(\max)}$	1.99	1.71	1.93

series. The effect of different norms is reflected in the estimates since the set of thresholds was fixed. As it is shown, the application of ℓ_∞ -norm, combined with the fixed set of thresholds used, achieves closest values with respect to the theoretical correlation dimension value for Lorenz attractor, 2.01.

B. Computational Time Demand of Novel Approaches to Correlation Dimension Estimates

In Section 2.2 a new technique based on matrix operations (MO) was introduced in order to compute correlation sums which represent the core of the correlation dimension algorithm. Nevertheless, in the paper no computational time cost was considered for the new proposed approaches for correlation dimension estimates. Table 5 shows the time required

TABLE 5: Computational time cost for correlation dimension estimates by all proposed approaches considering Lorenz attractor series and HRV signals in which ℓ_∞ -norm was applied. Data expressed as mean \pm standard deviation.

	Lorenz (5000 samples)	HRV (300 samples)
T_{D_2} (s) [10]	$(8.86 \pm 0.35)e5$	3314 ± 180
$T_{\widehat{D}_2}$ (s)	194 ± 29	4.66 ± 0.58
$T_{\widehat{D}_{2(\perp)}}(s)$	260 ± 38	214 ± 30
$T_{\widehat{D}_{2(\max)}}(s)$	194 ± 29	4.30 ± 0.56

for the correlation dimension estimates including that used in [10].

10 realizations of Lorenz attractor series were generated whose initial conditions were randomly chosen. It is noticeable that the time cost of $\widehat{D}_{2(\perp)}$ is higher compared to the others in both cases, the Lorenz series and the HRV signals (11 subjects), since it uses several sets of slope estimated to compute correlation dimension. Furthermore, the ratio between them is higher for the HRV signals than for the Lorenz series. Each of the different sets of thresholds is associated with an r value in an interval centred on the maximum slope for $m = 1$. This interval is defined as a decrease in 50% of the amplitude of the maximum in the SCF first derivative. The more abrupt the transition zone in the sigmoid, the lower the amount of starting points. Thus, each realization is done with a different number of points, varying the computational time.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

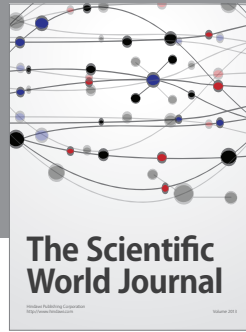
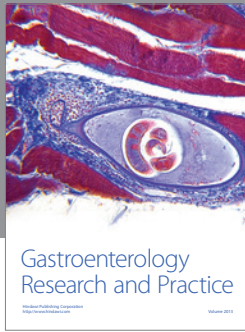
Acknowledgments

This work has been supported by the Ministerio de Ciencia e Innovación, Spain, and FEDER under Project TEC2010-21703-C03-02, and by ISCIII, Spain, through Project PI10/02851 (FIS).

References

- [1] Task Force of the European Society of Cardiology and The North American Society of Pacing and Electrophysiology, "Heart rate variability. Standards of measurement, physiological interpretation, and clinical use," *European Heart Journal*, vol. 17, no. 3, pp. 354–381, 1996.
- [2] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley, "Multifractal detrended fluctuation analysis of nonstationary time series," *Physica A*, vol. 316, no. 1–4, pp. 87–114, 2002.
- [3] A. Lempel and J. Ziv, "On the complexity of finite sequences," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 75–81, 1976.
- [4] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series," *Physica D*, vol. 16, no. 3, pp. 285–317, 1985.
- [5] S. M. Pincus and A. L. Goldberger, "Physiological time-series analysis: what does regularity quantify?" *American Journal of Physiology—Heart and Circulatory Physiology*, vol. 266, no. 4, pp. H1643–H1656, 1994.
- [6] S. Pincus and B. H. Singer, "Randomness and degrees of irregularity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 5, pp. 2083–2088, 1996.
- [7] A. Porta, S. Guzzetti, R. Furlan, T. Gneccchi-Rusccone, N. Montano, and A. Malliani, "Complexity and nonlinearity in short-term heart period variability: comparison of methods based on local nonlinear prediction," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 1, pp. 94–106, 2007.
- [8] S. Cerutti, G. Carrault, P. J. M. Cluitmans et al., "Non-linear algorithms for processing biological signals," *Computer Methods and Programs in Biomedicine*, vol. 51, no. 1–2, pp. 51–73, 1996.
- [9] D. Chamchad, V. A. Arkoosh, J. C. Horrow et al., "Using heart rate variability to stratify risk of obstetric patients undergoing spinal anesthesia," *Anesthesia & Analgesia*, vol. 99, no. 6, pp. 1818–1821, 2004.
- [10] L. Canga, A. Navarro, J. Bolea, J. M. Remartínez, P. Laguna, and R. Bailón, "Non-linear analysis of heart rate variability and its application to predict hypotension during spinal anesthesia for cesarean delivery," in *Proceedings of the Computing in Cardiology (CinC '12)*, pp. 413–416, Kraków, Poland, September 2012.
- [11] J. Bolea, R. Bailón, E. Rovira, J. M. Remartínez, P. Laguna, and A. Navarro, "Heart rate variability in pregnant women before programmed cesarean intervention," in *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*, M. L. R. Romero, Ed., vol. 41 of *IFMBE Proceedings*, pp. 710–713, Springer International, 2014.
- [12] P. Grassberger and I. Procaccia, "Characterization of strange attractors," *Physical Review Letters*, vol. 50, no. 5, pp. 346–349, 1983.
- [13] J. Theiler, "Spurious dimension from correlation algorithms applied to limited time-series data," *Physical Review A*, vol. 34, no. 3, pp. 2427–2432, 1986.
- [14] J. Theiler, "Estimating fractal dimension," *Journal of the Optical Society of America A*, vol. 7, no. 6, pp. 1055–1073, 1990.
- [15] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, UK, 2004.
- [16] G. Widman, K. Lehnertz, P. Jansen, W. Meyer, W. Burr, and C. E. Elger, "A fast general purpose algorithm for the computation of auto- and cross-correlation integrals from single channel data," *Physica D*, vol. 121, no. 1–2, pp. 65–74, 1998.
- [17] S. Zurek, P. Guzik, S. Pawlak, M. Kosmider, and J. Piskorski, "On the relation between correlation dimension, approximate entropy and sample entropy parameters, and a fast algorithm for their calculation," *Physica A*, vol. 391, no. 24, pp. 6601–6610, 2012.
- [18] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, D. Rand and L.-S. Young, Eds., vol. 898 of *Lecture Notes in Mathematics*, pp. 366–381, Springer, Berlin, Germany, 1981.
- [19] J. A. Lee and M. Verleysen, *Nonlinear Dimensionality Reduction*, Springer, Berlin, Germany, 2007.
- [20] J. P. Rigaut, "An empirical formulation relating boundary lengths to resolution in specimens showing "non-ideally fractal" dimensions," *Journal of Microscopy*, vol. 133, no. 1, pp. 41–54, 1984.

- [21] J. W. Dollinger, R. Metzler, and T. F. Nonnenmacher, “Bi-asymptotic fractals: fractals between lower and upper bounds,” *Journal of Physics A*, vol. 31, no. 16, pp. 3839–3847, 1998.
- [22] A. L. Navarro-Verdugo, F. M. Goycoolea, G. Romero-Meléndez, I. Higuera-Ciapara, and W. Argüelles-Monal, “A modified Boltzmann sigmoidal model for the phase transition of smart gels,” *Soft Matter*, vol. 7, no. 12, pp. 5847–5853, 2011.
- [23] R. Carvajal, M. Vallverdú, R. Baranowski, E. Orłowska-Baranowska, J. J. Zebrowski, and P. Caminal, “Dynamical non-linear analysis of heart rate variability in patients with aortic stenosis,” in *Proceedings of the Computers in Cardiology*, pp. 449–452, September 2002.
- [24] E. N. Lorenz, “Deterministic non-periodic flow,” *Journal of Atmospheric Science*, vol. 20, no. 2, pp. 130–141, 1963.
- [25] J. Mateo and P. Laguna, “Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal,” *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 3, pp. 334–343, 2003.
- [26] L. A. Aguirre, “A nonlinear correlation function for selecting the delay time in dynamical reconstructions,” *Physics Letters A*, vol. 203, no. 2-3, pp. 88–94, 1995.
- [27] H. S. Kim, R. Eykholt, and J. D. Salas, “Nonlinear dynamics, delay times, and embedding windows,” *Physica D*, vol. 127, no. 1-2, pp. 48–60, 1999.
- [28] A. M. Cyna, M. Andrew, R. S. Emmett, P. Middleton, and S. W. Simmons, “Techniques for preventing hypotension during spinal anaesthesia for caesarean section,” *Cochrane Database of Systematic Reviews*, vol. 18, no. 4, Article ID CD002251, 2006.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

