Multiscale entropy analysis of biological signals

Madalena Costa,[1,2] Ary L. Goldberger,[1] and C.-K. Peng[1]
[1]*Margret and H. A. Rey Institute for Nonlinear Dynamics in Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA*
[2]*Institute of Biophysics and Biomedical Engineering, Faculty of Sciences of the University of Lisbon, Campo Grande, 1749-016 Lisbon, Portugal*
(Received 1 July 2004; published 18 February 2005)

Traditional approaches to measuring the complexity of biological signals fail to account for the multiple time scales inherent in such time series. These algorithms have yielded contradictory findings when applied to real-world datasets obtained in health and disease states. We describe in detail the basis and implementation of the multiscale entropy (MSE) method. We extend and elaborate previous findings showing its applicability to the fluctuations of the human heartbeat under physiologic and pathologic conditions. The method consistently indicates a loss of complexity with aging, with an erratic cardiac arrhythmia (atrial fibrillation), and with a life-threatening syndrome (congestive heart failure). Further, these different conditions have distinct MSE curve profiles, suggesting diagnostic uses. The results support a general "complexity-loss" theory of aging and disease. We also apply the method to the analysis of coding and noncoding DNA sequences and find that the latter have higher multiscale entropy, consistent with the emerging view that so-called "junk DNA" sequences contain important biological information.

DOI: 10.1103/PhysRevE.71.021906 PACS number(s): 87.19.Hh, 05.40.Ca, 05.45.Tp

I. INTRODUCTION

Physiologic systems are regulated by interacting mechanisms that operate across multiple spatial and temporal scales. The output variables of these systems often exhibit complex fluctuations that are not simply due to "contaminative" noise but contain information about the underlying dynamics.

Two classical approaches to time series analysis are related to deterministic and stochastic mechanisms. A fundamental underpinning of the former approach is Takens' theorem [1,2], which states that it is possible to reach full knowledge of a high dimensional deterministic system by observing a single output variable. However, since experimental time series, even when generated by deterministic mechanisms, are most likely affected by dynamical noise, the purely deterministic approach may be of limited use. Nevertheless, for some practical applications, a low dimensional dynamics may be assumed and then the results tested for internal consistency [3].

The stochastic approach is aimed at quantifying the statistical properties of the output variables and developing tractable models that account for those properties. The diffusion process is a classic example of how a stochastic approach may contribute to the understanding of a dynamical system. At a "macroscopic" level, the diffusion equation can be derived from Fick's law and the principle of conservation of mass. Alternatively, at a "microscopic" level it is possible to derive the diffusion equation assuming that each particle can be modeled as a random walker, taking steps of length l in a given direction with probability p. The theory of Brownian motion, which is based on random walk models, together with experimental results, contributed to the understanding of the atomic nature of matter.

Time series generated by biological systems most likely contain deterministic and stochastic components. Therefore, both approaches may provide complementary information about the underlying dynamics. The method we use in this paper for the analysis of physiologic time series does not assume any particular mechanism. Instead, our method is aimed at comparing the degree of complexity of different time series. Such complexity-related metrics [4] have potentially important applications to discriminate time series generated either by different systems or by the same system under different conditions.

Traditional methods quantify the degree of regularity of a time series by evaluating the appearance of repetitive patterns. However, there is no straightforward correspondence between regularity, which can be measured by entropy-based algorithms, and complexity. Intuitively, complexity is associated with "meaningful structural richness" [5], which, in contrast to the outputs of random phenomena, exhibits relatively higher regularity. Entropy-based measures, such as the entropy rate and the Kolmogorov complexity, grow monotonically with the degree of randomness. Therefore, these measures assign the highest values to uncorrelated random signals (white noise), which are highly unpredictable but not structurally "complex," and, at a global level, admit a very simple description.

Thus, when applied to physiologic time series, traditional entropy-based algorithms may lead to misleading results. For example, they assign higher entropy values to certain pathologic cardiac rhythms that generate erratic outputs than to healthy cardiac rhythms that are exquisitely regulated by multiple interacting control mechanisms. Substantial attention, therefore, has been focused on defining a quantitative measurement of complexity that assigns minimum values to both deterministic/predictable and uncorrelated random/unpredictable signals [6]. However, no consensus has been reached on this issue.

Our approach to addressing this long-standing problem has been motivated by three basic hypotheses: (i) the com-

plexity of a biological system reflects its ability to adapt and function in an ever-changing environment; (ii) biological systems need to operate across multiple spatial and temporal scales, and hence their complexity is also multiscaled; and (iii) a wide class of disease states, as well as aging, which reduce the adaptive capacity of the individual, appear to degrade the information carried by output variables. Thus, loss of complexity may be a generic feature of pathologic dynamics. Accordingly, our approach to defining a complexity measurement focuses on quantifying the information expressed by the physiologic dynamics over multiple scales.

Recently, we introduced a new method, termed multiscale entropy (MSE) [7–11]. Due to the interrelationship of entropy and scale, which is incorporated in the MSE analysis, the results are consistent with the consideration that both completely ordered and completely random signals are not really complex. In particular, the MSE method shows that correlated random signals (colored noise) are more complex than uncorrelated random signals (white noise). Compared to traditional complexity measures, MSE has the advantage of being applicable to both physiologic and physical signals of finite length.

In this paper, we apply the MSE method to the study of (i) the cardiac interbeat interval time series, the output of a major physiologic system regulated by the involuntary autonomic nervous system; and (ii) biological codes. First, we seek to characterize changes in the complexity of cardiac dynamics due to aging and disease, during both wake and sleeping periods. This analysis is a major extension of our previous work [7] that focused on application of MSE to a more limited database. In addition, we address the question of applying the MSE method to binary sequences in order to study the complexity of coding versus noncoding human DNA sequences.

The structure of the paper is as follows. In Sec. II we provide the mathematical background for calculating the entropy rate and discuss its physical meaning. We also present a short description of the approximate entropy (A_E) and the sample entropy (S_E) algorithms, which have been widely used in the analysis of short, noisy physiologic time series. In Sec. III, we review the MSE method, which incorporates the S_E statistics, and discuss the results of applying the MSE method to white and $1/f$ noises. The analytical calculations of S_E for both types of noises are presented in Appendix A. In Sec. IV, we apply the MSE method to a cardiac interbeat interval database comprising recordings of healthy subjects, subjects with atrial fibrillation, an erratic cardiac arrhythmia, and subjects with congestive heart failure. We address the question of quantifying the information in MSE curves for possible clinical use. We further discuss the effects of outliers, white noise superimposed on a physiologic time series, and finite sample frequency values in Appendix B. In Sec. V, we apply the MSE method to binary sequences of artificial and biological codes, aimed at quantifying the complexity of coding and noncoding DNA sequences. Technical aspects of applying the MSE method to such discrete sequences are described in Appendix C. Section VI presents conclusions.

II. BACKGROUND

The entropy $H(X)$ of a single discrete random variable X is a measure of its average uncertainty. Shannon's entropy [12] is calculated by the equation

$$H(X) = -\sum_{x_i \in \Theta} p(x_i)\log p(x_i) = -E[\log p(x_i)], \tag{1}$$

where X represents a random variable with a set of values Θ and probability mass function $p(x_i)=P_r\{X=x_i\}$, $x_i \in \Theta$, and E represents the expectation operator. Note that $p \log p=0$ if $p=0$.

For a time series representing the output of a stochastic process, that is, an indexed sequence of n random variables, $\{X_i\}=\{X_1,\ldots,X_n\}$, with a set of values $\Theta_1,\ldots,\Theta_n$, respectively, and $X_i \in \Theta_i$, the joint entropy is defined as

$$H_n = H(X_1,X_2,\ldots,X_n) = -\sum_{x_1 \in \Theta_1} \cdots \sum_{x_n \in \Theta_n} p(x_1,\ldots,x_n)\log p(x_1,\ldots,x_n), \tag{2}$$

where $p(x_1,\ldots,x_n)=\mathrm{P_r}\{X_1=x_1,\ldots,X_n=x_n\}$ is the joint probability for the n variables $X_1,\ldots,X_n$.

By applying the chain rule to Eq. (2), the joint entropy can be written as a summation of conditional entropies, each of which is a non-negative quantity,

$$H_n = \sum_{i=1}^{n} H(X_i|X_{i-1},\ldots,X_1). \tag{3}$$

Therefore, one concludes that the joint entropy is an increasing function of n.

The rate at which the joint entropy grows with n, i.e., the entropy rate h, is defined as

$$h = \lim_{n\to\infty} \frac{H_n}{n}. \tag{4}$$

For stationary ergodic processes, the evaluation of the rate of entropy has proved to be a very useful parameter [2,5,13–17].

Let us consider a $\mathcal{D}$-dimensional dynamical system. Suppose that the phase space of the system is partitioned into hypercubes of content $\varepsilon^{\mathcal{D}}$ and that the state of the system is measured at intervals of time δ. Let $p(k_1,k_2,\ldots,k_n)$ denote the joint probability that the state of the system is in the hypercube k_1 at $t=\delta$, in the k_2 at $t=2\delta$, and in the hypercube k_n at $t=n\delta$. The Kolmogorov-Sinai (KS) entropy is defined as

$$H_{KS} = -\lim_{\delta\to 0}\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n\delta}\sum_{k_1,\ldots,k_n} p(k_1,\ldots,k_n)\log p(k_1,\ldots,k_n) = \lim_{\delta\to 0}\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n\delta}H_n. \tag{5}$$

For stationary processes [18], it can be shown that

$$\lim_{n\to\infty}\frac{H_n}{n} = \lim_{n\to\infty} H(X_n|X_{n-1},\ldots,X_1). \tag{6}$$

Then, by the chain rule, it is straightforward to show that

$$H_{KS} = \lim_{\delta \to 0} \lim_{\varepsilon \to 0} \lim_{n \to \infty} (H_{n+1} - H_n). \quad (7)$$

The state of a system at a certain instant t_i is partially determined by its history, $t_1, t_2, \ldots, t_{i-1}$. However, each new state carries an additional amount of new information. The KS entropy measures the mean rate of creation of information, in other words, the decrease of uncertainty at a receiver by knowing the current state of the system given the past history.

Numerically, only entropies of finite order n can be computed. As soon as n becomes large with respect to the length of a given time series, the entropy H_n is underestimated and decays toward zero. Therefore, Eq. (7) is of limited use to estimate the entropy of finite length "real-world" time series. However, several formulas have been proposed in an attempt to estimate the KS entropy with reasonable precision. Grassberger and Procaccia [15] suggested characterizing chaotic signals by calculating the K_2 entropy, which is a lower bound of the KS entropy.

Let $\{X_i\}=\{x_1, \ldots, x_i, \ldots, x_N\}$ represent a time series of length N. Consider the m-length vectors: $u_m(i) = \{x_i, x_{i+1}, \ldots, x_{i+m-1}\}$, $1 \leqslant i \leqslant N-m+1$. Let $n_i^m(r)$ represent the number of vectors $u_m(j)$ that are close to the vector $u_m(i)$, i.e., the number of vectors that satisfy $d[u_m(i), u_m(j)] \leqslant r$, where d is the Euclidean distance. $C_i^m(r) = n_i^m(r)/(N-m+1)$ represents the probability that any vector $u_m(j)$ is close to the vector $u_m(i)$. The average of the C_i^m, $C^m(r) = 1/(N-m+1)\Sigma_{i=1}^{(N-m+1)} C_i^m(r)$, represents the probability that any two vectors are within r of each other. K_2 is defined as

$$K_2 = \lim_{N \to \infty} \lim_{m \to \infty} \lim_{r \to 0} -\ln[C^{m+1}(r) - C^m(r)]. \quad (8)$$

Following the same nomenclature, Eckmann and Ruelle (ER) [2] defined the function $\Phi^m(r) = 1/(N-m+1)\Sigma_{i=1}^{N-m+1} \ln C_i^m(r)$, considering the distance between two vectors as the maximum absolute difference between their components: $d[u_m(i), u_m(j)] = \max\{|x(i+k) - x(j+k)| : 0 \leqslant k \leqslant m-1\}$. Note that $\Phi^{m+1}(r) - \Phi^m(r) \approx \Sigma_{i=1}^{N-m+1} \ln[C_i^m(r)/C_i^{m+1}(r)]$, represents the average of the natural logarithm of the conditional probability that sequences that are close to each other for m consecutive data points will still be close to each other when one more point is known. Therefore, Eckmann and Ruelle suggested calculating the KS entropy as

$$H_{\mathrm{ER}} = \lim_{N \to \infty} \lim_{m \to \infty} \lim_{r \to 0} [\Phi^m(r) - \Phi^{m+1}(r)]. \quad (9)$$

Although this formula has been useful in classifying low-dimensional chaotic systems, it does not apply to experimental data since the result is infinity for a process with superimposed noise of any magnitude [19]. For the analysis of short and noisy time series, Pincus [17] introduced a family of measures termed approximate entropy, $A_E(m,r)$, defined as

$$A_E(m,r) = \lim_{N \to \infty} [\Phi^m(r) - \Phi^{m+1}(r)]. \quad (10)$$

A_E is estimated by the statistics,

$$A_E(m,r,N) = \Phi^m(r) - \Phi^{m+1}(r). \quad (11)$$

A_E was not intended as an approximate value of ER entropy. Rather, A_E is a regularity statistic. It applies to "real-world" time series and, therefore, has been widely used in physiology and medicine [4]. Lower A_E values are assigned to more regular time series while higher A_E values are assigned to more irregular, less predictable, time series.

Recently, a modification of the A_E algorithm, sample entropy (S_E) [20], has been proposed. S_E has the advantage of being less dependent on time series length, and showing relative consistency over a broader range of possible r, m, and N values. Starting from the definition of the K_2 entropy, Richman and Moorman [20] defined the parameter

$$S_E(m,r) = \lim_{N \to \infty} -\ln \frac{U^{m+1}(r)}{U^m(r)}, \quad (12)$$

which is estimated by the statistic

$$S_E(m,r,N) = -\ln \frac{U^{m+1}(r)}{U^m(r)}. \quad (13)$$

The differences between $U^{m+1}(r)$ and $C^{m+1}(r)$, $U^m(r)$ and $C^m(r)$ result from (1) defining the distance between two vectors as the maximum absolute difference between their components; (2) excluding self-matches, i.e., vectors are not compared to themselves; and (3) given a time series with N data points, only the first $N-m$ vectors of length m, $u_m(i)$, are considered, ensuring that, for $1 \leqslant i \leqslant N-m$, the vector $u_{m+1}(i)$ of length $m+1$ is also defined. S_E is precisely equal to the negative of the natural logarithm of the conditional probability that sequences close to each other for m consecutive data points will also be close to each other when one more point is added to each sequence. Figure 1 illustrates how S_E values are calculated.

Note that

$$A_E(m,r,N) \cong \frac{1}{N-m} \sum_{i=1}^{N-m} \ln \frac{n_i^m}{n_i^{m+1}} \quad (14)$$

and

$$S_E(m,r,N) = \ln \frac{\sum_{i=1}^{N-m} n_i'^m}{\sum_{i=1}^{N-m} n_i'^{m+1}}, \quad (15)$$

where $n_i'^m$ differs from n_i^m to the extend that for S_E self-matches are not counted ($i \neq j$) and $1 \leqslant i \leqslant N-m$.

The difference between A_E and S_E can be related to the Renyi entropies, $S_R(q)$, which are defined by $S_R(q) = \ln(\Sigma_i p_i^q)/(1-q)$. A_E approximates the Renyi entropy of order $q=1$ (the usual Shannon entropy) and S_E the Renyi entropy of order $q=2$. The advantage of the latter is that the estimator [Eq. (15)] is unbiased [21].

Both S_E and A_E measure the degree of randomness (or inversely, the degree of orderliness) of a time series. However, as noted, there is no straightforward relationship between regularity, measured by entropy-based metrics, and

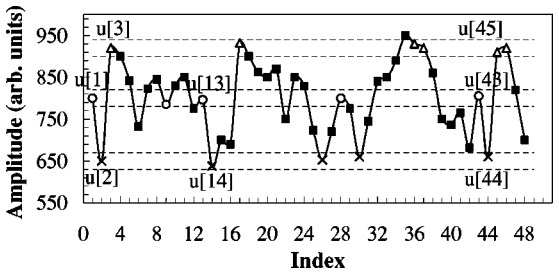

FIG. 1. A simulated time series $u[1],\ldots,u[N]$ is shown to illustrate the procedure for calculating sample entropy (S_E) for the case $m=2$ and a given positive real value r. Dotted horizontal lines around data points $u[1],u[2]$, and $u[3]$ represent $u[1]\pm r$, $u[2]\pm r$, and $u[3]\pm r$, respectively. Two data points match each other, that is, they are indistinguishable, if the absolute difference between them is $\leqslant r$. Typically, r varies between 10% and 20% of the time series SD. The symbol ○ is used to represent data points that match the data point $u[1]$. Similarly, the symbols × and △ are used to represent data points that match the data points $u[2]$ and $u[3]$, respectively. Consider the two-component ○-× template sequence $(u[1],u[2])$ and the three-component ○ -×- △ template sequence $(u[1],u[2],u[3])$. For the segment shown, there are two ○-× sequences, $(u[13],u[14])$ and $(u[43],u[44])$, that match the template sequence $(u[1],u[2])$, but only one ○-×-△ sequence that matches the template sequence $(u[1],u[2],u[3])$. Therefore, in this case, the number of sequences matching the two-component template sequences is two and the number of sequences matching the three-component template sequence is 1. These calculations are repeated for the next two-component and three-component template sequence, which are $(u[2],u[3])$ and $(u[2],u[3],u[4])$, respectively. The number of sequences that match each of the two- and three-component template sequences are again summed and added to the previous values. This procedure is then repeated for all other possible template sequences, $(u[3],u[4],u[5]),\ldots,(u[N\text{-}2],u[N\text{-}1],u[N])$, to determine the ratio between the total number of two-component template matches and the total number of three-component template matches. S_E is the natural logarithm of this ratio and reflects the probability that sequences that match each other for the first two data points will also match for the next point.

complexity [22]. An increase in entropy is usually but not always associated with an increase in complexity. For example, higher entropy values are assigned to randomized surrogate time series than to the original time series even when the original time series represent the output of complex dynamics with correlational structures on multiple spatiotemporal scales. However, the process of generating surrogate data is designed to destroy correlations and, consequently, degrades the information content of the original signal. In fact, entropy-based metrics are maximized for random sequences, although it is generally accepted that both perfectly ordered and maximally disordered systems possess no complex structures [23]. A meaningful physiologic complexity measure, therefore, should vanish for these two extreme states.

Of related note, when applied to physiologic data, both A_E and S_E algorithms assign higher entropy values to certain pathologic time series than to time series derived from free-running physiologic systems under healthy conditions [24]. However, pathologic time series represent the output of less adaptive (i.e., more impaired), and therefore, presumably, less complex systems [25,26]. One reason for obtaining these "nonphysiologic" results is the fact that A_E and S_E are based on a single scale. We note that both the KS entropy and the related A_E parameters depend on a function's one-step difference (e.g., $H_{n+1}-H_n$) and reflect the uncertainty of the next new point given the past history of the series. Therefore, these measures do not account for features related to structure and organization on scales other than the shortest one.

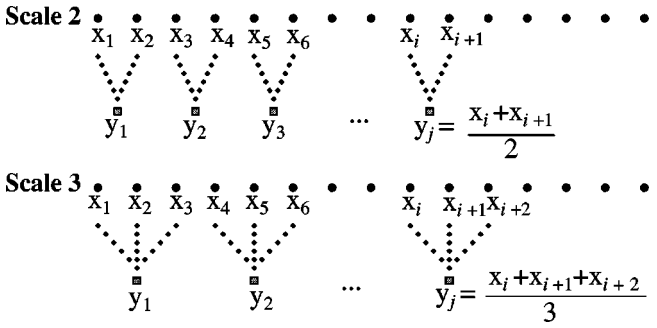

FIG. 2. Schematic illustration of the coarse-graining procedure. Adapted from Ref. [8].

For physical systems, Zhang [23,27] proposed a general approach to take into account the information contained in multiple scales. Zhang's complexity measure is a sum of scale-dependent entropies. It has the desirable property of vanishing in the extreme ordered and disordered limits, and is an extensive quantity. However, since it is based on Shannon's definition of entropy, Zhang's method requires a large amount of almost noise-free data, in order to map the data to a discrete symbolic sequence with sufficient statistical accuracy. Therefore, it presents obvious limitations when applied to free-running physiologic signals that typically vary continuously and have finite length.

To overcome these limitations, we [7] recently introduced the multiscale entropy (MSE) method, applicable both to physical and physiologic time series. Our method is based on Zhang's and Pincus's approach.

III. MULTISCALE ENTROPY (MSE) METHOD

Given a one-dimensional discrete time series, $\{x_1,\ldots,x_i,\ldots,x_N\}$, we construct consecutive coarse-grained time series, $\{y^{(\tau)}\}$, corresponding to the scale factor, τ. First, we divide the original times series into nonoverlapping windows of length τ; second, we average the data points inside each window (Fig. 2). In general, each element of a coarse-grained time series is calculated according to the equation

$$y_j^{(\tau)}=\frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau}x_i,\quad 1\leqslant j\leqslant N/\tau. \tag{16}$$

For scale one, the time series $\{y^{(1)}\}$ is simply the original time series. The length of each coarse-grained time series is equal to the length of the original time series divided by the scale factor, τ.

Finally, we calculate an entropy measure (S_E) for each coarse-grained time series plotted as a function of the scale

factor τ. We call this procedure multiscale entropy (MSE) analysis.

The MSE curves are used to compare the relative complexity of normalized time series (same variance for scale one) based on the following guidelines: (1) if for the majority of the scales the entropy values are higher for one time series than for another, the former is considered more complex than the latter; (2) a monotonic decrease of the entropy values indicates the original signal contains information only in the smallest scale.

Zhang defined complexity as the integral of all the scale-dependent entropies: $K=\int_1^N d\tau H(\tau)$, which for a discrete signal could be estimated by $K=\Sigma_{i=1}^N H(i)(N\rightarrow\infty)$. Due to the finite length of real-world time series, entropy can only be calculated for a finite range of scales. The sum to infinity is not feasible. Since different sets of entropy values can yield the same K value, we focus on the analysis of the MSE curves instead of assigning a single complexity value to each time series. Further, for application to biological systems, the MSE curve may provide useful insights into the control mechanisms underlying physiologic dynamics over different scales. We note, however, that an approximation of K for scales between one and twenty further supports the conclusions we present in this paper.

Unless otherwise specified, the values of the parameters used to calculate S_E are $N=2\times10^4$, $m=2$, and $r=0.15$.

The value of the parameter r is a percentage of the time series SD. This implementation corresponds to normalizing the time series. As a consequence, S_E results do not depend on the variance of the original time series, i.e., the absolute value of the data points, but only on their sequential ordering.

In general, however, the entropy measures reflect both the variance of a time series and its correlation properties. To illustrate this point, we examine two special cases where these two effects can be isolated. Case (1): Consider two uncorrelated random variables, X and Y, with set of values $\{x_1,x_2,\ldots,x_N\}$ and $\{y_1,y_2,\ldots,y_M\}$, respectively. Assuming that all values are equally probable, $p(x_i)=1/N$, the entropy of the random variables X is $H(X)=-\Sigma_{i=1}^N 1/N \log 1/N = \log N$. Similarly, $H(Y)=\log M$. If $N>M$, then $H(X)>H(Y)$. Therefore, for the same level of resolution, the larger the set of alphabet of a random variable, the larger its variance and the entropy value. Case (2): Consider a periodic signal with variance $|a|$ and a random signal with variance $|b|$, such that $|a|\gg|b|$. The entropy of a periodic signal is zero, since each data point occurs with probability 1. Therefore, the entropy of a periodic signal is never larger than the entropy of a random signal regardless of the variance of the signals.

With the exception of such very simple cases, it is not possible to weight separately the contributions of the SD and the correlation properties to the entropy value. Signals with higher variability and those that are more random tend to be more entropic. Nevertheless, the actual entropy value results from a complex combination of these two factors.

In the MSE method, r is set at a certain percentage (usually 15%) of the original time series SD, and remains constant for all scales [10,28]. We do not recalculate r for each coarse-grained time series. After the initial normalization, subsequent changes of the variance due to the coarse-graining procedure are related to the temporal structure of the original time series, and should be accounted for by the entropy measure. The initial normalization, however, insures that the MSE values assigned to two different time series are not a trivial consequence of possible differences between their variances but result from different organizational structures.

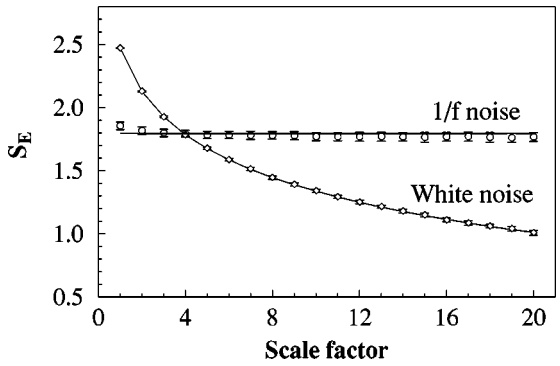

FIG. 3. MSE analysis of 30 simulated Gaussian distributed (mean zero, variance one) white and $1/f$ noise time series, each with 3×10^4 data points. Symbols represent mean values of entropy for the 30 time series and error bars the SD, which in average is 0.05 for white noise and 0.02 for $1/f$ noise. Lines represent numerical evaluation of analytic S_E calculation. Note that the differences between the mean values of S_E and the corresponding numerical values are less than 1%. SD is larger for $1/f$ noise time series because of nonstationarity. Adapted from Ref. [7]. (See Appendix A.)

We first applied the MSE method to simulated white and $1/f$ noises and compared the numerical results with the entropy values calculated analytically (Appendix A). Figure 3 presents the results. For scale one, a higher value of entropy is assigned to white noise time series in comparison with $1/f$ time series. However, while the value of entropy for the coarse-grained $1/f$ series remains almost constant for all scales, the value of entropy for the coarse-grained white noise time series monotonically decreases, such that for scales >4 it becomes smaller than the corresponding values for $1/f$ noise. This result is consistent with the fact that, unlike white noise, $1/f$ noise contains complex structures across multiple scales [23,27]. Note that in the case of white noise, as the length of the window used for coarse-graining the time series increases (i.e., the resolution decreases), the average value inside each window converges to a fixed value since no new structures are revealed on larger scales. Consequently, coarse-grained time series are progressively "smoothed out" and the standard deviation monotonically decreases with the scale factor. Therefore, the monotonic decrease of entropy with scale, which mathematically results from the decrease of standard deviation, reflects the fact that white noise has information only on the shortest scale. In contrast, for $1/f$ noise signals the average values of the fluctuations inside each window do not converge to a given value. In other words, the statistical properties of fluctuations within a window (e.g., 10 data points) are not the same as

those of the next window because new information is revealed at all scales. The MSE uses the average value of the fluctuations as the representative statistical property for each block and measures the irregularity of the block-to-block dynamics.

The discrepancy between the simulation and the analytical results is less than 0.5%. In Appendix B, we discuss how the time series length, N, and the values of parameters r and m affect S_E results for both white and $1/f$ noise time series. We further discuss the effects of uncorrelated noise and outliers on MSE results of cardiac interbeat interval time series.

IV. MSE ANALYSIS OF CARDIAC INTERBEAT INTERVAL TIME SERIES

We next apply the MSE method to the cardiac interbeat (RR) interval time series derived from 24 hour continuous electrocardiographic (ECG) Holter monitor recordings of healthy subjects, subjects with congestive heart failure, a life-threatening condition, and subjects with atrial fibrillation, a major cardiac arrhythmia.[1] We test the hypothesis that under free-running conditions, healthy interbeat interval dynamics are more complex than those with pathology during both daytime and nightime hours.

The data for the normal control group were obtained from 24 hour Holter monitor recordings of 72 healthy subjects, 35 men and 37 women, aged 54.6±16.2 years (mean±SD), range 20-78 years. ECG data were sampled at 128 Hz. The data for the congestive heart failure group were obtained from 24 hour Holter recordings of 43 subjects (28 men and 15 women) aged 55.5±11.4 years (mean±SD), range 22-78 years. New York Heart Association (NYHA) functional classification [30] is provided for each subject: 4 subjects were assigned to class I, 8 to class II, 17 to class III, and 14 to class III-IV. Fourteen recordings were sampled at 250 Hz and 29 recordings were sampled at 128 Hz. The data for the atrial fibrillation group were obtained from 10 hour Holter recordings sampled at 250 Hz of nine subjects. Datasets were filtered to exclude artifacts, premature ventricular complexes, and missed beat detections (see Appendix B). Of note, the inclusion of the premature ventricular complexes does not qualitatively change our analysis.

Representative time series of healthy, congestive heart failure, and atrial fibrillation group subjects are presented in Fig. 4.

When discussing the MSE results of cardiac interbeat interval time series, we refer to"large" and "small" time scales when the scales are larger or smaller than one typical respiratory cycle length, that is, approximately five cardiac beats.

In Fig. 5, we present the results of the MSE analysis of the RR interval time series for the three groups of subjects. We observe three different types of behaviors: (i) The entropy measure for time series derived from healthy subjects increases on small time scales and then stabilizes to a relatively constant value. (ii) The entropy measure for time series derived from subjects with congestive heart failure markedly decreases on small time scales and then gradually increases. (iii) The entropy measure for time series derived from subjects with atrial fibrillation [31] monotonically decreases, similar to white noise (Fig. 3).

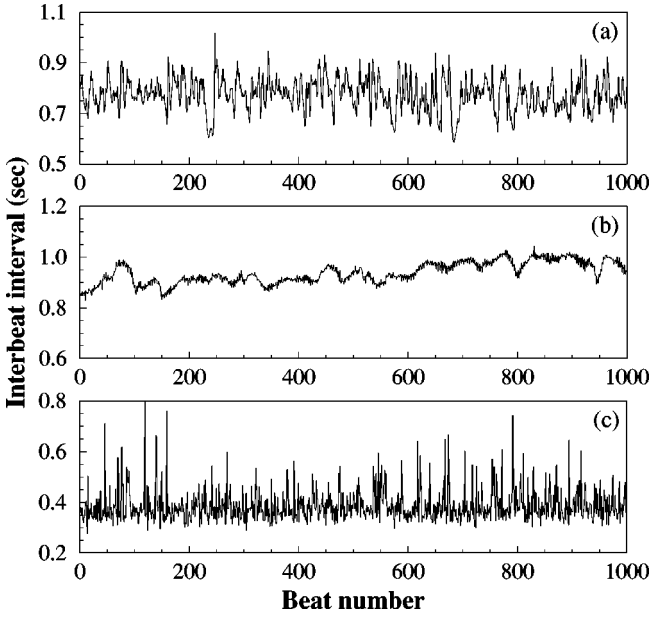

FIG. 4. Representative interbeat interval time series from (a) healthy individual (sinus rhythm), (b) subject with congestive heart failure, and (c) subject with atrial fibrillation, a highly erratic cardiac arrhythmia.

For scale one, which is the only scale considered by traditional single-scale based "complexity" methods, the entropy assigned to the heartbeat time series of subjects with atrial fibrillation and those with congestive heart failure is higher than the entropy assigned to the time series of healthy

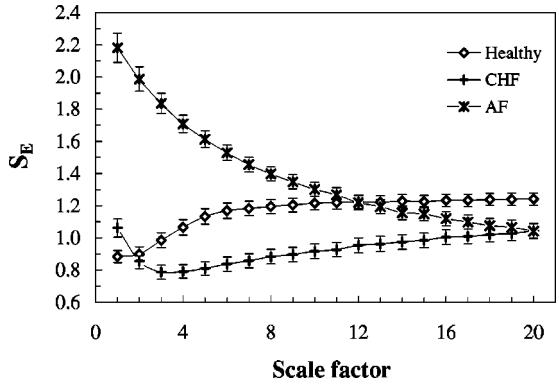

FIG. 5. MSE analysis of RR time series derived from long-term ECG recordings of healthy subjects in normal sinus rhythm, those with congestive heart failure (CHF) in sinus rhythm, and those with atrial fibrillation (AF). Symbols represent the mean values of entropy for each group and bars represent the standard error ($SE = SD/\sqrt{n}$), where n is the number of subjects). Parameters to calculate S_E are $m=2$ and $r=0.15$. Time series length is 2×10^4 beats. The S_E values from healthy subjects are significantly (t-test, $p<0.05$) higher than from CHF and AF subjects for scales larger than scale 2 and scale 20, respectively.

[1]All data analyzed here are available at http://physionet.org and have been described in Ref. [29].

subjects. In contrast, for sufficiently large scales, the time series of healthy subjects are assigned the highest entropy values. *Thus, the MSE method indicates that healthy dynamics are the most complex, contradicting the results obtained using the traditional S_E and A_E algorithms.*

The time series of subjects with AF exhibit substantial variability in beat-to-beat fluctuations. However, the monotonic decrease of the entropy with scale reflects the degradation of the control mechanisms regulating heart rate on larger time scales in this pathologic state.

The largest difference between the entropy values of coarse-grained time series from congestive heart failure and healthy subjects is obtained for time scale 5. On small time scales, the difference between the profiles of the MSE curves for these two groups may be due to the fact that the respiratory modulation of heart rate (respiratory sinus arrhythmia) has higher amplitude in healthy subjects than in subjects with congestive heart failure. Since entropy is a measure of regularity (orderliness), the higher the amplitude of the respiratory modulation, the lower the entropy values tend to be. However, the coarse-graining procedure filters out the periodic respiratory-related heart rate oscillations. Therefore, coarse-grained time series from healthy subjects on large time scales are likely more irregular (and are assigned higher entropy values) than the original time series.

For congestive heart failure subjects, the entropy of coarse-grained time series decreases from scales 1–3 and then progressively increases. This result suggests that for these subjects, the control mechanisms regulating heart rate on relatively short time scales are the most affected. However, this finding could also result from the measurement uncertainty of the interbeat intervals due to the finite sample frequency. Since time series from subjects with congestive heart failure have, in general, lower variance than time series from healthy subjects, the signal-to-noise ratio tends to be lower for datasets from heart failure subjects. We note that the MSE coarse-graining procedure progressively eliminates the uncorrelated random components such that the entropy of white noise coarse-grained time series monotonically decreases with scale (Fig. 3). Therefore, the monotonic decrease of the entropy values with heart failure over short time scales may be related to the relatively low signal-to-noise ratio.

We also find that the asymptotic value of entropy may not be sufficient to differentiate time series that represent the output of different dynamical processes. As seen in Fig. 5, for time scale 20, the value of the entropy measure for the heart failure (sinus rhythm) and atrial fibrillation time series is the same. However, these time series represent the output of very different types of cardiac dynamics. *Therefore, not only the specific values of the entropy measure but also their dependence on time scale need to be taken into account to better characterize the physiologic process.*

Next, to assess the effects of activity level, we compare the complexity of the RR intervals time series during sleep and wake periods for the different subject groups. Using the 24 h heartbeat interval time series of healthy and congestive heart failure subjects, the sleep and wake datasets were then obtained by extracting the segments of 2×10^4 consecutive data points ($\sim$5 h) with highest and lowest heart rate, respectively. Figures 6(a) and 6(b) show that during both the waking and sleeping periods, the highest entropy values on most time scales are assigned, in descending order, to the coarse-grained time series derived from healthy young subjects, healthy elderly subjects, and congestive heart failure subjects. These results further support the concept that under free-running conditions, the cardiac dynamics of healthy young subjects are the most complex and are consistent with the hypothesized loss of complexity with aging and disease [24].

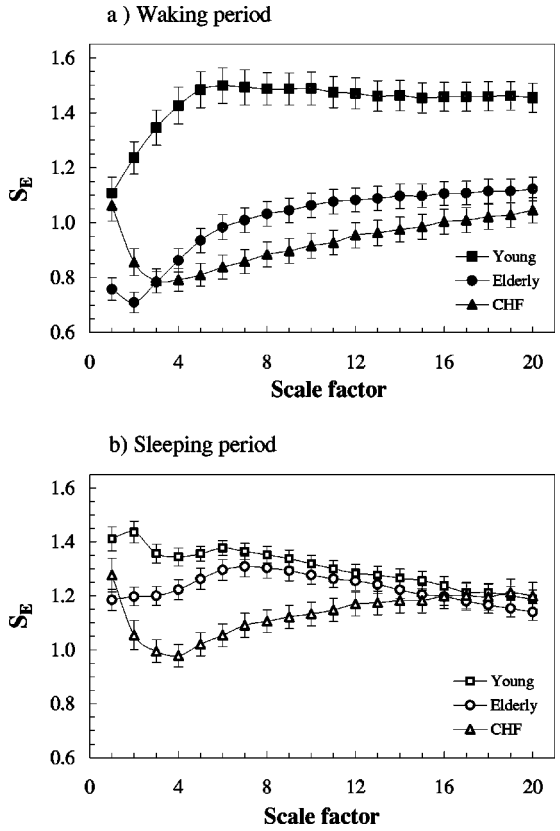

FIG. 6. MSE analysis of RR time series derived from 24 h ECG recordings of 27 healthy young subjects, aged 34.5±7.3 years (mean±SD), range 20 - 50 years, 45 healthy elderly subjects, aged 70±3.97 years, range 66 - 75 years, and 43 congestive heart failure (CHF) subjects, aged 55±11.6 years, range 22 - 78 years. (a) Waking period. For all scales the S_E values from healthy young subjects are significantly (t-test, $p<0.05$) higher than from CHF subjects. The S_E values from healthy young subjects are significantly higher than from healthy elderly subjects for scales larger than scale 1. The S_E values from healthy elderly subjects are significantly (t-test, $p<0.05$) higher than from CHF subjects for scales between scales 5 and 13, inclusively. (b) Sleeping period. Both the S_E values from healthy elderly and healthy young subjects are significantly (t-test, $p<0.05$) higher than from CHF subjects for scales between scales 2 and 11, inclusively. The S_E values from healthy young subjects are significantly higher than from healthy elderly subjects for scales shorter than scale 5. Symbols represent the mean values of entropy for each group and the bars represent the standard error. Parameters of S_E calculation are $m=2$ and $r=0.15$. Time series length is 2×10^4 beats.

Despite the fact that the entropy values for healthy elderly subjects are lower than those for healthy young subjects, the

profiles of MSE curves for both groups are similar, in particular over large time scales. Indeed, during sleep, a period of minimal activity, the difference between the entropy values of both groups is significant over only small time scales. These results are consistent with the known loss of high-frequency modulation of the cardiac rhythm with age [32], and suggest that the control mechanisms operating over small time scales, including the parasympathetic branch of the autonomic nervous system, are the most affected with aging. The monotonic decrease in entropy on large time scales for both young and elderly groups indicates that the coarse-grained time series become progressively more regular (less complex) than those corresponding to shorter time scales, which is compatible with a previous study [33] reporting a reduction in long-range correlations in healthy subjects during the sleeping period.

The MSE results for the waking and sleeping periods of each group of subjects are shown in Fig. 7. For both young and elderly healthy subjects, the profiles of the MSE curves corresponding to the waking and sleeping periods are qualitatively different from each other [Figs. 7(a) and 7(b)]. For subjects with congestive heart failure, however, there is only a shift of the entropy values but not a significant change in the profile of the MSE curves [Fig. 7(c)]. Thus, differences between the day versus night dynamics of subjects with a severe cardiac pathology are less marked than for healthy subjects. This loss of differentiation in the complexity of sleep/wake dynamics may be a useful new index of reduced adaptive capacity.

Further, we found that, contrary to the results obtained for healthy young subjects, in healthy elderly and congestive heart failure subjects, the coarse-grained time series obtained from the waking period have lower entropy than those obtained from the sleeping period. To the extent that aging and disease degrade adaptive capacity, environmental stimuli may exceed the system's reserve. This situation would be equivalent to what might occur if a young individual were subject to prolonged physical or other stress throughout the daytime hours.

Finally, to assist in clinical classification, we extracted two simple features of MSE curves, the slopes for small and large time scales, i.e., the slopes of the curves defined by S_E values between scale factors 1 and 5, and scale factors 6 and 20, respectively. Results for the healthy and congestive heart failure groups corresponding to the sleeping period are presented in Fig. 8. There is a good separation between the two groups. Considering other features of the MSE curves, in addition to these slopes, may further improve the separation. Alternatively, methods derived from pattern recognition techniques, e.g., Fisher's discriminant, may also be useful for clinical discrimination [9].

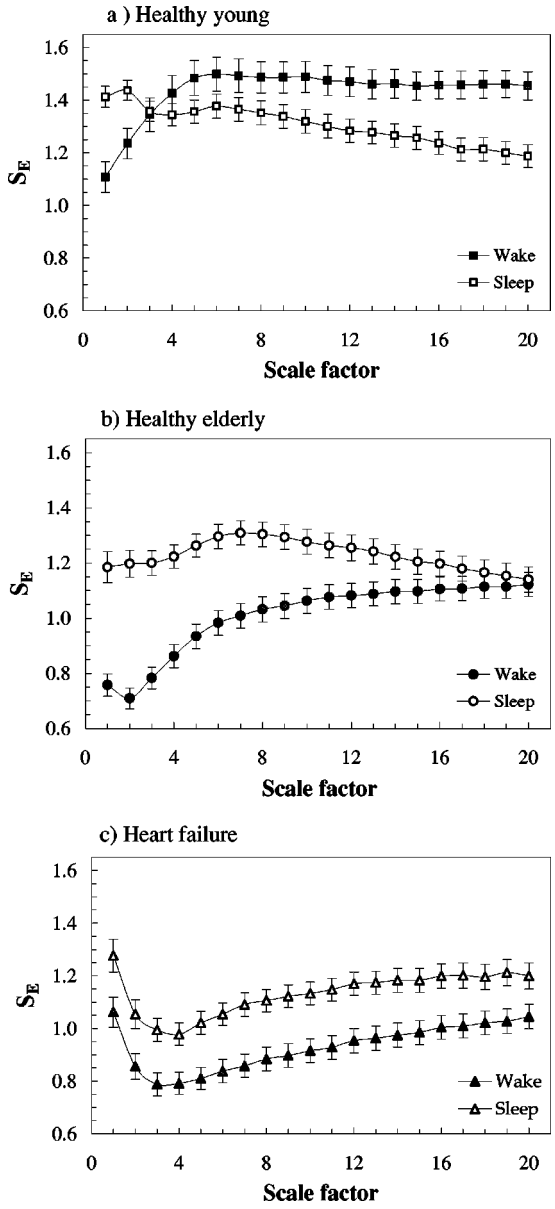

FIG. 7. MSE analysis of RR time series derived from 24 h ECG recordings during waking and sleeping periods. (a) Young healthy subjects. The S_E values for the waking period are significantly (t-test) higher ($p<0.05$) than for the sleeping period on scales larger than scale 7. (b) Elderly healthy subjects. The S_E values for the sleeping period are significantly (t-test) higher ($p<0.05$) than for the waking period on scales shorter than scale 16. (c) Congestive heart failure subjects. The S_E values for the sleeping period are significantly (t-test) higher ($p<0.05$) than for the waking period on all scales but scale 1. Symbols represent mean values of entropy for each group and the bars represent the standard error. Parameters of S_E calculation are $m=2$ and $r=0.15$. Time series length is 2×10^4 beats.

V. MSE ANALYSIS OF ARTIFICIAL AND BIOLOGICAL CODES

In all cells, from microbes to mammals, proteins are responsible for most structural, catalytic, and regulatory functions. Therefore, the number of protein-coding genes that an organism makes use of could be an indicator of its degree of complexity. However, several observations contradict this reasoning [34,35].

Large regions of DNA, which in humans account for about 97% of the total genome, do not code for proteins and were previously thought to have no relevant purpose. These regions have been referred to as "junk" DNA or gene

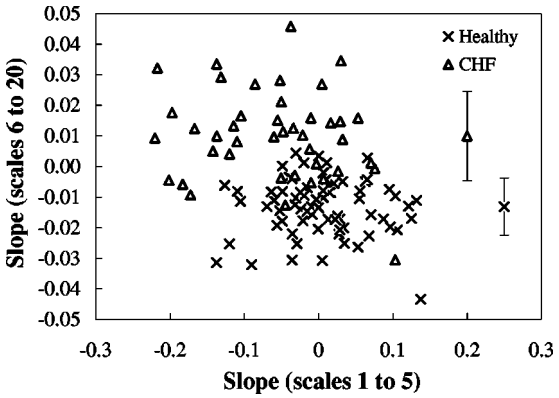

FIG. 8. Scatter plot of the slope of the MSE curves between scale factors 6 and 20 vs the slope of the MSE curves between scale factors 1 and 5, for healthy and congestive heart failure (CHF) groups during the sleeping period. For both groups, symbols with error bars represent the mean of *y*-axis values, and the error bars the corresponding SD. The groups are well separated ($p < 0.005$).

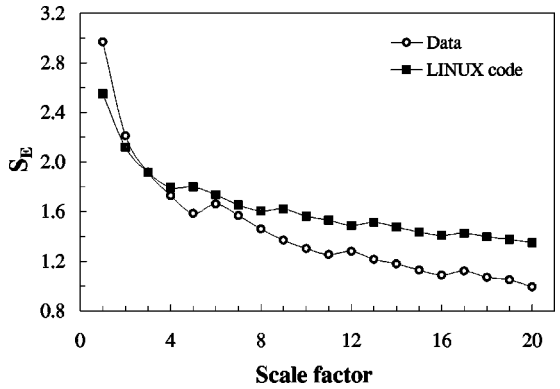

FIG. 9. MSE results for binary files of a computer executable program (LINUX kernel) and a compressed data file. The original binary file has only two symbols, 0 and 1. However, the number of symbols in coarse-grained sequences increases with the scale factor, which introduces a characteristic artifact on the MSE curves. In order to avoid this artifact, instead of the original sequences, we analyze a derived sequence, which is constructed as follows: we divide the original sequence into consecutive nonoverlapping segments, each with 128 data points, and then calculate the number of 1's (0's) within each segment. Some structural information is lost since the procedure is not a one-to-one mapping. The derived sequences are expected to be more regular than the original ones. However, this procedure does not alter the conclusions drawn from our analysis.

"deserts." However, these noncoding sequences are starting to attract increasing attention as more recent studies suggest that they may have an important role in regulation of transcription, DNA replication and chromosomal structure, pairing, and condensation.

Detrended fluctuation analysis [37–39] revealed that noncoding sequences contained long-range correlations and possessed structural similarities to natural languages, suggesting that these sequences could in fact carry important biological information. In contrast, coding sequences were found to be more like a computer data file than a natural language.

The biological implications of the presence of long-range correlations in noncoding sequences, their origin, and their nature are still being debated. Audit *et al.* [40,41] have investigated the relation between long-range correlations and the structure and dynamics of nucleosomes. Their results suggest that long-range correlations extending from 10 to 200 bp are related to the mechanisms underlying the wrapping of DNA in the nucleosomal structure.

Gene regulatory elements or enhancers are types of functional sequences that reside in noncoding regions. Until recently, enhancers were thought to be located near the genes that they regulate. However, subsequent *in vivo* studies [42,43] have demonstrated that enhancers and the genes to which they are functionally linked may be separated by more than thousands of bases. These results reinforce earlier evidence that the noncoding sequences contain biological information and further support the notion that there are several "layers" of information in genomic DNA.

In this section, we apply the MSE method to the analysis of the complexity of both coding and noncoding DNA sequences of human chromosomes.

Because of possible parallelisms between artificial and biological codes, we first considered two examples of artificial language sequences: the compiled version of the LINUX Operating System, an executable computer program, and a compressed nonexecutable computer data file, which can both be analyzed as binary sequences. Although both files contain useful information, the structure of that information is very different. The sequence derived from the executable program exhibits long-range correlations [38], while the sequence derived from the data file does not. These results indicate that the computer program, which executes a series of instructions and likely contains several loops running inside each other, possesses a hierarchical structure, in contrast to the computer data file. Therefore, the former is expected to be more complex than the latter.

When applied to discrete sequences (binary codes), the MSE results present a typical artifact due to the dependence of the entropy values on the size of the sequence alphabet, which we discuss in Appendix C.

MSE analysis of the nonbiological codes reveals (Fig. 9) the following. (i) For scale one, the sequence derived from the data file is assigned a higher entropy value than the sequence derived from the executable program. (ii) Between scales 2 and 6, the S_E measure does not separate the coarse-grained sequences of the two files. (iii) For scales larger than scale 6, the highest entropy values are assigned to coarse-grained sequences derived from the executable program file. Furthermore, the difference between S_E values assigned to coarsegrained sequences of the executable file and the computer data file increases with scale factor. These results indicate, as hypothesized, that the structure of the executable file is more complex than the structure of the data file. Of note, conventional (single scale) S_E and A_E algorithms applied to sequences of artificial languages fail to meaningfully quantify their overall complexity.

Finally, we apply the MSE method to the analysis of DNA sequences, likely one of the most complex natural information databases.

The DNA building units are four nucleotides. Two of them contain a purine base, adenine (A) or guanine (G), and

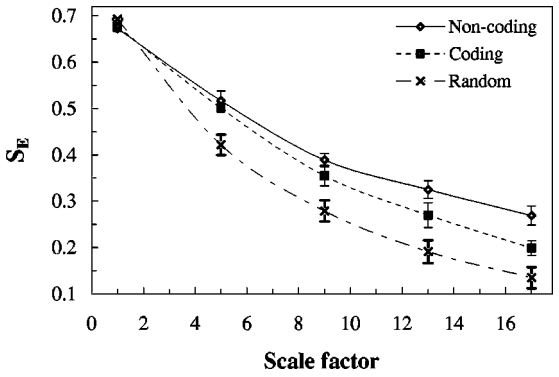

FIG. 10. MSE results for four coding, nine noncoding DNA sequences from human chromosome 22 and 30 binary random time series. All coding sequences with more than identified 4×10^3 bp were selected. The longest coding sequences has 6762 bp. All noncoding sequences with more than 6000 and fewer than 6050 bp were selected. The length of the random sequences is 6000 data points. The symbols and the error bars represent the S_E mean values and SD, respectively. Due to a typical artifact that affects the MSE results of discrete sequences (Appendix C), only the entropy values for scales 1, 5, 9, 13, and 17 are plotted. Note the higher complexity of the noncoding vs coding sequences ($p=0.006$ for scale 9). The lowest entropy values are assigned to the random (white noise: mean zero, variance 1) time series mapped to a binary sequence: 1 if $x_i>0$ and 0 if $x_i<0$.

the other two contain a pyrimidine base, cytosine (C) or thymine (T). There are many ways of mapping the DNA sequences to a numerical sequence that take into consideration different properties of the DNA sequences. For this application, we consider the purine-pyrimidine rule [37–39]. Given the original DNA sequence, bases A and G are mapped to number 1, and bases C and T are mapped to number -1.

In Fig. 10, we present the MSE results for selected coding and noncoding human DNA sequences. For scales larger than scale 5, S_E values for noncoding sequences are higher than for coding sequences. Consistently, for all scales but the first one, the lowest S_E values are assigned to coarse-grained time series derived from uncorrelated white noise mapped to a binary sequences. Comparable results were obtained from the analysis of coding versus noncoding sequences ($\geq 4\times10^3$ bp) of all human chromosomes. These results show that the structure of noncoding sequences is more complex than the structure of coding sequences analyzed here.

These findings support previous studies [37–39] suggesting a parallelism between executable computer programs and noncoding sequences, and data storing files and coding sequences. They also support the view that noncoding sequences contain important biological information. As pointed out by others [35,36,40,41], biological complexity and phenotype variations should relate not only to proteins, which are the main effectors of cellular activity, but also to the organizational structure of the control mechanisms responsible for the networking and integration of gene activity.

VI. LIMITATIONS AND FUTURE DIRECTIONS

The MSE method requires an adequate length of data to provide reliable statistics for the entropy measure on each scale. As discussed in Appendix B, for simulated white and $1/f$ noises, both the mean value of S_E and the SD increase as the length of the time series decreases. However, for all time series tested, the consistency of the results was preserved, i.e., given two time series, a and b, each with 3×10^4 data points, whenever S_E was higher (lower) for time series a than for time series b, the same result held if only 1×10^3 data points were considered.

The minimum number of data points required to apply the MSE method depends on the level of accepted uncertainty. Typically, we use time series with 2×10^4 data points for analyses extending up to scale 20, in which case the shortest coarse-grained time series has 1×10^3 data points.

Another important consideration is related to nonstationarity. To calculate S_E, one has to fix the value of a parameter that depends on the time series SD. Therefore, the results may be significantly affected by nonstationarities, outliers, and artifacts. As we discuss in Appendix C, removing local artifacts and a small percentage of outliers ($<2\%$) does not usually modify the structure of the time series and its related statistical properties. In contrast, attempts to remove nonlocal nonstationarities, e.g., trends, will most likely modify the structure of the time series over multiple time scales.

Further studies are needed to construct clinically useful indices for monitoring the complexity of biological systems, and for developing and testing the utility of complexity measures designed to quantify the degree of synchronization of two time series over multiple scales [20].

We note that the cardiac analyses reported here pertain to interbeat interval dynamics under free-running conditions. The high capability of healthy systems to adapt to a wide range of perturbations requires functioning in a multidimensional state space. However, under stress, the system is forced to work in a tighter regime. For example, during physical exercise, there is a sustained increase in heart rate and a decrease in the amplitude of the interbeat interval fluctuations in response to an increased demand for oxygen and nutrients. The dynamics is, therefore, limited to a subset of the state space. We anticipate that under a variety of stressed conditions, healthy systems will generate less complex outputs than under free-running conditions [11].

Finally, the potential applications of the MSE method to the study of artificial and biological codes, with attention to the effects of evolution on the complexity of genomic sequences, require systematic analysis.

VII. CONCLUSIONS

The long-standing problem of deriving useful measures of time series complexity is important for the analysis of both physical and biological systems. MSE is based on the observation that the output of complex systems is far from the extrema of perfect regularity and complete randomness. Instead, they generally reveal structures with long-range correlations on multiple spatial and temporal scales. These multiscale features, ignored by conventional entropy calculations, are explicitly addressed by the MSE method.

When applied to simulated time series, the MSE method shows that $1/f$ noise time series are more complex than

white noise time series. These results are consistent with the presence of long-range correlations in $1/f$ noise time series but not in white noise time series.

Physiologic complexity is associated with the ability of living systems to adjust to an ever-changing environment, which requires integrative multiscale functionality. In contrast, under free-running conditions, a sustained decrease in complexity reflects a reduced ability of the system to function in certain dynamical regimes possibly due to decoupling or degradation of control mechanisms.

When applied to the cardiac interbeat interval time series of healthy subjects, those with congestive heart failure and those with atrial fibrillation, the MSE method shows that healthy dynamics are the most complex. Under pathologic conditions, the structure of the time series variability may change in two different ways. One dynamical route to disease is associated with loss of variability and the emergence of more regular patterns (e.g., heart failure). The other dynamical route is associated with more random types of outputs (e.g., atrial fibrillation). In both cases, MSE reveals a decrease in system complexity.

Finally, we employed the MSE method to compare the complexity of an executable computer program versus a compressed nonexecutable computer data file, and selected coding versus noncoding DNA sequences. We found that the executable computer program has higher complexity than the nonexecutable computer data file, and similarly that the noncoding sequences are more complex than the coding sequences examined. Our results support recent *in vitro* and *in vivo* studies suggesting, contrary to the "junk DNA" theory, that noncoding sequences contain important biological information [44].

ACKNOWLEDGMENTS

We thank J. Mietus, I. Henry, and J. Healey for valuable discussions and assistance. We gratefully acknowledge support from the National Institutes of Health/National Center for Research Resources (P41-RR13622), the G. Harold and Leila Y. Mathers Charitable Foundation, the Centers for Disease Control and Prevention (H75-CCH119124), the NIH/NICHD (R01-HD39838), and the James S. McDonnell Foundation.

APPENDIX A: MSE RESULTS FOR WHITE AND $1/f$ NOISES

In this appendix, we provide detailed analytical derivations of MSE for two special cases: correlated and uncorrelated noises with Gaussian distributions. Linear Gaussian correlation is a necessary assumption to make the derivation possible. In general, it is difficult to derive analytical solutions for MSE of stochastic processes with nonlinear correlations.

First, we start with the case of uncorrelated noise (white noise). For the case $m=1$, S_E is the negative natural logarithm of the conditional probability that the distance between two data points is less than or equal to r (i.e., $|x_i-x_j|\leqslant r$) given that the distance between the two preceding data points is also less than or equal to r (i.e., $|x_{i-1}-x_{j-1}|\leqslant r$). Since there is no correlation between any data point and the preceding data points in white noise, S_E reduces to the negative natural logarithm of the probability that the distance between any two data points is less than or equal to r.

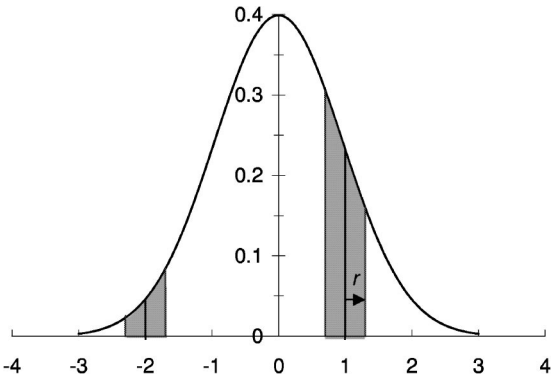

FIG. 11. Gaussian distribution. Shadowed areas centered at points −2 and 1 represent the probability that the distances between each of these points and any other point chosen randomly from the time series are less than or equal to r.

To be specific, the joint probability of a finite sequence of independent random variables is simply

$$p(x_1,x_2,\ldots,x_n)=\prod_{i=1}^{N}p(x_i). \tag{A1}$$

One can show that

$$\begin{aligned}P_r(|x_i-x_j|\leqslant r\,||x_{i-1}-x_{j-1}|\leqslant r)&=\frac{P_r(|x_i-x_j|\leqslant r\wedge|x_{i-1}-x_{j-1}|\leqslant r)}{P_r(|x_{i-1}-x_{j-1}|\leqslant r)}\\&=\frac{P_r(|x_i-x_j|\leqslant r)\times P_r(|x_{i-1}-x_{j-1}|\leqslant r)}{P_r(|x_{i-1}-x_{j-1}|\leqslant r)}\\&=P_r(|x_i-x_j|\leqslant r).\end{aligned}$$

Using this approach recursively, it can be proved that this result is valid for any m value, whenever the variables are independent. In this appendix, we adhere to the standard notations of using $P_r()$ for probability distributions and $p()$ for probability density functions.

To summarize, white noise is a random process such that all variables are independent. Therefore,

$$S_E=-\ln P_r(|x_j-x_i|\leqslant r). \tag{A2}$$

Next, we calculate the probability distribution $P_r(|x_j-x_i|\leqslant r)$.

For a given value of $\hat{x}$, the probability of finding other data points within the distance r from $\hat{x}$ is

$$P_r(|\hat{x}-x|\leqslant r)=\int_{\hat{x}-r}^{\hat{x}+r}p(x)dx. \tag{A3}$$

For example, if $x_i=1$ and $r=0.3$, (Fig. 11), $P_r(|1-x_j|\leqslant 0.3)$ is the area under the Gaussian curve between the vertical lines $x=0.7$ and $x=1.3$. Similarly, for $x_i=-2$ and the

same r value, $P_r(|2-x_j| \leq 0.3)$ is the area under the Gaussian curve between the vertical lines $x=-2.3$ and $x=-1.7$. Since x_i can assume any value between $-\infty$ and $+\infty$, $P_r(|x_i-x_j| \leq r)$ is the average area centered at all possible x_i values. In other words,

$$P_r(|x_j - x_i| \leq r) = \int_{-\infty}^{+\infty} \left\{ \int_{x_i-r}^{x_i+r} p(x_j)dx_j \right\} p(x_i)dx_i$$

$$= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{+\infty} \left\{ \int_{x_i-r}^{x_i+r} e^{-x_j^2/2\sigma^2} dx_j \right\} e^{-x_i^2/2\sigma^2} dx_i$$

$$= \frac{1}{2\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left\{ \mathrm{erf}\left(\frac{x_i + r}{\sigma\sqrt{2}} \right) - \mathrm{erf}\left(\frac{x_i - r}{\sigma\sqrt{2}} \right) \right\} \times e^{-x_i^2/2\sigma^2} dx_i,$$

where erf refers to the error function.

Without loss of generality, we considered a zero mean ($\mu=0$) Gaussian distribution. Coarse-grained white noise time series still have a zero mean Gaussian density because they are the output of a linear combination of Gaussian random variables. However, the variance decreases as the scale factor increases,

$$\sigma_\tau = \frac{\sigma}{\sqrt{\tau}}, \tag{A4}$$

where τ refers to the scale factor, σ_τ to the variance of the coarse-grained time series corresponding to scale τ, and σ to the variance of the original time series (scale 1). Consequently, the probability that the distance between two data points of the coarse-grained time series corresponding to scale τ is less than or equal to r is

$$P_r(|y_j^\tau - y_i^\tau| \leq r) = \frac{1}{2\sigma} \sqrt{\frac{1}{2\pi}} \int_{-\infty}^{+\infty} \left\{ \mathrm{erf}\left(\frac{y_i + r}{\sigma\sqrt{2/\tau}} \right) - \mathrm{erf}\left(\frac{y_i - r}{\sigma\sqrt{2/\tau}} \right) \right\} e^{-y_i^2 r/2\sigma^2} dy_i.$$

The above expression can be approximated numerically. We set the following conditions for our numerical calculation: (1) $dx \to \Delta x = 1/5000$; (2) the range of the integration is $[-3,3]=[-(N/2)\Delta x,(N/2)\Delta x]$, with $N=30\,000$. Thus, we have

$$\frac{1}{2} \sqrt{\frac{\tau}{2\pi}} \sum_{k=-N}^{N} \left\{ \mathrm{erf}\left(\frac{k\Delta x + r}{\sqrt{2/\tau}} \right) - \mathrm{erf}\left(\frac{k\Delta x - r}{\sqrt{2/\tau}} \right) \right\} \times e^{[-(k\Delta x)^2 \tau]/2} \Delta x,$$

The values obtained with the above formula are plotted in Fig. 3. These numerical values are in good agreement with those obtained by the MSE algorithm on simulated white noise time series.

Next, we show the MSE derivation for $1/f$ noise. Note that a random process with a power spectrum that decays as $1/f$ is correlated. In order to numerically calculate S_E for $1/f$ noise, we will show that there exists an orthogonal transformation that maps the correlated variables into a basis in which they are independent. The dimension of this basis reflects the extension of the system "memory."

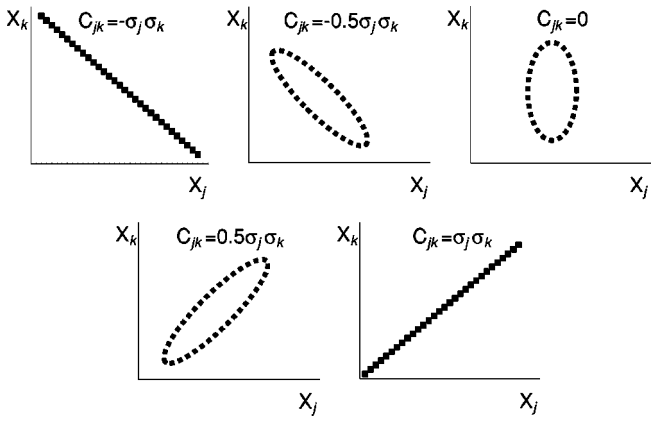

FIG. 12. Correspondence between the covariance and the shape of the contours of a bivariate Gaussian density function. If two random variables, X_j and X_k, are independent $[C_{jk}=C(X_j,X_k)=0]$, the shapes of the contours are ellipses with major and minor axes parallel to X_j and X_k axes, respectively. If the variables have equal variance ($\sigma_j=\sigma_k$), the shape of the contour is a circle. In contrast, if two variables are not independent, the shapes of the contours are ellipses with major and minor axes that are not aligned with the axes X_j and X_k.

Let us consider N random variables, $X_1, X_2, \ldots, X_N$, with mean values $\overline{X_j}$ for $j=1,\ldots,N$. Elements of the covariance matrix are defined by

$$C(X_j,X_k) = E[(X_j - \overline{X_j})(X_k - \overline{X_k})]. \tag{A5}$$

The diagonal elements are the variance of each random variable X_j, i.e., $C(X_j,X_j)=\sigma_j^2$ (see Fig. 12).

The covariance matrix is Hermitian since it is symmetric and all of its elements are real. Therefore, it has real eigenvalues whose eigenvectors form a unitary basis. Each of the eigenvectors, U_i, and the corresponding eigenvalues, λ_i, satisfy the equation

$$CU_j = \lambda_j U_j. \tag{A6}$$

Hence,

$$U_j^T C U_k = \lambda_k U_j^T U_k = \begin{cases} \lambda_k & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases}. \tag{A7}$$

Let U represent the matrix whose columns are the eigenvectors of the covariance matrix. Then,

$$U^T C U = \begin{bmatrix} \lambda_1 & 0 & \cdots & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & \cdots & \ddots & \cdots & 0 \\ 0 & \cdots & 0 & \lambda_{N-1} & 0 \\ 0 & \cdots & \cdots & 0 & \lambda_N \end{bmatrix} = \Lambda. \tag{A8}$$

We show next that $U^T C U$ is also the covariance matrix of the transformed vectors $Y=U^T X$, where $X=[X_1,X_2,\ldots,X_N]^T$,

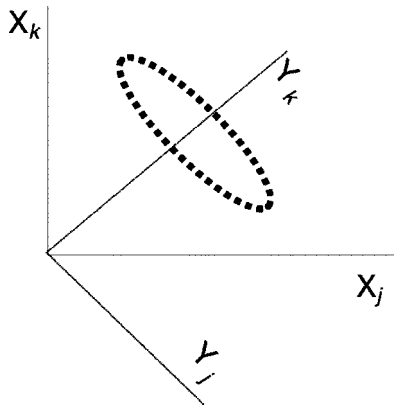

FIG. 13. The ellipse represents the contour of a bivariate Gaussian density function. The major and minor axes of the ellipse are not parallel to the axes X_j and X_k, meaning that the random variables are correlated in this frame. However, there exists a rotation that transforms the original frame into one defined by the axes Y_j and Y_k, which are aligned with the major and minor axes of the ellipse. Therefore, in this frame the original variables are uncorrelated.

$$\begin{aligned} U^T C U &= U^T E[(X-\overline{X})(X-\overline{X})^T]U = E[U^T(X-\overline{X})(X-\overline{X})^T U] \\ &= E[(U^T X - U^T\overline{X})(X^T U - \overline{X}^T U)] \\ &= E[(U^T X - U^T\overline{X})(U^T X - U^T\overline{X})^T] \\ &= E[(Y-\overline{Y})(Y-\overline{Y})^T]. \end{aligned}$$

Combining this result with Eq. (A8), we prove that all transformed variables are uncorrelated in the basis formed by the eigenvectors of the covariance matrix C. Furthermore, the variances, σ_j', of the transformed variables, Y_j, are $\sqrt{\lambda_j}$.

The physical meaning of the transformation U^T is illustrated in Fig. 13. U^T is an orthogonal transformation that amounts to a rotation of the original coordinate system into one defined by the eigenvectors of the covariance matrix, in which the transformed variables are independent.

The probability density function for an n-dimensional Gaussian random vector, X, is

$$p(X) = \frac{1}{\sqrt{(2\pi)^n |C|}} e^{[-(1/2)(X-\overline{X})^T C^{-1}(X-\overline{X})]}, \tag{A9}$$

where $|C|$ is the determinant of the covariance matrix.

For the transformed vector, $Y = U^T X$, the probability density function is

$$\begin{aligned} p(Y) &= \frac{1}{\sqrt{(2\pi)^n |\Lambda|}} e^{[-(1/2)(Y-\overline{Y})^T \Lambda^{-1}(Y-\overline{Y})]} \\ &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\lambda_i}} \exp - \frac{(Y_i - \overline{Y_i})^2}{2\lambda_i} = \prod_{i=1}^{N} p(Y_i), \end{aligned} \tag{A10}$$

where

$$p(Y_i) = \frac{1}{\sigma_i' \sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left(\frac{Y_i - \overline{Y_i}}{\sigma_i'} \right)^2 \right\}. \tag{A11}$$

In order to calculate the covariance matrix numerically, we limit the frequency range of the power spectral density, denoted as $S(\omega)$, of the $1/f$ noise signal to

$$S(\omega) = \begin{cases} K/\omega & \text{for } \omega_1 \leqslant \omega \leqslant \omega_2 \\ 0 & \text{otherwise,} \end{cases} \tag{A12}$$

where K is a constant. The upper and lower limits on frequency range are useful constraints for numerical calculation and also realistic in real-world applications where the resolution (sampling frequency of signal) and length of data are bounded.

The autocorrelation function, Φ, is obtained using the Wiener-Khintchine theorem,

$$\Phi(\tau) = \frac{K}{2\pi} \int_{\omega_1}^{\omega_2} \frac{\cos \omega\tau}{|\omega|} d\omega = \frac{K}{2\pi} \{ \mathrm{Ci}(\omega_2\tau) - \mathrm{Ci}(\omega_1\tau) \}, \tag{A13}$$

where τ represents the time lag and Ci is the cosine integral. The series expansion of the Ci is

$$\mathrm{Ci}(\tau) = \gamma + \ln(\tau) + \sum_{k=1}^{+\infty} \frac{(-1)^k \tau^{2k}}{(2k)!\, 2k}, \tag{A14}$$

where $\gamma = 0.5772\ldots$ is Euler's constant.

Therefore,

$$\Phi(\tau) = \frac{K}{2\pi} \left\{ \ln\left(\frac{\omega_2}{\omega_1} \right) + \sum_{k=1}^{+\infty} \frac{(-1)^k}{(2k)!\, 2k} \times [(\omega_2\tau)^{2k} - (\omega_1\tau)^{2k}] \right\}. \tag{A15}$$

The autocorrelation function is the autocovariance divided by the variance. For any ergodic process, as is the case of $1/f$ noise, the relation between the autocovariance function and the covariance matrix is

$$C = \begin{bmatrix} \Phi(0) & \Phi(\tau) & \Phi(2\tau) & \cdots & \Phi(N\tau) \\ \Phi(\tau) & \Phi(0) & \Phi(\tau) & \cdots & \Phi((N-1)\tau) \\ \Phi(2\tau) & \Phi(\tau) & \Phi(0) & \cdots & \Phi((N-2)\tau) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \Phi(N\tau) & \cdots & \cdots & \Phi(\tau) & \Phi(0) \end{bmatrix}. \tag{A16}$$

The eigenvalues of the covariance matrix are the variances of the transformed variables. Since the variables Y_i are independent, S_E is calculated using

$$p(Y_1) = \frac{1}{\sqrt{2\pi\lambda_1}} \exp\left(-\frac{[Y_1 - \overline{Y_1}]^2}{2\lambda_1} \right). \tag{A17}$$

We consider $k = \ln(\omega_1/\omega_2)$ for numerical calculation, which corresponds to normalizing the power spectrum. We also set $\omega_1 = 1/(2\Delta)$ and $\omega_2 = N$. The numerical calculation yields the value $S_E = 1.8$. We note that coarse-graining $1/f$ noise does not alter the correlation and the variance of the signal. Therefore, the S_E value calculated is valid for any scale.

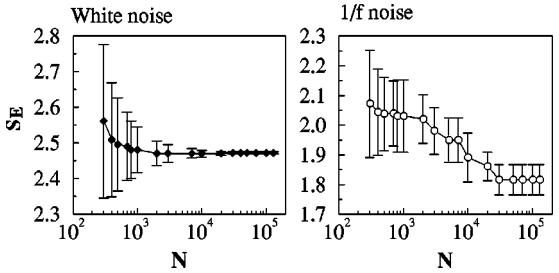

FIG. 14. S_E as a function of time series number of data points N. $r=0.15$ and $m=2$ for all time series. Symbols represent the mean values of S_E for 30 simulated white and $1/f$ noise time series, and the error bars represent the SD.

APPENDIX B: TECHNICAL ASPECTS OF MSE CALCULATIONS

1. Dependence on time series length and the values of parameters *m* and *r*

The MSE method uses the S_E family of statistics. Therefore, in this appendix we use simulated Gaussian distributed (mean zero, variance 1) white and $1/f$ noise time series to illustrate the effects on S_E of (i) the time series' finite length and (ii) the choice of parameters m and r.

Figure 14 shows that the mean value of S_E diverges as the number of data points decreases for both white and $1/f$ noise. However, since $1/f$ noise time series are not stationary, as the number of data points decreases, the discrepancy between the S_E value calculated numerically and the mean value for 30 simulated time series increases faster for $1/f$ noise than for white noise time series. For both types of noise, for $N=1\times10^5$, the discrepancy between the numerical and the mean value of S_E for simulated time series is less than 0.5%. However, for $N=1\times10^3$ the discrepancy between these values is approximately 12% in the case of $1/f$ noise but still less than 1% in the case of white noise. Furthermore, even for very large time series, the SD of S_E values for $1/f$ noise is never as small as for white noise. These results are due to the fact that stationarity is a basic requirement of S_E. The MSE method presents the same limitation. One possible solution to this problem is to decompose the original time signal into multiple "well-behaved" signals, each corresponding to different time scales.

We also note that as the number of data points decreases, the consistency of S_E results is progressively lost. Therefore, there is no guarantee that if S_E is higher for time series a than for time series b, both with N data points, the same result will hold if only N' data points are used to calculate S_E, in particular if $N \gg N'$ or $N' \gg N$.

We note that the coarse-graining procedure generates times series with a decreasing number of data points. However, coarse-grained time series are not a subset of the original time series. Instead, they contain information about the entire original time series. Therefore, the error due to the decrease of coarse-grained time series length is likely lower than that resulting from selecting a subset of the original time series.

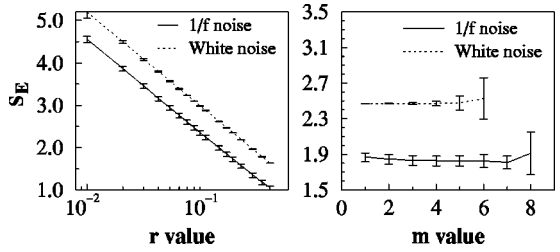

FIG. 15. S_E as a function of the parameter r (left plot) and m (right plot). $N=3\times10^4$ and $r=0.15$ for all time series. Symbols represent the mean values of S_E for 30 simulated $1/f$ and white noise time series, and error bars represent the SD.

As stated in Sec. II, the r value defines the similarity criterion used to compare vectors. If the absolute difference between any two matched vector components is larger than $r\times$SD, then the vectors are different; otherwise, they are considered equal. Theoretically, for continuous processes, r varies between 0 and 1; but for experimental time series, the recording resolution level determines the lowest possible r value. In any case, the actual r value determines the level of accepted noise, since for larger r values, fewer vectors are distinguishable. Figure 15 (left plot) shows that as the r value increases, the S_E value for both simulated $1/f$ and white noise time series decreases. Of note, the consistency of S_E values is preserved. Therefore, the SD of S_E values (error bars) reflects the scattering of values corresponding to different time series (intersubject variability).

Figure 15 (right plot) shows the variation of S_E with m value, i.e., the vector length. Between $m=1$ and $m=5$, the mean values of S_E vary less than 2% and the coefficient of variation (CV=SD/mean) is less than 3% for both types of noise. For larger m, both the S_E and the CV increase dramatically due to the finite number of data points, since longer and longer time series are required in order to calculate the frequency of the m and $(m+1)$-component vectors with sufficient statistical accuracy.

For a discussion of the optimal selection of m and r parameters, and the confidence intervals of S_E estimates, see [49]. We note that for $m=2$ and $r=0.15$, the discrepancies between the mean values of S_E for simulated time series and the numerically calculated values are less than 1% for both $1/f$ and white noises. This result suggests that for most practical applications, the error bars associated with computation of S_E values are likely smaller than the error bars related to experimental sources and also to inter- and intrasubject variability.

2. Effect of noise, outliers, and sample frequency

The output of an experiment may be contaminated by different types of noise. Here, we discuss the effects of MSE analysis of superimposing uncorrelated (white) noise on a physiologic time series. Common sources of uncorrelated noise for interbeat interval time series are the analog-digital conversion devices, whose accuracy depends both on the sample frequency and the number of bits used, and computer rounding errors. Figure 16 shows that (i) superimposing un-

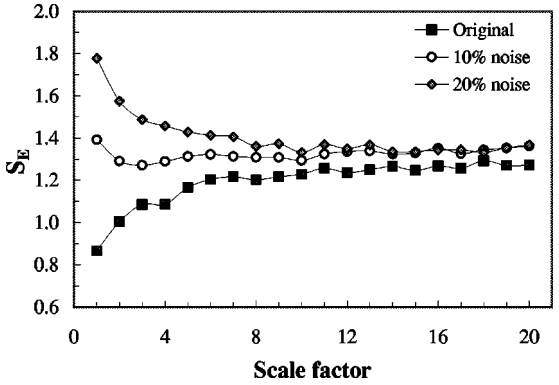

FIG. 16. Effects of different amounts of Gaussian white noise on MSE curves. The MSE curve labeled "original" corresponds to the MSE results for the RR intervals time series from a healthy subject.

correlated noise on a time series affects mainly the entropy values on small scales; (ii) the discrepancy between the entropy values assigned to the original time series and those assigned to time series with superimposed uncorrelated noise increases as the signal-to-noise ratio decreases; (iii) for small scales, S_E values monotonically decrease with scale factor similar to white noise time series. This effect becomes more prominent as the signal-to-noise ratio decreases.

Outliers may also affect S_E values because they change the time series SD and, therefore, the value of parameter r that defines the similarity criterion.

In the interbeat interval time series, two types of outliers are commonly found resulting from (i) missed beat detections by automated or visual electrocardiographic analysis, and (ii) recording artifacts [Fig. 18(a)]. These outliers do not have physiologic meaning. However, they may dramatically affect the entropy calculation if their amplitude is a few orders of magnitude higher than the mean value of the time series.

For the analysis of physiologic rhythm dynamics, cardiac beats not originating in the sinus node may be treated as outliers [Fig. 18(b)]. Of note, the amplitude of all cardiac (sinus and nonsinus) interbeat intervals is of the same order of magnitude. Therefore, the inclusion of a relatively low

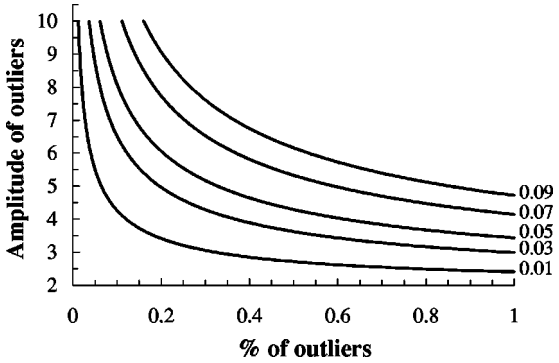

FIG. 17. Contour plot showing how the percentage of outliers and their amplitude (relative to the mean value of the time series) affects the variance of the time series. Lines connect pairs of values that change the variance by the same amount.

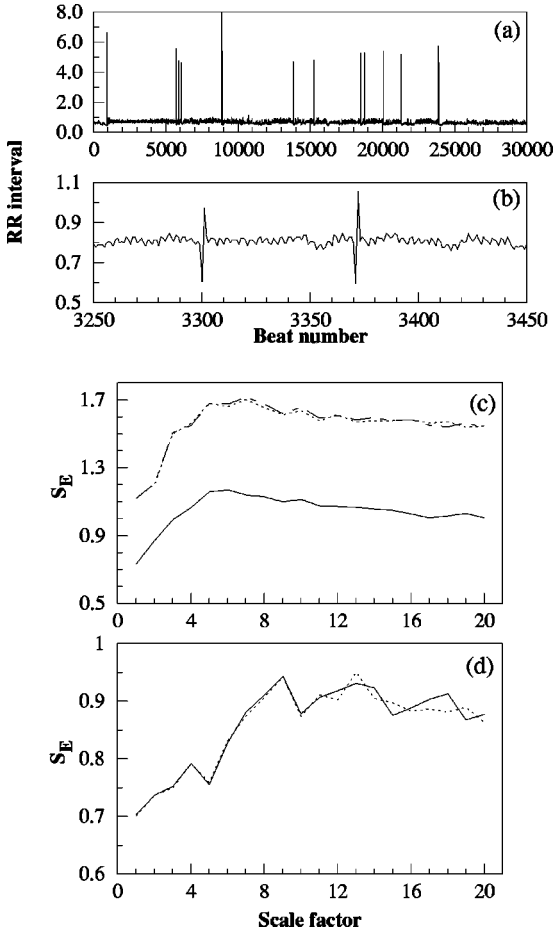

FIG. 18. (a) The interbeat interval time series of a young healthy subject with 15 outliers that represent artifacts or missed beat detections. Note that the absolute value of the outliers is much larger than the mean RR interval. (b) The interbeat interval time series of an elderly healthy subject with frequent premature ventricular complexes (PVCs) (two are represented in the figure). (c) MSE results for the time series shown in plot (a): the solid line is the MSE result for the unfiltered time series; the dotted line is the MSE results for the same time series excluding outliers; and the dashed line is the MSE result for the original time series but using an r value that is calculated by excluding the outliers. (d) MSE results for time series shown in plot (b): solid and dotted lines are the MSE results for unfiltered and filtered (PVCs removed) time series.

percentage of nonsinus beats should not significantly change the entropy values.

Consider a time series, X, with N data points, M of which are outliers with amplitude Δ. Let X' represent the time series that is obtained from the time series X by excluding the outliers. Assume that $M \ll N$ and that $\Delta = a\overline{X'}$, where $\overline{X'}$ is the time series mean value. It can be shown that $\sigma^2(X) - \sigma^2(X') = (a^2\epsilon - \epsilon^2 a^2 - 2\epsilon a)\mu(X')^2$, where $\epsilon = M/N$, and σ and μ are the time series SD and mean value, respectively.

Figure 17 shows that a small number of outliers with high amplitude has similar effects on the variance as a higher percentage of outliers with lower amplitude.

Figure 18(a) presents a time series with 0.05% outliers which account for an increase in the time series SD of about 44%. Figure 18(b) presents a time series with approximately

ten times more outliers than in Fig. 18(a). Since the amplitude of the outliers is of the same order of magnitude as the remaining data points, the difference between the SD of the time series which includes these outliers and that which excludes them is only 1%.

Changes of the time series SD proportionally affect the value of parameter r. Higher r values mean that fewer vectors will be distinguishable and that the time series will appear more regular. Figure 18(c) presents the MSE results for the unfiltered time series (a) (solid line) and the corresponding time series obtained by excluding the outliers (dotted line). As expected, the MSE curve corresponding to the unfiltered time series is lower than the MSE curve corresponding to the filtered time series.

The presence of a small percentage of outliers may significantly alter the SD but should not substantially modify the temporal structure of the time series. In Fig. 18(c), the dashed line represents the MSE results for the unfiltered time series obtained using the r value derived from the filtered time series. Note that when using the "correct" r value, the MSE curves for the unfiltered and the filtered time series overlap.

Figure 18(d) compares the MSE results for time series (b) and for the time series that results from excluding the outliers. The two MSE curves almost overlap, showing that the entropy measure is robust to the presence of a relatively small percentage of low-amplitude outliers.

For a time series sampled at frequency f, the temporal location of the actual heartbeat can be identified only up to an accuracy of $\Delta=1/f$. Each data point of a coarse-grained heartbeat interval time series is an average of consecutive differences. For example, $y_1^\tau=(RR_1+\cdots+RR_\tau)/\tau=[(t_2-t_1)+\cdots+(t_\tau-t_{\tau-1})]=(t_\tau-t_1)/\tau$. Therefore, the accuracy of averaged heartbeat intervals of coarse-grained time series is Δ/τ, i.e., the accuracy increases with scale.

S_E is underestimated for finite sample frequency values [48]. However, the discrepancy between the value of S_E calculated for a time series sampled at a finite frequency and the value of S_E corresponding to the limit $\lim_{\Delta\to 0}S_E$ decreases with scale. For analysis on small time scales, it may be important to consider a correction of this effect [48]. We note that the conclusions that we present in this paper are not altered by the value of sample frequency.

APPENDIX C: MSE ANALYSIS OF DISCRETE TIME SERIES

Here we discuss an important artifact that affects the MSE analysis of discrete time series, such as DNA sequences.

Let us consider an uncorrelated random variable, X, with alphabet $\Theta=\{0,1\}$. Both symbols occur with probability $1/2$.

All possible different two-component sequences built from the binary series are 00, 01, 10, and 11. Therefore, the alphabet of the coarse-grained time series corresponding to scale 2 is $\Theta_2=\{0,1/2,1\}$. The probabilities associated with the occurrence of the different values are $1/4$, $1/2$, and $1/4$, respectively. Let us consider that the r value used to calculate S_E is 0.5. In this case, only the distance between the coarse-grained values 0 and 1 (and not between values 0 and $1/2$, and between $1/2$ and 1) is higher that r. Therefore, the probability of distinguishing two data points randomly chosen from the coarse-grained time series, $P_r(|x_a-x_b|>r)$, is $p(0)\times p(1)=1/4\times 1/4=1/16=0.0625$.

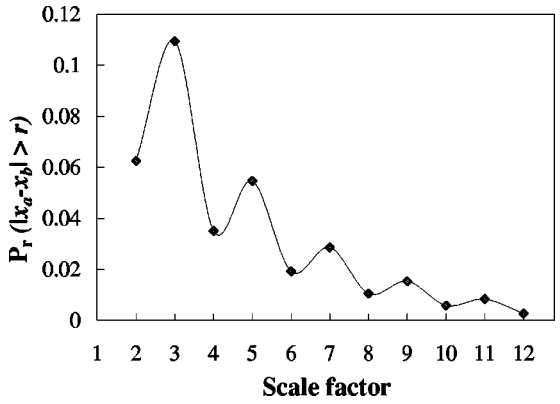

FIG. 19. Probability of distinguishing any two data points randomly chosen from the coarse-grained time series of binary discrete time series ($r=0.5$).

Similarly, there are eight different three-component sequences that can be built from the original binary series: 000, 001, 010, 100, 110, 011, 101, and 111. Consequently, the alphabet of the coarse-grained time series corresponding to scale 3 is $\Theta_2=\{0,1/3,2/3,1\}$ and the probabilities associated with the occurrence of each value are $1/8$, $3/8$, $3/8$, and $1/8$, respectively. For $r=0.5$, only the distances between the coarse-grained data points 0 and $2/3$, $1/3$ and 1, and 0 and 1 are higher than r. Therefore, $P_r(|x_a-x_b|>r)=p(0)\times p(2/3)+p(1/3)\times p(1)+p(0)\times p(1)=0.1094$.

Note that the probability of distinguishing two data points of the coarse-grained time series increases from scale 2 to scale 3 (Fig. 19). As a consequence, S_E also increases, contrary to both analytic and numerical results presented in Fig. 3. This artifact, which affects discrete time series, is due to the fact that the size of the alphabet of the coarse-grained time series increases with scale.

In general, for scale n, the alphabet set is $\Theta_n=\{i/n\}$ with $0\leq i\leq n$, and the corresponding probability set $\{p(i/n)\}$ is generated by the expression $n!/[2^n\times i!(n-i)!]$, $0\leq i\leq n$. The value of $P_r(|x_a-x_b|>r)$ is calculated by the equation

$$P_r(|x_a-x_b|>r)=\sum_{j=0}^{N-1}p(j/n)\sum_{i=i'}^{n}p(i/n), \qquad \text{(C1)}$$

where $i'=N+j+1$ if $n=2N$ (even scales) and $i'=N+j$ if $n=2N-1$ (odd scales).

Figure 19 shows how the probability varies with the scale factor. We note an attenuated oscillation, which as a consequence also shows up on the MSE output curve for the same time series. The period of this oscillation depends only on the r value.

To overcome this artifact, one approach is to select the scales for which the entropy values are either local minima or maxima of the MSE curve. We adopted this procedure in calculating the complexity of coding versus noncoding DNA sequences (Fig. 10). Note that for uncorrelated random bi-

nary time series (Fig. 19), and for $r=0.5$, the sequence of entropy values at odd or even scales monotonically decreases with scale factor, similar to the MSE curve for white noise time series, as described in Sec. III (Fig. 3).

An alternative approach is to map the original discrete time series to a continuous time series, for example by counting the number of symbols (1's or 0's) in nonoverlapping windows of length 2^n. Since this procedure is not a one-to-one mapping, some information encoded on the original time series is lost. Therefore, relatively long time series are required. We adopted this procedure in calculating the complexity of binary time series derived from a computer executable file and a computer data file (Fig. 9).

[1] F. Takens, in *Dynamical Systems and Turbulence*, edited by D. A. Rand and L. S. Young. Lecture Notes in Mathematics Vol. 898 (Springer, Berlin, 1981), p. 366.

[2] J.-P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).

[3] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, Physica D **58**, 77 (1992).

[4] S. M. Pincus, Ann. N.Y. Acad. Sci. **954**, 245 (2001), and references therein.

[5] P. Grassberger in *Information Dynamics*, edited by H. Atmanspacher and H. Scheingraber (Plenum, New York, 1991), p. 15.

[6] B.-Y. Yaneer, *Dynamics of Complex Systems* (Addison-Wesley, Reading, Massachusetts, 1997).

[7] M. Costa, A. L. Goldberger, and C.-K. Peng, Phys. Rev. Lett. **89**, 068102 (2002).

[8] M. Costa, A. L. Goldberger, and C.-K. Peng, Comput. Cardiol. **29**, 137 (2002).

[9] M. Costa and J. A. Healey, Comput. Cardiol. **30**, 705 (2003).

[10] M. Costa, A. L. Goldberger, and C.-K. Peng, Phys. Rev. Lett. **92**, 089804 (2004).

[11] M. Costa, C.-K. Peng, A. L. Goldberger, and J. M. Hausdorff, Physica A **330**, 53 (2003).

[12] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).

[13] R. Shaw, Z. Naturforsch. A **36**, 80 (1981).

[14] P. Grassberger and I. Procaccia, Physica D **56**, 189 (1983).

[15] P. Grassberger and I. Procaccia, Phys. Rev. A **28**, 2591 (1983).

[16] F. Takens, in *Proceedings of the 13th Colóquio Brasileiro de Matemática* (Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1983).

[17] S. M. Pincus, Proc. Natl. Acad. Sci. U.S.A. **88**, 2297 (1991).

[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991), p. 64.

[19] S. M. Pincus, I. M. Gladstone, and R. A. Ehrenkranz, J. Clin. Monit. **7**, 335 (1991).

[20] J. S. Richman and J. R. Moorman, Am. J. Physiol. **278**, H2039 (2000).

[21] P. Grassberger, T. Schreiber, and C. Schaffrath, Int. J. Bifurcation Chaos Appl. Sci. Eng. **1**, 521 (1991).

[22] D. P. Feldman and J. P. Crutchfield, Phys. Lett. A **238**, 244 (1998).

[23] Y.-C. Zhang, J. Phys. I **1**, 971 (1991).

[24] A. L. Goldberger, C.-K. Peng, and L. A. Lipsitz, Neurobiol. Aging **23**, 23 (2002).

[25] A. J. Mandell and M. F. Shlesinger in *The Ubiquity of Chaos*, edited by S. Krasner (American Association for the Advancement of Science, Washington, D.C., 1990), p. 35.

[26] M. P. Paulus, M. A. Geyer, L. H. Gold, and A. J. Mandell, Proc. Natl. Acad. Sci. U.S.A. **87**, 723 (1990).

[27] H. C. Fogedby, J. Stat. Phys. **69**, 411 (1992).

[28] V. V. Nikulin and T. Brismar, Phys. Rev. Lett. **92**, 089803 (2004).

[29] J. E. Mietus, C.-K. Peng, I. Henry, R. L. Goldsmith, and A. L. Goldberger, Heart **88**, 378 (2002).

[30] The New York Heart Association functional classification is used to characterize patients' limitations from left ventricular failure. Subjects assigned to class I can perform ordinary physical exercise with no limitations. Subjects assigned to class II are comfortable at rest but experience fatigue or shortness of breath when performing ordinary physical exercise. Class III subjects are also comfortable at rest but their ability to exercise is markedly reduced. Class IV comprises those subjects who have symptoms at rest.

[31] Time series derived from subjects with atrial fibrillation have statistical properties similar to those of white noise on shorter time scales ($\lesssim$200 s). For more details, see [45–47].

[32] K. K. L. Ho, G. B. Moody, C.-K. Peng, J. E. Mietus, M. G. Larson, D. Levy, and A. L. Goldberger, Circulation **96**, 842 (1997).

[33] A. Bunde, S. Havlin, J. W. Kantelhardt, T. Penzel, J.-H. Peter, and K. Voigt, Phys. Rev. Lett. **85**, 3736 (2000).

[34] T. Cavalier-Smith, in *The Evolution of Genome Size*, edited by T. Cavalier-Smith (Wiley, Chichester, U.K., 1985).

[35] J. S. Mattick, BioEssays **25**, 930 (2003).

[36] J. S. Mattick, EMBO Rep. **2**, 986 (2001).

[37] C.-K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, F. Sciortino, M. Simons, and H. E. Stanley, Nature (London) **356**, 168 (1992).

[38] C.-K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, R. N. Mantegna, M. Simons, and H. E. Stanley, Physica A **221**, 180 (1995).

[39] S. V. Buldyrev, A. L. Goldberger, S. Havlin, R. N. Mantegna, M. E. Matsa, C.-K. Peng, M. Simons, and H. E. Stanley, Phys. Rev. E **51**, 5084 (1995).

[40] B. Audit, C. Thermes, C. Vaillant, Y. d'Aubenton-Carafa, J. F. Muzy, and A. Arneodo, Phys. Rev. Lett. **86**, 2471 (2001).

[41] B. Audit, C. Vaillant, A. Arneodo, Y. d'Aubenton-Carafa, and C. Thermes, J. Mol. Biol. **316**, 903 (2002).

[42] L. A. Lettice *et al.*, Proc. Natl. Acad. Sci. U.S.A. **99**, 7548 (2002).

[43] M. A. Nobrega, I. Ovcharenko, V. Afzal, and E. M. Rubin, Science **302**, 413 (2003).

[44] Consider a time series with only two symbols: 0 and 1. The coarse-grained time series corresponding to scale τ contains the symbols $0, 1/\tau, \ldots, i/\tau, \ldots 1 (0 \leqslant i \leqslant \tau)$. If the time series is the output of a stochastic process without correlations and all values are equally probable, then the entropy of the process is $S=-\Sigma_{i=1}^{N} p_i \log p_i = \log N$, where N is the total number of data

points. Therefore, entropy monotonically increases as the number of symbols increases.

[45] J. Hayano, F. Yamasaki, S. Sakata, A. Okada, S. Mukai, and T. Fujinami, Am. J. Physiol. **273**, H2811 (1997).

[46] W. Zeng and L. Glass, Phys. Rev. E **54**, 1779 (1996).

[47] R. Balocchi, C. Carpeggiani, L. Fronzoni, C.-K. Peng, C. Michelassi, J. Mietus, and A. L. Goldberger, in *Methodology and Clinical Applications of Blood Pressure and Heart Rate Analysis*, edited by M. Rienzo, G. Mancia, G. Parati, A. Pedotti, and A. Zanchetti (Ios Press Inc., Amsterdam, 1999), p. 91.

[48] D. E. Lake and R. J. Moorman (private communication).

[49] D. E. Lake, J. S. Richman, M. P. Griffin, and J. R. Moorman, Am. J. Physiol. **283**, R789 (2002).