Multiscale Entropy to Distinguish Physiologic and Synthetic RR Time Series

M Costa[1,2], AL Goldberger[1], C-K Peng[1]

[1]Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, USA
[2]Institute of Biophysics and Biomedical Engineering, Faculty of Sciences of the University of Lisbon, Lisbon, Portugal

Abstract

We address the challenge of distinguishing physiologic interbeat interval time series from those generated by synthetic algorithms via a newly developed multiscale entropy method. Traditional measures of time series complexity only quantify the degree of regularity on a single time scale. However, many physiologic variables, such as heart rate, fluctuate in a very complex manner and present correlations over multiple time scales. We have proposed a new method to calculate multiscale entropy from complex signals. In order to distinguish between physiologic and synthetic time series, we first applied the method to a learning set of RR time series derived from healthy subjects. We empirically established selected criteria characterizing the entropy dependence on scale factor for these datasets. We then applied this algorithm to the CinC 2002 test datasets. Using only the multiscale entropy method, we correctly classified 48 of 50 (96%) time series. In combination with Fourier spectral analysis, we correctly classified all time series.

1. Introduction

Heart rate variability is the output of multiple physiologic control mechanisms that operate on a wide range of time scales. As a result, cardiac interbeat (RR) time series under healthy conditions have a complex temporal structure with multiscale correlations [1, 2]. In contrast, synthetic time series are most likely the output of simpler dynamical systems, and therefore, will be anticipated to have less complex temporal structures than their "true" biologic counterparts.

Classical entropy and physiologic complexity concepts do not have a straightforward correspondence. Entropy is related to the degree of "randomness" of a time series and it is maximum for completely uncorrelated random signals. Complexity is related to the underlying structure of a time series and its information content. An increase of the entropy assigned to a time series usually, but not always, corresponds to an increase of underlying system complexity. Entropy-based algorithms [3, 4] for measuring the complexity of physiologic time series have been widely used. They have proved to be useful in discriminating between healthy and disease states [5, 6], although some results may lead to misleading conclusions. For example, the entropy that these algorithms assign to time series derived of the ventricular response in atrial fibrillation is much higher than that assigned to sinus rhythm time series derived from healthy subjects. However, healthy systems generate much more complex outputs than diseased ones. Traditional algorithms are single-scale based and therefore fail to account for the multiple time scales inherent in physiologic systems. We have proposed a new method [7] to calculate multiscale entropy (MSE) from complex signals.

In 1991, Zhang [8, 9] proposed a new complexity measure applicable to physical systems. His measure, defined as a weighted sum of scale-dependent entropies, has the desirable property of yielding higher values for correlated noises than for uncorrelated ones. However, since it is based on Shannon's definition of entropy, it requires a huge number of almost noise-free data points [10]. Therefore, the possibility of applying Zhang's measure to real world biologic time series is very limited. In contrast, our related method is based on the approximate entropy (ApEn) family of parameters, which have been widely applied to physiologic and medical time series analysis [3].

2. Methods

We briefly describe the MSE method.

Given a time series, $\{x_1, \ldots, x_i, \ldots, x_N\}$, we first construct consecutive coarse-grained time series by averaging a successively increasing number of data points in non-overlapping windows (Figure1). Each element of the coarse-gained time series, $y_j^{(\tau)}$, is calculated accordingly to the equation:

$$y_j^{(\tau)} = 1/\tau \sum_{i=(j-1)\tau+1}^{j\tau} x_i. \qquad (1)$$

0276−6547/02 $17.00 © 2002 IEEE

Computers in Cardiology 2002;29:137−140.

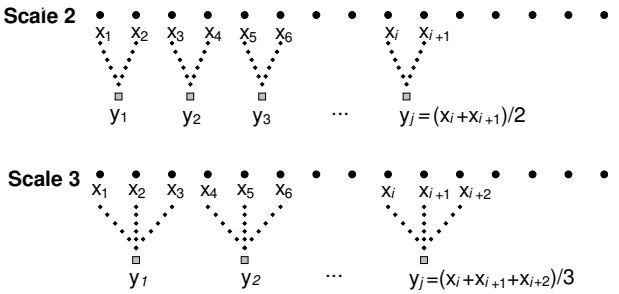

Figure 1. Schematic illustration of the coarse-graining procedure for scales 2 and 3.

where τ represents the scale factor and $1 \leq j \leq N/\tau$. For scale 1, the coarse-grained time series is simply the original time series.

Then, we calculate sample entropy (SampEn) [4], a refinement of the original ApEn statistics [3], for each coarse-grained time series plotted as a function of the scale factor τ.

The MSE method applied to white and $1/f$ noises, i.e., uncorrelated and correlated fluctuations (Figure 2), shows that for scale 1 the entropy for white noise is much higher than for $1/f$ noise [7]. However, while the entropy for $1/f$ noise remains almost constant for all scales, the entropy for white noise monotonically decreases with scale such that, for scales larger than 4 it is lower than the entropy for $1/f$ noise. This result is consistent with the fact that, unlike white noise, 1/f noise contains structures across multiple time scales.

The MSE method applied to the cardiac interbeat interval time series derived from young and elderly healthy subjects, subjects with congestive heart failure (CHF) and subjects with atrial fibrillation (AF), reveals that complexity degrades with disease and aging [7]. For scale one, AF times series are assigned the highest entropy value and CHF time series and time series derived from healthy subjects are assigned similar entropy values. However, for larger time scales, we verify that: a) the entropy for AF time series monotonically decreases similar to white noise, and for scales larger than 10 is lower than the entropy assigned to times series derived from healthy subjects; b) the entropy for CHF time series is lower than that for time series derived from healthy subjects on all time scales but the first one. In addition, the poorest separation between young and elderly healthy subjects occurs for scale one, the only scale that is traditionally studied. Therefore, MSE results are compatible with the concept that youthful healthy systems are the most complex ones.

In order to distinguish between the physiologic and the synthetic time series made available for the CinC 2002 challenge, we first applied the MSE method to a training set of cardiac interbeat interval time series derived from 20

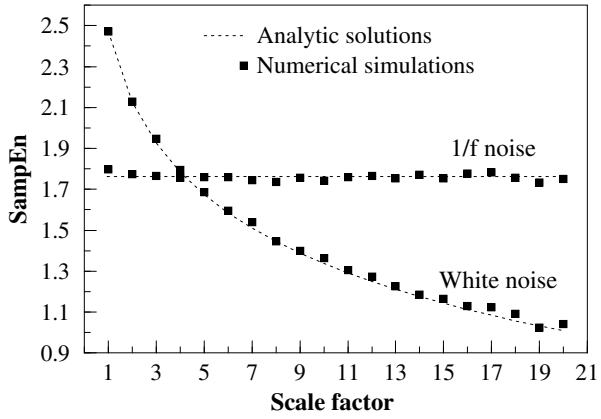

Figure 2. MSE analysis of Gaussian distributed white noise (mean zero, variance one) and $1/f$ noise. On the y-axis, the value of sample entropy (SampEn) [4] for the coarse-grained time series is plotted. Parameters are: $N = 3 \times 10^4$ data points, $m = 2$ and $r = 0.15$. The scale factor specifies the number of data points averaged to obtain each element of the coarse-grained time series. Symbols represent results of simulations and dotted lines represent analytic results. SampEn for coarse-grained white noise time series, is analytically calculated by the expression: $-\ln \int_{-\infty}^{+\infty} \frac{1}{2}\sqrt{(\frac{\tau}{2\pi})}[\,\mathrm{erf}(\frac{x+r}{\sqrt{(2/\tau)}}) - \mathrm{erf}(\frac{x-r}{\sqrt{(2/\tau)}})]\, e^{-\frac{1}{2}x^2\tau} dx$ (for any $m \geq 1$). τ and erf refer to the scale factor and to the error function, respectively. r is defined in Ref. [3]. For $1/f$ noise time series, the analytic value of SampEn is a constant. Adapted from Ref [7].

healthy elderly subjects, 10 males and 10 females (mean age$\pm$SD, 69 ± 3 years), and 20 healthy young subjects, 10 male and 10 female (mean age$\pm$SD, 32 ± 6 yr). Then, we empirically established selected criteria characterizing the entropy dependence on scale factor for these healthy subjects. Next, we applied the algorithm to the CinC 2002 test datasets.

3. Results

In Figure 3 we present the results of the MSE method for the training set that includes 20 healthy elderly subjects (mean age$\pm$SD, 69 ± 3 years) and 20 healthy young subjects (mean age$\pm$SD, 32 ± 6 years). Two types of curves are characteristic of healthy systems. For young subjects, the entropy for coarse-grained time series increases up to approximately time scale 5 and then stabilizes for larger time scales. For elderly subjects, entropy for coarse-grained time series initially decreases slightly and then progressively increases. For larger time scales it tends to stabilize. Using this training set we defined the range of physiologically meaningful entropy values. The upper and

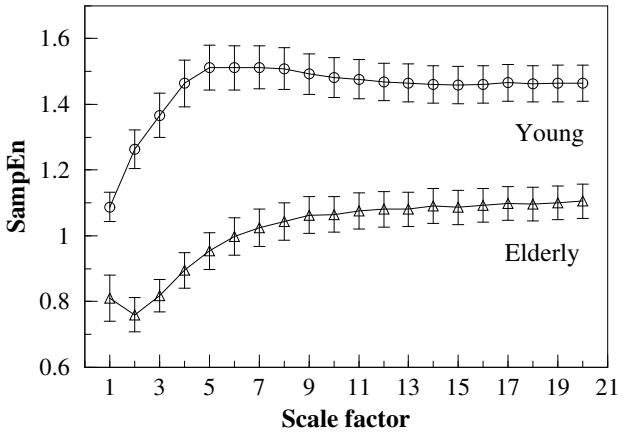

Figure 3. MSE analysis of the cardiac interbeat time series derived from 20 healthy young subjects and 20 healthy elderly subjects. Values are given as means ± standard error. Parameters for calculating SampEn are $m = 2$, $r = 0.15$ and $N = 3 \times 10^4$. For all time scales, the values of entropy for coarse-grained time series obtained from elderly subjects are significantly ($p < 0.005$; t-test) lower than those from young subjects. The poorest separation between groups is obtained for scale one, indicating the importance of calculating entropy over different scales. Adapted from Ref. [7].

the lower limits of this range are set as mean value $\pm 2SD$, respectively. MSE curves with similar patterns to those presented in Figure 3 and for which the entropy values are within the range defined by the physiologic intervals were considered as belonging to physiologic time series. In all other cases, the MSE curves were considered to have derived from synthetic time series.

We applied the MSE method to all CinC 2002 datasets. Results are presented in Figures 4 and 5. For easier interpretation of the results, these figures include also the mean values (symbols) and SD (error bars) of entropy for coarse-grained time series derived from our training set. (Instead of considering the results for young and elderly subjects separately, as in Figure 3, in Figures 4 and 5 mean values were calculated by pooling both sub-groups.)

The patterns of all MSE curves included in Figure 4 are similar to those presented in Figure 3. In addition, entropy values are within the limits defined by the physiologic range. Therefore these curves were identified as belonging to physiologic time series.

In Figure 5, top panel, we present all MSE curves in which the entropy monotonically decreases for more than 7 consecutive time scales. This pattern it similar to the one obtained with white noise, in which case the entropy is a monotonic decreasing function of the scale factor (Figure 2). The results suggest that all these time series have a common underlying random structure and therefore were classified as synthetic.

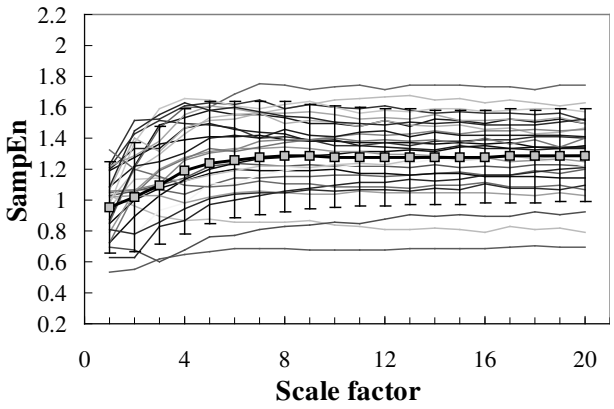

Figure 4. MSE results for time series from CinC 2002 Challenge identified as physiologic. Parameters for calculating SampEn are $m = 2, r = 0.15$ and $N = 4 \times 10^4$. Symbols and errors bars refer to mean and SD values of MSE results obtained with the training set considering both young and elderly healthy subjects.

In the bottom panel of Figure 5, we present the remaining MSE curves. For all these curves one or both of the following two situations occur: a) the entropy for at least one coarse-grained time series is out of the physiologic range; b) the entropy monotonically increases for large time scales defining a pattern not yet found for healthy physiologic systems. These curves were also classified as synthetic.

With the MSE method, 20 out of 22 synthetic time series and 28 out of 28 time series derived from healthy subjects were correctly identified, which yields a 96% success rate. (Two time series not identified by the MSE method could be excluded from the physiologic group based on the fact that their power spectra display a pure $1/f$ decay without any physiologic peak.) Comparable separation could not be achieved with traditional (single-scale) entropy measures. For example, the values of SampEn for 11 synthetic time series (scale 1 of Figures 4 and 5) overlap with those measured for time series derived from physiologic systems (scale 1 of Figure 3).

4. Conclusions

MSE analysis has been proposed to quantify the complexity of physical and physiologic time series. We previously applied MSE to correlated and uncorrelated noises and to physiologic time series under healthy and pathologic conditions [7]. Using the MSE method,

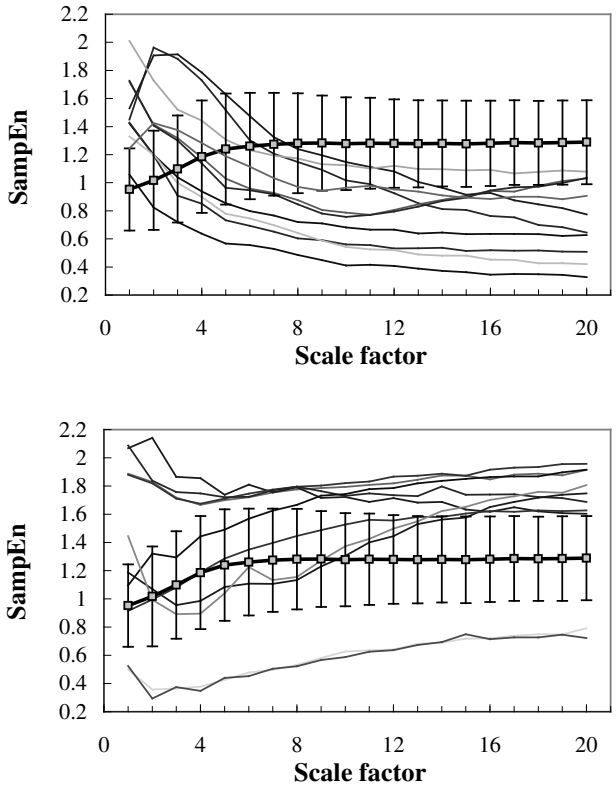

Figure 5. MSE results for time series from CinC 2002 Challenge identified as synthetic. Parameters for calculating SampEn are $m = 2$, $r = 0.15$ and $N = 4 \times 10^4$. For ease of visualization, results are grouped in two panels. Symbols and error bars represent mean and SD values of MSE results obtained for our training set which includes both young and elderly healthy subjects. In the top panel, 10 MSE curves are presented in which entropy for coarse-grained time series monotonically decreases for several time scales. In the bottom curve, all other time series identified as synthetic are presented. For all these MSE curves, entropy values either lie outside the defined physiologic range or monotonically increase for larger time scales.

correlated 1/f noise consistently shows higher complexity than white noise. Further, the complexity of heartbeat time series degrades with aging and disease. This finding is compatible with the unifying concept that physiologic complexity is fundamentally related to the adaptive capacity of the organism, requiring integrative and multiscale functionality. Finally, when applied to the CinC 2002 contest, the MSE method correctly identified the origin of 48 out of 50 time series. The MSE method seems to have the capacity to distinguish between time series generated by different mechanisms. Furthermore, it may be applied to a wide variety of other physiologic and physical time series.

References

[1] Goldberger AL, Peng CK, Lipsitz LA. What is physiologic complexity and how does it change with aging and disease? Neurobiol Aging 2002;23:23–26.

[2] Goldberger AL, Amaral LAN, Hausdorff JM, Ivanov PC, Peng CK, Stanley HE. Fractal dynamics in physiology: alterations with disease and aging. Proc Natl Acad Sci USA 2002;99(Suppl1):2466–272.

[3] Pincus SM. Approximate entropy as a measure of system complexity. Proc Natl Acad Sci USA 1991;88:2297–2301.

[4] Richman JS, Moorman JR. Physiologic time series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ 2000;278:H2039–2049.

[5] Pincus SM. Assessing serial irregularity and its implications for health. Ann N Y Acad Sci 2001;954:245–267, and references therein.

[6] Lake DE, Richman JS, Griffin MP, Moorman JR. Sample entropy analysis of neonatal heart rate variability. Am J Physiol Heart Circ 2002;283:R789–797.

[7] Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of complex physiologic time series. Phys Rev Lett 2002;89:068102.

[8] Zhang YC. Complexity and 1/f noise: a phase space approach. J Phys I France 1991;1:971–977.

[9] Fogedby HC. On the phase space approach to complexity. J Stat Phys 1992;69:411–425.

[10] Eckmann JP, Ruelle D. Ergodic theory of chaos and strange attractors. Rev Mod Phys 1985;57:617–656.

Address for correspondence:

Madalena Damasio Costa
Beth Israel Deaconess Medical Center
Kirtein Building, Room KB-30
330 Brookline Avenue, Boston, USA
tel./fax: ++1-617-667-2428/8052
madalena@mimic.bidmc.harvard.edu