A Characteristic Ridge-Ledge in Entropy Surfaces of Cardiovascular Time Series Estimated by the Norm Component Matrix Algorithm

S Żurek[1], P Castiglioni[2], M Kośmider[1], G Parati[3], P Guzik[4], J Piskorski[1]

[1]Institute of Physics, University of Zielona Góra, Zielona Góra, Poland
[2]Don C. Gnocchi Foundation, Milan, Italy
[3]Istituto Auxologico Italiano, Milan, Italy
[4]Department of Cardiology - Intensive Therapy and Internal Diseases, Poznan University of Medical Sciences, Poznań, Poland

Abstract

Sample entropy (SampEn) is a popular complexity measure in HRV analysis. SampEn is estimated by fixing the values of the embedding dimension "m" and distance threshold "r" and traditionally SampEn is calculated with m=2 and r=0.2 times the standard deviation of the series.

Attempts to extend the estimates to different (m,r) pairs are hampered by the high computational burden of the traditional algorithm. We recently proposed an extremely fast Norm Component Matrix (NCM) algorithm for SampEn calculation which allows analyzing whole ranges of (m,r) values leading to entropy surfaces. This paper is the first attempt to calculate these surfaces by NCM and to describe their properties for both synthetic and physiological time series.

We show that the entropy surface has a characteristic ridge-ledge structure associated with the randomness in the data.

1. Introduction

Approximate Entropy, *ApEn* [1], and Sample Entropy, *SampEn* [2], are popular methods for measuring the complexity of cardiovascular signals. They estimate the entropy of a series of N data points by calculating the probability that segments of "m" samples which are similar, i.e., within a given distance d, remain similar when the segments length increases to "$m+1$". The parameter "m" is called the embedding dimension, and the threshold distance d is expressed as a fraction r of the standard deviation (SD), of the time series: $d = r \times SD$.

Entropy estimation by *ApEn* or *SampEn* therefore depends on the choice of m and r. Previous analyses with both deterministic and stochastic processes suggested good statistical performances for *ApEn* by selecting m=2 and r between 0.10 and 0.25 [3, 4]. Following this initial suggestion, now most of heart-rate variability (HRV) studies calculate *ApEn* and *SampEn* with m=2 and r=0.20. However, this choice appears arbitrary, and recently it has been shown that arbitrary selections of r within the suggested range (0.10-0.25) may even lead to contradictory results in HRV analysis [5]. To eliminate this arbitrariness, selecting the r value which maximizes the entropy estimate has been suggested [6]. This approach removes ambiguities in the determination of signal complexity when synthesised processes are compared [6]. However, it remains unknown whether a maximum can be always detected univocally in *ApEn* and *SampEn* of real cardiovascular signals, independently from the choice of m. This is due to the high computational burden required for obtaining the structure of *ApEn(m,r)* and *SampEn(m,r)* with sufficient detail .

Very recently, however, our group explored the relationships between correlation dimension and entropy, and proposed a very fast algorithm for calculating *ApEn* and *SampEn* over a whole range of m and r values [7]. This algorithm allows describing in detail the structure of *ApEn(m,r)* or *SampEn(m,r)* surfaces even for real physiological time series of long duration.

The aim of this work is therefore to provide, for the first time, a description of the structure of *SampEn(m,r)* surfaces, for real heart rate data, assessing the feasibility of this approach for the quantification of entropy characteristics even for time series of long duration, as 24-h ECG Holter recordings.

2. Methods

The sample and approximate entropy parameters as well as the correlation dimension are closely related through the correlation sums:

$$C^m(r) = L_m^{-1} \sum_{i=1}^{L_m} C_i^m(r). \quad (1),$$

with

$$C_i^m(r) = (L_m - 1)^{-1} \sum_{j=1, j\neq i}^{L_m} \Theta(r - |\vec{v}_m(i) - \vec{v}_m(j)|), \quad (2)$$

where Θ is the *Heaviside* function:

$$\Theta(x) = \begin{cases} 0 & if\ x \geq 0 \\ 1 & if\ x < 0, \end{cases} \quad (3)$$

and the distance between two vectors $|\cdot|$ is defined as *maximum coordinate difference* [2]:

$$|\vec{v}_m(i) - \vec{v}_m(j)| = \max_{k=1,2,\ldots,m} (|u(i + (k-1)\tau) - u(j + (k-1)\tau)|). \quad (4)$$

The vectors compared in eq. 2 are $\vec{v}_{m,\tau}(i) = [u(i), u(i+\tau), u(i+2\tau), \ldots, u(i+(m-1)\tau)]$ so that $\vec{v}_{m,\tau}(i)$ consists of m consecutive points in the data set, τ is the time lag. As demonstrated in [7], from the above all the three complexity parameters may be extracted. SampEn is obtained through

$$SampEn(m, r, N) = lnC^m(r) - lnC^{m+1}(r). \quad (5)$$

2.1. NCM Algorithm

The Norm Component Matrix (NCM) algorithm is the computationally fastest method for simultaneously calculating the Correlation Dimension and *SampEn* as well as *ApEn* of a time series. It has been described in details in [7]. Briefly, this algorithm is based on building a look-up table using the following matrix

$$n_{i,j} = \|u_i - u_{i+(j+1)\cdot t}\|$$

where u is the component of the time series and t is time lag. For a signal consisting of N points the algorithm lookup table dimension is (N − τ − 1) × [(N − 1)/τ − 1]. The NCM algorithm uses the symmetry of this matrix to reduce the number of operations by half as well as removing redundancy from the calculations by noticing the cumulative character of the expression in equation (2). The next component of the algorithm, which is used to get the dependence of $C^m(r)$ on r, is replacing the looping operation over r between r_{min} and r_{max}, with an arithmetic calculation of the comparison of the norms (eq. 4) with the threshold within the Heaviside function (eq. 3). Apart from speeding up the calculations, this approach allows a very dense r sampling. All these steps are exact and no approximations are made. A full description of the algorithm together with the complete sources of a CPython implementation may be found in [7].

2.2. Entropy Surface

Data Collection. A 24-h ECG Holter recordings have been collected in 79 healthy volunteers (41 men, age 35 ±7.4 years, mean ±SD). Participants underwent careful history taking, physical examination and resting 12-lead ECG. They were all in sinus rhythm, between 55 and 90 beats/min. No one reported any chronic or acute disease within the last 3 months. No one was taking any medications (including oral contraceptives by female volunteers within the last 2 weeks) or was into endurance training.

Data Analysis. For all recordings sample entropy was calculated by the NCM algorithm across m values spanning from 1 to 20 and r from 0 to 5, thus producing entropy surfaces. The ridge position in the *(r,m)* plane was established by calculating the *SampEn(r)* maximum along the m cross-sections. Calculations were repeated after data shuffling to random order the RR time series with the use of a random number generator. All calculations were performed with in-house software written in Python (Python Foundation, Wolfeboro Falls, USA) and R (The R Foundation for Statistical Computing, Vienna, Austria).

3. Results

An example of the *SampEn* surface and the m cross-sections for a recording in the physiological order is presented in figure 1 and 2 and the first panel of figure 4. Figure 3 and the second panel of figure 4 are prepared for the same recordings after shuffling. What is characteristic is the ridge-ledge formation in the data in the physiological order and only a pronounced ridge in the shuffle data.

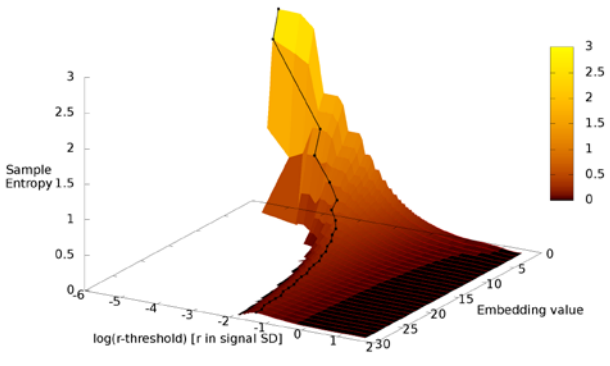

Figure 1. Entropy surface with entropy ridge marked as black solid line for 24-h RR series in one volunteer.

The "*ridge-ledge*" in *SampEn* surface is a clear ridge formed by the maximum values of *SampEn* along the *m* cross-section and either another, lower ridge or a ledge to the right of the main ridge. In fact, most of the 24-h recordings exhibit the same behavior: a ridge-ledge structure which disappears after shuffling, forming a single, much higher ridge.

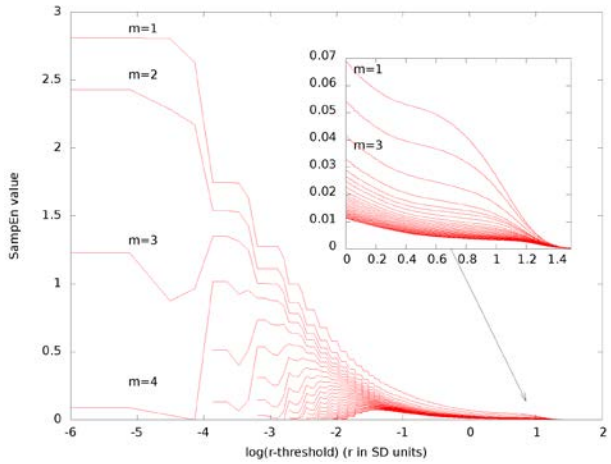

Figure 2. Cross-section of entropy surface calculated for 24-h RR series in one volunteer. A ledge at higher *r* values is visualized in inset. Each line corresponds to a different embedding *m*, from *m*=1 to *m*=30.

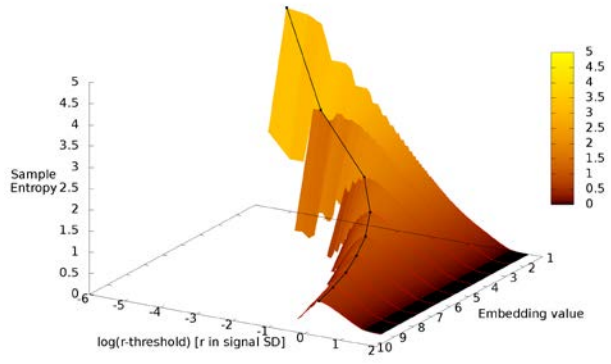

Figure 3. Entropy surface for the same data as in figure 1, after shuffling. *SampEn* values are much higher when compared to the values of the original 24h RR signal. Also, the ridge is right-shifted.

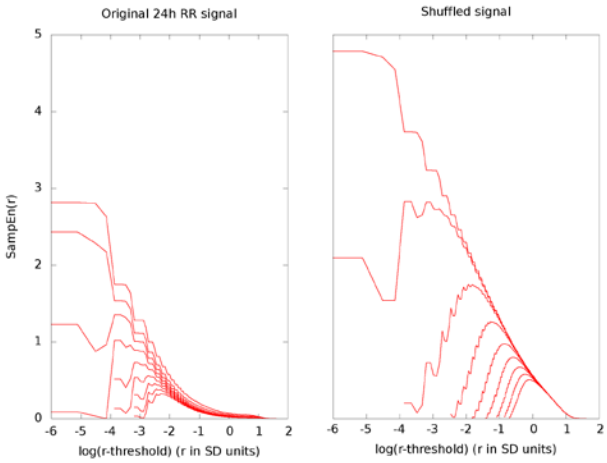

Figure 4. Cross-sections of the *SampEn* surfaces of Fig. 1 (*left*) and Fig 3 (*right*). In both panels a clear maximum is detectable in the series of consecutive values of embedding. The physiological time series *(left)* also exhibits the ridge-ledge structure. The ridge for the shuffled data (*right*) is much higher and is not accompanied by a ledge.

4. Discussion

In this paper we have introduced an application of the new NCM algorithm to the study of sample entropy surfaces. Our analysis reveals that RR time series from 24-h recordings have a distinctive ridge-ledge structure. After shuffling the data to a random order, only a single, pronounced ridge is visible. We may hypothesize that the main ridge expresses the degree of randomness. It is higher in random data than in physiological time series, where an autocorrelation structure lowers the ridge. However, our results also indicate that physiological mechanisms introduce a more complex structure in the form of a ledge, which may result from overlapping ridges.

The shape of the main ridge, its height in relation to random data and the presence of an accompanying ledge call for more comprehensive studies, both theoretical and data-analytic.

Acknowledgements

The first author (SŻ) is a scholar within Sub-measure 8.2.2 Regional Innovation Strategies, Measure 8.2 Transfer of Knowledge, Priority VIII Regional Human Resources for the Economy, Human Capital Operational Programme co-financed by the European Social Fund and state budget.

References

[1] Pincus SM, Goldberger AL. Physiological time-series analysis: what does regularity quantify? Am J Physiol Heart Circ Physiol 1994;266:H1643-56.

[2] Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol 2000;278:H2039–49.

[3] Pincus SM, Huang WM. Approximate entropy : statistical properties and application. Commun Statist-Theory Meth 1992;21:3061-3077.

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

[5] Castiglioni P, Di Rienzo M. How the threshold "r" influences approximate entropy analysis of heart-rate variability. Computers in Cardiology 2008;35:561-4.

[6] Lu S, Chen X, Kanters JK, Solomon IC, Chon KH Automatic selection of the threshold value R for approximate entropy. IEEE Trans Biomed Eng. 2008;55:1966-72.

[7] Żurek S, Guzik P, Pawlak S, Kośmider M, Piskorski J. On the relation between correlation dimension, approximate entropy and sample entropy parameters, and a fast algorithm for their calculation, Physica A, 2012; doi:10.1016/j.physa.2012.07.003 (*in press*).

Address for correspondence.

Sebastian Żurek
Institute of Physics, University of Zielona Góra
ul. Prof. Szafrana 4a, 65-516 Zielona Góra
S.Zurek@if.uz.zgora.pl