Applied Soft Computing 11 (2011) 2871–2879

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Complexity analysis of the biomedical signal using fuzzy entropy measurement

Hong-Bo Xie [a,*], Wei-Ting Chen [b], Wei-Xing He [a], Hui Liu [a]

[a] School of Electronic and Information Engineering, Jiangsu University, Xuefu Rd 301#, Zhenjiang, 212013, PR China
[b] Software Engineering Institute, East China Normal University, Shanghai, PR China

ARTICLE INFO

Article history:
Received 16 August 2009
Received in revised form 5 August 2010
Accepted 28 November 2010
Available online 4 December 2010

Keywords:
Fuzzy entropy
Sample entropy
Fuzzy membership function
Complexity
Time series
Biomedical signals

ABSTRACT

Exploiting the concept of fuzzy sets, a new time series complexity measure named fuzzy entropy was developed. In fuzzy entropy, the degree of similarity between vectors is defined on the basis of a fuzzy membership function and according to the shapes of the fuzzy membership function rather than by employing the conventional Heaviside function used in approximate entropy and sample entropy. Tests conducted on independent identically distributed uniform random numbers, mixture stochastic processes, the Rossler attractor, the Henon map, and sinusoidal signals showed that fuzzy entropy is superior to sample entropy in several respects, providing entropy definition in the case of small parameters, improving the degree of monotonicity, and being more robust to noise. The results of tests on a biomedical time series of electromyography signals illustrate the applicability of the proposed method.

1. Introduction

Complexity has moved from being a vague, non-scientific term referring to things we do not understand very well or at all to a more substantive technical term used in mathematics, computation, physics, chemistry, and biology [5]. Over recent decades, a significant debate has developed regarding the definition and use of the concept of complexity [19]. The concept of complexity in this paper is inspired to some degree by interest in deterministic chaos [19]. In qualitative terms, this concept refers to diversity of forms, to the emergence of coherent patterns out of randomness, and to a frequent switching ability among such patterns [3]. Complexity is inherently linked to several other concepts in chaos theory including entropy, randomness, and information theory. Analysis of time series data has suggested a direct link between chaos theory and the real world. Many measures have been developed over recent years to estimate the complexity of biomedical time series and provide new insights into physiological systems, such as Kolmogorov–Sinai (K–S) entropy [6], the Lyapunov exponent [15], Lempel–Ziv complexity [16], permutation entropy [1], the recurrence rate derived from recurrence quantification analysis [35], the correlation dimension [8], and the symplectic geometry spectrum [33]. However, it is widely known that biomedical signals are often short in data-length and easily contaminated by noises and most complexity measures usually require very long data sets to calculate reliable and convergent values [30], which may lead to spurious results when applied to short or irregular sequences of real experimental data [14].

To solve the problems of short data sets and noisy recordings in biomedical signal analysis, Pincus proposed a family of statistics called approximate entropy (ApEn) [20–24] aimed at measuring signal complexity, i.e., the presence of similar patterns in a time series. Given N points and tolerance r, ApEn(m,r,N) is approximately equal to the negative average natural logarithm of the conditional probability that two sequences that are similar for m points within the tolerance remain similar at the next point. As a statistic superior to most complexity measures [9,23,31], ApEn has been shown to have potential application to a wide range of biomedical signals such as hormone pulsatility, genetic sequences, respiratory patterns, heart rate variability, electrocardiograms, and electroencephalography. Nevertheless, ApEn always suggests more similarity than is present due to the counting of self-matches and is therefore biased.

To be free of the bias caused by self-matching, Richman and Moorman developed another related measure of time series complexity named sample entropy (SampEn) [25]. SampEn is largely independent of record length and displays relative consistency under circumstances where ApEn does not. For example, Richman and Moorman tested ApEn and SampEn statistics on independent identically distributed (i.i.d.) uniform random numbers consisting of different points and compared the results with the analytically calculated expected values. SampEn(2,r,N) very closely matches the

* Corresponding author. Tel.: +86 15951288938.
E-mail address: xiehb@sjtu.org (H.-B. Xie).

1568-4946/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2010.11.020

expected results for $r \geq 0.03$ and $N \geq 100$, whereas ApEn(2,r,N) differs markedly from expectations for $r < 0.2$ and $N < 1000$ [25]. The relative consistency of this measure means that if one set of records shows lower SampEn value than another with a set of parameters of a tolerance r and an embedding dimension m, it also has lower SampEn with other r and m parameters [25]. However, SampEn(m,r,N) is not defined in some cases for time series with small r and N values [25].

Fuzzy logic plays an important role in biomedical signal analysis fields such as pattern recognition, computer vision, machine learning, image analysis, communication, knowledge discovery, and data mining [2,7,10–12,28]. In the study reported in this paper, we formulated another related measure of time series complexity, i.e., fuzzy entropy (FuzzyEn), and applied it to characterize local muscle fatigue electromyography (EMG) signals. In the proposed measure, a fuzzy membership function is used to determine the degree of similarity between two vectors. To show the appropriateness and effectiveness of the proposed measure, FuzzyEn was compared with SampEn in several distinct settings including i.i.d. uniform random numbers, the mixture stochastic processes model (MIX), the Rossler attractor, the Henon map, and sinusoidal signals. FuzzyEn was then used to monitor the complexity decrement of surface EMG signals to identify local muscle fatigue in humans.

2. Methodology

2.1. Sample entropy

For an N-point normalized time series $\{u(i): 1 \leq i \leq N\}$ with mean zero and unit standard deviation, the following vector sequence can be formed [25]

$$X_i^m = \{u(i), u(i+1), \ldots, u(i+m-1)\} \quad 1 \leq i \leq N-m+1 \tag{1}$$

where X_i^m represents m consecutive u values commencing with the ith point and m is called the embedding dimension. The distance d_{ij}^m between X_i^m and X_j^m is defined as

$$d_{ij}^m = d[X_i^m, X_j^m] = \max_{k\in(0,m-1)} |u(i+k) - u(j+k)| \tag{2}$$

The average degree of similarity between X_i^m and its neighboring vectors X_j^m within tolerance r can then be defined as

$$B_r^m(i) = \frac{1}{N-m-1} \cdot \sum_{j=1,\, j\neq i}^{N-m} \Theta(d_{ij}^m - r) \tag{3}$$

where j ranges from 1 to $(N-m)$ and $j \neq i$ to exclude self-matches. The vector sequence X_i^m used for comparison is called the template.

In Eq. (3), Θ is the Heaviside function

$$\Theta(z) = \begin{cases} 1, & \text{if} \quad z \leq 0 \\ 0 & \text{if} \quad z > 0 \end{cases} \tag{4}$$

If the distance d_{ij}^m between X_i^m and X_j^m is not larger than the tolerance r, then a template match occurs. A specific case is a template self-match in which X_i^m is compared with itself and d_{ij}^m is equal to zero. As noted earlier, template self-matches have been excluded in the definition of SampEn [25].

The probability B_r^m that two vector sequences will match for m points is given by

$$B_r^m = \frac{1}{N-m} \cdot \sum_{i=1}^{N-m} B_r^m(i) \tag{5}$$

Similarly, an $(m+1)$-dimensional sequence $X_i^{m+1} = \{u(i), u(i+1), \ldots, u(i+m)\}$ can be formed and the average degree of similarity $A_r^m(i)$ can be defined in exactly the same way as $B_r^m(i)$ using X_i^{m+1}. The template match between $(m+1)$-dimensional vector sequences is called the template forward match. The probability A_r^m that two $(m+1)$-dimensional sequences will match is obtained in a similar way. The estimate SampEn of the time series (obtained in the limit of $N \to \infty$) is defined by [25]

$$\mathrm{SampEn}(m, r, N) = -\ln\left(\frac{A_r^m}{B_r^m}\right) \tag{6}$$

In the definition of SampEn, the similarity of vectors is based on the Heaviside function shown in Equation (4). The main feature of this Heaviside function is its provision of a step function that converts the value of the input into the activity to 0 or 1. This function is discontinuous because there is a “break” in it when its value goes from 0 to 1. It leads to a conventional kind of two-state classifier that judges an input pattern according to whether it satisfies certain precise properties required for membership of a given class. The contributions of all the data points inside the boundary are treated equally, while the data points just outside the boundary are ignored. The distance d_{ij}^m that is just greater than the tolerance r is not considered in B_r^m or A_r^m, and those that are less than the tolerance are treated equally. As a result, SampEn may not be discontinuous and may rise or fall dramatically when the tolerance r is slightly changed.

2.2. Fuzzy entropy

In the physical world, boundaries between classes may be ambiguous, and it is difficult to determine whether an input pattern belongs completely to a given class. The concept of “fuzzy sets” introduced by Zadeh in 1965 provides a means of characterizing such input–output relations in an environment of imprecision [34]. By introducing the concept of “membership degree” that has a fuzzy function $u_C(x)$ and associates each point x with a real number in the range of [0, 1], Zadeh's theory provided a mechanism for measuring the degree to which a pattern belongs to a given class: the nearer the value of $u_C(x)$ to unity, the higher the membership grade of x in the set $\boldsymbol{C}$. In FuzzyEn, we employed the fuzzy membership function $u(d_{ij}^m, r)$ to obtain a fuzzy measurement of the similarity between X_i^m and X_j^m based on their shapes.

According to the new similarity index discussed above, fuzzy entropy can be defined as follows. For an N-point normalized time series $\{u(i): 1 \leq i \leq N\}$, similar to the definition of SampEn, the form vector sequence is

$$\begin{aligned} X_i^m &= \{u(i), u(i+1), \ldots, u(i+m-1)\} - u0(i) \\ &\quad \times (i = 1, \ldots, N-m+1) \end{aligned} \tag{7}$$

However, X_i^m was generalized here by removing a baseline

$$u0(i) = \frac{1}{m}\sum_{j=0}^{m-1} u(i+j) \tag{8}$$

The distance d_{ij}^m between the vector X_i^m and its neighbor X_j^m is then defined as

$$\begin{aligned} d_{ij}^m &= d\left[X_i^m, X_j^m\right] = \max_{k\in(0,m-1)} |u(i+k) - u0(i) - (u(j+k) - u0(j))| \\ &\quad \times (i, j = 1 \sim N-m, j \neq i) \end{aligned} \tag{9}$$

The degree of similarity D_{ij}^m between X_i^m and X_j^m is determined by a fuzzy membership function

$$D_{ij}^m = u(d_{ij}^m, r) \tag{10}$$

Similar to the definition of SampEn, for each vector X_i^m, averaging all the degrees of similarity to its neighboring vectors X_j^m, gives an average degree of similarity

$$B_r^m(i) = \frac{1}{N-m-1}\sum_{j=1,j\neq i}^{N-m} D_{ij}^m \tag{11}$$

and the probability that two vector sequences will match is given by

$$B_r^m = \frac{1}{N-m}\sum_{i=1}^{N-m} B_r^m(i) \tag{12}$$

Similarly, the vector sequence $\{X_i^{m+1}\}$ can be formed and the average degree of similarity $A_r^m(i)$ and probability A_r^m can be defined as

$$A_r^m(i) = \frac{1}{N-m-1}\sum_{j=1,j\neq i}^{N-m} D_{ij}^{m+1} \tag{13}$$

$$A_r^m = \frac{1}{N-m}\sum_{i=1}^{N-m} A_r^m(i) \tag{14}$$

The FuzzyEn(m,r) estimation of a time series is defined as the negative natural logarithm of the deviation of B_r^m from A_r^m

$$\mathrm{FuzzyEn}(m, r) = \lim_{N\to\infty}(\ln B_r^m - \ln A_r^m) \tag{15}$$

For finite data sets, the following statistic measures fuzzy entropy

$$\mathrm{FuzzyEn}(m, r, N) = -\ln\left(\frac{A_r^m}{B_r^m}\right) \tag{16}$$

In practice, a Gaussian function, a Sigmoid function, a bell-shaped function, or any other fuzzy membership function can be chosen to describe the similarities between the two vectors as long as they uphold the following two properties: (1) continuity, so that similarity does not change abruptly; and (2) convexity, so that self-similarity is maximized. In the present study, the following Gaussian function was employed as the fuzzy membership function in calculating FuzzyEn.

$$u(d_{ij}^m, r) = \exp\left(\frac{-d_{ij}^2}{r}\right) \tag{17}$$

2.3. Simulated signals

The ability of the proposed method to capture different degrees of complexity was tested in several types of simulations to reproduce both stochastic and deterministic short time series. The simulated signals of i.i.d. uniform random numbers and MIX process were used to demonstrate the sensitivity of FuzzyEn to tolerance r, while the Rossler equation, the Henon map, and a sinusoidal series model were used to test the monotonicity of FuzzyEn when the complexity strength R of the models varied. Informally, the MIX(P) time series of N points, where P is between 0 and 1, is a sine wave where $N \times P$ randomly chosen points have been replaced by random noise [20,25].

The Rossler equation [27] is given by

$$\frac{dx}{dt} = -z - y \tag{18}$$

$$\frac{dy}{dt} = x + 0.15y \tag{19}$$

$$\frac{dz}{dt} = 0.20 + R(zx - 5.0) \tag{20}$$

Time series were obtained for complexity strength R ranging from 0.7 to 0.9 in steps of 0.01 by integration via an explicit time-step method with an increment of 0.005. The y values were recorded at intervals of $\Delta t = 2$. With monotonously increasing complexity strength R, the respective systems are given from a twice-periodic to chaotic limit cycle dynamics [20].

A parameterized version of the Henon map [26] is given by

$$x_{i+1} = Ry_i + 1 - 1.4x_i^2 \tag{21}$$

$$y_{i+1} = 0.3Rx_i \tag{22}$$

Time series for x_i were obtained for R ranging from 0.8 to 1.0 in steps of 0.01. With monotonously increasing complexity strength R, the respective systems correspond to deterministic chaos dynamics with more complex attractor [22].

The sinusoidal signals were generated with frequency varying from 5 Hz to 205 Hz in steps of 10 Hz.

$$x = \sin(2\pi(5 + 10R)t) \tag{23}$$

For the Rossler and Henon systems, time series were generated after a transient period of 5000 points. We added to each of these time series x_i i.i.d. samples of Gaussian noise η_i with different noise levels (NL) such that the resulting time series f_i is given by

$$f_i = x_i + \eta_i \tag{24}$$

Because the original time series x_i was standardized (centred and normalized by its standard deviation), the level of the added noise was quantified by the standard deviation of the noise time series η_i.

ApEn, SampEn, and the FuzzyEn measure proposed in this paper are all *pattern complexity* measures that quantify the regularity embedded in a time series [14,25,29]. In other words, they all estimate the probability that initially similar sequences in a dataset remain similar, within a given tolerance, in the next incremental comparison. A low entropy value indicates the time series is deterministic (low in complexity), whereas a high value indicates the data are subject to randomness (high in complexity) and are therefore difficult to predict. In other words, lower entropy values indicate more regular time series, whereas higher entropy values indicate more irregular time series [25,29].

In previous studies, the criterion used for evaluating the performance of ApEn and SampEn was often based on relative consistency which does not include a statistical analysis [20,25]. For example, Richman and Moorman compared the relative consistency of MIX(0.1) and MIX(0.9) processes using ApEn and SampEn, respectively. Graphically, plots of ApEn as a function of r for the two data sets intersected with each other, while plots of SampEn did not. According to the visual inspection, they then concluded that SampEn is a better measure than ApEn in comparing MIX(P) processes [25]. To compare these measures in terms of their ability to distinguish between different degrees of complexity, we designed an indicator that would yield higher value when the complexity measure showed better monotonicity, i.e., higher complexity value were obtained for signal with higher complexity strength. For every model, each measure of complexity c was estimated at monotonously increasing levels of complexity strength. We assumed that the total number of values of complexity strength considered was s, resulting in the statistics values c_i, $i = 1, 2, \ldots, s$. If c depends monotonically on complexity strength, then $c_i \le c_j$ for $i \le j$. To detect deviations from this rule, we defined the degree of

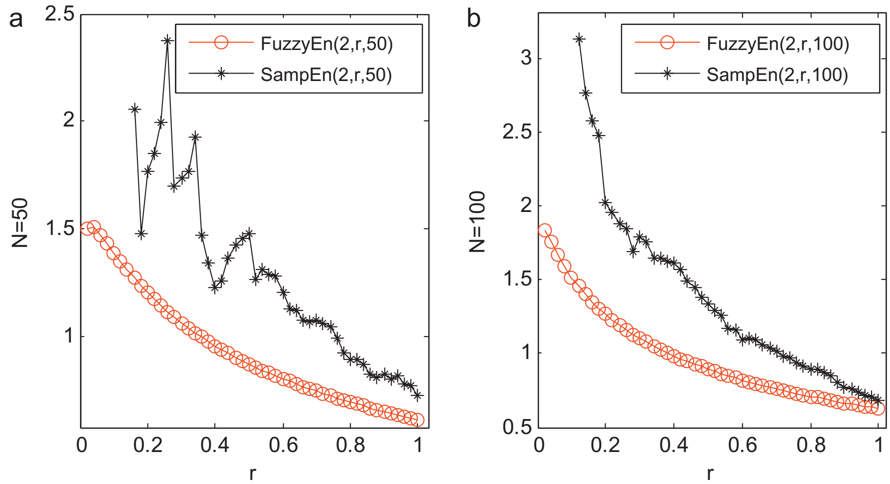

Fig. 1. FuzzyEn(2,r,N) and SampEn(2,r,N) as functions of r for i.i.d. uniform random numbers as r varies from 0.01 to 1.0 in steps of 0.01.

monotonicity (DoM) of the complexity measure as follows

$$\mathrm{DoM}(c) = \frac{2}{s(s-1)} \sum_{i=1}^{s-1} \sum_{j=i+1}^{s} \mathrm{sign}(c_j - c_i) \quad (25)$$

DoM(c) = 1 for a strictly monotonically increasing sequence (c_1, . . ., c_r) and is equal to −1 for a monotonically decreasing sequence. On the other hand, DoM(c) is not affected by the absolute values of c_i or by the way in which (c_1, . . ., c_r) increases (e.g., polynomial in i or exponential).

Richman and Moorman have shown that the performance of SampEn is superior to that of ApEn when they are applied to simulated signals [25]. Thus, we do not report the ApEn results for the simulated signal comparison. Pincus suggested to set the embedding dimension m to 2 or 3 and tolerance r to 0.1–0.3 in ApEn estimation [20]. Too large value of m is unfavorable due to the need of a very large N, which is hard to meet for a biomedical data set. Low value of tolerance r may result in salient influence from noise, while too large value of r should also be avoided for fear of information loss. In this study, we adopted Pincus' suggestions on the parameters m and r setting. The embedding dimension m was fixed at 2 in all the statistics estimations and the tolerance r was fixed at 0.1 in the noise performance comparison. Unless otherwise stated, 1000 data points were considered in all simulated examples. Both FuzzyEn and SampEn were calculated for the time series generated by the models described before, with increasing complexity strength and different noise levels in all the calculations. In the simulated experiments, complexity strength R for the Rossler equation varied from 0.7 to 0.9 in steps of 0.01, while complexity strength R for the Henon map varied from 0.8 to 1.0 in steps of 0.01. Complexity strength R for the sinusoidal signals varied from 5 to 205 in steps of 10. Thus, the total number of values of complexity strength (s) considered was 21 for each of the three simulation models. The s variable was set at 21 according to the recommendation of Janjarasjitt and Loparo [13]. We used levels of noise ranging from 5% to 50% with 5% steps for the quantitative analysis. The results were quantified by means of DoM, which offers an overall evaluation of the complexity measure to distinguish different degrees of complexity.

2.4. Real EMG signals

To illustrate the applicability of the proposed FuzzyEn measure, its performance in characterizing localized muscle fatigue using EMG signals was compared with that of ApEn and SampEn. EMG signals are electrical manifestations of the neuromuscular activation associated with muscle contraction [4,18,32]. Localized muscle fatigue is the term used to describe the situation in which a muscle has a reduced ability to produce force during sustained contractions, and is a phenomenon reflected in EMG signals through a decrease in power spectrum bandwidth [4,18,32]. A relationship has been demonstrated between the ApEn measure and the bandwidth of a signal's power spectrum, with larger ApEn values corresponding to broad band spectra and smaller ApEn values corresponding to peak spectra [22]. Thus, EMG signals of lower complexity are a characteristic feature of localized muscle fatigue.

The EMG signals analyzed in this paper were recorded during voluntary isometric contractions among 12 healthy human subjects. When the experiment began, the subject was asked to perform an elbow flexion against a fixed lever arm to 80% of their maximal voluntary contraction (MVC) and to maintain this value while visual feedback was provided via a torque reading on a computer screen. The test was stopped when the torque dropped to approximately 70% of the MVC, thus indicating that the muscle was fatigued. The gain of the EMG signal was 2000 with a bandwidth of 10–400 Hz. Signals were sampled at 1 kHz and stored by computer for further FuzzyEn and SampEn analysis.

An index of regression signal noise ration (RSNR), also termed SNR by the authors who first used it in muscle fatigue assessment [17], was applied to assess the variability of the complexity statistics in real EMG signal analysis. RSNR has been shown to be effective in evaluation of a muscle fatigue index and is defined as [17]:

$$\mathrm{RSNR} = \frac{\mathrm{RA}}{\sqrt{\frac{1}{F}\sum_{f=1}^{F}(I_f - \hat{I}_f)^2}} \quad (26)$$

where F represents the number of segments. The numerator 'RA' in this metric represents the range of the signal of interest and is calculated by taking the difference between the maximum and minimum of the best line of fit $\hat{I}$ for the outputs 'I'. The denominator represents the root mean square error of the outputs with respect to the best line of fit.

3. Results

3.1. Simulation results

The FuzzyEn and SampEn statistics were firstly tested on i.i.d. uniform random numbers. Fig. 1 shows the performance of FuzzyEn(2,r,N) and SampEn(2,r,N) on i.i.d. random numbers with

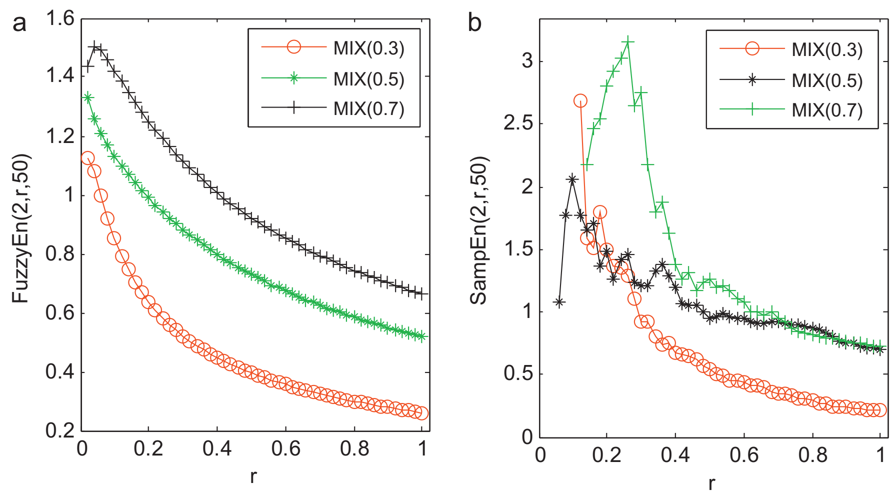

Fig. 2. The performance of FuzzyEn(2,r,N) and SampEn(2,r,N) statistics that measure the complexity of MIX(0.3), MIX(0.5), and MIX(0.7) for N = 50.

different lengths of N = 50 and N = 100. The tolerance r varied from 0.01 to 1.0 in steps of 0.01. As shown in Fig. 1(a) and (b), SampEn gave no entropy values when the tolerance r was smaller than about 0.15. Therefore, the problem of parameter limitation arose in calculating SampEn for the short data set. However, this problem did not affect the calculation of FuzzyEn, even for rather small r values (0.01 or below) when the fuzzy membership function was used to determine the degree of similarity. From Fig. 1, it can be seen that FuzzyEn changed continuously and smoothly. However, SampEn fluctuated significantly and exhibited discontinuity with small changes in the parameter r.

The FuzzyEn and SampEn statistics were then tested on the MIX(0.3), MIX(0.5), and MIX(0.7) processes. Fig. 2 shows the performance of FuzzyEn(2,r,N) and SampEn(2,r,N) on the three sequences with a length of N = 50. The tolerance r varied from 0.01 to 1.0 in steps of 0.01. As with the uniform random signal simulation, SampEn gave no value when r was smaller than about 0.15 for MIX(0.5) and MIX(0.7) or when r was smaller than about 0.08 for MIX(0.3). But this problem did not affect the calculation of FuzzyEn for rather small r values. The results for the length of N = 100 were similar to those for the length of N = 50 and are therefore not reported. Also, as in the uniform random signal simulation, FuzzyEn changed continuously and smoothly in MIX(P) stochastic processes, whereas SampEn still fluctuated significantly and exhibited discontinuity with small changes in the parameter r, a result that may degrade its value for practical applications.

The degree of monotonicity was assessed using the Rossler attractor, the Henon map, and sinusoidal signals at different frequencies because the degree of complexity in these systems could be controlled by the complexity strength parameter R. More complex Rossler systems and sinusoidal signals produce larger entropy values. This was validated by using FuzzyEn when no noise was added to the data set. This was also true for Henon systems in most cases, as indicated by the DoM of FuzzyEn on Rossler, Henon, and sinusoidal series being 1, 0.966, and 1, respectively. However, the DoM of SampEn on Rossler, Henon, and sinusoidal series was lower than that of FuzzyEn, with values of 0.834, 0.81, and 0.887, respectively, being observed, even when no noise was added to the data. The results demonstrate that FuzzyEn is more accurate in distinguishing Rossler, Henon, and sinusoidal series from each other with a different control parameter R.

We found that when noise was superimposed, it became difficult for both FuzzyEn and SampEn to distinguish system complexity. The system complexity measured may be reversed at some complexity strengths; our observations show that this effect became even more acute when the noise level was high. Figs. 3–5 show the dependence of FuzzyEn and SampEn on the complexity strength R of Rossler, Henon, and sinusoidal series, respectively,

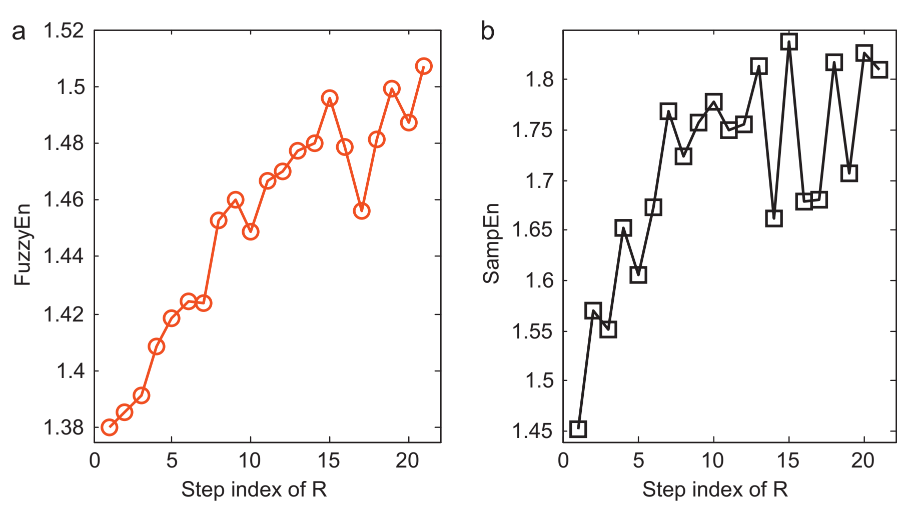

Fig. 3. The performance of FuzzyEn and SampEn statistics that distinguish between Rossler systems with different degrees of complexity. Additional noise with NL 0.2 is added. The parameters used for calculation are m = 2, r = 0.1, and N = 1000.

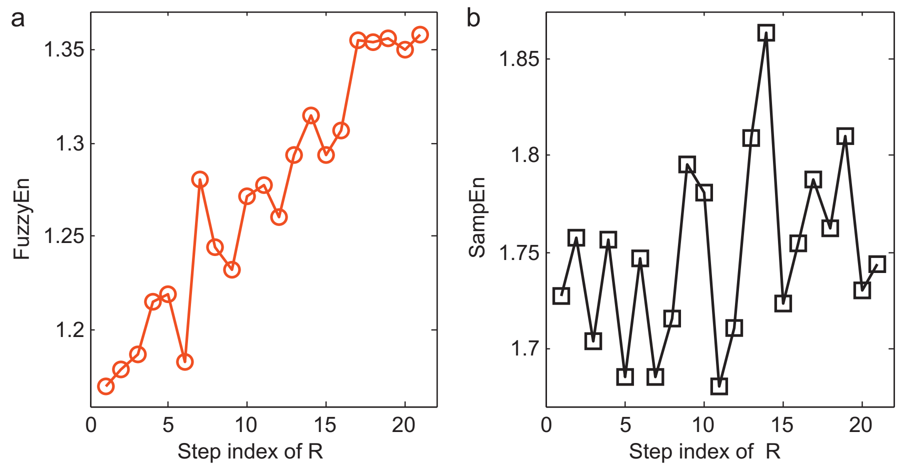

Fig. 4. The performance of FuzzyEn and SampEn statistics that distinguish between Henon systems with different degrees of complexity. Additional noise with NL 0.2 is added. The parameters used for calculation are $m = 2$, $r = 0.1$, and $N = 1000$.

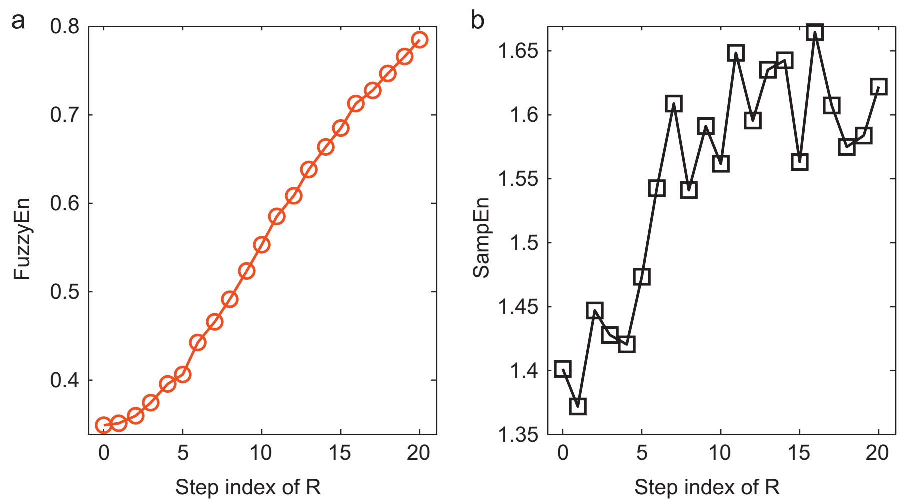

Fig. 5. The performance of FuzzyEn and SampEn statistics that distinguish between sinusoidal signals with different degrees of complexity. Additional noise with NL 0.2 is added. The parameters used for calculation are $m = 2$, $r = 0.1$, and $N = 1000$.

when noise level was 0.2. The figures demonstrate that SampEn fluctuated significantly at large R values for Rossler series, across the whole range of R values for Henon series, and at large R values for sinusoidal series, respectively. On the other hand, FuzzyEn increased rather steeply with a low degree of fluctuation when applied to each individual setting. To compare the two measures quantitatively, the degree of monotonicity was again used to enable us to track how the monotonicity obtained for noise-free systems deteriorated with an increasing noise level. Figs. 6–8 show the plots of DoM versus the noise level for the Rossler, Henon, and sinusoidal series, respectively. The DoMs of both FuzzyEn and SampEn decreased with an increasing noise level. However, the DoM of FuzzyEn was higher than that of SampEn for each system at each noise level. The results demonstrate that FuzzyEn is more robust than SampEn in series contaminated by noise. This is another important aspect to consider in choosing a suitable measure for real field data applications.

3.2. Experimental results

To monitor the changes in FuzzyEn, SampEn, and ApEn for EMG over time, the acquired EMG signal was first segmented into consecutive, non-overlapping epochs of 500 ms in length. Fig. 9 shows the typical raw EMG signals of the first and the last epochs acquired from subject 3. For each epoch, the EMG signal was normalized and the procedures described previously were followed in calculating FuzzyEn, SampEn, and ApEn. Fig. 10 shows the time courses of the three measures for subject 3. The least-squares error linear regression was fitted to each set of FuzzyEn, SampEn, and ApEn observations over the period of contraction to obtain their slope and intercept. The decreases in sequential FuzzyEn and SampEn

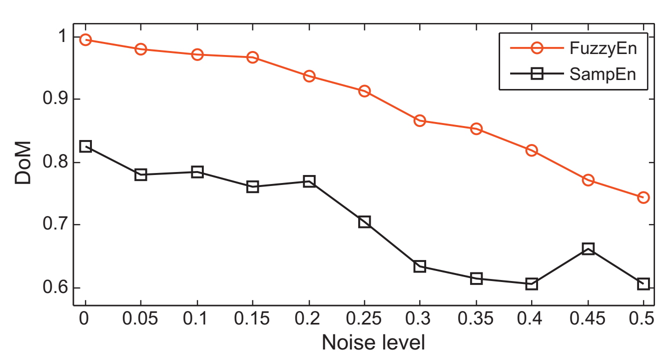

Fig. 6. Rossler systems: dependence of the degree of monotonicity of FuzzyEn and SampEn on the noise level. This figure also plots the case of noise-free systems below the smallest noise level of 0.05.

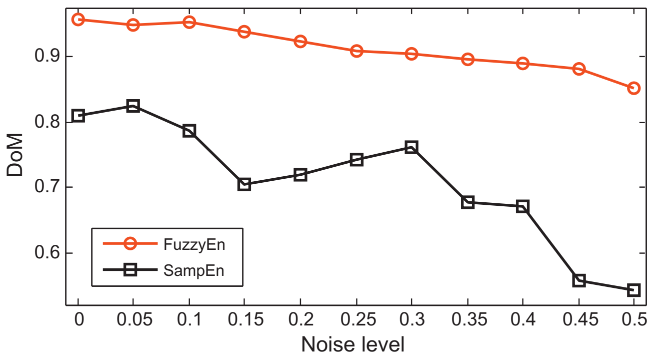

Fig. 7. Henon systems: dependence of the degree of monotonicity of FuzzyEn and SampEn on the noise level. This figure also plots the case of noise-free systems below the smallest noise level of 0.05.

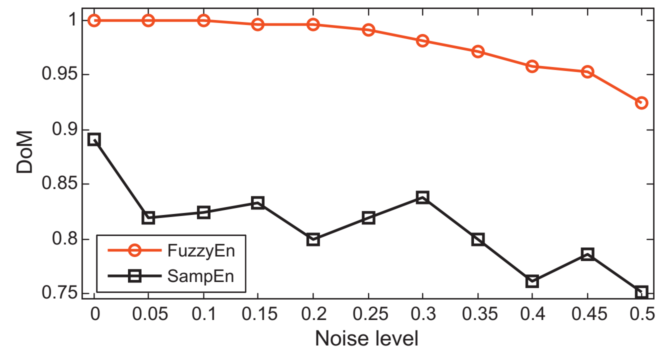

Fig. 8. Sinusoidal signals: dependence of the degree of monotonicity of FuzzyEn and SampEn on the noise level. This figure also plots the case of noise-free systems below the smallest noise level of 0.05.

estimates signified a downward trend that was confirmed by the statistically significant negative slope (ANOVA, $\alpha = 0.05$, $p < 0.001$). Only two different ApEn values were observed for different EMG epochs, indicating that this measure failed to detect changes in the complexity of EMG signals recorded during muscle fatigue. The results for all the other subjects were similar to the results shown in Fig. 10, indicating that FuzzyEn or SampEn might be suitable as indicators of the complexity of EMG signals in assessing muscle fatigue. Fig. 10 also indicates that FuzzyEn varied less than SampEn.

The RSNR metric was applied to both FuzzyEn and SampEn values estimated for each segment to gauge the relative performance of the two measures. Table 1 reports the RSNR values obtained from the test data for all subjects for both FuzzyEn and SampEn. Table 1 shows that the RSNR values of FuzzyEn regression lines are larger

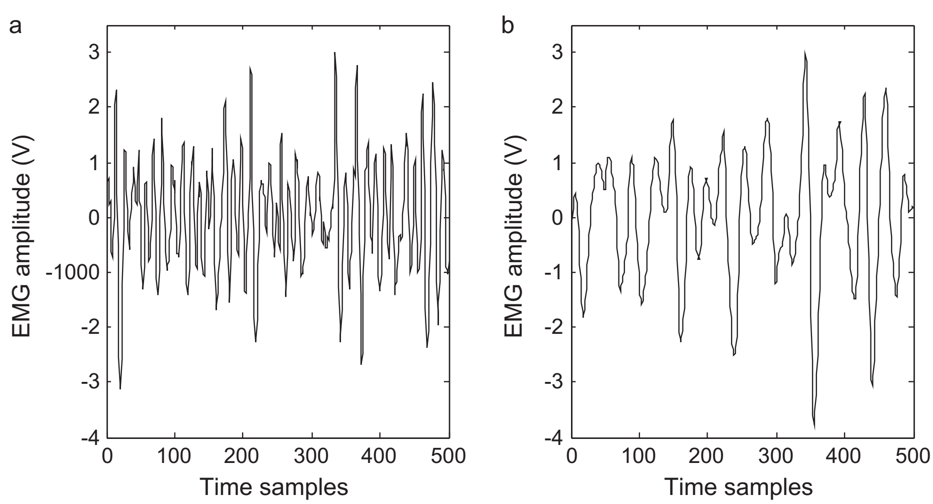

Fig. 9. The EMG signals recorded at the (a) beginning, (b) end of the static isometric contraction until exhaustion from the biceps for subject 3.

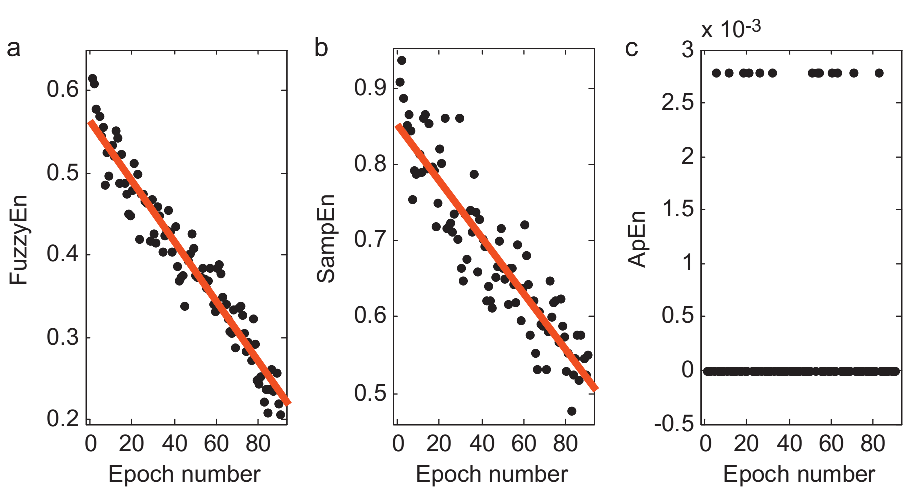

Fig. 10. Time courses of FuzzyEn (a), SampEn (b), and ApEn (c) for the EMG signals of subject 3 shown in Fig. 9. The analysis window was 500 ms. The linear regressions of FuzzyEn (a) and SampEn (b) are superimposed on the plot.

Table 1
The RSNRs of the linear regressions of FuzzyEn and SampEn for each subject.

Participant	RSNR of FuzzyEn	RSNR of SampEn
1	8.850	4.271
2	10.987	7.341
3	12.446	7.401
4	13.255	6.949
5	13.460	5.969
6	11.083	5.928
7	9.729	5.520
8	10.585	6.7221
9	9.838	5.658
10	8.834	6.594
11	8.817	5.648
12	10.517	6.698
Mean ± Std.	10.700 ± 1.640	6.225 ± 0.898

than that of SampEn among all subjects. A one-way ANOVA test indicated a significant difference between FuzzyEn and SampEn (ANOVA, $\alpha = 0.05$, $p < 0.0001$), showing that FuzzyEn is capable of tracking fatigue trends in EMG data and outperformed SampEn in all subjects with a large RSNR.

4. Discussion and conclusions

In this paper, we propose a new time series complexity measure, namely fuzzy entropy. Like ApEn and SampEn, FuzzyEn is the negative natural logarithm of the conditional probability that a data set of length N, having repeated itself for m points within a boundary r, will also repeat itself for $m+1$ points. The same modifications differentiating SampEn from ApEn are also adopted for FuzzyEn: (1) it excludes self-matches, i.e., vectors are not compared to themselves; (2) it considers only the first $(N-m)$ vectors of length m so that for $i \le N-m$, both X_i^m and X_i^{m+1} are defined. FuzzyEn and SampEn therefore have the same good properties that ApEn may lack under certain circumstances, such as greater independence from data-length and relative consistency [25].

Unlike ApEn and SampEn, in which the degree of similarity between two vectors is based on a Heaviside function, FuzzyEn employs a fuzzy membership function to bound the similarity of two vectors. In a Heaviside function, the boundary is rigid: the contributions of all the data points inside it are treated equally, whereas the data points just outside it are ignored. The hard boundary causes discontinuity, which may lead to abrupt changes in entropy values when the tolerance r changes slightly, and even to a failure to find a SampEn value if no template match can be found for a small tolerance r. In contrast, there is no rigid boundary in a fuzzy membership function. All data points, in which function membership is determined by point positions together with the parameter r, are considered to be members of the function. For a vector X_i^m, similarity is indicated by the fuzzy membership function value: the closer the neighboring vector X_j^m is, the more similar X_j^m is to X_i^m, and the degree of similarity between X_j^m and X_i^m is almost zero when X_j^m is distant from X_i^m. Under the rule given by the fuzzy membership function, FuzzyEn is continuous and will not change dramatically with a slight change in r.

We used complexity strength R as the control parameter in testing the extent to which FuzzyEn is able to distinguish between different degrees of complexity. This is essential in many field applications as the absolute value of complexity is rarely of interest, with attention instead focusing on the change in complexity between different states, times, or recording sites. Simulated results obtained using three different models demonstrated that FuzzyEn was superior to SampEn when differentiating between systems with different degrees of complexity. When Gaussian noise components were added to the noise-free data sets, FuzzyEn was more robust to the noise and exhibited a high degree of monotonicity. Experimental and real biomedical signals are always noisy, FuzzyEn should therefore be used to determine hidden complexity and should be more capable than other measures of handling biomedical data.

To illustrate the applicability of the proposed method, we analyzed changes in the complexity of EMG signals during sustained isometric contractions in human subjects using the three statistics. Prior studies have shown that significant spectrum compression occurs in EMG signals in the course of muscle fatigue [4,18]. ApEn, SampEn, and FuzzyEn are, in essence, event-counting statistics, where the events are instances of vectors similar to one another. As the spectra of fatigue EMG signals become more compressed, overall EMG signal patterns tend to become more similar. Accordingly, the number of template matches for m-dimension sequences increases. However, any further increase in the number of template matches can only lead to a smaller increase in the number of $(m+1)$-dimension sequence template forward matches. Thus, observed conditional probability decreases, i.e., the values of the three complexity measures are expected to decrease. Decreases were observed in both SampEn and FuzzyEn during the experiment. Moreover, the significantly large RSNR values observed for FuzzyEn validated the conclusions reached in the simulated tests and confirmed that FuzzyEn is more suitable for analyzing short and noisy time series. The results suggest that FuzzyEn would be a better choice than SampEn for estimating the slope of the regression line, a process used to quantify muscle fatigue. When ApEn was applied to the same fatigue EMG signals, most of the template matches and forward template matches in the different EMG epoch signals were self-matches, leading to the smallest difference between the probabilities B_r^m and A_r^m among the different EMG epochs. This may explain why ApEn could not detect changes in the complexity of the EMG signals.

In summary, based on the concept of fuzzy sets, fuzzy entropy is proposed to characterize the time series with different degrees of complexity. Both theoretical analysis and experimental tests showed that FuzzyEn is superior to SampEn and ApEn in evaluating complexity and can thus serve as a convenient and powerful tool for short noisy experimental time series.

Acknowledgements

We would like to thank the two anonymous reviewers for their constructive comments to improve the manuscript. This work is supported by the Jiangsu Natural Science Foundation (BK2009198), and Jiangsu University (07JDG40), PR China.

References

[1] C. Bandt, B. Pompe, Permutation entropy: a natural complexity measure for time series, Phys. Rev. Lett. 88 (2002) 174102.
[2] G. Castellano, C. Castiello, A.M. Fanelli, C. Mencar, Knowledge discovery by a neuro-fuzzy modeling framework, Fuzzy Sets Syst. 149 (1) (2005) 187–207.
[3] S. Drozdz, J. Kwapien, J. Speth, M. Wojcik, Identifying complexity by means of matrices, Physica A 314 (2002) 355–361.
[4] J. Duchene, F. Goubel, Surface electromyogram during voluntary contraction: processing tools and relation to physiological events, CRC Crit. Rev. Biomed. Eng. 21 (1993) 313–397.
[5] F. Eugene Yates, Complexity of a human being: changes with age, Neurobiol. Aging 23 (2002) 17–19.
[6] P. Faure, H. Korn, A new method to estimate the Kolmogorov entropy from recurrence plots: its application to neuronal signals, Physica D 122 (1998) 265–279.
[7] G.B. Ferrara, L. Delfino, F. Masulli, S. Rovetta, R. Sensi, A fuzzy approach to image analysis in HLA typing using oligonucleotide microarrays, Fuzzy Sets Syst. 152 (1) (2005) 37–48.
[8] P. Grassberger, I. Procaccia, Measuring the strangeness of strange attractors, Physica D 9 (1983) 189–208.
[9] U. Grouven, F.A. Beger, B. Schultz, A. Schultz, Correlation of narcotrend index, entropy measures, and spectral parameters with calculated propofol effect-site

concentrations during induction of propofol-remifentanil anaesthesia, J. Clin. Monitor. Comp. 18 (2005) 231–240.
[10] B. Homnan, W. Benjapolakul, Application of fuzzy inference to CDMA soft hand-off in mobile communication systems, Fuzzy Sets Syst. 144 (2) (2004) 345–363.
[11] E. Hüllermeier, Fuzzy methods in machine learning and data mining: status and prospects, Fuzzy Sets Syst. 156 (3) (2005) 387–406.
[12] M.Z. Jahromi, M. Taheri, A proposed method for learning rule weights in fuzzy rule-based classification systems, Fuzzy Sets Syst. 159 (4) (2008) 449–459.
[13] S. Janjarasjitt, K.A. Loparo, An approach for characterizing coupling in dynamical systems, Physica D 237 (2008) 2482–2486.
[14] S. Katsev, I.L. Heureux, Are Hurst exponents estimated from short or irregular time series meaningful? Comput. Geosci. 29 (2003) 1085–1089.
[15] L. Lara, Heuristic determination of the local Lyapunov exponents, Chaos Solitons Fractals 37 (2008) 1208–1213.
[16] A. Lempel, J. Ziv, Complexity of finite sequences, IEEE Trans. Inform. Theory 22 (1976) 75–81.
[17] D.T. MacIsaac, P.A. Parker, K.B. Englehart, D.R. Rogers, Fatigue estimation with a multivariable myoelectric mapping function, IEEE Trans. Biomed. Eng. 53 (4) (2006) 694–700.
[18] R. Merletti, M. Knaflitz, C. Deluca, Myoelectric manifestations of fatigue in voluntary and electrically elicited contractions, J. Appl. Physiol. 69 (1990) 1810–1820.
[19] G. Nicolis, I. Prigogine, Exploring Complexity: An Introduction, W.H. Freeman, New York, 1989.
[20] S.M. Pincus, Approximate entropy as a measure of system complexity, Proc. Natl. Acad. Sci. USA 88 (1991) 2297–2301.
[21] S.M. Pincus, Approximate entropy (ApEn) as a complexity measure, Chaos 5 (1995) 110–117.
[22] S.M. Pincus, Approximate entropy as a measure of irregularity for psychiatric serial metrics, Bipolar Disord. 8 (2006) 430–440.
[23] S.M. Pincus, A.L. Goldberger, Physiological time-series analysis: what does regularity quantify? Am. J. Physiol. Heart Circ. Physiol. 266 (1994) H1643–H1656.
[24] S.M. Pincus, D.L. Keefe, Quantification of hormone pulsatility via an approximate entropy algorithm, Am. J. Physiol. Endocrinol. Metab. 262 (1992) E741–E754.
[25] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate and sample entropy, Am. J. Physiol. Heart Circ. Physiol. 278 (2000) H2039–H2049.
[26] N. Radhakrishnan, B.N. Gangadhar, Estimating regularity in epileptic seizure time-series, IEEE Eng. Med. Biol. Mag. 17 (3) (1998) 89–94.
[27] O.E. Rossler, An equation for continuous chaos, Phys. Lett. A 57 (1996) 397–398.
[28] H. Tahani, J.M. Keller, Information fusion in computer vision using the fuzzy integral, IEEE Trans. Syst. Man Cybern. 20 (3) (1990) 733–741.
[29] M.E. Torres, L.G. Gamero, Relative complexity changes in time series using information measures, Physica A 286 (2000) 457–473.
[30] A. Wolf, J.B. Swift, H.L. Swinney, J.A. Vastano, Determining Lyapunov exponents from a time series, Physica D 16 (1985) 285–317.
[31] Y.G. Xu, Z.J. He, Research on comparison between approximate entropy and fractal dimension for complexity measure of signals, J. Vibration Shock 22 (2003) 25–27.
[32] H.B. Xie, Z.Z. Wang, Mean frequency derived via Hilbert-Huang transform with application to fatigue EMG signal analysis, Comput. Meth. Programs. Biomed. 82 (2) (2006) 114–120.
[33] H.B. Xie, Z.Z. Wang, H. Huang, Identification determinism in time series based on symplectic geometry spectra, Phys. Lett. A 342 (2005) 156–161.
[34] L.A. Zadeh, Fuzzy sets, Info. Control 8 (1965) 338–353.
[35] J.P. Zbilut, C.L. Webber, Embeddings and delays as derived from quantification of recurrence plots, Phys. Lett. A 171 (1992) 199–203.