

# Multiscale entropy analysis of biological signals: a fundamental bi-scaling law

Jianbo Gao<sup>1,2\*</sup>, Jing Hu<sup>2</sup>, Feiyan Liu<sup>1,3</sup> and Yinhe Cao<sup>1,2</sup>

<sup>1</sup> Institute of Complexity Science and Big Data Technology, Guangxi University, Nanning, China, <sup>2</sup> PMB Intelligence LLC, Sunnyvale, CA, USA, <sup>3</sup> School of Management, University of Chinese Academy of Sciences, Beijing, China

## OPEN ACCESS

### Edited by:

Tobias Alecio Mattei,  
Brain & Spine Center - InvisionHealth  
- Kenmore Mercy Hospital, USA

### Reviewed by:

Guillaume Lajoie,  
Max Planck Institute for Dynamics and  
Self-Organization, Germany  
Bailu Si,  
Chinese Academy of Sciences, China  
Xiaoli Li,  
Beijing Normal University, China

### \*Correspondence:

Jianbo Gao,  
Institute of Complexity Science and  
Big Data Technology, Guangxi  
University, 100 Daxue Road, Nanning,  
Guangxi 530005, China  
jbgao.pmb@gmail.com

**Received:** 14 December 2014

**Accepted:** 14 May 2015

**Published:** 02 June 2015

### Citation:

Gao J, Hu J, Liu F and Cao Y (2015)  
Multiscale entropy analysis of  
biological signals: a fundamental  
bi-scaling law.  
*Front. Comput. Neurosci.* 9:64.  
doi: 10.3389/fncom.2015.00064

Since introduced in early 2000, multiscale entropy (MSE) has found many applications in biosignal analysis, and been extended to multivariate MSE. So far, however, no analytic results for MSE or multivariate MSE have been reported. This has severely limited our basic understanding of MSE. For example, it has not been studied whether MSE estimated using default parameter values and short data set is meaningful or not. Nor is it known whether MSE has any relation with other complexity measures, such as the Hurst parameter, which characterizes the correlation structure of the data. To overcome this limitation, and more importantly, to guide more fruitful applications of MSE in various areas of life sciences, we derive a fundamental bi-scaling law for fractal time series, one for the scale in phase space, the other for the block size used for smoothing. We illustrate the usefulness of the approach by examining two types of physiological data. One is heart rate variability (HRV) data, for the purpose of distinguishing healthy subjects from patients with congestive heart failure, a life-threatening condition. The other is electroencephalogram (EEG) data, for the purpose of distinguishing epileptic seizure EEG from normal healthy EEG.

**Keywords:** scaling law, multiscale entropy analysis, fractal signal, heart rate variability (HRV), adaptive filtering

## 1. Introduction

Biological systems provide the definitive examples of highly integrated systems functioning at multiple time scales. Neurons function on a time scale of milliseconds. Circadian rhythms operate on time scale of hours, reproductive cycles occur on a time scale of weeks, and bone remodeling involves time scales of months. As an integrated system, each process interacts with faster and slower processes. Consequently, biosignals often are multiscaled (Gao et al., 2007)—depending upon the scale at which the signals are examined, they may exhibit different behaviors (e.g., nonlinearity, sensitive dependence on small disturbances, long memory, extreme variations, and nonstationarity), just as a great painting may exhibit various details and arouse a multitude of aesthetic feelings when appreciated at different distances, from different angles, under different illuminations, and under different moods.

With the rapid advance of sensing technology, complex data have been accumulating exponentially in all areas of life sciences. To better cope with such complex data, recently, Costa et al. (2005) have introduced an interesting method, the multiscale entropy (MSE) analysis. MSE has found numerous applications in various types of biosignal analysis, including fetal heart rate monitoring (Cao et al., 2006), assessment of EEG dynamical complexity in

Alzheimer's disease (Mizuno et al., 2010), classification of surface EMG of neuromuscular disorders (Istencic et al., 2010), heart rate analysis for predicting hospital mortality (Norris et al., 2008), and analysis of heart beat interval and blood flow for characterizing psychological dimensions in non-pathological subjects (Nardelli et al., 2015). MSE has also been extended to multivariate MSE (Ahmed and Mandic, 2011) and multiscale permutation entropy (Li et al., 2010). So far, however, no analytic analyses about MSE or multivariate MSE have been carried out. This has severely limited our basic understanding of MSE. For example, it has not been known whether MSE estimated using default parameter values and short data set is meaningful or not. Nor is it known whether MSE has any relation with other complexity measures, such as the Hurst parameter, which characterizes the correlation structure of the data.

To help gain insights into the above questions, and to guide more fruitful applications of MSE in diverse fields of life sciences, in this work, we report a fundamental bi-scaling law for MSE of the most popular model of biosignals, the fractal  $1/f$  type time series. As example applications, we will analyze heart rate variability (HRV) and electroencephalogram (EEG) data. With HRV, we will focus on distinguishing healthy subjects from patients with congestive heart failure (CHF), a life-threatening condition, as well as resolving an interesting debate (Wessel et al., 2003; Nikulin and Brismar, 2004) regarding the usefulness of MSE in distinguishing HRV of healthy subjects from that of patients with certain cardiac disease. With EEG, we will focus on distinguishing epileptic seizure EEG from normal healthy EEG.

## 2. Materials and Methods

### 2.1. Data

To illustrate the use of scaling analysis of MSE, in this paper, we analyze two types of data, heart rate variability (HRV), for the purpose of distinguishing healthy subjects from patients with congestive heart failure (CHF), and EEG, for the detection of epileptic seizures.

We downloaded two types of HRV data from the PhysioNet (MIT-BIH Normal Sinus Rhythm Database and BIDMC Congestive Heart Failure Database available at <http://www.physionet.org/physiobank/database/#ecg>), one for healthy subjects, and the other for subjects with CHF. The latter includes long-term ECG recordings from 15 subjects (11 men, aged 22 to 71, and 4 women, aged 54 to 63) with severe CHF (NYHA class 3–4). This group of subjects was part of a larger study group receiving conventional medical therapy prior to receiving the oral inotropic agent, milrinone. Further details about the larger study group can be found at the PhysioNet. The individual recordings of ECG are each about 20 h in duration, and contain two ECG signals each sampled at 250 samples per second with 12-bit resolution over a range of  $\pm 10$  millivolts. The other database are for 18 normal subjects. The individual recordings are each about 25 h in duration, each sampled at 128 samples per second. The HRV data analyzed here are the R-R intervals (in unit of second) derived from the ECG recordings.

The EEG database is downloaded at <http://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html>. The

database consists of three groups, H (healthy), E (epileptic subjects during a seizure-free interval), and S (epileptic subjects during seizure); each group contains 100 data segments, whose length is 4097 data points with a sampling frequency of 173.61 Hz. These data have been carefully examined by adaptive fractal analysis (Gao et al., 2011c) and scale-dependent Lyapunov exponent (Gao et al., 2006b, 2011b, 2012), for the same purpose of distinguishing epileptic seizure EEG from normal healthy EEG.

### 2.2. Methods

Entropy characterizes creation of information in a dynamical system. To facilitate derivation of a fundamental scaling law for MSE, we first rigorously define MSE and all related concepts.

Suppose that the  $F$ -dimensional phase space is partitioned into boxes of size  $\varepsilon^F$ . Suppose that there is an attractor in phase space and consider a transient-free trajectory  $\vec{x}(t)$ . The state of the system is now measured at intervals of time  $\tau$ . Let  $p(i_1, i_2, \dots, i_d)$  be the joint probability that  $\vec{x}(t = \tau)$  is in box  $i_1$ ,  $\vec{x}(t = 2\tau)$  is in box  $i_2, \dots$ , and  $\vec{x}(t = d\tau)$  is in box  $i_d$ . Let us now introduce the block entropy,

$$H_d(\varepsilon, \tau) = - \sum_{i_1, \dots, i_d} p(i_1, \dots, i_d) \ln p(i_1, \dots, i_d), \quad (1)$$

take the difference between  $H_{d+1}(\varepsilon, \tau)$  and  $H_d(\varepsilon, \tau)$ , and normalize it by  $\tau$ ,

$$h_d(\varepsilon, \tau) = \frac{1}{\tau} [H_{d+1}(\varepsilon, \tau) - H_d(\varepsilon, \tau)]. \quad (2)$$

Let

$$h(\varepsilon, \tau) = \lim_{d \rightarrow \infty} h_d(\varepsilon, \tau) \quad (3)$$

It is called the  $(\varepsilon, \tau)$ -entropy (Gaspard and Wang, 1993). Taking limits, we obtain the Kolmogorov-Sinai (K-S) entropy,

$$\begin{aligned} K &= \lim_{\tau \rightarrow 0} \lim_{\varepsilon \rightarrow 0} h(\varepsilon, \tau) \\ &= \lim_{\tau \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \lim_{d \rightarrow \infty} \frac{1}{\tau} [H_{d+1}(\varepsilon, \tau) - H_d(\varepsilon, \tau)] \end{aligned} \quad (4)$$

We now consider computation of the  $(\varepsilon, \tau)$ -entropy from a time series of length  $N$ ,  $x_1, x_2, \dots, x_N$ . As is well-known, the first step is to use the time delay embedding to construct vectors of the form:

$$V_i = [x_i, x_{i+L}, \dots, x_{i+(m-1)L}], \quad (5)$$

where  $m$ , the embedding dimension, and  $L$ , the delay time, can be chosen according to certain optimization criterion (Gao et al., 2007). Then one can employ the Cohen-Procaccia algorithm (Cohen and Procaccia, 1985) to estimate the  $(\varepsilon, \tau)$ -entropy. In particular, when it is evaluated at a fixed finite scale  $\hat{\varepsilon}$ , the resulting entropy is called the approximate entropy. To get better statistics from a finite time series, one may compute  $K_2(\varepsilon)$

using the Grassberger-Procaccia's algorithm (Grassberger and Procaccia, 1983):

$$K_2(\epsilon) = \lim_{m \rightarrow \infty} \frac{\ln C^{(m)}(\epsilon) - \ln C^{(m+1)}(\epsilon)}{mL\delta t} \quad (6)$$

where  $\delta t$  is the sampling time,  $C^{(m)}(\epsilon)$  is the correlation integral based on the  $m$ -dimensional reconstructed vectors  $V_i$  and  $V_j$ ,

$$C^{(m)}(\epsilon) = \lim_{N_v \rightarrow \infty} \frac{2}{N_v(N_v - 1)} \sum_{i=1}^{N_v-1} \sum_{j=i+1}^{N_v} H(\epsilon - \|V_i - V_j\|), \quad (7)$$

where  $N_v = N - (m - 1)L$  is the number of reconstructed vectors,  $H(y)$  is the Heaviside function (1 if  $y \geq 0$  and 0 if  $y < 0$ ).  $C^{(m+1)}(\epsilon)$  can be computed similarly based on the  $m + 1$ -dimensional reconstructed vectors. When we evaluate  $K_2(\epsilon)$  at a finite fixed scale  $\hat{\epsilon}$ , we obtain the sample entropy  $S_e$  (Richman and Moorman, 2000).

MSE analysis is based on the sample entropy  $S_e$ . The procedure is as follows. Let  $X = \{x_t : t = 1, 2, \dots\}$  be a covariance stationary stochastic process with mean  $\mu$ , variance  $\sigma^2$ , and autocorrelation function  $r(k), k \geq 0$ . Construct a new covariance stationary time series

$$X^{(b_s)} = \{x_t^{(b_s)} : t = 1, 2, 3, \dots\}, \quad b_s = 1, 2, 3, \dots,$$

by averaging the original series  $X$  over non-overlapping blocks of size  $b_s$ ,

$$x_t^{(b_s)} = (x_{tb_s-b_s+1} + \dots + x_{tb_s})/b_s, \quad t \geq 1. \quad (8)$$

MSE analysis involves (i) choosing a finite scale  $\hat{\epsilon}$  in phase space, and (ii) computing  $S_e$  from the original and the smoothed data  $X$  and  $X^{(b_s)}$  at the chosen scale  $\hat{\epsilon}$ . For convenience of later discussion, we denote  $K_2^{(b_s)}(\epsilon)$  for the correlation entropy of the smoothed data. When  $b_s = 1$ , it is the correlation entropy of the original data, and can be simply denoted as  $K_2(\epsilon)$ .

We emphasize that the length of the smoothed time series is only  $1/b_s$  of the original one. To fully resolve the scaling behavior of  $K_2(\epsilon)$ , the requirement on data length is quite stringent. A fundamental question is whether MSE calculated from short noisy data is meaningful or not.

### 3. Results

#### 3.1. Scaling for the MSE of Fractal Time Series

Among the most widely used models for biological signals, including HRV, EEG, and posture (Gao et al., 2011a), is the fractal time series with long memory, the so-called  $1/f^\alpha$ , or  $1/f^{2H-1}$ ,  $\alpha = 2H - 1$  processes, where  $0 < H < 1$  is called the Hurst parameter, whose value determines the correlation structure of the data (Gao et al., 2006a, 2007): when  $H = 1/2$ , the process is like the independent steps of the standard Brownian-motion; when  $H < 1/2$ , the process has anti-persistent correlations; when  $H > 1/2$ , the process has persistent correlations. Two special cases, white noise with  $H = 0.5$  and

$1/f$  process with  $H = 1$ , have been extensively used for the development of multivariate MSE (Ahmed and Mandic, 2011). In this subsection, we derive fundamental scalings for MSE of the ubiquitous  $1/f^{2H-1}$  noise.

A covariance stationary stochastic process  $X = \{X_t : t = 0, 1, 2, \dots\}$ , with mean  $\mu$ , variance  $\sigma^2$ , and autocorrelation function  $r(w), w \geq 0$ , is said to have long range correlation if  $r(w)$  is of the form Cox (1984)

$$r(w) \sim w^{2H-2}, \quad \text{as } w \rightarrow \infty, \quad (9)$$

where  $0 < H < 1$  is the Hurst parameter. When  $1/2 < H < 1$ ,  $\sum_w r(w) = \infty$ , leading to the term long range correlation. Note the  $X$  time series has a power spectral density  $1/f^{2H-1}$ . Its integration,  $\{y_t\}$ , where  $y_t = \sum_{i=1}^t x_i$ , is called a random walk process which is nonstationary with power-spectral density (PSD)  $1/f^{2H+1}$ . Being  $1/f$  processes, they cannot be aptly modeled by Markov processes or ARIMA models (Box and Jenkins, 1976), since the PSD for those processes are distinctly different from  $1/f$ . To adequately model  $1/f$  processes, fractional order processes has to be used. The most popular is the fractional Brownian motion model Mandelbrot (1982), whose increment process is called the fractional Gaussian noise (fGn). The importance and popularity of fGn in modeling various types of noises in science and engineering motivates us to focus our analysis on it when deriving the bi-scaling law.

$1/f^{2H-1}$  noises are self-similar, with the autocorrelation for the original data and the smoothed data (defined by Equation 8) being the same (Gao et al., 2006a, 2007). This signifies that there must exist a simple relation between  $K_2^{(b_s)}(\epsilon)$  and  $K_2(\epsilon)$ . To find this relation, we note that the variance,  $\text{var}(X^{(b_s)})$ , of the smoothed data, and the variance,  $\sigma^2$ , of the original data, are related by the following simple and elegant scaling law (Gao et al., 2006a, 2007),

$$\text{var}(X^{(b_s)}) = \sigma^2 b_s^{2H-2} \quad (10)$$

Equation (10) states that the scale  $\epsilon$  for the original data is transformed to a smaller scale  $b_s^{H-1}\epsilon$  for the smoothed data. Using the self-similarity property of the  $1/f^{2H-1}$  noise, we therefore obtain,

$$K_2^{(b_s)}(b_s^{H-1}\epsilon) = K_2(\epsilon) \quad (11)$$

Since for stationary random processes,  $K_2(\epsilon)$  diverges when  $\epsilon \rightarrow 0$ , Equation (11) states that  $K_2^{(b_s)}(b_s^{H-1}\epsilon)$  can be obtained from  $K_2(\epsilon)$  by shifting downward the curve for  $K_2(\epsilon)$ . How much  $K_2(\epsilon)$  should be shifted depends on the functional form for  $K_2(\epsilon)$ , which we shall find out momentarily.

First we note that for 1-D independent random variables, which correspond to  $H = 1/2$ ,  $h(\epsilon, \tau) \sim -\ln \epsilon$  (Gaspard and Wang, 1993). Therefore,  $K_2(\epsilon) \sim -\ln \epsilon$ . In fact, for any stationary noise process, irrespective of its correlation structure, we always have  $C^{(m)}(\epsilon) \sim \epsilon^{-m}$ ,  $\epsilon \rightarrow 0$ , therefore,

$$K_2(\epsilon) \sim -\ln \epsilon, \quad \epsilon \rightarrow 0 \quad (12)$$

Equation (12) is, however, not adequate for us to understand the scaling of  $K_2(\epsilon)$  on finite scales. To gain more insights, we resort to the rate distortion function or the Shannon-Kolmogorov (SK) entropy (Berger, 1971; Gaspard and Wang, 1993). It is thought to diverge with  $\epsilon$  in the same way as the  $(\epsilon, \tau)$ -entropy and  $K_2(\epsilon)$  (Gaspard and Wang, 1993).

Suppose we wish to approximate the random signal  $X(t)$  by  $Z(t)$  according to

$$\rho(X, Z) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \langle [X(t) - Z(t)]^2 \rangle dt \leq \epsilon^2 \quad (13)$$

where  $\langle \cdot \rangle$  denotes averaging. Equation (13) may be considered a partition of the phase space containing the random signal  $X(t)$  by centering around  $X(t)$ . Denote the conditional probability density for  $Z$  given  $x$  by  $q(z|x)$ . The mutual information  $I(q)$  between  $X$  and  $Z$  is a functional of  $q(z|x)$ ,

$$I(q) = \int \int dx dz p(x) q(z|x) \ln[q(z|x)/q(z)]. \quad (14)$$

The SK  $(\epsilon, \tau)$ -entropy is

$$H_{SK}(\epsilon, \tau, T) = \text{Inf}_{q \in Q(\epsilon)} I(q) \quad (15)$$

where  $Q(\epsilon)$  is the set of all conditional probabilities  $q(z|x)$  such that Condition (13) is satisfied. The SK  $(\epsilon, \tau)$ -entropy per unit time is then

$$h_{SK}(\epsilon, \tau) = \lim_{T \rightarrow \infty} H_{SK}(\epsilon, \tau, T)/T \quad (16)$$

For stationary Gaussian processes,  $h_{SK}(\epsilon, \tau)$  can be readily computed by the Kolmogorov formula (Berger, 1971; Kolmogorov, 1956). In the case of a discrete-time process, it reads

$$\epsilon^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min[\theta, \Phi(\omega)] d\omega \quad (17)$$

$$h_{SK}(\epsilon, \tau) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \max\{0, \ln[\Phi(\omega)/\theta]\} d\omega \quad (18)$$

where  $\Phi(\omega)$  is the PSD of the process and  $\theta$  is an intermediate variable.

We now evaluate the SK entropy for a popular model of  $1/f^{2H-1}$  noise, the fractional Gaussian noise (fGn). It is a stationary Gaussian process with PSD  $1/\omega^{2H-1}$ . Since we are primarily interested in small  $\epsilon$ , we may choose the intermediate variable  $\theta \leq \Phi(\omega)$ . Let us denote  $\Phi(\omega) = B(H)\omega^{1-2H}$ , where  $B(H)$  is a factor depending on  $H$ . When  $H = 1/2$ , it equals the variance of the noise  $\sigma_{H=1/2}^2$ . Using Equations (17) and (18), we immediately have

$$h_{SK}(\epsilon) = A(H) - \ln \epsilon \quad (19)$$

where

$$A(H) = \frac{1-2H}{2} (\ln \pi - 1) + \frac{1}{2} \ln B(H) \quad (20)$$

If we assume fGn of different  $H$  to have the same variance, then  $\int_0^\pi \Phi(\omega) d\omega$  is a constant independent of  $H$ .  $A(H)$  can then be written as

$$A(H) = \frac{1}{2} \ln \sigma_{H=1/2}^2 + \frac{1}{2} [\ln(2-2H) - (1-2H)] \quad (21)$$

$A(H)$  is maximal when  $H = 1/2$ . However, when  $H$  is not close to 0 or 1, the term  $\frac{1}{2} [\ln(2-2H) - (1-2H)]$  is negligibly small, signifying that  $h_{SK}(\epsilon)$  cannot readily classify fGn of different  $H$ .

Since  $h_{SK}(\epsilon)$  and  $K_2(\epsilon)$  diverge in the same fashion (Gaspard and Wang, 1993), using Equation (12) to determine the prefactor, we have a scaling for finite  $\epsilon$

$$K_2(\epsilon) \sim -\ln \epsilon \quad (22)$$

Combining Equations (22) and (11), we arrive at a fundamental bi-scaling law for  $K_2^{(b_s)}(\epsilon)$  for fractal time series:

$$K_2^{(b_s)}(\epsilon) \sim (H-1) \ln b_s - \ln \epsilon \quad (23)$$

To verify the above bi-scaling law, and more importantly, to gain insights into the relative importance of the two scale parameters  $b_s$  and  $\epsilon$  in MSE analysis, we numerically perform MSE analysis of fGn processes with different  $H$ . A few examples are shown in **Figures 1, 2**. The computations are done with  $2^{14}$  points and  $m = 2$ . We observe excellent bi-scaling relations, thus verifying Equation (23). Recalling our earlier comment that  $K_2(\epsilon)$  itself is not very useful for distinguishing fGn of different  $H$ , **Figure 2** clearly shows that the scaling  $K_2^{(b_s)}(\epsilon) \sim (H-1) \ln b_s$  can aptly separate fGn processes of different  $H$ . In fact,  $H$  values estimated from **Figure 2** are fully consistent the values of  $H$  chosen in simulating the fGn processes. This analysis thus has demonstrated the major advantage of the scale parameter  $b_s$  over  $\epsilon$  for the study of fGn processes using MSE. It has also made it clear that MSE is a highly non-trivial extension of the sample entropy, and more generally, the correlation entropy  $K_2(\epsilon)$ .

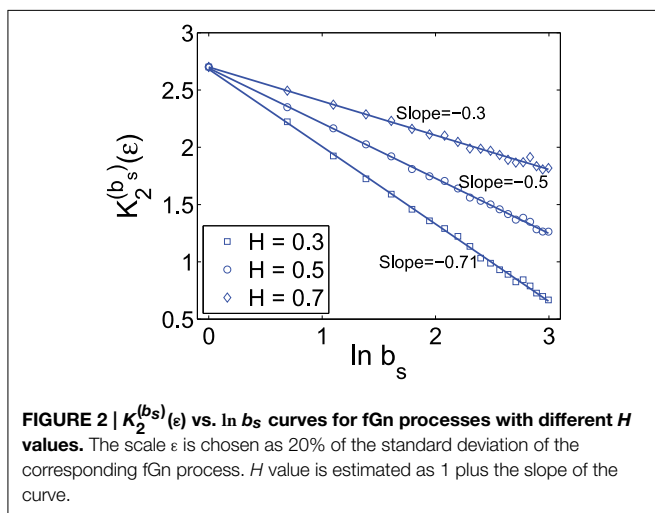
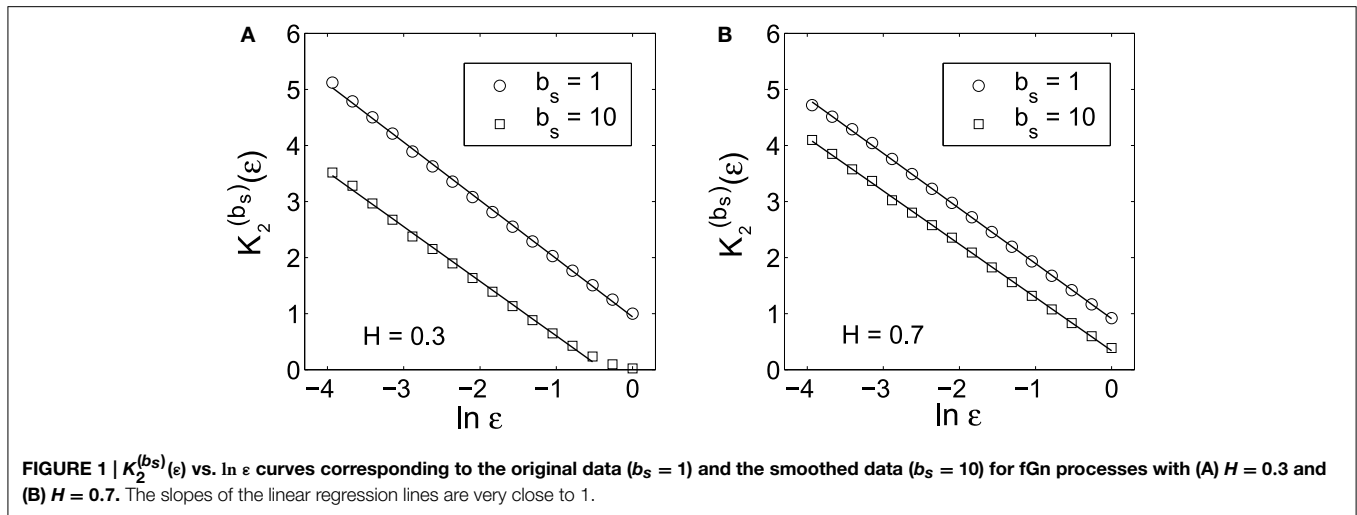
While Equation (23) is fundamental for MSE, it can also help us better understand the behavior of multivariate MSE, which is shown in numerical simulations to be almost constant for  $1/f$  processes with  $H = 1$ , and decays in a well-defined fashion for white noise, where  $H = 1/2$ , and some randomized data derived from experimental data possibly with correlations (Ahmed and Mandic, 2011). The reason is very clear. For  $1/f$  process,  $H = 1$ , and therefore, MSE or multivariate MSE does not vary with the scale parameter  $b_s$ . For white noise or some derived randomized data,  $H = 1/2$ , and therefore, MSE or multivariate MSE decays with the scale parameter  $b_s$  in a well-defined fashion,

$$K_2^{(b_s)}(\epsilon) \sim -\frac{1}{2} \ln b_s, \quad \text{or} \quad b_s \sim e^{-2K_2(\epsilon)}. \quad (24)$$

One can readily check that the MSE curve for white noise shown in Ahmed and Mandic (2011) is fully consistent with the formula derived here.

### 3.2. Heart Rate Variability Data Analysis

As an important application of MSE, we analyze HRV data for the purpose of distinguishing healthy subjects from patients with



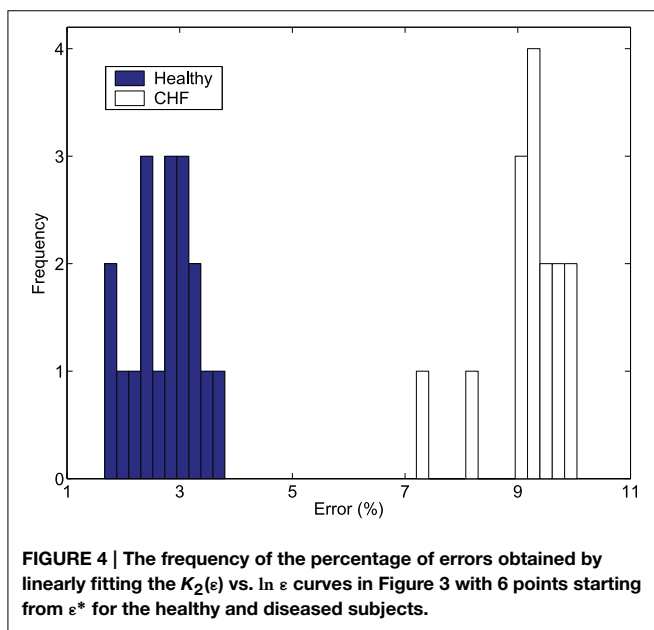
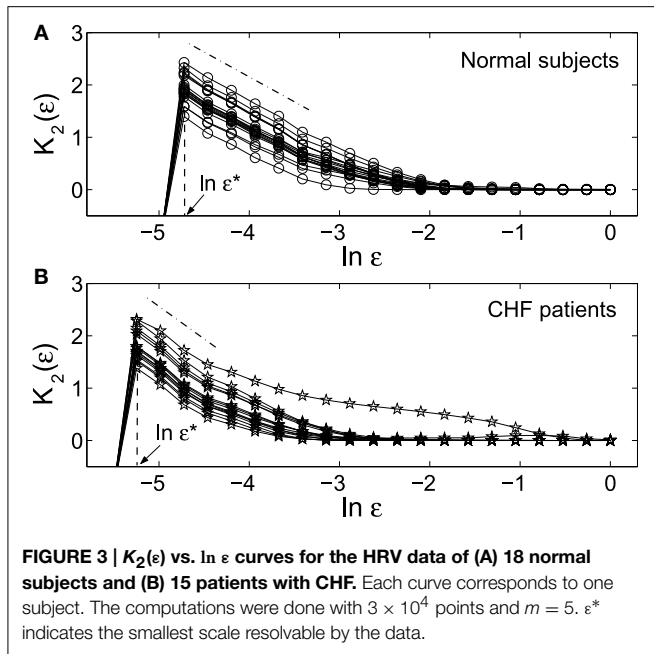
CHF, a life-threatening condition. This is an important issue. We refer to (Hu et al., 2009, 2010) and references therein for the background. Note that part of the data examined here were analyzed in prior work (Ivanov et al., 1999; Barbieri and Brown, 2006), for the same purpose. We analyze all 33 datasets here. For ease of comparison, we take the first  $3 \times 10^4$  points of both groups of HRV data for analysis. Note that based on different  $b_s$  parameter, MSE was not very good at separating the two groups (Hu et al., 2010). This instigated a debate on whether MSE was useful or not for analyzing HRV (Wessel et al., 2003; Nikulin and Brismar, 2004). To resolve this interesting debate, and more importantly, to satisfactorily separate the two groups of HRV data, we shall focus on the dependence of MSE on the scale parameter  $\epsilon$  in the following discussions.

Since earlier studies find HRV data to be nonstationary, having  $1/f$  spectrum with anti-persistent long-range correlations and multifractality (see Ivanov et al., 1999 and references therein), we analyze the increment processes of the HRV data. **Figure 3** shows  $K_2(\epsilon)$  vs.  $\ln \epsilon$  curves for the two groups of HRV data. We observe:

(i) On small scales,  $K_2(\epsilon)$  vs.  $\ln \epsilon$  curves for both groups of HRV data show good scaling behavior. As a consequence, one can expect a scaling relation between  $K_2^{(b_s)}(\epsilon)$  and  $\ln b_s$  (Equation 23). This is indeed so. The results, being very similar to that shown in **Figure 2**, are not shown here, however. (ii) The scaling of  $K_2(\epsilon)$  vs.  $\ln \epsilon$  is better and longer for the normal HRV data. (iii) As indicated by  $\epsilon^*$  in the figure, the smallest scale resolvable by the HRV data of the healthy subjects is much larger than that of the diseased subjects.

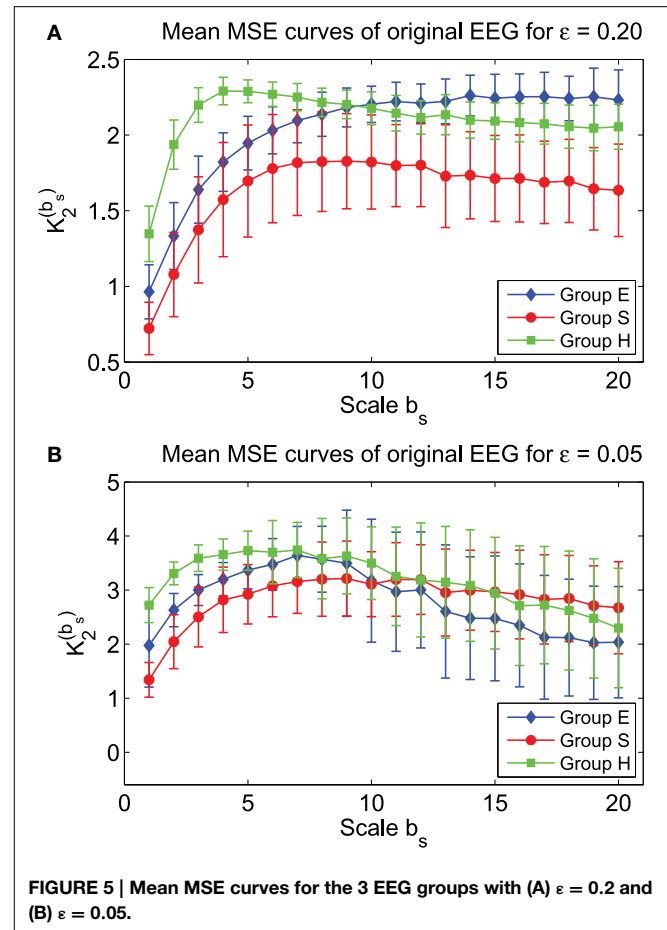
We now discuss how to use MSE to distinguish the healthy subjects from patients with CHF. We have found (i) The curves  $K_2^{(b_s)}(\epsilon)$  vs.  $b_s$  averaged over all the subjects within the two groups are different, just as reported in Costa et al. (2005). However, such curves are not very useful for separating the two groups as a diagnostic tool, as pointed out in Nikulin and Brismar (2004). The fundamental reason is of course that the Hurst parameter  $H$  is not very effective in distinguishing healthy subjects from patients with HRV, as quantitatively analyzed in Hu et al. (2010). (ii) The smallest resolvable scale,  $\epsilon^*$ , completely separates the healthy subjects from patients with CHF, as shown by **Figure 3**. Note the scale parameter  $\epsilon$  is a generalization of the concept variance (or standard deviation). The observation made by Nikulin and Brismar (2004) that a variance-like parameter is better than MSE with varying block size parameter  $b_s$  in distinguishing healthy subjects from patients with HRV is most appropriately interpreted as the following: the parameter  $b_s$  is less important than the scale parameter  $\epsilon$ . This is somewhat the opposite of the case for  $1/f$  noise analyzed in the last section.

To more clearly see how much more advantageous  $\epsilon$  is over  $b_s$  in distinguishing healthy subjects from patients with HRV, we examine how the scaling  $K_2(\epsilon) \sim -\ln \epsilon$  can be used for this purpose. We have found that the errors obtained by linearly fitting the  $K_2(\epsilon)$  vs.  $\ln \epsilon$  curves of **Figure 3** are much smaller for the normal HRV data than for those of CHF patients and also can completely separate the healthy subjects from patients with CHF. This is shown in **Figure 4**. Therefore, the scale parameter  $\epsilon$  is indeed more important than  $b_s$ .



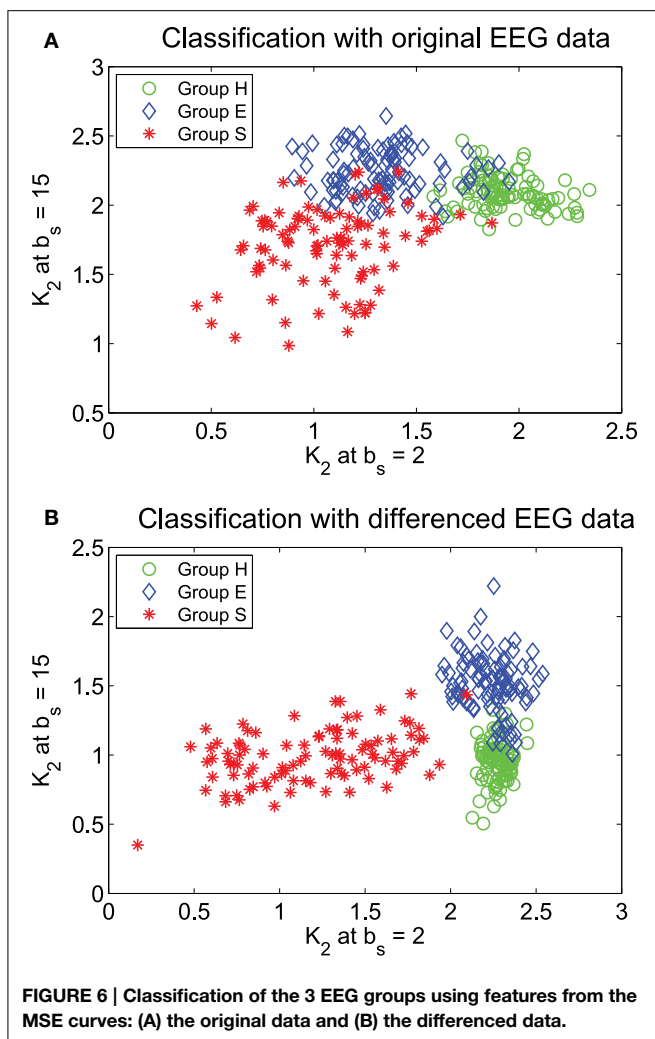
### 3.3. Epileptic Seizure Detection Through MSE of EEG

Epilepsy is a common and debilitating brain disorder. It is characterized by intermittent seizures. During a seizure, the normal activity of the central nervous system is disrupted. The concrete symptoms include abnormal running/bouncing fits, clonus of face and forelimbs, or tonic rearing movement as well as simultaneous occurrence of transient EEG signals such as spikes, spike and slow wave complexes or rhythmic slow wave bursts. Clinical effects may include motor, sensory, affective, cognitive, automatic and physical symptomatology. To make medications effective, timely detection of seizure is very important. In the



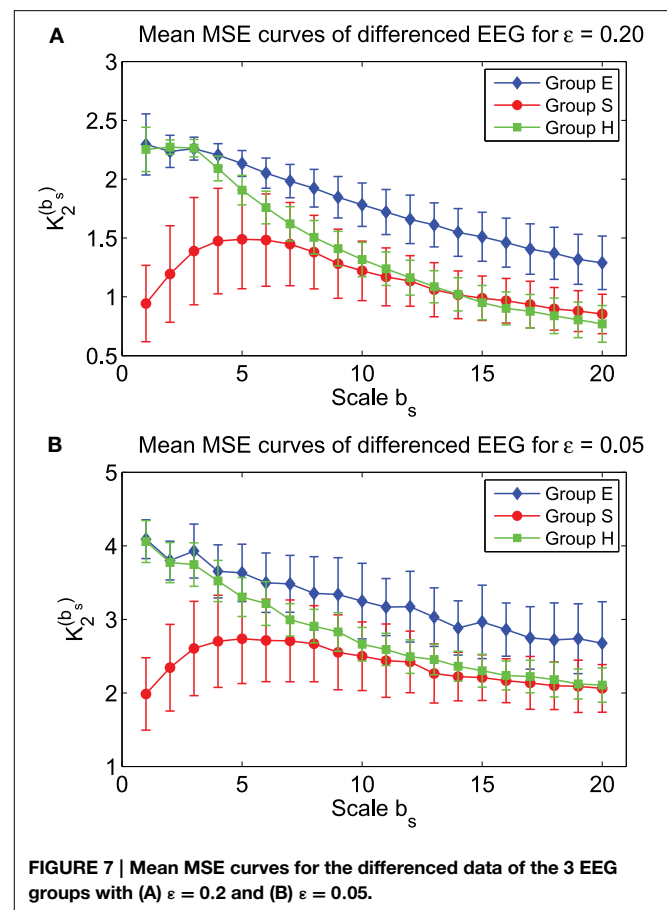
past several decades, considerable efforts have been made to detect/predict seizures through nonlinear analysis of EEGs. For a list of the major nonlinear methods proposed for seizure detection, we refer to Gao and Hu (2013) and references therein. In particular, the three groups of EEG data analyzed here, H (healthy), E (epileptic subjects during a seizure-free interval), and S (epileptic subjects during seizure), were examined by adaptive fractal analysis (Gao et al., 2011c) and scale-dependent Lyapunov exponent (Gao et al., 2012), and excellent classification was achieved.

To examine how well MSE characterizes the three groups of EEG data, we have plotted in Figure 5 the mean MSE curves for the three groups, for two parameter values of the phase space scale,  $\epsilon$ . We observe that they separate very well. Indeed, statistical test shows that the separations are significant. In particular, for the scale parameter in the phase space  $\epsilon = 0.2$ , the MSE curve for the S group lies well below the other 2 curves. One may be tempted to equate this as smaller complexity of the seizure EEG. However, such an interpretation is informative only relative to the specific  $\epsilon$  chosen here, which is 0.2. When  $\epsilon = 0.05$ , the red curve for seizure EEG actually lie above the other 2 curves for larger  $b_s$ . In fact, if one can pause a moment and think twice, one would realize that such interpretations are not too helpful for clinical applications, since MSE can vary substantially within and across the groups.



We have tried to use MSE at specific  $b_s$  values to classify the three groups of EEG. Guided by the mean MSE curves in **Figure 5**, we have found that when  $\varepsilon = 0.2$ , if only two  $b_s$  can be used, then  $b_2 = 2$  and 15 are the optimal values. The result of the classification is shown in **Figure 6A**. We observe that there are some overlaps between groups H (healthy) and E (epileptic subjects during a seizure-free interval), as well as E and S (epileptic subjects during seizure). Intuitively, this is reasonable. Overall, the classification is not very satisfactory. How may we improve the accuracy of the classification?

Recall that in fractal scaling analysis of EEG, EEG data are found to be equivalent to random walk processes, but not noise or increment processes (Gao et al., 2011c). The latter amounts to a differentiation of the random walk processes. Since the basic scaling law derived here is for noise or increment process, but not for random walk processes, it suggests us to try to compute MSE from the differenced data of EEG, defined by  $y_i = x_i - x_{i-1}$ , where  $x_i$  is the original EEG signal. The mean MSE curves for the differenced data of EEG are shown in **Figure 7**, again for two  $\varepsilon$  values. We observe that the separation between the mean MSE curves becomes wider. Indeed, classification of the 3 EEG groups



now is much improved, as shown in **Figure 6B**. It should be noted however that the accuracy of the classification is still slightly worse than using other methods, such as adaptive fractal analysis (Gao et al., 2011c) and scale-dependent Lyapunov exponent (Gao et al., 2012).

## 4. Conclusion and Discussion

To better understand MSE, we have derived a fundamental bi-scaling relation for the MSE analysis. While MSE analysis normally only focuses on the scale parameter  $b_s$  with  $\varepsilon$  more or less arbitrarily chosen, our analysis of fGn and HRV data clearly demonstrates that both scale parameters are important—in the case of HRV analysis, the  $\varepsilon$  is more important, while in the case of  $1/f$  noise, the  $b_s$  parameter is more important. In fact, we have shown (Hu et al., 2010) that MSE, when used with  $\varepsilon$  fixed, is not very effective in distinguishing healthy subjects from patients with HRV. The accuracy achieved when we focus on the scaling of  $K_2(\varepsilon) \sim -\ln \varepsilon$  is not only much higher, but also comparable to that using the scale-dependent Lyapunov exponent (SDLE) (Gao et al., 2006a, 2007, 2013), as reported by Hu et al. (Hu et al., 2010). The fundamental reason of course is that SDLE has a similar scaling as  $K_2(\varepsilon) \sim -\ln \varepsilon$ .

We have also computed MSE for the original as well as the differenced data of the three EEG groups, H (healthy), E (epileptic subjects during a seizure-free interval), and S (epileptic

subjects during seizure), and found that mean MSE curves for the three groups are well separated. The classification of the 3 EEG groups using MSE at two specific scale parameters  $b_s$  is reasonably good, and is better for the differenced data than for the original EEG data. This strongly suggests that EEG data are like random walk processes. However, even with the differenced data of EEG, the classification is still not as accurate as using adaptive fractal analysis (Gao et al., 2011c) and scale-dependent Lyapunov exponent (Gao et al., 2011a). One of the reasons for this inferiority lies in the difference in the range of scales covered by these three multiscale methods. Adaptive fractal analysis and scale-dependent Lyapunov exponent both cover the entire range of scales presented in the EEG data. However, with the length of the EEG data, which is only 4097 points for each data set, MSE can only cover a moderate range

of scales, with the largest  $b_s$  only around 20, since with  $b_s = 20$ , the smoothed data is already only 200 points long. Our analysis here has raised an important question: how do we use MSE to analyze short data? We conjecture that it may be beneficial to focus on the scaling of  $K_2(\epsilon) \sim -\ln \epsilon$ , or develop new smoothing schemes, by introducing a parameter equivalent to  $1/b_s$  but without sacrificing the length of the smoothed data.

## Acknowledgments

One of the authors (JG) is grateful for the generous support by National Institute for Mathematical and Biological Synthesis (NIMBIO) at the University of Tennessee to attend Heart Rhythm Disorders Investigative Workshop.

## References

- Ahmed, M. U., and Mandic, D. P. (2011). Multivariate multiscale entropy: a tool for complexity analysis of multichannel data. *Phys. Rev. E* 84:061918. doi: 10.1103/PhysRevE.84.061918
- Barbieri, R., and Brown, E. N. (2006). Analysis of heartbeat dynamics by point process adaptive filtering. *IEEE Trans. Biomed. Eng.* 53, 4–12. doi: 10.1109/TBME.2005.859779
- Berger, T. (1971). *Rate Distortion Theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Box, G. E. P., and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control, 2nd Edn*. San Francisco, CA: Holden-Day.
- Cao, H. Q., Lake, D. E., Ferguson, J. E., Chisholm, C. A., Griffin, M. P., and Moorman, J. R. (2006). Toward quantitative fetal heart rate monitoring. *IEEE Trans. Biomed. Eng.* 53, 111–118. doi: 10.1109/tbme.2005.859807
- Cohen, A., and Procaccia, I. (1985). Computing the Kolmogorov entropy from time series of dissipative and conservative dynamical systems. *Phys. Rev. A* 31, 1872–1882. doi: 10.1103/PhysRevA.31.1872
- Costa, M., Goldberger, A. L., and Peng, C. K. (2005). Multiscale entropy analysis of biological signals. *Phys. Rev. E* 71:021906. doi: 10.1103/PhysRevE.71.021906
- Cox, D. R. (1984). “Long-range dependence: a review,” in *Statistics: An Appraisal*, eds H. A. David and H. T. Davis (Ames: The Iowa State University Press), 55–74.
- Gao, J. B., Cao, Y. H., Tung, W. W., and Hu, J. (2007). *Multiscale Analysis of Complex Time Series — Integration of Chaos and Random Fractal Theory, and Beyond*. Hoboken, NJ: Wiley.
- Gao, J. B., Hu, J., Tung, W. W., Cao, Y. H., Sarshar, N., and Roychowdhury, V. P. (2006a). Assessment of long range correlation in time series: how to avoid pitfalls. *Phys. Rev. E* 73:016117. doi: 10.1103/PhysRevE.73.016117
- Gao, J. B., Hu, J., Tung, W. W., and Cao, Y. H. (2006b). Distinguishing chaos from noise by scale-dependent Lyapunov exponent. *Phys. Rev. E* 74:066204. doi: 10.1103/PhysRevE.74.066204
- Gao, J. B., Hu, J., Buckley, T., White, K., Hass, C. (2011a). Shannon and Renyi entropies To classify effects of mild traumatic brain injury on postural sway. *PLoS ONE* 6:e24446. doi: 10.1371/journal.pone.0024446
- Gao, J. B., Hu, J., and Tung, W. W. (2011b). Complexity measures of brain wave dynamics. *Cogn. Neurodynamics* 5, 171–182. doi: 10.1007/s11571-011-9151-3
- Gao, J. B., Hu, J., and Tung, W. W. (2011c). Facilitating joint chaos and fractal analysis of biosignals through nonlinear adaptive filtering. *PLoS ONE* 6:e24331. doi: 10.1371/journal.pone.0024331
- Gao, J. B., Hu, J., and Tung, W. W. (2012). Entropy measures for biological signal analysis. *Nonlin. Dynamics* 68, 431–444. doi: 10.1007/s11071-011-0281-2
- Gao, J. B., Gurbaxani, B. M., Hu, J., Heilman, K. J., Emauele, V. A., Lewis, G. F. et al. (2013). Multiscale analysis of heart rate variability in nonstationary environments. *Front. Comput. Physiol. Med.* 4:119. doi: 10.3389/fphys.2013.00119
- Gao, J. B., and Hu, J. (2013). Fast monitoring of epileptic seizures using recurrence time statistics of electroencephalography. *Front. Comput. Neurosci.* 7:122. doi: 10.3389/fncom.2013.00122
- Gaspard, P., and Wang, X. J. (1993). Noise, chaos, and  $(\epsilon, \tau)$ -entropy per unit time. *Phys. Rep.* 235, 291–343. doi: 10.1016/0370-1573(93)90012-3
- Grassberger, P., and Procaccia, I. (1983). Estimation of the Kolmogorov entropy from a chaotic signal. *Phys. Rev. A* 28, 2591–2593. doi: 10.1103/PhysRevA.28.2591
- Hu, J., Gao, J. B., and Tung, W. W. (2009). Characterizing heart rate variability by scale-dependent Lyapunov exponent. *Chaos* 19, 028506. doi: 10.1063/1.3152007
- Hu, J., Gao, J. B., Tung, W. W., and Cao, Y. H. (2010). Multiscale analysis of heart rate variability: a comparison of different complexity measures. *Ann. Biom. Eng.* 38, 854–864. doi: 10.1007/s10439-009-9863-2
- Istencik, R., Kaplanis, P. A., Pattichis, C. S., Zazula, D. (2010). Multiscale entropy-based approach to automated surface EMG classification of neuromuscular disorders. *Medi. Biol. Eng. Comput.* 48, 773–781. doi: 10.1007/s11517-010-0629-7
- Ivanov, P. C., Amaral, L. A. N., Goldberger, A. L., Havlin, S., Rosenblum, M. G., Struzik, Z. R. (1999). Multifractality in human heartbeat dynamics. *Nature* 399, 461–465. doi: 10.1038/20924
- Kolmogorov, A. N. (1956). On the Shannon theory of information transmission in the case of continuous signals. *IRE Trans. Inf. Theory* 2, 102–108. doi: 10.1109/TIT.1956.1056823
- Li, D. A., Li, X. L., Liang, Z. H., Voss, L. J., and Sleight, J. W. (2010). Multiscale permutation entropy analysis of EEG recordings during sevoflurane anesthesia. *J. Neural Eng.* 7:046010. doi: 10.1088/1741-2560/7/4/046010
- Mandelbrot, B. B. (1982). *The Fractal Geometry of Nature*. San Francisco, CA: Freeman.
- Mizuno, T., Takahashi, T., Cho, R. Y., Kikuchi, M., Murata, T., Takahashi, K., et al. (2010). Assessment of EEG dynamical complexity in Alzheimer's disease using multiscale entropy. *Clin. Neurophysiol.* 121, 1438–1446. doi: 10.1016/j.clinph.2010.03.025
- Nardelli, M., Valenza, G., Cristea, I. A., Gentili, C., Cotet, C., David, D., et al. (2015). Characterizing psychological dimensions in non-pathological subjects through autonomic nervous system dynamics. *Front. Comput. Neurosci.* 9:37. doi: 10.3389/fncom.2015.00037
- Nikulin, V. V., and Brismar, T. (2004). Comment on “Multiscale entropy analysis of complex physiologic time series.” *Phys. Rev. Lett.* 92:089803. doi: 10.1103/PhysRevLett.92.089803



- Norris, P. R., Anderson, S. M., Jenkins, J. M., Williams, A. E., and Morris, J. A. Jr. (2008). Heart rate multiscale entropy at three hours predicts hospital mortality in 3,154 Trauma patients. *Shock* 30, 17–22. doi: 10.1097/SHK.0b013e318164e4d0
- Richman, J. S., and Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* 278, H2039–H2049. Available online at: <http://ajpheart.physiology.org/content/278/6/H2039>
- Wessel, N., Schirdewan, A., and Kurths, J. (2003). Intermittently decreased beat-to-beat variability in congestive heart failure *Phys. Rev. Lett.* 91:119801. doi: 10.1103/PhysRevLett.91.119801

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Gao, Hu, Liu and Cao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.