

# *Entropy measures for biological signal analyses*

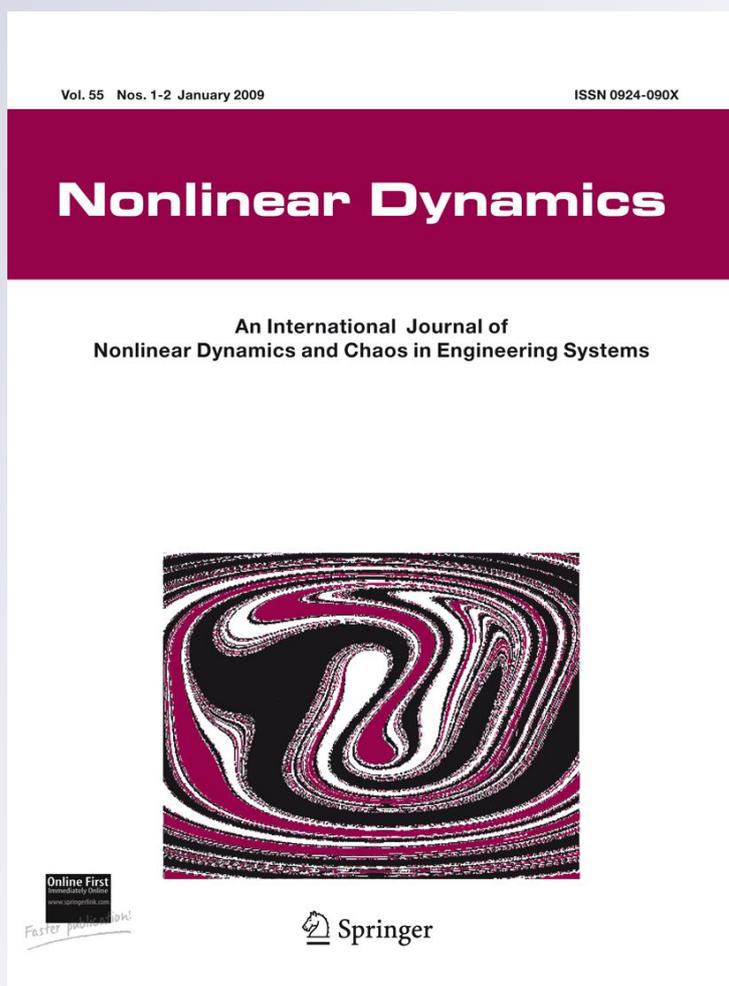
**Jianbo Gao, Jing Hu & Wen-wen Tung**

## **Nonlinear Dynamics**

An International Journal of Nonlinear  
Dynamics and Chaos in Engineering  
Systems

ISSN 0924-090X  
Volume 68  
Number 3

Nonlinear Dyn (2012) 68:431-444  
DOI 10.1007/s11071-011-0281-2



**Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media B.V.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.**

# Entropy measures for biological signal analyses

Jianbo Gao · Jing Hu · Wen-wen Tung

Received: 4 May 2011 / Accepted: 19 November 2011 / Published online: 10 December 2011  
© Springer Science+Business Media B.V. 2011

**Abstract** Entropies are among the most popular and promising complexity measures for biological signal analyses. Various types of entropy measures exist, including Shannon entropy, Kolmogorov entropy, approximate entropy (ApEn), sample entropy (SampEn), multiscale entropy (MSE), and so on. A fundamental question is which entropy should be chosen for a specific biological application. To solve this issue, we focus on scaling laws of different entropy measures and introduce an ensemble forecasting framework to find the connections among them. One critical component of the ensemble forecasting framework is the scale-dependent Lyapunov exponent (SDLE), whose scaling behavior is found to be the richest among all the entropy measures. In fact, SDLE contains all the essential information of other entropy measures, and can act

as a unifying multiscale complexity measure. Furthermore, SDLE has a unique scale separation property to aptly deal with nonstationarity and characterize high-dimensional and intermittent chaos. Therefore, SDLE can often be the first choice for exploratory studies in biology. The effectiveness of SDLE and the ensemble forecasting framework is illustrated by considering epileptic seizure detection from EEG.

**Keywords** Approximate entropy · Sample entropy · Scale-dependent Lyapunov exponent · Seizure detection

## 1 Introduction

Entropies are among the most popular complexity measures for analyzing biological signals. This is for a very good reason: Fundamentally speaking, entropies of various kinds characterize the rate of creation of information in dynamical systems, while biological systems, such as gene regulatory networks, protein interaction networks, cells, neurons, or organisms (e.g., heart, brain), all are complicated dynamical systems with uncertainties, irrespective of their relative size, and thus continuously generate entropies. Indeed, many researches using entropy for biological applications have shown tremendous promise. For example, recently, approximate entropy (ApEn) [1] has been used to assess postural instability following mild traumatic brain injury (MTBI) [2, 3], which is a sig-

---

J. Gao (✉)  
PMB Intelligence LLC, PO Box 2077, West Lafayette,  
IN 47996, USA  
e-mail: jbgao.pmb@gmail.com

J. Gao  
Mechanical and Materials Engineering, Wright State  
University, Dayton, OH 45435, USA

J. Hu  
Affymetrix, Inc., 3380 Central Expressway, Santa Clara,  
CA 95051, USA

W.-w. Tung  
Department of Earth & Atmospheric Sciences, Purdue  
University, West Lafayette, IN 47907, USA

nificant issue in health sciences, costing nearly \$60 billion annually in the United States alone [4–6]. Even though ApEn has a number of weaknesses compared with sample entropy (SampEn) [7], it is already considerably more effective than conventional methods for assessing postural instability [8]. One can readily see that other entropy measures could be even more effective for this purpose.

Besides ApEn and SampEn, there are many other entropies, including Shannon entropy [9], Kolmogorov entropy [10], correlation entropy [10], multiscale entropy (MSE) [11], permutation entropy [12], lempel-Ziv complexity [13, 14], and so on. For a researcher, especially a newcomer, targeting at a specific biological problem, a natural question would be which entropy measure should be chosen for the task. Moreover, in a clinical setting, often, a single number, called *score*, is overwhelmingly favored. When is this possible?

The above questions cannot be satisfactorily answered without a proper understanding of the entropy measures. To expedite progress of many ongoing exciting researches in biology using entropies, and stimulate new researches, in this semireview article, we examine a number of fundamental issues:

- (1) What are the connections among the different entropy measures?
- (2) How may entropies reliably be computed from noisy biological data?
- (3) How can analytic scaling laws for entropies be derived to guide experimental data analyses?

Our ultimate goal is to be able to choose the most appropriate form(s) of entropy to use in a specific application and aptly interpret analysis results.

The remainder of the paper is organized as follows. In Sect. 2, we briefly describe the EEG data to be analyzed here and the state-of-the-art of epileptic seizure detection. In Sect. 3, we introduce Shannon and Renyi entropies in physical space. In Sect. 4, we introduce entropies in phase space, and discuss scaling laws to better understand the difference between ApEn and SampEn. In Sect. 5, we explore the connections among various types of entropies through an ensemble forecasting point of view, and explain how the scale-dependent Lyapunov exponent (SDLE) [15, 16] brings a proper closure to the quantification of information flow in phase space. Finally in Sect. 6, we briefly summarize main results of the paper, discuss the distinc-

tion between entropy value and complexity level of a dynamical system, and present future perspectives.

## 2 Data

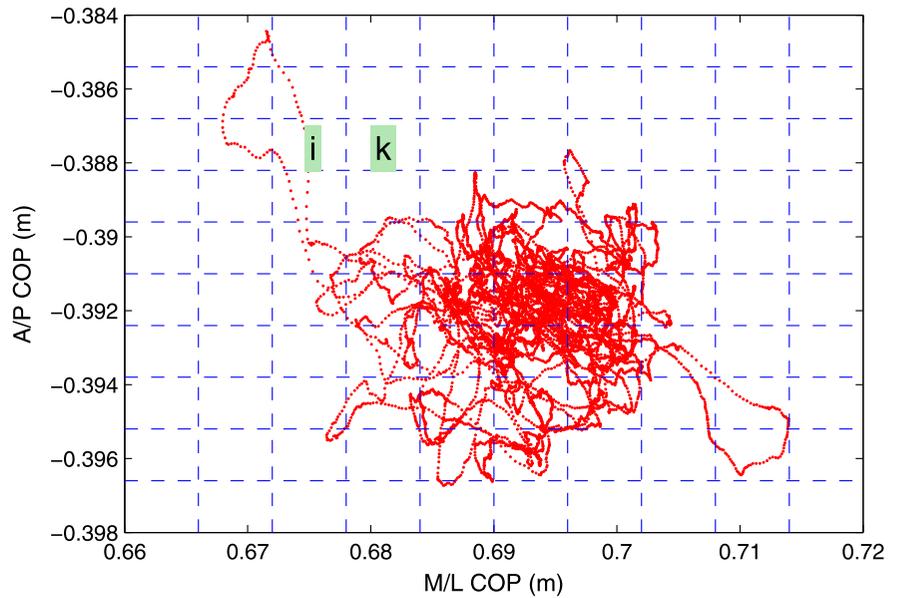
For ease of illustration, we shall analyze a database of EEG that can be downloaded at <http://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html>. The database consist of three groups, H (healthy), E (epileptic subjects during a seizure-free interval), and S (epileptic subjects during seizure); each group contains 100 data segments, whose length is 4097 data points with a sampling frequency of 173.61 Hz. The details of the data can be found at [17]. Note part of these data were analyzed in [18], which introduced a complicated framework of first decomposing EEG data into 5 subbands, then calculating linear and nonlinear measures from the subbands, including the largest Lyapunov exponent and correlation dimension, and finally using neural network or other machine learning based tools to classify which state an EEG signal belongs to. In [19], we have shown that various types of entropies and other complexity measures are similarly effective in forewarning epileptic seizures. The result of [18] thus suggests that the commonly used entropies measures, including ApEn and SampEn, when used alone, may not detect seizures as accurately as the sophisticated approached developed in [18]. Using multiscale approaches, however, the task can be readily handled. One effective multiscale approach is the adaptive fractal analysis introduced in [20], which is based on a versatile adaptive filter [21–23]. The approach of [18] but is much simpler. In Sect. 5, we shall develop another similarly simple and effective multiscale approach based on SDLE.

## 3 Shannon and Renyi entropies of biological trajectories

To facilitate discussions of various types of dynamical entropies that are defined in phase space, it is beneficial to first define Shannon and Renyi entropies in physical space. For illustration purpose, we examine a biological trajectory.

In a 2-D plane, the simplest biological trajectory is a straight line. The other extreme is the Drunkard's walk: Image that one is completely drunken, blindfolded, spun around a number of times, and let go.

**Fig. 1** COP trajectory for a subject on day 6 after concussion



The Drunkard's trajectory is the famous Brownian motion (Bm), which also characterizes the trajectory of a pollen particle in a solution. One can easily imagine that an arbitrary trajectory lies in between a Bm and a straight line.

A special type of biological trajectory is traced out by the Center of Pressure (COP), when one is standing still in a static balance test. Such a test is valuable for assessing postural instability following MTBI, which results from neuropathologic changes due to a direct or indirect blow to the head. An example of a COP trajectory is shown in Fig. 1, where A/P and M/L denote Anterior/Posterior and Medial/Lateral balance, respectively. For more details of the data, see [24].

To compute Shannon and Renyi entropies for a biological trajectory, assume that the trajectory has visited  $m$  unit areas, with the  $i$ -th unit area being visited by  $n_i$  times. This is schematically shown in Fig. 1. Note that an empty unit area, such as that denoted by  $k$  in Fig. 1 is irrelevant. Let the trajectory be  $N$  points long. Then the probability  $p_i$  that the  $i$ -th unit area being visited is  $n_i/N$ . The Shannon entropy is defined by

$$I = - \sum_{i=1}^m p_i \log p_i, \tag{1}$$

where the unit for  $I$  is a bit or baud corresponding to base 2 or  $e$  in the logarithm. Without loss of generality, we shall choose base  $e$ .

Renyi entropy is a generalization of Shannon entropy. It is defined by

$$I_q^R = \frac{1}{1-q} \log \left( \sum_{i=1}^m p_i^q \right). \tag{2}$$

$I_q^R$  has a number of interesting properties:

- When  $q = 1$ ,  $I_1^R$  is the Shannon entropy:  $I_1^R = I$ .
- $I_0^R = \log(m)$  is the topological entropy, which is just the logarithm of the area traced out by a sway trajectory.
- If  $p_1 = p_2 = \dots = p_m = \frac{1}{m}$ , then for all real valued  $q$ ,  $I_q^R = \log(m)$ .
- In the case of unequal probability,  $I_q^R$  is a non-increasing function of  $q$ . In particular, if we denote

$$p_{\max} = \max_{1 \leq i \leq m} (p_i), \quad p_{\min} = \min_{1 \leq i \leq m} (p_i),$$

then

$$\lim_{q \rightarrow -\infty} I_q^R = -\log(p_{\min}),$$

$$\lim_{q \rightarrow \infty} I_q^R = -\log(p_{\max}).$$

Albeit simple, Shannon and Renyi entropies have been found to be able to readily detect postural instability 1–2 weeks post-concussion, and thus are very promising for clinical applications. For more details, we refer to [24].

### 4 Dynamical entropies and information flow

In the previous section, we discussed Shannon and Renyi entropies in a physical space. Now we discuss dynamical entropies and information flow in a phase space. To set up a proper stage to consider information flow, a few words about the concept of phase space will be helpful. For simplicity, let us consider a system with two state variables,  $X_1$  and  $X_2$ . When monitoring the motion of the system in time, one can plot out the waveforms for  $X_1(t)$  and  $X_2(t)$ . Alternatively, one can monitor the trajectory defined by  $(X_1(t), X_2(t))$ , where the time,  $t$ , appears as an implicit parameter. The space spanned by  $X_1$  and  $X_2$  is called the phase space (or state space).

The dimension of the phase space is the number of degrees of freedom of the system. When the system is modeled by partial differential equations (PDEs) or stochastic differential equations, the dimension of the phase space is infinite. However, the dimension of the attractor of the dynamics could still be quite low, even if the phase space is infinite-dimensional. In fact, irrespective of the dimension of the phase space, a stable fixed point has dimension zero, a limit cycle has dimension one, and low-dimensional chaos, whose motion appears random because of exponential divergence in the phase space,  $dr_t \sim dr_0 e^{\lambda_1 t}$ , where  $dr_0$  is the initial small separation between two nearby orbits,  $dr_t$  is the distance between them at time  $t$ , and  $\lambda_1 > 0$  is the largest positive Lyapunov exponent, also has a small dimension.

In many biological applications, only available data may be just a scalar time series. In such a situation, one can reconstruct a phase space using the following time-delay embedding [25–27]:

$$V_i = [x(i), x(i + L), \dots, x(i + (m - 1)L)], \quad (3)$$

where  $\{x(i), i = 1, 2, \dots\}$  is the time series,  $m$  and  $L$  are the embedding dimension and the delay time, respectively. For a stochastic process, which is infinite-dimensional, the primary role of embedding is to transform a self-affine stochastic process (e.g., the units of the two axes are different) to a self-similar process in a phase space. Therefore, specific values of  $m$  and  $L$  are not important, so long as  $m > 1$ . In fact, often,  $m = 2$  may best resolve the scaling behavior from a finite time series. For a chaotic time series, the issue of embedding becomes more challenging—both  $m$  and  $L$  have to be chosen optimally. While there

are many methods to determine  $L$  (see, for example, Sect. 13.1.3 of [15]), basically there are only two kinds of methods to systematically determine both  $m$  and  $L$ . One type is geometrical, called the false nearest neighbor method [28, 29]. The other is dynamical, based on a quantification of the smoothness of the reconstructed dynamics [30, 31]. The values of  $m$  and  $L$  given by the two algorithms are quite consistent.

In the following, we assume a proper phase space has been set up based on available data.

#### 4.1 General considerations on information flow in a phase space

Extending our discussions in Sect. 2 to calculate the entropy of the motion of a dynamical system, we partition the phase space into small boxes of size  $\varepsilon$ , compute the probability  $p_i$  that box  $i$  is visited by the trajectory, and finally calculate Shannon entropy using (1). For many systems, when  $\varepsilon \rightarrow 0$ , information linearly increases with time [32]:

$$I(\varepsilon, t) = I_0 + Kt, \quad (4)$$

where  $I_0$  is the initial entropy and  $K$  is the *Kolmogorov–Sinai (KS) entropy* (to be more precisely defined shortly). Suppose that the system is initially in a particular region of the phase space, and that all initial probabilities are zero except the corresponding probability for that region, which is 1. Therefore,  $I_0 = 0$ .

We now consider three cases of dynamical systems (i) deterministic, nonchaotic, (ii) deterministic, chaotic, and (iii) random. For case (i), during the time evolution of the system, phase trajectories remain close together. After a time  $T$ , nearby phase points are still close to each other, and can be grouped into some other small region of the phase space. Therefore, there is no change in information. For case (ii), due to exponential divergence, the number of phase space region available to the system after a time  $T$  is  $N \propto e^{(\sum \lambda^+)T}$ , where  $\lambda^+$  are positive Lyapunov exponents. Assuming that all of these regions are equally likely, then  $p_i(T) \sim 1/N$ , and the information function becomes

$$I(T) = - \sum_{i=1}^N p_i(T) \ln p_i(T) = \left( \sum \lambda^+ \right) T. \quad (5)$$

Therefore,  $K = \sum \lambda^+$ . More generally, if these phase space regions are not visited with equal probability,

then

$$K \leq \sum \lambda^+. \tag{6}$$

Grassberger and Procaccia [10], however, suggest that equality usually holds. Finally, for case (iii), we can easily envision that after a short time, the entire phase space may be visited. Therefore,  $I \sim \ln N$ . When  $N \rightarrow \infty$ , we have  $K = \infty$ .

### 4.2 Basic definitions of entropies

Consider a dynamical system with  $F$  degrees of freedom. 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  $i_d$ . We now can use (1) to compute the block entropy,

$$H_d(\varepsilon, \tau) = - \sum_{i_1, \dots, i_d} p(i_1, \dots, i_d) \ln p(i_1, \dots, i_d). \tag{7}$$

It is on the order of  $d\tau K$ ; then we take the difference between  $H_{d+1}(\varepsilon, \tau)$  and  $H_d(\varepsilon, \tau)$ , and normalize it by  $\tau$ ,

$$h_d(\varepsilon, \tau) = \frac{1}{\tau} [H_{d+1}(\varepsilon, \tau) - H_d(\varepsilon, \tau)]. \tag{8}$$

Let

$$h(\varepsilon, \tau) = \lim_{d \rightarrow \infty} h_d(\varepsilon, \tau). \tag{9}$$

It is called the  $(\varepsilon, \tau)$ -entropy [33]. When  $\tau$  is fixed, such as taken as the sampling time, it may be written as  $h(\varepsilon)$ . Taking proper limits in (9), we obtain the Kolmogorov–Sinai entropy,

$$K = \lim_{\tau \rightarrow 0} \lim_{\varepsilon \rightarrow 0} h(\varepsilon, \tau) = \lim_{\tau \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \lim_{d \rightarrow \infty} \frac{1}{\tau} [H_{d+1}(\varepsilon, \tau) - H_d(\varepsilon, \tau)] \tag{10}$$

$$= - \lim_{\tau \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \lim_{d \rightarrow \infty} \frac{1}{d\tau} \times \sum_{i_1, \dots, i_d} p(i_1, \dots, i_d) \ln p(i_1, \dots, i_d). \tag{11}$$

The KS entropy can be extended to order- $q$  Renyi entropies:

$$K_q = - \lim_{\tau \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \lim_{d \rightarrow \infty} \frac{1}{d\tau} \frac{1}{q-1} \times \ln \sum_{i_1, \dots, i_d} p^q(i_1, \dots, i_d) \tag{12}$$

when  $q \rightarrow 1$ ,  $K_q \rightarrow K$ . In the case of unequal probabilities, Renyi entropy of order- $q$  is a non-increasing function of  $q$ .

### 4.3 Numerical computations

There are three interesting ways to calculate the KS entropy from a time series. One is to first estimate all the positive Lyapunov exponents, and then use the summation of these exponents to estimate the KS entropy. Alternatively, one can estimate  $K$  by approximating the probabilities  $p(i_1, \dots, i_m)$ , as proposed by Cohen and Procaccia [34] and Eckmann and Ruelle [35]. Let the length of a time series be  $N$ , and let the  $m$ -dimensional embedding vectors based on (3) be explicitly denoted by  $V_i^{(m)}$ . For ease of presentation, assume the delay time to be 1. Let  $n_i^{(m)}(\varepsilon)$  be the number of vectors  $V_j^{(m)}$  satisfying  $\|V_j^{(m)} - V_i^{(m)}\| \leq \varepsilon$ . Cohen and Procaccia [34] noted that

$$C_i^{(m)}(\varepsilon) = n_i^{(m)} / (N - m + 1)$$

approximates the probability  $p(i_1, \dots, i_m)$  for boxes of size  $2\varepsilon$ . They then proposed to estimate  $H_m(\varepsilon)$  by

$$H_m(\varepsilon) = - \frac{1}{N - m + 1} \sum_i \ln C_i^{(m)}. \tag{13}$$

Then

$$h(\varepsilon, \tau) = \lim_{m \rightarrow \infty} \frac{1}{\delta t} [H_{m+1}(\varepsilon) - H_m(\varepsilon)], \tag{14}$$

where  $\delta t$  is the sampling time, and

$$K = \lim_{\tau \rightarrow 0} \lim_{\varepsilon \rightarrow 0} h(\varepsilon, \tau). \tag{15}$$

Note that with the delay time  $L = 1$ , when the maximum norm,

$$\|V_j^{(m)} - V_i^{(m)}\| = \max_{0 \leq k \leq m-1} |x(i+k) - x(j+k)|,$$

is used,  $H_{m+1}(\varepsilon) - H_m(\varepsilon)$  is the logarithm of the conditional probability

$$|x(i + m) - x(j + m)| \leq \varepsilon,$$

given that

$$|x(i + k) - x(j + k)| \leq \varepsilon, \quad k = 0, 1, \dots, m - 1$$

averaged over  $i$ . Therefore, the maximum norm may simplify numerical computations, especially when data are random without correlations. Also note that when  $h(\varepsilon, \tau)$  is evaluated at a fixed finite scale  $\hat{\varepsilon}$ , the resulting entropy is called the approximate entropy ApEn [1].

The third method is to approximate  $K$  by estimating the correlation entropy  $K_2$  (i.e., Renyi entropy of order 2) using the Grassberger–Procaccia’s algorithm [10]:

$$K_2(\varepsilon) = \lim_{m \rightarrow \infty} \frac{\ln C^{(m)}(\varepsilon) - \ln C^{(m+1)}(\varepsilon)}{L\delta t}, \quad (16)$$

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

$$C^{(m)}(\varepsilon) = \lim_{N_v \rightarrow \infty} \frac{2}{N_v(N_v - 1)} \times \sum_{i=1}^{N_v-1} \sum_{j=i+1}^{N_v} H(\varepsilon - \|V_i - V_j\|), \quad (17)$$

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  $y < 0$ ).  $C^{(m+1)}(\varepsilon)$  can be computed similarly based on the  $m + 1$ -dimensional reconstructed vectors. When we evaluate  $K_2(\varepsilon)$  at a finite fixed scale  $\hat{\varepsilon}$ , we obtain the sample entropy SampEn [7].

Three comments are in order: (1) KS entropy is the limit of  $(\varepsilon, \tau)$ -entropy or  $h(\varepsilon)$ ; (2) both  $h(\varepsilon)$  and the correlation entropy  $K_2(\varepsilon)$  are functions of the scale parameter  $\varepsilon$ ; (3) ApEn and SampEn, being  $h(\varepsilon)$  and  $K_2(\varepsilon)$  evaluated at a fixed  $\varepsilon$  (such as 15% or 20% of the standard deviation (STD) of the data), are special cases. To avoid using too many different terms, in the following, we may also denote  $h(\varepsilon)$  by ApEn( $\varepsilon$ ), and  $K_2(\varepsilon)$  by SampEn( $\varepsilon$ )—explicit dependence of ApEn and SampEn on  $\varepsilon$  emphasizes that they are functionals

of the scale parameter  $\varepsilon$  rather than a single number—the latter may be denoted as ApEn(15% STD), SampEn(15% STD), etc.

To better appreciate the scalings for ApEn and SampEn, it is helpful to consider a sequence of independent identically distributed (IID) random variables (RVs). When one uses maximum norm, in the case of SampEn,  $C^{(m)}(\varepsilon)$  amounts to the probability that  $|x_{i+k} - x_{j+k}| \leq \varepsilon, k = 1, \dots, m$ . Let  $z = |u - v|$ , where  $u, v$  are IID RVs having the same distribution as  $x$ , we can readily obtain the probability density function  $p(z)$  for  $z$ . Then

$$C^{(m)}(\varepsilon) = \left[ \int_0^\varepsilon p(z) dz \right]^m \quad (18)$$

and

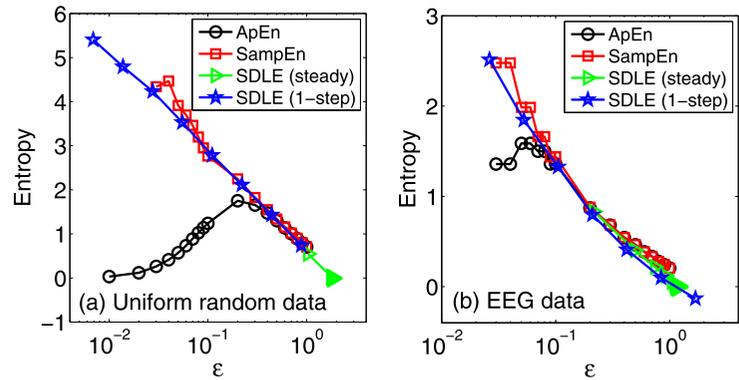
$$\begin{aligned} K_2(\varepsilon) &= \ln C^{(m)}(\varepsilon) - \ln C^{(m+1)}(\varepsilon) \\ &= -\ln \left[ \int_0^\varepsilon p(z) dz \right]. \end{aligned} \quad (19)$$

In the case of IID RVs uniformly distributed in the unit interval,  $p(z) = 2(1 - z), 0 \leq z \leq 1$ . Then

$$K_2(\varepsilon) \sim -\ln \varepsilon. \quad (20)$$

It is clear that the same scaling should also characterize ApEn. Therefore, theoretically, ApEn( $\varepsilon$ ) and SampEn( $\varepsilon$ ) are equivalent. The major difference between them is in the computations. An example is shown in Fig. 2(a), where we observe that the scaling for ApEn( $\varepsilon$ ) is much shorter than that for SampEn( $\varepsilon$ ). Almost identical results can be obtained with Gaussian white noise, and similar inferiority of ApEn( $\varepsilon$ ) to SampEn( $\varepsilon$ ) is also true for EEG data, as shown in Fig. 2(b). Therefore, ApEn( $\varepsilon$ ) has to be considered worse than SampEn( $\varepsilon$ ). It is thus fortunate that MSE, a popular method for coping with the complexities of biological data, is based on SampEn, not ApEn. MSE is SampEn computed from the original and the filtered data whose length is  $1/b_s$  that of the original data, where  $b_s$  is the parameter for nonoverlapping moving average. In [36], we have derived, for time series with long memory, a fundamental biscaling law of MSE, one for the scale  $\varepsilon$  in phase space, the other for the block size  $b_s$  used for smoothing. The biscaling law of MSE indicates that so far as the fractal property is concerned, direct fractal analysis would be more advantageous than computing MSE.

**Fig. 2** ApEn, SampEn, and SDLE for (a) 1000 uniform RVs and (b) a normal EEG signal. The scale  $\varepsilon$  is normalized by the standard deviation of the original data



**5 Understanding information flow through ensemble forecasting**

5.1 General considerations

Ensemble forecasting has been commonly employed by the major operational weather prediction facilities worldwide. Its most basic form consists of conducting multiple numerical predictions using slightly different initial conditions that are all plausible given the past and current set of observations, or measurements. When a dynamical system is chaotic, the information contained in the initial conditions will be rapidly lost as the system evolves. An example is shown in Fig. 3 for the Lorenz '63 system [37]:

$$\begin{aligned}
 dx/dt &= -10(x - y), \\
 dy/dt &= -xz + 28x - y, \\
 dz/dt &= xy - \frac{8}{3}z,
 \end{aligned}
 \tag{21}$$

where the initial ensembles, consisting of 2,500 normally distributed members, are indicated as pink. As time goes by, this tight cluster quickly spreads, yielding the red, green, and blue regions at  $t = 2, 4,$  and  $6,$  respectively.

The ensemble forecasting point of view can help deepen our understanding of various types of entropies. For this purpose, we first note that in ergodic systems, the operation  $d \rightarrow \infty$  in (9) may be replaced by ensemble average. This means that in general, the rate of creation of new information, or the temporal derivative of Shannon entropy  $I$ , can be described by [38]

$$dI(\varepsilon_t)/dt = h(\varepsilon_t),
 \tag{22}$$

where  $\varepsilon$  is written as  $\varepsilon_t$  to emphasize that in the case of finite scales, the size of the ensembles is a function of time. Alternatively, we have

$$I(\varepsilon_t) = I_0 + \int_0^t h(\varepsilon_t) dt.
 \tag{23}$$

Another interesting way to quantify information loss in ensemble forecasting is by *relative entropy* [39–41]. Let  $\{p_i, i = 1, 2, \dots\}$  be the probabilities associated with the forecast ensembles in the  $i$ th box and  $\{q_i\}$  be the invariant measure or the equilibrium distribution. The discrepancy between them can be quantified by the relative entropy defined by

$$R = \sum_i p_i \ln\left(\frac{p_i}{q_i}\right),
 \tag{24}$$

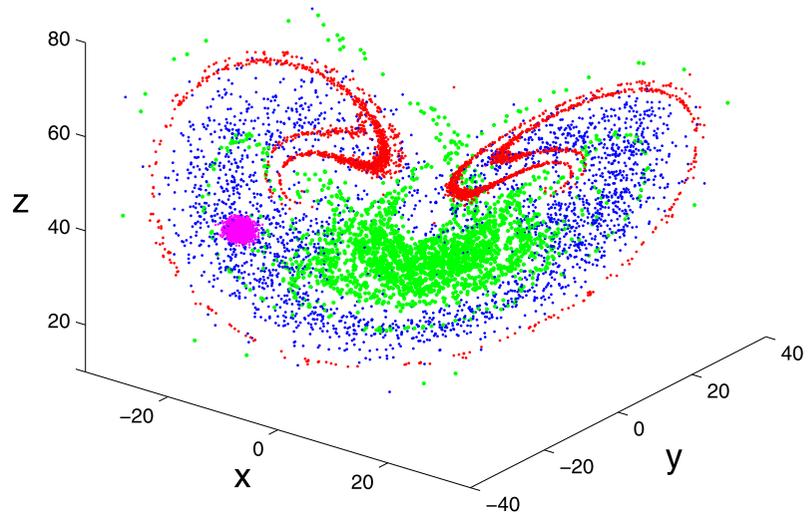
which is always nonnegative and is zero if and only if the two distributions are the same.

The major result we can prove for the relative entropy is the following equation:

$$R(t) = I_{\max} - \int_0^t h(\varepsilon_t) dt
 \tag{25}$$

where  $I_{\max}$  is the maximal prior information available to the system with a partition using boxes of size  $\varepsilon_0$ . The proof, while being tremendously helpful for understanding, may be difficult for some readers on a first read. Therefore, it is left to the [Appendix](#). Since  $\int_0^t h(\varepsilon_t) dt$  is the major term that appears in both (23) and (25), we see that the Shannon entropy and the relative entropy approaches are essentially equivalent. More importantly, we observe that in order to compute  $I(t)$  or  $R(t)$ , we need to know the scaling law of  $h(\varepsilon_t)$ , as well as the temporal evolution of  $\varepsilon$ . To bring

**Fig. 3** (Color online) Ensemble forecasting in the chaotic Lorenz system: 2,500 ensemble members, initially represented by the pink color, evolve to those represented by the red, green, and blue colors at  $t = 2, 4,$  and  $6$  units



a satisfactory closure to this issue, we need the concept of scale-dependent Lyapunov exponent (SDLE) which we discuss in the next subsection.

### 5.2 Quantifying ensemble forecasting using SDLE

SDLE is defined in a phase space through consideration of an ensemble of trajectories. Denote the initial separation between two nearby trajectories by  $\varepsilon_0$ , and their average separation at time  $t$  and  $t + \Delta t$  by  $\varepsilon_t$  and  $\varepsilon_{t+\Delta t}$ , respectively. The SDLE  $\lambda(\varepsilon_t)$  is defined through equations [15, 16]

$$\varepsilon_{t+\Delta t} = \varepsilon_t e^{\lambda(\varepsilon_t)\Delta t}, \quad \text{or} \quad \lambda(\varepsilon_t) = \frac{\ln \varepsilon_{t+\Delta t} - \ln \varepsilon_t}{\Delta t}. \tag{26}$$

Equivalently, we have a differential equation for  $\varepsilon_t$ ,

$$\frac{d\varepsilon_t}{dt} = \lambda(\varepsilon_t)\varepsilon_t, \quad \text{or} \quad \frac{d \ln \varepsilon_t}{dt} = \lambda(\varepsilon_t). \tag{27}$$

To compute SDLE, we can start from an arbitrary number of shells,

$$\varepsilon_k \leq \|V_i - V_j\| \leq \varepsilon_k + \Delta\varepsilon_k, \quad k = 1, 2, 3, \dots, \tag{28}$$

where  $V_i, V_j$  are reconstructed vectors,  $\varepsilon_k$  (the radius of the shell) and  $\Delta\varepsilon_k$  (the width of the shell) are arbitrarily chosen small distances ( $\Delta\varepsilon_k$  is not necessarily a constant). Then we monitor the evolution of all pairs of

points  $(V_i, V_j)$  within a shell and take average. Equation (26) can now be written as

$$\lambda(\varepsilon_t) = \frac{\langle \ln \|V_{i+t+\Delta t} - V_{j+t+\Delta t}\| - \ln \|V_{i+t} - V_{j+t}\| \rangle}{\Delta t}, \tag{29}$$

where  $t$  and  $\Delta t$  are integers in unit of the sampling time, and the angle brackets denote average within a shell. For chaotic systems, a simple condition,

$$|i - j| > W, \tag{30}$$

where  $W$  is a threshold value on the order of the embedding window size,  $(m - 1)L$ , needs to be imposed to ensure that vectors  $V_j$  are along the most unstable direction, instead of the tangential direction, of  $V_i$ . However, such a condition is not needed when the dynamics of a biological system are largely random. In fact, in those situations, it is useful to consider two types of SDLE, which may be called 1-step SDLE and steady-state SDLE. The 1-step SDLE is simply the SDLE computed with  $t = 1$  for the series of shells. The steady-state SDLE are the SDLE curves properly converged—while inequality (30) is often sufficient, when data is very random,  $t > 1$  may already be adequate.

SDLE has distinctive scaling laws for different types of time series. We list three here:

- (1) For clean chaos on small scales, and noisy chaos with weak noise on intermediate scales,

$$\lambda(\varepsilon) = \lambda_1, \tag{31}$$

where  $\lambda_1$  is the largest positive Lyapunov exponent.

- (2) For clean chaos on large scales where memory has been lost and for noisy chaos (including noise-induced chaos [42–44]) on small scales,

$$\lambda(\varepsilon) \sim -\gamma \ln \varepsilon, \tag{32}$$

where  $\gamma > 0$  is a parameter, controlling the speed of loss of information.

- (3) For random  $1/f^{2H+1}$  processes, where  $0 < H < 1$  is called the Hurst parameter which characterizes the correlation structure of the process: depending on whether  $H$  is smaller than, equal to, or larger than  $1/2$ , the process is said to have anti-persistent, short-range, or persistent long-range correlations [15, 45],

$$\lambda(\varepsilon) \sim \varepsilon^{-1/H}. \tag{33}$$

To better understand the above scaling laws, let us derive (32) for IID RVs. Using the maximum norm, it is easy to see that if the initial separation of two nearby trajectories is  $\varepsilon$ , at the next time step, the separation will be close to the mean of the random variable  $Z$  used for deriving (20). Therefore, (32) holds with  $\gamma = 1$ . The 1-step and steady-state SDLE curves for the uniform RVs and EEG data of Fig. 2 are also shown there. We observe that for noise, the scaling range for the steady-state SDLE is much shorter than that of the 1-step SDLE and that the value of the steady-state SDLE is close to zero. We also note that the 1-step SDLE contains all the essential information of SampEn—in fact, its scaling behavior is even better than SampEn, since the SDLE curve is smoother than that for SampEn.

### 5.3 Integral form of SDLE for characterizing biosignals

Now that we have found the connection among different entropy measures, we can safely conclude that the more suitable entropy forms to use in practice are  $K_2(\varepsilon)$  and SDLE. The latter has a few additional features that make it more advantageous than other entropy measures. The most important of such features is *scale-separation*: the regular (e.g., periodic or quasiperiodic) aspects of the dynamics only contribute to the structure of SDLE around the characteristic scales  $\varepsilon_\infty$ , which are the scales where SDLE is

close to zero, therefore, the major scalings of (31)–(33) will largely be intact if a physiological data contain a significant rhythmic component—in fact, this is the very mechanism that SDLE can readily deal with nonstationarity and detect intermittent chaos [15, 46–49]. Because of scale-separation, one can often readily extract a few important parameters from SDLE to solve a practical problem. This is not feasible with other entropy measures.

To extract useful parameters to characterize a complex signal, it is often convenient to use an integral form of SDLE:

$$\ln \varepsilon_t = \ln \varepsilon_0 + \int_0^t \lambda(\varepsilon_t) dt. \tag{34}$$

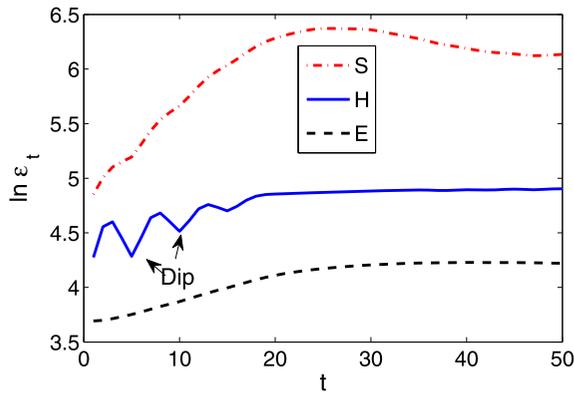
Clearly, (34) parallels (23) and (25). More importantly, (34) brings a closure to information flow characterized by (23) and (25)–(34) gives the critical information on how the scale  $\varepsilon$  evolves with time. In fact, for random data or chaotic systems with only one positive Lyapunov exponent,  $\ln \varepsilon_t$  plays the role of the information content  $I$  described by (23)—this can be better appreciated if one reexamines Fig. 2(a).

To better appreciate the significance of (34), let us demonstrate how dynamical features derived from SDLE can facilitate accurate detection of epileptic seizures from EEG, which was briefly described in Sect. 2. For ease of comparison with the results of [18], we shall work on the same data sets.

Figure 4 shows three  $\ln \varepsilon_t$  vs.  $t$  curves, belonging to the groups H, E, and S, respectively. Noting that  $\ln \varepsilon_\infty$  plays the role of  $I_{\max}$ —the larger it is, the more information an EEG signal has, we can conclude that seizure EEG has the richest information.

For the purpose of detecting epileptic seizures, besides  $\ln \varepsilon_\infty$ , many other features can be derived from SDLE. We discuss two more here. One is the slope of  $\ln \varepsilon_t$  in the increasing part of the curve. Noting that local slopes of the curve is the very SDLE, therefore, this is the SDLE averaged across the scale. For convenience, we may choose the part of the curve from  $t = 1$ –20. Re-examining Fig. 2(b), we see that this slope may be interpreted as the approximate entropy (or sample entropy) averaged over a wide range of scales (i.e., y-axis of the curve) where dynamical evolution is vigorous. This slope may be denoted as  $SDLE_s$ .

The other feature we wish to explore is the parameter “Dip” schematically indicated in Fig. 4. It occurs



**Fig. 4** Typical  $\ln \varepsilon_t$  vs.  $t$  curves for three types of EEG data:  $H$  denotes EEG segment from a healthy subject,  $E$  denotes EEG segment from an epileptic subject during a seizure-free interval, while  $S$  denotes EEG segment from an epileptic subject during seizure

at multiples of the delay time for the embedding, i.e., at  $t = L, 2L, \dots$  (especially at  $t = L$ ) [50] and can be used to estimate measurement noise [51]. It can be estimated as

$$\text{Dip} = \frac{1}{2}(\ln \varepsilon_{L-2} + \ln \varepsilon_{L+2}) - \ln \varepsilon_L. \tag{35}$$

The reason we have chosen  $\varepsilon_{L\pm 2}$  instead of  $\varepsilon_{L\pm 1}$  is that  $\varepsilon_t$  may still be dipped at  $t = L \pm 1$ . In the following discussion, we have chosen  $L = 5$  for the embedding. It turns out the results presented below do not depend much on the specific choice of  $L$ , so far as  $L \neq 1$  (when  $L = 2$ , (35) may be modified accordingly). It is clear that this is also a dynamical quantity. Note that although Dip is very small for S and E in Fig. 4, for some data sets of S and E, Dip could be large, if those data were contaminated by noise.

How well can the parameters  $\ln \varepsilon_\infty$  and Dip separate the three EEG groups? It is shown in Fig. 5(a). We observe that the three groups almost completely separate. This means that with just these two parameters, the classification accuracy of these three groups can already be very high. For example, if one uses the dashed straight line for separating group S from groups (E, H), then the positive detection rate is 98%, while false detection rate is 0.5%, since there are only two red stars and one green square to the left and right of the line, respectively. If one slightly pushes the line to the left, then detection accuracy can be 100%, while false detection rate only increases slightly. The separation of the groups E and H is also very good, except

that now a linear classification scheme is not sufficient. Obviously, a neural network type tool can play a significant role here.

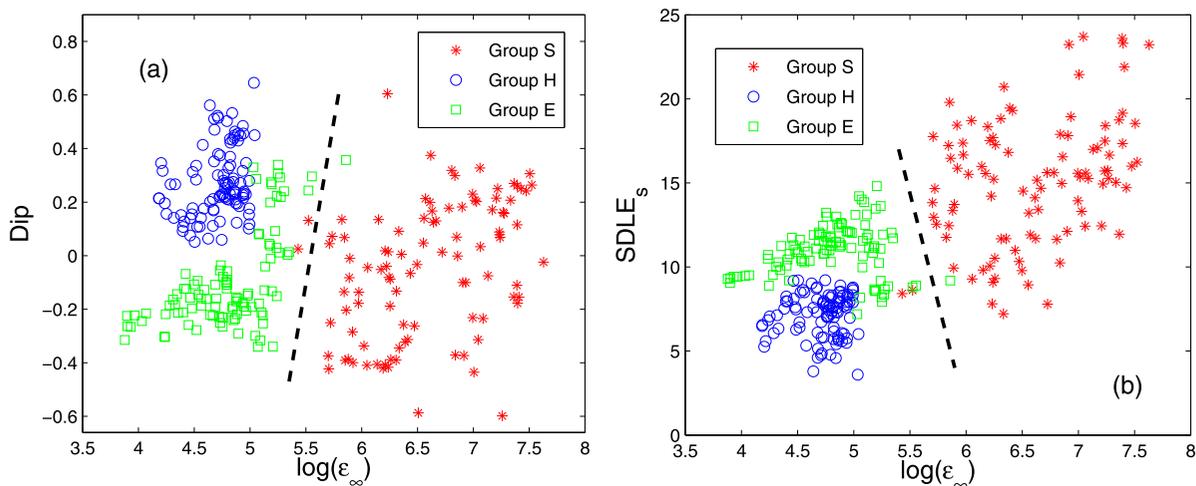
It turns out  $\ln \varepsilon_\infty$  and  $\text{SDLE}_s$  can be similarly effective in classifying the three groups, as shown in Fig. 5(b). More interestingly, we observe that the circles for the group H are basically above the squares for the group E in Fig. 5(a), but are below the squares in Fig. 5(b). This means  $\text{SDLE}_s$  and Dip are fully complement for the purpose of separating the groups H and E. Therefore, if we use all these three parameters, the classification will be more robust and has even higher accuracy.

### 6 Concluding discussions

Entropy measures have played, and will be playing leading roles in the analyses of biological signals. To expedite progress of many ongoing exciting researches in biology using entropies and stimulate new ones, in this paper, we have carefully examined the connections among the major dynamical entropy measures. We have found that  $h(\varepsilon, \tau)$  and  $K_2(\varepsilon)$  are fundamental: when  $\varepsilon$  is fixed, they are reduced to ApEn and SampEn, respectively; when proper limit is taken (i.e., both  $\varepsilon$  and  $\tau \rightarrow 0$ ), they yield the KS entropy. More importantly, by introducing an ensemble forecasting framework and carefully considering the scaling behavior of the entropies, we have found that various dynamical entropies are connected to the information content of the system through integral relations.

We can now answer all the motivating questions raised in Sect. 1:

- The scaling behavior of  $\text{ApEn}(\varepsilon)$  is worse than that of  $\text{SampEn}(\varepsilon)$ , as shown in Fig. 2, therefore,  $\text{SampEn}(\varepsilon)$  should always be chosen over  $\text{ApEn}(\varepsilon)$ .
- The computation of SDLE is simpler than that of  $\text{SampEn}(\varepsilon)$  or  $\text{ApEn}(\varepsilon)$ , since only one embedding is needed for computing SDLE, but two (i.e., with embedding dimension  $m$  and  $m + 1$ ) are needed for computing  $\text{SampEn}(\varepsilon)$  or  $\text{ApEn}(\varepsilon)$ . Therefore, when data length is limited, SDLE would be preferred.
- While computationally the simplest, the scaling laws for SDLE are the richest. In fact, commonly used complexity measures, including various types of entropies, can be obtained from SDLE by choosing specific  $\varepsilon$  [19]. More importantly, due to the

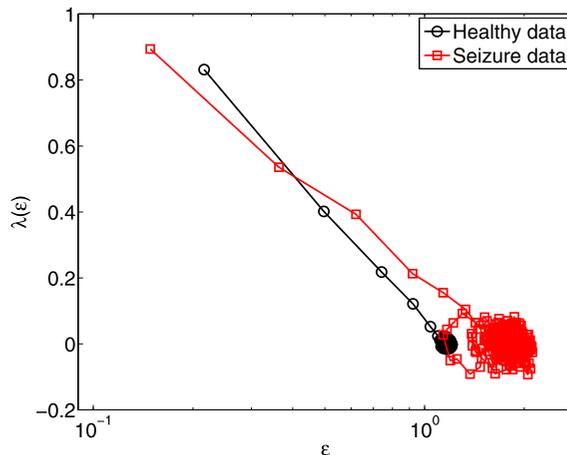


**Fig. 5** Epileptic seizure detection using different features, (a)  $\ln \epsilon_\infty$  and the “Dip” parameter, (b)  $\ln \epsilon_\infty$  and the slope parameter  $SDLE_s$

unique scale-separation property, SDLE can aptly deal with nonstationarity and characterize high-dimensional as well as intermittent chaos [15, 46–49]. Therefore, SDLE can often be the first choice in exploratory studies.

At this point, we would like to emphasize that there exist entropy measures that do not depend on the scale parameter. Such measures include permutation entropy [12] and lempel-Ziv complexity [13, 14]. Interestingly, they amount to steady-state SDLE evaluated on small scales [19]. Also, note that the biscaling law for MSE of fractal time series derived in [36] indicates that whenever the Hurst parameter  $H$  is of key importance, direct fractal analysis would be more advantageous than computing MSE.

Finally, let us resolve whether an entropy value can be equated to the complexity level of a biological system. We show here that a simple positive answer is not always obtainable. An example is shown in Fig. 6 for the steady-state SDLE curves for a normal and a seizure EEG data set. We observe that the two curves cross at a specific scale  $\epsilon^*$ —when  $\epsilon < \epsilon^*$ , the SDLE for the seizure EEG is smaller than that of normal EEG; when  $\epsilon > \epsilon^*$ , the opposite is true. The equivalence of SDLE and SampEn (see Fig. 2(b)) makes it clear that such a crossover behavior also exists if SampEn is used instead of SDLE. Can we say that normal EEG is more complex than seizure EEG, or vice versa? Clearly, there is no simple answer to such a question.



**Fig. 6** Steady-state SDLE curves for normal and seizure EEG data. The scale  $\epsilon$  is normalized by the standard deviation of the corresponding EEG data

In fact, without explicitly referring to the scale parameter, the issue is meaningless.

Most entropy metrics discussed here measure the randomness of the data, and thus are called deterministic complexity measures [15]. Is it possible to construct a structural complexity measure, which is maximized for neither high nor low randomness? The answer is positive. It is the steady-state SDLE. As shown in Fig. 2(a), the value of the steady-state SDLE and its scaling range are both close to zero. This means IID RVs, or more general, purely random data have

very low SDLE. For regular data, SDLE will always be close to zero. Therefore, the steady-state SDLE will be minimized for both purely random and regular data. This is the very essential ingredient of structural complexity. This issue will be an interesting future direction to pursue.

**Appendix: Evolution of relative entropy**

To prove (25), let us focus on the evolution of one particular ensemble occupying a space of size  $\varepsilon_0$  that is around one of the chosen initial conditions,  $\vec{x}(i)$ . Denote it by Ensemble( $i$ ). We can partition the phase space with boxes of size  $\varepsilon_0$  to include Ensemble( $i$ ) in one single box  $i$ . With this setting,  $p_k = 1$  only when  $k = i$  and zero elsewhere. Therefore, initially, for Ensemble( $i$ ), the relative entropy is

$$R^{(i)}(t = 0) = -\ln q_i, \tag{36}$$

where for clarity, the superscript ( $i$ ) in the relative entropy is used to indicate that this is for Ensemble( $i$ ). At time  $\tau$ , Ensemble( $i$ ) will have evolved into a different region in phase space, with some directions expanded and some others shrunk. If the dynamics are complicated, it will then occupy a number of boxes that are used to partition the phase space. For ease of presentation, we consider two cases of  $\varepsilon_0$ : infinitesimal and finite. Let us first consider the case of infinitesimal  $\varepsilon_0$ . For chaotic systems, assume that the dynamics of the system around  $\vec{x}(i)$  have  $J$  positive local Lyapunov exponents,  $\lambda_1^{(i)} \geq \lambda_2^{(i)} \geq \dots \geq \lambda_J^{(i)} > 0$ , where the superscript  $i$  is used to indicate that these Lyapunov exponents are local Lyapunov exponents near  $\vec{x}(i)$ . Edges along the  $J$  Lyapunov directions should have expanded. The size of the ensemble now becomes  $\varepsilon_\tau = \varepsilon_0 e^{\lambda_1^{(i)} \tau}$ , where  $\lambda_1^{(i)}$  is the largest positive local Lyapunov exponent, and the number of phase space boxes  $N$  available to Ensemble( $i$ ) becomes  $N \propto e^{(\sum_j \lambda_j^{(i)}) \tau}$  or equivalently,  $e^{K^{(i)} \tau}$ , where  $K^{(i)}$  is the summation of the  $J$  local positive Lyapunov exponents and may be called local KS entropy. In the case that the invariant measure is smooth on its support, then so long as  $\varepsilon_\tau$  is small, we may assume that the phase space regions available to Ensemble( $i$ ) at time  $\tau$  are equally likely. Indexing the boxes covering these

regions by  $l = 1, 2, \dots$ , then the relative entropy becomes

$$R^{(i)}(t = \tau) = -\sum_l \ln q_{il}^{(\tau)} - K^{(i)} \tau, \tag{37}$$

where we have used  $q_{il}^{(\tau)}$  to denote the invariant measure in phase space regions available to Ensemble( $i$ ) at time  $\tau$ .

Next, we consider the case of finite  $\varepsilon_0$ . Following the argument leading to (22), we readily see that (37) should be replaced by

$$R^{(i)}(t = \tau) = -\sum_l \ln q_{il}^{(\tau)} - h(\varepsilon_\tau)^{(i)} \tau, \tag{38}$$

where  $h(\varepsilon_\tau)^{(i)} = h(\varepsilon_\tau, \tau)^{(i)}$  is identified as the local  $(\varepsilon, \tau)$ -entropy pertinent to Ensemble( $i$ ) at time  $\tau$ .

We now consider a collection of ensembles. Loss of information in each ensemble may be described by (36)–(38). Averaging over all the ensembles, and assuming that initial conditions represent the invariant measure so that both  $\langle -\ln q_i \rangle$  and  $\langle -\sum_l \ln q_{il}^{(\tau)} \rangle$  are equal to the maximal prior information,  $I_{\max}$ , available to the system with a partition using boxes of size  $\varepsilon_0$ , we arrive at (25).

**References**

1. Pincus, S.M.: Approximate entropy as a measure of system complexity. Proc. Natl. Acad. Sci. USA **88**, 2297–2301 (1991)
2. Cavanaugh, J.T., Guskiewicz, K.M., Giuliani, C., Marshal, S., Mercer, V., Stergiou, N.: Detecting altered postural-control after cerebral concussion in athletes with normal postural stability. Br. J. Sports Med. **39**, 805–811 (2005)
3. Sosnoff, J.J., Broglio, S.P., Shin, S.H., Ferrara, M.S.: Previous mild traumatic brain injury and postural-control dynamics. J. Athl. Train. **46**, 85–91 (2011)
4. National Center for Injury Prevention and Control: Report to Congress on Mild Traumatic Brain Injury in the United States: Steps to Prevent a Serious Public Health Problem. Centers for Disease Control and Prevention, Atlanta (2003)
5. Finkelstein, E.A., Corso, P.S., Miller, T.R.: The Incidence and Economic Burden of Injuries in the United States. Oxford University Press, New York (2006)
6. Aubry, M., Cantu, R., Dvorak, J., Graf-Baumann, T., Johnston, K.M., Kelly, J., Lovell, M., McCrory, P., Meeuwisse, W.H., Schamasch, P., Concussion in Sport (CIS) Group: Summary and agreement statement of the 1st international symposium on concussion in sport. Clin. J. Sport Med. **12**, 6–11 (2002)

7. Richman, J.S., Moorman, J.R.: Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol., Heart Circ. Physiol.* **278**, H2039–H2049 (2000)
8. McCrory, P., Meeuwisse, W., Johnston, K., Dvorak, J., Aubry, M., Molloy, M., Cantu, R.: Consensus statement on concussion in sport: The 3rd International Conference on Concussion in Sport held in Zurich, November 2008. *J. Athl. Train.* **44**, 434–448 (2009)
9. Cover, T.M., Thomas, J.A.: *Elements of Information Theory*. Wiley, New York (1991)
10. Grassberger, P., Procaccia, I.: Estimation of the Kolmogorov entropy from a chaotic signal. *Phys. Rev. A* **28**, 2591–2593 (1983)
11. Costa, M., Goldberger, A.L., Peng, C.K.: Multiscale entropy analysis of biological signals. *Phys. Rev. E* **71**, 021906 (2005)
12. Cao, Y.H., Tung, W.W., Gao, J.B., Protopopescu, V.A., Hively, L.M.: Detecting dynamical changes in time series using the permutation entropy. *Phys. Rev. E* **70**, 046217 (2004)
13. Lempel, A., Ziv, J.: On the complexity of finite sequences. *IEEE Trans. Inf. Theory* **22**, 75–81 (1976)
14. Hu, J., Gao, J.B., Principe, J.C.: Analysis of biomedical signals by the Lempel-Ziv complexity: The effect of finite data size. *IEEE Trans. Biomed. Eng.* **53**, 2606–2609 (2006)
15. Gao, J.B., Cao, Y.H., Tung, W.W., Hu, J.: *Multiscale Analysis of Complex Time Series—Integration of Chaos and Random Fractal Theory, and Beyond*. Wiley, New York (2007)
16. Gao, J.B., Hu, J., Tung, W.-W., Cao, Y.H.: Distinguishing chaos from noise by scale-dependent Lyapunov exponent. *Phys. Rev. E* **74**, 066204 (2006)
17. Andrzejak, R.G., Lehnertz, K., Rieke, C., Mormann, F., David, P., Elger, C.E.: Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Phys. Rev. E* **64**, 061907 (2001)
18. Adeli, H., Ghosh-Dastidar, S., Dadmehr, N.: A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy. *IEEE Trans. Biomed. Eng.* **54**, 205–211 (2007)
19. Gao, J.B., Hu, J., Tung, W.W.: Complexity measures of brain wave dynamics. *Cogn. Neurodyn.* **5**, 171–182 (2011)
20. Gao, J.B., Hu, J., Tung, W.W.: Facilitating joint chaos and fractal analysis of biosignals through nonlinear adaptive filtering. *PLoS ONE* **6**(9), e24331 (2011). doi:[10.1371/journal.pone.0024331](https://doi.org/10.1371/journal.pone.0024331)
21. Hu, J., Gao, J.B., Wang, X.S.: Multifractal analysis of sunspot time series: The effects of the 11-year cycle and Fourier truncation. *J. Stat. Mech.* 2009/02/P02066
22. Gao, J.B., Sultan, H., Hu, J., Tung, W.W.: Denoising nonlinear time series by adaptive filtering and wavelet shrinkage: A comparison. *IEEE Signal Process. Lett.* **17**, 237–240 (2010)
23. Tung, W.W., Gao, J.B., Hu, J., Yang, L.: Recovering chaotic signals in heavy noise environments. *Phys. Rev. E* **83**, 046210 (2011)
24. Gao, J.B., Hu, J., Buckley, T., White, K., Hass, C.: Shannon and Renyi entropies to classify effects of mild traumatic brain injury on postural sway. *PLoS ONE* **6**(9), e24446 (2011). doi:[10.1371/journal.pone.0024446](https://doi.org/10.1371/journal.pone.0024446)
25. Packard, N.H., Crutchfield, J.P., Farmer, J.D., Shaw, R.S.: Geometry from a time series. *Phys. Rev. Lett.* **45**, 712–716 (1980)
26. Takens, F.: Detecting strange attractors in turbulence. In: Rand, D.A., Young, L.S. (eds.) *Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, vol. 898, pp. 366–381. Springer, Berlin (1981)
27. Sauer, T., Yorke, J.A., Casdagli, M.: Embedology. *J. Stat. Phys.* **65**, 579–616 (1991)
28. Liebert, W., Pawelzik, K., Schuster, H.G.: Optimal embedding of chaotic attractors from topological considerations. *Europhys. Lett.* **14**, 521–526 (1991)
29. Kennel, M.B., Brown, R., Abarbanel, H.D.I.: Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **45**, 3403–3411 (1992)
30. Gao, J.B., Zheng, Z.M.: Local exponential divergence plot and optimal embedding of a chaotic time series. *Phys. Lett. A* **181**, 153–158 (1993)
31. Gao, J.B., Zheng, Z.M.: Direct dynamical test for deterministic chaos and optimal embedding of a chaotic time series. *Phys. Rev. E* **49**, 3807–3814 (1994)
32. Atmanspacher, H., Scheingraber, H.: A fundamental link between system theory and statistical mechanics. *Found. Phys.* **17**, 939–963 (1987)
33. Gaspard, P., Wang, X.J.: Noise, chaos, and  $(\varepsilon, \tau)$ -entropy per unit time. *Phys. Rep.* **235**, 291–343 (1993)
34. Cohen, A., Procaccia, I.: Computing the Kolmogorov entropy from time series of dissipative and conservative dynamical systems. *Phys. Rev. A* **31**, 1872–1882 (1985)
35. Eckmann, J.P., Ruelle, D.: Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* **57**, 617–656 (1985)
36. Gao, J.B., Hu, J., Tung, W.W.: Multiscale entropy analysis of biological signals: A fundamental bi-scaling law. *Europhys. Lett.* (submitted)
37. Lorenz, E.Z.: Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141 (1963)
38. Gao, J.B., Tung, W.W., Hu, J.: Quantifying dynamical predictability: The pseudo-ensemble approach (in honor of Professor Andrew Majda's 60th birthday). *Chin. Ann. Math., Ser. B* **30**, 569–588 (2009)
39. Kleeman, R.: Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.* **59**, 2057–2072 (2002)
40. Abramov, R., Majda, A., Kleeman, R.: Information theory and predictability for low frequency variability. *J. Atmos. Sci.* **62**, 65–87 (2005)
41. Haven, K., Majda, A., Abramov, R.: Quantifying predictability through information theory: Small sample estimation in a non-Gaussian framework. *J. Comput. Phys.* **206**, 334–362 (2005)
42. Gao, J.B., Hwang, S.K., Liu, J.M.: When can noise induce chaos? *Phys. Rev. Lett.* **82**, 1132–1135 (1999)
43. Gao, J.B., Chen, C.C., Hwang, S.K., Liu, J.M.: Noise-induced chaos. *Int. J. Mod. Phys. B* **13**, 3283–3305 (1999)
44. Hwang, K., Gao, J.B., Liu, J.M.: Noise-induced chaos in an optically injected semiconductor laser. *Phys. Rev. E* **61**, 5162–5170 (2000)
45. Gao, J.B., Hu, J., Tung, W.W., Cao, Y.H., Sarshar, N., Roychowdhury, V.P.: Assessment of long range correlation in time series: How to avoid pitfalls. *Phys. Rev. E* **73**, 016117 (2006)

46. Gao, J.B., Hu, J., Tung, W.W., Zheng, Y.: Multiscale analysis of economic time series by scale-dependent Lyapunov exponent. *Quant. Finance* (2011). doi:[10.1080/14697688.2011.580774](https://doi.org/10.1080/14697688.2011.580774)
47. Gao, J.B., Hu, J., Mao, X., Tung, W.W.: Detecting low-dimensional chaos by the “noise titration” technique: Possible problems and remedies. *Chaos Solitons Fractals* (in press)
48. Hu, J., Gao, J.B., Tung, W.W.: Characterizing heart rate variability by scale-dependent Lyapunov exponent. *Chaos* **19**, 028506 (2009)
49. Hu, J., Gao, J.B., Tung, W.W., Cao, Y.H.: Multiscale analysis of heart rate variability: A comparison of different complexity measures. *Ann. Biomed. Eng.* **38**, 854–864 (2010)
50. Cellucci, C.J., Albano, A.M., Rapp, P.E., Pittenger, R.A., Josiassen, R.C.: Detecting noise in a time series. *Chaos* **7**, 414–422 (1997)
51. Hu, J., Gao, J.B., White, K.D.: Estimating measurement noise in a time series by exploiting nonstationarity. *Chaos Solitons Fractals* **22**, 807–819 (2004)