Recurrence quantification analysis, pp. 45–64,
coordonné par C. Webber et N. Marwan, Springer, Berlin (2014)

Estimating Kolmogorov Entropy from Recurrence Plots

Philippe Faure and Annick Lesne

Abstract Kolmogorov entropy, actually an entropy rate h, has been introduced in chaos theory to characterize quantitatively the overall temporal organization of a dynamics. Several methods have been devised to turn the mathematical definition into an operational quantity that can be estimated from experimental time series. The method based on recurrence quantitative analysis (RQA) is one of the most successful. Indeed, recurrence plots (RPs) offer a trajectory-centered viewpoint circumventing the need of a complete phase space reconstruction and estimation of the invariant measure. RP-based entropy estimation methods have been developed for either discrete-state or continuous-state systems. They rely on the statistical analysis of the length of diagonal lines in the RP. For continuous-state systems, only a lower bound K_2 can be estimated. The dependence of the estimated quantity on the tunable neighborhood radius ε involved in constructing the RP, termed the ε-entropy, gives a qualitative information on the regular, chaotic or stochastic nature of the underlying dynamics. Although some caveats have to be raised about its interpretation, Kolmogorov entropy estimated from RPs offers a simple, reliable and quantitative index, all the more if it is supplemented with other characteristics of the dynamics.

Philippe Faure
CNRS UMR 7102, Neurobiologie des Processus Adaptatifs, Université Pierre et Marie Curie-Paris 6, 4 place Jussieu, 75252 Paris Cedex 05, France, e-mail: philippe.faure@snv.jussieu.fr

Annick Lesne
CNRS UMR 7600, Laboratoire de Physique Théorique de la Matière Condensée, Université Pierre et Marie Curie-Paris 6, 4 place Jussieu, 75252 Paris Cedex 05, France, e-mail: lesne@lptmc.jussieu.fr

1 Introduction

1.1 Entropy

The quantity today termed "Kolmogorov entropy" has been introduced for deterministic dynamical systems by Kolmogorov [1] and one year later by Sinai [2] hence it is sometimes called Kolmogorov-Sinai entropy or metric entropy. This quantity is similar to the entropy rate introduced in a different context by Shannon [3, 4] for symbolic sequences. They should not be confused with Kolmogorov complexity, which is another name for algorithmic complexity [5], nor with thermodynamic entropy, which is a different concept, only slightly and indirectly related through statistical mechanics and Boltzmann entropy [6].

Kolmogorov entropy, henceforth denoted h, is a global quantity providing a quantitative measure of the overall temporal organization of the dynamics. The initial motivation of Kolmogorov was to investigate whether some dynamical systems were isomorphic or not. He demonstrated that entropy is an invariant of the dynamics, i.e. it is preserved upon any isomorphism. To assess that two dynamical systems are non isomorphic, it is then sufficient to show that their entropies differ. Kolmogorov entropy is in fact an entropy rate. For a discrete-state source, Sec. 2, a direct and simple interpretation is provided by the asymptotic equipartition property. This property states that n-words, i.e. sequences of length n produced by the source, asymptotically separate in two classes: typical and non-typical n-words. Typical n-words have asymptotically the same probability e^{-nh} (hence the name "equipartition property") and their number scales as e^{nh}, while all the other n-words have a vanishing probability and in this respect no chance to be observed. Entropy has a meaning for both deterministic and stochastic dynamics. We will see, Sec. 4.1, that in the case of continuous-state dynamical systems, the intermediary steps of the computation at finite resolution ε in the phase space provide an auxiliary ε-entropy $h(\varepsilon)$, whose dependence with respect to ε reflects the deterministic or stochastic nature of the dynamics. To date, h has been used to characterize linguistic data [7], DNA sequences [8, 9], behavioral sequences [10, 11], speech analysis [12, 13], or spike emission in neurons [14, 15]. Estimating Kolmogorov entropy from experimental data is not an easy task. We will see that using recurrence plots allows a simple and reliable estimation of this value, or at least a lower bound.

1.2 Recurrence Plots

Several methods have been devised to turn the mathematical definition of Kolmogorov entropy into an operational quantity that can be estimated from experimental time series. The method based on recurrence plots (RPs) has proven to be an efficient one. RPs have been introduced as a graphical representation of a dynamical system well-suited for data analysis [16]. They consist in square binary matrices, depicting how the trajectory repeats itself: a black dot (i,j) is drawn in the plot when a recurrence is observed in the trajectory for times i and j. Patterns of activity can be identified in the organization of the recurrences along vertical or diagonal lines [17]. RPs thus provide a useful framework for visualizing the time evolution of a system and for discerning subtle transitions or drifts in the dynamics. Beyond being a graphical tool, RPs offer a means for the quantitative analysis of the underlying dynamics, what is summarized by the words "recurrence quantification analysis" (RQA) [18, 19, 20]. In particular, the graphical power of RPs will not be the point in computing Kolmogorov entropy, except if used for prior detection of non stationarity in the data prompting to windowing the series. Recurrence quantification analysis is particularly powerful because it relies on a trajectory-centered exploration of the phase space. In contrast to methods involving a blind partition of the phase space, containing many regions of negligible weight that considerably slow down the computation for no significant gain in accuracy, the recurrence pattern naturally samples important regions of the phase space, thus reducing computational complexity and time. This is analogous in spirit to Monte-Carlo-Markov-Chain sampling of the integration domain for computing an integral, instead of using a Riemann discretization of the integration domain.

As in a number of nonlinear analysis methods, it is necessary for building a RP to first reconstruct a multidimensional signal from a single observed variable. The reconstruction amounts to consider a series $(\mathbf{x}_i)_{i\geq 0}$ of embedded vectors $\mathbf{x}_i$ instead of the original one-dimensional time series $(u_t)_{t\geq 0}$. Their definition involves an embedding dimension m and a time delay τ, according to:

$$\mathbf{x}_i = (u_{i\tau}, u_{(i+1)\tau}, \ldots, u_{(i+m-1)\tau}) \tag{1}$$

The embedding of a one-dimensional signal in a multidimensional phase space and the choice of the involved time delay are standard procedures presented in Chapter 1 (see also [21]). The choice of m and τ is critical for any subsequent analysis since an inappropriate choice can either "insufficiently unfold" the high-dimensional dynamic or lead to false positive indications of chaos. An additional parameter in the RP is the radius ε of the neighborhoods defining recurrence. We will see that the dimension m is involved jointly with the length l of the diagonal lines, namely the relevant quantity is $m+l$, hence there is no need to consider m as an independent parameter.

RPs portray the dynamics of the embedded signals in the form of dots interspersed in a square matrix. Let $\mathbf{x}_i$ be the i-th point on the reconstructed trajectory, of length N, describing the system in an m-dimensional space. A recurrence plot is the $N \times N$ matrix in which a dot is placed at (i,j) whenever $\mathbf{x}_j$ is close to $\mathbf{x}_i$, i.e. whenever the distance $d(\mathbf{x}_i, \mathbf{x}_j)$ is lower that a given cutoff value ε. Different metrics can be used, for instance Euclidian distance or Maximum norm. The RP then contains N^2 black or white dots, see Fig. 1A(Left). The black dots represent the recurrence of the dynamical process determined with a given resolution ε, and their organization characterizes the recurrence properties of the dynamics. A vertical line of length l starting from a dot (i,j) means that the trajectory starting from $\mathbf{x}_j$ remains close to $\mathbf{x}_i$ during $l-1$ time steps. A diagonal black line of length l starting from a dot (i,j) means that trajectories starting from $\mathbf{x}_i$ and $\mathbf{x}_j$ remain close during $l-1$ time steps, Fig. 1A(Right). Diagonal segments in the RP (excluding the main diagonal) can be counted and plotted as an histogram according to their length. We will see below that this histogram, and in particular the slope α fo the linear region in the log-log plot (i.e. for l large enough) is the basis for estimating Kolmogorov entropy. It should be noted that changing the embedding dimension m amounts to shift by a few units the length of the diagonal lines, which does not affect the slope α. Hence the estimated value for Kolmogorov entropy will be independent of the embedding dimension, as shown in [22].

2 Discrete-State Signals

2.1 Messages, Symbolic Sequences and Discrete Sources

Information theory is concerned with the analysis of messages written with letters from a given alphabet [3]. This symbolic setting relates to continuous-state dynamical systems through encoding continuous states into discrete ones. Such encoding is an acknowledged approach allowing to prune irrelevant information, to improve statistics by reducing the dimension of the sequence and overall to simplify the system description without altering its essential dynamical properties [23]. For discrete-time dynamics in a continuous phase space, symbolic sequences can be obtained from the discretization of continuous-valued trajectories using a partition of the phase space in subsets A_w [24]. Each trajectory $(z_i)_{i\geq 0}$ is associated with a symbolic sequence $(w_i)_{i\geq 0}$ describing the array of visited subsets according to $z_i \in A_{w_i}$. The partition is said to be generating when the knowledge of the semi-infinite symbolic sequence$(w_i)_{i\geq 0}$ fully determines a unique initial condition $z_0 \in A_{w_0}$ in the continuous phase space. In this special case, the symbolic encoding is asymptotically faithful, with no loss of information compared to the continuous-valued trajectory (a loss of information nevertheless occurs when considering trajectories of finite length). However, generating partitions are very rare, existing only for sufficiently chaotic dynamical systems [25]. Even if generating partitions exist, a constructive method to determine them may not be available [26]. Discretization has then to be done using an a priori chosen partition of the phase space, with a main issue being to make the proper choice [27, 28].

Often in practice the system phase space is not fully formalized, think for instance of behavioral sequences recorded with a CCD camera. Encoding is then achieved in an heuristic way. Typically, discrete states are defined by partitioning the values of a few relevant observables. In the example of behavioral sequences, when considering the velocity V of a moving individual to be the relevant observable, time steps where $V < V_c$ will be coded 0 and time steps where $V \geq V_c$ will be coded 1, transforming the video recording of the individual into a binary sequence.

Another situation is the case where data are intrinsically discrete, e.g. language or DNA sequences [29, 30]. In this case, and more generally in the information-theoretic terminology, one speaks of a symbolic source generating messages, instead of discrete-state dynamics generating trajectories.

2.2 Entropy Rate of Symbolic Sequences

In information theory, entropy rate has been basically introduced to characterize languages modeled by Markov chains of increasing orders [3]. It measures the time-average information per symbol needed to transmit messages. It has been later demonstrated to coincide with Kolmogorov entropy (up to a factor relating ln and $\log_2$) in the case where the symbolic sequences originate from the phase-space discretization of a dynamical system according to a generating partition [31, 32]. Formally, it is defined as the limit of normalized block-entropies or block-entropy differences (theorem 5 in [3]). Block entropy H_n is defined as the Shannon entropy of the n-word distribution $p_n(.)$, namely $H_n = -\sum_{\bar{w}_n} p_n(\bar{w}_n) \log_2 p_n(\bar{w}_n)$ where the sum runs over the set of n-words $\bar{w}_n$. H_n increases with n, while the sequence of increments $h_n = H_{n+1} - H_n$ is a decreasing positive sequence. The increments h_n and the normalized quantities $\tilde{h}_n = H_n/n$ have the same limit $h = \lim_{n\to\infty} h_n = \lim_{n\to\infty} \tilde{h}_n$ (if it exists), which defines the entropy rate h. The definition of entropy h as a rate is thus far more than a mere normalization by a duration: it involves a time integration, by considering words of increasing length. The above information-theoretic definition can be reformulated in terms of temporal correlations, which reduce the amount of information required to retrieve a message. In other words, h reflects the temporal organization of the dynamics, taking small values when the dynamics has a strongly correlated structure. h thus provides an integrated measure of the overall temporal correlations present in the dynamics, and $1/h$ can roughly be seen as a correlation time. It is to note that h_n corresponds to the entropy rate of the $(n-1)$-th order Markov approximation of the source, involving only n-point joint probabilities [3]. In the above definition, we used the binary logarithm $\log_2$ to match information-theoretic usage and Shannon definition; h is then expressed in bits per time unit. It is straightforward to replace $\log_2$ by the Neperian logarithm ln so as to exactly recover Kolmorogov entropy.

Estimation of h is currently based on the above definition [10, 33, 34], with h_n appearing as a better estimator than $\tilde{h}_n$ (although both are unbiased). In practice, due to the finite size N of the data sequence, the estimated value $\hat{H}_n$ plotted as a function of the block size n saturates to $\log_2 N$. If h has a non trivial value, this plot displays a linear region of slope h, and the crossover to the asymptotic plateau occurs around $n^* = (\log_2 N)/h$ [34, 35]. Another method uses the identity for ergodic sources between the entropy rate h and the Lempel-Ziv complexity, defined for a single sequence and computed using compression algorithms [36, 37, 38]. The latter method may perform better, in particular for short sequences [35]. Actually a proper implementation requires a two-step estimation. The first step is to obtain a rough estimate of h, delineating the validity and the performance of these alternative methods, giving in particular a lower bound on the sequence length required for the block-entropy method to be reliable. Then a refined estimate of h is obtained using the best of the two methods [35]. A third alternative is provided by RP representation of the symbolic sequences, as we will see below, Sec. 2.4.

2.3 Shannon-McMillan-Breiman Theorem

Both the interpretation of the above-defined quantity h and its estimation via RP, Sec. 2.4, relies on the Shannon-McMillan-Breiman theorem [39]. This theorem, established as theorem 3 in [3] and further improved by McMillan then Breiman [40, 4, 41], states that the number of typical n-words (i.e. n-words that have the same statistical properties corresponding to the almost sure behavior) scales like e^{nh} as $n \to \infty$, where the exponent h is the entropy rate of the source. A corollary of this theorem is the asymptotic equipartition property, stating that the probability $p_n(\bar{w}_n)$ of every typical n-word $\bar{w}_n$ takes asymptotically the same value e^{-nh}. Although containing the core idea of the asymptotic equipartition property, this finite-size statement has been made more rigorous on mathematical grounds. Its formulation requites to introduce random variables $\hat{\mathscr{P}}_n$ depending on the realization $\bar{w}$ of the whole symbolic sequence according to $\hat{\mathscr{P}}_n(\bar{w}) = p_n(w_0, \ldots, w_{n-1})$. The asymptotic equipartition property then writes

$$\lim_{n\to\infty} (-1/n) \ln \hat{\mathscr{P}}_n \to h \qquad \text{in probability} \tag{2}$$

This means that for any $\delta > 0$ and $\varepsilon > 0$ arbitrary small, there exists a word threshold size $n^*(\delta, \varepsilon)$ such that $\text{Prob}(\{\bar{w}, p_n(w_0, \ldots, w_{n-1}) > e^{n(-h+\delta)}\}) < \varepsilon$ and $\text{Prob}(\{\bar{w}, p_n(w_0, \ldots, w_{n-1}) < e^{n(-h-\delta)}\}) < \varepsilon$ for any $n \geq n^*(\delta, \varepsilon)$, or equivalently in terms of n-word subsets, $p_n(\{\bar{w}_n, p_n(\bar{w}_n) > e^{n(-h+\delta)}\}) < \varepsilon$ and $p_n(\{\bar{w}_n, p_n(\bar{w}_n) < e^{n(-h-\delta)}\}) < \varepsilon$. As a corollary of this result, the number of typical n-words $\bar{w}_n$ for which the asymptotic equipartition property $p_n(\bar{w}_n) \sim e^{-nh}$ holds scales as e^{nh} for n large enough, providing another interpretation of h.

Shannon-McMillan-Breiman theorem will be the basis of the entropy estimation method from RPs. It justifies that all the observed n-words belong to the set of typical words since non-typical ones are too rare to be ever observed. An important caveat is the asymptotic nature of this theorem, making its application to finite words and finite sequences a questionable extrapolation. However, numerical experiments show that the asymptotic regime is reached rapidly and the theorem yields the correct dominant behavior even for moderate values of n (lower than 10).

2.4 *RP-based Estimation of the Entropy (per Unit Time)*

Our starting point will be the representation of the data as a m-RP describing the recurrence of m-words: a black dot (i,j) means that $\bar{w}_m(i) = \bar{w}_m(j)$ where $\bar{w}_m(i) = (w_i, \dots, w_{i+m-1})$ denotes the m-words starting at time i in the original sequence. The integer m thus appears as an embedding dimension. Observing a diagonal line of length l starting in (i,j) in the m-RP means that the two $(m+l-1)$-words starting at times i and j coincide: $\bar{w}_{m+l-1}(i) = \bar{w}_{m+l-1}(j)$. Such a line corresponds to a diagonal line of length $l+1$ in the $(m-1)$-RP and a single dot in the $(m-l+1)$-RP. In fact, all the quantities that may be introduced regarding the statistics of diagonal lines are relative to a given realization of the m-RP and they depend not only on m but also on the sequence length N and its realization $\bar{w}$, which will be skipped for simplicity. We will assume that the size of the m-RP, or equivalently the sequence length N, is large enough to identify quantities computed in one realization of the m-RP and their statistical average, based on the assumed ergodicity of the dynamics. We will also assume (and numerically check) that the asymptotic probability estimate given by Shannon-McMillan-Breiman theorem, that centrally involves the entropy rate h of the source, is valid at the leading order for the considered words.

We consider a length l large enough, so that a $(m+l-1)$-word occurs at most twice and Shannon-McMillan-Breiman theorem approximately holds. The probability of double occurrence of a typical $(m+l-1)$-word, $(N-l-m+1)e^{-h(m+l-1)}$, multiplied by the number $e^{h(m+l-1)}$ of these non-identical typical words yields the number $\nu_m^{(N)}(l)$ of diagonal (and possibly overlapping) segments of length l in the upper triangle of the m-RP, not counting the main diagonal line:

$$\nu_m^{(N)}(l) = \frac{(N-l-m+2)(N-m-l+1)}{2}\, e^{-h(m+l-1)} \tag{3}$$

Note that this histogram of line lengths l is currently denoted $H_D(l)$ (see e.g. Chapter 1)); we adopt the notation $\nu(l)$ to avoid any confusion with block entropies H_n.

Since $m+l \ll N$ we may identify $N-l-m+2$ and $N-l-m+1$ with N, getting:

$$\nu_m^{(N)}(l) \sim (N^2/2)\, e^{-h(m+l-1)} \tag{4}$$

A semi-log plot of $\nu_m^{(N)}(l)$ with respect to l will exhibit a slope $-h$ in its linear region. While the theoretical derivation of the scaling behavior is done for $\nu_m^{(N)}(l)$, numerical implementation is more easily done in practice using the number $\eta_m^{(N)}(l)$ of diagonal lines of length exactly equal to l in the m-RP, or the cumulative number $\phi_m^{(N)}(l)$ of diagonal lines of total length larger or equal than l (Fig. 1B). After normalization by the total number of diagonal lines, $\phi_m^{(N)}(l)$ coincides with the cumulative probability $p^c(l)$ introduced in Chapter 1. However, $\phi_m^{(N)}(l)$ and $p^c(l)$ satisfy the same scaling laws, which allows to circumvent the normalization issue. We will henceforth work with the raw number $\phi_m^{(N)}(l)$. A diagonal line of total length $l+r$ yields $r+1$ (partly overlapping) diagonal stretches of length l contributing to

$\nu_m^{(N)}(l)$. It follows that $\nu_m^{(N)}(l) = \sum_{r\geq 0}(r+1)\,\eta_m^{(N)}(l+r)$ [39]. The scaling behavior of $\phi_m^{(N)}(l)$ is then derived using $\nu_m^{(N)}(l) - \nu_m^{(N)}(l+1) \approx -d\nu_m^{(N)}(l)/dl = h\nu_m^{(N)}(l)$:

$$\phi_m^{(N)}(l) \equiv \sum_{r\geq 0} \eta_m^{(N)}(l+r) \approx -\frac{d\nu_m^{(N)}(l)}{dl} \sim (hN^2/2)\, e^{-h(m+l-1)} \tag{5}$$

A semi-log representation of the number $\phi_m^{(N)}(l)$ of diagonal lines of total length larger than l as a function of the length l would also have a slope $-h$ in its linear region, which provides a direct way to estimate h from RP, as presented in Fig. 1. Note finally that the average length of the diagonal lines in the m-RP expresses:

$$\langle D \rangle = \frac{\sum_{l\geq 1} l\, \eta_m^{(N)}(l)}{\sum_{l\geq 1} \eta_m^{(N)}(l)} = \frac{\nu_m^{(N)}(1)}{\nu_m^{(N)}(1) - \nu_m^{(N)}(2)} \tag{6}$$

The second equality is obtained using the change of variable $l = r+1$ in $\nu_m^{(N)}(1)$ and $l = r+2$ in $\nu_m^{(N)}(2)$ where the expressions for $\nu_m^{(N)}(1)$ and $\nu_m^{(N)}(2)$ have been obtained by plugging $l = 1$ and $l = 2$ in the identity $\nu_m^{(N)}(l) = \sum_{r\geq 0}(r+1)\,\eta_m^{(N)}(l+r)$. At the leading order, the scaling $\sum_{r\geq 0} \eta_m^{(N)}(l+r) \sim (hN^2/2)\, e^{-h(m+l-1)}$ yields $\langle D \rangle \sim e^h/(e^h - 1)$ and even $\langle D \rangle \sim 1/h$ if $h \ll 1$, which gives an intuitive interpretation of $1/h$ as a characteristic time (correlation time) of the source. Note that the sums include single dots ($l = 1$), which is not always the case in RP. As noted before, the relevant quantity in a m-RP is $m+l$. A line of length 1 in a m-dimensional space is a line of length 2 in a $(m-1)$-dimensional space, hence there is no obvious reason to discriminate single dots in a m-RP.

3 Continuous-State Dynamics

3.1 Kolmogorov Entropy for a Continuous State System

For dissipative dynamical systems, attractors provide a global picture of the long term behavior. A more quantitative representation of the latter is given by the invariant measure on the attractor, that is, the probability measure invariant upon the action of the dynamics. It describes how frequently a given trajectory visits any particular region of the state space. The state space can be divided into a finite number of intervals (one dimension) or boxes (two or more dimensions) which defines a finite partition $\mathscr{A} = \{A_i, i = 1,\ldots,m\}$. The frequency at which a trajectory visits these specific boxes thus gives a partial insight into this invariant measure.

Let us first consider discrete-time dynamics, i.e. dynamics generated by maps. As explained in Sec. 2.1, a trajectory $(z_i)_{i\geq 1}$ in the continuous phase space can be encoded by a n-word $\bar{w}_n = (w_1,\ldots,w_n)$, meaning that the trajectory successively visits the regions $A_{w_1},\ldots,A_{w_n}$, with $z_i \in A_{w_i}$. Denoting $P_n(\bar{w}_n)$ the probability of

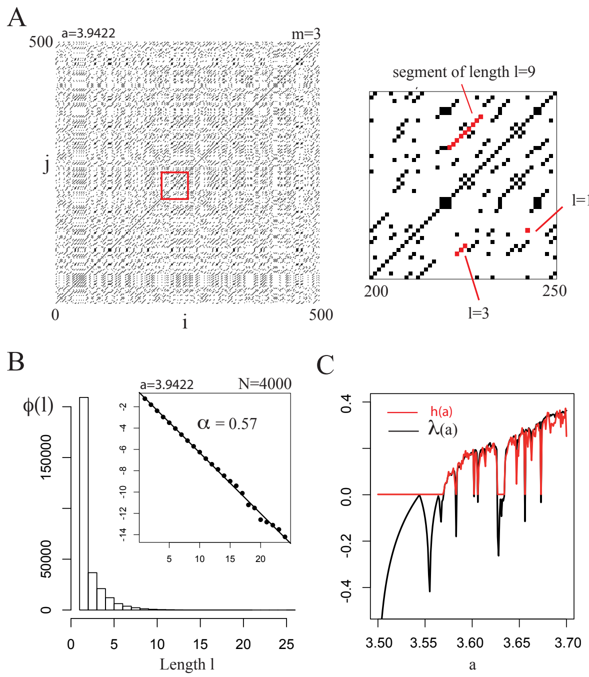

Fig. 1 Method for constructing length histograms of diagonal line segments. (A, Left) RP obtained from a logistic map $g_a(z) = az(1-z)$ in $[0,1]$ with $a = 3.9422$ and state-discretized using a symbolic encoding based on the simple rule: if $z_i > 0.5$ then $w_i = 1$ else $w_i = 0$. The embedding dimension is $m = 3$, hence a dot (i, j) means that the 3-words (w_i, w_{i+1}, w_{i+2}) and (w_j, w_{j+1}, w_{j+2}) are identical. (A, Right) Detail of the RP with diagonal line segments of various length l. (B) Number $\phi_m^{(N)}(l)$ of diagonal lines of total length longer or equal to l, counted in the RP obtained from $(w_i)_{i=1,\dots,4000}$. (Inset) Semi-log representation of $\phi_m^{(N)}(l)$; the absolute value $-\alpha$ of the slope of the fitting line yields an estimation of h. (C) Comparison of the RP-estimated value of $h(a)$ (red line) with Lyapunov exponent value $\lambda(a)$ (black line). Trajectories used to estimate $h(a)$ have a length $N = 2000$ with $3.5 \le a \le 3.7$. A negative value of $\lambda(a)$ corresponds to an entropy value $h(a) = 0$, whereas Pesin equality ensures $h(a) = \lambda(a)$ for positive values of $\lambda(a)$.

a n-word $\bar{w}_n$, equal to the measure of the set of points whose n-step trajectory is encoded by $\bar{w}_n$, the n-block entropy for this partition $\mathscr{A}$ is

$$H_n(\mathscr{A}) = -\sum_{\bar{w}_n} P_n(\bar{w}_n) \ln P_n(\bar{w}_n) \tag{7}$$

where the sum runs over the set of all possible n-words. The relationship to Shannon block-entropies is obvious. As for Shannon entropy rate, Kolmogorov entropy (also termed Kolmogorov-Sinai entropy [1, 2] or metric entropy) is defined as the limit

$$h = \sup_{\mathscr{A}} \lim_{n\to\infty} h_n(\mathscr{A}) \tag{8}$$

The supremum is reached for special partitions, called "generating partitions", when they exist (Sec. 2.1). However, this powerful theorem is rarely applicable in practice as we do not know how to construct generating partitions except for unimodal maps and certain maps of the planes. In general, the supremum over the partitions $\mathscr{A}$ is reached at infinitely refining the partitions. It is enough to consider a sequence of partitions $\mathscr{A}_\varepsilon$ whose boxes have a typical linear size ε. Denoting $h_n(\varepsilon) \equiv h_n(\mathscr{A}_\varepsilon)$ and $h(\varepsilon) = \lim_{n\to\infty} h_n(\varepsilon)$, it comes $h = \lim_{\varepsilon\to 0} h(\varepsilon)$.

In the case of continuous-time dynamics, an additional parameter is the time delay τ involved in the reconstruction of the dynamics. As explained in the introduction, the first step is the reconstruction of a m-dimensional trajectory $(\mathbf{x}_i)_{i=1,\ldots,N}$ of length N from the continuous-state experimental trajectory $u(t)$, according to $\mathbf{x}_i = [u(i\tau), u((i+1)\tau), \ldots, u((i+m-1)\tau)]$. We denote $H_{n,\tau}(\varepsilon)$ the n-block entropy corresponding to the sequence $(\mathbf{x}_i)_{i=1,\ldots,N}$. The difference $H_{n+1,\tau}(\varepsilon) - H_{n,\tau}(\varepsilon)$ is the average information needed to predict which box of the partition $\mathscr{A}_\varepsilon$ will be visited at time $(n+1)\tau$, given the n boxes visited up to $n\tau$. Definition of Kolmogorov entropy is then similar to the discrete-time case, except for an additional limit $\tau \to 0$

$$h = \lim_{\tau\to 0} \frac{1}{\tau} \lim_{\varepsilon\to 0} \lim_{n\to\infty} \left(h_{n,\tau}(\varepsilon)\right) \tag{9}$$

Generalized entropies $h^{(q)}$ can be defined according to [42, 43]

$$h^{(q)} = -\lim_{\tau\to 0} \frac{1}{\tau} \lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{q-1} \ln \sum_{\bar{w}_n} [P_n(\bar{w}_n)]^q \tag{10}$$

For $q = 1$, we have $h = h^{(1)}$, and it can be demonstrated that $h^{(q)} \geq h^{(q')}$ for every $q' > q$. The inequality is strict as soon as the attractor has a nontrivial multifractal structure. In particular, $h^{(2)}$, currently denoted K_2, is a lower bound for Kolmogorov entropy. This property is centrally used in the estimation of h. Indeed, the mathematical definition of Kolmogorov entropy for a continuous-state dynamics involves several non-commuting limits, which prevent any direct implementation. Actually, only $K_2 = h^{(2)}$ can be extracted from experimental time series, either by a method based on phase space reconstruction and the computation of a correlation integral [42, 44], or by a method based on RQA [45], both presented below.

3.2 Grassberger and Procaccia Method for Computing K_2

A first method for estimating K_2 from experimental data has been developed by Grassberger and Procaccia [42]. The estimation method is based on the computation of the correlation integral of the reconstructed trajectory in embedding dimension m

$$C^m(\varepsilon) = \frac{1}{N} \sum_{i=1}^{N} C_i^m(\varepsilon) \tag{11}$$

where $C_i^m(\varepsilon)$ is the number of time indices j $(1 \leq j \leq N)$ for which $d(\mathbf{x}_i, \mathbf{x}_j) \leq \varepsilon$. Note that the quantity $C^m(\varepsilon)$ can be viewed as the average probability that two trajectories visit jointly a given sequence of m boxes of the partition $\mathscr{A}_\varepsilon$. This quantity provides a direct access to the entropy bound $K_2 \leq h$ according to:

$$\tau K_2 = \lim_{\varepsilon \to 0} \lim_{m \to \infty} \ln \frac{C^m(\varepsilon)}{C^{m+1}(\varepsilon)} \tag{12}$$

Since $C_i^m(\varepsilon)$ is the number of black dots in line i when the RP is defined at resolution ε, $C^m(\varepsilon)/N$ is also the recurrence density at resolution ε and embedded dimension m. This establish a simple link between RPs and correlation integrals. However a method based on an estimation of the density would ignore the spatial organization of dots in the RP, which contain important information about the dynamics. It would moreover require to explicitly construct RPs for any embedding dimension m.

3.3 RP-Based Method for Computing K_2

RQA, here based on diagonal line statistics, offers an alternative method for computing K_2, with the advantage that it will be enough to consider the reconstructed sequence $(\mathbf{x}_i)_{i=1,\ldots,N}$ for a single embedding dimension m [45]. Indeed, as in the discrete-state case, changing the embedding dimension only shifts the diagonal line statistics, since a diagonal line of length l in dimension m corresponds to a diagonal line of length $l-1$ in dimension $m+1$. Let us denote $\nu_\varepsilon(l)$ the number of diagonal segments of length l, possibly included in longer segments. A diagonal line thus contributes by all l-segments that can be delineated in it. For instance, a diagonal line of four points contributes to $\nu_\varepsilon(l=2)$ by 3 segments, to $\nu_\varepsilon(l=3)$ by 2 segments and to $\nu_\varepsilon(l=4)$ by a single segment (itself). As in the discrete case, it can be shown that

$$\nu_\varepsilon(l) = \text{const.} e^{-l\alpha(\varepsilon)} \tag{13}$$

hence the exponent $\alpha(\varepsilon)$ can be obtained by fitting the linear part of the log histogram of $\nu_\varepsilon(l)$. The expected limiting behavior is:

$$\lim_{\varepsilon \to 0} \alpha(\varepsilon) = \tau K_2 \tag{14}$$

Actually, as presented in Fig.1 in the discrete-state case, the cumulative number $\phi_\varepsilon(l)$ of diagonal lines of total length equal or larger than l displays the same scaling behavior, and could be easier to extract from the RP. A noticeable point is that for a discrete-state system, it follows from Shannon-McMillan-Breiman theorem and the ensuing asymptotic estimate of n-word probability $p_n(.)$ that all $h^{(q)}$ coincide with h, in particular $K_2 = h$, so that the results of the present Sec. 3.3 are fully consistent with those of Sec. 2.4. The main difference between this approach and the Grassberger-Procaccia method is that the convergence of $\alpha(\varepsilon)$ is now studied in terms of the $\varepsilon \to 0$ limit, and not by increasing the embedding dimension $m \to \infty$.

The resulting advantage is that the distribution of the distances between points need to be calculated only once, for constructing a single RP, instead of for each new embedding dimension.

4 Some Factors Influencing Computation of Entropy

4.1 The Notion of ε-Entropy for Analyzing Noisy Dynamics

The presence of the limits $\varepsilon \to 0$ (resolution ε in the phase space) and $n \to \infty$ (length n of words, i.e. trajectory segments) in the entropy definition has the consequence that the estimated entropy is, at best, an approximation. In particular, for small resolutions ε, the number of neighboring points become too small to get reliable statistics. This limitation becomes more critical as the word-length n increases or the length N of the time series decreases. Overall, the major constraint imposed by the limit $\varepsilon \to 0$ is due to the noise inherent to experimental data. It prevents any analysis below a given resolution at which deterministic structures are destroyed. This resolution threshold is directly related to the variance of the noise. The choice of an optimal threshold ε a priori depends on the considered time series but Thiel et al. [46] suggested that a value of $\varepsilon = 5\sigma$ (where σ is the fluctuation level in the signal) is appropriate for a wide class of processes. Letellier suggested the value $\varepsilon = \sigma\sqrt{m}/10$ where m is the dimension of the embedding [47]. Effect of noise is illustrated in Fig. 2. where the convergence of the slope $\alpha(\varepsilon)$ towards h is depicted for a Hénon map system with additive Gaussian noise. For a noise value equal to 0 there is a clear logarithmic convergence towards K_2. As soon as noise increases, a threshold of divergence $\varepsilon_{\mathrm{div}}(\zeta)$ appears, with a related upward swing of the curves. Such a result is consistent with the fact that $\alpha(\varepsilon)$ converges towards K_2 in a chaotic map, while it diverges as $\varepsilon \to 0$ in a stochastic dynamics [32, 48, 5]. As a consequence, a plot like that of Fig. 2A, provides complete information on the system, with a good estimate of K_2 in the range $\varepsilon > \varepsilon_{\mathrm{div}}(\zeta)$. In contrast, it only gives information on the noise component if $\varepsilon < \varepsilon_{\mathrm{div}}(\zeta)$.

4.2 Non-stationarity

A key assumption in entropy estimation is the statistical stationarity of the source. When dealing with experimental data, people are confronted to a trade-off between the requirement of recording long time series and the non-stationarity of real systems. Non-stationarity can produce spurious identification of chaos [21], hence it has triggered the development of statistical methods (such as surrogates) to test the obtained results. Development of methods allowing to analyze non stationary time series is thus important [49]. RPs allow to get a visual assessment of the assumption

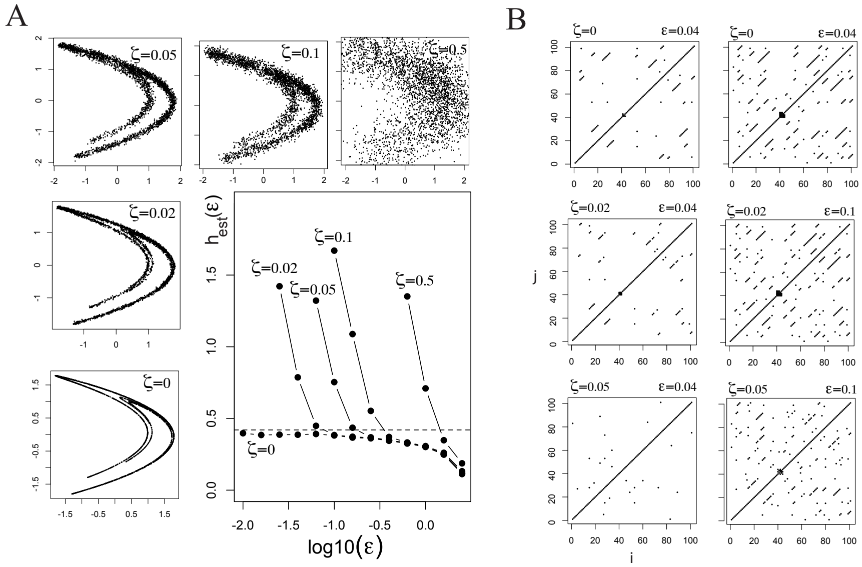

Fig. 2 Plot of the RP-estimated ε-entropy $h_{est}(\varepsilon) = \alpha(\varepsilon)$ as a function of the resolution ε (neighborhood radius) for different noise levels ζ. (A) Time series of $N = 2000$ points obtained from a Hénon map (with a = 1.4, b = 0.3) for which the theoretical entropy rate is $h = 0.42$. Gaussian white noise with standard deviation ζ was added, with ζ varying from 0 to 0.5. The RP-estimated value $h_{est}(\varepsilon) = \alpha(\varepsilon)$ is represented as a function of ε for the five times series. As soon as the noise level increases, the convergence of $h_{est}(\varepsilon)$ to the theoretical value of K_2 (horizontal dashed line) as ε tends to 0, which indicates chaos, is interrupted by an upward swing at some value $\varepsilon_{\text{div}}(\zeta)$, reflecting the stochastic component of the dynamics. For a given level of noise, the convergence can only be inferred from the inflexion point in the computed curve. (B) RP for different values of the noise level ζ and resolution ε.

of stationarity. Homogeneity of the RP gives a support of sequence stationarity. Else, non-stationary features like the presence of a drift in the evolution law (reflecting in inhomogeneous lower right and upper left corners, compared to the RP core), the occurrence of transitions (reflecting in disruptions within the RP) or periodicities (reflecting in periodic patterns in the RP) can be easily detected. This is illustrated in Fig. 3B where we plotted the recurrence pattern of a time series obtained by concatenating two time series from logistic maps with different values of the parameter a. The RP shows a clearly heterogeneous organization in four quadrants illustrating the modification of temporal properties when passing from the first sequence to the second one. In such non-stationary situations, statistical analysis and in particular entropy estimation should be restricted to time windows where the RP is statistically homogeneous. As the shift points are usually not known in advance, a sliding window is used in practice. The optimal size of the sliding window has to be determined in a preliminary step, either by visual inspection of the RP, or by more quantitative image analysis techniques to determine the typical size of statistically homogeneous regions in the RP, or by invoking additional knowledge about the system. We have numerically implemented the sliding window procedure on a binary discretized trajectory, initially generated by a logistic map whose control parameter a slowly increases by small steps in the course of evolution (Fig. 3A). This increase is slow enough for a quasi-stationary approximation to make sense. It allows to consider an entropy rate $h(a)$ corresponding to the instantaneous value of a and characterizing the non-stationary dynamics during the associated transient stage. Fig. 3C shows that the evolution of the entropy rate $h(a)$ as a varies, although very irregular, can be faithfully captured by entropy RP-estimation in a sliding window.

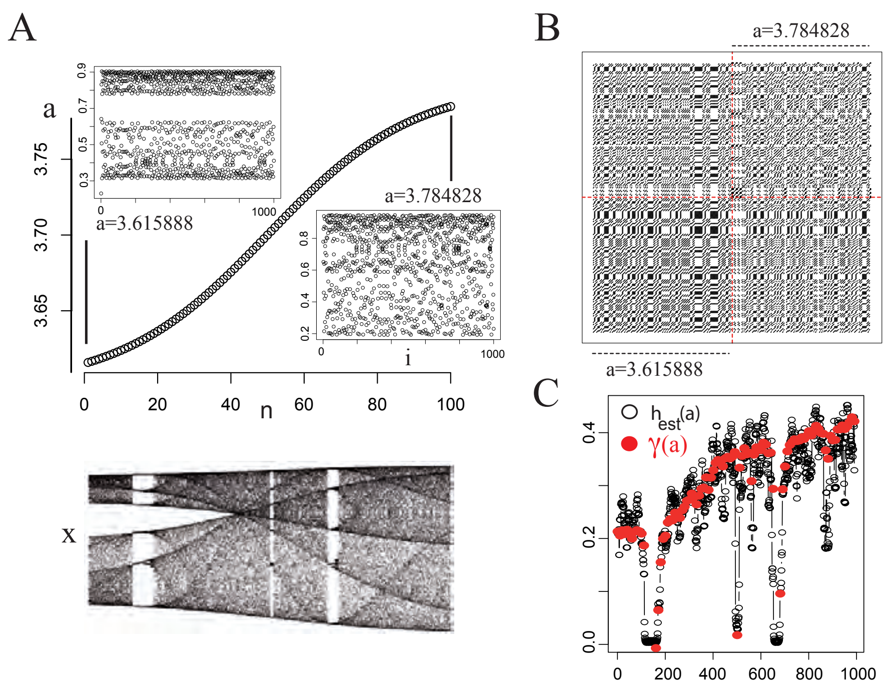

Fig. 3 Entropy estimation for non-stationary dynamics (A, Top Left) Variation of a from 3.6 to 3.8, according to a sigmoidal function. The value of a in the n-th part of the simulation is given by $a_n = 3.6 + \frac{2}{10*(1+\exp(-0.5*n))}$ with $n = 1, \dots, 100$. (A, Bottom Left) Time series obtained from concatenation of 100 time series $z_{a_n}(i)_{i=1,\dots,1000}$ for $n = 1, \dots, 100$ from logistic map $g_a(z) = az(1-z)$, with a varying from 3.6 to 3.8, as can be read on the above curve. The last point of a time series $z_{a_n}(i)$ is used as the initial point to calculate the following $z_{a_{n+1}}(i)$ series. (A, Right) Two of these time series for $a = 3.615888$ and $a = 3.784828$. (B) Recurrence plot obtained by concatenating the two time series illustrated in (A, Right). (C) Variation of h along the non-stationary dynamics. $h(a)$ is estimated on successive overlapping windows of length $N = 1000$ with a shift of 100 time steps. For each value of a, estimated value $h_{est}(a)$ of $h(a)$ (black points) is superimposed with the value of Lyapunov exponent $\lambda(a)$ (red points).

5 Discussion

For continuous-state dynamics, whose attractor has a non trivial multifractal structure, only a lower bound K_2 of the Kolmogorov entropy h can be estimated from experimental data. Discrete-state dynamics has a simpler structure, for which the different entropies $h^{(q)}$ coincide, hence it is possible to estimate $h = K_2$. However, discretizing the continuous states does not solve the issue, since what will be estimated is the Kolmogorov entropy of an approximation of the continuous-state dynamics, missing the multifractal structure of the attractor, and providing anyhow a lower bound on the actual Kolmogorov entropy of the original dynamical system. More generally, discretization is associated with a loss of information about the states, and all the discretization procedures presented above are quite sensitive to the contamination by an additive noise. However, quantitative analysis, and specifically entropy estimation, is expected to be statistically more faithful when performed on symbolic sequences.

Kolmogorov entropy rate h should not be confused with the Shannon entropy of the length distribution of black diagonal lines [17] or white diagonal lines [47]. Black diagonal lines correspond to the recurrence of a segment of trajectory, that is, two stretches of trajectory remaining close one to the other during l time steps if the diagonal length is l. Intuitively, it is thus expected in case of a chaotic dynamics that their average length scales as the inverse of the maximal Lyapunov exponent. Numerical evidences, for some 1D maps, seem to suggest that these Shannon entropies display the same behavior than the maximal Lyapunov exponent, coinciding for hyperbolic 1D maps with the Kolmogorov entropy rate h. However, to our knowledge, an analytical and general link between Kolmogorov entropy rate and the Shannon entropy of the length distribution of black or white diagonal lines is still lacking.

We have seen that h roughly measures the range of temporal correlations. In other words, the time during which the behavior of the system can be predicted is proportional to $1/h$. More rigorously, $h = \lambda$ when there is a single strictly positive Lyapunov exponent λ. In the case where two or more Lyapunov exponents are strictly positive, what is termed hyperchaos, Kolmogorov entropy determines only a lower bound on the sum of positive Lyapunov exponents according to $h \leq \sum_{\lambda \geq 0} \lambda$. This inequality, known as the Pesin inequality, turns into an equality under the condition of uniform hyperbolicity (typically Anosov and Axiom A systems). As a consequence, if h approaches zero, the system becomes fully predictable (for example the case of periodic dynamics). On the other hand, a finite positive h indicates the presence of chaos while h diverges for a stochastic dynamics. Entropy is also closely related to recurrence times [50, 39]. In the discrete case, Wyner-Ziv theorem states that the minimal recurrence time at the level of m-words (i.e. the smallest time t such that $\bar{w}_m(0) = \bar{w}_m(t)$) behaves asymptotically as e^{mh} [51]. This theorem can be exploited either to give an interpretation of h in terms of recurrence times, or conversely to estimate h from the recurrence times. Note that recurrence times involved in this theorem are minimal recurrence times given by the vertical distance to the main diagonal (first bisector), that is, the length of the white vertical line separating a point (i,i) from its first recurrence, associated with some dot (i,j) [39, 52]

Overall the use of a RP-based, finite-size estimation of the Kolmogorov entropy for characterizing an experimental dynamics is clearly limited. Two types of limiting factors can be distinguished. The first type refers to the dynamics of the investigated system and the very nature of Kolmogorov entropy. Although entropy is an absolute dynamical invariant, it characterizes only stationary regimes, and the superimposition of slow and fast components in the dynamics [32] does not straightforwardly reflect in the entropy. Another set of limiting factors is related to the nature of the data and estimation issues. Experimental data are generally associated with short time series, high-dimensional underlying dynamics, noise and non-stationarity. These factors (some of which have been discussed above) all limit the accuracy and reliability of the entropy estimation. Furthermore, except in the case of symbolic dynamics, only a lower bound K_2 can be obtained. Estimating Kolmogorov entropy is however appropriated for comparison purposes. In data analysis, entropy is often used for comparing the real system with null models, through a comparison of their entropies. Moreover, entropy is a unifying concept insofar as it applies to both deterministic and stochastic dynamics. Such a feature alleviates data analysis from the need for assessing the deterministic nature of the dynamics and makes the same estimation procedure valid in both cases. Finally, entropy provides an overall quantification of the complexity of the dynamics but it cannot answer questions about a specific moment, nor about a specific region of the phase space. This gap can be filled with further analysis of the RPs, which possibly provides information about what happens in a localized region of the phase space [50]. While RPs are suitable for statistical analysis of sequences (extraction of average or integrated features like the entropy or the average recurrence time) they also allow to keep track of the temporal location of specific events, hence allowing to visually evidence and locate dynamic transitions and more generally non-stationary features of the evolution.

References

1. A.N Kolmogorov. A new metric invariant of transient dynamical systems and automorphisms in lebesgue spaces (in russian). *Doklady Akademii Nauk SSSR*, 119:768–771, 1958.
2. Sinai Ya.G. On the concept of entropy for a dynamic system. *Doklady Akademii Nauk SSSR*, 124:768–771, 1959.
3. C.E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:479–423, 623–656, 1948.
4. T.M. Cover and J.A. Thomas. *Elements of information theory*. Wiley, New York, 1991.
5. G. Nicolis and P. Gaspard. Toward a probabilistic approach to complex systems. *Chaos, Solitons and Fractals*, 4:41–57, 1994.
6. A. Lesne. Shannon entropy: a rigorous mathematical notion at the crossroads between probability, information theory, dynamical systems and statistical physics. *Mathematical Structures in Computer Science*, 24, 2014.
7. W. Ebeling and T. Pöschel. Entropy and long-range correlations in literary english. *Europhysics Letters*, 26:241–246, 1994.

8. H. Herzel, W. Ebeling, and A.O. Schmitt. Entropies of biosequences: The role of repeats. *Physical Review E*, 50:5061–5071, 1994.
9. C.K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, M. Simons, and Stanley H.E. Finite-size effects on long-range correlations: Implications for analyzing DNA sequences. *Physical Review E*, 47:3730–3733, 1993.
10. M.P Paulus, M.A. Geyer, L.H. Gold, and A.J. Mandell. Application of entropy measures derived from ergodic theory of dynamical systems to rat locomotor behavior. *Proceedings of the National Academy of Sciences USA*, 87:723–727, 1990.
11. P. Faure, H. Neumeister, D.S. Faber, and H. Korn. Symbolic analysis of swimming trajectories reveals scale invariance and provides model for fish locomotion. *Fractals*, 11:233–243, 2003.
12. K. Doba, L. Pezard, A. Lesne, V. Christophe, and J.L. Nandrino. Dynamics of emotional expression in autobiographical speech of patients with anorexia nervosa. *Psychological Reports*, 101:237–249, 2007.
13. K. Doba, J.L. Nandrino, A. Lesne, J. Vignau, and L. Pezard. Organization of the narrative components in the autobiographical speech of anorexic patients: A statistical and non-linear dynamical analysis. *New Ideas in Psychology*, 26:295–308, 2008.
14. S.P. Strong, R.B. Koberle, R.R. de Ruyter van Steveninck, and W. Bialek. Entropy and information in neural spike trains. *Physical Review Letters*, 80:197–200, 1998.
15. J.M. Amigo, J. Szczepanski, E. Wajnryb, and M.V. Sanchez-Vives. Estimating the entropy rate of spike trains via Lempel-Ziv complexity. *Neural Computation*, 16:717–736, 2004.
16. J.P. Eckmann, S. Kamphorst, and D. Ruelle. Recurrence plots of dynamical systems. *Europhysics Letters*, 5:973–977, 1987.
17. L.L. Trulla, A. Giuliani, J.P Zbilut, and C.L.Jr. Webber. Recurrence quantification analysis of the logistic equation with transients. *Physics Letters A*, 223:255–260, 1996.
18. C.L Webber and J.P. Zbilut. Dynamical assessment of physiologicalsystems and states using recurrence plot strategies. *Journal of Applied Physiology*, 76:965–973, 1994.
19. C.L Webber and J.P. Zbilut. Recurrence quantifications: feature extractions from recurrence plots. *International Journal of Bifurcation and Chaos*, 17:3467–3475, 2007.
20. N. Marwan, M.C. Romano, M. Thiel, and J. Kurths. Recurrence plots for the analysis of complex systems. *Physics Reports*, 438:237–329, 2007.
21. H. Kantz and T. Schreiber. *Nonlinear time series analysis*. Cambridge University Press, Cambridge, 1997.
22. M. Thiel, M.C. Romano, P.L. Read, and J. Kurths. Estimation of dynamical invariants without embedding by recurrence plots. *Chaos*, 14:234–243, 2004.
23. C.S. Daw, C.E.A. Finney, and E.R. Tracy. A review of symbolic analysis of experimental data. *Review of Scientific Instruments*, 74:916–930, 2003.
24. D Lind and B. Marcus. *An introduction to symbolic dynamics and coding*. Cambridge University Press, Cambridge, 1995.
25. J. Guckenheimer and P. Holmes. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*. Springer, Berlin, 1983.
26. R.L. Davidchack, Y.C. Lai, E.M. Bollt, and Dhamala M. Estimating generating partitions of chaotic systems by unstable periodic orbits. *Physical Review E*, 61:1353–1356, 2000.
27. E.M. Bollt, T. Stanford, Y.C. Lai, and K. Zyczkowski. Validity of threshold-crossing analysis of symbolic dynamics from chaotic time series. *Physical Review Letters*, 85:3524–3527, 2000.
28. E.M. Bollt, T. Stanford, Y.C. Lai, and K. Zyczkowski. What symbolic dynamics do we get with a misplaced partition? On the validity of threshold crossings analysis of chaotic time-series. *Physica D*, 154:259–286, 2001.
29. W. Ebeling and G. Nicolis. Word frequency and entropy of symbolic sequences: a dynamical perspective. *Chaos, Solitons and Fractals*, 2:635–650, 1992.
30. H. Herzel and I. Grosse. Measuring correlations in symbol sequences. *Physica A*, 216:518–542, 1995.
31. R. Badii and A. Politi. Thermodynamics and complexity of cellular automata. *Physical Review Letters*, 78:444–447, 1997.
32. P. Castiglione, M. Falcioni, A. Lesne, and A. Vulpiani. *Chaos and coarse-graining in statistical mechanics*. Cambridge University Press, Cambridge, 2008.

33. J. Kurths, A. Voss, A. Witt, P. Saparin, H.J. Kleiner, and N. Wessel. Quantitative analysis of heart rate variability. *Chaos*, 5:88–94, 1995.
34. T. Schürmann and P. Grassberger. Entropy estimation of symbol sequences. *Chaos*, 6:414–427, 1996.
35. A. Lesne, J.L. Blanc, and L. Pezard. Entropy estimation of very short symbolic sequences. *Physical Review E*, 79:046208, 2009.
36. A. Lempel and J. Ziv. On the complexity of finite sequences. *IEEE Transactions on Information Theory*, 22:75–81, 1976.
37. J. Ziv and A. Lempel. A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory*, 23:337–343, 1977.
38. J. Ziv and A. Lempel. Compression of individual sequences by variable rate coding. *IEEE Transactions in Information Theory*, 24:530–536, 1978.
39. P. Faure and A. Lesne. Recurrence plots for symbolic sequences. *International Journal of Bifurcation and Chaos*, 20:1731–1749, 2010.
40. L Breiman. The individual ergodic theorem of information theory. *The Annals of Mathematical Statistics*, 28:809–811, 1957.
41. B. McMillan. The basic theorems of information theory. *The Annals of Mathematical Stastistics*, 24:196–219, 1953.
42. P. Grassberger and I. Procaccia. Computing the Kolmogorov entropy from a chaotic signal. *Physical Review A*, 28:2591–2593, 1983.
43. A Cohen and I. Procaccia. Computing the Kolmogorov entropy from time signals of dissipative and conservative dynamical systems. *Physical Review A*, 31:1872–1882, 1985.
44. J.P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics*, 57:617–656, 1985.
45. P. Faure and H. Korn. A new method to estimate the kolmogorov entropy from recurrence plots: its application to neuronal signals. *Physica D*, 122:265–279, 1998.
46. M. Thiel, M.C. Romano, J. Kurths, R. Meucci, E. Allaria, and T. Arecchi. Influence of observational noise on the recurrence quantification analysis. *Physica D*, 171:138, 2002.
47. C. Letellier. Estimating the Shannon entropy: recurrence plots versus symbolic dynamics. *Physical Review Letters*, 96:254102, 2006.
48. P. Gaspard and X.J. Wang. Noise, chaos, and (ε, τ)-entropy per unit time. *Physics Reports*, 235:291, 1993.
49. M.C. Casdagli. Recurrence plots revisited. *Physica D*, 108:12–44, 1997.
50. M.S. Baptista, E.J. Ngamga, P.R.F. Pinto, M. Brito, and J. Kurths. Kolmogorov-Sinai entropy from recurrence times. *Physics Letters A*, 374:1135–1140, 2010.
51. A.D. Wyner and J. Ziv. Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression. *IEEE Transactions on Information Theory*, 35:1250–1258, 1989.
52. E.J. Ngamga, D.V. Senthilkumar, A. Prasad, P. Parmananda, N. Marwan, and J. Kurths. Distinguishing dynamics using recurrence-time statistics. *Physical Review E*, 85:026217, 2012.