International Journal of Bifurcation and Chaos, Vol. 20, No. 3 (2010) 775–785

DOI: 10.1142/S021812741002606X

GENERALIZED STATISTICAL COMPLEXITY MEASURE

OSVALDO A. ROSSO
Centre for Bioinformatics, Biomarker Discovery and Information-Based Medicine and Hunter Medical Research Institute, School of Electrical Engineering and Computer Science, The University of Newcastle, University Drive, Callaghan, NSW, 2308, Australia

Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Pabellón II, Ciudad Universitaria, 1428 Ciudad Autonoma de Buenos Aires, Argentina

LUCIANA DE MICCO and HILDA A. LARRONDO
Departamentos de Fisica y de Ingenieria Electronica, Facultad de Ingeniería, Universidad Nacional de Mar del Plata, Juan B. Justo 4302, 7600 Mar del Plata, Argentina

MARÍA T. MARTÍN and ANGEL PLASTINO
Instituto de Física-IFLP-CCT La Plata-Conicet, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, C.C. 727, 1900 La Plata, Argentina

Received November 14, 2008; Revised January 27, 2009

A generalized Statistical Complexity Measure (SCM) is a functional that characterizes the probability distribution P associated to the time series generated by a given dynamical system. It quantifies not only randomness but also the presence of correlational structures. We review here several fundamental issues in such a respect, namely, (a) the selection of the information measure $\mathcal{I}$; (b) the choice of the probability metric space and associated distance $\mathcal{D}$; (c) the question of defining the so-called generalized disequilibrium $\mathcal{Q}$; (d) the adequate way of picking up the probability distribution P associated to a dynamical system or time series under study, which is indeed a fundamental problem. In this communication we show (point *d*) that sensible improvements in the final results can be expected if the underlying probability distribution is "extracted" via appropriate consideration regarding causal effects in the system's dynamics.

Keywords: Complexity measure.

1. Statistical Complexity Measures

An information measure $\mathcal{I}$ can primarily be viewed as a quantity that characterizes a given probability distribution. $\mathcal{I}[P]$ is regarded as the measure of the uncertainty associated to the physical processes described by the probability distribution $P = \{p_j, j = 1, \ldots, N\}$, with N the number of possible states of the system under study. If $\mathcal{I}[P] = 0$

we are in a position to predict with certainty which of the N possible outcomes will actually take place. Our knowledge of the underlying process described by the probability distribution is in this instance maximal. On the other hand, we are ignorant of what gets maximal if $\mathcal{I}[P] = \mathcal{I}[P_e] \equiv \mathcal{I}_{\max}$; P_e being the uniform distribution. These two extreme circumstances of (i) maximum foreknowledge ("perfect order") and (ii) maximum ignorance (or maximum "randomness") can, in a sense, be regarded as "trivial" ones. We define for a given probability distribution P and its associate information measure $\mathcal{I}[P]$, an amount of "disorder" H in the fashion

$$\mathcal{H}[P] = \mathcal{I}[P]/\mathcal{I}_{\max}. \tag{1}$$

We have thus $0 \leq \mathcal{H} \leq 1$.

It follows that a definition of *Statistical Complexity Measure* (SCM) must not be made in terms of just "disorder" or "information". A proper SCM needs to adopt some kind of distance $\mathcal{D}$ of the given P to the equilibrium distribution P_e of the accessible states of the system [Lopez-Ruiz *et al.*, 1997; Martín *et al.*, 2003; Lamberti *et al.*, 2004]. This motivates defining the "disequilibrium" as

$$\mathcal{Q}[P] = \mathcal{Q}_0 \cdot \mathcal{D}[P, P_e], \tag{2}$$

where $\mathcal{Q}_0$ is a normalization constant and $0 \leq \mathcal{Q} \leq 1$. The disequilibrium $\mathcal{Q}$ would reflect on the systems "architecture", being different from zero if there exist "privileged", or "more likely" states among the accessible ones. Consequently, we will adopt the following functional product form for the SCM introduced originally by Lopez-Ruiz, Mancini and Calbet (LMC) [Lopez-Ruiz *et al.*, 1997].

$$\mathcal{C}[P] = \mathcal{H}[P] \cdot \mathcal{Q}[P]. \tag{3}$$

This quantity reflects on the delicate interplay extant between the amount of information stored in the system and its disequilibrium.

Following the Shannon–Kinchin paradigm, we define $\mathcal{I}$ in terms of entropies. $\mathcal{H}$ thus refers to different entropic functional forms [Martín *et al.*, 2006; Rosso *et al.*, 2006] and $P \equiv \{p_1, \ldots, p_N\}$ is a discrete distribution:

(a) Shannon, $\mathcal{H}_1^{(S)} = \mathrm{S}_1^{(S)}[P]/\mathrm{S}_1^{(S)}[P_e]$,

$$\mathrm{S}_1^{(S)}[P] = -\sum_{j=1}^{N} p_j \, \ln(p_j); \tag{4}$$

(b) Tsallis, $\mathcal{H}_q^{(T)} = \mathrm{S}_q^{(T)}[P]/\mathrm{S}_q^{(T)}[P_e]$,

$$\mathrm{S}_q^{(T)}[P] = \frac{1}{(q-1)} \sum_{j=1}^{N} [p_j - (p_j)^q]; \tag{5}$$

(c) escort-Tsallis $\mathcal{H}_q^{(G)} = \mathrm{S}_q^{(G)}[P]/\mathrm{S}_q^{(G)}[P_e]$,

$$\mathrm{S}_q^{(G)}[P] = \frac{1}{(q-1)} \left\{ 1 - \left[\sum_{j=1}^{N} (p_j)^{1/q} \right]^{-q} \right\}; \tag{6}$$

(d) Renyi $\mathcal{H}_q^{(R)} = \mathrm{S}_q^{(R)}[P]/\mathrm{S}_q^{(R)}[P_e]$,

$$\mathrm{S}_q^{(R)}[P] = \frac{1}{(1-q)} \ln \left\{ \sum_{j=1}^{N} (p_j)^q \right\}. \tag{7}$$

The quantity q is a so-called deformation parameter ($q = 1$ for Shannon's instance). In the $q \to 1$ limit all the above expressions coincide with that for the Shannon-measure.

As for the metrics and its induced distance $\mathcal{D}$ entering in $\mathcal{Q}$ definition one faces a panoply of choices. For $P_i \equiv \{p_1^{(i)}, \ldots, p_N^{(i)}\}$, with $i = 1, 2$ discrete distributions, we limit our considerations here to [Martín *et al.*, 2006; Rosso *et al.*, 2006]

(a) Euclidean norm $\mathcal{D}_E$ in $\mathbb{R}^N$ [Lopez-Ruiz *et al.*, 1997],

$$\mathcal{D}_E[P_1, P_2] = \|P_1 - P_2\|_E = \sum_{j=1}^{N} \{p_j^{(1)} - p_j^{(2)}\}^2; \tag{8}$$

(b) Wootters's distance $\mathcal{D}_W$ [Martín *et al.*, 2003],

$$\mathcal{D}_W[P_1, P_2] = \cos^{-1} \left\{ \sum_{j=1}^{N} (p_j^{(1)})^{1/2} \cdot (p_j^{(2)})^{1/2} \right\}; \tag{9}$$

(c) relative Kullback entropies $\mathrm{K}_q^{(\kappa)}$ [Rosso *et al.*, 2006],

$$\mathrm{K}_1^{(S)}[P_1|P_2] = \sum_{j=1}^{N} p_j^{(1)} \cdot \ln \left(\frac{p_j^{(1)}}{p_j^{(2)}} \right), \tag{10}$$

$$\mathrm{K}_q^{(T)}[P_1|P_2] = \frac{1}{(q-1)} \sum_{j=1}^{N} (p_j^{(1)})^q \cdot \{(p_j^{(2)})^{1-q} - (p_j^{(1)})^{1-q}\}, \tag{11}$$

$$\mathrm{K}_q^{(G)}[P_1|P_2] = \frac{1}{(q-1)} \sum_{j=1}^{N} \frac{p_j^{(1)}}{(A[P_1])^q} \cdot \left\{ \left[\frac{(p_j^{(2)})^{1/q}}{A[P_2]} \right]^{1-q} - \left[\frac{(p_j^{(1)})^{1/q}}{A[P_1]} \right]^{1-q} \right\}, \quad (12)$$

$$\mathrm{K}_q^{(R)}[P_1|P_2] = \frac{1}{(q-1)} \ln \left\{ \sum_{j=1}^{N} (p_j^{(1)})^q (p_j^{(2)})^{1-q} \right\}, \quad (13)$$

for $q \neq 1$ and $A[P] = \sum_{j=1}^{N} (p_j)^{1/q}$;

(d) Jensen divergences $\mathcal{J}_q^{(\kappa)}$ [Rosso *et al.*, 2006],

$$\mathcal{J}_q^{(\kappa)}[P_1, P_2] = \frac{1}{2}\mathrm{K}_q^{(\kappa)} \left[P_1 \left| \frac{(P_1 + P_2)}{2} \right. \right] + \frac{1}{2}\mathrm{K}_q^{(\kappa)} \left[P_2 \left| \frac{(P_1 + P_2)}{2} \right. \right], \quad (14)$$

with $\kappa =$ Shannon (S) ($q = 1$), Tsallis (T), escort-Tsallis (G) and Renyi (R).

On the basis of the LMC-functional product form we obtain then a family of SCMs for each of four distinct disorder measures and disequilibriums just enumerated, namely,

$$\mathcal{C}_{\nu,q}^{(\kappa)}[P] = \mathcal{H}_q^{(\kappa)}[P] \cdot \mathcal{Q}_q^{(\nu)}[P], \quad (15)$$

with $\kappa =$ S, T, G, R for a fixed q. In Shannon's instance ($\nu =$ S) we have, of course, $q = 1$. The index $\nu = \mathcal{D}_E, \mathcal{D}_W, \mathrm{K}_q^{(\kappa)}, \mathcal{J}_q^{(\kappa)}$ tells us that the disequilibrium is to be evaluated with the appropriate distance measure.

It is important that, for $\nu = \mathrm{K}_q^{(\kappa)}$, the SCM family becomes $\mathcal{C}_q^{(\kappa)}[P] = (1 - \mathcal{H}_q^{(\kappa)}[P]) \cdot \mathcal{H}_q^{(\kappa)}[P]$, which thus becomes the generalized functional form proposed by Shiner *et al.* [1999] for the statistical complexity measure. One could raise the objection that in this case this complexity family becomes just a set of comprising simple functions of the entropy, implying that it might not contain new information vis-a-vis the entropic measure of order. We emphasize that the remaining members of the family $\mathcal{C}_{\nu,q}^{(\kappa)}$ (with $\nu \neq \mathrm{K}_q^{(\kappa)}$) are not just a function of the entropy. On the contrary, for a given $\mathcal{H}_q^{(\kappa)}$-value, an ample range of SCMs can be obtained, from a minimum one $\mathcal{C}_{\min}$ up to a maximal value $\mathcal{C}_{\max}$. Evaluation of $\mathcal{C}_{\nu,q}^{(\kappa)}$ yields, consequently, new information according to the peculiarities of the pertinent probability distribution. A general procedure to obtain the bounds $\mathcal{C}_{\min}$ and $\mathcal{C}_{\max}$ corresponding to the generalized $\mathcal{C} = \mathcal{H} \cdot \mathcal{Q}$ -family is given in [Martín *et al.*, 2006].

In statistical mechanics one is often interested in isolated systems characterized by an initial, arbitrary, and discrete probability distribution. Evolution towards equilibrium is to be described, as the overriding goal. At equilibrium, the distribution is the equiprobability distribution P_e. In order to study the time evolution of the SCM a diagram of $\mathcal{C}$ versus time t can then be used. But, as we know, the second law of thermodynamics states that in isolated system entropy grows monotonically with time ($d\mathcal{H}/dt \geq 0$). This implies that $\mathcal{H}$ can be regarded as an arrow of time, so that an equivalent way to study the time evolution of the SCM is to plot $\mathcal{C}$ versus $\mathcal{H}$. In this way, the normalized entropy-axis substitutes for the time-axis.

If $\kappa =$ S ($q = 1$) and $\nu = \mathcal{D}_E$ we recover the statistical complexity measure definition given originally by Lopez-Ruiz, Mancini and Calbet (LMC) [Lopez-Ruiz *et al.*, 1997], $\mathcal{C}_{\mathrm{LMC}} = \mathcal{C}_{E,1}^{(S)}$. It has been pointed out by Crutchfield and co-workers [Feldman & Crutchfield, 1998] that the LMC measure is marred by some troublesome characteristics that we list below:

- it is neither an intensive nor an extensive quantity.
- it vanishes exponentially in the thermodynamic limit for all one-dimensional, finite-range systems. The above authors forcefully argue that a reasonable SCM should
- be able to distinguish among different degrees of periodicity;
- vanish only for periodicity unity.

Finally, and with reference to the ability of the LMC measure to adequately capture essential dynamical aspects, some difficulties have also been encountered [Anteneodo & Plastino, 1996]. For example, use of the product functional form for the generalized SCM makes it impossible to overcome the second deficiency mentioned above. In previous works we have shown that, after performing some suitable changes in the definition of the disequilibrium (utilization of either Wootters' distance [Martín *et al.*, 2003] or Jensen's divergence

[Lamberti *et al.*, 2004] one is in a position to obtain a generalized SCM that is:

(i) able to grasp essential details of the dynamics,
(ii) an intensive quantity, and
(iii) capable of discerning among different degrees of periodicity.

An important point in the evaluation of the generalized SCM is that of properly determining the underlying probability distribution function (PDF) P (associated to a given dynamical system or time series). This is an often neglected issue that indeed deserves detailed consideration. Why? Because the probability distribution P and the sample space Ω are inextricably linked. Many schemes have been proposed for a proper selection of the probability space (Ω, P). We can mention, among others: (a) procedures based on amplitude statistics [De Micco *et al.*, 2008], (b) binary symbolic dynamics [Mischaikow *et al.*, 1999], (c) Fourier analysis [Powell & Percival, 1979] and, (d) wavelet transform [Rosso & Mairal, 2002]. Their applicability depends on particular characteristics of the data such as stationarity, length of the time series, variation of the parameters, level of noise contamination, etc. In all these cases the global aspects of the dynamics can be somehow captured, but the different approaches are not equivalent in their ability to discern all the relevant physical details. One must also acknowledge the fact that the above techniques are introduced in a *rather ad hoc fashion and are not directly derived from the dynamical properties themselves of the system under study,* as can be adequately achieved, for instance, by recourse to the Bandt–Pompe methodology [Bandt & Pompe, 2002]. This requires suitable partitions of a D-dimensional embedding space that will, it is hoped, reveal important details concerning the ordinal structure of a given one-dimensional time series.

The Bandt–Pompe method for evaluating the probability distribution P is based on the details of the system's attractors-reconstruction procedure. *Causal information* is, consequently, properly incorporated into the "building-up" process that yields (Ω, P). The Bandt–Pompe probability distribution is the only one among those in popular use that takes into account the temporal structure of the time series generated by the physical process under study. A notable result from the Bandt–Pompe approach is a notorious improvement in the performance of the information quantifiers, like entropy and statistical complexity measures, obtained using the probability distribution P generated by their algorithm [Rosso *et al.*, 2007]. Of course, one must assume with Bandt and Pompe that (i) our system is weakly stationary and (ii) enough data are available for a correct attractor reconstruction.

2. Application to Logistic Map

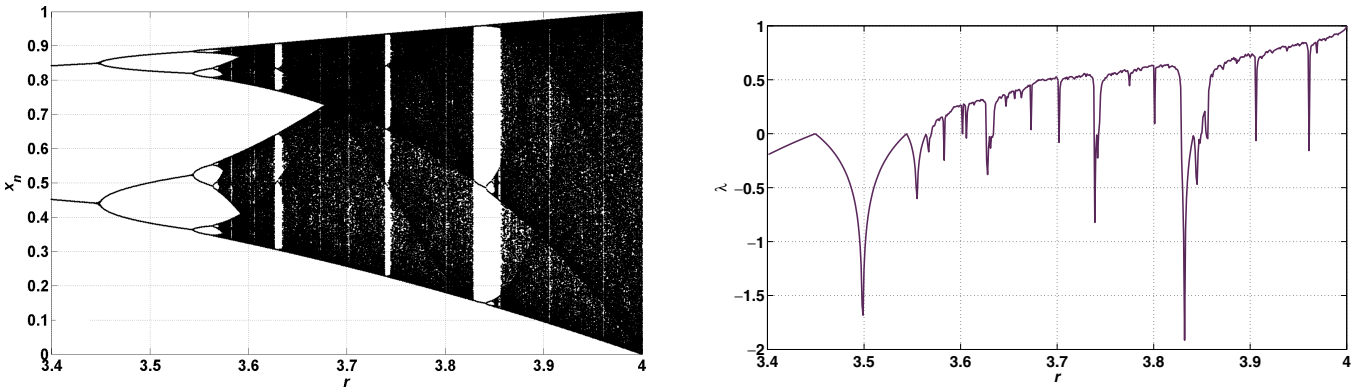
The logistic map constitutes a paradigmatic example, often employed as a testing-ground in order to illustrate new concepts in the treatment of dynamical systems. In such a vein we discuss here the application of the generalized statistical complexity measures (see Eq. (15)).

We deal then with the map $F : x_n \to x_{n+1}$ [Sprott, 2004; Ott *et al.*, 1994], described by the ecologically motivated, dissipative system described by the first order difference equation

$$x_{n+1} = r \cdot x_n \cdot (1 - x_n) \tag{16}$$

with $0 \leq x_n \leq 1$ and $0 < r \leq 4$. Figure 1(a) depicts the well-known bifurcation diagram for the logistic map for $3.4 \leq r \leq 4.0$ while, in Fig. 1(b), the corresponding Lyapunov exponent, λ, is also shown.

Let us briefly review, with reference to Figs. 1(a) and 1(b), some exceedingly well known results for this map that we need in order to put into an appropriate perspective the properties of our family of generalized statistical complexity measures. For values of the control parameter $1 < r < 3$ there exists only a single steady-state solution. Increasing the control parameter past $r = 3$ forces the system to undergo a period-doubling bifurcation. Cycles of period $8, 16, 32, \ldots$ occur and, if r_n denotes the value of r where a 2^n cycle first appears, the r_n converge to a limiting value $r_\infty \cong 3.5699456$ [Sprott, 2004; Ott *et al.*, 1994]. As r grows some more, a quite rich, and well-known structure arises. In order to be in a position to better appreciate at once the long-term behavior for all values of r lying between 3.4 and 4.0, we plot the pertinent bifurcation diagram in Fig. 1(a). We immediately note there the cascade of further period-doubling that occurs as r increases, until, at r_∞, the maps become chaotic and the attractors change from comprising a finite set of points to becoming an infinite set. For $r > r_\infty$ the orbit-diagram reveals an "strange" mixture of order and chaos. The large window beginning near $r = 3.8284$ contains a stable period-3 cycle. In Fig. 1(b) we see that the nonzero Lyapunov characteristic exponent λ remains negative



(a) (b)

Fig. 1. (a) Bifurcation diagram and (b) Lyapunov exponent (λ), for the logistic map as function of parameter r with step $\Delta r = 0.0003$. Time series with total length $M = 10^7$ data.

for $r < r_\infty$. We notice that λ approaches zero at the period-doubling bifurcation. The onset of chaos is apparent near $r \cong 3.5699$, where λ first becomes positive. As stated above, for $r > r_\infty$ the Lyapunov exponent increases globally, except for the dips one sees in the windows of periodic behavior. Notice the particularly large dip due to the period-3 window near $r = 3.8284$.

3. Temporal Information and Methodologies for Getting the Pertinent PDFs

In the present work, we compare the results of using Information Theory quantifiers (Entropy and Generalized Statistical Complexity) after a proper evaluation of the all important probability distribution function (PDF) is made by recourse to one of the three distinct approaches in coordinate space, namely, (i) histograms of the x_i-values, (ii) binary representations, and (iii) the Bandt–Pompe technique; the temporal information-content of each being discussed. Note also, that if the PDF evaluation is made in the frequency space, similar results are obtained (not shown). If the PDF is obtained by recourse of the Fast Fourier Transform (FFT), no temporal information is included. Contrary, in the case of Discrete Wavelet Transform used for determination of the PDF, frequency and temporal information are included in the standard way (for details see [Rosso *et al.*, 2006; Rosso & Mairal, 2002]).

3.1. *PDF based on histograms*

In order to extract a PDF via amplitude-statistics, divide first the interval $[0, 1]$ into a finite number *nbin* of nonoverlapping subintervals A_i: $[0, 1] = \bigcup_{i=1}^{\text{nbin}} A_i$ and $A_i \bigcap A_j = \emptyset \, \forall \, i \neq j$. One then employs the usual histogram-method, based on counting the relative frequencies of the time series values within each subinterval. It should be clear that the resulting PDF lacks information regarding temporal evolution. The only pieces of information we have here are the x_i-values that allow one to assign inclusion within a given bin, ignoring just where they are located (this is, the subindex i).

3.2. *PDF based on binary representation*

Following the original work of Lopez-Ruiz *et al.* [1997], for each parameter value, r, the dynamics of the logistic map was reduced to a binary sequence (0 if $x \leq (1/2)$; 1 if $x > (1/2)$) and binary strings of length $L = 12$ without overlap were considered as states of the system. The concomitant probabilities are assigned according to the frequency of occurrence after running over 10^7 iterations.

In this case some temporal information is, on average, retained. Why? Because we face a two-step procedure in this approach. In making the binary assignment (1st step), part of the temporal information is lost by averaging, but, in considering words of length L (without overlap) in a second step we are indeed respecting such "averaged" temporal information.

3.3. *PDF based on Bandt and Pompe methodology*

To use the Bandt and Pompe [2002] methodology for evaluating the probability distribution P associated to the time series (dynamical system)

under study, one starts by considering partitions of the pertinent D-dimensional space that will hopefully "reveal" relevant details of the ordinal-structure of a given one-dimensional time series $\{x_t : t = 1, \ldots, M\}$ with embedding dimension $D > 1$. We are interested in "ordinal patterns" of order D [Bandt & Pompe, 2002; Keller & Sinn, 2005] generated by

$$(s) \mapsto \left(x_{s-(D-1)}, x_{s-(D-2)}, \ldots, x_{s-1}, x_s\right), \quad (17)$$

which assign to each time s the D-dimensional vector of values at times $s, s-1, \ldots, s-(D-1)$. Clearly, the greater the D-value, the more information on the past is incorporated into our vectors. By the "ordinal pattern" related to the time (s) we mean the permutation $\pi = (r_0, r_1, \ldots, r_{D-1})$ of $(0, 1, \ldots, D-1)$ defined by

$$x_{s-r_{D-1}} \leq x_{s-r_{D-2}} \leq \cdots \leq x_{s-r_1} \leq x_{s-r_0}. \quad (18)$$

In order to get a unique result we set $r_i < r_{i-1}$ if $x_{s-r_i} = x_{s-r_{i-1}}$. Thus, for all $D!$ possible permutations π of order D, the probability distribution $P = \{p(\pi)\}$ is defined by

$$p(\pi) = \frac{\sharp\{s | s \leq M - D + 1; (s), \texttt{ has type } \pi\}}{M - D + 1}. \quad (19)$$

In this expression, the symbol $\sharp$ stands for "number".

The Bandt–Pompes methodology is not restricted to time series representative of low-dimensional dynamical systems but can be applied to any type of time series (regular, chaotic, noisy, or reality based), with a very weak stationary assumption (for $k = D$, the probability for $x_t < x_{t+k}$ should not depend on t [Bandt & Pompe, 2002]). One also assumes that enough data are available for a correct attractor-reconstruction. Of course, the embedding dimension D plays an important role in the evaluation of the appropriate probability distribution because D determines the number of accessible states $D!$ Also, it conditions the minimum acceptable length $M \gg D!$ of the time series that one needs in order to work with reliable statistics.

4. Results and Discussion

Time series with total length $M = 10^7$ data were generated for the logistic map for parameter value r in the range $3.4 \leq r \leq 4.0$ with step of $\Delta r = 0.0003$. For the evaluation of PDF-histogram we consider $N = \text{nbin} = 2^{12}$ bins [De Micco *et al.*, 2008]. In the case of PDF-binary the word length was fixed to $L = 12$ bits ($N = 2^{12}$) and for the case of PDF-Bandt and Pompe, the embedding dimension was fixed to $D = 6$ ($N = 6! = 720$).

The behavior ("degree of chaoticity") of the Lyapunov exponent λ as a function of the parameter r is displayed in Fig. 1(b). From this figure, we see that λ and, as a result, the associated degree of chaoticity grows globally with r (since there are many periodic windows where λ drops to negative values) , reaching a maximum at $r = 4$. One would expect that a sensible statistical complexity measure should accompany such a global growth. In other words, a reasonable statistical complexity measure should take very small values for $r < r_\infty$ and then grow globally together with the degree of chaoticity.

In Fig. 2 we depict the normalized Shannon entropy, $\mathcal{H}_1^{(S)}$ evaluated for the logistic map as a function of the parameter r considering either (i) an histogram-determined PDF ($\mathcal{H}_{\text{hist}} \equiv \mathcal{H}_1^{(S)}|_{\text{histogram}}$), (ii) a binary ($\mathcal{H}_{\text{bin}} \equiv \mathcal{H}_1^{(S)}|_{\text{binary}}$) one or, (iii) a Bandt–Pompe PDF ($\mathcal{H}_{\text{BP}} \equiv \mathcal{H}_1^{(S)}|_{\text{Bandt-Pompe}}$). We observe in all instances an abrupt entropy growth around $r > r_\infty \cong 3.5699$. After we pass this point, the entropy displays a trend to increase, taking its maximum value at $r = 4$. The several "drops" in entropic values in the interval $r_\infty < r < 4$ correspond to the periodic windows, as can be easily confirmed comparing with the bifurcation diagram and with the Lyapunov exponent depicted in Fig. 1.

It is interesting to observe that, in the PDF-histogram-instance, the entropy for $r = 4$ is almost unity ($H_{\text{hist}} \simeq 0.983$). The reason is that the invariant measure of the logistic map is in this case given by an almost constant function, namely,

$$\rho(x) = \begin{cases} \dfrac{1}{\pi\sqrt{x(1-x)}} & 0 < x < 1 \\ 0 & \text{other } x \end{cases}. \quad (20)$$

Results with a PDF-binary are even worse ($H_{\text{bin}} = 1$), because the invariant measure of Eq. (20) is symmetric around $x = 0.5$. Only the PDF-Bandt and Pompe leads to the realistic value $H_{\text{BP}} \simeq 0.6$. Furthermore, the doubling period-cascade is lost if a PDF-binary is used.

We depict in Fig. 3 the statistical complexity measure $\mathcal{C}_{\mathcal{J},1}^{(S)}$ as evaluated with the three different PDF-methodologies outlined above. In all cases

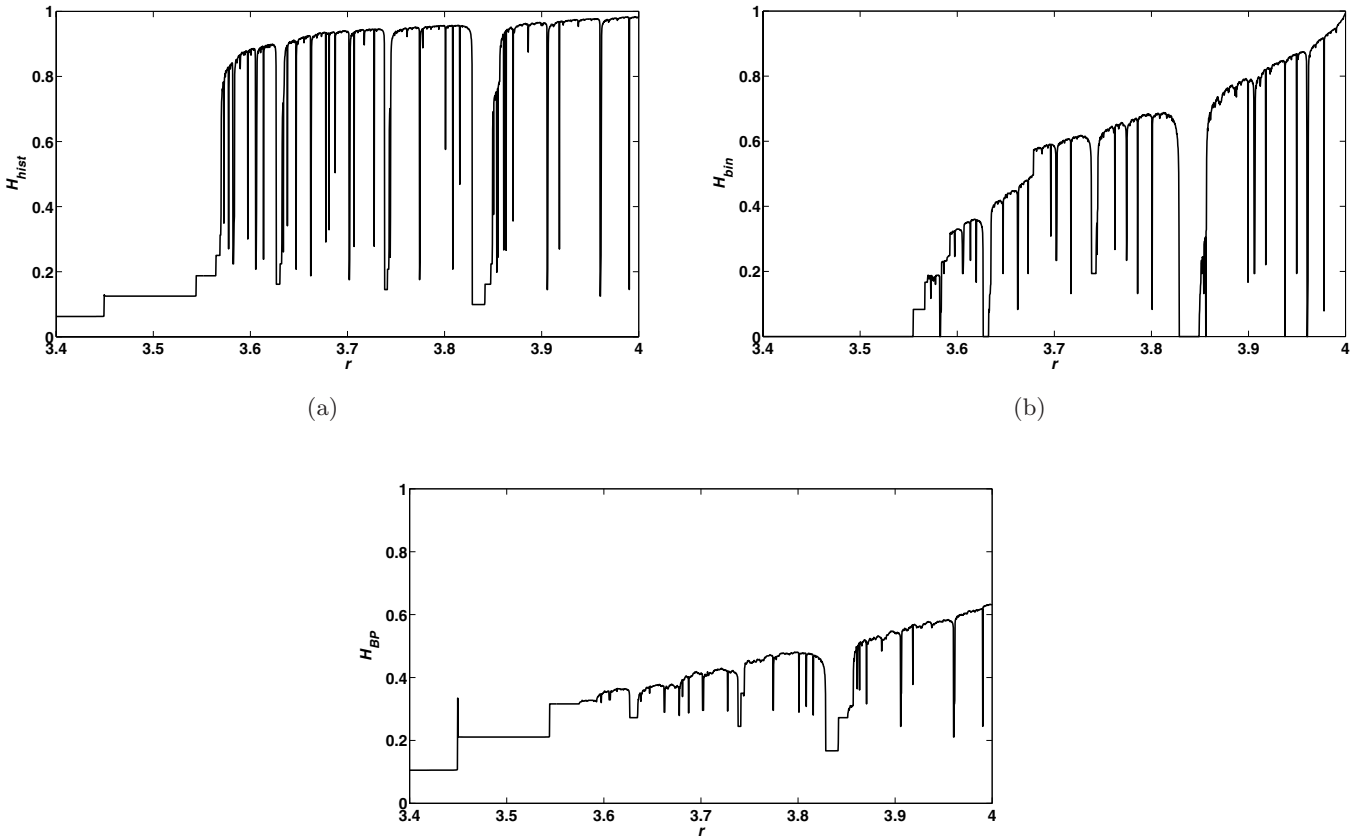


Fig. 2. Normalized Shannon entropy for the logistic map as function of parameter r with step $\Delta r = 0.0003$ (time series with total length $M = 10^7$ data) evaluated in the following instances: (a) PDF-histogram ($N_{\rm bin} = 2^{12}$), (b) PDF-binary ($L = 12$), (c) PDF-Bandt and Pompe ($D = 6$).

we note an abrupt statistical complexity growth around $r > r_\infty$. After we pass this point, the corresponding complexity measure behaves in quite different fashion. Notwithstanding the fact that in all instances the complexity measure decreases, it remains constant within the periodic windows (consider for example the period three window $r \in [3.8284, 3.8570]$). Also, for different periodic windows the complexity values differ according to the kind of periodicity-degree of that window. For the statistical complexity evaluated using PDF-histograms one sees that, after $r > r_\infty$, an overall decreasing tendency becomes evident. A minimum value exists at $\mathcal{C}_{\rm hist} \equiv \mathcal{C}^{(S)}_{\mathcal{J},1}|_{\rm histogram} \cong 0$ for $r = 4$. The several peaks observed in the region $r_\infty < r < 4$ signal a local complexity growth. Comparison with the bifurcation diagram (see Fig. 1) indicates that these peaks are correlated with periodic windows. As a consequence, they signal the transition between different dynamical behaviors, i.e. from chaotic to periodic ones.

The statistical complexity evaluated with a PDF-binary exhibits a parabolic behavior according to the variation of the parameter $r_\infty < r < 4$, with a value $\mathcal{C}_{\rm bin} \equiv \mathcal{C}^{(S)}_{\mathcal{J},1}|_{\rm binary} \cong 0$ for $r = 4$. Globally, the decay of $\mathcal{C}_{\rm bin}$ is more swift the closer we approach the value $r = 4$. Note also that for this region the statistical complexity grows in the inter-windows region and rapidly falls within the periodic windows.

In the instance of a PDF evaluated with the Bandt and Pompe methodology, the statistical complexity measure $\mathcal{C}_{\rm BP} \equiv \mathcal{C}^{(S)}_{\mathcal{J},1}|_{\rm Bandt-Pompe}$ exhibits overall increasing values as a function of the parameter r, adopting its maximum value for $r = 4$, which corresponds to the case of totally developed chaos. Also we observe drops in its values, that are associated with periodic windows. The behavior of the normalized Shannon entropy ($\mathcal{H}_{\rm BP}$) and of the statistical complexity measure ($\mathcal{C}_{\rm BP}$) take into account in a natural way, as emphasized above, "time causality", which allows in fact for the

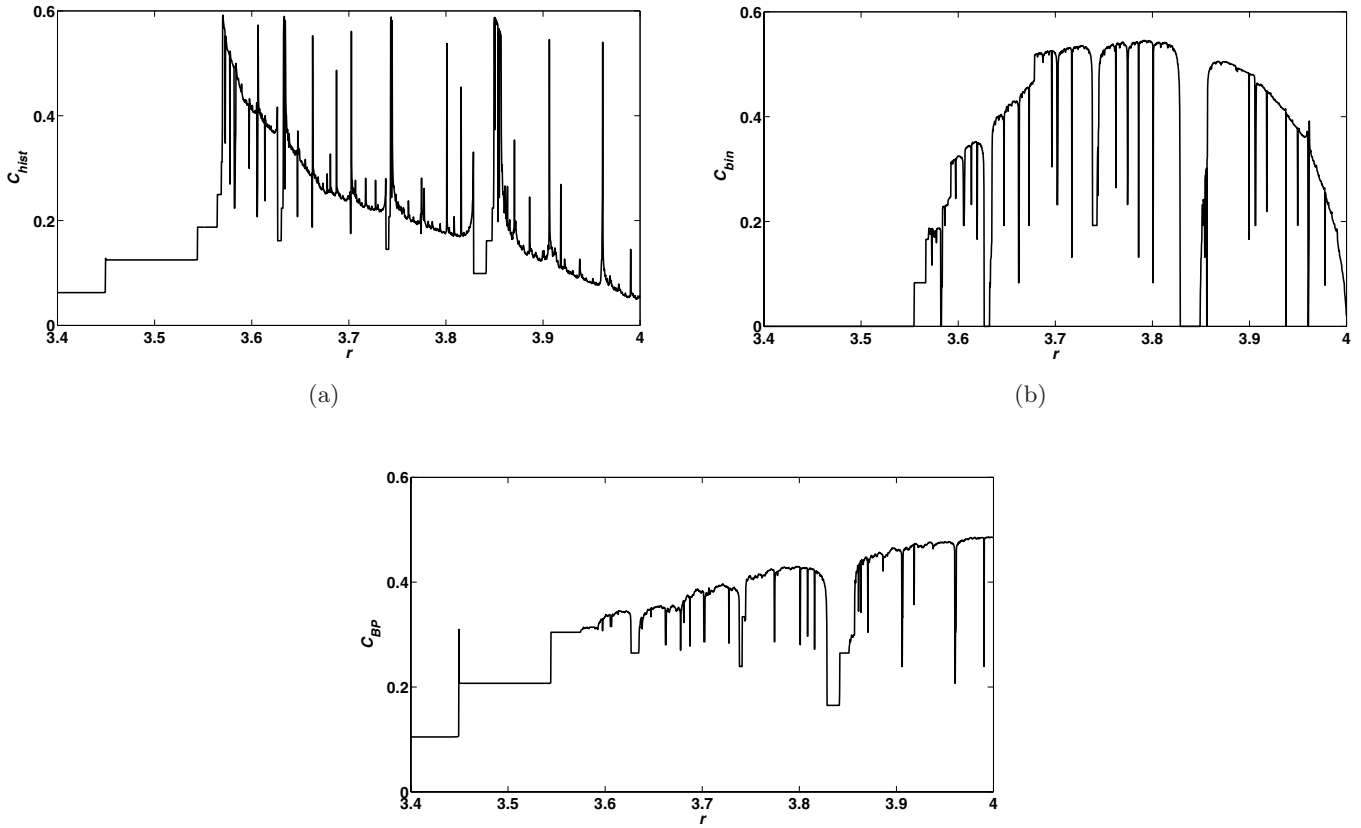


Fig. 3. Statistical complexity for the logistic map as function of parameter r with step $\Delta r = 0.0003$ (time series with total length $M = 10^7$ data) evaluated in the following instances: (a) PDF-histogram ($N_{\rm bin} = 2^{12}$), (b) PDF-binary ($L = 12$), (c) PDF-Bandt and Pompe ($D = 6$).

ability to distinguish between chaotic and stochastic dynamics [Rosso *et al.*, 2007]. Indeed, if the time causality is not totally taken into account, as is the case of both the PDF-histogram and the PDF-binary, one has $\mathcal{H} \cong 1$ and $\mathcal{C} \cong 0$ for both kinds of dynamics, in contradiction to what happens using the Bandt and Pompe methodology to determine the PDF, for which the situation $\mathcal{H} \cong 1$ together with $\mathcal{C} \cong 0$ is only reached for the stochastic (noise) dynamics [Rosso *et al.*, 2007]. In plain words, "chaos is not noise" even if they share some common characteristics. In fact, chaos is representative of deterministic processes, and thus time-causality constitutes an important facet that must be taken into account for a proper characterization.

The above described special characteristics of the PDF obtained using Bandt and Pompe's methodology can provide deeper insight if we consider the emergence of the so-called "forbidden patterns" [Amigó *et al.*, 2007; Zanin, 2008]. When dealing with random time series, every permutation pattern should have the same probability of appearance. Therefore, when $N \to \infty$, the pertinent PDF should be a uniform distribution. However, as shown by Amigó *et al.* [2007] not all ordering patterns can be effectively materialized into orbits for a given one-dimensional map, which, in a sense, makes these "virtual" patterns "forbidden". Even worse, the existence of these "forbidden" ordering patterns is a persistent dynamical feature (always!). Since a truly random dynamics cannot, obviously, have forbidden patterns, one can draw the interesting conclusion that their existence is indeed an indicator of deterministic orbit generation [Amigó *et al.*, 2007; Zanin, 2008]. In Fig. 4, we depict the number of forbidden patterns ($D = 6$) for the logistic map as a function of the parameter r within the range $3.4 \leq r \leq 4$. See the decreasing tendency in the number of forbidden patterns as a function of r, with a minimum value at $r = 4$. The fluctuation in the values of forbidden patterns for the different periodic windows and chaotic zones can be associated with the numerical precision attained in the evaluation of Eq. (18), which defines not only the

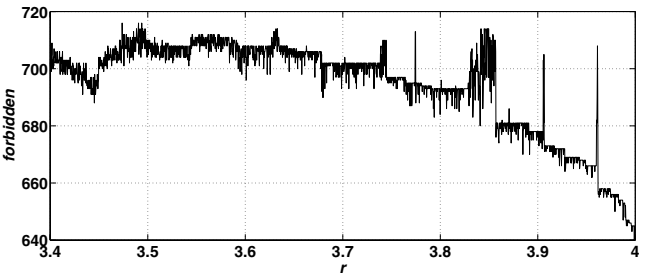
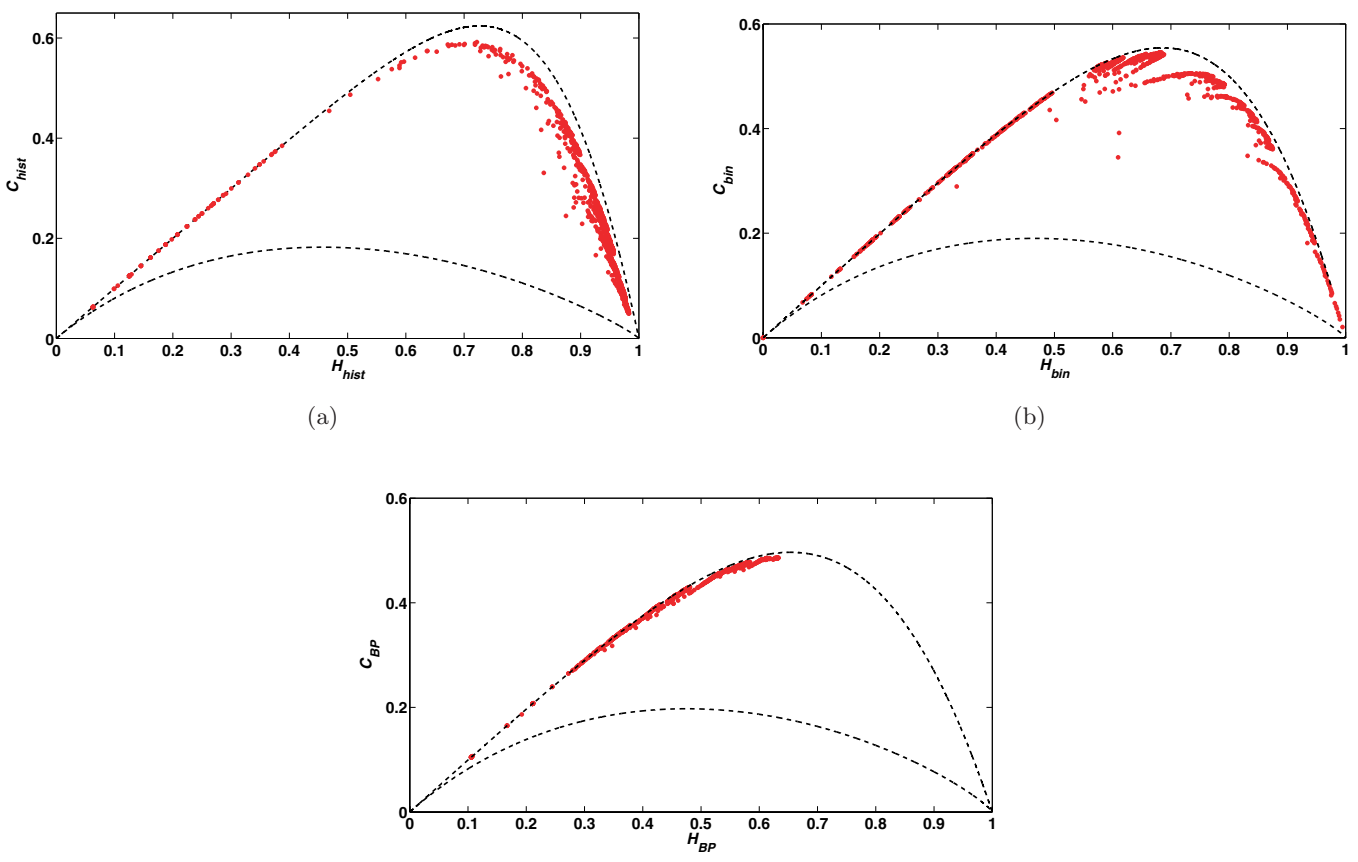


Fig. 4. Forbidden patterns for the logistic map as function of parameter r with step $\Delta r = 0.0003$ (time series with total length $M = 10^7$ data) as function of parameter r. Embedding dimension $D = 6$.

patterns but also the finite number of realization-data one considers.

Finally, we display in Fig. 5 the $\mathcal{H} \times \mathcal{C}$-plane for the three different instances of PDF-evaluation under consideration, in the r-range $3.4 \leq r \leq 4$ (the control parameter does not explicitly appear in the graph, of course). The two continuous curves represent, respectively, the maximum, $\mathcal{C}_{\max}$, and minimum $\mathcal{C}_{\min}$, statistical complexity values evaluated as explained in [Martín *et al.*, 2006] for the corresponding N (number of degrees of freedom). Note that, for the case of periodic windows, if $\mathcal{H} < \mathcal{H}^*$ ($\mathcal{H}^*_{\text{hist}} \approx 0.4$, $\mathcal{H}^*_{\text{bin}} \approx 0.3$ and $\mathcal{H}^*_{\text{BP}} \approx 0.25$) we can ascertain that the Lyapunov exponent $\lambda < 0$, while for $\mathcal{H} > \mathcal{H}^*$ we observed that $\lambda > 0$, which entails chaotic behavior. As evidenced by Fig. 5, in all the instances of PDF-evaluation, periodic behaviors exhibit lower values than chaotic ones,



(a) (b)

(c)

Fig. 5. Entropy-complexity plane for the logistic map (parameter r with step $\Delta r = 0.0003$, time series with total length $M = 10^7$ data) for: (a) PDF-histogram ($N_{\text{bin}} = 2^{12}$), (b) PDF-binary ($L = 12$), (c) PDF-Bandt and Pompe ($D = 6$). We also display the maximum and minimum possible values of the statistical complexity (segmented lines).

as demanded in [Feldman & Crutchfield, 1998]. The "Lyapunov"-based criteria mentioned above is satisfied by the three PDF evaluation instances as well. For a physicist, the Jensen–Shannon statistical complexity measure evaluated with the Bandt and Pompe PDF is the best one because it is an intensive quantity (which is not the case when the disequilibrium is evaluated in terms of Wootters distance) and also clearly distinguishes deterministic (chaos) from stochastic process [Rosso *et al.*, 2007]. The importance of the $\mathcal{H} \times \mathcal{C}$-plane [Martín *et al.*, 2006] resides in the fact that this kind of diagrams yield information of a system independently of the values that the different control parameters may adopt. The bounds yield also information that depends on the particular characteristics of a given system (for instance, the existence of global extrema), or on the peculiarities of the system's internal configuration for which such extrema can be obtained.

Similar results, to the previously described are obtained when the other functional forms of entropy and generalized statistical complexity are considered. Results for the logistic map considering PDF-binary as function of the variation of the parameter r for Tsallis and Renyi were presented in [Martín *et al.*, 2006].

5. Conclusions

We conclude that the following remarks are in order when a statistical complexity measure (SCM) is going to be used for characterizing a time series:

- Special care is required in the selection of the probability distribution function (PDF). If determinism is an important feature to be taken into account, the Bandt and Pompe prescription has advantages over other choices.
- A statistical complexity measure depends also on the selection of an appropriate disequilibrium $\mathcal{Q}$. A complete discussion of this issue is not given in this paper, but previous works support the choice of the Jensen Shannon Divergence as that having the best properties [Lamberti *et al.*, 2004].
- Chaotic time series are located near the $\mathcal{C}_{\max}$ curve in any of the $\mathcal{H} \times \mathcal{C}$ representations planes. But the Bandt and Pompe prescription locates chaos very near from the top, with high complexity and $\mathcal{H}$-values near 0.5. The other "noncausal" prescriptions place deterministic systems close to truly random ones (near $\mathcal{C} = 0$ and $\mathcal{H} = 1$).
- The number of forbidden ordering patterns is another quantifier that yields deterministic behavior in the full chaos case $r = 4$. However, it does not exhibit the bifurcation diagram in as good a fashion as $\mathcal{C}_{\mathrm{BP}}$ does.

Acknowledgments

This work was partially supported by the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina (PIP 5687/05, PIP 6036/05) and ANPCyT, Argentina (PICT 11-21409/04). O. A. Rosso gratefully acknowledges support from Australian Research Council (ARC) Centre of Excellence in Bioinformatics and School of Electrical Engineering and Computer Science, The University of Newcastle, Australia. O. A. Rosso is very grateful to the organizers of the workshop, in particular, to Prof. Dr. Hector Mancini, for their very kind hospitality during his stay in Pamplona, Spain.

References

Amigó, J. M., Zambrano, S. & Sanjúan, M. A. F. [2007] "True and false forbidden patterns in deterministic and random dynamics," *Europhys. Lett.* **79**, 50001.

Anteneodo, C. & Plastino, A. R. [1996] "Some features of the Lopez-Ruiz–Mancini–Calbet (LMC) statistical measure of complexity," *Phys. Lett. A* **223**, 348-354.

Bandt, C. & Pompe, B. [2002] "Permutation entropy: A natural complexity measure for time series," *Phys. Rev. Lett.* **88**, 174102.

De Micco, L., Gonzalez, C. M., Larrondo, H. A., Martín, M. T., Plastino, A. & Rosso, O. A. [2008] "Randomizing nonlinear maps via symbolic dynamics," *Physica A* **387**, 3373–3383.

Feldman, D. P. & Crutchfield, J. P. [1998] "Measures of statistical complexity: Why?" *Phys. Lett. A* **238**, 244–252.

Keller, K. & Sinn, M. [2005] "Ordinal analysis of time series," *Physica A* **356**, 114–120.

Lamberti, P. W., Martín, M. T., Plastino, A. & Rosso, O. A. [2004] "Intensive entropic non-triviality measure," *Physica A* **334**, 119–131.

Lopez-Ruiz, R., Mancini, H. L. & Calbet, X. [1997] "A statistical measure of complexity," *Phys. Lett. A* **209**, 321–326.

Martín, M. T., Plastino, A. & Rosso, O. A. [2003] "Statistical complexity and disequilibrium," *Phys. Lett. A* **311**, 126–132.

Martín, M. T., Plastino, A. & Rosso, O. A. [2006] "Generalized statistical complexity measures: Geometrical and analytical properties," *Physica A* **369**, 439–462.

Mischaikow, K., Mrozek, M., Reiss, J. & Szymczak, A. [1999] "Construction of symbolic dynamics from experimental time series," *Phys. Rev. Lett.* **82**, 1114–1147.

Ott, E., Sauer, T. & Yorke, J. A. [1994] *Coping with Chaos* (Wiley, NY).

Powell, G. E. & Percival, I. C. [1979] "A spectral entropy method for distinguishing regular and irregular motion of hamiltonian systems," *J. Phys. A: Math. Gen.* **12**, 2053–2071.

Rosso, O. A. & Mairal, M. L. [2002] "Characterization of time dynamical evolution of electroencephalographic records," *Physica A* **312**, 469–504.

Rosso, O. A., Martín, M. T., Figliola, A., Keller, K. & Plastino, A. [2006] "EEG analysis using wavelet-based information tools," *J. Neurosci. Meth.* **153**, 163–182.

Rosso, O. A., Larrondo, H. L., Martín, M. T., Plastino & Fuentes, M. A. [2007] "Distinguishing noise from chaos," *Phys. Rev. Lett.* **99**, 154102.

Shiner, J. S., Davison, M. & Landsberg, P. T. [1999] "Simple measure for complexity," *Phys. Rev. E* **59**, 1459–1464.

Sprott, J. C. [2004] *Chaos and Time Series Analysis* (Oxford University Press, Oxford).

Zanin, M. [2008] "Forbidden patterns in financial time series," *Chaos* **18**, 013119.