International Journal of Bifurcation and Chaos, Vol. 20, No. 6 (2010) 1731–1749
© World Scientific Publishing Company
DOI: 10.1142/S0218127410026794

RECURRENCE PLOTS FOR SYMBOLIC SEQUENCES

PHILIPPE FAURE
UMR 7102 Neurobiologie des Processus Adaptatifs,
Équipe Neurophysiologie et Comportement Boîte 14,
Université Pierre et Marie Curie, 9 quai St Bernard,
75005, Paris, France
philippe.faure@snv.jussieu.fr

ANNICK LESNE
Institut des Hautes Études Scientifiques
Le Bois-Marie, 35 route de Chartres, 91440,
Bures-sur-Yvette, France

Laboratoire de Physique Théorique de la Matière Condensée UMR 7600,
Boîte 121, Université Pierre et Marie Curie, 4 place Jussieu,
75252 Paris Cedex 05, France
lesne@ihes.fr

Received November 13, 2008; Revised July 17, 2009

This paper introduces an extension of recurrence analysis to symbolic sequences. Heuristic arguments based on Shannon–McMillan–Breiman theorem suggest several relations between the statistical features of "symbolic" recurrence plots and the entropy per unit time of the dynamics; their practical efficiency for experimental sequences of finite length is checked numerically on two paradigmatic models, namely discretized logistic map trajectories and Markov chains, and also on experimental behavioral sequences. Specific advantages of RP analysis are presented, among which is the detection of nonstationary features.

Keywords: Recurrence plots; symbolic sequences; entropy rate; Shannon–McMillan–Breiman theorem; behavioral sequences.

1. Introduction

Recurrence plots (RPs) have been introduced by Eckmann *et al.* [1987] for continuous-valued time series to visually represent and investigate a key feature of their underlying nonlinear dynamics, that is, recurrence; they are now an acknowledged tool for nonlinear time series analysis [Marwan *et al.*, 2007]. On the other hand, in several contexts, the relevant time series under consideration are symbolic sequences originating either from a dynamical system with intrinsically discrete states or from a proper encoding that turns the original continuous-valued trajectories into discrete ones [Badii & Politi, 1999; Daw *et al.*, 2003].

We here propose to extend RPs to symbolic analysis. Application of RPs in a symbolic context has already been developed for analyzing the recurrence of order patterns, namely the symbolic encoding of a continuous-valued trajectory $(x_t)_{t\geq 0}$ according to the ordering of a few successive terms, e.g. (x_t, x_{t+1}, x_{t+2}); this approach is very efficient for quantifying dynamic features of bivariate time series (especially when the two components of the bivariate series take their values in different spaces and are not directly comparable) [Bandt *et al.*,

2008; Groth, 2005]. We shall here consider more general symbolic sequences. The methodology that we develop here is in particular motivated by behavioral studies, where encoding the observed behavior (e.g. video-recorded behavior) into a symbolic sequence is an acknowledged approach allowing to prune irrelevant information and to alleviate the impact of individual variability [Faure *et al.*, 2003; Maubourguet *et al.*, 2008]. Symbolic sequences are also currently used in case of short time series, where turning to a symbolic sequence by a suitable partition of the phase space is a first means to weaken finite-size effects and extract statistically meaningful features [Lesne *et al.*, 2009]. In such symbolic frameworks, the system state is described in a coarser way and so is the notion of recurrence; in counterpart, a recurrence appears now to be a clear-cut event, namely the identity of two sequence stretches ("words") hence involving no arbitrary closeness threshold while introducing additional symmetry properties. We show several fruitful outcomes allowed by this approach, e.g. investigation of the recurrence of patterns ("words") through higher-order RPs, dissection of the contribution of each specific pattern through analysis of partial RPs, estimation of the entropy per unit time of a stationary source, and analysis of nonstationary dynamics.

The ensuing "symbolic" RPs are closely related to word statistics and as such, their interpretation could be based on information-theoretic results. In particular, we derive an entropy estimator related to the cumulative length distribution of diagonal lines, appearing as the discrete analogue of the acknowledged entropy estimation method from RPs in a continuous phase space [Faure & Korn, 1998; Marwan *et al.*, 2007] and possibly circumventing the statistical stationarity required in standard symbolic methods of entropy estimation [Lesne *et al.*, 2009; Letellier, 2006] (Sec. 2). Nevertheless, these analytical results rely on limit theorems valid for infinite-length sequences and words, hence their applicability for practical purposes is questionable. Accordingly, we have implemented numerical simulations in order to appreciate finite-size effects and check the accuracy and practical efficiency of the proposed entropy estimation procedure. We thus investigated two paradigmatic models, logistic map trajectories possibly spoiled by experimental noise or affected with a slow drift in the dynamical parameter and Markov chains with tunable characteristic times (Sec. 3). Technical details on numerical simulations are given in an Appendix. We also investigated the efficiency of symbolic RPs and associated entropy estimator for analyzing experimental data, namely behavioral sequences encoding mouse displacement in an open-field (Sec. 4). We conclude with some perspectives for exploiting additional features of symbolic RPs and widening the scope of their applications (Sec. 5).

2. Theoretical Basis

2.1. *General properties of symbolic RPs*

2.1.1. *Definition*

Data are given under the form of a symbolic sequence $\overline{x} = (x_i)_{i=1,\ldots,N}$ of length N, written with an alphabet of k-symbols. The discrete analogue of m-embedding is to consider m-words, denoted $w_{m,i} = [x_i, \ldots, x_{i+m-1}]$. (We do not speak of m as a delay nor refer to embedding [Marwan *et al.*, 2007] because our methodology also intends to deal with intrinsically symbolic sequences, e.g. genomic sequences or behavioral sequences.) A dot (i, j) is present in the recurrence plot relative to m-words (henceforth termed m-RP) if and only if m-words starting at times i and j coincide: $w_{m,i} = w_{m,j}$. A recurrence (i, j) is thus an intrinsic feature (i.e. not depending on some observer choices) of the symbolic sequence at the given time scale m. The number of m-words that can be extracted from the sequence is $N - m + 1$, hence the m-RP is a square matrix of size $(N - m + 1) \times (N - m + 1)$. In 1-RPs, at most k nonidentical column patterns can be seen, each associated to a symbol x; the number of dots in a column associated with x gives the symbol occurrence N_x in the sequence, and N_x/N provides an unbiased estimate of the symbol frequency $p_1(x)$. In 2-RPs, there is at most k^2 nonidentical column patterns, each associated to a 2-word xy; the number of dots in a column associated with xy gives the word occurrence N_{xy} in the sequence, and N_{xy}/N_x provides an unbiased estimate of the transition probability from x to y while $N_{xy}/(N - 1)$ estimates the word probability $p_2(xy)$. In fact, some of these 2-words might be absent if they are forbidden or of very low probability. We shall see below (Sec. 2.2.4) a quantitative statement about the absence of some m-words in a m-RP, all the more valid since m is large, and making the number of actually observed m-words far lower than its upper bound k^m. Note that in experimental

reconstruction and exploitation of RPs, an additional reduction is observed, namely the absence of some words due to undersampling, that will be all the more stringent that m is small.

All RPs are obviously symmetric with respect to the (fully occupied) diagonal $\{(i,i)\}$. But discrete RPs exhibit more systematic symmetries than their continuous counterparts due to the exactness of recurrences: indeed if (i,j_1) and (i,j_2) are recurrences, then (j_1,j_2) is also a recurrence. Lines and columns associated to the same word w are identical, so that the contribution (if any) of the m-word w to the m-RP (what we shall term the partial $\mathrm{RP}_m(w)$) has a square lattice motif. This very strong symmetry property can be used in "error correction", since a line and column sharing a point (i,j) should be identical, unless this recurrence is spurious, leading either to eliminate the recurrence (i,j) or to correct the line j or column i.

2.1.2. *Recurrence of specific words and partial RPs*

A recurrence plot for m-words is the superposition of disjoint recurrence plots $\mathrm{RP}_m(w)$, one for each possible m-word w. The idea is to dissect the respective contribution of each word to the complete RP_m (see for instance the implementation illustrated in Figs. 5(d), 5(e) and 6(f)). Indeed, each of these "partial" recurrence plots $\mathrm{RP}_m(w)$ gives access to the statistics of recurrence time of w (e.g. its minimum, maximium, mean, standard deviation, median, and cumulative distribution). It has a quite trivial structure, since all the nonempty columns and lines are identical, describing the successive occurrences of the considered word: as already underlined above, it is isomorphic to a regular (but in general nonperiodic) square grid. (Note that such partial RPs have no simple continuous counterpart.) The fraction of occurrences of the m-word w in the sequence $\overline{x}$ of length N can be estimated as

$$\hat{f}_m^{(N)}(w|\overline{x}) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \delta(w, w_{m,i})$$

where $\delta(w,w')$ is the Kronecker symbol, equal to 1 if $w = w'$ else 0, and $w_{m,i}$ the ith m-word in the sequence $\overline{x}$.

A mere application of the law of large numbers ensures that

$$\lim_{N\to\infty} \hat{f}_m^{(N)}(w|\overline{x}) = p_m(w) \quad \text{almost surely}$$

(i.e. for any typical sequence $\overline{x}$) where $p_m(w)$ is the probability of emission of the m-word w by the source. It would be directly given by the fraction of dots in any line or row of the partial plot $\mathrm{RP}_m(w)$.

This limit statement is valid except in the case where the source exhibits very strong (nonintegrable) correlations. In most situations, correlations are integrable and do not spoil the law of large numbers; they only slow down the convergence towards the asymptotic limit, lowering the convergence rate by a factor proportional to the correlation range (at least m here since words overlap), as if the length of the sequence were reduced by the same factor [Lesne *et al.*, 2009]. This statement can equally be seen as a consequence of Birkhoff pointwise theorem under the assumption that the source is stationary and ergodic. A recurrence is observed in (i,j) associated with the word w if $\delta(w,w_{m,i})\delta(w,w_{m,j}) = 1$, hence the probability of observing such a recurrence is $\langle \delta(w,w_{m,i})\delta(w,w_{m,j})\rangle$ that in general differs from $\langle \delta(w,w_{m,i})\rangle\langle \delta(w,w_{m,j})\rangle = p_m(w)^2$: at this local level, correlations matter. On the contrary, the (estimator of the) overall density of recurrences associated with the m-word w writes

$$\begin{aligned}\hat{\rho}_m^{(N)}(w|\overline{x}) &= \frac{1}{(N-m+1)^2} \\ &\quad \times \sum_{i=1}^{N-m+1} \sum_{j=1}^{N-m+1} \delta(w,w_{m,i})\delta(w,w_{m,j}) \\ &= [\hat{f}_m^{(N)}(w|\overline{x})]^2\end{aligned}$$

so that almost surely

$$\begin{aligned}\rho_m(w) &= \lim_{N\to\infty} \hat{\rho}_m^{(N)}(w|\overline{x}) \\ &= \lim_{N\to\infty} [\hat{f}_m^{(N)}(w|\overline{x})]^2 = p_m(w)^2\end{aligned}$$

Hence at the level of the whole $\mathrm{RP}_m(w)$, $\rho_m(w) = p_m(w)^2$ and correlations no longer matter. The partial density $\hat{\rho}_m^{(N)}(w|\overline{x})$ of recurrences estimated from the sequence $\overline{x}$ and the total density of recurrences $\hat{\rho}_m^{(N)} = \sum_w \rho_m^{(N)}(w|\overline{x})$ (where the sum runs over all the possible m-words) are thus asymptotically trivial, simply related to the m-word frequencies; correlations do not influence the limiting result but only the rate of almost sure convergence towards the asymptotic limit $N\to\infty$. Essential information about the source correlations and temporal organization of the sequence thus lies in intermediate-range structures exhibited by the RPs, and in the relative organization and positioning of partial

recurrence plots. This information reflects in particular in diagonal lines, an essential feature of the whole m-RP that will be exploited in the following (see Sec. 2.3).

In this subsection, we have explicitly indicated that the recurrence fraction in a line $\hat{f}_m^{(N)}(w|\overline{x})$ and the density $\hat{\rho}_m^{(N)}(w|\overline{x})$ depend not only on the recurrent word w but also on the experimental sequence $\overline{x}$ from which the RP is drawn; this dependence on $\overline{x}$ is ubiquitous in all the statistical quantities relative to the RPs and will be henceforth omitted for alleviating the notations.

2.2. *Three theorems*

The quantitative analysis and interpretation of RPs statistical features, in particular the design of estimation procedures for the source entropy rate h, will be rooted in three theorems of information theory. These theorems are valid under the assumptions that the symbolic source is stationary and ergodic and that the source entropy per unit time h does not vanish ($h > 0$). But they do not require any assumption about the deterministic or stochastic nature of the source hence, like the very notion of entropy per unit time h on which they rely, they equally apply to sequences generated by discrete-state stochastic processes or encoding of a deterministic dynamics.

2.2.1. *Asymptotic equipartition property (Shannon–McMillan–Breiman)*

Shannon–McMillan–Breiman theorem [Breiman, 1957; Cover & Thomas, 2006; McMillan, 1953; Shannon, 1948] states that the number of typical m-words (i.e. that have the same properties corresponding to almost sure behavior) behaves like e^{mh} as $m \to \infty$ where the exponent h is the entropy rate of the source. This entropy rate h is currently defined as the limit $h = \lim_{m\to\infty} H_m/m = \lim_{m\to\infty}(H_{m+1} - H_m)$ where H_m is the Shannon entropy of the m-word distribution $p_m(\cdot)$, namely $H_m = -\sum_w p_m(w) \ln p_m(w)$ where the sum runs over all the possible m-words. A corollary of the Shannon–McMillan–Breiman theorem is the *Asymptotic Equipartition Property*, stating that the probability $p_m(w)$ of a typical m-word w takes asymptotically the value e^{-mh} common to all typical m-words, hence the name "equipartition". Formally, the statement has to be made more rigorous since the limiting behavior of the probabilities when $m \to \infty$ is still a function of m. Introducing the random variables $\hat{P}_m$ (depending on the whole realization $\overline{x}$ of the symbolic sequence) such that $\hat{P}_m(\overline{x}) = p_m(x_0, \ldots, x_{m-1})$, the asymptotic equipartition property writes more precisely

$$\lim_{m\to\infty} \left(-\frac{1}{m}\right) \ln \hat{P}_m \to h \quad \text{in probability} \qquad (1)$$

i.e. for any $\delta > 0$ and $\epsilon > 0$ (arbitrary small), there exists a word-size threshold $m^*(\delta, \epsilon)$ such that $\text{Prob}(\{\overline{x}, p_m(x_0, \ldots, x_{m-1}) > e^{m(-h+\delta)}\}) < \epsilon$ and $\text{Prob}(\{\overline{x}, p_m(x_0, \ldots, x_{m-1}) < e^{m(-h-\delta)}\}) < \epsilon$ for any $m \geq m^*(\delta, \epsilon)$, or equivalently, in terms of m-word subset, $p_m(\{w, p_m(w) > e^{m(-h+\delta)}\}) < \epsilon$ and $p_m(\{w, p_m(w) < e^{m(-h-\delta)}\}) < \epsilon$. Another corollary of Shannon–McMillan–Breiman theorem is to describe quantitatively how h accounts in an effective way for the correlations present within the sequence. Namely, the "effective" probability of a new symbol, knowing the sequence of length l that precedes, is asymptotically ($l \to \infty$) either e^{-h} or 0 whether the ensuing $(l+1)$-word is typical or not, instead of being equal to the symbol frequency as it is the case when there are no correlations within the sequence.

Shannon–McMillan–Breiman theorem and its corollaries will be the basis of the entropy estimation method from RPs developed and numerically checked in the next sections. Beforehand, we shall state two other limiting theorems and give their interpretation in terms of RPs; their practical applicability and scope would similarly deserve systematic investigations and numerical checks in future works.

2.2.2. *Recurrence time (Wyner–Ziv)*

Recurrence time theorem due to Wyner and Ziv (Theorem 1.a in [Wyner & Ziv, 1989]) states that the minimal recurrence time $\hat{T}_m$ at the level of m-words (that is, the smallest time t such that $(x_0, x_1, \ldots, x_{m-1}) = (x_t, x_{t+1}, \ldots x_{t+m-1})$) behaves as e^{mh} when $m \to \infty$, namely

$$\lim_{m\to\infty} \left(\frac{1}{m}\right) \ln \hat{T}_m \to h \quad \text{in probability} \qquad (2)$$

Note that from its very definition $\hat{T}_m$ is a random variable (a fact indicated by the hat) depending on the whole realization $\overline{x}$ of the symbolic sequence; statistical stationarity of the source ensures that $\tilde{T}_m$ defined as that the smallest time t such that $(x_0, x_1, \ldots, x_{m-1}) = (x_{-t}, x_{-t+1}, \ldots x_{-t+m-1})$

exhibits the same statistical features as $\hat{T}_m$. The statement (2) means that for any $\delta > 0$ and any $\epsilon > 0$ (arbitrary small), there exists a word-size threshold $m^*(\delta, \epsilon)$ such that for any $m \geq m^*(\delta, \epsilon)$, $\text{Prob}[\hat{T}_m > e^{m(h+\delta)}] < \epsilon$ and $\text{Prob}[\hat{T}_m < e^{m(h-\delta)}] < \epsilon$.

2.2.3. *Recurrence length (Wyner–Ziv)*

The same authors proved a related result (Theorem 1.b in [Wyner & Ziv, 1989]) that can be reformulated in our RP context as follows: let $\hat{L}^{(N)}$ be the length of the longest diagonal line in a RP of size $N \times N$ for single symbols. Like $\hat{T}_m, \hat{L}^{(N)}$ is a random variable over the sequence space. Its meaning is that of a "recurrence length" i.e. the length of the largest recurrent word. Then

$$\lim_{N\to\infty} \frac{\ln N}{\hat{L}^{(N)}} \to h \quad \text{in probability} \tag{3}$$

meaning that for any $\delta > 0$ and $\epsilon > 0$ (arbitrary small), there exists a size threshold $N^*(\delta, \epsilon)$ such that for $N \geq N^*(\delta, \epsilon)$, $\text{Prob}[\hat{L}^{(N)} > (\ln N)/(h-\delta)] < \epsilon$ and $\text{Prob}[\hat{L}^{(N)} < (\ln N)/(h+\delta)] < \epsilon$. In other words, $(\ln N)/\hat{L}^{(N)} \approx h$ with an arbitrarily small precision δ and an arbitrarily small probability ϵ of making a wrong statement for N large enough. The same asymptotic statement obviously holds as well for the length $\hat{L}_m^{(N)} = \hat{L}^{(N)} + m - 1$ of the longest diagonal line (if any) in the $(N - m + 1) \times (N - m + 1)$ m-RP$\hat{L}_m^{(N)}$.

2.2.4. *Practical meaning and use of these theorems*

Shannon–McMillan–Breiman and Wyner–Ziv theorems at the same time justify to treat all the observed m-words on the same footing and provide the value of the associated uniform probability, recurrence time and recurrence length; we can indeed assume, according to the very definition of typicality, that all the observed m-words are typical for m enough large, since nontypical ones are too rare to be ever observed. Nevertheless, an important caveat is the asymptotic nature of these theorems, making their application to finite words and finite sequences an *a priori* questionable extrapolation: simulations are required to appreciate the actual impact of finite-size effects, that motivates the numerical study presented in Sec. 3.

A first exploitation of these theorems is to give a self-consistent assessment of the quality and faithfulness of statistical analysis of the sequence and associated RPs. For large m and large N, an estimate of the average number of occurrences in the sequence of a typical m-word is $Np_m(w) \sim Ne^{-mh}$. This estimate leads to distinguish two regimes in constructing the m-RP:

— *good statistics*, if $Ne^{-mh} \gg 1$: each typical m-word is present a large number of times in the sequence; in particular all the typical words will be encountered in the m-RP;
— *bad statistics*, if $Ne^{-mh} \ll 1$: each typical m-word is present at most a few times in the sequence, and most often is not represented; in particular, multiple occurrences will be rare.

Accordingly, a trade-off has to be done in the choice of the word-length m between the applicability of asymptotic theorems ($m \to \infty$) and the statistical quality of the sampling ($e^{mh} \ll N$) hence of the estimations. When this trade-off can be achieved, that requires long enough sequences, the number of non-identical lines in a m-RP is expected to behave as e^{mh}. It means that in a m-RP with m large enough, the maximal number k^m of nonidentical columns reduces to e^{mh}, each one being the signature of a typical word in the RP, and accounting for its occurrence pattern in the observed sequence.

2.3. *Estimation of the entropy (per unit time) of the source*

A striking feature of RPs are diagonal lines. In a continuous context, going beyond the mere visual observation and analyzing their statistics yields a powerful entropy estimation method [Faure & Korn, 1998; Marwan *et al.*, 2007]. We shall here demonstrate that it is also the case in a discrete context. A diagonal line of length l starting in (i, j) in the m-RP means that $w_{m+l-1,i} = w_{m+l-1,j}$, namely the coincidence of two $(m+l-1)$-words. It corresponds to a diagonal line of length $l + 1$ in the $(m - 1)$-RP and a single dot in the $(m - l + 1)$-RP. All the quantities introduced below 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 $\overline{x}$ (not mentioned explicitly). We shall first assume that Shannon–McMillan–Breiman theorem, which centrally involves the entropy rate h of the source, is valid at the leading order, and that the size of

the m-RP (or equivalently the sequence length N) is large enough so that the law of large numbers applies and allows to identify quantities computed in one realization of the m-RP and their statistical average. This yields potential ways, presented below, of estimating the entropy per unit time h; their validity and efficiency in practical situations will be investigated numerically in the next section for map-generated trajectories and Markov chain realizations and in Sec. 4 for experimental behavioral sequences.

2.3.1. *Number of diagonal lines of a given length*

Considering a large enough diagonal length l so that at most double occurrences of $(m+l-1)$-word arise, the probability of double occurrence of a typical $(m+l-1)$-word multiplied by the number $e^{h(m+l-1)}$ of these (nonidentical) typical words yields the number $\nu_m^{(N)}(l)$ of diagonal lines of length l in the upper triangle of the m-RP, notwithstanding the main diagonal line:

$$\nu_m^{(N)}(l) = \frac{(N-l-m+2)(N-m-l+1)}{2} \times e^{-h(l+m-1)} \tag{4}$$

(It comes: $\nu_m^{(N)}(l) = \sum_{q\geq 2}(q(q-1)/2)e^{h(m+l-1)} C^q_{N-m-l-q+2} e^{-(m+l-1)qh}[1-e^{-(m+l-1)h}]^{N-m-l-q+2}$ when multiple occurrences are no longer neglected.) In this counting, a diagonal line of length $(l+r)$ contributes $(r+1)$ times to $\nu_m^{(N)}(l)$ since it corresponds to the double occurrence of $(r+1)$ typical and nonidentical $(m+l-1)$-words. Introducing the number $\eta_m^{(N)}(l)$ of diagonal lines of length exactly equal to l, we get:

$$\nu_m^{(N)}(l) = \sum_{r\geq 0}(r+1)\,\eta_m^{(N)}(l+r) \tag{5}$$

Considering the difference $\nu_m^{(N)}(l) - \nu_m^{(N)}(l+1)$ straighforwardly yields the number $\phi_m^{(N)}(l)$ of diagonal lines of length larger than l:

$$\phi_m^{(N)}(l) \equiv \sum_{r\geq 0}\eta_m^{(N)}(l+r) = \nu_m^{(N)}(l) - \nu_m^{(N)}(l+1) \tag{6}$$

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

$$\nu_m^{(N)}(l) \sim \left(\frac{N^2}{2}\right) e^{-h(l+m-1)} \tag{7}$$

so that a semi-log plot of $\nu_m^{(N)}(l)$ with respect to l will exhibit a slope $-h$ in its linear region, and

$$\ln\left(\frac{\nu_m^{(N)}(l)}{\nu_m^{(N)}(l+1)}\right) = h + \text{h.o.} \tag{8}$$

Using $\nu_m^{(N)}(l) - \nu_m^{(N)}(l+1) \approx -d\nu_m^{(N)}(l)/dl$ finally yields the scaling behavior of the number $\phi_m^{(N)}(l)$:

$$\phi_m^{(N)}(l) \equiv \sum_{r\geq 0}\eta_m^{(N)}(l+r) \approx -\frac{d\nu_m^{(N)}(l)}{dl} \sim \left(\frac{hN^2}{2}\right) e^{-h(l+m-1)} \tag{9}$$

The probability $\Phi_m(l)$ of observing diagonal lines of length larger than l is obtained by a mere normalization of $\phi_m^{(N)}(l)$ according to:

$$\Phi_m(l) \equiv \frac{\phi_m^{(N)}(l)}{\phi_m^{(N)}(l=1)} = \frac{\sum_{r\geq 0}\eta_m^{(N)}(l+r)}{\sum_{r\geq 0}\eta_m^{(N)}(1+r)} \quad \text{so that } \Phi_m(1)=1 \tag{10}$$

Its l-dependence obviously satisfies the same scaling law hence it does not matter whether we consider a number $\phi_m^{(N)}(l)$ or a probability $\Phi_m^{(N)}(l)$ when investigating their scaling behavior.

2.3.2. *Average length*

The number $\eta_m^{(N)}(l)$ of diagonal lines of length exactly equal to l gives straighforwardly the average length of the diagonal lines in the m-RP:

$$\langle l_m\rangle = \frac{\sum_{l\geq 1} l\,\eta_m^{(N)}(l)}{\sum_{l\geq 1}\eta_m^{(N)}(l)} = \frac{\nu_m^{(N)}(1)}{\nu_m^{(N)}(1)-\nu_m^{(N)}(2)} \tag{11}$$

At the leading order, the scaling $\sum_{r\geq 0}\eta_m^{(N)}(l+r) \sim (hN^2/2)e^{-h(l+m-1)}$ yields $\langle l_m\rangle \sim e^h/(e^h-1)$ and even $\langle l_m\rangle \sim 1/h$ if $h \ll 1$, giving an interpretation of $1/h$ as a characteristic time (correlation time) of the source. (Recall that labeling along the axes of RPs corresponds to time, and line lengths to time intervals).

2.3.3. *Maximal length*

Writing that the probability of double occurrence for a given word multiplied by the number e^{-mh} of possible m-words is lower than 1, we get a threshold

of single occurrence

$$\frac{(N-m+1)(N-m)}{2}e^{-mh}(1-e^{-mh})^{N-m-1} \leq 1 \tag{12}$$

yielding (using $m \ll N$ and $e^{-mh} \ll 1$) the maximal length $m^* - 1$ of words occurring more than once in the sequence:

$$N^2 e^{-mh} \leq 2 \quad \text{i.e. } m \geq m^* = \frac{2\ln N}{h} + \text{h.o.} \tag{13}$$

Let us denote $\hat{L}_m^{(N)}$ the largest diagonal length in the m-RP. It means that no $(m+l-1)$-word with $l > \hat{L}$ has double (or multiple) occurrences. This maximal length of a diagonal line is simply related with the above threshold m^* according to $\hat{L}_m^{(N)} = m^* - m$. We finally get:

$$\begin{aligned} m + \hat{L}_m^{(N)} &= \frac{2\ln N}{h} + \text{h.o.} \quad \text{i.e.} \\ h &= \frac{2\ln N}{m + \hat{L}_m^{(N)} - 1} + \text{h.o.} \end{aligned} \tag{14}$$

Another method comes from the theorem (3) of Wyner and Ziv about the length $\hat{L}^{(N)}$ of the longest diagonal line, Sec. 2.2.3: it can be exploited in practice by plotting $(\ln N)/\hat{L}^{(N)}$ as a function of N in order to evidence a stabilization at the value h at large N.

3. Numerical Experiments

In order to check in practice the validity and accuracy of entropy estimation from the statistical analysis of m-RP diagonal lines, we performed four series of numerical experiments:

(1) entropy estimation from symbolic encoding of logistic map trajectories;
(2) entropy estimation from symbolic encoding of logistic map trajectories in the presence of a variable amount of experimental noise;
(3) entropy estimation from symbolic encoding of logistic map trajectories in the case where the control parameter of the evolution law experiences a slow drift;
(4) entropy estimation from Markov chains with various numbers k of states and various characteristic times.

In these four cases, the parameters of the underlying dynamics are tunable, and the exact value of the entropy h per unit time (that depends on these parameters) is either reachable by an independent simulation (cases 1–3) or known analytically (case 4). We extracted from the RPs of the simulated trajectories the number $\phi_m^{(N)}(l)$ of diagonal lines of length longer or equal to l and plot its logarithm as a function of l (semi-log plot). From the asymptotic scaling behavior $\phi_m^{(N)}(l) \sim e^{-h(m+l-1)}$ demonstrated in Sec. 2.3, Eq. (9), we expect that the modulus of the slope of this monotonically decreasing curve in the region where it is linear (i.e. once the asymptotic regime is reached and before finite-sampling estimation issue sets in) yields an estimation of h. The purpose of our numerical experiments is to check the accuracy and robustness of this estimator for finite-length sequences, in other words to check quantitatively whether the range of practical validity of $\phi_m^{(N)}(l)$ scaling behavior could include real data. Note that the choice of m does not matter in this entropy estimation since diagonal-line lengths are related when considering two RPs associated with different word-lengths m and $m+q$ in a way that does not affect the scaling relation (explicitly, the histograms are related according to $\phi_{m+q}^{(N)}(l) = \phi_m^{(N)}(l+q)$).

3.1. *Logistic maps*

We first investigated the logistic map $g_a(z) = az(1-z)$ on the interval $[0,1]$ with binary encoding associated with the generating partition $[0,1/2] \cup]1/2,1]$ (see Appendix). The implementation of the RP estimation procedure for logistic map trajectories is presented in Fig. 1. Very different types of dynamics, ranging from periodic to fully chaotic, are encountered in a complicated and intermingled way when a is varied between 3 and 4. In such a case where the symbolic sequence follows from the phase space discretization of continuous-valued trajectories by means of a generating partition, the entropy rate of the symbolic source coincides with the metric entropy $h(a)$ of the underlying dynamics and, according to Pesin equality [Castiglione *et al.*, 2008], with the Lyapunov exponent $\gamma(a)$. The exact value $h(a)$ can thus be computed by computing $\gamma(a)$ on a long enough run, and the quality of the proposed estimation procedure checked in all deterministic dynamic regimes by comparing the estimated and exact values of $h(a)$ for all values of a. The results presented in Figs. 2(a) and 2(b) show a very good agreement since even the fine details of the curve $a \to h(a)$ are captured by the estimator. We have as expected obtained identical results with a quaternary encoding (see Appendix).

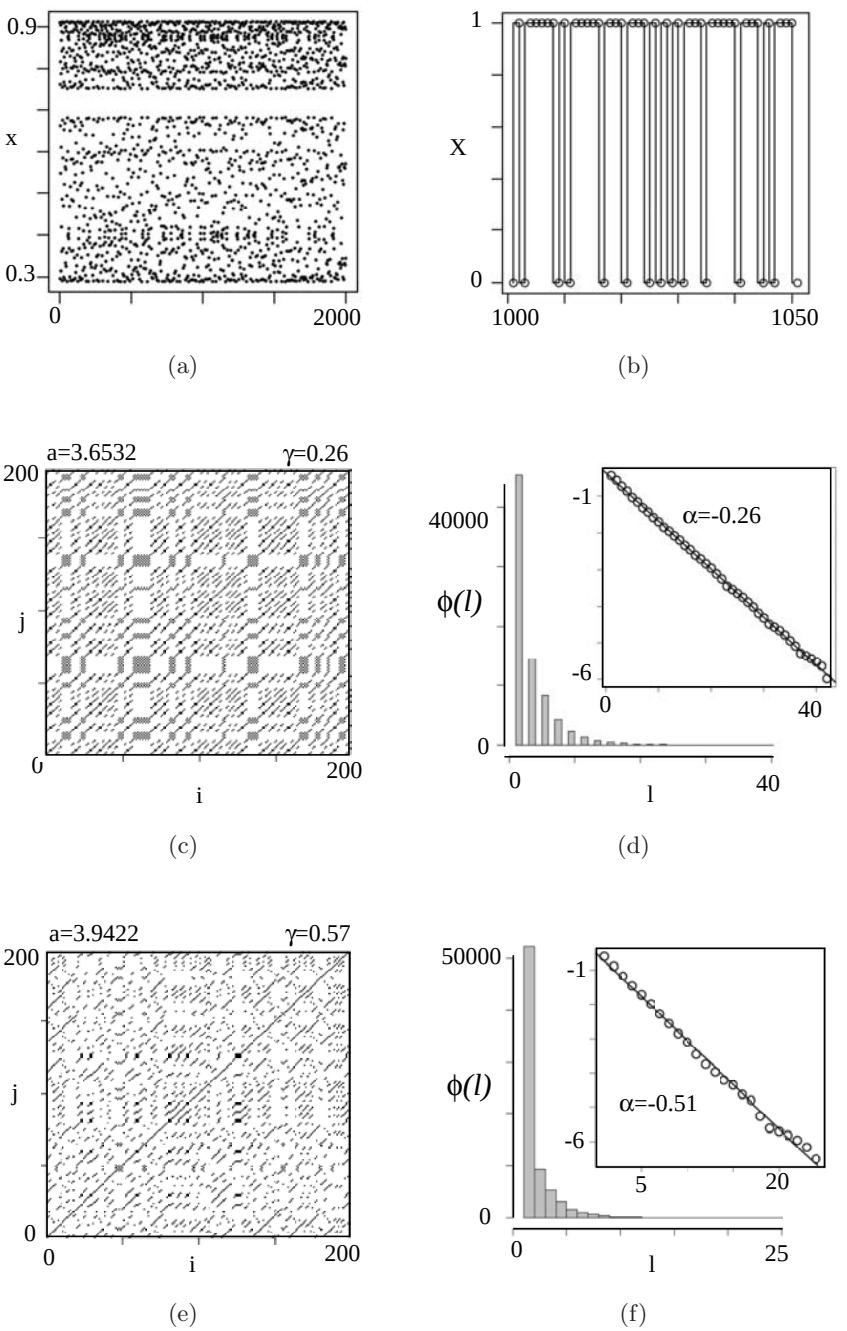

Fig. 1. Principle of entropy estimation from RP of logistic map trajectories' symbolic encoding. (a) Time series $(z_i)_{i=1,\ldots,N}$ of length $N = 2000$ obtained from logistic map $g_a(z) = az(1-z)$ in $[0,1]$ with $a = 3.6532$ and (b) corresponding symbolic sequence $(x_i)_{i=1,\ldots,N}$, written with an alphabet of $k = 2$ symbols. Symbolic encoding of logistic map trajectories is made using the simple rule: if $z_i > 0.5$ then $x_i = 1$ else $x_i = 0$. (c) RP obtained from $(x_i)_{i=1,\ldots,200}, a = 3.6532$ and $m = 4$. γ indicates the exact value of the entropy per unit of time calculated from an independent simulation as the Lyapunov exponent value. (d) Histogram of the number $\phi_m^{(N)}(l)$ of diagonal lines of length longer or equal to l, counted in the RP obtained from $(x_i)_{i=1,\ldots,2000}$. (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. (e, f) Same as (c, d), with $a = 3.9422$.

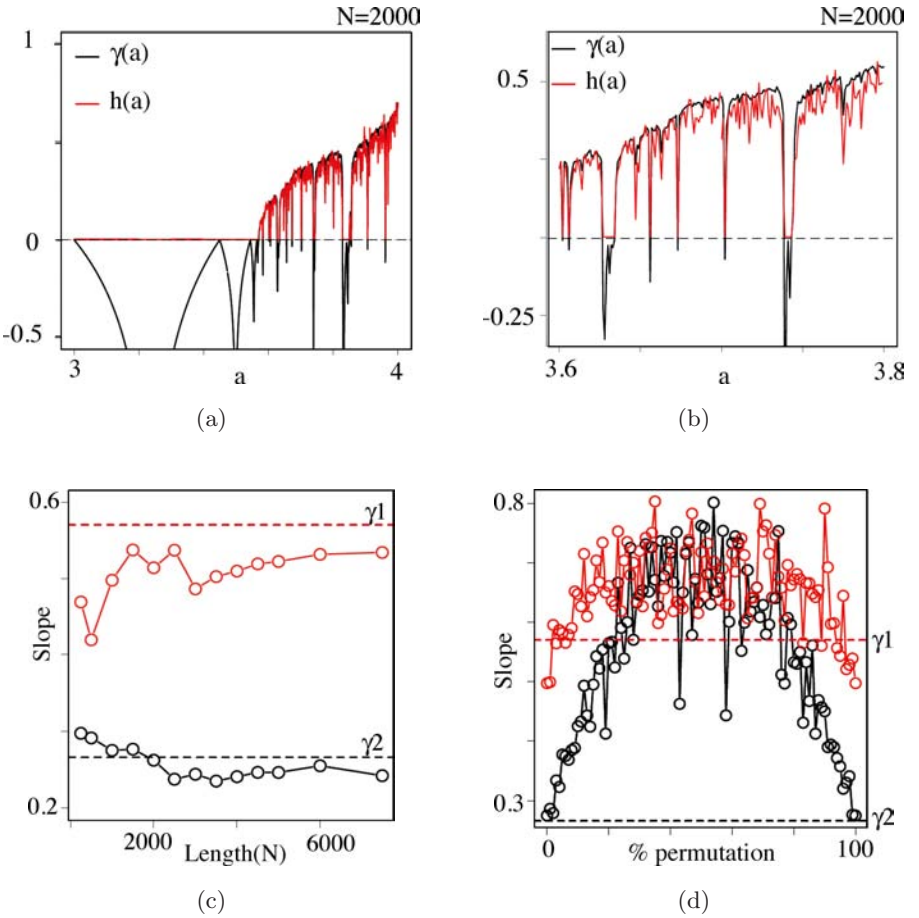

Fig. 2. Accuracy and robustness of entropy estimation. (a) Comparison of the estimated value of $h(a)$ from RP (red line) compared with Lyapunov exponent value $\gamma(a)$ (black line). $h(a)$ is estimated from symbolic encoding of logistic map trajectories of length $N = 2000$ where a is varied from 3 to 4 with a step of 0.001. (b) Enlargement of the left graph, with $3.6 \leq a \leq 3.8$. Note that a negative value of $\gamma(a)$ corresponds to an entropy value $h(a) = 0$, whereas Pesin equality ensures $h(a) = \gamma(a)$ for positive values of $\gamma(a)$. (c) Variation of estimated $h(a)$ with the length N of the symbolic sequence for two values of a, $a = 3.6532$ (black circles) and $a = 3.9422$ (red circles). Values of the corresponding Lyapunov exponents are indicated by horizontal dashed lines. N is varied from 250 to 7000. (d) Robustness of $h(a)$ estimation of the symbolic sequence of length $N = 2000$ for two values of a, $a = 3.6532$ (black circles) and $a = 3.9422$ (red circles) in the presence of experimental noise. The initial time series is modified by randomly inverting ($1 \to 0$ and $0 \to 1$) a given percentage (% permutation) of symbols, corresponding to the noise strength. % permutation is varied from 0 to 1 with step of 0.01. Values of the corresponding Lyapunov exponents are indicated by horizontal dashed lines.

Investigating the size dependence of the estimation quality, Fig. 2(c), shows that the sequence length has to be larger than $N = 1000$ to get an accurate estimation of $h(a)$.

3.2. *Noise effect on entropy estimation*

We also investigated numerically the robustness of the entropy estimation procedure with respect to experimental noise, that is, noise spoiling the recorded sequence but not amplified by the dynamics. We mimicked the effect of such a noise by randomly flipping symbols with a tunable probability ϵ, corresponding to the strength of the noise (see Appendix) in sequences obtained by encoding logistic maps trajectories, for several values of the control parameter a. As expected, the quality of the estimate markedly decreases when the noise strength increases [Fig. 2(d)], up to reaching the maximal value $\ln 2$ for the entropy rate of a binary sequence with no track of the underlying deterministic value. Note the spurious symmetry of the curves with respect to the flipping probability $\epsilon = 1/2$ (see Appendix): $\epsilon > 1/2$ amounts to flipping the mirror sequence $x'_i = 1 - x_i$ (having the same entropy and symmetric statistical features) with a probability $1 - \epsilon$.

3.3. *Entropy estimation for a nonstationary dynamics*

A key assumption in entropy estimation is the statistical stationarity of the source: RPs allow to get a visual support of this assumption or conversely warn that it might fail. RPs visually hint at nonstationary 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). In such

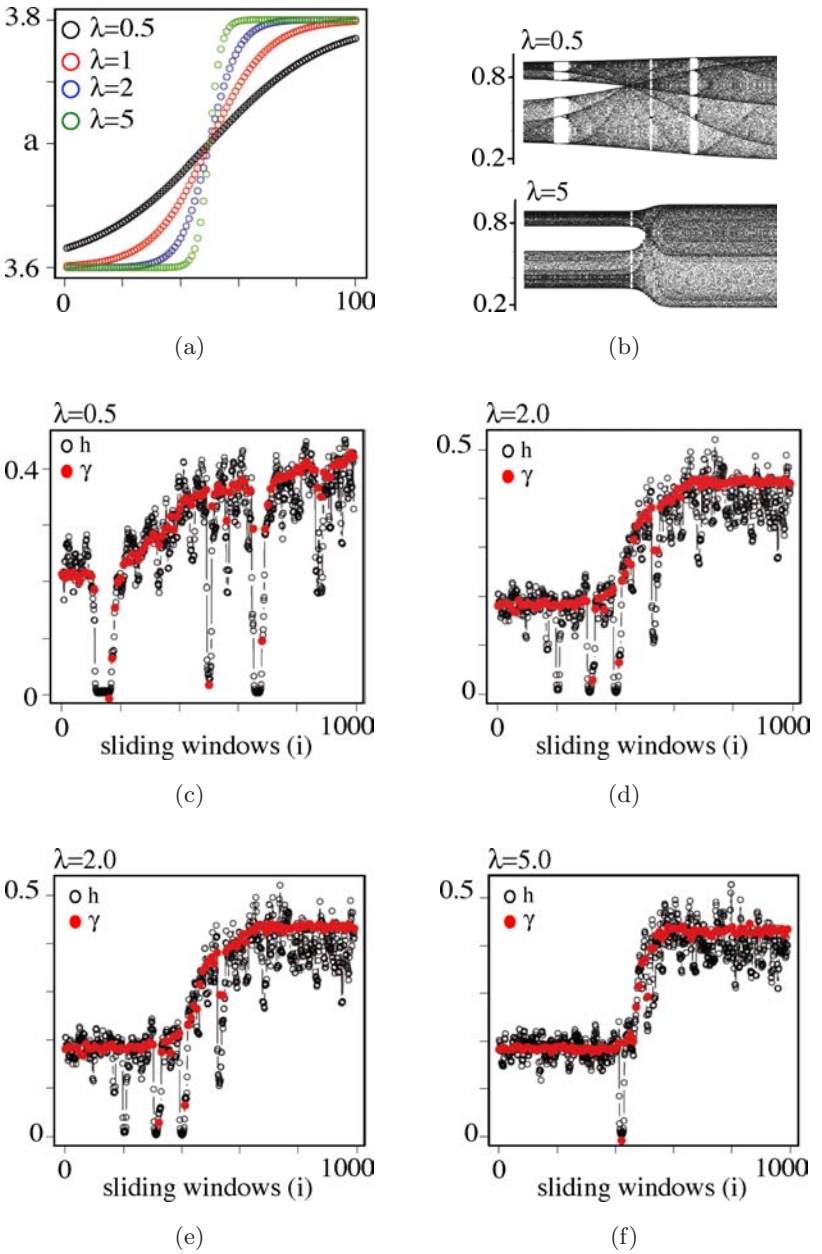

Fig. 3. Entropy estimation for nonstationary dynamics. (a) Variation of a from 3.6 to 3.8 in the course of the simulation, following a sigmoidal function. Value of a in the nth stage of the simulation is given by $a_n = 3.6 + (2/(10 * (1 + \exp(-\lambda n))))$ with $\lambda = 0.5, 1, 2, 5$ and $n = 1, \ldots, 100$. (b) Time series obtained from concatenation of time series $[z_{a_n}(i)]_{i=1,\ldots,1000}$ from logistic map $g_{a_n}(z) = a_n z(1-z)$ for $n = 1, \ldots, 100$, with a_n varying from 3.6 to 3.8 as explained in (a), for $\lambda = 0.5$ (top) and $\lambda = 5$ (bottom). The last point ($i = 1000$) of a time series $z_{a_n}(i)$ is used as the initial point ($i = 1$) to calculate the following $z_{a_{n+1}}(i)$ series. (c)–(f) Variation of h along the nonstationary dynamics. h is estimated on successive overlapping windows of length $N = 1000$ with a shift of 100 time steps. (c) Estimate of h (black circles), superimposed with the value of Lyapunov exponent (red points) calculated for each value of $a_n, n = 1, \ldots, 100$, calculated here with $\lambda = 0.5$. (d) Same as (c) with $\lambda = 1.0$, (e) $\lambda = 2.0$, (f) $\lambda = 5.0$.

nonstationary situations, statistical analysis and in particular entropy estimation should be restricted to time windows where the RP is statistically homogeneous. We have numerically implemented this procedure, taking as a benchmark trajectories generated by a logistic map whose control parameter a slowly increases by small steps δa in the course of evolution; this increase is slow enough for a quasi-stationary approximation to make sense and allow to consider an entropy rate $h(a)$ corresponding to the instantaneous value of a and characterizing the nonstationary dynamics during the associated transient stage. It amounts for the dynamical system to scan the curve $a \to h(a)$ with a resolution corresponding to step-sizes δa. Figure 3 shows that the ensuing evolution of the entropy rate, although very irregular, can be faithfully captured by entropy estimation in a sliding window. In order to mimic experimental situations, the analysis of numerical trajectories is done without involving *a priori* knowledge of the underlying nonstationarity: the size of the sliding window and the number of time steps by which it is recursively shifted are chosen arbitrarily (in practical situations, additional knowledge or hint about the system could be invoked in this choice). We have checked (data not shown) the robustness of the entropy estimation with respect to specific values of these observation parameters within an appropriate range that can be determined from visual inspection of the RP or, adaptively, from the outcome of the local analysis. This robustness supports the consistency and relevance of the local estimation procedure.

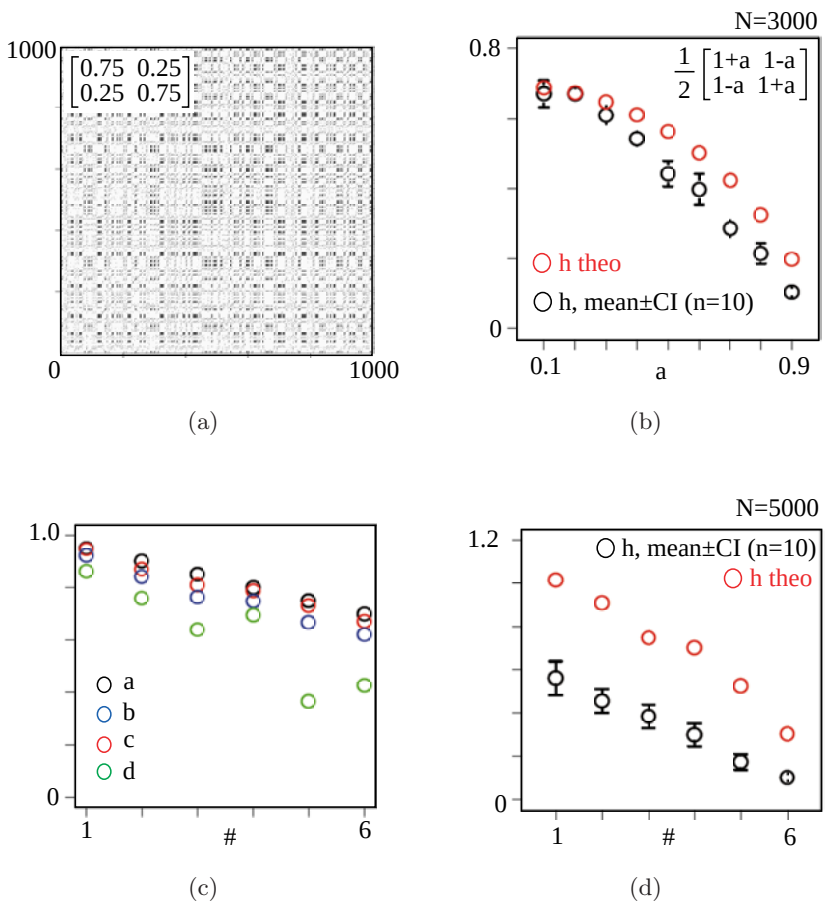

Fig. 4. Entropy estimation for Markov chains. (a) RP from a two-symbol Markov chain. The transition matrix is given in inset. (b) Comparison of the estimated value of $h(a)$ from RP (black points, mean $\pm 95\%$ confidence interval, average over $n = 10$ trajectories of length $N = 3000$) compared with the exact value of $h(a)$ computed from the transition matrix as the average conditional entropy (red points). General form of the matrix is given in inset, where $a = e^{-1/\tau}$ gives the characteristic time τ of the Markov chain; a is varied from 0.1 to 0.9. (c, d) Entropy estimate for five-symbols Markov chains. (c) Trajectories of length $N = 5000$ are generated from six different matrices (see text and Appendix) with six quadruplets (a, b, c, d). (d) Corresponding estimated value of $h(a)$ from RP (black points, mean $\pm 95\%$ confidence interval, average over $n = 10$ chains) and exact value calculated analytically from the transition matrix as the average conditional entropy (red points).

3.4. *Markov chains*

The dynamics of a Markov chain of order 1 is fully determined by the transition matrix R, that is, the matrix whose element R_{ij} is the conditional probability of transition from symbol i to symbol j; each line of R thus sums up to 1. Denoting H_m the Shannon entropy of m-words, its entropy rate h expresses exactly as $H_2 - H_1$ or equivalently as the average conditional entropy (that is, the average with respect to the stationary distribution $p_{\rm stat}$ of the conditional entropies i.e. Shannon entropies of the lines $(R_{ij})_j$ each seen as a probability distribution):

$$h = -\sum_i p_{\rm stat}(i) \sum_j R_{ij} \ln R_{ij} \qquad (15)$$

where $\sum_j R_{ij} = 1$ and $\sum_i p_{\rm stat}(i) R_{ij} = p_{\rm stat}(j)$. With such an analytical benchmark, we were in a position to investigate the quality of entropy estimation from RPs. We considered two families of Markov chains (see Appendix) with respectively 2 and 5 states and tunable eigenvalues, corresponding to tunable characteristic times (equivalently correlation times in the stationary regime or relaxation times towards the stationary state since it can be shown that they coincide [Gaveau *et al.*, 1999]).

Results for a two-state Markov chain are shown in Figs. 4(a) and 4(b) and exhibit good agreement given the limited length of the sequences. Finite-size effects are as expected stronger for a Markov chain with five states, as shown in Figs. 4(c) and 4(d). Our systematic numerical study evidences a significant negative bias that might be due to undersampling: estimated entropy underestimates the actual value. This bias here appears to be lower at low h, for more correlated sequences; additional finite-size effects might come from the relative influence of the transient regime before the stationary distribution is reached. But we underline that trends are preserved and the qualitative behavior of the entropy as a function of the control parameters of the dynamics is captured.

4. Applications to Experimental Sequences

Tracing, from analysis of spontaneous activity, mouse cognitive functions and decision making, and alteration of these functions in mutant mice is a current challenge in behavior analysis. Open-field behavior has been studied in rodents for long time and such experiment is today a central element in the battery of tests used in laboratories. Establishing a quantitative description of such behavior is then important but a detailed description of the structure of the animal displacement in the open-field is generally ignored. We have developed an analysis of open-field behavior based on symbolic analysis. Encoding behavioral records into symbolic sequences is an acknowledged approach allowing to eliminate information nonpertinent to the investigated issue, to improve the statistics by reducing the dimension of the phase space and to reduce the part of individual variability. Animal trajectories where position and velocity are recorded at each step, and represented by a sequence of symbols {PI,CI,PAp,PAc,CA}, that correspond to Activity or Inactivity in the Periphery or in the Center of the arena. Two different symbols PAp and PAc are introduced to distinguish peripheral movement that follows inactivity in the periphery (PI), and peripheral movement that follows central movement (CA). This distinction have been made in order to obtain a first-order Markov dynamics [Maubourguet *et al.*, 2008].

We first described transitions among the five states (the residence time in each state is not taken into account here [Maubourguet *et al.*, 2008]) in mice trajectories by transition matrices having the general parametrized form:

$$R(i) = \begin{pmatrix} 0.0 & 0.03 & 0.97 & 0.0 & 0.0 \\ 0.12 & 0.0 & 0.0 & 0.88 & 0.0 \\ a = 0.64 - i * 0.05 & 0.0 & 0.0 & 0.36 + i * 0.05 & 0.0 \\ 0.0 & b = 0.27 + i * 0.05 & 0.0 & 0.0 & 0.73 - i * 0.05 \\ c = 0.76 - i * 0.05 & 0.0 & 0.0 & 0.24 + i * 0.05 & 0.0 \end{pmatrix} \qquad (16)$$

and we analyzed the modification of the entropy per unit time with its parameter i. We generated trajectories of length $N = 500$ (the typical length of experimental sequences) for $i = 0$ to 10, constructed for each of them the corresponding m-RP with $m = 4$ and extracted the cumulative histogram $\phi_m^{(N)}(l)$ of the length

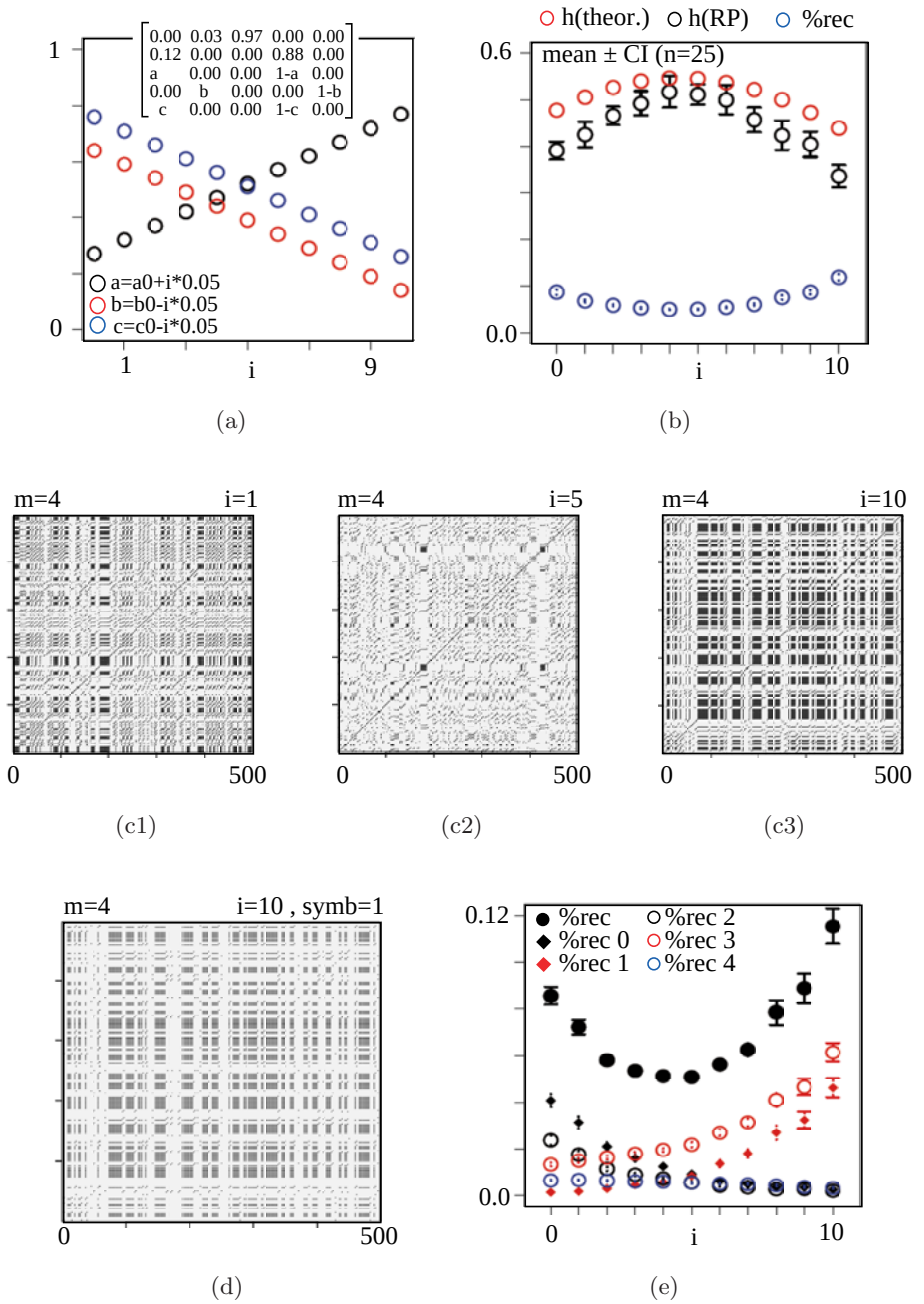

Fig. 5. Entropy estimation for a five-state Markov chain model of behavioral sequences. (a) General form $R(i)$ of the transition matrix (in inset) and values of (a, b, c) used to generate trajectories of length $N = 500$. (b) Comparison of the estimated value of $h(a)$ from RP (black points, mean $\pm 95\%$ confidence interval, average over $n = 25$ trajectories) compared with the exact value of $h(a)$ computed from the transition matrix as the average conditional entropy (red points). The fraction % rec of recurrence points in the RP (for $m = 4$) is also represented (blue points). (c1, c2, c3) Three RPs illustrating the recurrence pattern for values $i = 1, 5$ and 10 of the matrix parameter. (d) The recurrence plot for 4-words such as that represented for $i = 10$ in (c3) is the superposition of five disjoint recurrence plots $\mathrm{RP}_4(w_j)$ with $j = 0, 1, 2, 3, 4$, each corresponding to words of length $m = 4$ that begin with symbol j. The partial $\mathrm{RP}_4(w_{j=1})$ is represented on the left. The percentage of recurrence %rec can accordingly be dissociated into five parts, $\%\mathrm{rec} = \sum_{j=0}^{4} \%\mathrm{rec}(j)$ where $\%\mathrm{rec}(j)$ is the percentage of recurrence in $\mathrm{RP}_4(w_j)$. Variation with i of %rec and $\%\mathrm{rec}(j)$ for $j = 0, 1, 2, 3, 4$ is represented on the right.

diagonal lines. Variation of a, b, and c [Fig. 5(a)] encompasses a large range, allowing to cover most of the cases observed in experimental situations, e.g. with different mouse strains. Variation of theoretical entropy obtained from the matrix $R(i)$ is well recovered by entropy estimation from RP diagonal lines. The percentage of recurrence (%rec) is also a good indicator of the modification of the

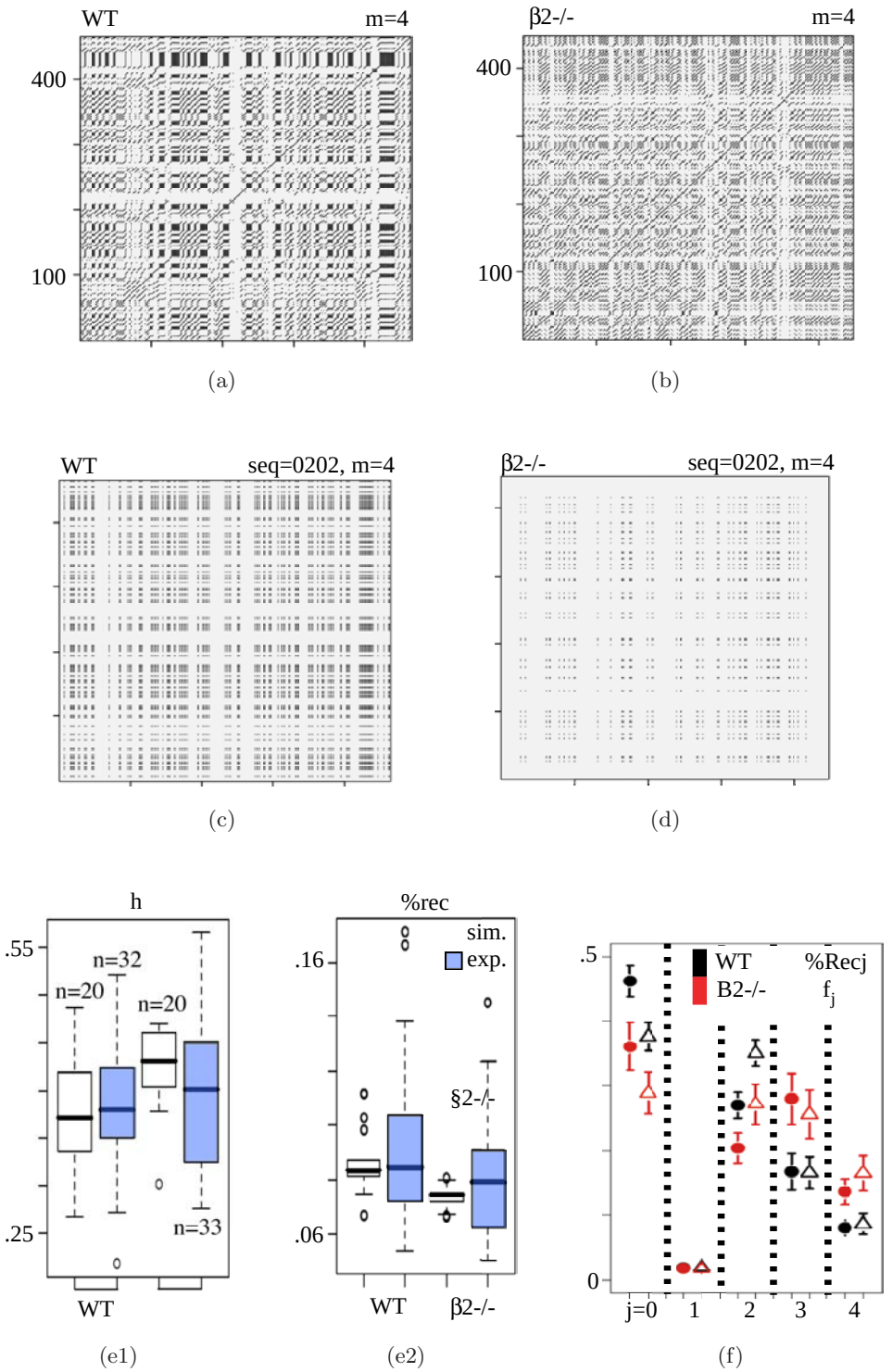

Fig. 6. Analysis of experimental data. (a) Example of recurrence plot from wild type (WT) and (b) β2-/- mutant mice (right) symbolic trajectories. (c) Partial $\mathrm{RP}_m(w)$ for $m = 4$ and $w = [0202]$ in WT and (d) β2-/- animals, illustrating that checkerboard structure characterizing WT symbolic trajectories are mainly repeated [02] sequences. (e1) Comparison of the estimated entropy $h(a)$ and (e2) %rec from simulated and experimental RPs ($m = 4$). Simulated data are obtained using the transition matrix estimated from experimental data. n indicates the number of trajectories, i.e. the number of mice or simulation runs in each group. (f) Comparison in WT (black points) and β2-/- mice (red points) of the distribution of $\%\mathrm{rec}(j)$ (bullets) and of c_j (triangles). $\%\mathrm{rec}(j)$ is calculated in $\mathrm{RP}_m(w_j)$ for $j = 0, 1, 2, 3, 4$ and normalized by the total recurrence %rec. c_j correspond to the normalized square percentage of symbols $j = 0, 1, 2, 3, 4$ in the sequence, namely $c_j = N_j^2 / \sum_{i=0}^{4} N_i$ (averaged over several trajectories) with N_j being the total number of symbols j in the sequence. The rationale of this comparison is the fact that $\%\mathrm{rec}(j)/\%\mathrm{rec} = c_j$ for a sequence of independent and identically distributed random variables.

recurrence structure when the transition probabilities are modified [Fig. 5(b)]. Typical RPs obtained for $i = 1$, 5 and 10 [Fig. 5(c)] illustrate the modification of checkerboard structures: Generally, square clusters in a m-RP reveal the presence of repeats of a r-word w^* (with $r \leq m$); most often, rather than fully-filled clusters, we observe checkerboards, composed of staggered rows and lines that are not continuous but composed of interspersed dots in diagonal arrays, revealing the recurrence of a r-word with now $r > m$. This staggered-row structure can also be encountered in continuous RPs where it reveals a transient and recurring quasi-periodicity. The meaning of checkerboards as regards the underlying dynamics is here more delicate to assess than for continuous RPs; in particular, the association of checkerboards with quasi-equilibria (in case of diagonal checkerboard) or intermittency (several aligned off diagonal checkerboards) that is often valid in continuous RPs is less straightforward for symbolic RPs. The repetition of a symbol (or word) has a weaker meaning than a quasi-equilibrium, revealing only a succession of features associated with the same symbol (or word). Thus, there is no general statement going beyond the equivalence between the observation of a checkerboard and the occurrence (in case of a diagonal checkerboard) or recurrence (in case of a off-diagonal checkerboard) of a repeat sequence $[w, \ldots, w]$ where w is either a single symbol or a word; further interpretation has to be done on a case-by-case basis. For instance, in the present behavioral study, it would indicate phases of perseveration, during which the animal performs several times in a row the same behavioral pattern corresponding to the recurring m-word.

Figures 5(d) and 5(e) implement the notion of partial RPs. The left part of the figure represents the total RP for 4-words, RP_4, which is the superposition of five partial recurrence plots $\mathrm{RP}_4(w_j)$ describing the recurrence of 4-words beginning by a given symbol $j = 0, \ldots, 4$. The right part of the figure represents the percentage of recurrence of each subset of words w_j as a function of the parameter i of the dynamics. The interest of this analysis is to dissect which words give the dominant contribution to the full RP. Going beyond a global view on recurrences is specially important in a behavioral context, where we seek to extract specific behavioral patterns, in order to understand what determines the specific moments where animals "make a decision" and change their movement.

RPs of experimental data obtained from wild-type mice (WT, $n = 32$ mice) are compared with those obtained in knock-out mutant mice (β2-/-, $n = 33$ mice). The terminology β2-/- for the mutant mice is the standard way of meaning that both alleles of the gene coding for the β2 subunit of the acetylcholine receptor have been knocked out hence are no longer expresed in the animals. RP of WT animals [Fig. 6(a)] for $m = 4$ shows a number of checkerboard structures. Investigating more precisely the contribution of the different partial $\mathrm{RP}_4(w)$ shows that checkerboards mainly originate from the repetition of sequences [0,2] i.e. [PI, PAp]; this observation evidences that typical behavior in WT animals consists of alternating movements and pauses at the periphery intermingled with sequences of center movement. In β2-/- mice checkerboards are significantly reduced [Figs. 6(c) and 6(d)].

We also compared the percentage of recurrent points in WT and β2-/- mice [Fig. 6(e)]. %rec corresponds to what is formally and less straight forwardly denoted $\hat{\rho}_m^{(N)}$ in Sec. 2.1.2, with $\hat{\rho}_m^{(N)} = \sum_w \hat{\rho}_m^{(N)}(w)$ where the sum runs over all possible m-words. Letting the sum run only over m-words beginning by the symbol j yields a partial percentage $\%\mathrm{rec}(j)$, such that $\%\mathrm{rec} = \sum_{j=0}^{4} \%\mathrm{rec}(j)$. Without any correlation between the successive symbols, we would have $\%\mathrm{rec}(j)/\%\mathrm{rec} = [\hat{f}_1(j)]^2 / \sum_{i=0}^{4} [\hat{f}_1(i)]^2$ where $\hat{f}_1(j)$ is the observed frequency of the symbol j in the sequence or equivalently $\%\mathrm{rec}(j)/\%\mathrm{rec} = N_j^2 / \sum_{i=0}^{4} N_i^2 \equiv c_j$ where N_j is the number of occurrences of the symbol j in the sequence. The discrepancy between the two ratios observed in a given experiment [Fig. 6(f)] reflects the presence of correlations and the fact that recurrence of words (corresponding here to behavioral sequences) differs from a succession of symbol recurrences. In the example presented here, recurrence of symbol $j = 0$ noticeably increases the probability of recurrence of 4-words starting with $j = 0$, evidencing a short-range correlation between $j = 0$ and the subsequent symbols, in agreement with the fact that the relevant dynamic model allowing to reproduce the recorded data is a Markov chain rather than a sequence of independent and identically distributed random variables.

5. Conclusion

In this paper, we introduced an extension of RPs for symbolic sequences. In practice, symbolic RPs are specially useful when there is no known underlying

continuous-state dynamics and the data are intrinsically symbolic (e.g. behavioral sequences and genomic sequences). Our main point was to assess the validity of using RPs for analyzing such data. We developed heuristic arguments and checked on both simulated and experimental data that such symbolic RPs offer a faithful and accurate way for estimating the entropy rate h of the underlying dynamics, even for relatively short sequences. As in the continuous counterpart, estimation is based on the analysis of the length cumulative distribution of diagonal lines. We used the logistic maps (trajectories in a continuous phase space and their symbolic encoding using a generating partition) as a well-known benchmark to check that symbolic RPs actually give access to the entropy per unit time, with no need to rely on the underlying continuous trajectories. Entropy rate h is relevant for 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. This is illustrated above on a special instance of discrete-state stochastic processes, namely Markov chains.

The underlying theoretical result is the Shannon–Breiman–McMillan theorem. It states a striking feature of the word statstics, known as asymptotic equipartition property, namely that for a word-length m large enough, only two classes of m-words are to be distinguished: those that have a negligible probability to be observed, in particular that will never be observed in experiments or simulations (non typical words) and the typical words, having all the same probability of occurrence e^{-mh} controlled by the entropy rate h of the source. Not only does it support to treat all the observed (hence typical) words on the same footing, but it also gives access to h. What we have evidenced in this paper through a systematic numerical investigation, is the nontrivial fact that this asymptotic equipartition regime sets in very rapidly, for moderate values of m (here $m = 4$) and yields efficient ways of estimating the entropy rate h even for short sequences, for instance, of length N between 1000 and 3000 in case of symbolic encoding of an underlying deterministic dynamics. Finite-size effects are expectedly stronger for stochastic trajectories, flawing the estimation with a systematic bias (h is systematically underestimated due to undersampling) and an accurate entropy estimate would require longer sequences; nevertheless, even for short sequences, the trend of the entropy as a function of the control parameters of the dynamics is preserved, hence the estimated value can be used for classification or discrimination between several dynamics.

The acknowledged benefit of estimating entropy is to get a quantitative and integrated appreciation of the strength of correlations and temporal organization of the dynamics. The counterpart of this very advantage is the fact that entropy is only an asymptotic feature, giving only an overall appreciation of the dynamic correlations. A possible way yet to be explored, to get further information, and capture and quantify dynamic correlations at different scales, would be to perform a multiscale analysis and adapt to symbolic RPs the method of windowed RPs introduced in [Casdagli, 1997]. The idea is to consider square cells of linear size q (an integer) and to compute the local density $\rho_{IJ}^{(\Delta)}$ of recurrence in the cell (I, J) as a sum of m-word contributions $\rho_{IJ}^{(\Delta)} = \sum_w \rho_{IJ}^{(\Delta)}(w)$. This coarse description involves an average over time scales smaller than q while keeping track of the exact recurrences of m-words. The result can be represented either using a grayscale or introducing a threshold on $\rho_{IJ}^{(\Delta)}$ for considering that (I, J) is a "meta-recurrence" [Casdagli, 1997]. As mentioned in Sec. 2.1.2, a noticeable feature revealed by this multiscale analysis is the fact that local correlations average out at the level of the whole RP (correlations simply slow down the convergence toward $\rho_m(w)$ of the w-recurrence densities $\hat{\rho}_m^{(N)}(w)$ as $N \to \infty$) whereas they are likely to have a non-negligible effect at intermediary scales q; these correlations would thus reflect in local heterogeneities and, quantitatively, in the discrepancy between $\rho_{IJ}^{(\Delta)}(w)$ and $q^2\rho_m(w)$.

A strength of RPs is to be suitable for statistical analysis of symbolic sequences (extraction of average or integrated features like the entropy or the average recurrence time) while keeping track of the temporal location, hence also allowing to evidence visually and locate dynamic transitions and more generally nonstationary features of the evolution. Indeed, entropy is not always the more meaningful index, and has not a direct interpretation in terms of dynamic structure, e.g. structure of a decision process in a behavioral context. For instance, several very different dynamics, from a behavioral viewpoint, might have the same entropy. But several other specific RP features are not yet fully exploited. Additional quantification indices, focusing on intermediate-scale features (motifs,

checkerboards, recurrence of a specific subset of words, e.g. words beginning by a given symbol) are yet to be developed. For instance, checkerboard structures would deserve a deeper quantification and interpretation since they are likely to give unique information about the dynamic organization at intermediate scales, out of reach of other methods of analysis. m-RPs would also allow to investigate the recurrence of a motif $w - gap - w'$ where w and w' are two different or identical m-words. Denoting g the distance between the beginning of w and w', the recurrence of the motif reflects in the correlation between the occupancy of the points (i, j) and $(i + g, j + g)$. A quantitative assessment would then be obtained by computing the ratio $r(i, g)$, at a given i, of the number of pairs of recurrences (i, j) and $(i + g, j + g)$ divided by the number of occupied pairs (i, j). Indeed, this ratio $r(i, g)$ equals 1 when there is a perfect correlation and the word w always followed at a distance g by the word w'. Also nonstationary features (e.g. sharp transitions, slow drift) can be detected if the appropriate tools are developed. In particular, we have shown that symbolic RPs analysis allows to compute entropy of symbolic sequences without the requirement of statistical stationarity: the RP methodology presented here allows to capture a slow variation in the (local) entropy per unit time following from a slow drift in a control parameter of the dynamics. Finally, an interesting research direction is to determine in what respect the coarse-grained description of a symbolic RP and its quantification would allow to perform a multiscale analysis of the dynamics, as it could be sometimes done in a continuous context [Casdagli, 1997].

References

Badii, R. & Politi, A. [1999] *Complexity. Hierarchical Structures and Scaling in Physics* (Cambridge University Press, Cambridge).

Bandt, C., Groth, A., Marwan, N., Romano, M. C., Thiel, M., Rosenblum, M. & Kurths, J. [2008] "Analysis of bivariate coupling by means of recurrence," *Mathematical Methods in Time Series Analysis and Digital Image Processing*, eds. Dahlhaus, R., Kurths, J., Maas, P. & Timmer, J. (Springer, Berlin, Heidelberg), pp. 153–182.

Breiman, L. [1957] "The individual ergodic theorem of information theory," *Ann. Math. Statist.* **28**, 809–811; Correction **31**, 809–810.

Casdagli, M. C. [1997] "Recurrence plots revisited," *Physica D* **108**, 12–44.

Castiglione, P., Falcioni, M., Lesne, A. & Vulpiani, A. [2008] *Chaos and Coarse-Graining in Statistical Mechanics* (Cambridge University Press, Cambridge).

Collet, P., Galves, A. & Leonardi, F. [2007] "Random perturbations of stochastic chains with unbounded variable length memory," ArXiv:math/0707.2796.

Cover, T. M. & Thomas, J. A. [2006] *Elements of Information Theory* (Wiley, NY).

Daw, C. S., Finney, C. E. A. & Tracy, E. R. [2003] "A review of symbolic analysis of experimental data," *Rev. Sci. Instrum.* **74**, 916–930.

Eckmann, J. P., Kamphorst, S. O. & Ruelle, D. [1987] "Recurrence plots of dynamical systems," *Europhys. Lett.* **4**, 973–977.

Faure, P. & Korn, H. [1998] "A new method to estimate the Kolmogorov entropy from recurrence plots: Its application to neuronal signals," *Physica D* **122**, 265–279.

Faure, P., Neumeister, H., Faber, D. S. & Korn, H. [2003] "Symbolic analysis of swimming trajectories reveals scale invariance and provides model for fish locomotion," *Fractals* **11**, 233–243.

Gaveau, B., Lesne, A. & Schulman, L. S. [1999] "Spectral signatures of hierarchical relaxation," *Phys. Lett. A* **258**, 222–228.

Groth, A. [2005] "Visualization of coupling in time series by order recurrence plots," *Phys. Rev. E* **72**, 046220.

Lesne, A., Blanc, J. L. & Pezard, L. [2009] "Entropy estimation of very short symbolic sequences," *Phys. Rev. E* **79**, 046208.

Letellier, C. [2006] "Estimating the Shannon entropy: recurrence plots versus symbolic dynamics," *Phys. Rev. Lett.* **96**, 254102.

Marwan, N., Romano, M. C., Thiel, M. & Kurths, J. [2007] "Recurrence plots for the analysis of complex systems," *Phys. Rep.* **438**, 237–329.

Maubourguet, N., Lesne, A., Changeux, J. P., Maskos, U. & Faure, P. [2008] "Behavioral sequence analysis reveals a novel role for $\beta 2^*$ nicotinic receptors in exploration," *PloS Comput. Biol.* **4**, e1000229.

McMillan, B. [1953] "The basic theorems of information theory," *Ann. Math. Stastist.* **24**, 196–219.

Shannon, C. [1948] "A mathematical theory of communication," *Bell System Tech. J.* **27**, 379–423.

Wyner, A. D. & Ziv, J. [1989] "Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression," *IEEE Trans. Inform. Th.* **35**, 1250–1258.

Appendix A

A.1. *Logistic map simulations*

For deterministic dynamics generated in a continuous phase space by a map g, symbolic sequences result from the discretization of continuous-valued

trajectories using a partition of the phase space in k subsets $A_x^{(1)}$, with $z_i \in A_{x_i}^{(1)}$ being the positions along the continuous trajectory underlying the symbolic sequence $\overline{x}$. The partition is said to be generating when the knowledge of the semi-infinite symbolic sequence $(x_i)_{i\geq 0}$ fully determines a unique initial condition $z_0 \in A_{x_0}^{(1)}$ in the continuous phase space. In this 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 encoding trajectories of finite length. On the other hand, quantitative analysis, and specifically entropy estimation, is expected to be statistically more faithful when performed on symbolic sequences, i.e. trajectories in a finite phase space.

For an unimodal map g of the interval, a generating partition is known: $A_0 \cup A_1$ where $A_0 = [0, z^*[$ and $A_1 = [z^*, 1]$ if the critical point is z^*, in particular $z^* = 1/2$ for logistic maps. Other generating partitions are obtained from the basic one by considering $A_{xx'}^{(2)} = A_x \cap g^{-1}(A_{x'})$ (leading to a 4-symbol partition) or more generally $A_{x_1,\ldots,x_m}^{(m)} = A_{x_1} \cap g^{-1}(A_{x_2}) \cap \cdots \cap g^{m-1}(A_{x_m})$ (leading to a 2^m-symbol partition). It amounts to code a point $z_1 \in [0,1]$ by the m-word describing to which half-interval belong the successive images $z_1, z_2 = g(z_1), \ldots, z_m = g(z_m)$. Binary and quaternary encodings are expected to give identical results as regards entropy estimation. Indeed, symbols in the quaternary encoding correspond to 2-words in the binary encoding; more generally, m-words in the quaternary encoding correspond to $(m+1)$-words in the binary encoding and diagonal lines of length l corresponds to diagonal lines of length $l+1$. Accordingly, the scaling of the histogram of diagonal-line length with respect to l, hence the resulting entropy estimation, will not depend on the choice of m. We have checked this point (data not shown) hence present only the binary case results in the main text.

A.2. *Perturbing the symbolic sequences by a controlled amount of noise*

Currently, the way to add noise on a binary sequence is to flip each symbol with a (small) probability ϵ, see e.g. [Collet *et al.*, 2007]. Obviously, such a procedure is valid only for small ϵ since for $\epsilon = 1$, we do not get a fully disordered sequence but rather the mirror-symmetric sequence of the original one. A more proper way to generate a noisy sequence would be to proceed in two steps: for each symbol, consider it with a probability ϵ then, if it is considered, redraw its value with an unbiased probability 1/2 (a fully random sequence would now be obtained for $\epsilon = 1$). In contrast to the first method, this procedure is not restricted to binary sequences but would be efficient for any symbolic sequence, whatever the size of the alphabet. Nevertheless, it is computationally more heavy and the issue of entropy estimation robustness only makes sense for moderate levels of experimental noise, for which the first procedure for adding noise is sufficient (we do not expect to have any faithful access to the the entropy of the underlying deterministic dynamics when experimental noise is strong); we thus implemented the first method.

For $\epsilon \ll 1$, the probability to destroy an existing recurrence is $2m\epsilon$ for the simple flip and $m\epsilon$ when using the second option for adding noise (up to $\mathcal{O}(\epsilon^2)$, that amounts to ignore double flips). It is difficult to assess analytically the probability of false recurrence (it would require to assess the number in the noiseless sequence of quasi-recurrences where only one symbol misses to get a true recurrence) but given a noiseless sequence, we could ensure that this probability scales as ϵ.

A.3. *Markov chain simulations*

We devise the transition matrix R (with the convention that R_{ij} is the conditional probability of a transition from i to j) as $R = SDS^{-1}$ where D is the diagonal matrix of eigenvalues ($1 \geq a \geq b \geq c \geq d \geq \cdots$), the lines of the matrix S^{-1} are the right eigenvectors of R and the columns of matrix S are the left eigenvectors of R; hence the first line of S^{-1}, corresponding to the stationary distribution, has strictly positive components if R is nondegenerate (i.e. if 1 is a single isolated eigenvalue) and is normalized to 1, the other lines sum up to 0, and the first column of S has all its components equal to 1. Additional conditions arise on the eigenvalues to ensure the positivity of all the components of the transition matrix R. Considering 2×2 matrices with equiprobable states in the stationary regime:

$$S^{-1} = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & -0.5 \end{pmatrix} \quad S = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \qquad D = \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix} \tag{A.1}$$

it becomes:

$$R(a) = \frac{1}{2} \cdot \begin{pmatrix} 1+a & 1-a \\ 1-a & 1+a \end{pmatrix} \tag{A.2}$$

The characteristic time of the stochastic dynamics generated by $R(a)$ is related to the nontrivial eigenvalue $a > 0$ according to $\tau(a) = 1/\ln(1/a)$. The exact value of the entropy per unit time $h(a)$ can be computed analytically as the average conditional entropy:

$$h(a) = -\left(\frac{1+a}{2}\right)\ln\left(\frac{1+a}{2}\right) - \left(\frac{1-a}{2}\right)\ln\left(\frac{1-a}{2}\right) \tag{A.3}$$

Considering 5×5 matrices with equiprobable states in the stationary regime and

$$D = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & a & 0 & 0 & 0 \\ 0 & 0 & b & 0 & 0 \\ 0 & 0 & 0 & c & 0 \\ 0 & 0 & 0 & 0 & d \end{pmatrix} \tag{A.4}$$

it becomes:

$$R(a,b,c,d) = \frac{1}{5} \cdot \begin{pmatrix} 1+4a & 1-a & 1-a & 1-a & 1-a \\ 1-4a+3b & 1+a+3b & 1+a-2b & 1+a-2b & 1+a-2b \\ 1-3b+2c & 1-3b+2c & 1+2b+2c & 1+2b-3c & 1+2b-3c \\ 1-2c+d & 1-2c+d & 1-2c+d & 1+3c+d & 1+3c-4d \\ 1-d & 1-d & 1-d & 1-d & 1+4d \end{pmatrix} \tag{A.5}$$

requiring the consistency conditions: $1 - 4a + 3b \geq 0, 1 - 3b + 2c \geq 0$ and $1 - 2c + d \geq 0$ on the eigenvalues (assuming $1 > a \geq b \geq c \geq d \geq 0$) in order that $R(a,b,c,d)$ is actually a transition matrix. The eigenvalues give the relaxation times, the largest one being τ such that $a = e^{-1/\tau}$. Moreover, it can be shown (using the spectral decomposition of the transition matrix) that correlation functions $C_{B_1,B_2}(t)$ of pairs (B_1, B_2) of observables decrease exponentially fast to 0 with correlation times that coincide with the relaxation times, thus widening the interpretation of the eigenvalues [Gaveau *et al.*, 1999]. As in the case of a 2×2 matrix, the exact value of the entropy per unit time can be computed analytically as a function of a, b, c and d.