Non-sequential recursive pair substitutions and numerical entropy estimates in symbolic dynamical systems

Lucio M. Calcagnile∗, Stefano Galatolo† and Giulia Menconi‡

July 21, 2009

Abstract
We numerically test the method of non-sequential recursive pair substitutions to estimate the entropy of an ergodic source. We compare its performance with other classical methods to estimate the entropy (empirical frequencies, return times, Lyapunov exponent). We considered as a benchmark for the methods several systems with different statistical properties: renewal processes, dynamical systems provided and not provided with a Markov partition, slow or fast decay of correlations. Most experiments are supported by rigorous mathematical results, which are explained in the paper.

1 Introduction
We investigate a symbolic substitution method as a tool to estimate entropy of an ergodic source. The entropy we deal with is the Shannon entropy of finite-alphabet stationary stochastic processes, in particular those that can be obtained as a symbolic model of a dynamical system.

Throughout the paper, we shall refer to this method as Non-Sequential Recursive Pair Substitution (NSRPS). The idea of applying recursive pair substitutions to symbolic sequences was first proposed by Jimenez-Montaño, Ebeling and others (see [EJM]), but it was put into the formal context of

∗Scuola Normale Superiore, Pisa, Italy: l.calcagnile@sns.it
†Dipartimento di Matematica Applicata, Università di Pisa, Italy: galatolo@mail.dm.unipi.it
‡Istituto Nazionale di Alta Matematica, Roma, Italy
probability theory and studied more deeply by Grassberger [Gr] in 2002 and Benedetto, Caglioti, Gabrielli [BCG] in 2006.

We now briefly explain how the NSRPS method works.

Let us suppose to have a finite-state stationary source, that is a device providing infinite sequences of symbols \(x_0x_1x_2\ldots\) where each \(x_i\) is an element of a finite alphabet \(A\), in such a way that the probability of receiving a given finite string does not vary with time. Given a sequence from such a source, the NSRPS method prescribes to individuate the pair (or one of the pairs) of symbols of maximal frequency and to substitute all its non-overlapping occurrences with a new symbol \(\alpha \notin A\). For example, given the sequence

\[
011010111011000111011010011\ldots,
\]

taken from a source \(\mu\) for which \(\mu(01)\) is the highest among the probabilities of symbol pairs, we substitute the pair 01 with the new symbol 2, thus obtaining

\[
21221120211212021\ldots.
\]

In the case the pair to substitute is made up of two equal symbols, not all the occurrences are to be substituted, but only the non-overlapping ones. For example, given the sequence

\[
00110100001010001000001100001\ldots,
\]

we substitute the pair 00, obtaining

\[
211012210120122011221\ldots.
\]

Starting from a source \(\mu\) with alphabet \(A = \{0, 1\}\), after the first substitution we shall have a new source with alphabet \(A_1 = \{0, 1, 2\}\) and a measure \(\mu_1\) on the finite strings inherited from \(\mu\). We can then go on repeating the steps, introducing new symbols 3, 4, \ldots and obtaining new sources \(\mu_2, \mu_3, \ldots\).

The main theorem about the NSRPS method (Theorem 2.4) says that the entropy \(h\) of an ergodic source \(\mu\), which is defined by

\[
h(\mu) = \lim_{k \to \infty} -\frac{1}{k} \sum_{\text{length}(x) = k} \mu(x) \log_2 \mu(x),
\]

can be calculated, in the limit for the number \(N\) of substitutions which approaches infinity, knowing only the probabilities according to \(\mu_N\) of the individual symbols and of the pairs in the new sources, after many substitutions. We remark that the hypotheses of substituting at each step one of the pairs with the maximum probability is a sufficient but not necessary condition for the conclusion of the main theorem [2.4] to hold (see [BCG]).
Numerical results about the use of this method for the estimation of the entropy of the English language were sketched in [Gr]. Here we show a first systematic comparison of this method with other classical ones, by performing several experiments on artificial sequences. We will mainly use symbolic sequences constructed by dynamical systems.

The use of symbolic models of dynamical systems as a benchmark for this kind of study is motivated by the following two important features:

- dynamical systems can produce strings with many kinds of nontrivial statistical features (slow decay of correlations, no Markov structure, and so on...)
- the dynamical/geometrical properties of the system under consideration often allow the entropy of the system to be estimated (sometime rigorously calculated) by some other method (Lyapunov exponents and geometrical properties of the invariant measure e.g.) whose results can be compared with the estimation done by symbolic methods.

In order to judge the precision and the speed of the entropy estimating algorithm suggested by the NSRPS method, we shall compare it with other three much used entropy estimating methods. Two of them apply to symbolic sequences. They are the empirical frequencies method and the return times method. Finally, in the case of ergodic transformations, we calculate the Lyapunov exponent which converges very fast and will be considered as a reference value for the entropy. The use of these numerical estimators will be supported by rigorous mathematical results, which will be explained in the paper.

In section 2 we formally present the NSRPS method and state the main theorem about it. In section 3 we recall some basic notions of symbolic dynamics. In section 4 we give a review of rigorous results supporting the estimation of entropy by the other methods we chose: empirical frequencies, return times and Lyapunov exponent. In section 5 we discuss the details of the implementation of the above methods and the reasons of some arbitrary choice we could not avoid. In section 6 we present the experimental results, with some tables and figures.

2 Non-Sequential Recursive Pair Substitutions (NSRPS)

In this section we briefly recall from [BCC] definitions and main results on the NSRPS method. We introduce the terms and the notations which are
fundamental to state the main theorem. We omit all the technical details and the proofs, which the interested reader can find in [BCG].

We recall from the introduction that the method we study is applied to symbolic sequences which are supposed to come from a finite-state stationary source.

Let us call our finite alphabet $A$ and denote with $A^* = \cup_{k=1}^{\infty} A^k$ the collection of all finite words in the alphabet $A$. A word $w \in A^*$ has length $|w|$ and, if $|w| = k$, it will also be indicated with $w^k = w_1 \ldots w_k$.

Let $x, y \in A$, $\alpha \notin A$ and $A_1 = A \cup \{\alpha\}$.

**Definition 2.1.**

A pair substitution is a function $G = G_{xy}^\alpha : A^* \rightarrow A_1^*$ which is defined by recursively substituting all the non-overlapping occurrences of the pair $xy$. More precisely, $Gw$ is defined substituting in $w$ the first occurrence from left of $xy$ with $\alpha$ and repeating this procedure to the end of the sequence.

We consider again the example sketched in the introduction and show some general notation. Given the sequence

$$w = 0110111011000111011010011 \in \{0, 1\}^*,$$

performing the substitution $01 \mapsto 2$ leads to

$$G_{01}^2(w) = 2122112100211212021 \in \{0, 1, 2\}^*.$$

We indicate with $\mathcal{E}(A)$ the set of all the stationary ergodic measures on $A^\mathbb{Z}$, the only ones we shall deal with. If $\mu \in \mathcal{E}(A)$ and $w \in A^*$, we shall use the notation $\mu(w)$ to indicate the $\mu$-measure of the cylinder set $[w_1, \ldots, w_k] = \cap_{i=1}^{k} \{X_i = w_i\}$, where the $X_i$’s are the random variables which describe the stochastic process.

The map $G = G_{xy}^\alpha$ naturally induces a map $G = G_{xy}^\alpha : \mathcal{E}(A) \rightarrow \mathcal{E}(A^*_1)$, as the following theorem shows. We indicate with $\sharp\{s \subseteq r\}$ the number of occurrences of a subword $s$ in a word $r$.

**Theorem 2.2.** If $\mu \in \mathcal{E}(A)$ and $s \in A_1^*$, then the limit

$$G_\mu(s) = \lim_{n \rightarrow \infty} \frac{\sharp\{s \subseteq G(w^n)\}}{|G(w^n)|}$$

exists and is constant $\mu$ almost everywhere in $w$. Furthermore, the values $\{G_\mu(s)\}_{s \in A_1^*}$ are the marginals of an ergodic measure on $A_1^\mathbb{Z}$.

It is obvious that a pair substitution shortens the sequence it is applied to. The following proposition gives an average quantification of this shortening.
Proposition 2.3. If $x \neq y$ then

$$Z_{xy}^\mu \overset{\text{def}}{=} \lim_{n \to \infty} \frac{n}{|G(w_1^n)|} = \frac{1}{1 - \mu(xy)} \quad (\mu \text{ a. e. in } w). \quad (2.1)$$

If $x = y$ then

$$Z_{xx}^\mu \overset{\text{def}}{=} \lim_{n \to \infty} \frac{n}{|G(w_1^n)|} = \frac{1}{1 - \sum_{k=2}^{\infty} (-1)^k \mu(x^k)} \quad (\mu \text{ a. e. in } w), \quad (2.2)$$

where $x^k$ is the string made up of $k$ symbols $x$.

We now recall the definition of entropy of a process.

Given $\mu \in \mathcal{E}(A)$ and $n \geq 1$, the quantity

$$H_n(\mu) = -\sum_{|w|=n} \mu(w) \log_2 \mu(w)$$

is the $n$-th order entropy.

The $n$-th order conditional entropy is defined as

$$h_n(\mu) = H_{n+1}(\mu) - H_n(\mu).$$

It can be shown (see [Sh]) that the quantities $h_n(\mu)$ and $H_n(\mu)/n$ converge to the same value, which is the Shannon entropy of the process $\mu$:

$$h(\mu) = \lim_{n \to \infty} h_n(\mu) = \lim_{n \to \infty} H_n(\mu)/n. \quad (2.3)$$

2.1 The main theorem

Intuitively, after a pair substitution the information is more concentrated, with respect to the original sequence.

After several substitutions, the most important blocks (the most frequent ones) are concentrated into symbols and the value of the entropy can be calculated by applying the standard formula with short blocks ($H_k$ with small $k$).

This can be formulated in precise terms (see [BCG], Theorem 3.2 and Corollary 2.6) and suggests that a sequence of substitutions might asymptotically transfer all the information to the distribution of the pairs and individual symbols. This is precisely the content of the main theorem.

To state it, we define the following objects:

- the alphabets $A_N = A_{N-1} \cup \{\alpha_N\}$ where $\alpha_N \notin A_{N-1}$ and $A_0 = A$;
- the maps $G_N = G_{x_Ny_N}^{α_N} : A_{N-1}^* \rightarrow A_N^*$, where $x_N, y_N \in A_{N-1}$;
- the maps between measures $G_N = G_{x_Ny_N}^{α_N}$;
- the measures $μ_N = G_Nμ_{N-1}$, with $μ_0 = μ$;
- the quantities $Z_N = Z_{x_Ny_N}^{μ_{N-1}}$ and $Z_N = Z_N \ldots Z_1$.

Theorem 2.4. ([BCG], Theorem 3.2) If
\[ \lim_{N \to \infty} Z_N = +\infty \]
then
\[ h(μ) = \lim_{N \to \infty} \frac{h_1(μ_N)}{Z_N}. \] (2.4)

Theorem 2.5. If at each step $N$ the pair $x_Ny_N$ is a pair with the maximum frequency among all the pairs of symbols of $A_{N-1}$, then
\[ \lim_{N \to \infty} Z_N = +\infty. \]

Theorems 2.4 and 2.5 combined guarantee that, by performing at each step the substitution of a pair with maximum probability, the entropy of the original ergodic process is approximated by the 1-st order conditional entropy, which takes into consideration only the distribution of the single symbols and of the pairs of symbols. In this sense, through this method “the ergodic process becomes 1-Markov in the limit”.

In practical utilizations of the above theorem we have access to the statistical properties of the source by measuring the empirical frequency of digit sequences in the experimental data we have. Given a sequence $x_1x_2 \ldots x_n$, the empirical distribution of the (overlapping) $k$-blocks $a^k_1$ is defined naturally by
\[ p_k(a^k_1| x_n^1) = \frac{\# \{ i \in [1, n-k+1] : x^i+i+k-1 = a^k_1 \} }{n-k+1} \] (2.5)
and its empirical $k$-entropy is defined by
\[ \tilde{H}_k(x^n_1) = -\sum_{w=x^1_1} p_k(w|x^n_1) \log_2 p_k(w|x^n_1). \]

Let us call $G$ the substitution operation on the maximal frequency pair (if there are more than one string of maximal frequency, the lexicographic order is used). By ergodicity, it is possible to rephrase the above theorem into a statement which is more similar to what can be practically done on long strings coming from the source:
Corollary 2.6. If $\mu$ is ergodic, for almost each $\omega \in A^\mathbb{N}$

$$h(\mu) = \lim_{n \to \infty} \lim_{l \to \infty} \frac{\bar{H}_2(G^n(\omega'_1)) - \bar{H}_1(G^n(\omega'_1))}{\bar{Z}_n(\omega'_1)}$$

(2.6)

where $\bar{Z}_n(\omega'_1) = \frac{l}{|G^n(\omega'_1)|}$ is the shortening rate after $n$ substitutions.

Proof. Let $\omega$ be a typical realization of the system. Since the system is ergodic $\lim_{l \to \infty} \bar{H}_k(G^n(\omega'_1)) = H_k(\mu_n)$, hence $\lim_{l \to \infty} \bar{H}_2(G^n(\omega'_1)) - \bar{H}_1(G^n(\omega'_1)) = h_1(\mu_n)$. Moreover, in the same way, when $n$ is fixed and $l \to \infty$, $\bar{Z}_n(\omega'_1) \to \bar{Z}_n$ and the corollary follows from the above Theorem 2.4.

□

3 Symbolic dynamics

In this section we briefly recall the basic notions about symbolic dynamics and Kolmogorov-Sinai entropy. We already defined the entropy of a symbolic process. Entropy may be defined also for measure-preserving transformations. This will be done by associating symbolic sequences with the orbits of the transformation. Let us more precisely recall the definition of Kolmogorov-Sinai entropy $h_\mu$ of a map $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ having an ergodic invariant measure $\mu$.

Let $\alpha = \{A_1, \ldots, A_k\}$ be a finite measurable partition of $X$. Let $\Omega$ be the product space $\{1, 2, \ldots, k\}^\mathbb{N}$, so that an element of $\Omega$ is a sequence $\omega = (\omega_n)_{n=0}^\infty$, where $\omega_n \in \{1, 2, \ldots, k\}$ for all $n$.

It is possible to translate in a standard way the dynamics of $(X, \mathcal{B}, \mu, T)$ into the dynamics of the space $\Omega$, which is provided with the Borel $\sigma$-algebra $\mathcal{B}(\Omega)$ generated by the cylinder sets and the left shift transformation $\sigma$. Let us define a map $\phi_\alpha : (X, \mathcal{B}) \to (\Omega, \mathcal{B}(\Omega))$ by

$$(\phi_\alpha(x))_n = \omega_n \quad \text{if} \ T^n x \in A_{\omega_n}.$$ 

so that the $n$-th coordinate of $\phi_\alpha(x)$ is the alphabet letter corresponding to the element of the partition $\alpha$ which $T^n x$ belongs to.

It holds $\phi_\alpha(Tx) = \sigma(\phi_\alpha x), \forall x \in X$. Furthermore, the map $\phi_\alpha$ is measurable and naturally transports the measure $\mu$ on $(\Omega, \mathcal{B}(\Omega))$ defined by setting for every measurable $E \subseteq \Omega$, $\nu(E) = \mu(\phi_\alpha^{-1}E)$.

Notice that in general the map $\phi_\alpha$ is not invertible, thus it does not always give an isomorphism. However, if the partition $\alpha$ is generating, that is the sets of the form $A_1 \cap T^{-1}A_2 \cap \ldots \cap T^{-(m-1)}A_m$ generate the $\sigma$-algebra $\mathcal{B}$, then the map $\phi_\alpha$ gives an isomorphism between $(X, \mathcal{B}, \mu, T)$ and $(\Omega, \mathcal{B}(\Omega), \nu, \sigma)$.
If $\alpha$ and $\beta$ are two measurable partitions of $(X, \mu, T)$, their joint partition $\alpha \lor \beta$ is the set $\{A \cap B \mid A \in \alpha \text{ and } B \in \beta\}$. If $T$ is a measurable and non-singular function and $\alpha$ is a partition, then $T^{-1}\alpha$ is the partition defined by the subsets $\{T^{-1}A \mid A \in \alpha\}$.

Given the partition $\alpha = \{A_1, \ldots, A_k\}$ we shall denote the Shannon entropy of the partition by

$$H(\alpha) = -\sum_{i=1}^{k} \mu(A_i) \log_2(\mu(A_i)) .$$

The entropy of the map $T$ with respect to the partition $\alpha$ is:

$$h_\mu(T, \alpha) = \lim_{n \to +\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i} \alpha\right).$$

The Kolmogorov-Sinai entropy of the dynamical system $(X, \mu, T)$ is

$$h_\mu(T) = \sup_{\alpha} h_\mu(T, \alpha),$$

where the supremum is taken over all the finite partitions.

There exist partitions whose entropy is the Kolmogorov-Sinai entropy of the map.

**Theorem 3.1** (Kolmogorov). Consider a dynamical system $(X, \mu, T)$. If $\alpha$ is a generating partition with respect to the map $T$, then

$$h_\mu(T) = h_\mu(T, \alpha).$$

The existence of a generating partition for a dynamical system is assured by the following theorem.

**Theorem 3.2** (Krieger Generator Theorem [Kr]). For an ergodic dynamical system $(X, \mu, T)$ on a Lebesgue space $X$, such that $h_\mu(T) < \infty$, there exists a finite generating partition $\alpha$.

The identification of a generating partition is generally a challenging task. In the following, we shall provide some examples of generating partitions in specific cases.
4 Estimating entropy from samples

When a process’ invariant measure is explicitly known, we could in principle estimate the entropy by applying the definition. On the other hand, when we are not given explicit knowledge of the measure, we are often not able to know exactly the entropy of the process and the problem of entropy estimation arises. A usual approach to this problem is considering long sample sequences, which are looked at as parts of infinite typical sequences and thus representing the statistical features of the system. To such samples several entropy estimating algorithms can be applied.

We shall compare the estimating algorithm suggested by the NSRPS method with two others, which we shall call the empirical frequencies (briefly, EF) method and the return times (briefly, RT) method. We remark that these methods can be applied directly to the symbolic sequence without having any other information on the source. For the ergodic transformations of the unit interval we shall use another estimating algorithm which does not apply to symbolic processes: the approximation of the Lyapunov exponent. We remark that the estimation of entropy by this method uses some additional information on the system (the derivative of the map, which is calculated at each step of the dynamics, and the dimension of the invariant measure).

Each estimation algorithm is supported by rigorous results, as it will be shown in the following sections and will be implemented in its simplest form.

We end remarking that, while experimental examples contained in this paper are long artificial trajectories mostly coming from dynamical systems. When working on short sequences (for instance finite realization of some biophysical process or experiment), surrogate analysis and a suitable correction of the estimator can be useful in order to take into account fluctuations of entropy or implicit bias on the chosen estimator (see e.g. [MEPR], [BHM]).

4.1 Empirical frequencies (EF)

To estimate entropy directly by the definition, a simple procedure consists in determining the empirical distribution $p_k$ of the overlapping $k$-blocks and taking $\frac{H_k(p_k)}{k}$ as an estimate for $h$. If $k$ is fixed and the length of the sample sequence $n$ tends to infinity, then $\frac{H_k(p_k)}{k}$ almost surely converges to $\frac{H_k(\mu_k)}{k}$, which tends to $h$ as $k \to \infty$. Theorem 4.1 below guarantees that these two limits can be performed together with $k(n) \sim \log_2 n$.

Given the sequence $x_1x_2 \ldots x_n$, the empirical distribution $p_k(\cdot|x^n_t)$ of the overlapping $k$-blocks is defined as in (2.5).
**Theorem 4.1.** If \( \mu \) is an ergodic measure of entropy \( h > 0 \), if \( k(n) \to \infty \) as \( n \to \infty \) and if \( k(n) \leq \frac{\log n}{h} \), then

\[
\lim_{n \to \infty} \frac{1}{k(n)} H_{k(n)}(x_n^1) = h, \quad a. \ s.
\]

For the proof and further details see [Sh], Theorem II.3.5 and Remark II.3.6. We remark that the same result holds for non-overlapping distributions. The reason why we chose to consider the overlapping one is to enrich the statistic as much as possible, as it will be explained in section 5.

### 4.2 Return times (RT)

Ornstein and Weiss proved an interesting result which links entropy and the so-called return times for ergodic processes. They showed in [OW] that the logarithm of the waiting time until the first \( n \) terms of a sequence \( x \) occur again in \( x \) is almost surely asymptotic to \( nh \).

**Definition 4.2.** Given a sequence \( x \) taken from an ergodic process, we define the \( n \)-th return time as

\[
R_n(x) = \min \{m \geq 1 : x_{m+n+1} = x_1^n \}.
\]

**Theorem 4.3.** If \( \mu \) is an ergodic process with entropy \( h \), then

\[
\lim_{n \to \infty} \frac{1}{n} \log_2 R_n(x) = h, \quad a. \ s.
\]

For the original proof see [OW], for an alternative one see [Sh], Theorem II.5.1.

### 4.3 Lyapunov exponent

If we are interested in the estimation of the entropy of a one dimensional system a powerful tool is the Lyapunov exponent.

Let us consider a map \( T : [0, 1] \to [0, 1] \) having an ergodic invariant measure \( \mu \). We define its Lyapunov exponent by

\[
\lambda_\mu = \int_0^1 \log_2 T' \, d\mu.
\]

Under some assumptions (see below) this quantity is related to the fractal dimension \( HD(\mu) \) of \( \mu \) and the entropy \( h_\mu \) of the system by the formula
HD(\mu) = \frac{h_\mu}{\lambda_\mu}. \text{ Hence if we know } HD(\mu) \text{ and estimate } \lambda_\mu \text{ numerically, we obtain an estimation for } h_\mu. 

Let us give a precise statement for one dimensional systems (see \[LY\] for a generalization to multidimensional systems). A map \(T : [0,1] \to [0,1]\) is called piecewise monotonic if there is a sequence \(\{Z_i\}_{i \in \mathbb{N}}\) of disjoint open subintervals of \([0,1]\) such that \(T|_{Z_i}\) is strictly monotone and continuous for each \(i\).

Let us consider the set \(E_Z = \cap_{i \in \mathbb{N}} T^{-i} (\cup_{j \in \mathbb{N}} Z_j)\), where all iterates of \(T\) are in the open intervals. Let \(\mu\) be an invariant ergodic measure such that \(\mu(E_Z) = 1\). Let us consider its Lyapunov exponent \(\lambda_\mu\) and its K-S entropy \(h_\mu\). Let us denote by \(H_D(X)\) the Hausdorff dimension of a subset \(X \subset [0,1]\).

The Hausdorff dimension \(H_D(\mu)\) of a measure \(\mu\), is defined as the infimum

\[ H_D(\mu) = \inf_{\{X\} = 1} (H_D(X)) \]

of the dimension of full measure sets.

Let us consider the \(p\)-variation of a function \(f : [0,1] \to \mathbb{R}\) on a subinterval \([a,b]\) defined by:

\[ \text{var}^p_{[a,b]}(f) = \sup \left\{ \sum_{i=1}^m |f(x_{i-1}) - f(x_i)|^p : m \in \mathbb{N}, a \leq x_0 < \ldots < x_m \leq b \right\}. \]

We say that the derivative of a piecewise monotonic map has bounded \(p\)-variation if there is a function \(g\) such that \(g(x) = 0\) on \([0,1] \setminus E_Z\), \(g = T'\) on each \(Z_i\) and \(\text{var}^p_{[0,1]}(g) < \infty\).

**Theorem 4.4 \([HR]\).** Let \(T\) be a map on \([0,1]\) with finitely many monotonic pieces and a derivative of bounded \(p\)-variation for some \(p \geq 0\). If \(\mu\) is an ergodic invariant measure with Lyapunov exponent \(\lambda_\mu > 0\), then

\[ HD(\mu) = \frac{h_\mu}{\lambda_\mu}. \]

In many of the systems we will study we have that the invariant measure \(\mu\) we are interested to consider is absolutely continuous with respect to the Lebesgue measure with a regular (bounded variation or continuous) density, hence \(HD(\mu) = 1\).

The Lyapunov exponent will be then numerically estimated with a Birkhoff average along a typical orbit of the system, hence giving

\[ h_\mu = \int_0^1 \log_2 T' d\mu = \lim_{n \to \infty} \frac{\sum_{i=1}^n \log_2 (T'(T^i(x_0)))}{n} \]

for \(\mu\text{-a.e. } x_0\), by the ergodic theorem. Experimental results indicate that this limit converges very fast and gives a very good estimation for \(h_\mu\).
5 Computer simulations

Concerning the results of the computer simulations, some comments are due on the way we implemented the entropy estimating algorithms.

- About empirical frequency estimation, in our simulations we could not consider blocks much longer than 23 bits. This is because the algorithm takes a time which grows exponentially in the length of the blocks considered. The empirical distribution of blocks of various lengths was calculated on the entire symbolic sequence.

- The return times method was performed by calculating the return times of strings long up to \( \log_2 n \), where \( n \) is the length of the symbolic sequences. Moreover, in order to have more reliable results, for every binary sequence we considered not only the return times of the initial strings \( x_1^k \), but also of \( x_2^{k+1}, x_3^{k+2}, \ldots, x_{1000}^{k+999} \), and took the average of their logarithms, hence what we measure is an average return time indicator.

- In the implementation of the NSRPS method, at every step the substitution with a new symbol of a pair with maximum probability was performed, then we calculated the conditional entropy of order 1 and the inverse of the mean shortening \( Z_N \) estimating the entropy according to Corollary 2.6.

The implementation of the substitutions method did not show meaningful computational constraints, since performing a pair substitution requires a very short time. Nevertheless, there is one algorithmic question to be answered: the identification of a stop condition.

For the estimation of the entropy with NSRPS, at the moment we have not an analogous of Theorem 4.1 hence we have to find how many substitutions it is convenient to made on a finite sample string. We had to understand when to stop the substitutions before the sequence becomes too short and consequently the statistics becomes too poor. We chose to stop when the following condition has occurred:

\( \text{StopCond}: \) the substituted pair has frequency \(< 0.02. \)

The stop condition above is somewhat artificial and has no intrinsic relation with the symbolic process. In all the cases we studied we knew the true entropy or estimated it quite precisely by means of the Lyapunov exponent, so that we could understand when the approximation through the pair substitutions method was good. In all our processes, for which we took symbolic samples long 15 millions bits, it seems that
few tens of pair substitutions are enough for the estimate to become
more or less constant when considering the first three decimal digits.
Obviously, when the process is independent or 1-Markov at most one
pair substitution is needed in order to have a very precise estimate of
the entropy. On the contrary, processes which have long memory prop-
erties need many pair substitutions. The stop condition we used does
not take into account the memory properties of the process, so that
it lets the algorithm performing unnecessary pair substitutions in low-
Markov cases and stops it before useful substitutions in long-memory
processes. Although a threshold lower than 0.02 in StopCond could
improve the estimates, the goal is to find some criterion, both user-
independent and sequence-dependent, which determines for each case
the most appropriate number of substitutions to perform.

6 Experiments

We now describe the transformations of the unit interval generating the sym-
bolic sequences to which we applied the entropy estimating algorithms.

6.1 Maps

We considered a few maps of the interval, to which we applied the construc-
tion explained in section 3 to obtain symbolic sequences.

6.1.1 Piecewise expanding maps

We considered a piecewise expanding map $E$, defined by

$$E = \begin{cases} \frac{4x}{3-2x} & \text{if } x \in [0, \frac{1}{2}] \\ \frac{2x-1}{2-x} & \text{if } x \in \left[\frac{1}{2}, 1\right] \end{cases},$$

which is discontinuous in $\frac{1}{2}$ and has two surjective branches (see Figure 6.1). It holds $E'(x) > k$ for all $x$, where $k > 1$ is a constant. As it is well
known (see e. g. [Vi]), a map of this kind has a unique absolutely continuous
invariant measure with dimension 1. Moreover, Theorem 4.4 applies and we
can estimate the entropy by the Lyapunov exponent. A generating partition
for $E$ is $\{[0, \frac{1}{2}], [\frac{1}{2}, 1]\}$ (see [Bu], Exercise 3.4).

We show the results of the entropy estimates in Table 6.1 and Figure 6.1.

The NSRPS method gives the best estimate. Though, the substitutions
themselves have no particular role, since the map seems to be 1-Markov (the
| map | $h_{\text{Lyap}}$ | $h_{\text{EF}}$ | $h_{\text{RT}}$ | $h_{\text{NSRPS}}$ ($N_{\text{sub}}$) |
|-----|------------------|----------------|----------------|----------------------------------|
| $E$ | 0.8673           | 0.865          | 0.838          | 0.867 (17)                       |

Table 6.1: Entropy estimates for the piecewise expanding map $E$. The values $h_{\text{Lyap}}$, $h_{\text{EF}}$, $h_{\text{RT}}$ and $h_{\text{NSRPS}}$ are the entropy estimates as Lyapunov exponent or by empirical frequencies, return times, NSRPS, respectively. $N_{\text{sub}}$ is the number of pair substitutions executed when the stop condition StopCond occurs.

**PIECEWISE EXPANDING MAP**

![Figure 6.1: Piecewise expanding map $E$ and entropy estimates by means of empirical frequencies, return times and NSRPS. The straight line corresponds to the Lyapunov exponent value.](image)

First value calculated with the substitutions algorithm is already very close to the true entropy.
6.1.2 Lorenz-like maps

Another example of map with two non-surjective branches is a Lorenz-like map (similar maps are involved in the study of the famous Lorenz system) defined by

\[
Lx = \begin{cases} 
1 - \left(\frac{-6x + 3}{4}\right)^{\frac{3}{4}} & \text{if } x \in [0, \frac{1}{2}] \\
\left(\frac{6x - 3}{4}\right)^{\frac{3}{4}} & \text{if } x \in [\frac{1}{2}, 1]
\end{cases}
\]

The derivative of \(L\) is uniformly greater than 1 for all \(x \in [0, 1] \setminus \{\frac{1}{2}\}\) and \(L'(\frac{1}{2}) = +\infty\) (see Figure 6.2).

As for the previous piecewise expanding map \(E\), the Lorenz-like map \(L\) has a unique absolutely continuous invariant measure with dimension 1 (see [Vi]). Theorem 4.4 does not apply in this case because the derivative is not bounded and hence has not \(p\)-bounded variation. However the usual relation between entropy and Lyapunov exponent holds and can be recovered by [St]. Moreover, the natural partition \(\{[0, \frac{1}{2}], [\frac{1}{2}, 1]\}\) is generating (see again [Bu]).

In Table 6.2 and Figure 6.2 the results obtained for the map \(L\) are shown.

| map | \(h_{\text{Lyap}}\) | \(h_{\text{EF}}\) | \(h_{\text{RT}}\) | \(h_{\text{NSRPS}} (N_{\text{sub}})\) |
|-----|----------------|----------------|----------------|----------------------------------|
| \(L\) | 0.7419 | 0.764 | 0.723 | 0.748 (17) |

Table 6.2: Entropy estimates for the Lorenz-like map \(L\). \(N_{\text{sub}}\) is the number of pair substitutions executed when the stop condition \(\text{StopCond}\) occurs.

The Lorenz-like map \(L\) appears not to be 1-Markov. In fact, from the plot relative to NSRPS in Figure 6.2 it can be noticed that the best value is not the first estimated, that is simply the 1-st order conditional entropy \(h_1\). Instead, there are pair substitutions that significantly improve the approximation of the entropy. These substitutions are those which condense more information than others. Furthermore, this is one of the cases in which a few more pair substitutions after condition \(\text{StopCond}\) occurs give a better estimate.

6.1.3 Logistic maps

The logistic maps are of the form

\[
A_\lambda x = \lambda x(1 - x), \quad 1 \leq \lambda \leq 4
\]

We took \(\lambda = 4, 3.8\) and 3.6 (the graph of \(A_{3.8}\) is shown in Figure 6.3 (map)). For all these three maps, the partition \(\{[0, \frac{1}{2}], [\frac{1}{2}, 1]\}\) is generating (see [Bu]).
LORENZ-LIKE MAP

Figure 6.2: Lorenz-like map $L$ and entropy estimates by means of empirical frequencies, return times and NSRPS. The straight line corresponds to the Lyapunov exponent value.

For $\lambda = 4$ there is a unique invariant measure, which is ergodic and absolutely continuous with respect to Lebesgue and whose density is $\rho(x) = \frac{1}{\pi \sqrt{x(1-x)}}$. Furthermore, the dynamical system $([0, 1], \mathcal{B}([0, 1]), \rho(x)dx, L_4)$ is isomorphic to the shift on the Bernoulli process with alphabet $\{0, 1\}$ and parameter $\frac{1}{2}$. Thus, for the entropy it holds $h(L_4) = 1$.

About the maps $L_{3.8}$ and $L_{3.6}$ we remark that the assumptions of Theorem 4.4 still hold and the dimension of the invariant measure is estimated to be very close to 1 (see [Sp]). Hence we assume to be reasonable to estimate the entropy by the Lyapunov exponent.

In Table 6.3 we summarize the final entropy estimates obtained with the four methods for the three logistic maps, while in Figure 6.3 we show in graphical form the complete results for the map $L_{3.8}$.

For the map $L_4$ the NSRPS method does not require any substitution
Table 6.3: Entropy estimates for the logistic maps $\Lambda$. $N_{\text{sub}}$ is the number of pair substitutions executed when the stop condition StopCond occurs.

**LOGISTIC MAP**

| map  | $h_{\text{Lyap}}$ | $h_{\text{EF}}$ | $h_{\text{RT}}$ | $h_{\text{NSRPS}} (N_{\text{sub}})$ |
|------|------------------|-----------------|-----------------|----------------------------------|
| $\Lambda_4$ | 1.0000 | 0.997 | 0.959 | 1.000 (17) |
| $\Lambda_{3.8}$ | 0.6234 | 0.652 | 0.610 | 0.628 (18) |
| $\Lambda_{3.6}$ | 0.2646 | 0.348 | 0.314 | 0.269 (18) |

Figure 6.3: Logistic map for $\lambda = 3.8$ and entropy estimates by means of empirical frequencies, return times and NSRPS. The straight line corresponds to the Lyapunov exponent value.

to correctly estimate the entropy up to the sixth decimal digit. This is no surprise, since the symbolic process associated with $\Lambda_4$ is independent.

Instead, for the map $\Lambda_{3.8}$ it happens that, similarly to the NSRPS case
of the map \(L\) (see Figure 6.2), there are pair substitutions which are more important than others in approximating the value of the entropy, as it can be noticed in Figure 6.3.

The entropy estimating algorithms give for the map \(A_{3,6}\) results that are qualitatively similar to those of \(A_{3,8}\).

### 6.1.4 Manneville-Pomeau maps

Manneville maps exhibit dynamics with long range correlations. They are defined by

\[ M_z x = x + x^z \pmod{1}, \quad z \in \mathbb{R}^+ \]

Such maps have great interest in physics and possess different characteristics as the exponent \(z\) varies. We focused our attention on the values \(1 < z < 2\), for which the maps admit a unique absolutely continuous invariant probability measure (with unbounded density). For these parameters, the system has power law decay of correlations, and the rate is slower and slower as \(z\) approaches 2 (see [Vi], section 3 e.g.). In this case the system has “long memory” and to estimate entropy by the empirical frequencies we would need long blocks. For \(z \geq 2\) the absolutely continuous invariant measure is no longer finite. We also remark that since those maps have bounded variation derivative, in the cases where the absolutely continuous invariant measure is finite we can again estimate the entropy by the Lyapunov exponent. We took values of \(z\) which go very close to 2: \(z_1 = \frac{3}{2}\), \(z_2 = \frac{7}{4}\), \(z_3 = \frac{15}{8}\), \(z_4 = \frac{31}{16}\), \(z_5 = \frac{63}{32}\), \(z_6 = \frac{127}{64}\) (see the plot of \(M_{z_4}\) in Figure 6.4). For all \(1 \leq i \leq 6\) it holds \(M'_{z_i}(x) > 1\) for all \(x \in [0,1]\) and \(M'_{z_4}(0^+) = 1\). For these maps the natural partitions \(\{[0,c_i],[c_i,1]\}\), where \(c_i \in [0,1]\) is that value such that \(M_{z_i}(c_i^-) = 1\) and \(M_{z_i}(c_i^+) = 0\), are obviously generating.

The presence in 0 of an indifferent fixed point is the main responsible for the peculiar behaviour of the Manneville maps. When, starting from a random point \(x_0\), after a certain number \(n\) of iterations the point \(M^n_{z_4} x_0\) happens to be very close to 0, the subsequent iterations remain very close to 0 for a long time. This fact translates in having many consecutive zeros in the binary symbolic sequence associated with the orbit of \(x_0\). These strings of zeros can be long even hundreds of thousands of bits or more. The closer to 2 is the exponent \(z\), the longer and more frequent these strings.

In carrying out the simulations for the Manneville maps and commenting their results, one cannot ignore the peculiarities of these maps. It turns out that the symbolic sequences we generated are too short to reflect the general characteristics of the maps. If in a sequence of 15 millions bits there happen to be groups of consecutive zeros that are hundreds of thousands of bits long, then the results obtained from such a sequence cannot be completely
reliable. The usual approach to this problem is considering many sequences, generated from different initial random points, and taking the averages of the estimates. For the map \( M_{zi} \) we considered \( 2^i \) sequences, with \( 1 \leq i \leq 6 \). Still, the values obtained from the various sequences are quite different, so that we cannot consider completely reliable the averages as well.

Bearing in mind these considerations, we report in graphic form the results for the Manneville map \( M_{z4} = z + z^{16} \) (mod 1) (see Figure 6.4), while the results for all the six Manneville maps considered are shown in Table 6.4.

| map   | \( h_{\text{Lyap}} \) | \( h_{\text{EF}} \) | \( h_{\text{RT}} \) | \( h_{\text{NSRPS}} (N_{\text{sub}}) \) |
|-------|------------------|----------------|-----------------|------------------|
| \( M_{2} \) | 0.811            | 0.804          | 0.821           | 0.813 (18)       |
| \( M_{3} \) | 0.519            | 0.522          | 0.558           | 0.511 (20)       |
| \( M_{3} \) | 0.314            | 0.340          | 0.442           | 0.322 (19)       |
| \( M_{4} \) | 0.228            | 0.244          | 0.444           | 0.226 (21)       |
| \( M_{5} \) | 0.175            | 0.234          | 0.400           | 0.216 (21)       |
| \( M_{6} \) | 0.168            | 0.214          | 0.358           | 0.196 (21)       |

Table 6.4: Entropy estimates for the Manneville maps \( M_{zi} \). \( N_{\text{sub}} \) is the average number of pair substitutions executed when the stop condition \( \text{StopCond} \) occurs.

For each Manneville map that we studied (except for \( M_{2} \)), the entropy estimates obtained through the NSRPS method were clearly the closest to the true entropy (which we assumed to be equal to the average Lyapunov exponent), although they were not as close as for the other maps or processes (see section 6.2).

### 6.1.5 A skew product

We consider an example of a two dimensional system having long range correlations which is quite different from the Manneville map. Let us consider the following map \( S : [0, 1]^2 \rightarrow [0, 1]^2 \) defined by

\[
S(x, y) = (Ex, y + \alpha \phi(x) \mod 1)
\]  

(6.1)

where: \( \phi(x) = \begin{cases} 
1 & \text{if } x \geq \frac{1}{2} \\
0 & \text{if } x < \frac{1}{2}
\end{cases} \), \( \alpha \) is a diofantine irrational and \( E \) is the one dimensional piecewise expanding map considered in section 6.1.1. In the
Figure 6.4: Manneville map with $z = \frac{31}{16}$ and entropy estimates by means of empirical frequencies, return times and NSRPS. The straight line corresponds to the Lyapunov exponent value.

system the $x$ coordinate is subjected to a chaotic transformation, while the $y$ is rotated according to the value of $x$. Such systems preserve an absolutely continuous invariant measure and are mixing. Some estimations for the decay of correlations are given in [Dol].

We partitioned the unit square in four equal squares $Q_1, ..., Q_4$ having a common vertex at $(\frac{1}{2}, \frac{1}{2})$.

The entropy of $S$ with respect to the partition $\{Q_1, ..., Q_4\}$ is the same as the entropy of $E$, indeed the rotation has zero entropy and a symbolic orbit for the two dimensional system can be constructed by the information given by its symbolic orbit for the one dimensional map $E$ and the information relative to the rotation part. Although the entropy is the same, its estimation is much more complicated, as the experiments show.

In figure 6.5 we consider the case where $\alpha = \frac{1+\sqrt{5}}{2}$ is the golden ratio. The
empirical frequencies and the substitutions seem to converge to a value which is slightly greater than the true entropy. The return time instead seems to better approximate the entropy in this case.

**SKEW PRODUCT**

Figure 6.5: Results for the skew product $S$: entropy estimates by means of empirical frequencies, return times and NSRPS. The straight line corresponds to the entropy value.

### 6.2 Renewal processes

Apart from the symbolic sequences obtained from ergodic transformations of the unit interval, we considered sequences taken from the so-called renewal processes.

A renewal process is a stationary process with alphabet $\{0, 1\}$ for which the distances between consecutive ones are independent and identically distributed random variables. When a symbol ‘1’ occurs, the sequence forgets
all its past and the probability of having the next ‘1’ after \( j \) bits is \( p_j \), where \( 0 \leq p_j \leq 1 \) and \( \sum_{j=1}^{\infty} p_j = 1 \).

We considered such renewal processes, with \( p_1 = p_2 = \ldots = p_{2^k} = \frac{1}{2^k} \), \( 5 \leq k \leq 9 \), which we shall indicate with \( RP_{2^k} \).

For these renewal processes, the value of their entropy can be calculated exactly. We recall in fact that the entropy of a process is the number of bits per symbol that are necessary to describe the process itself. The quantity \( C = -\sum j p_j \log_2 p_j \) represents the number of bits that one needs to describe the process of the jumps between consecutive ones. In other words, \( C \) is the entropy of a random variable which describes the length of the jumps. If \( n \) is large, with \( n \) jumps (\( nC \) bits) we describe a sequence long about \( n \bar{L} \) symbols, where \( \bar{L} \) is the average length of the jumps. Thus,

\[
h(RP_{2^k}) \approx \frac{nC}{n \bar{L}} = \frac{\sum_{j \geq 1} j p_j \log_2 p_j}{\sum_{j \geq 1} j p_j}.
\]

In our cases, where \( p_1 = \ldots = p_{2^k} = \frac{1}{2^k} \) and \( p_j = 0 \) for \( j > 2^k \), we have

\[
h(RP_{2^k}) = \frac{2k}{2^k + 1}.
\]

In Table 6.5 we show the results of the entropy estimates for the renewal processes \( RP_{2^k} \) and those of \( RP_{32} \) are also plotted in Figure 6.6.

| map   | \( h \)    | \( h_{EF} \) | \( h_{RT} \) | \( h_{NSRPS} (N_{sub}) \) |
|-------|------------|--------------|--------------|-----------------------------|
| \( RP_{32} \) | 0.303030 | 0.320        | 0.272        | 0.303067 (11)              |
| \( RP_{64} \) | 0.184615 | 0.196        | 0.153        | 0.184793 (22)              |
| \( RP_{128} \) | 0.108527 | 0.115        | 0.110        | 0.108498 (25)              |
| \( RP_{256} \) | 0.062257 | 0.066        | 0.055        | 0.062239 (18)              |
| \( RP_{512} \) | 0.035088 | 0.037        | 0.039        | 0.035112 (16)              |

Table 6.5: Entropy estimates for the renewal processes \( RP_{2^k} \). \( N_{sub} \) is the number of pair substitutions executed when the stop condition StopCond occurs.

For this process, the substitutions method gives an excellent approximation of the entropy already after five pair substitutions. After these substitutions all the memory of the process has been transferred to the distribution of the pairs, so that the sequence has become 1-Markov.
7 Conclusions and final remarks

The performance of the three symbolic methods is summarized in Figure 7.1. Summarizing, NSRPS results to be the method that best approximates the entropy value. To this aim, it is a fast and light computational tool that may be used also for systems having low entropy or long range correlations where other statistical methods fail.

This paper shows for the first time a comparison in entropy estimation among NSRPS and other well-known methods. The results also open some further questions about NSRPS:

- how to prove an analogous of Theorem 4.1 for NSRPS giving a sufficient number of substitutions in function of the length of the string?
Figure 7.1: Entropy estimates for the maps $\Lambda_{3.8}$, $E$, $L$, $M_{31}$ and the renewal process $RP_{32}$; symbol empty □ refers to the NSRPS value under StopCond condition; full □ refers to the EF value and ◦ refers to the RT method. Straight lines show the entropy values (for the maps they are the estimated Lyapunov exponents).

- are there other meaningful substitution methods (different from the recipe given in Theorem 2.5) that may be proved to be (at least) sufficient for Theorem 2.4 to hold?

- can the joint use of NSRPS and Lyapunov exponent (which are both fast converging and fastly computable) together with Theorem 4.4 give a particularly good method to numerically estimate the Hausdorff dimension of an attractor?

- concerning the applications of NSRPS to non-artificial processes, such as literary texts, biological sequences (DNA, proteins) and time series in general, what interesting features of the driving dynamics may be extracted?

- NSRPS method might be the core of some data compression algorithm (see [LaMo]). This should pave the way to some investigations towards its compression capabilities in comparison with other well-known algorithms. We remark that data compression procedures have also been successfully used as entropy estimators (see e.g. [BBGMV] and [GKB]).

Acknowledgements. The work of G.M. was supported by a post-doc
research scholarship “Compagnia di San Paolo” awarded by the Istituto Nazionale di Alta Matematica “F. Severi”.

References

[AA] V. I. Arnol’d, A. Avez, Problèmes ergodiques de la mécanique classique, Gauthier-Villars, Paris, 1967.

[BBGMV] V. Benci, C. Bonanno, S. Galatolo, G. Menconi, M. Virgilio, Dynamical systems and computable information, Discrete and Continuous Dynamical Systems - B 4, 4 (2004), 935–960.

[BCG] D. Benedetto, E. Caglioti, D. Gabrielli, Non-sequential recursive pair substitution: some rigorous results, Journal of Statistical Mechanics (2006) P09011.

[BHM] J. Bonachela, H. Hinrichsen, M. Muoz Entropy estimates of small data sets J. Phys. A Math. Theor. 41 202001 (2008)

[Bu] J. Buzzi, A minicourse on entropy theory on the interval, arXiv:math/0611337 2006.

[Dol] D Dolgopyat On mixing properties of compact group extensions of hyperbolic systems , Israel J. Math. , 130 (2002) 157-205.

[EJM] W. Ebeling, M. A. Jiménez-Montaño, On grammars, complexity, and information measures of biological macromolecules, Mathematical Biosciences 52 (1980), 53–71.

M. A. Jiménez-Montaño, On the syntactic structure of protein sequences and the concept of grammar complexity, Bulletin of Mathematical Biology 42 (1984), 641–659.

M. A. Jiménez-Montaño, W. Ebeling, T. Pöschel, SYNTAX: a computer program to compress a sequence and to estimate its information content, arXiv:cond-mat/0204134 2002.

P. E. Rapp, I. D. Zimmermann, E. P. Vining, N. Cohen, A. M. Albano, M. A. Jiménez-Montaño, The algoritmic complexity of neural spike trains increases during focal seizures, The Journal of Neuroscience 14 (1994), 4731–4739.

[GKB] Y. Gao, I. Kontoyiannis, E. Bienenstock, Estimating the Entropy of Binary Time Series: Methodology, Some Theory and a Simulation Study, Entropy 10, 2 (2008), 71–99.
[Gr] P. Grassberger, *Data compression and entropy estimates by non-sequential recursive pair substitution*, [arXiv:physics/0207023](http://arxiv.org/abs/physics/0207023), 2002.

[HR] F. Hofbauer, P. Raith, *The Hausdorff dimension of an ergodic invariant measure for a piecewise monotonic map of the interval*, Canadian Mathematical Bulletin 35 (1992), 84–98.

[MEPR] Jimenez-Montano M.A., Ebeling W., Pohl T., Rapp P.E., ”Entropy and complexity of finite sequences as fluctuating quantities”, BioSystems 64 (2002) 23-32

[Kr] W. Krieger, *On entropy and generators of measure-preserving transformations*, Transactions of the American Mathematical Society 149 (1970), 453–464.

[LaMo] Larsson, N.J. Moffat, A., *Off-line dictionary-based compression*, Proceedings of the IEEE, 88, 11 (2000), 1722-1732.

[LeMi] F. Ledrappier, M. Misiurewicz, *Dimension of invariant measures for maps with exponent zero*, Ergodic Theory and Dynamical Systems 5 (1985), 595–610.

[LY] F. Ledrappier, L.-S. Young, *The metric entropy of diffeomorphisms. II. Relations between entropy, exponents and dimension*, Annals of Mathematics 122 (1985), 540–574.

[OW] D. Ornstein, B. Weiss, *How sampling reveals a process*, The Annals of Probability 18 (1990), 905–930.

[Sh] P. C. Shields, *The ergodic theory of discrete sample paths*, American Mathematical Society, Providence, RI, 1996.

[Sp] J. C. Sprott, G. Rowlands, *Improved correlation dimension calculation*, International Journal of Bifurcation and Chaos 11 (2001), 1865–1880.

[St] T. Steinberger, *Local dimension of ergodic measures for two-dimensional Lorenz transformations*, Ergodic Theory and Dynamical Systems 20 (2000), 911–923.

[Vi] M. Viana, *Stochastic dynamics of deterministic systems*, Lecture Notes XXI Brazilian Mathematics Colloquium, Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1997.