On Lempel-Ziv complexity for multidimensional data analysis

Steeve Zozor, Philippe Ravier, Olivier Buttelli

On Lempel-Ziv complexity for multidimensional data analysis

S. Zozor [a], P. Ravier [b] and O. Buttelli [c]

[a] *Laboratoire des Images et des Signaux*
Rue de la Houille Blanche
B.P. 46
38420 Saint Martin d'Hères Cedex
France
Phone # +33 4 76 82 64 23
Fax # +33 4 76 82 63 84
E-mail: Steeve.Zozor@lis.inpg.fr

[b] *Laboratoire d'Électronique, Signaux, Images, 12 rue de Blois, B.P. 6744, 45067 Orléans Cedex 2, France*

[c] *Laboratoire Activité Motrice et Conception Ergonomique, Rue de Vendôme, B.P. 6237, 45062 Orléans Cedex 2, France*

Abstract

In this paper, a natural extension of the Lempel-Ziv complexity for several finite-time sequences, defined on finite size alphabets is proposed. Some results on the defined joint Lempel-Ziv complexity are given, as well as properties in connection with the Lempel-Ziv complexity of the individual sequences. Also, some links with Shannon entropies are exhibited and, by analogy, some derived quantities are proposed. Lastly, the potential use of the extended complexities for data analysis is illustrated on random boolean networks and on a proposed multidimensional extension of the minority game.

Key words: Complexity measures, Lempel-Ziv complexity, Shannon Entropy, nonlinear deterministic multidimensional systems, random boolean network, minority game.
PACS: 87.10.+e, 02.50.Le, 64.60.Cn, 02.50.Sk

1 Introduction

Many physical, biological or financial signals result from the dynamics of great dimensional systems. As an example, in biology, the cells of the cardiac tissue

exchange ions with the extracellular field by a nonlinear reaction, and with the connected cells by (linear) diffusion. Hence, electric waves propagate on the tissue, and the electrocardiogram (ECG) is the electric field produced by the propagation and measured by an electrode. Another example can be found in the collective actions of the genes, generating the production of proteins in certain quantities. In finance, one can take the example of the variation in the price of an asset, resulting from the collective action of the buyers and sellers. In all these examples, the challenge is to describe or extract meaningful information from the data [1,2], to characterize, to detect or to classify different pathologies [3–10], etc.

Due to the complex origin of the signals, researchers generally use tools from either information theory or from dynamic nonlinear processes. The goal is to characterize as well as possible the degree of organization of the measured signals. The first approach is statistical, and the tools generally employed are entropies [4,6,11–13] (or "approximate entropies" or multi-resolution entropies), correlation measures [3], spectral analysis [14], etc. The second approach is based on the fact that the underlying mechanisms producing the measured signals are generally deterministic and nonlinear. This is clear in the example of ECG, even if measurements may be corrupted by noise. The nonlinear tools generally employed to study the signals come from chaos theory such as Lyapunov exponents or dimensions (fractal, etc.) [1,5,9]. Other tools often employed come from the concept of complexity in the sense of Kolmogorov, more particularly Lempel-Ziv complexity [2,8,10,15–18]. Several complexity measures for data analysis have already been proposed as presented in [11] and in the references therein. However, in spite of the terminology, the approach is generally more likely statistical. In this paper, we will focus on Lempel-Ziv complexity. The motivation for this is that the data generally studied in biomedical engineering have a nonlinear and deterministic origin. Furthermore, tools such as Lyapunov exponents are difficult to evaluate [1] and require long time-computation. In contrast, the Lempel-Ziv complexity contains the notion of complexity in the deterministic sense (Kolmogorov sense) as well as in a statistical sense (Shannon sense), and can be computed with a low cost. In the literature, this tool is generally employed to analyze mono-dimensional signals. In this paper we show that contrary to what has been previously claimed [8], this tool can be used for multidimensional signals. The goal is to use the natural extension of the Lempel-Ziv complexity to characterize the "regularity" of a system using several signals.

Section 2 recalls the basics on Lempel-Ziv complexity (LZC). Section 3 then shows how the LZC can naturally be extended for multidimensional sequences.

[1] reconstruction of the phase space, with several estimations to determine the embedding dimension and the optimal delay; then, estimation of the whole Lyapunov spectrum or just of some exponents (positive, max...)

In this section we will give some properties of multidimensional LZC in connection with "marginal" LZCs. Some parallels and differences with Shannon entropy will be exhibited. In section 4 we will illustrate how the extended LZC can capture the spatiotemporal organization of the signal delivered by a multidimensional nonlinear and nonlinearly coupled system. The illustration will be done for two particular binary systems: random boolean networks and an extension of the minority game. Finally, section 5 will give some conclusions and perspectives.

2 Basics on the Lempel-Ziv complexity

Consider a sequence (or a word) $S = s_1 \dots s_n$ of length n where each letter s_i is taken from an alphabet $\mathcal{A}$ of finite size α. Kolmogorov in 1965 defined the complexity of such a sequence as the size of the smallest binary program that can produce this sequence [19]. The complexity in the sense of Kolmogorov seems very general and computer-dependent. It was shown that even if a Kolmogorov complexity can be defined up to a constant, the evaluation of such a complexity in a finite time is not guaranteed [19]. Several years later, in their seminal work [20], A. Lempel and J. Ziv proposed to define a complexity in the sense of Kolmogorov, but limiting their definition to programs based on two operations: recursive copy and paste operations. They defined two fundamental notions to evaluate the Lempel-Ziv complexity (LZC): production and reproduction operations. As in the original paper, we will denote SQ the concatenation of two words S and Q, $\ell(S)$ the length of a sequence S, $S(i,j)$ the sub-sequence $s_i s_{i+1} \dots s_j$ of S and π the operation suppressing the last letter of a sequence, *i.e.* $S\pi = S(1, \ell(S) - 1)$.

Reproduction: An extension $R = SQ$ of a sequence S is said reproducible from S if the sequence Q is in the vocabulary of $SQ\pi$, *i.e.* if Q is a sub-sequence of $SQ\pi$. As an example, for $S = 101$ and $Q = 010$, $R = SQ = 101010$ is a reproducible extension of S since $q_1 = s_2 = r_2$, $q_2 = s_3 = r_3$ and $q_3 = q_1 = r_4$ previously copied, and Q is a sub-sequence of $SQ\pi = 10101$. In other words, SQ can be reproduced from S by recursive copy and paste operations. As in the initial paper, we will denote the reproduction by $S \longrightarrow R$ and the index $p \leq \ell(S)$ such that $Q = R(p, \ell(Q) + p - 1)$ is called the pointer for the reproduction $S \longrightarrow R$.

Production: A non-empty sequence S is said to be producible from a prefix $S(1,j)$ if $S(1,j) \longrightarrow S\pi$ and $j < \ell(S)$. From the previous example, it can be seen that $S = 101$ produces $SQ = 101010$ since S can reproduce $SQ\pi = 10101$. But S also produces $T = 101011$ while S does not reproduce this sequence. The difference between reproduction and production is that in production the last letter can come from a supplementary copy-paste but can also be "new".

We will denote production $S(1,j) \Longrightarrow S$ and $S(1,j)$ is called a basis of S.

To understand how a program can build a sequence using these two operations, consider a given word S. An index h_2 can be found such that $s_1 \Longrightarrow S(1,h_2)$. $h_2 = 2$ always works, and if $s_2 = s_1$ one can also choose $h_2 = 3$. If $s_2 = s_1$ and $s_3 = s_2$, then $h_2 = 4$ also works; and so on. Then, there is an index h_3 so that $S(1,h_2) \Longrightarrow S(1,h_3)$. And so on. Thus, S can be built as $s_1 = S(1,h_1) \Longrightarrow S(1,h_2) \Longrightarrow S(1,h_3) \cdots \Longrightarrow S(1,h_m) = S$ [20]. The process $Hi(S) = S(1,h_1)S(h_1+1,h_2) \ldots S(h_{m-1}+1,h_m)$ is called the history of the production and the sub-sequences $S(h_{i-1}+1,h_i)$ are called the components of the process. The size of the process is then defined as the number of components of the process, *i.e.* $c_{Hi}(S) = h_m$. Following the idea of Kolmogorov, Lempel and Ziv sought the shortest production possible process. Hence they defined a complexity in the sense of Lempel-Ziv as

$$c(S) = \min_{Hi \in \{\text{histories of } S\}} c_{Hi}(S) \tag{1}$$

This quantity is not precisely a complexity since it does not give directly the program and its size that can produce S, but $c(S)$ is directly linked to the size of such a program. As an illustration of LZC, for the sequence $S = 0100100100$ the process $0 \Longrightarrow 01 \Longrightarrow 0100 \Longrightarrow 0100100 \Longrightarrow 010010010 \Longrightarrow 0100100100$ is of size 6 while the minimal process $0 \Longrightarrow 01 \Longrightarrow 0100 \Longrightarrow 0100100100$ is only of size 4. In this example, it can be seen that in the minimal process each $S(1,h_{i+1})$, except the last one, cannot be reproduced but only produced by $S(1,h_i)$. In [20], Lempel and Ziv defined as exhaustive components $S(h_{i-1}+1,h_i)$ all the components that can be produced but not reproduced by $S(1,h_{i-1})$ and a history of a process is said to be exhaustive if all the components are exhaustive, except possibly the last one. Then, they showed that an exhaustive history is unique and that the number of components of the exhaustive history is precisely the LZC [20]. This last remark led to algorithms to evaluate the LZC of a sequence (see *e.g.* [21]). With the procedure to evaluate the LZC, it can be seen that if the length of each component, each pointer and the last letter of each component are stored, the complete sequence can easily be retrieved [2].

In conclusion, the LZC contains the notion of complexity in the sense of Kolmogorov. Furthermore, if the sequence $S_{(n)} = s_1 \ldots s_n$ is random and if the source is stationary and ergodic it can be shown that $\lim_{n\to+\infty} c(S_{(n)})\frac{\log(n)}{n} = H(S)$ where $H(S)$ is the entropy rate of the source [19,20]: this result says that the LZC also contains a notion of "average information quantity" in the Shannon sense.

[2] Notice that it is the idea used in many well known compression algorithms, such as "gzip". Other parsing schemes also exist [22], leading to several variants of compression algorithms.

Since the LZC tries to capture a degree of redundancy, or patterns that are similar in a sequence, this tool seems interesting for the analysis of sequences that appear complex but that may hide some simple underlying behaviors. In this way, the LZC has been proposed to analyze chaotic sequences [21]. Furthermore, since the LZC can also be viewed as an estimation of Shannon entropy, it seems that it can also be used for non deterministic signals. This tool seems to bridge the two above- mentioned approaches. However, although the LZC has already been employed for data analysis [8,10,15–18], it was only used for scalar sequences. In the next section, we will show that the LZC is in fact also naturally defined for vectorial sequences and thus can be applied for multidimensional data analysis. Furthermore we will draw a parallel with Shannon entropy, pointing out some similarities, but also some differences.

3 Multidimensional Lempel-Ziv complexity

As far as we know, the first attempt to use the LZC for spatiotemporal data analysis was made by Kaspar and Schuster [21]. In their paper, they proposed to analyze spatiotemporal signals defined on a finite size alphabet by computing a spatial LZC at each time t, *i.e.* for N sequences $x_i(t)$, $i = 0, \ldots, N-1$, $c(t) = c(x_0(t) \ldots x_{N-1}(t))$. Then, with their approach, the time evolution of the spatial complexity $c(t)$ can be analyzed to deduce a spatiotemporal behavior. As an example, if the spatial LZC decreases with time, it can be concluded that the signal/system tends to be more and more organized. However, such an approach remains only spatial, and the temporal links are omitted. To better understand this fact, if the signal has only two components (*e.g.* from two sensors), by definition the spatial LZC remains constant and equal to 2 at each time t. Hence, the evolution of the spatial LZC fails in describing the spatiotemporal behavior of the signal: in this example the spatial dimension is not large enough.

A more natural approach is to extend directly the LZC for vectorial data. This can be done naturally by extending the alphabet. To this aim, consider k sequences $X_i = x_{i,1} \ldots x_{i,n}$ for $i = 0, \ldots, k-1$, where the letters are respectively in the alphabets $\mathcal{A}_0, \ldots, \mathcal{A}_{k-1}$ of respective sizes $\alpha_0, \ldots, \alpha_{k-1}$. Consider now the sequence $Z = z_1 \ldots z_n$, defined on the extended alphabet $\mathcal{B} = \mathcal{A}_0 \times \cdots \times \mathcal{A}_{k-1}$ of size $\alpha_0 \ldots \alpha_{k-1}$, where the components are the k-uplets $z_j = (x_{0,j}, \ldots, x_{k-1,j})$. Notice that Z is a sequence of n k-uplets and not a sequence of $k \times n$ letters: sequence Z does not result from a letter mixing approach. Since the equality relation holds for k-uplets, the production and reproduction operations defined by Lempel and Ziv hold for k-uplets. As a conclusion, all the work of Lempel and Ziv remains valid for vectorial sequences, although not explicitly spelled out in their paper [20]. But once again, it is important to notice that for vectorial sequences the alphabet has

no scalar elements but only k-uplet elements. Hence, we can define what we will call the joint Lempel-Ziv complexity of sequences $X_0, \ldots, X_{k-1}$ by

$$c(X_0, \ldots, X_{k-1}) = c(Z) \tag{2}$$

Furthermore, if the alphabets are the same and are of the form $\mathcal{A} = \{0, \ldots, \alpha - 1\}$, we can also define a sequence $Z = z_1 \ldots z_n$ considering that each z_j has the $x_{i,j}$ as α-ary decomposition, *i.e.* $z_j = \sum_{i=0}^{k-1} x_{i,j}\alpha^i$. Defining joint LZC of the X_i as that of Z is exactly similar to the previous definition. The last assumption is not restrictive since bijections can be found from the $\mathcal{A}_k$ to a subset of $\mathcal{A}$ of size α_k to achieve such a case, provided that $\alpha \geq \max_k \alpha_k$. Moreover, using this approach, the algorithm proposed in [21] to evaluate the LZC can still be used, comparing scalars. With this definition, the LZC of multidimensional signals can then be viewed as a joint LZC and analyzed regarding the LZC of its components. The first property of the joint LZC is its obvious symmetry by permutation of the X_j:

Property 1 *The joint LZC is invariant by any permutation σ of $\{0, \ldots, k-1\}$, i.e.*

$$c(X_0, \ldots, X_{k-1}) = c(X_{\sigma(0)}, \ldots, X_{\sigma(k-1)}) \tag{3}$$

PROOF. Consider the exhaustive history of the vectorial sequence $Z = z_1 \ldots z_n$ where $z_j = (x_{0,j}, \ldots, x_{k-1,j})$, *i.e.* $E_Z(Z) = Z(1, h_1) \ldots Z(h_{m-1} + 1, h_m)$. Hence, whatever $i = 2, \ldots, m$, we have both $Z(1, h_{i-1}) \Longrightarrow Z(1, h_i)$ and $Z(1, h_{i-1}) \not\rightarrow Z(1, h_i)$. Thus, for any i, let us denote p_i the pointer $p_i \leq h_{i-1}$ such that both $Z(h_{i-1} + 1, h_i - 1) = Z(p_i, p_i + h_i - h_{i-1} - 2)$ and $z_{h_i} \neq z_{p_i + h_i - h_{i-1} - 1}$. Furthermore, for any $q_i \leq h_{i-1}$, $Z(h_{i-1} + 1, h_i) \neq Z(q_i, q_i + h_i - h_{i-1} - 1)$. The direct consequence is for all j, $X_j(h_{i-1} + 1, h_i - 1) = X_j(p_i, p_i + h_i - h_{i-1} - 2)$ and $(x_{0,h_i}, \ldots, x_{k-1,h_i}) \neq (x_{0,p_i+h_i-h_{i-1}-1}, \ldots, x_{k-1,p_i+h_i-h_{i-1}-1})$. Furthermore, whatever $q_i \leq h_{i-1}$ there is at least one $j_{q_i} \in \{0, \ldots, k-1\}$ so that $X_{j_{q_i}}(h_{i-1} + 1, h_i) \neq X_{j_{q_i}}(q_i, q_i + h_i - h_{i-1} - 1)$. The equality between the subsequences and the inequality between the k-uplets remain unchanged by any permutation σ in $\{0, \ldots, k-1\}$ and the inequality between the sequences holds with $\sigma(j_{q_i})$, that finishes the proof. □

Due to this property, the joint LZC is uniquely defined. Note that a definition using a sample mixing approach would destroy this symmetry (*e.g.* $X_0 = 00001$ and $X_1 = 10100$ leads to $c(X_0, X_1) = c(X_1, X_0) = 4 \neq c(x_{0,1}x_{1,1} \ldots x_{0,5}x_{1,5}) = c(0100010010) = 5 \neq c(x_{1,1}x_{0,1} \ldots x_{1,5}x_{0,5}) = c(1000100001) = 4$). Furthermore, by a sample mixing approach, one can feel that the temporal links of a sequence will be less well captured than by using this natural approach.

Since the joint LZC is simply an LZC as defined by Lempel and Ziv, it possesses all the properties of the LZC given in [20]. The first of them is the asymptotic link with Shannon entropy:

Property 2 *Consider k random sequences $X_0, \ldots, X_{k-1}$ that are jointly stationary and ergodic. Hence*

$$\lim_{n \to +\infty} c(X_0, \ldots, X_{k-1}) \frac{\log(n)}{n} = H(X_0, \ldots, X_{k-1}) \tag{4}$$

where $H(X_0, \ldots, X_{k-1})$ is the joint entropy rate of the sequences.

PROOF. Consider the sequence $Z = z_1 \ldots z_n$ where $z_j = (x_{0,j}, \ldots, x_{k-1,j})$ Since the X_i are jointly stationary and ergodic, Z is stationary and ergodic. On the extended alphabet, from [20], $\lim_n c(Z) \frac{\log(n)}{n} = H(Z) = \lim_n \frac{E[-\log(\Pr[z_1,\ldots,z_n])]}{n}$. Since $\Pr[z_1, \ldots, z_n] = \Pr[x_{0,1}, \ldots, x_{k-1,1}, \ldots, x_{0,n}, \ldots, x_{k-1,n}]$ we obtain that $H(Z) = H(X_0, \ldots, X_{k-1})$ that finishes the proof. □

This complexity clearly exhibits the fact that both spatial and temporal relations are captured by the joint LZC, at least statistically and asymptotically.

Notice now that in the LZC sense, a multidimensional signal is obviously more complex than each single component. This can be summarized as follows.

Property 3 *The joint LZC of k sequences $X_0, \ldots, X_{k-1}$ is greater than or equal to the LZC of each component, that reads*

$$c(X_0, \ldots, X_{k-1}) \geq \max(c(X_0), \ldots, c(X_{k-1})) \tag{5}$$

PROOF. Consider sequence Z of the k-uplets $(x_{0,j}, \ldots, x_{k-1,j})$ and its exhaustive history $E_Z(Z) = Z(1, h_1) \ldots Z(h_{m-1}+1, h_m)$. Since for all i we have $Z(1, h_{i-1}) \Longrightarrow Z(1, h_i)$, obviously for all j, $X_j(1, h_{i-1}) \Longrightarrow X_j(1, h_i)$. As a consequence $E_Z(X_j)$ is also a production process of X_j. Using the definition of the LZC, $c(X_j) \leq c_{E_Z}(X_j) = c(Z)$. □

This property, as well as the previous one, are shared with the Shannon entropies.

Notice that $Z(1, h_{i-1}) \not\rightarrow Z(1, h_i)$ does not imply the same property for all X_j; this property holds at least for one of the X_j. Hence, the exhaustive history of

Z is not necessarily exhaustive for the X_j and it can even be not exhaustive for any X_j. The consequence is that the equality says nothing more in terms of links between the joint sequence and the marginal one. The following property give some links in particular cases.

Property 4 *Consider X and Y two sequences defined in alphabets $\mathcal{A}_x$ and $\mathcal{A}_y$ respectively. If there is a bijection σ from $\mathcal{A}_x$ to $\mathcal{A}_y$ such that $y_i = \sigma(x_i)$ for all $i = 1, \ldots, n$, then*

$$c(X, Y) = c(X) = c(Y) \tag{6}$$

A particular case is

$$c(X, X) = c(X) \tag{7}$$

The reciprocal property is false.

PROOF. Consider first the exhaustive history of X, $E_X(X) = X(1, h_1) \ldots X(h_{m-1} + 1, h_m)$. Since $X(1, h_{i-1}) \Longrightarrow X(1, h_i)$ and $X(1, h_{i-1}) \not\rightarrow X(1, h_i)$, there is a pointer $p_i \leq h_{i-1}$ such that $X(h_{i-1} + 1, h_i - 1) = X(p_i, p_i + h_i - h_{i-1} - 2)$, that $x_{h_i} \neq x_{p_i+h_i-h_{i-1}-1}$ and whatever $q_i \leq h_{i-1}$, $X(h_{i-1} + 1, h_i) \neq X(q_i, q_i + h_i - h_{i-1} - 1)$. Clearly $(\sigma(x_{h_{i-1}+1}), \ldots, \sigma(x_{h_i-1})) = (\sigma(x_{p_i}), \ldots, \sigma(x_{p_i+h_i-h_{i-1}-2}))$ and the bijectivity of function σ yields that $\sigma(x_{h_i}) \neq \sigma(x_{p_i+h_i-h_{i-1}-1})$ and that whatever $q_i \leq h_{i-1}$, $(\sigma(x_{h_{i-1}+1}), \ldots, \sigma(x_{h_i})) \neq (\sigma(x_{q_i}), \ldots, \sigma(x_{q_i+h_i-h_{i-1}-1}))$. As a conclusion, $E_X(Y)$ is an exhaustive history of Y and then clearly $c(X) = c(Y)$. Now, consider Z of components $z_j = (x_j, y_j)$. We immediately obtain that $Z(h_{i-1} + 1, h_i - 1) = Z(p_i, p_i + h_i - h_{i-1} - 2)$, that $z_{h_i} \neq z_{p_i+h_i-h_{i-1}-1}$ and that whatever $q_i \leq h_{i-1}$, $Z(h_{i-1} + 1, h_i) \neq Z(q_i, q_i + h_i - h_{i-1} - 2)$. As a consequence, $E_X(Z)$ is also an exhaustive history of Z, that finishes the proof. Concerning non reciprocity, consider the example of two binary strings, $X = 010010$ and $Y = 011011$. We obtain $c(X) = c(Y) = c(X, Y)$ but Y cannot be predicted through X with a bijection from $\{0, 1\}$ to itself. $\square$

This property can naturally be extended for more than two sequences. Furthermore, this property is similar to that of Shannon entropy. Indeed, for two random sources X and Y, if and only if $\Pr[Y|_X] = 1$ (*i.e.* Y completely determined by X), then $H(X, Y) = H(X)$. But contrary to Shannon entropy, $c(X) = c(X, Y)$ does not imply that X is completely determined by Y. Because of the non reciprocity of this property for the LZC, some care will have to be taken in the results of the joint LZC analysis, when an analogy is made with Shannon entropy.

To go farther with the analogy between Shannon entropy and LZC, one can

define conditional LZC (CLZC) and informational LZC (ILZC) by

$$\begin{cases} c(Y|X) & = c(X,Y) - c(X) \\ I_c(X;Y) & = c(X) + c(Y) - c(X,Y) = c(X) - c(X|Y) = c(Y) - c(Y|X) \end{cases} \tag{8}$$

and so on for higher order informational LZC (as for the Shannon entropies [23]). Since the joint LZC and the marginal LZC tend to the joint and marginal Shannon entropies for stationary ergodic sequences, this holds for the conditional LZC and for informational LZC, respectively toward conditional entropy and mutual information. Furthermore, property 3 leads to the positivity of the conditional complexity as for the conditional Shannon entropy.

In many classification problems or source separation problems, the Kullback-Leibler divergence, and particularly mutual information, is used as a separation criterion. In a certain sense, mutual information can be understood as a distance between the joint density of two sequences and independent sequences sharing the same marginal densities. Using the LZC, by analogy with the Shannon entropy, I_c can be expected to be understood as a divergence between two sequences. However, the ILZC I_c can be negative, *e.g.* $X = 00100100100$ and $Y = 01010100000$ leads to $I_c(X,Y) = -1$. This non-property of non-negativity is another difference with mutual information in the Shannon sense. Furthermore, the non-reciprocal of property 4 and the non existence of triangular inequality does not permit to build a metric with $\rho = c(X|Y) + c(Y|X)$ as can be done, in a certain sense, with Shannon entropy (see [19]), *e.g.* $X = 00110$, $Y = 01000$ and $Z = 01010$ leads to $4 = \rho(X,Z) > \rho(X,Y) + \rho(Y,Z) = 1 + 1 = 2$.

The LZC tool seems interesting in itself for the analysis of complicated multi-dimensional data (biomedical, economic,...), since it makes a kind of bridge between nonlinear tools for deterministic data analysis and information theory tools. However, this bridge is not complete because of the limited analogies with Shannon entropy. As a conclusion, even if this tool is interesting in itself to characterize or to classify signals, the results obtained need to be interpreted with great caution .

4 Data analysis using the joint LZC

In this section, the use of the joint LZC and the informational LZC (ILZC) for data analysis will be illustrated on two examples. The first example illustrates how the joint LZC can capture a spatiotemporal degree of organization of a random boolean network (RBN), while the second example will show the interest of using ILZC on a proposed multidimensional variant of the minority

game (MG). The RBN as well as the MG produce boolean signals, and can be naturally studied by LZCs.

4.1 Illustration on a random boolean network

An RBN, also called Kauffman network, is given by N binary automata, where each element or cell i is spatially connected to K other cells; K is called the connectivity of the network [24]. Then, the temporal evolution is given by maps $f_i : \{0,1\}^K \to \{0,1\}$ such that $x_i(t+1) = f_i(x_{i_1}(t), \ldots, x_{i_K}(t))$. The connections of the cells are directional, *i.e.* one of the i_l can be equal to j even if $j_m \neq i$ whatever m. The initial state $(x_0(1), \ldots, x_{N-1}(1))$ is randomly chosen in $\{0,1\}^N$ where 1 (resp. 0) is drawn with probability p (resp. $1-p$). The maps f_i are also randomly drawn from all the 2^{2^K} possible functions $\{0,1\}^K \times \{0,1\}$ with the same probability p to draw 1. In other words, the value of $f_i(0, \ldots, 0)$ is drawn in $\{0,1\}$, ..., the value of $f_i(1, \ldots, 1)$ is drawn in $\{0,1\}$, for all the i. This is done to initialize the process, and once built, these functions remain unchanged during the time-evolution of the process. Hence, for a given choice of map, a specific network is considered. The probability p is called the bias of the network [24]. It is shown in [24,25] that, according to the values of (p, K), the behavior of the network can produce order or disorder. Furthermore, it is shown that the network can be controlled by freezing at each time t, $F(t)\gamma N$ cells, where γ is the maximum proportion of cells that can be frozen and where $F(t) \in [0\ ;\ 1]$ is a control function. The cells that can be frozen are randomly chosen, but this choice is time-independent. Here again, at the beginning of the process, the cells that will be potentially frozen are randomly chosen and this choice remains unchanged during the evolution of the process. The $\gamma N F(t)$ cells that are frozen are the first $\gamma N F(t)$ cells that were drawn. Furthermore, the frozen states are randomly chosen in the initialization of the process and remain the same during all the process. As a conclusion, even if there are random choices in the network, all the random variables are drawn in the initialization, thus the spatiotemporal evolution is clearly deterministic.

This RBN has been proposed as a description of discrete genetic network models. In this model, the two possible states for the gene represent if the transcription process is active (*e.g.* state 1) or not (*e.g.* state 0). Hence, the control has been proposed for the modeling of external forcing (biological rhythms, etc.). More information about the RBN can be found in [24,25] and in the references therein.

First of all, in [24,25] it is shown that according to the values of K and p, the network can exhibit an ordered behavior, or a more complex one, akin to a sort of spatiotemporal chaos. More rigorously, since the number of states for the

cells is finite and equals 2^N, the behavior must be periodic with a period less or equal to 2^N. The chaos-like behavior is possible, analyzing the RBN during a time sufficiently small compared to 2^N. In particular, it is shown that when $K < K_c = \frac{1}{2p(1-p)}$ the system exhibits order and when $K > K_c$ it exhibits chaos-like behavior. Here, we have chosen to analyze such a coupled oscillator, in order to use the joint LZC. Thus, we have built an RBN, for different values of p and K, and for different realizations of maps (*i.e.* several specific networks), initial conditions, etc. For each vectorial signal (*i.e.* one realization for the set (p, K)), the joint LZC of the N oscillators is computed and is denoted $c(p, K)$. Figure 1 represents, then, the average LZC $\langle c(p, K)\rangle$ over the realizations. This figure clearly shows that under the critical curve $K = \frac{1}{2p(1-p)}$, the mean joint LZC is low, corresponding to a quite spatiotemporally regular signal, while the mean joint LZC is high over the critical curve. This result exhibits that the joint LZC is able to capture the degree of organization of such a coupled system, at least on average. Furthermore, since the signal cannot be rigorously chaotic, an approach using a Lyapunov exponent seems not rigorous for the analysis of an RBN, even if it is possible by analogy with continuous systems as shown in [24,26]. As is concluded for chaotic signals in [21], the LZC can capture similar information to a Lyapunov exponent, but its evaluation has a lower cost than that of a Lyapunov exponent. Here, we illustrate that this holds for the joint LZC previously presented.

The second illustration concerns a controlled RBN. An RBN was generated, with a free evolution in the beginning, then the RBN is controlled during a finite time and with a periodic control function, and there is again a free evolution. To analyze the spatiotemporal complexity of this system, we have first evaluated at each time a spatial LZC as proposed by Kaspar and Schuster [21] for spatiotemporal analysis. This spatial LZC is compared with the LZC evaluated on a sliding window of N_w samples (sliding step of 1 sample) for the first and the middle cells, *i.e.* at each time t the LZCs $c(x_1(t - N_w + 1) \dots x_1(t))$ and $c(x_{(N+1)/2}(t - N_w + 1) \dots x_{(N+1)/2}(t))$ (*e.g.* for N odd) are computed. Lastly the joint LZC on the sliding window was evaluated, *i.e.* at each time t the joint LZC $c(x(t - N_w + 1) \dots x(t))$ is computed, where $x(t)$ is the N-uplet $(x_1(t), \dots, x_N(t))$). The results are depicted in figure 2. In these figures it can be seen that, for this example, the spatial LZC does not capture the spatiotemporally organized area in the RBN. One can also see that the 1-dimensional LZC is sometimes able to capture the more organized area, when the periodic behavior concerns the chosen cell for the analysis. However, the analyzed cell can also have a periodic behavior during the non-controlled period, that does not reflect the spatiotemporally non-organized area. Lastly, the joint LZC is clearly able to detect the organized period and to differentiate it from the non-organized areas. In particular, the behavior of the joint LZC reflects the controlled function. Since this joint LZC takes simultaneously into account all the cells, during the non-controlled period, its value is large, reflecting the globally non-organized area. Figure 2 also depicts

the averaged spatial LZC and the averaged joint LZC (sliding window) over 500 realizations of the controlled RBN. It can be seen in these figures that on average while the spatial LZC does not capture the controlled area, the joint LZC does. This is due to the fact that for each realization, the frozen cells are not the same and the spatial links vary from one realization to the other. Since the RBN is only spatially analyzed, the temporal periodicities cannot be captured by the spatial LZC. Some regularities can be captured only when the frozen cells present regularities in their spatial choices, but on average there are not regularities. Hence on average, the spatial LZC cannot exhibit low values when the RBN is more organized. In contrast, since the joint LZC takes into account simultaneously all the cells, on average it can clearly distinguish the organized period: in each realization, even if there is not necessarily spatial periodic behavior, the N-uplet has a periodically organized behavior in time. The mean marginal LZCs are not pictured here, but they exhibit the same kind of behavior as the joint LZC. Indeed, since during the control a given cell can exhibit periodic behavior for some realizations and not for others, by averaging over the realizations the non time-periodic LZC (or periodic but not linked to that of the control) are smoothed, resulting in a periodic mean marginal LZC. During the free evolution, the marginal LZC is also smoothed by averaging over the realizations. However, since for the joint LZC the RBN is globally taken into account, during the control the regularity of the RBN is better captured, the effect is more pronounced for the average joint LZC than for each marginal LZC.

4.2 Illustration on the minority game

The minority game consists of N agents where each agent i makes a binary decision $a_i(t) \in \{0,1\}$ at time t. The goal for each agent is to be in the minority, *i.e.* agent i is in the minority if the number of agents taking the same decision $a_i(t)$ is less than $\frac{N}{2}$ [27–29]. Let us denote by $\mu(t)$ the decision of the minority of agents at time t. To make a decision, each agent looks at the m previous minorities $(\mu(t-m),\ldots,\mu(t-1))$ where m is the memory of the game. Each agent possesses a look up table of s functions $f_k : \{0,1\}^m \longrightarrow \{0,1\}$ called strategies. There are 2^{2^m} possible strategies and each agent has the same number s of strategies f_{i_k}, $i = 0,\ldots,N-1$, $k = 0,\ldots,s-1$. The i_k are generally different from one agent to another, but some strategies can possibly be shared by several agents. A gain G_{i_k} is attached to each strategy f_{i_k}. At each time, the gains of all the strategies that would have led to being in the minority are increased by 1, *i.e.* at time t, if $f_{i_k}(\mu(t-m),\ldots,\mu(t-1)) = \mu(t)$, G_{i_k} is increased by 1 point. Hence, at each time, to make its decision, agent i looks at all the strategies at his disposal and chooses the one which has the greatest gain, *i.e.* $i_l = \mathrm{Argmax}_{i_k} G_{i_k}$. The decision of agent i is then $a_i(t) = f_{i_l}(\mu(t-m),\ldots,\mu(t-1))$. If for agent i several strategies shared

the same maximal gain over the G_{i_k}, $k = 0, \ldots, s-1$, the chosen strategy is randomly (uniformly) chosen over these optimal strategies. To initiate the game, one only has to randomly choose the s strategies of each agent, to randomly choose an action $a_i(0)$ for each agent, and an initial memory string $(\mu(-m), \ldots, \mu(-1))$. This game is often used for social, biological, ecological or financial modeling, etc (see [2,27–31] and ref. therein). The game can be easily understood in the case of financial modeling. Say that decision 0 corresponds to buying an asset, and 1 to selling an asset. It is clear that a player prefers to sell an asset while the majority buys this asset, and vice-versa. Several variants have been proposed. For example, a strategy may be rewarded by taking into account the "quality" of the minorities. It was shown that the behavior of such a game can be characterized by the number $L_1(t)$ of agents choosing 1 at each time. In particular the (time-averaged) variance σ_1^2 of L_1, known as volatility in finance, has been shown to be a good characteristic of the behavior of the MG [27,29,30]. It characterizes the total waste of the game. This variance decreases as m increases until a critical size m_c of the memory m, and increases to attain the same variance as that of a game with agents playing completely randomly (symmetric phase for $m < m_c$, and then phase transition and symmetry breaking [30]). However, it was emphasized that this characterization is insensitive to the introduction in the game of agents playing with a random history [2]. In this variant of the game, where a proportion of agents play using the real history, and the other agents play using a random history, the behavior of σ_1^2 remains unchanged as the proportion of "random agents" increases. Rajkovic showed in [2] that the Lempel-Ziv complexity of the minority sequence $c(\mu)$ shows the opposite behavior to volatility, with the same critical size of memory, for MG. But as the proportion of agents playing with a random history increases, $c(\mu)$ exhibits the modified nature of the game. Intuitively, this is due to the fact that if all the agents play with a random history, there is no time-organization in the game.

A multidimensional extension of the game has been proposed in [32] using a three-level structure. It consists of several MG in parallel, but where a connection is made through the payoffs of the strategies, by taking into account the minorities of the whole game. Here, to show how a multidimensional LZC can be used in the MG, we first propose the following simpler multidimensional variant of the game. We consider N players playing with N_a assets. For asset j, agent i makes a decision using the m last minorities of asset j, but also taking into account the m_a last minorities of K other assets. K will be called the connectivity of the multidimensional minority game (MMG). The K connections $k_{i,j,0}, \ldots k_{i,j,K-1} \in \{0, \ldots, N_a - 1\}$ of agent i and variable j are randomly chosen in the initiation of the game and an asset can possibly be connected to itself (in this case there is an index l so that $k_{i,j,l} = j$). Then at any time t, the decision of agent i for the asset j is made using the data $(\mu_j(t-m), \ldots, \mu_j(t-1), \mu_{k_{i,j,0}}(t-m_a), \ldots, \mu_{k_{i,j,0}}(t-1), \ldots, \mu_{k_{i,j,K-1}}(t-1))$. For $K = 0$, it is as if N_a MG were played independently, but for $K \neq 0$,

there are clearly spatiotemporal links. As for the standard MG, for each asset the agents possess s strategies in the functions $\{0,1\}^{m+Km_a} \longrightarrow \{0,1\}$. Each strategy for each asset and of each agent possesses a gain that evolves in time (one point won for a good strategy, 0 otherwise).

In our illustrations, the MMG is composed of $N = 101$ agents, playing on $N_a = 5$ assets, with $s = 2$ strategies for each agent and each asset. An additional memory $m_a = 1$ and $m_a = 2$ for each connected asset is considered, and the time duration of the game is 10000 samples. Figure 3 depicts the normalized volatility $\frac{\sigma_1^2}{N}$, as a function of the memory m, considering only the first asset of the MMG, in the case of an additional connecting memory $m_a = 1$ and $m_a = 2$; Figure 4 describes the normalized LZC of the minority sequence for the first asset $c(\mu_0)\,\frac{\log_2(T)}{T}$; Figure 5 represents the normalized LZC of the minority sequence for the first asset $c(\mu_0)\,\frac{\log_2(T)}{T}$; the normalized joint LZC of all the assets $c(\mu_0,\ldots,\mu_{N_a-1})\,\frac{\log_2(T)}{T}$ is plotted figure 6 for $m_a = 1$ and $m_a = 2$; Lastly, the behavior of the normalized informational LZC of the first two assets $I_c(\mu_0,\mu_1)\,\frac{\log_2(T)}{T}$ and of all the assets $I_c(\mu_0,\ldots,\mu_{N_a-1})\,\frac{\log_2(T)}{T}$ is given in figure 7. In these figures, the solid line with stars represents the result given by a totally random game. For $K = 0$, since the first asset comes from a standard MG, the behavior of $\frac{\sigma_1^2}{N}$ is that described in [27,29]. As shown in [2] the LZC $c(\mu_0)$ shows the opposite behavior, leading to an equivalent characterization of the game. $c(\mu_0,\mu_1)$ has also the same behavior. In this case, m_a does not change the behavior of the curves, since with no connection this additional memory is not taken into account. As K increases, the volatility exhibits a behavior that is increasingly similar to that of a totally random game. Since the volatility is similar for the five games, the curve would have the same shape on average over the assets. Although this average variable still represents the total waste of the game, it is not efficient to capture the self-organization hidden in this MMG. The LZC of the first asset, $c(\mu_0)$, leads to the same conclusion, as well as the joint LZC of the first two assets, $c(\mu_0,\mu_1)$, and the joint LZC of all the assets, $c(\mu_0,\ldots,\mu_{N_a-1})$. Furthermore these characteristics seem insensitive to the additional memory m_a. However, some fluctuations of the LZCs as a function of m can be seen in these curves. Since the joint LZCs and the marginal ones have the same shape, this suggests that the combination of marginal and joint LZC can reveal the self-organized characteristic of the proposed MMG. Figure 7 depicts the two particular combinations represented by the informational LZC: $I_c(\mu_0,\mu_1) = c(\mu_0) + c(\mu_1) - c(\mu_0,\mu_1)$ and $I_c(\mu_0,\ldots,\mu_{N_a-1}) = \sum_{l=1}^{N_a}(-1)^{l+1}\sum_{i_1<\ldots<i_l} c(\mu_{i_1},\ldots,\mu_{i_l})$ (see [23] for the analogy with Shannon entropy). In both cases, it is clear that the behavior of the ILZC is far from that of a random game. Moreover, some differences can clearly be seen for different values of connectivity K. In conclusion, the ILZC can clearly reveal the self-organized behavior of the proposed MMG. Since the purpose of the paper is not to study in detail this MMG, no further analyzis is given here. However, it will be interesting to investigate if some critical

memory sizes in this MMG, taking into account both m and m_a or not, would be exhibited through the ILZC, as is the case for the LZC of a standard MG with the memory m [30].

5 Discussion

The Lempel-Ziv complexity has long been known and has been extensively used in the compression domain. However, this tool is not so extensively employed for data analysis. Recent studies present this tool as a potential tool for both analyzing biomedical sequences [10,16–18] and signals from complex systems such as the minority game [2]. In this paper we show that the LZC can naturally be understood for vectorial sequences, and then used to analyze multidimensional signals. We propose, then, derived quantities such as informational LZC, by analogy with Shannon entropy, and we have shown that these derived quantities can also be used for data analysis. Notice, however, that the use of the LZCs seems interesting for the analysis of signals which are naturally coded on a finite size alphabet. It is the case for random boolean networks or for the different variants of the minority game. For real world signals, one can envisage analyzing DNA sequences that are naturally coded on an alphabet of size 4. But it seems at present more difficult to use this tool for continuous-state signals such as ECG, even if some attempts have already been made. In the latter case, a crucial point is the quantification of the signals: static or dynamic quantization [11]? How many levels of quantization [11]? Very often, empirical rules have been chosen [10,18], but the choice may have consequences on the LZC of the quantized signals. Working on these points seems very challenging. Another challenge would be also to investigate links between signals of different natures, such as electrocardiograms, electromyograms (etc.) via the joint LZC or derived quantities.

References

[1] M. Rajković, Extracting meaningful information from financial data, Physica A 287 (3-4) (2000) 383–395.

[2] M. Rajković, Z. Mihailović, Quantifying complexity in the minority game, Physica A 325 (1-2) (2003) 40–47.

[3] G. W. Botteron, J. M. Smith, A technique for measurment of the extent of spatial organization of atrial activation during atrial fibrillation in the intact human heart, IEEE Transactions on Biomedical Engineering 42 (6) (1995) 548–555.

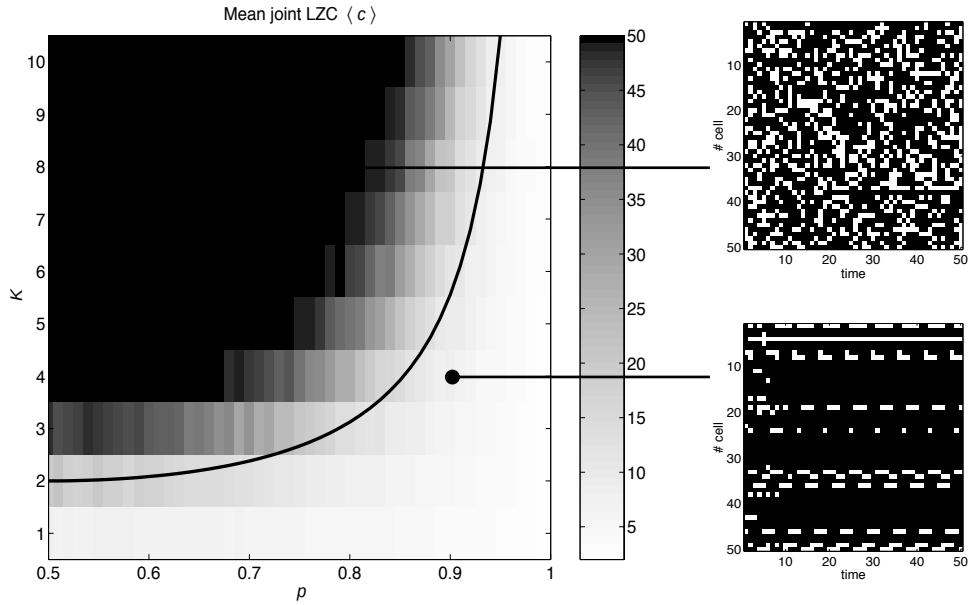

Fig. 1. Mean joint LZC for RBNs, as a function of the bias p and the connectivity K. The RBNs contain $N = 50$ cells, the duration is $T = 50$ and c has been averaged using 100 realizations. The solid line represents the critical line $K_c = \frac{1}{2p(1-p)}$. The two figures on the right give two particular illustrations of an RBN, respectively for $(K, p) = (8, .7)$ and $(4, .9)$. The black (resp. white) points correspond to the value 1 (resp. 0).

[4] C. Cysarz, H. Bettermann, P. V. Leeuwen, Entropies of short binary sequences in heart period dynamics, AJP Heart and Circulatory Physiology 278 (6) (2000) H2163–H2172.

[5] M. I. Owis, A. H. Abou-Zied, A.-B. M. Youssef, Y. M. Kadah, Study of features based on nonlinear dynamical modeling in ECG arrhythmia detection and classification, IEEE Transactions on Biomedical Engineering 49 (7) (2002) 733–736.

[6] R. Q. Quiroga, J. Arnhold, K. Lehnertz, P. Grassberger, Kulback–Leibler and renormalized entropies: Applications to electroencephalograms of epilepsy patients, Physical Review E 62 (6) (2000) 8380–8386.

[7] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, Heart rate variability: Standards of measurement, physiological interpretation, and clinical use, Circulation 93 (3) (1996) 1043–1065.

[8] N. V. Thakor, Y.-S. Zhu, K.-Y. Pan, Ventricular tachycardia and fibrillation detection by a sequantial hypothesis testing algorithm, IEEE Transactions on Biomedical Engineering 37 (9) (1990) 837–843.

[9] Y. Wang, Y.-S. Zhu, N. V. Thakor, Y.-H. Xu, A short-time multifractal approach for arrhythmia detection based on fuzzy neural network, IEEE Transactions on Biomedical Engineering 48 (9) (2001) 989–995.

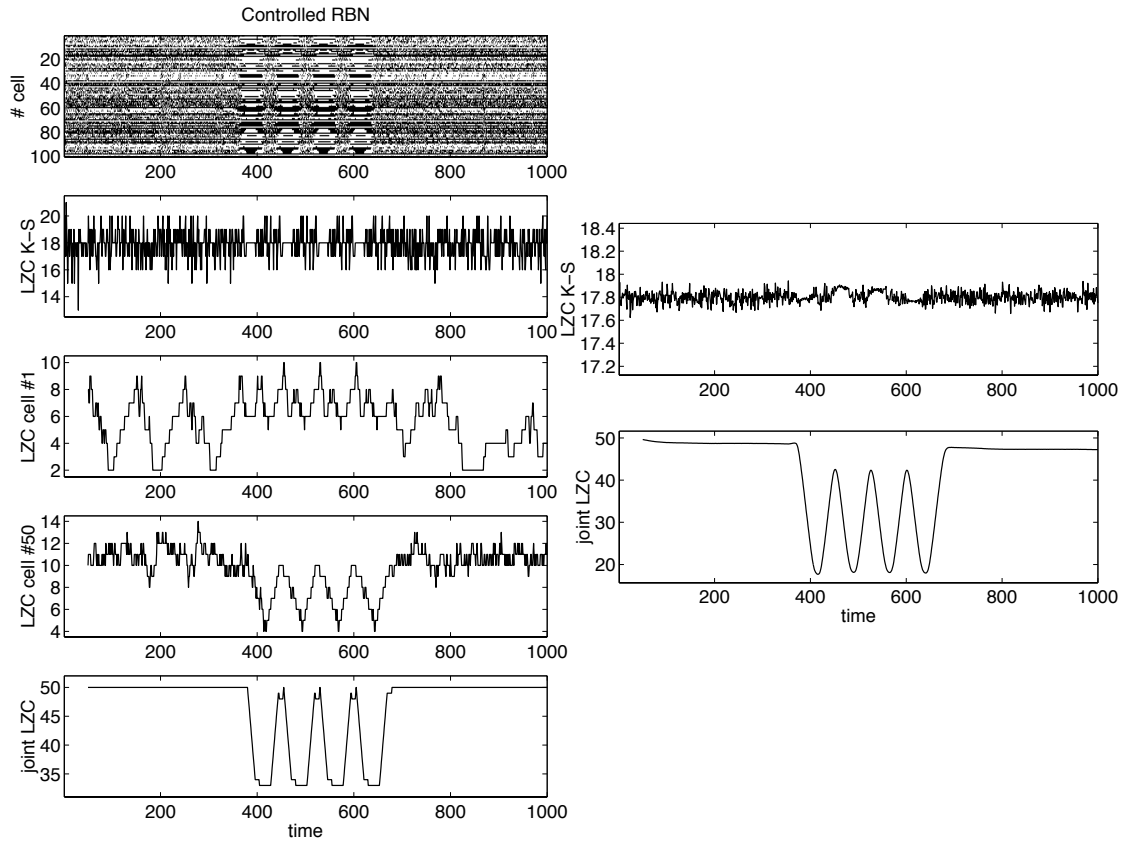

Fig. 2. Top-left figure: controlled RBN; Second-left figure: spatial LZC at each time (as proposed by Kaspar and Schuster); Third and fourth left figures: LZC of the first and 50th cells on a sliding window; Bottom-left figure: joint LZC for all the cells, on a sliding window. Right-figures: respectively spatial LZC and mean joint LZC on a sliding window, averaged over several realizations of the RBN. The parameters are $N = 100$ cells, the connectivity is $K = 3$, the bias is $p = .5$ the maximum proportion of frozen cells is $\gamma = .75$ and the control function is $F(t) = \sin(\pi(t - 350)/75)$ for $t \in [350\,;\, 649]$. The size of the window is $N_w = 50$ samples and the window slide each sample (*i.e.* the joint LZC at each time t is $c(t) = c(x(t - N_w + 1) \dots x(t))$ where $x(t)$ is the N-uplet $(x_1(t), \dots, x_N(t))$). The averaged curves have been obtained on 500 realizations

[10] X.-S. Zhang, Y.-S. Zhu, N. V. Thakor, Z.-Z. Wang, Detecting ventricular tachycardia and fibrillation by complexity measure, IEEE Transactions on Biomedical Engineering 46 (5) (1999) 548–555.

[11] W. Ebeling, L. Molgedey, J. Kurths, U. Schwarz, Entropy, complexity, predictability and data analysis of time series and letter sequences, in: Theory of Disaster, A. Bundle and H.-J. Schellnhuber Edition, Springer Verlag, Berlin, 2000.

[12] S. Pincus, Approximate entropy (ApEn) as a complexity measure, Chaos 5 (1) (1995) 110–117.

[13] M. E. Torres, L. G. Gamero, Relative complexity changes in time series using information measures, Physica A 286 (3/4) (2000) 457–473.

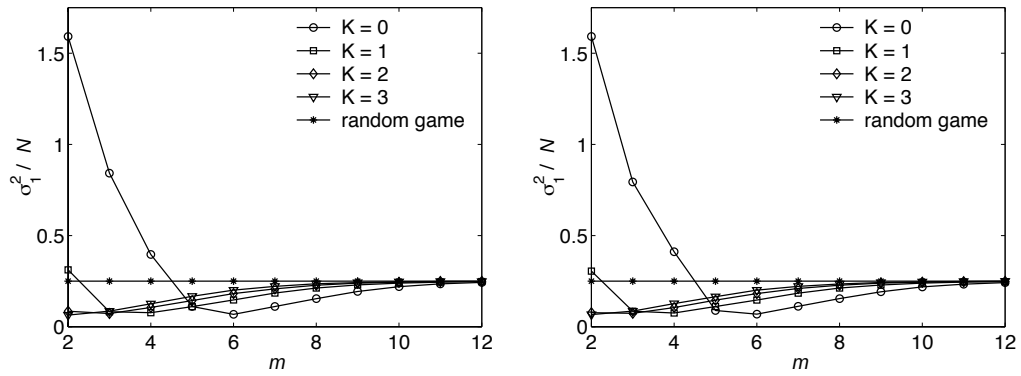

Fig. 3. Normalized volatility for the first asset of the MMG, $\frac{\sigma_1^2}{N}$, as a function of the individual memory m, for the additional memory of the connection $m_a = 1$ (left) and $m_a = 2$ (right), for the connectivities $K = 0, 1, 2, 3$. The parameters are $N = 101$ players, $s = 2$ strategies per player, $N_a = 5$ assets and the duration is 10000 samples. The estimation is made on 50 realizations of the MMG.

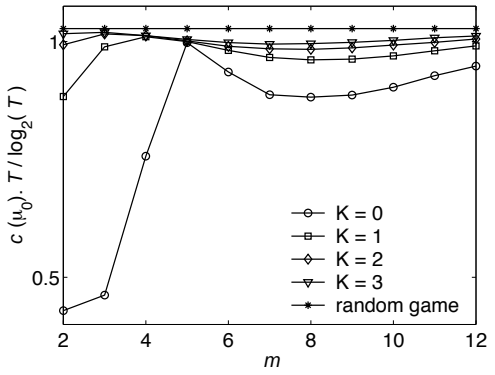

Fig. 4. Normalized LZC of the first asset of the MMG, as a function of m, for $m_a = 1$ and in the same conditions as in figure 3.

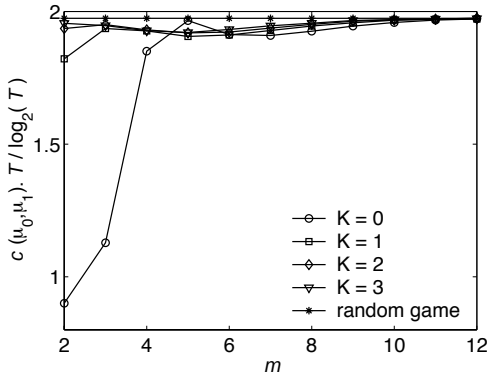

Fig. 5. Normalized joint LZC of the first two assets, as a function of m, for $m_a = 1$ and in the same conditions as in figure 3.

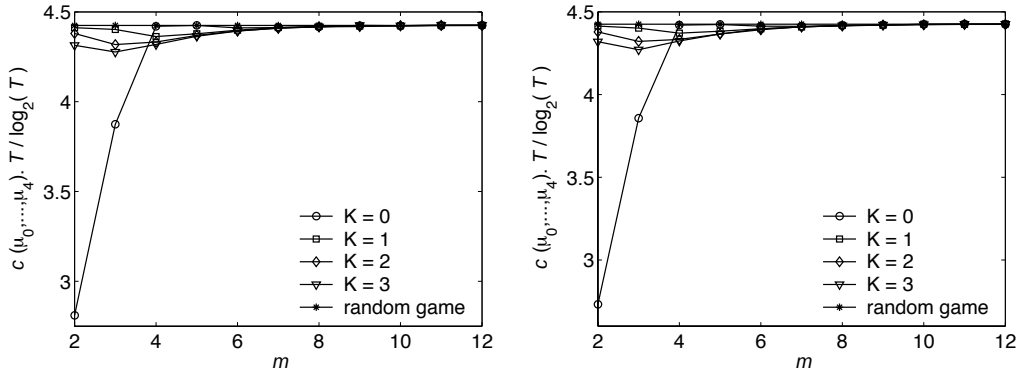

Fig. 6. Normalized joint LZC of all the assets, as a function of m, in the same conditions as in figure 3, for $m_a = 1$ (left) and $m_a = 2$ (right).

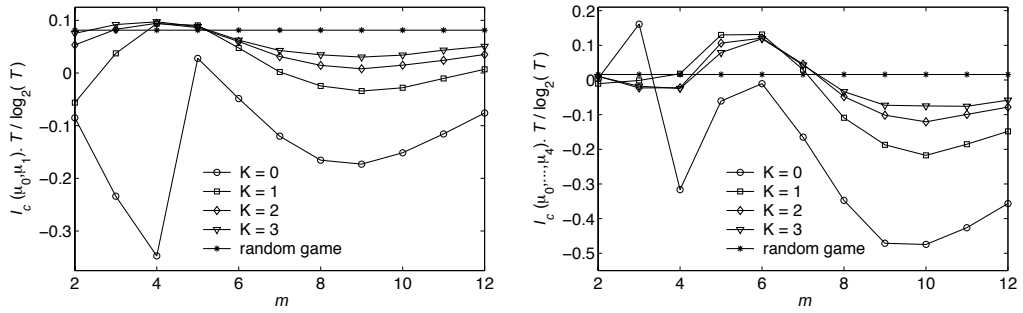

Fig. 7. Normalized informational LZC as a function of m, for the first two assets and $m_a = 1$ (left), and for all the assets and $m_a = 2$ (right). the other parameters are the same as those of figure 3.

[14] T. H. Evrett, J. R. Moorman, L.-C. Kok, J. G. Akar, D. E. Haines, Assessment of global atrial fibrillation organization to optimize timing of atrial defibrillation, Circulation 103 (23) (2001) 2857–2861.

[15] N. Radhakrishnan, B. N. Gangadhar, Estimating regularity in epileptic seizure time-series data – a complexity-measure approach, IEEE Engineering in Medicine and Biology Magazine 17 (3) (1998) 89–94.

[16] N. Radhakrishnan, Quantifying physiological data with Lempel – Ziv complexity – certain issues, IEEE Transactions on Biomedical Engineering 49 (11) (2002) 1371–1373.

[17] J. Szczepański, J. M. Amigó, E. Wajnryb, M. V. Sanchez-Vives, Application of Lempel – Ziv complexity to the analysis of neural discharges, Network: Computation in Neural Systems 14 (2) (2003) 335–350.

[18] X.-S. Zhang, R. J. Roy, E. W. Jensen, EEG complexity as a measure of depth of anesthesia for patients, IEEE Transactions on Biomedical Engineering 48 (12) (2001) 1424–1433.

[19] T. M. Cover, J. A. Thomas, Elements of Information Theory, John Wiley & Sons, New-York, 1991.

[20] A. Lempel, J. Ziv, On the complexity of finite sequences, IEEE Transactions on Information Theory 22 (1) (1976) 75–81.

[21] F. Kaspar, H. G. Schuster, Easily calculable measure for the complexity of spatiotemporal patterns, Physical Review A 36 (2) (1987) 842–848.

[22] J. Ziv, A. Lempel, Compression of individual sequences via variable-rate coding, IEEE Transactions on Information Theory 24 (5) (1978) 530–536.

[23] H. Matsuda, Physical nature of higher-order mutual information: Intrinsic correlations and frustration, Physical Review E 62 (3) (2000) 3096–3102.

[24] F. J. Ballesteros, B. Luque, Random boolean networks response to external periodic signals, Physica A 313 (3-4) (2002) 289–300.

[25] R. V. Solé, B. Luque, Phase transitions and antichaos in generalized kauffman networks, Physics Letters A 196 (5-6) (1995) 331–334.

[26] B. Luque, R. Solé, Lyapunov exponents in random boolean networks, Physica A 284 (1-2) (2000) 33–45.

[27] D. Challet, Y.-C. Zhang, Emergence of cooperation and organization in an evolutionary game, Physica A 246 (3-4) (1997) 407–418.

[28] Y.-C. Zhang, Modeling market mechanism with evolutionary games, Europhysics News 29 (2) (1998) 51–54.

[29] R. Savit, R. Manuca, R. Riolo, Adaptive competition, market efficiency, and phase transition, Physical Review Letters 82 (10) (1999) 2203–2206.

[30] M. Marsili, D. Challet, Continuum time limit and stationary states of the minority game, Physical Review E 64 (5).

[31] D. Challet, M. Marsili, G. Ottino, Shedding light on el farol, Physica A 332 (2004) 469–482.

[32] C. Földy, Z. Somogyvári, P. Érdi, Hierarchically organized minority games, Physica A 323 (2003) 735–742.