COMPLEXITY CHARACTERIZATION OF DYNAMICAL SYSTEMS THROUGH PREDICTABILITY*

Fabio Cecconi, Massimo Falcioni and Angelo Vulpiani

Dipartimento di Fisica Università di Roma "La Sapienza"
INFM, Unità di Roma1 and SMC Center,
P.le Aldo Moro 2, 00185 Roma, Italy

(Received November 27, 2002)

Some aspects of the predictability problem in dynamical systems are reviewed. The deep relation among Lyapunov exponents, Kolmogorov–Sinai entropy, Shannon entropy and algorithmic complexity is discussed. In particular, we emphasize how a characterization of the unpredictability of a system gives a measure of its complexity. A special attention is devoted to finite-resolution effects on predictability, which can be accounted with suitable generalization of the standard indicators. The problems involved in systems with intrinsic randomness is discussed, with emphasis on the important problems of distinguishing chaos from noise and of modeling the system.

PACS numbers: 45.05.+x, 05.45.–a

All the simple systems are simple in the same way, each complex system has its own complexity (freely inspired by *Anna Karenina* by Lev N. Tolstoy)

1. Introduction

The possibility to predict future states of a system stands at the foundations of scientific knowledge with an obvious relevance both from a conceptual and applicative point of view. The perfect knowledge of the evolution law of a system may induce the conclusion that this aim could be attained. This classical deterministic point of view was claimed by Laplace [1]: once the evolution laws of the system are known, the state at a certain time t_0 completely determines the subsequent states for every time $t > t_0$. However it is well established now that in some systems, full predictability cannot

* Presented at the XV Marian Smoluchowski Symposium on Statistical Physics, Zakopane, Poland, September 7–12, 2002.

be accomplished in practice because of the unavoidable uncertainty in the initial conditions. Indeed, as already stated by Poincaré, long-time predictions are reliable only when the evolution law does not amplify the initial uncertainty too rapidly. Therefore, from the point of view of predictability, we need to know how an error on the initial state of the system grows in time. In systems with great sensitive dependence on initial conditions (deterministic chaotic systems) errors grows exponentially fast in time, limiting the ability to predict the future states.

A branch of the theory of dynamical systems has been developed with the aim of formalizing and characterizing the sensitivity to initial conditions. The Lyapunov exponent and the Kolmogorov–Sinai entropy are the two main indicators for measuring the rate of error growth and information production during a deterministic system evolution. A complementary approach has been developed in the context of information theory, data compression and algorithmic complexity theory and it is rather clear that the latter point of view is closely related to the dynamical systems one. If a system is chaotic, then its predictability is limited up to a time which is related to the first Lyapunov exponent, and the time sequence by which we encode one of its chaotic trajectories cannot be compressed by an arbitrary factor, *i.e.* is algorithmically complex. On the contrary, the coding of a regular trajectory can be easily compressed (*e.g.*, for a periodic trajectory it is sufficient to have the sequence for a period) so it is "simple".

In this paper we will discuss how unpredictability and algorithmic complexity are closely related and how information and chaos theory complete each other in giving a general understanding of complexity in dynamical processes. In particular, we shall consider the extension of this approach, nowadays well established in the context of low dimensional systems and for asymptotic regimes, to high dimensional systems with attention to situations far from asymptotic (*i.e.* finite time and finite observational resolution) [2].

2. Two points of view

2.1. Dynamical systems approach: Characteristic Lyapunov exponents

The characteristic Lyapunov exponents are somehow an extension of the linear stability analysis to the case of aperiodic motions. Roughly speaking, they measure the typical rate of exponential divergence of nearby trajectories and, thus, contain information on the growing rate of a very small error on the initial state of a system.

Consider a dynamical system with an evolution law given, *e.g.*, by the differential equation

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{F}(\boldsymbol{x}) \; ; \tag{2.1}$$

we assume that $\boldsymbol{F}$ is smooth enough that the evolution is well-defined for time intervals of arbitrary extension, and that the motion occurs in a bounded region of the phase space. We intend to study the separation between two trajectories, $\boldsymbol{x}(t)$ and $\boldsymbol{x}'(t)$, starting from two close initial conditions, $\boldsymbol{x}(0)$ and $\boldsymbol{x}'(0) = \boldsymbol{x}(0) + \delta\boldsymbol{x}(0)$, respectively.

As long as the difference between the trajectories, $\delta\boldsymbol{x}(t) = \boldsymbol{x}'(t) - \boldsymbol{x}(t)$, remains small (infinitesimal, strictly speaking), it can be regarded as a vector, $\boldsymbol{z}(t)$, in the tangent space. The time evolution of $\boldsymbol{z}(t)$ is given by the linearized differential equations:

$$\frac{dz_i(t)}{dt} = \sum_{j=1}^{d} \left.\frac{\partial F_i}{\partial x_j}\right|_{\boldsymbol{x}(t)} z_j(t) \; . \tag{2.2}$$

Under rather general hypothesis, Oseledec [3] proved that for almost all initial conditions $\boldsymbol{x}(0)$, there exists an orthonormal basis $\{\boldsymbol{e}_i\}$ in the tangent space such that, for large times,

$$\boldsymbol{z}(t) = \sum_{i=1}^{d} c_i \boldsymbol{e}_i \mathrm{e}^{\lambda_i \, t} \; , \tag{2.3}$$

where the coefficients $\{c_i\}$ depend on $\boldsymbol{z}(0)$. The exponents $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ are called *characteristic Lyapunov exponents* (LEs). If the dynamical system has an ergodic invariant measure, the spectrum of LEs $\{\lambda_i\}$ does not depend on the initial condition, except for a set of measure zero with respect to the natural invariant measure.

Equation (2.3) describes how a d-dimensional spherical region of the phase space, with radius ϵ centered in $\boldsymbol{x}(0)$, deforms, with time, into an ellipsoid of semi-axes $\epsilon_i(t) = \epsilon \exp(\lambda_i t)$, directed along the $\boldsymbol{e}_i$ vectors. Furthermore, for a generic small perturbation $\delta\boldsymbol{x}(0)$, the distance between the reference and the perturbed trajectory behaves as

$$|\delta\boldsymbol{x}(t)| \sim |\delta\boldsymbol{x}(0)| \, \mathrm{e}^{\lambda_1 \, t} \, [1 + O\left(\exp -(\lambda_1 - \lambda_2)t\right)] \, .$$

If $\lambda_1 > 0$ we have a rapid (exponential) amplification of an error on the initial condition. In such a case, the system is chaotic and, *de facto*, unpredictable on the long times. Indeed, if the initial error amounts to $\delta_0 = |\delta\boldsymbol{x}(0)|$, and we purpose to predict the states of the system with a certain tolerance Δ (not too large), then the prediction is reliable just up to a *predictability time* given by

$$T_{\mathrm{p}} \sim \frac{1}{\lambda_1} \ln\left(\frac{\Delta}{\delta_0}\right) \; . \tag{2.4}$$

This equation shows that $T_{\rm p}$ is basically determined by the largest Lyapunov exponent, since its dependence on δ_0 and Δ is logarithmically weak. Because of its preeminent role, λ_1 is often referred as "the Lyapunov exponent", and denoted by λ.

2.2. Information based approach

In experimental investigations of physical processes, the access to a system occurs only through a measuring device which produces a time record of a certain observable, *i.e.* a sequence of data. In this regard a system, whether or not chaotic, generates messages and may be regarded as a source of information whose properties can be analyzed through the tools of information theory.

The characterization of the information contained in a sequence can be approached in two very different frameworks. The first considers a specific message (sequence) as belonging to the ensemble of all the messages that can be emitted by a source, and defines an average information content by means of the average compressibility properties of the ensemble [4]. The second considers the problem of characterizing the universal compressibility (*i.e.* ensemble independent) of a specific sequence and concerns the theory of algorithmic complexity and algorithmic information theory [5,6]. For the sake of self-consistency we briefly recall the concepts and ideas about the Shannon entropy [4], that is the basis of whole information theory.

2.2.1. Shannon entropy

Consider a source that can output m different symbols; denote with s_t the symbol emitted by the source at time t and with $P(C_N)$ the probability that a given word $C_N = (s_1, s_2, \ldots, s_N)$, of length N, is emitted $P(C_N) = P(s_1, s_2, \ldots, s_N)$. We assume that the source is stationary, so that, for the sequences $\{s_t\}$, the time translation invariance holds: $P(s_1, \ldots, s_N) = P(s_{t+1}, \ldots, s_{t+N})$. We introduce the N-block entropies

$$H_N = -\sum_{\{C_N\}} P(C_N) \ln P(C_N)\,, \tag{2.5}$$

for stationary sources the limit

$$\lim_{N\to\infty} \frac{H_N}{N} = h_{\rm Sh} \tag{2.6}$$

exists and defines the Shannon entropy $h_{\rm Sh}$ which quantifies the richness (or "complexity") of the source emitting the sequence. This can be precisely expressed by the first theorem of Shannon–McMillan [7] that applies to stationary ergodic sources: The ensemble of N-long subsequences, when N is

large enough, can be partitioned in two classes, $\Omega_1(N)$ and $\Omega_0(N)$ such that all the words $C_N \in \Omega_1(N)$ have the same probability $P(C_N) \sim \exp(-N h_{\rm Sh})$ and

$$\sum_{C_N \in \Omega_1(N)} P(C_N) \to 1\,, \qquad \text{while} \sum_{C_N \in \Omega_0(N)} P(C_N) \to 0\,, \qquad \text{for } N \to \infty\,. \tag{2.7}$$

The meaning of this theorem is the following. An m-states process admits, in principle, m^N possible sequences of length N. However the number of typical sequences, $N_{\rm eff}(N)$, effectively observable (*i.e.* those belonging to $\Omega_1(N)$) is

$$N_{\rm eff}(N) \sim \exp(N h_{\rm Sh})\,. \tag{2.8}$$

Note that $N_{\rm eff} \ll m^N$ if $h_{\rm Sh} < \ln m$. The entropy per symbol, $h_{\rm Sh}$, is a property of the source. However, because of the ergodicity $h_{\rm Sh}$ can be obtained by analyzing just one single sequence in the ensemble of the typical ones, and it can also be viewed as a property of each typical sequence.

In information theory, expression (2.8) is somehow the equivalent of the Boltzmann equation in statistical thermodynamics: $S \propto \ln W$, being W the number of possible microscopic configurations and S the thermodynamic entropy, this justifies the name "entropy" for $h_{\rm Sh}$.

The relevance of the Shannon entropy in information theory is given by the fact that $h_{\rm Sh}$ sets the maximum compression rate of a sequence $\{s_1, s_2, s_3, \ldots\}$. Indeed a theorem of Shannon states that, if the length T of a sequence is large enough, there exists no other sequence (always using m symbols), from which it is possible to reconstruct the original one, whose length is smaller than $(h_{\rm Sh}/\ln m)T$ [4]. In other words, $h_{\rm Sh}/\ln m$ represents the maximum allowed compression rate. The relation between Shannon entropy and data compression problems is well illustrated by considering the optimal coding (Shannon–Fano) to map $\mathcal{N}$ objects (*e.g.* the N-words C_N) into sequences of binary digits $(0, 1)$ [8]. Denoting with ℓ_N the binary length of the sequence specifying C_N, we have

$$\lim_{N\to\infty} \frac{\langle \ell_N \rangle}{N} = \frac{h_{\rm Sh}}{\ln 2}\,, \tag{2.9}$$

i.e., in a good coding, the mean length of a N-word is equal to N times the Shannon entropy, apart from a multiplicative factor, since in the definition (2.6) of $h_{\rm Sh}$ we used the natural logarithm and here we want to work with a two symbol code.

2.2.2. The Kolmogorov–Sinai entropy

After the introduction of the Shannon entropy we can easily define the Kolmogorov–Sinai entropy which is the analogous measure of complexity

applied to dynamical systems. Consider a trajectory, $\boldsymbol{x}(t)$, generated by a deterministic system, sampled at the times $t_j = j\,\tau$, with $j = 1, 2, 3, \ldots$. Perform a finite partition $\mathcal{A}$ of the phase space with the finite number of symbols $\{s\}_{\mathcal{A}}$ enumerating the cells of the partition. The time-discretized trajectory $\boldsymbol{x}(t_j)$ determines a sequence $\{s(1), s(2), s(3), \ldots\}$, whose meaning is clear: at the time t_j the trajectory is in the cell labeled by $s(j)$. To each subsequence of length $N \cdot \tau$ one can associate a word of length N: $W_j^N(\mathcal{A}) = (s(j), s(j+1), \ldots, s(j+(N-1)))$. If the system is ergodic, as we suppose, from the frequencies of the words one obtains the probabilities by which the block entropies $H_N(\mathcal{A})$ are calculated:

$$H_N(\mathcal{A}) = -\sum_{\{W^N(\mathcal{A})\}} P(W^N(\mathcal{A})) \ln P(W^N(\mathcal{A})). \tag{2.10}$$

The probabilities $P(W^N(\mathcal{A}))$, computed by the frequencies of $W^N(\mathcal{A})$ along a trajectory, are essentially dependent on the stationary measure selected by the trajectory. The entropy per unit time of the trajectory with respect to the partition $\mathcal{A}$, $h(\mathcal{A})$, is defined as follows:

$$h_N(\mathcal{A}) = \frac{1}{\tau} \lim_{N\to\infty} \frac{1}{N} H_N(\mathcal{A})\,. \tag{2.11}$$

Notice that, for the deterministic systems we are considering, the entropy per unit time does not depend on the sampling time τ [9]. The KS-entropy (h_{KS}), by definition, is the supremum of $h(\mathcal{A})$ over all possible finite partitions [9,10]

$$h_{\mathrm{KS}} = \sup_{\mathcal{A}} h(\mathcal{A})\,. \tag{2.12}$$

The extremal character of h_{KS} makes every computation based on the definition (2.12), impossible in the majority of practical cases. In this respect, a useful tool would be the Kolmogorov–Sinai theorem, through which one is granted that $h_{\mathrm{KS}} = h(\mathcal{G})$ if $\mathcal{G}$ is a generating partition. A partition is said to be generating if every infinite sequence $\{s_n\}_{n=1,\ldots,\infty}$ corresponds to a single initial point. However the difficulty now is that, with the exception of very simple cases, we do not know how to construct a generating partition. We only know that, according to the Krieger theorem [11], there exists a generating partition with k elements such that $\mathrm{e}^{h_{\mathrm{KS}}} < k \leq \mathrm{e}^{h_{\mathrm{KS}}} + 1$. Then, a more tractable way to define h_{KS} is based upon considering the partition $\mathcal{A}_\epsilon$ made up by a grid of cubic cells of edge ϵ, from which one has

$$h_{\mathrm{KS}} = \lim_{\epsilon\to 0} h(\mathcal{A}_\epsilon)\,. \tag{2.13}$$

We expect that $h(\mathcal{A}_\epsilon)$ becomes independent of ϵ when $\mathcal{A}_\epsilon$ is so fine to be "contained" in a generating partition.

For discrete time maps what has been exposed above is still valid, with $\tau = 1$ (however, Krieger's theorem only applies to invertible maps).

The important point to note is that, for a truly stochastic (*i.e.* non-deterministic) system, with continuous states, $h(\mathcal{A}_\epsilon)$ is not bounded and $h_{\mathrm{KS}} = \infty$.

2.2.3. Algorithmic complexity

The Shannon entropy establishes a limit on how efficiently the ensemble of messages emitted by a source can be coded. However, we may wonder about the compressibility properties of a single sequence with no reference to its belonging to an ensemble. That is to say, we are looking for an universal characterization of its compressibility or, it is the same, an universal definition of its information content. This problem can be addressed through the notion of *algorithmic complexity*, that concerns the difficulty in reproducing a given string of symbols.

Everybody agrees that the binary digits sequence

$$01110100010110010110 \ldots \tag{2.14}$$

is, in some sense, more random than

$$10101010101010101010 \ldots \tag{2.15}$$

The notion of algorithmic complexity, independently introduced by Kolmogorov [5], Chaitin [12] and Solomonov [6], is a way to formalize the intuitive idea of randomness of a sequence.

Consider, for instance, a binary digit sequence (this does not constitute a limitation) of length N, $q_N = (i_1, i_2, \ldots, i_N)$, generated by a certain computer code on a given machine $\mathcal{M}$. The algorithmic complexity (or algorithmic information content) $K_{\mathcal{M}}(N)$ of q_N is the bit-length of the shortest computer program able to give q_N and to stop afterward. Of course, such a length depends not only on the sequence but also on the machine. However, Kolmogorov [5] proved the existence of a universal computer, $\mathcal{U}$, able to perform the same computation that a program p makes on $\mathcal{M}$, with a modification of p that depends only on $\mathcal{M}$. This implies that for all finite strings:

$$K_{\mathcal{U}}(N) \leq K_{\mathcal{M}}(N) + C_{\mathcal{M}}\,, \tag{2.16}$$

where $K_{\mathcal{U}}(N)$ is the complexity with respect to the universal computer and $C_{\mathcal{M}}$ depends only on the machine $\mathcal{M}$. We can consider the algorithmic complexity with respect to a universal computer dropping the $\mathcal{M}$-dependence in the symbol for the algorithmic complexity, $K(N)$. The reason is that we

are interested in the limit of very long sequences, $N \to \infty$, for which one defines the algorithmic complexity per unit symbol:

$$\mathcal{C} = \lim_{N\to\infty} \frac{K(N)}{N}, \tag{2.17}$$

that, because of (2.16), is an intrinsic quantity, *i.e.* independent of the machine.

Now coming back to the N-sequences (2.14) and (2.15), it is obvious that the latter can be obtained with a minimal program of length $O(\ln N)$ and therefore when taking the limit $N \to \infty$ in (2.17), one obtains $\mathcal{C} = 0$. Of course $K(N)$ cannot exceed N, since the sequence can always be generated by a trivial program (of bit length N)

$$\text{“PRINT } i_1, i_2, \ldots, i_N\text{”} . \tag{2.18}$$

Therefore, in the case of a very irregular sequence, *e.g.*, (2.14), one expects $K(N) \propto N$ (*i.e.* $\mathcal{C} \neq 0$), and the sequence is named complex (*i.e.* of non zero algorithmic complexity) or random.

Algorithmic complexity cannot be computed, and the un-computability of $K(N)$ may be understood in terms of Gödel's incompleteness theorem [12]. Beyond the problem of whether or not $K(N)$ is computable in a specific case, the concept of algorithmic complexity brings an important improvement to clarify the vague and intuitive notion of randomness.

Between the Shannon entropy, h_{Sh}, and the algorithmic complexity, there exists the straightforward relationship

$$\lim_{N\to\infty} \frac{\langle K(N)\rangle}{H_N} = \frac{1}{\ln 2}, \tag{2.19}$$

where $\langle K(N)\rangle = \sum_{C_N} P(C_N) K_{C_N}(N)$, being $K_{C_N}(N)$ the algorithmic complexity of the N-words, in the ensemble of sequences, C_N, with a given distribution of probabilities, $P(C_N)$. Therefore the expected complexity $\langle K(N)/N\rangle$ is asymptotically equal to the Shannon entropy (modulo the $\ln 2$ factor). It is important to stress again that, apart from the numerical coincidence of the values of $\mathcal{C}$ and $h_{\mathrm{Sh}}/\ln 2$, there is a conceptual difference between the information theory and the algorithmic complexity theory. The Shannon entropy essentially refers to the information content in a statistical sense, *i.e.* it refers to an ensemble of sequences generated by a certain source. The algorithmic complexity defines the information content of an individual sequence [13].

The notion of algorithmic complexity can be also applied to the trajectories of a dynamical system. This requires the introduction of finite open coverings of the phase space, the corresponding encoding of trajectories into

symbolic sequences, and the searching of the supremum of the algorithmic complexity per symbol at varying the coverings [14]. Brudno's and White's theorems [15, 16] state that the complexity $\mathcal{C}(\boldsymbol{x})$ for a trajectory starting from the point $\boldsymbol{x}$, is

$$\mathcal{C}(\boldsymbol{x}) = \frac{h_{\mathrm{KS}}}{\ln 2}\,, \tag{2.20}$$

for almost all $\boldsymbol{x}$ with respect to the natural invariant measure. The factor $\ln 2$ stems again from the conversion between natural logarithms and bits.

This result indicates that the KS-entropy quantifies not only the richness of a dynamical system but also the difficulty of describing its typical sequences.

2.3. Algorithmic complexity and Lyapunov exponent

Let us consider a $1d$ chaotic map

$$x(t+1) = f(x(t))\,. \tag{2.21}$$

The transmission of the sequence $\{x(t),\ t = 1, 2, \ldots, T\}$, accepting only errors smaller than a tolerance Δ, is carried out by using the following strategy [18]:

1. Transmit the rule (2.21): for this task one has to use a number of bits independent of the sequence length T.

2. Specify the initial condition $x(0)$ with a precision δ_0 using a finite number of bits which is independent of T.

3. Let the system evolve till the first time τ_1 such that the distance between two trajectories, that was initially $\delta x(0) = \delta_0$, equals Δ and then specify again the new initial condition $x(\tau_1)$ with precision δ_0.

4. Let the system evolve and repeat the procedure (2–3), *i.e.* each time the error acceptance tolerance is reached specify the initial conditions, $x(\tau_1 + \tau_2)$, $x(\tau_1 + \tau_2 + \tau_3)$ ..., with precision δ_0. The times $\tau_1, \tau_2, \ldots$ are defined as follows: putting $x'(\tau_1) = x(\tau_1) + \delta_0$, τ_2 is given by the minimum time such that $|x'(\tau_1 + \tau_2) - x(\tau_1 + \tau_2)| \geq \Delta$ and so on.

Following the steps 1–4, the receiver can reconstruct, with a precision Δ, the sequence $\{x(t)\}$, by simply iterating on a computer the evolution law (2.21) between 1 and $\tau_1 - 1$, τ_1 and $\tau_1 + \tau_2 - 1$, and so on. The amount of bits necessary to implement the above transmission (1–4) can be easily computed. For simplicity of notation we introduce the quantities

$$\gamma_i = \frac{1}{\tau_i} \ln \frac{\Delta}{\delta_0} \tag{2.22}$$

which can be regarded as a sort of *effective* Lyapunov exponents [19, 20]. The LE λ can be written in terms of $\{\gamma_i\}$ as follows

$$\lambda = \langle \gamma_i \rangle = \frac{\sum_i \tau_i \gamma_i}{\sum_i \tau_i} = \frac{1}{\overline{\tau}} \ln \frac{\Delta}{\delta_0}\,, \tag{2.23}$$

where

$$\overline{\tau} = \frac{1}{N} \sum \tau_i\,,$$

is the average time after which we have to transmit the new initial condition. Note that to obtain λ from the γ_i's requires the average (2.23), because the transmission time, τ_i, is not constant. If T is large enough the number of transmissions, N, is $T/\overline{\tau} \simeq \lambda T/\ln(\Delta/\delta_0)$. Therefore, noting that in each transmission, a reduction of the error from Δ to δ_0 requires the employ of $\ln_2(\Delta/\delta_0)$ bits, the total amount of bits used in the transmission is

$$\frac{T}{\overline{\tau}} \ln_2 \frac{\Delta}{\delta_0} = \frac{\lambda}{\ln 2} T\,. \tag{2.24}$$

In other words the number of bits for unit time is proportional to λ.

In more than one dimension, we have simply to replace λ with $h_{\rm KS}$ in (2.24), because the above transmission procedure has to be repeated for each of the expanding directions.

3. Limitation of the Lyapunov exponent and Kolmogorov–Sinai entropy

Lyapunov exponents and KS-entropy are properly defined only in specific asymptotic limits: very long times and arbitrary accuracy. However, predictability problem in realistic situations entails considering finite time intervals and limited accuracy. The first obvious way for quantifying the predictability of a physical system is in terms of the *predictability time* $T_{\rm p}$, *i.e.* the time interval on which one can typically forecast the system. A simple argument suggests

$$T_{\rm p} \sim \frac{1}{\lambda} \ln \left(\frac{\Delta}{\delta_0} \right)\,. \tag{3.1}$$

However, the above relation is too naive to be of practical relevance, in any realistic system. Indeed, it does not take into account some basic features of dynamical systems. The Lyapunov exponent is a global quantity, because it measures the average rate of divergence of nearby trajectories. In general there exist finite-time fluctuations and their probability distribution functions (pdf) is important for the characterization of predictability. The

generalized Lyapunov exponents have been introduced with the purpose to take into account such fluctuations [19, 20]. Moreover, the Lyapunov exponent is defined for the linearized dynamics, *i.e.*, by computing the rate of separation of two infinitesimally close trajectories. On the other hand, in measuring the predictability time (3.1) one is interested in a finite tolerance Δ, because the initial error δ_0 is finite. A recent generalization of the Lyapunov exponent to *finite size* errors extends the study of the perturbation growth to the nonlinear regime, *i.e.* both δ_0 and Δ are not infinitesimal [21].

3.1. Growth of non infinitesimal perturbations

We discuss now an example where the Lyapunov exponent is of little relevance for characterizing the predictability. This problem can be illustrated by considering the following coupled map model:

$$\begin{cases} \boldsymbol{x}(t+1) & = \boldsymbol{R}\,\boldsymbol{x}(t) + \varepsilon \boldsymbol{h}(y(t))\,, \\ y(t+1) & = G(y(t))\,, \end{cases} \tag{3.2}$$

where $\boldsymbol{x} \in \mathbb{R}^2$, $y \in \mathbb{R}^1$, $\boldsymbol{R}$ is a rotation matrix of arbitrary angle θ, $\boldsymbol{h}$ is a vector function and G is a chaotic map. For simplicity we consider a linear coupling $\boldsymbol{h}(y) = (y, y)$ and the logistic map $G(y) = 4y(1-y)$.

For $\varepsilon = 0$ we have two independent systems: a regular and a chaotic one. Thus the Lyapunov exponent of the $\boldsymbol{x}$ subsystem is $\lambda_x(\varepsilon = 0) = 0$, *i.e.*, it is completely predictable. On the contrary, the y subsystem is chaotic with $\lambda_y = \lambda_1 = \ln 2$. The switching on of a small coupling ($\varepsilon > 0$) yields a single three-dimensional chaotic system with a positive global Lyapunov exponent

$$\lambda = \lambda_y + O(\varepsilon)\,. \tag{3.3}$$

A direct application of (3.1) would give

$$T_{\mathrm{p}}^{(x)} \sim T_{\mathrm{p}} \sim \frac{1}{\lambda_y}\,, \tag{3.4}$$

but this result is clearly unacceptable: the predictability time for $\boldsymbol{x}$ seems to be independent of the value of the coupling ε. This is not due to an artifact of the chosen example, indeed, the same argument applies to many physical situations [22]. A well known example is the gravitational three body problem, with one body (asteroid) much smaller than the other two (planets). When the gravitational feedback of the asteroid on the two planets is neglected (restricted problem), one has a chaotic asteroid in the regular field of the planets. As soon as the feedback is taken into account (*i.e.* $\varepsilon > 0$ in the example) one has a non-separable three body system with a positive LE. Of course, intuition correctly suggests that, in the limit of small asteroid

mass ($\varepsilon \to 0$), a forecast of the planet motion should be possible even for very long times. The apparent paradox arises from the misuse of formula (3.1), strictly valid for tangent vectors, to the case of non infinitesimal regimes. As soon as the errors become large, the full nonlinear evolution of the three body system has to be taken into account. This situation is clearly illustrated by the model (3.2) in figure 1. The evolution of $\delta\boldsymbol{x}$ is given by

$$\delta\boldsymbol{x}(t+1) = \boldsymbol{R}\delta\boldsymbol{x}(t) + \varepsilon\delta\boldsymbol{h}(y)\,, \tag{3.5}$$

where, with our choice, $\delta\boldsymbol{h} = (\delta y, \delta y)$. At the beginning, both $|\delta\boldsymbol{x}|$ and δy grow exponentially. However, the available phase space for y is finite and the uncertainty reaches the saturation value $\delta y \sim O(1)$ in a time $t \sim 1/\lambda_1$. At larger times the two realizations of the y variable are completely uncorrelated and their difference δy in (3.5) acts as a noisy term. As a consequence, the growth of the uncertainty on $\boldsymbol{x}$ becomes diffusive with a diffusion coefficient proportional to ε^2 [22]

$$|\delta\boldsymbol{x}(t)| \sim \varepsilon t^{1/2} \tag{3.6}$$

so that:

$$T_{\mathrm{p}}^{(x)} \sim \varepsilon^{-2}\,. \tag{3.7}$$

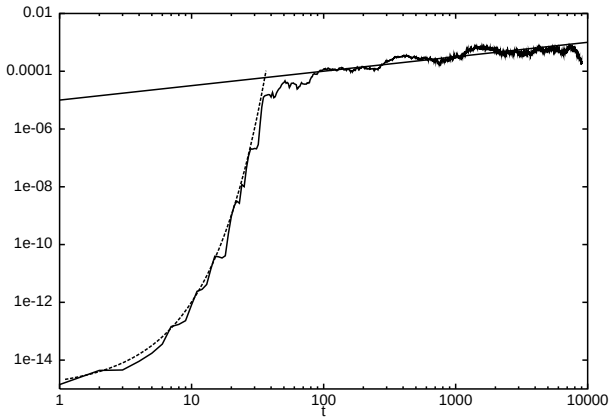

Fig. 1. Growth of error $|\delta\boldsymbol{x}(t)|$ for the coupled map (3.2). The rotation angle is $\theta = 0.82099$, the coupling strength $\varepsilon = 10^{-5}$ and the initial error only on the y variable is $\delta y = \delta_0 = 10^{-10}$. Dashed line $|\delta\boldsymbol{x}(t)| \sim \mathrm{e}^{\lambda_1 t}$ where $\lambda_1 = \ln 2$, solid line $|\delta\boldsymbol{x}(t)| \sim t^{1/2}$.

This example shows that, even in simple systems, the Lyapunov exponent can be of little relevance for the characterization of the predictability.

In more complex systems, in which different scales are present, one is typically interested in forecasting the large scale motion, while the LE is

related to the small scale dynamics. A familiar example of that is weather forecast: despite the LE of the atmosphere is indeed rather large, due to the small scale convective motion, large-scale weather predictions are possible for about 10 days [23,24]. It is thus natural to seek for a generalization of the LE to finite perturbations from which one can obtain a more realistic estimation for the predictability time. It is worth underlining the important fact that finite errors are not confined in the tangent space but are governed by the complete nonlinear dynamics. In this sense the extension of the LE to finite errors will give more information on the system.

Aiming to generalize the LE to non infinitesimal perturbations let us now define the Finite Size Lyapunov Exponent (FSLE) [21]. Consider a reference $\boldsymbol{x}(t)$ and a perturbed trajectory $\boldsymbol{x}'(t)$, such that $|\boldsymbol{x}'(0) - \boldsymbol{x}(0)| \sim \delta$. One integrates the two trajectories and computes the time $\tau_1(\delta, r)$ necessary for the separation $|\boldsymbol{x}'(t) - \boldsymbol{x}(t)|$ to grow from δ to $r\delta$. At time $t = \tau_1(\delta, r)$ the distance between the trajectories is rescaled to δ and the procedure is repeated in order to compute $\tau_2(\delta, r), \tau_3(\delta, r) \ldots$.

The threshold ratio r must be $r > 1$, but not too large in order to avoid contributions from different scales in $\tau(\delta, r)$. A typical choice is $r = 2$ (for which $\tau(\delta, r)$ is properly a "doubling" time) or $r = \sqrt{2}$. In the same spirit of the discussion leading to Eqs (2.22) and (2.23), we may introduce an effective finite size growth rate:

$$\gamma_i(\delta, r) = \frac{1}{\tau_i(\delta, r)} \ln r \,. \tag{3.8}$$

After having performed $\mathcal{N}$ error-doubling experiments, we can define the FSLE as

$$\lambda(\delta) = \langle \gamma(\delta, r) \rangle_t = \left\langle \frac{1}{\tau(\delta, r)} \right\rangle_t \ln r = \frac{1}{\langle \tau(\delta, r) \rangle_e} \ln r \,, \tag{3.9}$$

where $\langle \tau(\delta, r) \rangle_e$ is

$$\langle \tau(\delta, r) \rangle_e = \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \tau_n(\delta, r) \,, \tag{3.10}$$

see [25] for details. In the infinitesimal limit, the FSLE reduces to the standard Lyapunov exponent

$$\lim_{\delta \to 0} \lambda(\delta) = \lambda_1 \,. \tag{3.11}$$

In practice this limit means that $\lambda(\delta)$ displays a constant plateau at λ_1 for sufficiently small δ (Fig. 2). For finite value of δ the behavior of $\lambda(\delta)$ depends on the details of the non linear dynamics. For example, in the model (3.2)

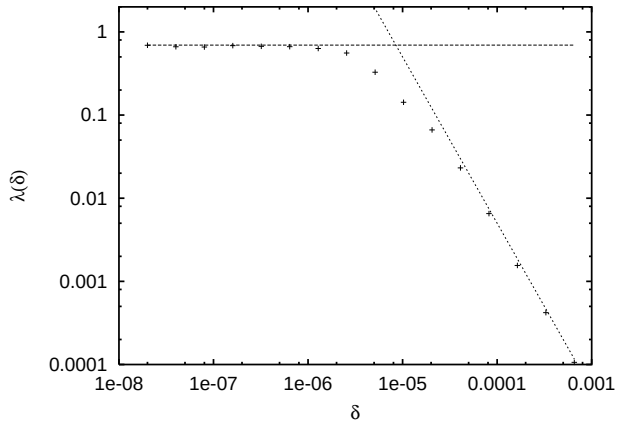

Fig. 2. $\lambda(\delta)$ as a function of δ for the coupled map (3.2) with $\varepsilon = 10^{-5}$. The perturbation has been initialized as in Fig. 1. For $\delta \to 0$, $\lambda(\delta) \simeq \lambda_1$ (horizontal line). The dashed line shows the behavior $\lambda(\delta) \sim \delta^{-2}$.

the diffusive behavior (3.6), by simple dimensional arguments, corresponds to $\lambda(\delta) \sim \delta^{-2}$. Since the FSLE measures the rate of divergence of trajectories at finite errors, one might wonder whether it is just another way to look at the average response $\langle \ln(|\boldsymbol{x}'(t) - \boldsymbol{x}(t)|)\rangle$ as a function of time. The answer is negative, because taking the average at fixed time is not the same as computing the average doubling time at *fixed scale*, as in (3.9). This is particularly clear in the case of strongly intermittent system, in which $|\delta\boldsymbol{x}(t)|$ can be very different in each realization. In the presence of intermittency, averaging over different realizations at fixed times can produce a spurious regime due to the superposition of exponential and diffusive contributions by different samples at the same time [25]. The FSLE method can be easily applied to data analysis [26]. For other approaches addressing the problem of non-infinitesimal perturbations see [27,28].

3.2. The ϵ-entropy

For most systems, the computation of Kolmogorov–Sinai entropy (2.12) is practically impossible, because it involves the limit on arbitrary fine resolution and infinite times. However, in the same philosophy of the FSLE, by relaxing the requirement of arbitrary accuracy, one can introduce the ϵ-entropy which measures the amount of information for reproducing a trajectory with finite accuracy ϵ in phase-space. Roughly speaking the ϵ-entropy can be considered the counterpart, in information theory, of the FSLE. Such a quantity was originally introduced by Shannon [4], and by Kolmogorov [29]. Recently Gaspard and Wang [30] made use of this concept to characterize a large variety of processes.

We start with a continuous-time variable $\boldsymbol{x}(t) \in \mathbb{R}^d$, which represents the state of a d-dimensional system, we discretize the time by introducing an interval τ and we consider the new variable

$$\boldsymbol{X}^{(m)}(t) = (\boldsymbol{x}(t), \boldsymbol{x}(t+\tau), \ldots, \boldsymbol{x}(t+(m-1)\tau)) \,. \tag{3.12}$$

Of course $\boldsymbol{X}^{(m)}(t) \in \mathbb{R}^{md}$ and it corresponds to the trajectory which lasts for a time $T = m\tau$.

In data analysis, the space where the state of the system lives is unknown and usually only a scalar variable $u(t)$ can be measured. Then, one considers vectors $(u(t), u(t+\tau), \ldots, u(t+m\tau-\tau))$, that live in $\mathbb{R}^m$ and allow a reconstruction of the original phase space, known as delay embedding in the literature [31–33], and it is a special case of (3.12). Introduce now a partition of the phase space $\mathbb{R}^d$, using cells of edge ϵ in each of the d directions. Since the region where a bounded motion evolves contains a finite number of cells, each $\boldsymbol{X}^{(m)}(t)$ can be coded into a word of length m, out of a finite alphabet:

$$\boldsymbol{X}^{(m)}(t) \longrightarrow W^m(\epsilon, t) = (i(\epsilon, t), i(\epsilon, t+\tau), \ldots, i(\epsilon, t+m\tau-\tau)) \,, \tag{3.13}$$

where $i(\epsilon, t+j\tau)$ labels the cell in $\mathbb{R}^d$ containing $\boldsymbol{x}(t+j\tau)$. From the time evolution one obtains, under the hypothesis of ergodicity, the probabilities $P(W^m(\epsilon))$ of the admissible words $\{W^m(\epsilon)\}$. We can now introduce the (ϵ, τ)-entropy per unit time, $h(\epsilon, \tau)$ [4]:

$$h(\epsilon, \tau) = \frac{1}{\tau} \lim_{m \to \infty} \frac{1}{m} H_m(\epsilon, \tau) \,, \tag{3.14}$$

where H_m is the block entropy of blocks (words) with length m:

$$H_m(\epsilon, \tau) = - \sum_{\{W^m(\epsilon)\}} P(W^m(\epsilon)) \ln P(W^m(\epsilon)) \,. \tag{3.15}$$

For the sake of simplicity, we ignored the dependence on details of the partition. To make $h(\epsilon, \tau)$ partition-independent one has to consider a generic partition of the phase space $\{\mathcal{A}\}$ and to evaluate the Shannon entropy on this partition: $h_{\mathrm{Sh}}(\mathcal{A}, \tau)$. The ε-entropy is thus defined as the infimum over all partitions for which the diameter of each cell is less than ε [30]:

$$h(\varepsilon, \tau) = \inf_{\mathcal{A}:\mathrm{diam}(\mathcal{A}) \leq \varepsilon} h_{\mathrm{Sh}}(\mathcal{A}, \tau) \,. \tag{3.16}$$

Note that the time dependence in (3.16) is trivial for deterministic systems, and that in the limit $\epsilon \to 0$ one recovers the Kolmogorov–Sinai entropy

$$h_{\mathrm{KS}} = \lim_{\epsilon \to 0} h(\epsilon, \tau) \,.$$

4. Characterization of complexity and system modeling

In the previous Sections, we discussed the characterization of dynamical behaviors when the evolution laws are known either exactly or with some degree of uncertainty. In experimental investigations, however, only time records of some observable are available, while the equation of motion for the observable are generally unknown. The predictability problem of this latter case, at least from a conceptual point of view, can be treated as if the evolution laws were known. Indeed, in principle, the embedding technique allows for a reconstruction of the phase space [31–33]. Nevertheless there are rather severe limitations for high dimensional systems [34] and even in low dimensional ones non trivial features appear in the presence of noise [32]. In this Section we show that an entropic analysis at different resolution scales provides a pragmatic classification of a signal and gives suggestions for modeling of systems. In particular we illustrate, using some examples, how quantities such as the ϵ-entropy or the FSLE can display a subtle transition from the large to the small scales. A negative consequence of this is the difficulty in distinguishing, only from data analysis, a genuine deterministic chaotic system from one with intrinsic randomness [35]. On the other hand, the way the ϵ-entropy or FSLE depends on the (resolution) scale, allows for a classification of the stochastic or chaotic character of a signal, and this gives some freedom in modeling the system.

4.1. How random is a random number generator?

The "true character" of the number sequence $(x_1, x_2, \ldots)$ obtained by a (pseudo) random number generator (PRNG) on a computer is an issue of paramount importance in computer simulations and modeling. One would like to have a sequence with a random character as much as possible, but is forced to use deterministic algorithms to generate $(x_1, x_2, \ldots)$. This subsection is mainly based on the paper [36]. A simple and popular PRNG is the multiplicative congruent one:

$$\begin{aligned} z_{n+1} &= N_1 z_n \mod N_2\,, \\ x_{n+1} &= \frac{z_{n+1}}{N_2}\,, \end{aligned} \tag{4.1}$$

with an integer multiplier N_1 and modulus N_2. The $\{z_n\}$ are integer numbers from which one hopes to generate sequence of random variables $\{x_n\}$, which are uncorrelated and uniformly distributed in the unit interval. A first problem arises from the periodic nature of the rule (4.1) as a consequence of its discrete nature. Note that the rule (4.1) can be interpreted also as a deterministic dynamical system, *i.e.*

$$x_{n+1} = N_1 x_n \mod 1\,, \tag{4.2}$$

which has a uniform invariant measure and a KS entropy $h_{\rm KS} = \lambda = \ln N_1$. When imposing the integer arithmetics of Eq. (4.1) onto this system, we are, in the language of dynamical systems, considering an unstable periodic orbit of Eq. (4.2), with the particular constraint that, to achieve the period $N_2 - 1$ (*i.e.* all integers $< N_2$ should belong to the orbit of Eq. (4.1)), it has to contain all values k/N_2, with $k = 1, 2, \cdots, N_2 - 1$. Since the natural invariant measure of Eq. (4.2) is uniform, such an orbit represents the measure of a chaotic solution in an optimal way. Every sequence of a PRNG is characterized by two quantities: its period $\mathcal{T}$ and its positive Lyapunov exponent λ, which is identical to the entropy of a chaotic orbit of the equivalent dynamical system. Of course, a good random number generator must have a very large period, and as large as possible entropy.

It is natural to ask how this apparent randomness can be reconciled with the facts that (a) the PRNG is a deterministic dynamical systems (b) it is a discrete state system. If the period is long enough, on shorter times only point (a) matters and it can be discussed in terms of the behavior of the ϵ-entropy, $h(\epsilon)$. At high resolutions ($\epsilon \leq 1/N_1$), it seems rather reasonable to think that the true deterministic chaotic nature of the congruent rule shows up, and, therefore, $h(\epsilon) \simeq h_{\rm KS} = \ln N_1$. On the other hand, for $\epsilon \geq 1/N_1$, one expects to observe the "apparent random" behavior of the system, *i.e.* $h(\epsilon) \sim \ln(1/\epsilon)$, see Fig 3.

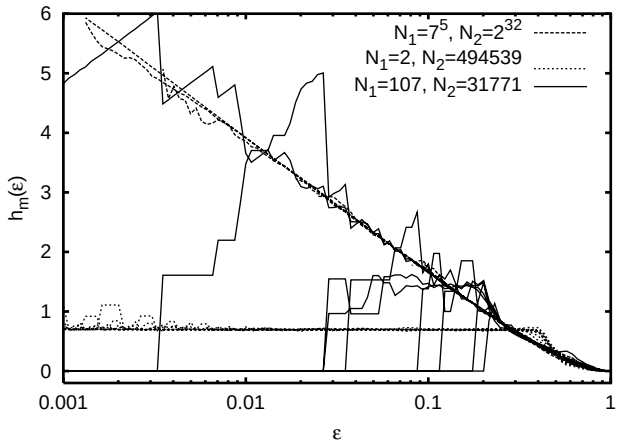

Fig. 3. The ϵ-entropies, $h_m(\epsilon)$, at varying the embedding dimension m for the multiplicative congruential random number generator Eq. 4.1 for different choices of N_1 and N_2.

4.2. High dimensional systems

We discuss an example of high-dimensional system with a non-trivial behavior at varying the resolution scales, namely the emergence of nontrivial collective behavior.

Let us consider a globally coupled map (GCM) defined as follows

$$x_n(t+1) = (1-\varepsilon) f_a(x_n(t)) + \frac{\varepsilon}{N} \sum_{i=1}^{N} f_a(x_i(t)) \,, \tag{4.3}$$

where N is the total number of elements, and $f_a(u)$ is a chaotic map on the interval $[0,1]$, depending on the control parameter a.

The evolution of a macroscopic variable, *e.g.*, the center of mass

$$m(t) = \frac{1}{N} \sum_{i=1}^{N} x_i(t) \,, \tag{4.4}$$

upon varying ε and a in Eq. (4.3), displays different behaviors [38]:

(a) *Standard Chaos*: $m(t)$ obeys a Gaussian statistics with a standard deviation $\sigma_N = \sqrt{\langle m(t)^2 \rangle - \langle m(t) \rangle^2} \sim N^{-1/2}$;

(b) *Macroscopic Periodicity*: $m(t)$ is a superposition of a periodic function and small fluctuations $O(N^{-1/2})$;

(c) *Macroscopic Chaos*: $m(t)$ exhibits an irregular motion, as seen by plotting $m(t)$ *vs* $m(t-1)$. The plot sketches a structured function (with thickness $\sim N^{-1/2}$), and suggests a chaotic motion for $m(t)$.

In the case of *macroscopic chaos*, the center of mass is expected to evolve with typical times longer than the characteristic time $1/\lambda_1$ of the full dynamics (microscopic dynamics); λ_1 being the Lyapunov exponent of the GCM. Indeed, conceptually, macroscopic chaos for GCM can be thought of as the analogous of the hydro-dynamical chaos for molecular motion. In spite of a huge microscopic Lyapunov exponent ($\lambda_1 \sim 1/\tau_c \sim 10^{11} s^{-1}$, τ_c is the collision time), one can have rather different behaviors at a hydrodynamical (coarse grained) level: regular motion ($\lambda_{\rm hydro} \leq 0$) or chaotic motion ($0 < \lambda_{\rm hydro} \ll \lambda_1$). In principle, if the hydrodynamic equations were known, a characterization of the macroscopic behavior would be possible by means of standard dynamical system techniques. However, in generic CML there are no general systematic methods to build up the macroscopic equations, apart from particular cases [37]. We recall that for chaotic systems, in the limit of infinitesimal perturbations $\delta \to 0$, one has $\lambda(\delta) \to \lambda_1$, *i.e.*

$\lambda(\delta)$ displays a plateau at the value λ_1 for sufficiently small δ. However, for non infinitesimal δ, one can expect that the δ-dependence of $\lambda(\delta)$ may give information on the characteristic time-scales governing the system, and, hence, it could be able to characterize the macroscopic motion. In particular, at large scales ($\delta \gg 1/\sqrt{N}$), the fast microscopic components saturate and $\lambda(\delta) \approx \lambda_M$, where λ_M can be fairly called the "macroscopic" Lyapunov exponent.

The FSLE has been determined by looking at the evolution of $|\delta m(t)|$, which has been initialized at the value $\delta m(t) = \delta_{\min}$ by shifting all the elements of the unperturbed system by the quantity $\delta_{\min}$ (*i.e.* $x_i'(0) = x_i(0) + \delta_{\min}$), for each realization. The computation has been performed by choosing the tent map as local map, but similar results can be obtained for other maps [38, 39].

The main result can be summarized as follows:

- at small δ ($\ll 1/\sqrt{N}$), where N is the number of elements, the "microscopic" Lyapunov exponent is recovered, *i.e.* $\lambda(\delta) \approx \lambda_{\text{micro}}$,
- at large δ ($\gg 1/\sqrt{N}$), another plateau $\lambda(\delta) \approx \lambda_{\text{macro}}$ appears, which can be much smaller than the microscopic one.

The emerging scenario is that, at a coarse-grained level, *i.e.* $\delta \gg 1/\sqrt{N}$, the system can be described by an "effective" hydro-dynamical equation (which in some cases can be low-dimensional), while the "true" high-dimensional character appears only at very high resolution, *i.e.*

$$\delta \leq \delta_{\text{c}} = O\left(\frac{1}{\sqrt{N}}\right).$$

4.3. Diffusion in deterministic systems and Brownian motion

Consider the following map which generates a diffusive behavior on the large scales [40]:

$$x_{t+1} = [x_t] + F\left(x_t - [x_t]\right), \tag{4.5}$$

where $[x_t]$ indicates the integer part of x_t and $F(y)$ is given by:

$$F(y) = \begin{cases} (2+\alpha)y & \text{if } y \in [0, 1/2] \\ (2+\alpha)y - (1+\alpha) & \text{if } y \in [1/2, 1]. \end{cases} \tag{4.6}$$

The largest Lyapunov exponent λ can be obtained immediately: $\lambda = \ln|F'|$, with $F' = dF/dy = 2+\alpha$. One expects the following scenario for $h(\epsilon)$:

$$h(\epsilon) \approx \lambda \qquad \text{for} \quad \epsilon < 1, \tag{4.7}$$

$$h(\epsilon) \propto \frac{D}{\epsilon^2} \qquad \text{for} \quad \epsilon > 1, \tag{4.8}$$

where D is the diffusion coefficient, $\langle (x_t - x_0)^2 \rangle \approx 2D\,t$ for large t. Consider now a stochastic system, namely a noisy map

$$x_{t+1} = [x_t] + G\left(x_t - [x_t]\right) + \sigma \eta_t\,, \tag{4.9}$$

where $G(y)$, as shown in Fig. 4, is a piece wise linear map which approximates the map $F(y)$, and η_t is a stochastic process uniformly distributed in the interval $[-1, 1]$, and no correlation in time. When $|dG/dy| < 1$, as is the case we consider, the map (4.9), in the absence of noise, gives a non-chaotic time evolution.

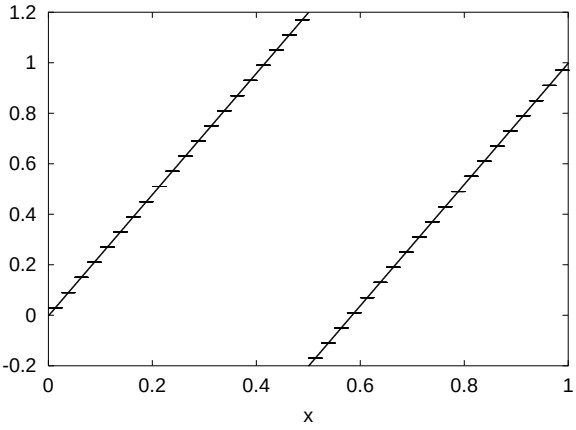

Fig. 4. The map $F(x)$ (4.6) for $\alpha = 0.4$ is shown with superimposed the approximating (regular) map $G(x)$ (4.9) obtained by using 40 intervals of slope 0.

Now we compare the finite size Lyapunov exponent for the chaotic map (4.5) and for the noisy one (4.9). In the latter the FSLE has been computed using two different realizations of the noise. In Fig. 5 we show $\lambda(\epsilon)$ *versus* ϵ for the two cases. The two curves are practically indistinguishable in the region $\epsilon > \sigma$. The differences appear only at very small scales $\epsilon < \sigma$ where one has a $\lambda(\epsilon)$ which grows with ϵ for the noisy case, remaining at the same value for the chaotic deterministic case.

Both the FSLE and the (ϵ, τ)-entropy analysis show that we can distinguish three different regimes observing the dynamics of (4.9) on different length scales. On the large length scales $\epsilon > 1$ we observe diffusive behavior in both models. On length scales $\sigma < \epsilon < 1$ both models show chaotic deterministic behavior, because the entropy and the FSLE are independent of ϵ and larger than zero. Finally on the smallest length scales $\epsilon < \sigma$ we see stochastic behavior for the system (4.9), *i.e.* $h(\varepsilon) \sim -\ln(\varepsilon)$, while the system (4.5) still shows chaotic behavior.

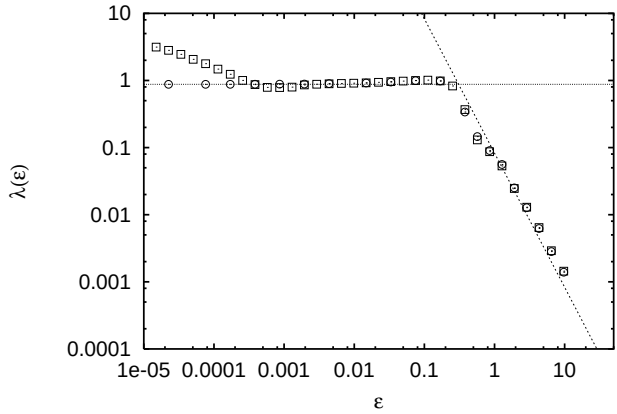

Fig. 5. Lyapunov exponent $\lambda(\epsilon)$ versus ϵ obtained for the map $F(y)$ (4.6) with $\alpha = 0.4$ ($\circ$) and for the noisy (regular) map (4.9) ($\Box$) with 10^4 intervals of slope 0.9 and $\sigma = 10^{-4}$. Straight lines indicate the Lyapunov exponent $\lambda = \ln 2.4$ and the diffusive behavior $\lambda(\epsilon) \sim \epsilon^{-2}$.

4.4. On the distinction between chaos and noise

The above examples show that the distinction between chaos and noise can be a highly non trivial task, which makes sense only in very peculiar cases, *e.g.*, very low dimensional systems. Nevertheless, even in this case, the entropic analysis can be unable to recognize the "true" character of the system due to the lack of resolution. Again, the comparison between the diffusive map (4.5) and the noisy map (4.9) is an example of these difficulties. For $\sigma \leq \epsilon \leq 1$ both the system (4.5) and (4.9), in spite of their "true" character, will be classified as chaotic, while for $\epsilon \geq 1$ both can be considered as stochastic.

In high-dimensional chaotic systems, with N degrees of freedom, one has typically $h(\epsilon) = h_{\mathrm{KS}} \sim O(N)$ for $\epsilon \leq \epsilon_{\mathrm{c}}$ (where $\epsilon_{\mathrm{c}} \to 0$ as $N \to \infty$) while for $\epsilon \geq \epsilon_{\mathrm{c}}$, $h(\epsilon)$ decreases, often with a power law [30]. Since also in some stochastic processes the ϵ-entropy obeys a power law, this can be a source of confusion.

These kind of problems are not abstract ones, as a recent debate on "microscopic chaos" demonstrates [41–43]. The detection of microscopic chaos by data analysis has been recently addressed in a work of Gaspard *et al.* [41]. These authors, from an entropic analysis of an ingenious experiment on the position of a Brownian particle in a liquid, claim to give an empirical evidence for microscopic chaos. In other words, they state that the diffusive behavior observed for a Brownian particle is the consequence of chaos at a molecular level. Their work can be briefly summarized as follows: from a long ($\approx 1.5 \times 10^5$ data) record of the position of a Brownian particle they

compute the ϵ-entropy with the Cohen–Procaccia method [44] from which they obtain:

$$h(\epsilon) \sim \frac{D}{\epsilon^2}\,, \tag{4.10}$$

where D is the diffusion coefficient. Then, *assuming* that the system is deterministic, and making use of the inequality $h(\epsilon > 0) \leq h_{\mathrm{KS}}$, they conclude that the system is chaotic. However, their result does not give a direct evidence that the system is deterministic and chaotic. Indeed, the power law (4.10) can be produced with different mechanisms:

1 a genuine chaotic system with diffusive behavior, as the map (4.6);

2 a non chaotic system with some noise, as the map (4.9), or a genuine Brownian system;

3 a deterministic linear non chaotic system with many degrees of freedom (see for instance [45]);

4 a "complicated" non chaotic system as the Ehrenfest wind-tree model where a particle diffuses in a plane due to collisions with randomly placed, fixed oriented square scatters, as discussed by Cohen *et al.* [42] in their comment to Ref. [41].

It seems to us that the weak points of the analysis in Ref. [41] are:

(a) the explicit assumption that the system is deterministic;

(b) the limited number of data points and therefore limitations in both resolution and block length.

The point (a) is crucial, without this assumption (even with an enormous data set) it is not possible to distinguish between 1 and 2. One has to say that in the cases 3 and 4 at least in principle it is possible to understand that the systems are "trivial" (*i.e.* not chaotic) but for this one has to use a huge number of data. For example Cohen *et al.* [42] estimated that in order to distinguish between 1 and 4 using realistic parameters of a typical liquid, the number of data points required has to be at least $\sim 10^{34}$.

Concluding, we have the apparently paradoxical result that "complexity" helps in the construction of models. Basically, in the case in which one has a variety of behaviors at varying the scale resolution, there is a certain freedom on the choice of the model to adopt. For some systems the behavior at large scales can be realized both with chaotic deterministic models or suitable stochastic processes. From a pragmatic point of view, the fact that in certain stochastic processes $h(\epsilon) \sim \epsilon^{-\alpha}$ can be indeed extremely useful for modeling such high-dimensional systems. Perhaps, the most relevant

case in which one can use this freedom in modeling is the fully developed turbulence whose non infinitesimal (the so-called inertial range) properties can be successfully mimicked in terms of multi-affine stochastic process (see Ref. [46]).

5. Concluding remarks

The guideline of this paper has been *the interpretation of different aspects of the predictability of a system as a way to characterize its complexity.*

We have discussed the relation between chaoticity, the Kolmogorov–Sinai entropy and algorithmic complexity. As clearly exposed in the seminal works of Alekseev and Yakobson [14] and Ford [17], the time sequences generated by a system with sensitive dependence on initial conditions have non-zero algorithmic complexity. A relation exists between the maximal compression of a sequence and its KS-entropy. Therefore, one can give a definition of complexity, without referring to a specific description, as an intrinsic property of the system.

The study of these different aspects of predictability constitutes a useful method for a quantitative characterization of "complexity", suggesting the following equivalences:

$$\textbf{Complex} = \textbf{Uncompressible} = \textbf{Unpredictable} \tag{5.1}$$

The above point of view, based on dynamical systems and information theory, quantifies the complexity of a sequence considering each symbol relevant but it does not capture the structural level. Let us clarify this point with the following example. A binary sequence obtained with a coin tossing is, from the point of view adopted in this review, complex since it cannot be compressed (*i.e.* it is unpredictable). On the other hand such a sequence is somehow trivial, *i.e.* with low "organizational" complexity. It would be important to introduce a quantitative measure of this intuitive idea. The progresses of the research on this intriguing and difficult issue are still rather slow. We just mention some of the most promising proposals as the logical depth and the sophistication [47].

We thank G. Boffetta, A. Celani, D. Vergni and M. Cencini for the long collaboration and many fruitful discussions on the subject of the paper.

REFERENCES

[1] S. Laplace, *Essai philosophique sur les probabilités*, Courcier, Paris 1814.

[2] G. Boffetta, M. Cencini, M. Falcioni, A. Vulpiani, *Phys. Rep.* **356**, 367 (2002).

[3] V.I. Oseledec, *Trans. Mosc. Math. Soc.* **19**, 197 (1968).

[4] C.E. Shannon, *The Bell System Technical J.* **27**, 623 (1948); **27**, 379 (1948).

[5] A.N. Kolmogorov, *Prob. Info. Trans.* **1**, 1 (1965).

[6] R.J. Solomonoff, *Inform. Contr.* **7**, 1 (1964); **7**, 224 (1964).

[7] A.I. Khinchin, *Mathematical Foundations of Information Theory*, Dover, New York 1957.

[8] D. Welsh, *Codes and Cryptography*, Clarendon Press, Oxford 1989.

[9] P. Billingsley, *Ergodic Theory and Information*, Wiley, New York 1965.

[10] J.P. Eckmann, D. Ruelle, *Rev. Mod. Phys.* **57**, 617 (1985).

[11] W. Krieger, *Trans. Am. Math. Soc.* **149**, 453 (1970).

[12] G.J. Chaitin, *Information, Randomness and Incompleteness*, 2nd edition, World Scientific, Singapore 1990.

[13] P. Grassberger, *Int. J. Theor. Phys.* **25**, 907 (1986).

[14] V.M. Alekseev, M.V. Yakobson, *Phys. Rep.* **75**, 287 (1981).

[15] A.A. Brudno, *Trans. Moscow Math. Soc.* **44**, 127 (1983).

[16] H. White, *Erg. Theory Dyn. Syst.* **13**, 807 (1993).

[17] J. Ford, *Physics Today*, **36**, 40 (1983).

[18] G. Paladin, M. Serva, A. Vulpiani, *Phys. Rev. Lett.* **74**, 66 (1995).

[19] H. Fujisaka, *Prog. Theor. Phys.* **70**, 1264 (1983).

[20] R. Benzi, G. Paladin, G. Parisi, A. Vulpiani, *J. Phys.* **A18**, 2157 (1985).

[21] E. Aurell, G. Boffetta, A. Crisanti, G. Paladin, A. Vulpiani, *Phys. Rev. Lett.* **77**, 1262 (1996).

[22] G. Boffetta, G. Paladin, A. Vulpiani, *J. Phys.* **A29**, 2291 (1996).

[23] E.N. Lorenz, *Tellus* **21**, 3 (1969).

[24] A. Monin, *Weather Prediction as a Problem in Physics*, MIT Press, Moscow 1973.

[25] V. Artale, G. Boffetta, A. Celani, M. Cencini, A. Vulpiani, *Phys. Fluids* **A9**, 3162 (1997).

[26] G. Boffetta, A. Crisanti, F. Paparella, A. Provenzale, A. Vulpiani, *Physica D* **116**, 301 (1998).

[27] H. Kantz, T. Letz, *Phys. Rev.* **E61**, 2533 (2000).

[28] A. Torcini, P. Grassberger, A. Politi, *J. Phys.* **A27**, 4533 (1995).

[29] A.N. Kolmogorov, *IRE Trans. Inf. Theory*, **1**, 102 (1956).

[30] P. Gaspard, X.J. Wang, *Phys. Rep.* **235**, 291 (1993).

[31] F. Takens, "Detecting strange attractors in turbulence" in *Dynamical Systems and Turbulence (Warwick 1980)*, Vol. 898 of *Lecture Notes in Mathematics*, D.A. Rand and L.-S. Young (eds.), pg. 366, Springer-Verlag, Berlin 1980.

[32] H. Kantz, T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge 1997.

[33] H.D.I. Abarbanel, *Analysis of Observed Chaotic Data*, Springer-Verlag, New York 1996.

[34] P. Grassberger, *Phys. Scr.* **40**, 346 (1989).

[35] M. Cencini, M. Falcioni, H. Kantz, E. Olbrich, A. Vulpiani, *Phys. Rev.* **E62**, 427 (2000).

[36] H. Kantz, E. Olbrich, *Physica A* **280**, 34 (2000).

[37] A.S. Pikovsky, J. Kurths, *Physica D* **76**, 411 (1994).

[38] M. Cencini, M. Falcioni, D. Vergni, A. Vulpiani, *Physica D* **130**, 58 (1999).

[39] T. Shibata, K. Kaneko, *Phys. Rev. Lett.* **81**, 4116 (1998).

[40] M. Schell, S. Fraser, R. Kapral, *Phys. Rev.* **A26**, 504 (1982).

[41] P. Gaspard, M.E. Briggs, M.K. Francis, J.V. Sengers, R.W. Gammon, J.R. Dorfman, R.V. Calabrese, *Nature* **394**, 865 (1998).

[42] C. Dettman, E. Cohen, H. van Beijeren, *Nature* **401**, 875 (1999).

[43] P. Grassbeger, T. Schreiber, *Nature* **401**, 875 (1999).

[44] A. Cohen, I. Procaccia, *Phys. Rev.* **A31**, 1872 (1985).

[45] P. Mazur, E. Montroll, *J. Math. Phys.* **1**, 70 (1960).

[46] L. Biferale, G. Boffetta, A. Celani, A. Crisanti, A. Vulpiani, *Phys. Rev.* **E57**, R6261 (1998).

[47] R. Badii, A. Politi, *Complexity. Hierarchical Structures and Scaling in Physics*, Cambridge University Press, Cambridge, UK, 1997.