KOLMOGOROV ENTROPY FROM TIME SERIES USING INFORMATION-THEORETIC FUNCTIONALS

*Milan Paluš**

Abstract: A technique for identification and quantification of chaotic dynamics in experimental time series is presented. It is based on evaluation of information-theoretic functionals - redundancies which, estimated from data generated by a low-dimensional chaotic dynamical system, have specific properties reflecting a positive information production rate. This rate, measured by metric (Kolmogorov-Sinai) entropy, can be directly estimated from the redundancies.

Key words: *Kolmogorov-Sinai (metric) entropy, time series analysis, dynamical systems, ergodic theory, information theory*

Received: December 17, 1996
Revised and accepted: May 23, 1997

1. Introduction

Recent results in the theory of nonlinear dynamical systems and deterministic chaos, which are applicable to the analysis of experimental time series (see e.g. [1], [2]), have brought a new alternative to generally used linear stochastic methods and significantly changed theoretical paradigms in interpretation of obtained results. Purely phenomenological parameters used in stochastic methods have been replaced in this new approach by invariants characterizing dynamical properties of systems under study. Extraction of these invariants may be regarded as the first step to building a model of system dynamics. Applications of this approach in the analysis of experimental time series have been reported in many fields of natural and social sciences.

A significant portion of work devoted to classification of chaotic behaviour is oriented to estimations of geometric or static invariants such as dimensions or the number of degrees of freedom. Even some algorithms proposed for the estimation of dynamical entropies are derived from dimensional algorithms [3], [4], [5] and are related to the concept of Rényi entropies [6]. A. N. Kolmogorov, who introduced the theoretical concept of classification of dynamical system by information

*Milan Paluš
Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic, E-mail: mp@uivt.cas.cz

rates [7], was inspired by information theory and generalized the notion of the entropy of an information source [7], [8]. A possibility to use ideas and methods from the information theory in the field of nonlinear dynamics applied to analysis of experimental data was demonstrated by Shaw [9], [10] and Fraser [11], [12], [13].

This paper considers a method for estimating the metric (Kolmogorov-Sinai) entropy from experimental time series by using information-theoretic functionals, as proposed by Fraser [11], [12], [13]. The method itself represents an "interface" between the ergodic theory of dynamical systems and the information theory. While in his original work [11], Fraser analyses information aspects of chaotic dynamics in detail, here we concentrate on attributes of dynamical systems studied in the ergodic theory, such as mixing and generating partitions, and we demonstrate how they are reflected in behaviour of information-theoretic functionals estimated from chaotic data. Necessary elements of the ergodic theory are introduced in Section 2. It is based on books [14], [15], [16], [17], in which further details and proofs of theorems can be found. Basic concepts of the information theory are presented in Section 3. For more details we refer to the books [18], [19], [20], [21], [22]. The relation between the entropy of dynamical systems and information functionals is explained in Section 4. The technique for estimating the metric entropy, based on this relation, is described in Section 5 and its application to well-known chaotic systems is presented in Section 6. In Section 7, further applications of this approach are presented. Influence of additive noise on quantification and identification of chaotic dynamics by the proposed information theoretic method is studied in Section 8. The conclusion is given in Section 9.

2. Elements of Ergodic Theory

We will study long-term average behaviour of systems. Let a collection of all states of a system form a space $\mathcal{S}$. Time evolution of the system is represented by a transformation T: $\mathcal{S} \to \mathcal{S}$, where Tx is taken as the state at time 1 of the system which at time 0 is in state x. In the case of continuous time one can consider a one-parameter family $\{T_t; t \in \Re\}$ ($\Re$ is the set of real numbers) of maps of $\mathcal{S}$ into itself. Suppose that $T_{t+s} = T_t T_s$, so that $\{T_t; t \in \Re\}$ is a flow on $\mathcal{S}$. We will restrict on such cases when $\mathcal{S}$ is a measure space and T is a measure preserving transformation.

Let $\mathcal{B}$ be a σ-algebra of measurable subsets of $\mathcal{S}$ and let μ be a countably additive non-negative set function (measure) on $\mathcal{B}$ such that $\mu(\mathcal{S}) = 1$ and such that $\mathcal{B}$ contains all subsets of measure 0. Then $\mathcal{S}$, $\mathcal{B}$ and μ form a complete probability space $(\mathcal{S}, \mathcal{B}, \mu)$.

Let T: $\mathcal{S} \to \mathcal{S}$ be a measurable map, i.e., $T^{-1}\mathcal{B} = \mathcal{B}$; and $\forall E \in \mathcal{B}$: $\mu(T^{-1}E) = \mu(E)$ holds. Then the system $(\mathcal{S}, \mathcal{B}, \mu, T)$ is called the measure preserving transformation (abbreviated m.p.t.).

This is the typical "set-up" used in the ergodic theory. A reader more familiar with the theory of dynamical systems, or "differentiable dynamics" is used to begin with a diffeomorphism T on a differentiable manifold $\mathcal{S}$. Then the question of existence of an invariant measure μ, i.e., of the measure for which T is a m.p.t., arises. This problem is discussed, e.g., in [17], [23]; here we simply suppose that such a measure exists.

The following properties of m.p.t.'s are important for further considerations here:

Definition 2.1. A m.p.t. $(\mathcal{S}, \mathcal{B}, \mu, T)$ is called ergodic if $\forall B \in \mathcal{B}$ with $\mu(B) > 0$:

$$\mu(\bigcup_{n=1}^{\infty} T^{-n}B) = 1.$$

Theorem 2.1. A m.p.t. $(\mathcal{S}, \mathcal{B}, \mu, T)$ is ergodic iff $\forall A, B \in \mathcal{B}$:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) = \mu(A)\mu(B).$$

Definition 2.2. Let $(\mathcal{S}, \mathcal{B}, \mu, T)$ be a m.p.t. of a probability space $(\mathcal{S}, \mathcal{B}, \mu)$.

(i) T is weakly mixing if $\forall A, B \in \mathcal{B}$:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| = 0.$$

(ii) T is strongly mixing if $\forall A, B \in \mathcal{B}$:

$$\lim_{n\to\infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B).$$

Clearly, every strongly mixing transformation is weakly mixing and every weakly mixing transformation is ergodic.

Theorem 2.2. Let $(\mathcal{S}, \mathcal{B}, \mu)$ be a measure space and let $\mathcal{A}$ be a semi-algebra that generates $\mathcal{B}$. Let T: $\mathcal{S} \to \mathcal{S}$ be a m.p.t. Then:

(i) T is ergodic iff $\forall A, B \in \mathcal{A}$:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) = \mu(A)\mu(B),$$

(ii) T is weakly mixing iff $\forall A, B \in \mathcal{A}$:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| = 0, \text{ and}$$

(iii) T is strongly mixing iff $\forall A, B \in \mathcal{A}$:

$$\lim_{n\to\infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B).$$

Theorem 2.3. Let T be a m.p.t. of a probability space $(\mathcal{S}, \mathcal{B}, \mu)$ and $\mathcal{Z}^+$ be the set of positive whole numbers. Then T is weakly mixing iff for every pair of elements $A, B \in \mathcal{B}$ there is a subset $J(A, B) \subset \mathcal{Z}^+$ of density zero such that

$$\lim_{n \to \infty} \mu(T^{-n} A \cap B) = \mu(A)\mu(B), \text{ for } n \notin J(A, B).$$

Intuitively we can say that T is strongly mixing if for any set A the sequence of sets $T^{-n}A$ is asymptotically independent of any set B. Ergodicity means $T^{-n}A$ is independent of B on average, for each pair of sets $A, B \in \mathcal{B}$. Considering Theorem 2.3. we can say that for weakly mixing T and for any $A \in \mathcal{B}$, the sequence $T^{-n}A$ becomes asymptotically independent of any other set $B \in \mathcal{B}$ provided we neglect a few instants of time. Theorem 2.2. gives us the possibility to apply these considerations also to elements of finite partitions of $\mathcal{S}$.

Definition 2.3. A partition of $(\mathcal{S}, \mathcal{B}, \mu)$ is a disjoint collection of elements of $\mathcal{B}$ which union is $\mathcal{S}$.

Definition 2.4. Let $\alpha = \{A_1, \ldots, A_k\}$ be a finite partition of $(\mathcal{S}, \mathcal{B}, \mu)$. Then the entropy of the partition α is

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

We will be interested in finite partitions, however, all the following considerations can be extended to countable partitions with finite entropy.

Definition 2.5. Let $\alpha = \{A_1, \ldots, A_n\}$ and $\beta = \{B_1, \ldots, B_m\}$ be partitions of $(\mathcal{S}, \mathcal{B}, \mu)$. Then

(i) $T^{-1}\alpha$ is the partition $\{T^{-1}A_1, \ldots, T^{-1}A_n\}$.

(ii) β is a refinement of α, written $\beta \geq \alpha$, if each B_j is, up to a set of measure 0, a subset of some A_i .

(iii) $\alpha \vee \beta$ (the least common refinement of α and β) is the partition $\{A_i \cap B_j, i = 1, \ldots, n; j = 1, \ldots, m\}$.

(iv) The least common refinement $\bigvee_{i=1}^{n} \alpha_i$ of the partitions $\alpha_1, \ldots, \alpha_n$ is:

$$\bigvee_{i=1}^{n} \alpha_i = \alpha_1 \vee \alpha_2 \vee \ldots \vee \alpha_n.$$

(v) The conditional entropy $H(\alpha|\beta)$ of α given β is

$$H(\alpha|\beta) = -\sum_{i,j} \mu(A_i \cap B_j) \log \frac{\mu(A_i \cap B_j)}{\mu(B_j)},$$

omitting the j-terms with $\mu(B_j) = 0$.

Definition 2.6. Let α be a finite partition of $(\mathcal{S}, \mathcal{B}, \mu)$ and T be a m.p.t. on $(\mathcal{S}, \mathcal{B}, \mu)$. The entropy of the transformation T with respect to the partition α is

$$h(T, \alpha) = \lim_{k \to \infty} \frac{1}{k} H(\bigvee_{i=0}^{k-1} T^{-i}\alpha).$$

Definition 2.7. Let T be a m.p.t. on $(\mathcal{S}, \mathcal{B}, \mu)$. The entropy of the transformation T is

$$h(T) = \sup_{\alpha} h(T, \alpha),$$

where the supremum is taken over all finite partitions of $(\mathcal{S}, \mathcal{B}, \mu)$. The entropy $h(T)$ is called metric entropy or measure-theoretic entropy or Kolmogorov-Sinai entropy.

Theorem 2.4. Let $(\mathcal{S}, \mathcal{B}, \mu, T)$ be a m.p.t. and $\alpha = \{A_1, \ldots, A_k\}$ be a finite partition of $(\mathcal{S}, \mathcal{B}, \mu)$. Then

$$h(T, \alpha) \leq H(\alpha) \leq \log k.$$

Theorem 2.5. Let α and β be partitions of $(\mathcal{S}, \mathcal{B}, \mu)$ and let $(\mathcal{S}, \mathcal{B}, \mu, T)$ be a m.p.t. If $\alpha \leq \beta$ then $h(T, \alpha) \leq h(T, \beta)$.

Theorem 2.6. Let $(\mathcal{S}, \mathcal{B}, \mu, T)$ be a m.p.t. and ξ be a finite partition of $(\mathcal{S}, \mathcal{B}, \mu)$. Then

$$h(T, \xi) = \lim_{n \to \infty} H(\xi | \bigvee_{i=1}^{n} T^{-i}\xi).$$

Theorem 2.7. Let $(\mathcal{S}, \mathcal{B}, \mu, T)$ be a m.p.t. Then

$$h(T^k) = |k| h(T), \ \forall k \in \mathcal{Z}.$$

For continuous flow T_t on $(\mathcal{S}, \mathcal{B}, \mu)$ the equality

$$h(T_t) = |t| h(T_1), \ \forall t \in \Re$$

holds.

Definition 2.8. A finite partition α of $(\mathcal{S}, \mathcal{B}, \mu)$ is called a generating partition (generator) with respect to a m.p.t. $(\mathcal{S}, \mathcal{B}, \mu, T)$ if

$$\bigvee_{i=-\infty}^{\infty} T^{-i}\alpha = \mathcal{B}$$

up to sets of measure 0.

Theorem 2.8. (Kolmogorov-Sinai theorem) If α is a generator with respect to $(\mathcal{S}, \mathcal{B}, \mu, T)$, then $h(T) = h(T, \alpha)$.

There are several important theorems about the existence of generating partitions, e.g.:

Theorem 2.9. (Krieger generator theorem) If T is an ergodic m.p.t. on a Lebesgue space with $h(T) < \infty$, then T has a finite generator.

3. Elements of Information Theory

Definition 3.1. Let X, Y be (continuous) random variables with probability distribution densities $p_X(x)$ and $p_Y(y)$.

(i) The entropy of distribution of a single variable, say X, is

$$H(X) = -\int p_X(x) \log p_X(x) dx.$$

(ii) For the joint distribution $p_{X,Y}(x, y)$ of X and Y the joint entropy is

$$H(X,Y) = -\int\int p_{X,Y}(x,y) \log p_{X,Y}(x,y) dxdy.$$

(iii) The conditional entropy $H(X|Y)$ of X given Y is

$$H(X|Y) = -\int\int p_{X,Y}(x,y) \log \frac{p_{X,Y}(x,y)}{p_Y(y)} dxdy.$$

(iv) The entropy of a distribution of a discrete random variable Z with values z_i and probability distribution $p(z_i)$, $i = 1, \ldots, k$, is

$$H(Z) = -\sum_{i=1}^{k} p(z_i) \log p(z_i).$$

Definitions (ii) and (iii) for discrete variables can be derived straightforwardly.

Definition 3.2. The mutual information $I(X;Y)$ quantifying the average amount of information which the variable X bears about the variable Y is

$$I(X;Y) = H(X) + H(Y) - H(X,Y).$$

Theorem 3.1. $I(X;Y) = 0$ iff $p_{X,Y}(x,y) = p_X(x)p_Y(y)$, i.e., iff X and Y are statistically independent.

Generalization of the definition of the joint entropy for n variables $X_1, \ldots, X_n$ is straightforward:

Definition 3.3. The joint entropy of distribution $p_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ of the n variables $X_1, \ldots, X_n$ is

$$H(X_1, \ldots, X_n) = -\int \ldots \int p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \log p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \ldots dx_n.$$

The mutual information will be generalized in two ways:

Definition 3.4. The redundancy $R(X_1; \ldots; X_n)$ of the variables $X_1, \ldots, X_n$, quantifying the average amount of common information contained in the variables $X_1, \ldots, X_n$, is defined as

$$R(X_1; \ldots; X_n) = H(X_1) + \ldots + H(X_n) - H(X_1, \ldots, X_n).$$

Definition 3.5. The marginal redundancy $\varrho(X_1, \ldots, X_{n-1}; X_n)$ quantifying the average amount of information about the variable X_n contained in the variables $X_1, \ldots, X_{n-1}$, is defined as

$$\varrho(X_1, \ldots, X_{n-1}; X_n) = H(X_1, \ldots, X_{n-1}) + H(X_n) - H(X_1, \ldots, X_n).$$

The following relations between redundancies and entropies can be obtained by a simple manipulation:

Theorem 3.2. $\varrho(X_1, \ldots, X_{n-1}; X_n) = R(X_1; \ldots; X_n) - R(X_1; \ldots; X_{n-1})$
Theorem 3.3. $\varrho(X_1, \ldots, X_{n-1}; X_n) = H(X_n) - H(X_n | X_1, \ldots, X_{n-1})$

4. Numerical Set-up and Relation between $h(T)$ and ϱ

In practical applications one deals with a time series $\{y(t)\}$ considered as a realization of a stochastic process $\{Y(t)\}$, which is stationary and ergodic. Then, due to ergodicity, all the subsequent information-theoretic functionals can be estimated by using time averages instead of ensemble averages, and the variables X_i are substituted as

$$X_i = y(t + (i-1)\tau), \tag{1}$$

where τ is a time lag. Due to stationarity, the redundancies

$$R^n(\tau) \equiv R(y(t); y(t+\tau); \ldots; y(t + (n-1)\tau)) \tag{2}$$

and

$$\varrho^n(\tau) \equiv \varrho(y(t), y(t+\tau), \ldots, y(t + (n-2)\tau); y(t + (n-1)\tau)) \tag{3}$$

are functions of the number n of variables and the time lag τ, and are independent of t.

A measure preserving system $(\mathcal{S}, \mathcal{B}, \mu, T)$ can correspond to a stochastic process $\{Y(t)\}$:
If T: $\mathcal{S} \to \mathcal{S}$ is a m.p.t., the orbit $\{T^n s;\ n \in \mathcal{Z}\}$ of a point $s \in \mathcal{S}$ represents a single evolution of the system. The σ-algebra $\mathcal{B}$ is thought of as a family of observable events with the T-invariant measure μ specifying the (time-independent) probabilities of their occurrence. A measurable function f: $\mathcal{S} \to \Re$ represents a measurement made on the system: $f(s)$, $f(Ts)$, $f(T^2 s)$, ... are values of a physically observable variable measured in successive instants of time – the experimental time series.

Conversely, any stationary stochastic process corresponds to a measure preserving system in a standard way: One can construct a map Φ mapping variables of a stochastic process to a sequence of points $\{q_n\}$ of a measure space $\mathcal{Q}$ and define shift transformation σ on the sequence $\{q_n\}$ as $\sigma q_n = q_{n+1}$. Due to stationarity of the original process, such a system is a m.p.t. (For more details see Ref. [14].)

Thus we can identify entropies of distributions of stochastic variables with partition entropies of the corresponding measure space and especially the conditional entropies from Theorems 2.6. and 3.3. Then, for $n \to \infty$ we have:

$$\varrho^n(\tau) \approx A_\xi - h(T_\tau, \xi), \tag{4}$$

where A_ξ is a parameter independent of n and τ (and, clearly, dependent on the partition ξ), and $h(T_\tau, \xi)$ is the entropy of (continuous) transformation T_τ with respect to the partition ξ, corresponding to the probability distribution $p(x_i)$.

Let ξ be a generating partition with respect to T (more precisely with respect to T_τ for a large enough range of τ's – see Sec. 5.2), then, considering Theorems 2.7, 2.8 we have

$$\lim_{n\to\infty} \varrho^n(\tau) = A - |\tau| h(T_1). \tag{5}$$

This assertion was originally conjectured by Fraser [11], [12], [13]. The mutual information was used for the estimation of metric entropy of one-dimensional maps firstly by Shaw [10]. Here we present a detailed study of this estimation technique and bring further numerical support for this conjecture.

Remark: There is an exact equality relation between the partition entropy and the entropy of the related distribution [24], provided they are defined on the same measure space. Here we want to estimate metric entropy of a dynamical system evolving in an n-dimensional state space $\mathcal{S}$ from a sequence of one-dimensional stochastic variables, providing they are images (projections) of a single trajectory of the dynamical system (see considerations above) mapped by a map f: $\mathcal{S} \to \Re$.

5. Estimation Technique

We will study the behaviour of $\varrho^n(\tau)$ as a function of τ and n for low-dimensional chaotic processes. Let us consider that the studied time series has been generated by an m-dimensional dynamical system, i.e., there is a m.p.t. $(\mathcal{S}, \mathcal{B}, \mu, T_\tau)$ and T_τ is a continuous-time dynamical system fulfilling the conditions of the existence and uniqueness theorem [25, 26] and a particular trajectory of T_τ is mapped from $\mathcal{S}$ to $\Re$. There is a unique trajectory passing through each point $s \in \mathcal{S}$ so that the evolution on the particular trajectory is fully determined by one m-dimensional point $s \in \mathcal{S}$. On the other hand, according to the theorem of Takens [27], m-tuples of m successive samples $y(t), \ldots, y(t + (m-1)\tau)$ form a mapping of the process $\{Y(t)\}$ to a space $\mathcal{Q}$, so that the sequence $\{q_i\}$ of the images of the m-tuples $y(t), \ldots, y(t + (m-1)\tau)$ is topologically equivalent to the original trajectory $\{s_i\}$ in $\mathcal{S}$. Hence a particular m-tuple $y(t), \ldots, y(t + (m-1)\tau)$ is equivalent to a point from $\mathcal{S}$ and thus it determines the rest of the series $\{y(t)\}$. This means that only the redundancies $\varrho^n(\tau)$ for $n \le m$ should be finite and for $n > m$ the redundancy $\varrho^n(\tau)$ should diverge. This is, however, theoretic consideration providing infinite precision. In experimental and numerical practice, measurement noise and finite precision cause that all estimated redundancies $\varrho^n(\tau)$ are finite and increasing with n. We can only suppose that the increase with n of $\varrho^n(\tau)$ for $n > m$ is lower than for $n \leq m$ and it is independent of τ.

Intuitively we can explain this supposition by the fact that adding another variable to n variables, $n < m$, the common information measured by $\varrho^n(\tau)$ is increased by specific dynamical information, i.e., the increase $\varrho^{n+1}(\tau) - \varrho^n(\tau)$ depends on n and τ. The addition of another variable when $n > m$ is, considering the increase $\varrho^{n+1}(\tau) - \varrho^n(\tau)$ of the common information, (approximately) equivalent to the addition of a "noise term" contributing only non-specific information related to noise and finite precision. (All these considerations can be "biased" by an actual amount of noise in the studied data, see Figs. 5 and 10.) Therefore for $n > m$ we expect that the curves $\varrho^n(\tau)$ as functions of τ have the same shape, they are just shifted. (I.e., $\varrho^{n+1}(\tau) \approx \varrho^n(\tau)+$ const.) Thus the limit behaviour of $\varrho^n(\tau)$ for $n \to \infty$ in the case of an m-dimensional dynamical system is attained for very small n, actually for $n = m+1, m+2, \ldots$.

Let us consider that the probability distribution $p(x_1, \ldots, x_n)$ used in the estimation of $\varrho^n(\tau)$ corresponds to a generating partition of the studied m-dimensional dynamical system for a certain range of τ, then the limit behaviour (5) of $\varrho^n(\tau)$, i.e., $\varrho^n(\tau) \approx A - \tau h(T_1)$ is attained for $n = m+1, m+2, \ldots$. And this is actually the behaviour of $\varrho^n(\tau)$ for low-dimensional dynamical systems. The range of τ for which marginal redundancies approach the linearly decreasing function is usually bounded by some τ_1 and τ_2, i.e., the equality $\varrho^n(\tau) \approx A - \tau h(T_1)$ holds for τ: $\tau_1 < \tau < \tau_2$. Before considering what determines these bounds we need to explain the technique for estimating the redundancies.

5.1 Redundancy algorithm

Practical computation of mutual information and redundancies of continuous variables is always connected with the problem of quantization. By the quantization we understand a definition of finite-size boxes covering the state space. The probability distribution is then estimated as relative frequencies of the occurrence of data samples in particular boxes. A naive approach to estimate the redundancies of continuous variables would be the use of the finest possible quantization given, e.g., by a computer memory or measurement precision. We must remember, however, that we usually have a finite number N of data samples. Hence, using a quantization that is too fine, the estimation of entropies and redundancies can be heavily biased: Estimating the joint entropy of n variables using q marginal bins one obtains q^n boxes covering the state space. If q^n approaches the number N of data samples, or even $q^n > N$, the estimate of $H(X_1, \ldots, X_n)$ can be equal to $\log(N)$, or, in any case, it can be determined more by a number of data samples and/or by a number of distinct data values than by a structure in the data, i.e., by properties of the system under study. We say, in such a case, that the data are overquantizied. (We will see that even a "natural" quantization of experimental data given by an A/D converter is usually too fine for reliable estimation of the redundancies.)

Emergence of overquantization is given by the number of boxes covering the state space, i.e., the higher the space dimension (the number of variables), the lower the number of marginal quantization levels that can cause the overquantization. Recalling Def. 3.4. of the redundancy of n variables, one can see that while the estimate of the joint entropy can be overquantizied, i.e., saturated on a value given by the number of the data samples and/or by the number of distinct data values, the estimates of the individual entropies are not and they increase with fining the quantization. Thus the overquantization causes an overestimation of redundancy $R^n(\tau)$ and obscures its dependence on τ.

Recalling $\varrho^n(\tau) = R^n(\tau) - R^{n-1}(\tau)$, one can see that the overquantization causes an overestimation of the marginal redundancy and, moreover, attenuation of its decrease with increasing τ. Further the overquantization can lead to a paradoxical unreal result of $\varrho^n(\tau)$ increasing with τ, which formally implies negative metric entropy (Figs. 1, 2, 3).

Therefore one must be very careful in defining the quantization. Fraser & Swinney [28] have proposed an algorithm for constructing locally data-dependent quantization (for details see [28]). In our computations we use a simple box-counting method with marginal equiquantization. It means that the marginal boxes are not defined equidistantly but so that there is approximately the same number of data points in each marginal bin. The choice of the number of bins is, however, crucial. In [29] we have proposed that computing R^n of n variables, the number of marginal bins should not exceed the $n+1$-st root of the number of the data samples. By extensive numerical experimentation we have found that this was the strongest rule necessary for preventing the overquantization and there are special cases when finer quantizations give unbiased results. Actually, it depends on the "level of chaoticity" (measured, e.g., by the metric entropy) or the level of mixing of the system under study. The weaker the mixing (the lower "chaoticity") the finer the quantization can be. (Cf. the results for the Lorenz and Rössler systems – see Sec. 6 and Figs. 1, 2, 4). Probably no general rule exists for determining ideal quantization for arbitrary data. Therefore we propose to compute redundancies for several numbers q of (equi)quantization levels around the recommended value $q = \sqrt[n+1]{N}$.

Defining the quantization boxes, we construct a partition of the experimental state space for which we estimate the probability distribution. Above we conjectured that this partition corresponds to a partition in the original state space $\mathcal{S}$ of the m.p.t. which generated the data. For simplicity, in the following considerations we will not distinguish these partitions.

5.2 A region of linear $\varrho^n(\tau)$ decrease

Now we can return to the problem of the bounds τ_1, τ_2 determining the region $\tau_1 < \tau < \tau_2$ in which $\varrho^n(\tau) \approx A - \tau h(T_1)$. Let us consider there are two m.p.t.'s R and L on a measure space $(\mathcal{S}, \mathcal{B}, \mu)$ and let ρ and λ be the "most coarse" generators of R and L, respectively. (Clearly, each refinement of a generator is a generator.) Let L be "more chaotic" than R, i.e. we mean $h(L) > h(R)$. Then we can intuitively say that the sequence

$$\bigcup_{i=1}^{\infty} R^{-i}\rho$$

is changing (refining) less, or more slowly than the sequence

$$\bigcup_{i=1}^{\infty} L^{-i}\lambda$$

and that L is able to generate $\mathcal{B}$, i.e.,

$$\bigcup_{i=1}^{\infty} L^{-i}\lambda = \mathcal{B} \text{ up to sets of zero measure,}$$

from more coarse partitions than R, i.e., $\lambda \leq \rho$, or even λ is not a generator with respect to R. Hence in computation of $\varrho^n(\tau)$ from series generated by R, for attaining the limit behaviour (5) we need a finer partition (quantization) than for series generated by L. (On the other hand, as we stated above, there is a greater "tolerance" to overquantization for R than for L.)

As an example for these considerations there are the Lorenz and Rössler systems, where $h(L) = 1.31$ bits/time unit, $h(R) = 0.13$ bit/time unit, and generating partition for L emerge from $q = 16$, and for R from $q = 48$ (marginal quantization bins), see Sec. 6, Figs. 1, 2 and 4.

Futhermore, consider a time-continuous transformation (dynamical system) T_τ. Let $\tau > 0$, according to Theorem 2.7. the equality $h(T_\tau) = \tau h(T_1)$ holds. For any $\tau < \tau_1$ we have

$$h(T_\tau) < h(T_{\tau_1}),$$

and, in analogy with the above transformations R and L, if ξ is a generating partition for T_{τ_1}, it need not be generating for T_τ, $\tau < \tau_1$. Hence the limit behaviour (5) of $\varrho^n(\tau)$ for a chosen partition ξ can start from some $\tau_1 > 0$, which decreases with refining the partition ξ.

For the first approach to the right boundary τ_2 let us recall Theorem 2.2. It states that for an ergodic m.p.t. T, the sequence of sets $T^{-i}a$, $a \in \xi$, becomes asymptotically and "on average" independent of any set from ξ. We can go further: For an ergodic transformation, weak mixing is a generic property [15] and considering Theorem 2.3. we can say, providing we neglect a few instants of time, that strong mixing is a generic property of ergodic systems. Then according to Theorems 2.2., 2.3., the sequence $T^{-i}a$, for $i \to \infty$, is independent of any set from ξ. This means that $\varrho^n(\tau) = 0$ for τ large enough and for any partition used. In practice, however, the right boundary τ_2 of the region of the linear decrease of $\varrho^n(\tau)$ is smaller than the lags for which $\varrho^n(\tau)$ vanishes.

Let us recall that $h(T,\xi) \leq H(\xi)$ and for a generating partition ξ also $h(T) \leq H(\xi)$. Hence if ξ is generating for T_{τ_1}, it can be generating only for a bounded interval of τ's for which the inequality $\tau h(T_1) < H(\xi)$ holds. In the case $h(T_{\tau_2}) = H(\xi)$, T_{τ_2} behaves, with respect to the partition ξ, like a Bernoulli (IID) process [14], [15] and $\varrho^n(\tau_2)$ should vanish. In fact, however, when τ is approaching τ_2 ($h(T_{\tau_2}) = H(\xi)$), the partition ξ looses its generating property and the linear decrease of $\varrho^n(\tau)$ stops for $\tau < \tau_2$. Then $\varrho^n(\tau)$ either decreases slowly to or stops on some "numerical zero" value. Clearly, for partition η, $\eta > \xi$, there is a right bound τ_3, $\tau_3 > \tau_2$, because $H(\eta) > H(\xi)$ and the region of linear $\varrho^n(\tau)$ decrease extends.

This phenomenon is illustrated in Fig. 3. There are the redundancies $\varrho^5(\tau)$ of the Lorenz system data for increasing number q of the marginal quantization levels ($q = 4$, 6, 8, 10, 16, 20, 32, reading from the bottom to the top), i.e., for fining the partition. The increasing straight line is the graph of $h(L_\tau) = \tau h(L_1) = 1.31\tau$, i.e., it gives the value of the metric entropy of L_τ against τ. Clearly, $\varrho^n(\tau = 0) = H(\xi)$ (or $H(q)$, i.e., the entropy of the partition). We can see that the linear region $\varrho^5(\tau) \approx A - \tau h(L_1)$ extends with fining the partition, slightly to the left and more apparently to the right, but it always ends before τ_2, $h(L_{\tau_2}) = H(\xi)$. For $q = 32$ overquantization effects emerge.

The above phenomenon can be interpreted in terms of uncertainty of prediction of states of a system with metric entropy $h(T_1)$. Suppose we know the state of the system at time $\tau = 0$ with a precision given by a partition ξ (i.e., we know in which a, $a \in \xi$, the system state lies). The prediction uncertainty of a system state increases with time as $h(T_\tau) = \tau h(T_1)$. The maximum uncertainty which can be reached is given by $H(\xi)$, i.e., the entropy of the partition ξ. This is the prediction uncertainty of a system state without knowledge of previous system states. Thus the time τ_2 for which $h(T_{\tau_2}) = H(\xi)$ is the time of the total loss of system memory as observed with precision given by the partition ξ.

6. Numerical Results

Figure 1 illustrates the typical time-lag dependence of marginal redundancies $\varrho^n(\tau)$ computed from time series generated by the x-component of the Lorenz system (in the Appendix referred to as "Lorenz a"), for four different quantization levels: a) $q = 4$ is the quantization insufficient for the limit behaviour (5) to appear, b) the quantization $q = 16$ defines the partition which is generating for the range of τ: $0.3 < \tau < 1.4$, which is sufficient for reliable estimation of the metric entropy. In the case c) $q = 40$ the linear region (5) for $\varrho^4(\tau)$ is extended to $\tau = 1.9$, but the results for $\varrho^5(\tau)$ are biased due to overquantization. With $q = 64$ (d) the overquantization distorted both $\varrho^4(\tau)$ and $\varrho^5(\tau)$.

Figure 2 depicts $\varrho^n(\tau)$ computed in the same conditions as those in Fig. 1, but from the data generated by the Rössler system. One can see that for this "less chaotic" system both the quantizations $q = 4$ (a) and $q = 16$ (b) are not fine enough for $\varrho^n(\tau)$ to approach the behaviour (5). Using quantizations $q = 40$ (c) and $q = 64$ (d) the correct value of the metric entropy (0.13 bit/time unit) can be estimated from $\varrho^4(\tau)$, while $\varrho^5(\tau)$ suffers from the overquantization, esp. for $q = 64$ (d).

Figure 3 illustrates dependence of the region of the linear $\varrho^n(\tau)$ decrease for $\varrho^5(\tau)$ of the Lorenz system on the number of the marginal quantization levels. This figure was discussed above (Sec. 5.2).

As one can see, the objective assessment of the region of linear $\varrho^n(\tau)$ decrease (5) for a particular quantization (and a particular data) is not simple and straightforward. Therefore we use the following strategy: Looking at the redundancies $\varrho^n(\tau)$ computed with various numbers of quantization levels we define a reasonable range of τ and using this range for all the quantizations we estimate the metric

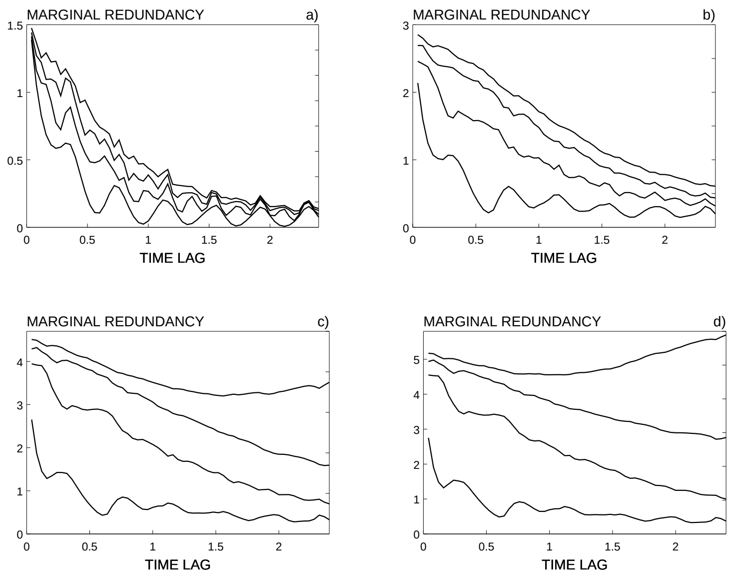

Fig. 1 *Time lag τ plots of marginal redundancies $\varrho^n(\tau)$ for the Lorenz system ("Lorenz a" - see Appendix) computed with different numbers q of marginal (equi)quantization levels:* **a)** $q = 4$, **b)** $q = 16$, **c)** $q = 40$, **d)** $q = 64$. *Four different curves in each figure represent different numbers n of lagged series, n = 2, 3, 4 and 5, reading from the bottom to the top. Redundancies are in bits and time lags in units of the system "time" variable ("time units").*

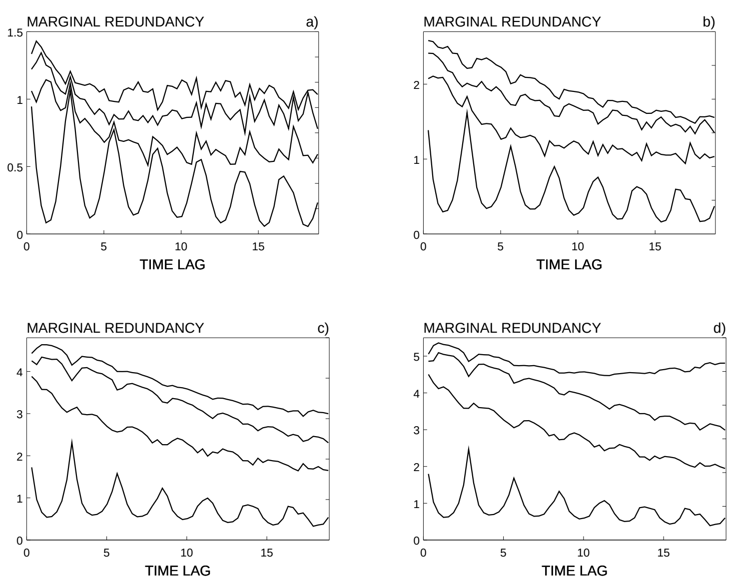

Fig. 2 *The same τ-plots of marginal redundancies as in Fig. 1 but for the Rössler system.*

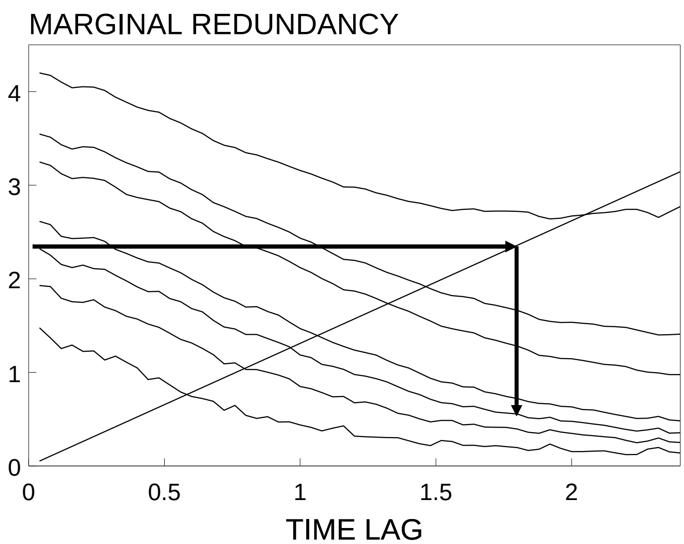

Fig. 3 *Time lag plots of the marginal redundancy $\varrho^5(\tau)$ for the Lorenz system, computed with various numbers q of marginal quantization levels, q = 4, 6, 8, 10, 16, 20 and 32, reading from the bottom to the top. The increasing straight line represents dependence of values of the metric entropy of the Lorenz system sampled with the lag τ against this lag. The thick lines with arrows indicate how to find the lag τ for which the metric entropy of the system (sampled with this lag) is equal to the partition entropy for particular partition (quantization). Note that the region of the linear decrease of $\varrho^n(\tau)$ always finishes before this point.*

entropy by a robust linear regression [30]. These estimates are typically underestimated for lower numbers q of the quantization levels and increase with increasing q, and then, when a partition given by a particular q becomes generating for a sufficient range of τ, the estimates of the metric entropy approximately saturate on the correct value. By further increasing q, the overquantization can occur leading to the decrease of the entropy estimates. (See discussion in Sec. 5.1.) Clearly, with an insufficient amount of data, the overquantization can occur before the partition is generating for a sufficient range of τ (or for a studied system in general), i.e., fining the partition, the estimates of the metric entropy increase to a maximum which is lower than the correct value of the system metric entropy, and after emerging of the overquantization they begin to decrease. Examples of these types of behaviour are given in Fig. 4. Figure 4 a) illustrates the estimation of the metric entropy of the Lorenz (a) system. Estimates obtained from 1,024,000 data samples (the upper line, the particular values denoted by the asterisks) saturate from $q = 16$ on the value of about 1.32 bits per time unit. For $q > 40$ the estimates decrease due to overquantization. The estimates obtained from 102,400 data samples (the lower line, the particular values denoted by the crosses) reach the maximum value 1.15 at $q = 12$ and then decrease, again due to overquantization. These results were obtained from $\varrho^4(\tau)$, for $n = 5$ the overquantization occurred earlier.

Figure 4 b) presents the same estimation results as Fig. 4 a), but for the Rössler system. Recalling the above discussion (Sec. 5.2) considering "less" and "more" chaotic systems, one can compare the numbers q of the quantization levels, i.e., the "fineness" of the "most coarse" partitions which are generating with respect to the Lorenz ($q = 16$) and the Rössler ($q = 48$) systems and how these systems are vulnerable to overquantization – the Lorenz data are overquantizied from $q = 40$, while the Rössler data from $q = 80$, considering the same number 1,024,000 of data samples.

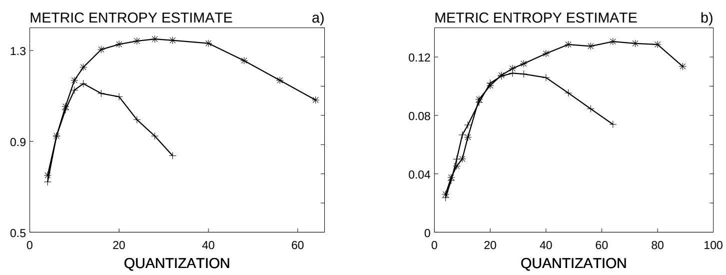

Fig. 4 *Dependence of the estimate of the metric entropy of a) the Lorenz system ("Lorenz a") and b) the Rössler system on the number of marginal quantization levels. The number of the data samples used in the estimation was 1,024,000 (the upper line with the asterisks for the estimated values) and 102,400 (the lower line with the crosses for the estimated values). Note that the scales are different. The metric entropy is in bits per time unit.*

Using the well-known theorem of Pesin [31], which states that the metric entropy of a dynamical system is equal to the sum of its positive Lyapunov exponents, we can assess the validity of this method for estimating the metric entropy by comparing our values (obtained using 1,024,000 samples) with the values of the positive Lyapunov exponent published for the systems examined here. (See Appendix for definition of the systems.) The results follow (the first number is our estimate of the metric entropy with its standard deviation, the second number is the positive Lyapunov exponent (both in bits per time unit) with related reference):
Lorenz a: 1.32 ± 0.04; 1.31 [32],
Lorenz b: 2.20 ± 0.07; 2.16 [33],
Rössler: 0.129 ± 0.002; 0.13 [33].
These results support the assertion (5).

It was found above that a large amount of data was necessary to obtain a reliable estimate of the metric entropy from $\varrho^n(\tau)$. In experimental practice such data amounts cannot usually be recorded. Also in such cases, however, estimations of the redundancies $\varrho^n(\tau)$ can be useful in two important tasks:

- "Relative quantification" of studied system(s) by using so called coarse-grained entropy rates [51], which are not suitable for estimating the metric entropy, but provide a classification of different systems (system states) equivalent to the classification based on the metric entropy.

- Qualitative characterization of data (systems) under study, namely applications in nonlinearity tests [49], [48], and in the assessment of the possible existence of low-dimensional chaos by searching for a region of linear decrease of $\varrho^n(\tau)$ – see Refs. [47, 52] and the next section.

7. Qualitative Characterization of Experimental Time Series

Analyzing experimental data, before their quantification the question of their qualitative character should be solved, i.e., one should ask "Is it reasonable to hypothesize that the data were generated by a low-dimensional chaotic dynamical system?" A number of papers of the type "Evidence of the existence of chaos by obtaining a finite value of the correlation dimension" have been published during the last decade. As far as we know any strict evidence of chaos which can be obtained from experimental data is not possible. All the typical properties of chaotic systems reflected in time series, generated by these systems, are necessary, not sufficient conditions for chaos. Hence we cannot speak about the evidence of chaos, but only about signatures of chaos which are detectable in experimental data.

The correlation (or other type of) dimension represents a geometrical signature of chaos, i.e., it characterizes the geometry (of the system attractor) but not the dynamics. Therefore extracting from the data only the geometric signature, one can erroneously consider some special type of stochastic dynamics, formally exhibiting finite correlation dimension [36], to be chaotic. To prevent this mistake dynamical signatures like the metric entropy and/or Lyapunov exponents should be applied.

We have done a number of extensive numerical studies of the possibility to correctly detect chaotic dynamics from time series. Except for the redundancy method, discussed here, we used the Cohen-Procaccia algorithm for estimating the metric entropy [4], and for estimating the Lyapunov exponents we applied an algorithm based on the higher-order fitting of the Jacobian matrices [39], [40]. It was reported that these and similar methods [4], [5], [41], [42] were able to estimate the correct values of the metric entropy/Lyapunov exponents from significantly smaller amounts of data than necessary for the above redundancy method. We have found, however, that without any a priori knowledge about a system underlying data, these methods can give even qualitatively incorrect results, i.e., positive value of the metric entropy or Lyapunov exponent for non-chaotic data. This cannot happen estimating the redundancies $\varrho^n(\tau)$ even from limited amounts of data. The qualitative character of an underlying system can be assessed by a simple look at the time-lag traces of the marginal redundancies $\varrho^n(\tau)$. For the illustration we present results obtained from the following data (in addition to the above Lorenz and Rössler systems):

1. Torus (two-periodic) time series without and with additive uniformly distributed noise (Figs. 5 a and 5 b, respectively).

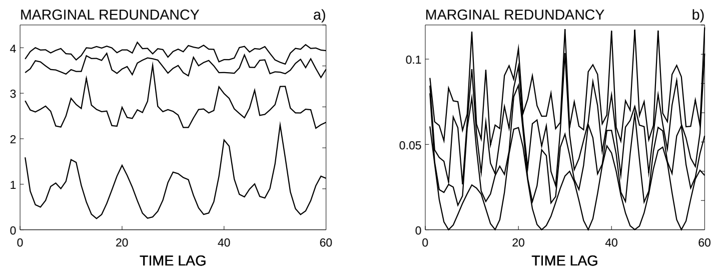

Fig. 5 *Marginal redundancy $\varrho^n(\tau)$, n = 2-5, for* **a)** *torus (2-periodic) time series and* **b)** *the same torus series jammed by 50% of additive uniformly distributed noise. Time lag is in number of samples and redundancies are in bits. Note that the scales on the ordinates are different.*

2. Time series generated by the double-Monod system [43] in periodic (Fig. 6 a), quasiperiodic (Fig. 6 b) and chaotic (Fig. 7) states (for details see Appendix).

3. Time series generated by dynamics on a "strange nonchaotic attractor" (Fig. 8), introduced by Grebogi et al. [44]. (By "strange" the authors describe fractal geometry of the attractor, "nonchaotic" means that this system has one zero and two negative Lyapunov exponents, i.e., zero metric entropy.)

4. Gaussian random processes with the same spectra as the Lorenz and Rössler systems investigated above, obtained by forward FFT (fast Fourier transform [30]) of the original data, followed by randomization of phases and backward FFT into the time domain (Fig. 9).

Time-lag plots of the marginal redundancies $\varrho^n(\tau)$ for these time series are presented in Figs. 5–9. We can see that for all the non-chaotic systems there is no tendency to the linearly decreasing trend of $\varrho^n(\tau)$, i.e., no specific "production of information" such as in chaotic systems. It means that the estimate of the metric entropy is clearly zero and there is no dynamical signature of chaos. This holds also for a random process with the same spectrum as the Rössler system (Fig. 9 b), representing highly correlated "coloured" noise. In the case of the random process with the same spectrum as the Lorenz system (Fig. 9 a) the redundancies decrease quickly to zero by an exponential or power-law way. On the other hand, in the chaotic state of the double-Monod system the region of linear decrease of $\varrho^n(\tau)$ for $n = 4, 5$ is apparent, i.e., the metric entropy of this system is clearly positive (Fig. 7). We can see that this "redundancy analysis" is useful for qualitative characterization of experimental time series. (See also [29], [47].)

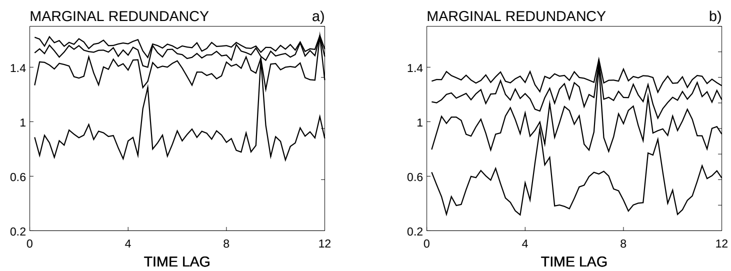

Fig. 6 *Marginal redundancy* $\varrho^n(\tau)$*, n = 2-5, for* **a)** *periodic and* **b)** *quasiperiodic state of the double Monod system. Time lag is in the system "time" variable and the redundancies are in bits.*

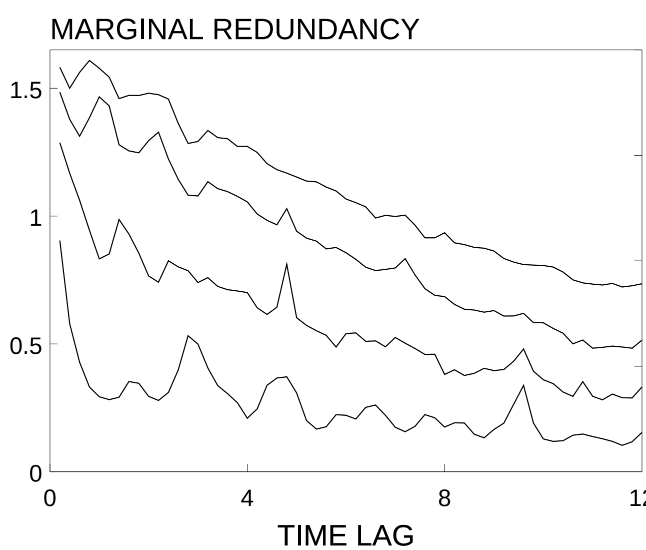

Fig. 7 *Marginal redundancy* $\varrho^n(\tau)$*,* $n = 2 - 5$*, for the chaotic state of the double Monod system. The time lag is in the system "time" and the redundancies are in bits.*

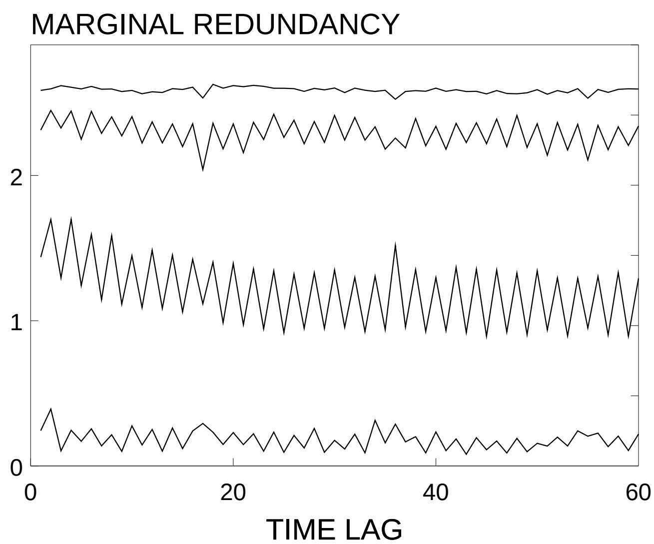

Fig. 8 *Marginal redundancy $\varrho^n(\tau)$, $n = 2 - 5$, for the "strange non-chaotic attractor" (see Appendix). The time lag is in the number of samples and the redundancies are in bits.*

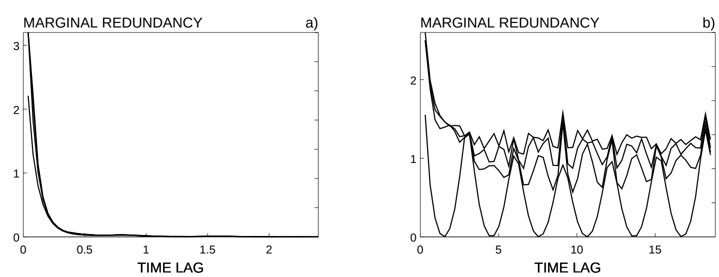

Fig. 9 *Marginal redundancy $\varrho^n(\tau)$, $n = 2 - 5$ for the stochastic processes with the same spectra as* **a)** *the Lorenz system and* **b)** *the Rössler system. The time lag is in the system "time" and the redundancies are in bits.*

8. Influence of Noise

The influence of additive noise contaminating chaotic time series on the estimation of the metric entropy from $\varrho^n(\tau)$ and on the character of the time-lag dependence of the marginal redundancy in general, has been studied by adding various portions of uniformly distributed noise to chaotic time series. Two typical examples of the marginal redundancies $\varrho^n(\tau)$ for the noisy Lorenz data are presented in Fig. 10. In general, additive noise induces a decrease of magnitudes of the redundancies. (See also Fig. 5, note that the scales are different.) The noise also induces a decrease of the estimates of the metric entropy. At the first sight, this is a paradoxical result, but it can be explained again in terms of prediction uncertainty of system states. As we have stated above, in any experimental/numerical situation there is a maximum uncertainty (given by the partition entropy) which can be reached by the "production of information" during the evolution of a chaotic system. In noisy data there is some basic uncertainty level which is higher than in noise-free data (i.e., the redundancies are lower). And this causes a decrease of the effective rate of the dynamical increase of the uncertainty. This holds, however, only with some moderate amount of noise, when the behaviour (5) of $\varrho^n(\tau)$ can be found (Fig. 10 a, 10% of noise). Starting from some critical amount of noise (which is individual for different chaotic systems) no "typically chaotic" behaviour (5) of $\varrho^n(\tau)$ can be detected. The marginal redundancy $\varrho^n(\tau)$ decreases by an exponential or power-law way (Fig. 10 b, 25% of noise), as in the case of an ergodic stochastic process. The metric entropy cannot be estimated in this case, neither such noisy chaotic data can be recognized from stochastic data. (At least by using this redundancy method.) Hence we can conclude that the production of information, the typical property of chaotic dynamical systems, can be detected from noisy data only up to a certain level of noisiness of the data. Over this level of noise the chaotic properties of the system are obscured by noise and the system cannot be recognized from a stochastic process.

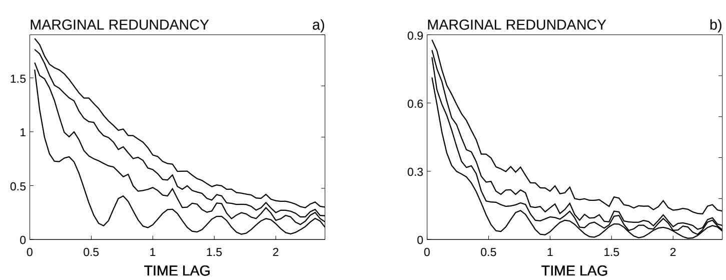

Fig. 10 *Marginal redundancy $\varrho^n(\tau)$, n = 2-5, for the Lorenz time series contaminated by* **a)** *10% and* **b)** *25% of additive uniformly distributed noise. The time lag is in the system "time" and the redundancies are in bits. Note that scales on ordinates are different.*

9. Conclusion

A method for analysis of experimental time series suitable for identification and quantification of underlying chaotic dynamics has been presented. It is based on examination of time-lag dependence of marginal redundancy, the information-theoretic functional computed from studied time series and its lagged versions. The theoretic approach, consisting of a translation of properties of chaotic dynamical systems from the language of ergodic theory to that of information theory, is followed by an extensive numerical study supporting the assertions derived theoretically. It has been demonstrated that this method is able to discern chaotic dynamics from (noisy) (quasi-)periodic or stochastic processes and provides a quantitative characterization of chaotic systems by measuring their information production rates in terms of the metric (Komogorov-Sinai) entropy.

The technique presented above, unlike the entropy algorithms proposed in [3], [4], [5], is not related to any dimensional algorithm and is also independent of any method for extracting the Lyapunov exponents. Moreover, it is based on the quantification of a macroscopic property – the information production rate, while the other mentioned methods are related to microscopic properties of systems under study and can be heavily biased by finite precision and measurement noise. Therefore we believe that this "redundancy analysis" can play an important role in the analysis of dynamical data.

Acknowledgements

The author would like to thank V. Albrecht, I. Dvořák and A. Fraser for stimulating discussions. The support from the Grant Agency of the Czech Republic (grant No. 205/97/0921) is gratefully acknowledged.

This paper contains a part of the author's CSc (PhD) thesis "Entropy and Information Properties of Chaotic Systems and EEG", Institute of Computer Science, Czechoslovak Academy of Sciences, Prague, 1992.

Appendix

The time series related to the systems of ordinary differential equations were generated by the numerical integration based on the Bulirsch-Stoer method [30], p.563, of the Rössler system [45]:

$$(dx/dt, dy/dt, dz/dt) = (-z - y, x + 0.15y, 0.2 + z(x - 10)),$$

with the integration step 0.314 and accuracy 0.0001;
the Lorenz system [46] – "Lorenz a:"

$$(dx/dt, dy/dt, dz/dt) = (10(y - x), x(28 - z) - y, xy - 8z/3),$$

with the integration step 0.04 and accuracy 0.0001;
"Lorenz b:"

$$(dx/dt, dy/dt, dz/dt) = (16(y-x), x(45.92-z)-y, xy-4z),$$

with the integration step 0.02 and accuracy 0.0001;
the double Monod system [43]:

$$(dx/dt, dy/dt, dz/dt) = (1+\epsilon\sin(\omega t)-x-A, A-y-B, B-z),$$

where $A = 5xy/((8/115)+x)$, $B = 2yz/((9/46)+y)$ and
$\epsilon = 0$ in the periodic state,
$\epsilon = 3/5$, $\omega = 4\pi$ in the quasiperiodic state, and
$\epsilon = 3/5$, $\omega = 5\pi/6$ in the chaotic state,
always with the integration step 0.2 and accuracy 0.0001. In all cases, the component x was recorded.

The time series from the "strange nonchaotic attractor" [44] was obtained by iterating the system:

$$\Theta_{n+1} = (\Theta_n + 2\pi\omega) mod(2\pi)$$

$$u_{n+1} = \Lambda(u_n\cos(\Theta) + v_n\sin(\Theta))$$

$$v_{n+1} = -0.5\Lambda(u_n\cos(\Theta) - v_n\cos(\Theta))$$

where $\omega = (\sqrt{5}-1)/2$ and $\Lambda = 2/(1+u_n^2+v_n^2)$. Component Θ was recorded.

Two-periodic noisy data were generated according to the following formula:

$$Y(t) = (R_1 + R_2\sin(\omega_2 t+\phi))\sin(\omega_1 t) + \xi,$$

where $R_1 : R_2 = 5 : 4$, $\omega_1 : \omega_2 = 10 : 9$, $\phi = 1.3\pi$, ξ are random numbers uniformly distributed between $-\Xi$ and Ξ. The term "50% of noise" means that $R_1 : R_2 : \Xi = 5 : 4 : 9$. For the torus series without noise the same formula with $\xi = 0$ holds.

References

[1] Mayer-Kress G., ed.: Dimensions and entropies in chaotic systems. Springer, Berlin, 1986.

[2] Abraham N.B., Albano A.M., Passamante A., Rapp P.E., eds.: Measures of Complexity and Chaos. Plenum Press, New York, 1989.

[3] Grassberger P., Procaccia I.: Phys. Rev. A **28**, 1983, 2591.

[4] Cohen A., Procaccia I.: Phys. Rev. A **31**, 1985, 1872.

[5] Termonia Y., Phys. Rev. A **29**, 1984, 1612.

[6] Renyi A.: In: Trans. 2nd Prague Conf. on Information Theory, Statistical Decision Function, Random Processes. Academia, Prague, 1959, p.549.

[7] Kolmogorov A.N.: Dokl. Akad. Nauk SSSR 124, 1959, 754.

[8] Sinai Ya.G.: Dokl. Akad. Nauk SSSR 124, 1959, 768.

[9] Shaw R.S.: Naturforsch. 36A, 1981, 80.

[10] Shaw R.S.: The dripping faucet as a model chaotic system. Aerial, Santa Cruz, 1985.

[11] Fraser A.M.: IEEE Trans. Information Theory **35**, 1989, 245.

[12] Fraser A.M.: In: [1], p.82.

[13] Fraser A.M.: In: [2], p.117.

[14] Petersen K.: Ergodic theory. Cambridge University Press, Cambridge, 1983.

[15] Walters P.: An introduction to ergodic theory. Springer, Berlin, 1982.

[16] Sinai Ya. G.: Introduction to ergodic theory. Princeton University Press, Princeton, 1976.

[17] Cornfeld I.P., Fomin S.V., Sinai Ya. G.: Ergodic theory. Springer, New York, 1982.

[18] Shannon C.E., Weaver W.: The mathematical theory of communication. University of Illinois Press, Urbana, 1964.

[19] Gallager R.G.: Information theory and reliable communication. Wiley, New York, 1968.

[20] Khinchin A.I.: Mathematical foundations of information theory. Dover Publications, New York, 1957.

[21] Billingsley P.: Ergodic theory and information. Wiley, New York, 1965.

[22] Kullback S.: Information theory and statistics. J. Wiley, New York, 1959.

[23] Eckmann J.P., Ruelle D.: Rev. Mod. Phys. **57**, 1985, 617.

[24] Martin N.F.G., England J.W.: Mathematical theory of entropy. Addison-Wesley, London, 1981, (in Russian: Mir, Moscow, 1988).

[25] Kamke E.: Differentialgleichungen Lösungmethoden und Lösungen, Leipzig, 1959, (in Russian: Nauka, Moscow, 1971).

[26] Arnold A.I.: Ordinary differential equations. MIT Press, Cambridge, 1073.

[27] Takens F.: In: D.A. Rand, D.S. Young, eds., Dynamical Systems and Turbulence (Warwick 1980), Lecture Notes in Mathematics 898. Springer, Berlin, 1981, p.365.

[28] Fraser A.M., Swinney H.L.: Phys. Rev. A **33**, 1986, 1134.

[29] Paluš M., Albrecht V., Dvořák I.: Information theoretic test for nonlinearity in time series. Phys. Lett. A **175**, 1993, 203-209.

[30] Press W.H., Flannery B.P., Teukolsky S.A., Vetterling W.T.: Numerical recipes: the art of scientific computing. Cambridge Univ. Press, Cambridge, Mass., 1986.

[31] Pesin Ya. B.: Russian Math. Surveys, (4), **32**, 1977, 55. Also in: Ya.G. Sinai (ed.): Dynamical Systems. Collection of Papers. (World Scientific, Singapore, 1991).

[32] Aizawa Y.: Progr. Theor. Phys. **68**, 1982, 64.

[33] Wolf A., Swift J.B, Swinney H.L., Vastano J.A.: Physica D **16**, 1985, 285.

[34] Grassberger P., Procaccia I.: Phys. Rev. Lett. **50**, 1983, 346.

[35] Grassberger P., Procaccia I.: Physica D **9**, 1983, 189.

[36] Osborne A.R., Provenzale A.: Physica D **35**, 1989, 357.

[37] Paluš M., Dvořák I., David I.: In: I. Dvořák, A.V. Holden, (eds.), Mathematical approaches to brain functioning diagnostics. Manchester University Press, Manchester and New York, 1991, p. 369.

[38] Takens F.: In: G.I. Barenblatt, G. Ioos, D.D. Joseph (eds.): Nonlinear dynamics and turbulence. Pittman, 1982.

[39] Briggs K.: Phys. Lett. A **151**, 1990, 27.

[40] Bryant P., Brown R., Abarbanel H.D.I.: Phys. Rev. Lett. **65**, 1990, 1523.

[41] Stoop R., Parisi J.: Physica D **50**, 1991, 89.

[42] Zeng X., Eykholt R., Pielke R.A.: Phys. Rev. Lett. **66**, 1991, 3229.

[43] Kot M., Sayler G.S., Schultz T.W.: Bull. Math. Biol. **54**, 4, 1992, 619.

[44] Grebogi C., Ott E., Pelikan S., Yorke J.A.: Physica D **13**, 1984, 261.

[45] Rössler O.E.: Phys. Lett. A **57**, 1976, 397.

[46] Lorenz E.N.: J. Atmos. Sci. **20**, 1963, 130.

[47] Paluš M.: Identifying and quantifying chaos by using information-theoretic functionals. In: Weigend A.S., Gershenfeld N.A. (eds.): Time Series Prediction: Forecasting the Future and Understanding the Past. Santa Fe Institute Studies in the Sciences of Complexity, Proc. **XV**, Addison–Wesley, Reading, Mass., 1993, 387-413.

[48] Paluš M., Novotná D.: Testing for nonlinearity in weather records. Phys.Lett. A, **193**, 1994, 67-74.

[49] Paluš M.: Testing for nonlinearity using redundancies: Quantitative and qualitative aspects. Physica D **80**, 1995, 186-205.

[50] Paluš M.: Detecting nonlinearity in multivariate time series. Phys. Lett. A **213**, 1996a, 138-147.

[51] Paluš M.: Coarse-grained entropy rates for characterization of complex time series. Physica D **93**, 1996b, 64-77.

[52] Paluš M.: Nonlinearity in normal human EEG: cycles, temporal asymmetry, nonstationarity and randomness, not chaos. Biol. Cybern. **75**, 5, 1996c, 1996, 389-396.