Generalized redundancies for time series analysis

Dean Prichard[1,2a] and **James Theiler**[2,3]

[1] *Department of Physics, University of Alaska, Fairbanks, AK 99775*
[2a] *Complex Systems Group, Theoretical Division;* [2b] *Center for Nonlinear Studies; and*
[2c] *Astrophysics and Radiation Measurement Group, Nonproliferation and International Security Division, Los Alamos National Laboratory, Los Alamos, NM 87545*
[3] *Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501*

(Draft: February 17, 1995)

ABSTRACT

Extensions to various information-theoretic quantities (such as entropy, redundancy, and mutual information) are discussed in the context of their role in nonlinear time series analysis. We also discuss "linearized" versions of these quantities and their use as benchmarks in tests for nonlinearity. Many of these quantities can be expressed in terms of the generalized correlation integral, and this expression permits us to more clearly exhibit the relationships of these quantities to each other and to other commonly used nonlinear statistics (such as the BDS and Green-Savit statistics). Further, numerical estimation of these quantities is found to be more accurate and more efficient when the the correlation integral is employed in the computation. Finally, we consider several "local" versions of these quantities, including a local Kolmogorov-Sinai entropy, which gives an estimate of variability of the short-term predictability.

1 Introduction

In Shaw's influential (and prize-winning) essay, chaotic dynamical systems were interpreted as generators of information [1]. This information-theoretic viewpoint has led to a number of algorithms and statistics which characterize strange and chaotic attractors in terms of their information generating properties. Foremost among these is the Kolmogorov-Sinai (K-S) entropy [1, 2], which measures the average production of Shannon [3] entropy per unit time. Generalizations of this quantity have been described [4–6] which measure the production rate of Renyi [7] entropies.[1] Various measures of dependence have also been employed, such as mutual information [1, 8] (between two variables) and redundancy [9, 10] (among several variables). Another commonly used information-theoretic statistic is the information dimension [11] which describes how information in a strange attractor scales with characteristic coarse grain size.

These information-theoretic measures are not the only tools available for characterizing nonlinear systems. Measures for characterizing chaos are usually categorized as either "static" or "dynamical," though this distinction is actually somewhat artifical for time series applications, since time-delay embeddings [12] effectively encode dynamical information directly into the static attractor.

So-called static measures rely only on the invariant measure of the attractor: this includes various flavors of dimension [6,11,13–18] as well as a variety of statistics [19–24] based on the correlation integral of Grassberger and Procaccia [14]. A growing number of measures are being developed for characterizing some of the topological properties of the attractor [25–28]. Statistics which employ dynamical information directly include Lyapunov exponents [29–31], Kolmogorov-Sinai (K-S) entropy, and nonlinear prediction error [32–36]. There is as well a growing class of statistics which purport to measure determinism directly [37–42].

Many of these measures are related. For example, the Pesin identity relates Lyapunov exponents to Kolmogorov-Sinai entropy [2], and the Kaplan-Yorke conjecture relates Lyapunov exponents to information dimension [43]. Further, properties of the unstable periodic orbits, which form the basis of many of the topological methods, can also be used to estimate dimensions and entropies [44–48].[2]

The relationship between the Shannon [3] and Renyi [7] entropies and the generalized correlation integral [6] has been pointed out by a number of authors [49–51], and this relation extends to mutual information and redundancies as well. This relationship permits us to suggest information-theoretic analogues to a number of statistics [19,22,23]. The generalized correlation integral also provides a numerically efficient way to compute these quantities. Finally, by using the relationship between the Shannon entropy and the generalized correlation integral [6,17], it is possible to write down "local" versions of information-theoretic statistics; we will investigate the applicability of these new statistics for time series analysis.

In section 2, we review information-theoretic measures and methods for estimating them from data. Section 3 describes "linearized" versions of these measures; in general, the linearized statistic is the value that would be obtained if the time series were a linear (gaussian) process. Section 4 discusses the expression of these statistics in terms of the generalized cor-

[1] The Renyi entropies are parameterized by a real-valued q which can vary from $-\infty$ to ∞; the value $q = 1$ corresponds to the Shannon entropy.

[2] However, unlike the topological methods, these methods do not use the linking of the orbits.

relation integral ($C_q(r)$ with $q = 1$), and in section 5, we use the expressions of sections 3 and 4 to extend the information theoretic quantities to $q \neq 1$, and in particular to demonstrate the increased numerical efficiency that arises when computations are performed for the $q = 2$ statistics. In section 6, we describe some applications to time series data, and in section 7, we discuss how these statistics relate to others already in use by the time series community. The last section is a little speculative; it discusses "local" versions of a number of statistics.

2 Entropy, mutual information and redundancies (a brief review)

Many information-theoretic statistics can be expressed in terms of Shannon's original expression for the entropy [3] in a partition of probabilities $p_1, \ldots, p_M$ with $p_1 + \cdots + p_M = 1$; that is: $H = -\sum_{i=1}^{M} p_i \log p_i$. This value quantifies the average information gained from observing which of $i = 1, \ldots, M$ actually occurred. For a time series $x(t)$, with $t = 1, \ldots, N$, we can use the Shannon entropy to quantify the average information gained with each finite-precision measurement of $x(t)$. This is often implemented with a box counting approach. A partition size δ is chosen, corresponding to the accuracy of the measurement, and and the data x (we use x here as an abbreviation for $x(t)$, $t = 1, \ldots, N$) are discretized into integers $y = 1, \ldots, M$ depending on what bin of size δ they fall into. In this case, the Shannon entropy is:

$$H_1(x, \delta) = H_1(y) = -\sum_{y=1}^{M} p(y) \log_2[p(y)] \tag{1}$$

where $p(y)$ is the probability of being in the y^{th} bin. For two time series x_1 and x_2 one can define a joint entropy as:

$$H_1(x_1, x_2, \delta) = H_1(y_1, y_2) = -\sum_{y_1}\sum_{y_2} p(y_1, y_2) \log_2[p(y_1, y_2)] \tag{2}$$

and for m variables the entropy (often called block entropy) is:

$$H_1(x_1, \ldots, x_m, \delta) = H_1(y_1, \ldots, y_m) = -\sum_{y_1, \ldots, y_m} p(y_1, \ldots, y_m) \log_2[p(y_1, \ldots, y_m)]. \tag{3}$$

One can also define an entropy for continuous variables:

$$H_1(\vec{x}) = -\int p(\vec{x}) \log_2[p(\vec{x})] d\vec{x} \tag{4}$$

where $\vec{x} = (x_1, \ldots, x_m)$. However, this definition has some unusual properties, for one it depends on the coordinate system which is used. For example, if $\vec{z} = a\vec{x}$ then $H(\vec{z}) = H(\vec{x}) + \log(a)$, also $\lim_{\delta \to 0} H(\vec{x}, \delta) \neq H(\vec{x})$. The fact that the entropy depends on the partition size allows one to define an information dimension [11], based on the average scaling of the amount of information required to specify a point in the state space within an accuracy of δ. The information dimension is given by:

$$D_1 = \lim_{r \to 0} \frac{-H(\vec{x}, \delta)}{\log \delta}. \tag{5}$$

To quantify the dependence between time series, a commonly used measure is the mutual information

$$I_1(x_1; x_2, \delta) = H_1(x_1, \delta) + H_1(x_2, \delta) - H_1(x_1, x_2, \delta). \tag{6}$$

The mutual information is a measure of many bits one can predict about x_2 given a measurement of x_1 with an accuracy of δ. If x_1 and x_2 are independent the mutual information is zero, while if x_2 is completely dependent on x_1 then $I_1(x_1;x_2) = H_1(x_2)$. Also note that $I_1(x_1;x_2) = I_1(x_2;x_1)$ and $I_1(x_1,x_1) = H_1(x_1)$. The mutual information for continuous variables is coordinate independent, unlike the entropy, and assuming that there is a small amount of noise in the data $\lim_{\delta\to 0} I_1(x_1;x_2,\delta) = I_1(x_1;x_2)$. This is because for δ smaller than the noise scale, both $H_1(x_1,\delta)$ and $H_1(x_2,\delta)$ will scale as $-\log\delta$, while $H(x_1,x_2,\delta)$ will scale as $-2\log\delta$. For a noiseless deterministic system one can use the scaling of the mutual information with δ to define a "mutual information dimension" [52, 53].

The m dimensional extension of mutual information is called redundancy [9],

$$R_1(x_1;\ldots;x_m,\delta) = \sum_{i=1}^{m} H_1(x_i,\delta) - H_1(x_1,\ldots,x_m,\delta) \tag{7}$$

where $\vec{x}(t) = (x_1(t), x_2(t), \ldots, x_m(t))$ can be either a multivariate signal or a time delay embedding [12] $\vec{x}(t) = (x(t), x(t-\tau), \ldots, x(t-(m-1)\tau))$; for delay coordinates we have:

$$R_1(x_1;\ldots;x_m,\delta) = mH_1(x_1,\delta) - H_1(x_1,\ldots,x_m,\delta). \tag{8}$$

To quantify the amount of information about x_m contained in $x_1, x_2, \ldots, x_{m-1}$ a quantity called marginal redundancy (R_1') is used [9]:

$$R_1'(x_1,\ldots,x_{m-1};x_m,\delta) = R_1(x_1;\ldots;x_m,\delta) - R_1(x_1;\ldots;x_{m-1},\delta). \tag{9}$$

If x_m is independent of $x_1, x_2, \ldots, x_{m-1}$ then the marginal redundancy is zero, while if x_m is completely dependent on $x_1, x_2, \ldots, x_{m-1}$ then $R_1'(x_1,\ldots,x_{m-1};x_m) = H_1(x_m)$.

The redundancies and marginal redundancies, like the mutual information, are only scale independent for noisy systems. However the quantity

$$\lim_{m\to\infty}[H_1(x_1,\ldots,x_m,\delta) - H_1(x_1,\ldots,x_{m-1},\delta)] \tag{10}$$

has the opposite behavior, that is, for a deterministic system it is scale independent (both terms scale as $-D_1\log\delta$, for small δ, so the overall expression does not depend on δ), while for a noisy system the expression scales as $-\log\delta$. A related quantity is the Kolmogorov-Sinai (K-S) entropy, which is a measure of the mean rate of information creation by the system. Given a time delay embedding $\vec{x}(t) = (x_1(t), x_2(t), \ldots, x_m(t)) = (x(t), x(t-\tau), \ldots, x(t-(m-1)\tau)$, the K-S entropy is:

$$\begin{aligned} K_1 &= \lim_{\substack{m\to\infty \\ \delta\to 0}} [H_1(x_m|x_1,\ldots,x_{m-1},\delta)]/\tau \\ &= \lim_{\substack{m\to\infty \\ \delta\to 0}} [H_1(x_1,\ldots,x_m,\delta) - H_1(x_1,\ldots,x_{m-1},\delta)]/\tau. \end{aligned} \tag{11}$$

The K-S entropy can also be related to the marginal redundancy [9]; for small time delays τ we have,

$$\lim_{m\to\infty} R_1'(\tau) = H_1(x_1) - \tau K_1. \tag{12}$$

This suggests another way to estimate the K-S entropy is:

$$K_1 \approx \lim_{m\to\infty} \frac{R_1'(\tau = t_1) - R_1'(\tau = t_2)}{t_2 - t_1}. \tag{13}$$

That is, for a deterministic system, the K-S entropy can be obtained from the slope of the marginal redundancy R_1' versus time delay τ curves for large embedding dimensions. Further, if $x(t)$ is completely dependent on the previous values of $x(t-\tau)\ldots x(t-(m-1)\tau)$ then $K_1 = 0$.

3 Linear redundancies

Paluš *et al.* [10] define "linear redundancies" which are derived from the continuous versions of above formulas for the special case of a multivariate gaussian distribution

$$p(\vec{x}) = \frac{|\boldsymbol{\xi}|^{1/2}}{(2\pi)^{m/2}} e^{-\frac{1}{2}\sum_{i,j}^{m} \xi_{ij} x_i x_j} \tag{14}$$

where $|\cdot|$ is the determinant, and ξ_{ij} is an element of the matrix $\boldsymbol{\xi}$, which is the inverse of the covariance matrix $\boldsymbol{\Xi}$, with elements given by: $\Xi_{ij} = \langle x_i(t) x_j(t)\rangle$, where it is assumed that $\langle x_i(t)\rangle = 0$ Combining Eq. (14) and Eq. (4) we find that the "linearized" entropy is given by:

$$\begin{aligned} \mathcal{H}_1(\vec{x}) &= \log_2\left(\frac{(2\pi)^{m/2}}{|\boldsymbol{\xi}|^{1/2}}\right) + \frac{1}{2\log_e(2)} \int \left(\sum_{i,j}^{m} \xi_{ij} x_i x_j\right) p(\vec{x}) d\vec{x} \\ &= \frac{m}{2}\log_2(2\pi e) + \frac{1}{2}\log_2|\boldsymbol{\Xi}| \end{aligned} \tag{15}$$

and the linearized redundancy[3] is:

$$\mathcal{R}(\vec{x}) = \frac{1}{2}\sum_{i}^{m} \log_2(\Xi_{ii}) - \frac{1}{2}\log_2|\boldsymbol{\Xi}|. \tag{16}$$

This is equivalent to the form given in Paluš *et al.* [10] since for a symmetric matrix $|\boldsymbol{\Xi}| = \prod_j^m \lambda_j$ where λ_j are the eigenvalues of $\boldsymbol{\Xi}$. Paluš *et al.* also define what they call a marginal linear redundancy by:

$$\mathcal{R}'(x_1, x_2, \ldots, x_{m-1}; x_m) = \mathcal{R}(x_1; x_2; \ldots; x_m) - \mathcal{R}(x_1; x_2; \ldots; x_{m-1}). \tag{17}$$

One of the main reasons for computing these linearized statistics is that they provide a qualitative way to assess the role of linear correlations in the estimate of an actual information-theoretic quantity. Paluš [54] has advocated this approach as complimentary to the more formal methods of statistical hypothesis testing, such as the use of surrogate data [55].

[3] One can also define nonlinear statistics based on "local" versions of the linearized statistics.

4 Relation to C_1

The most straightforward way to estimate the quantities defined above is to use a box counting approach: the m dimensional space is divided into a number of boxes of size δ. By counting the number of points n_i in the i^{th} box, the probability can be estimated as $p_i \approx n_i/N$ where N is the total number of points.[4] A number of authors have used refinements to this procedure, by adapting the size of the boxes depending on the local density [8–10].

It has been shown by Liebert and Schuster [49] that $H_1(\vec{x}, r)$ can be related to $C_1(\vec{x}, r)$, the generalized correlation integral of order 1. Instead of estimating probabilities $p(\vec{x}, \delta)$ within boxes of size δ, one can calculate probabilities $P(\vec{x}, r)$ in regions of radius r about each point. The two are related by:

$$\sum_{\substack{i \\ \text{(bins)}}} p_i(\vec{x}, \delta) \log_2[p_i(\vec{x}, \delta)] = \frac{1}{N} \sum_{\substack{t \\ \text{(datapoints)}}} \log_2[p_{i(t)}(\vec{x}, \delta)] \approx \frac{1}{N} \sum_t \log_2[P_t(\vec{x}, r)] = \log_2 C_1(\vec{x}, r) \tag{18}$$

where $p_i(\vec{x}, \delta)$ is the probability of being in the the i^{th} bin (a box of diameter δ), $i(t)$ is the box that the t^{th} data point is in, and $P_t(\vec{x}, r)$ is the probability of being in the box of radius r (or diameter $\delta = 2r$) centered at the point $\vec{x}(t)$. Since the two boxes are the same size and very close to each other (they overlap), we can heuristically justify the relation $p_{i(t)}(\delta) \approx P_t(r)$ for $\delta = 2r$ (using maximum norm).[5]

A natural estimate of $P_t(r)$ is given by $B(\vec{x}(t), r)$, [6] which is the fraction of data points (excluding $\vec{x}(t)$ itself) within r of $\vec{x}(t)$. That is,

$$P_t(\vec{x}, r) \approx B(\vec{x}(t), r) = \frac{\mathrm{n}_t}{N} = \frac{1}{N} \sum_{\substack{s \\ s \neq t}} \Theta(r - \|\vec{x}(t) - \vec{x}(s)\|) \tag{19}$$

where Θ is the Heaviside function, n_t is the number of points within a radius r of $\vec{x}(t)$, and $\|\cdot\|$ is some measure of distance (we use maximum norm). The generalized correlation integral of order 1 is a geometric average of the $B(\vec{x}(t), r)$ probabilities, and is given by

$$\log C_1(\vec{x}, r) = \frac{1}{N} \sum_t \log B(\vec{x}(t), r). \tag{20}$$

Note that Eq. (19) is a simple form of kernel density estimation [58], often this would be written in the form:

$$P_t(\vec{x}, r) \approx \frac{1}{N} \sum_{\substack{s \\ s \neq t}} K\left(\frac{\|\vec{x}(t) - \vec{x}(s)\|}{r}\right) \tag{21}$$

where $K(z) = 1$ if $z < 1$ and otherwise is zero. This particular kernel is far from optimal (in fact, Silverman [58] calls it the "naive estimator"). However, even this crude form of kernel

[4] For small n_i Grassberger [56] has derived a correction to the formula, which for the Shannon entropy is $p_i \log_2[p_i] \approx (n_i/N)\left[\log_2(N) - \Psi(n_i) + (-1)^{n_i+1}/(n_i+1)\right]$ where $\Psi(x)$ is the digamma function (see also Wolpert and Wolf [57]).

[5] For maximum norm, a radius r corresponds to a box size (diameter) $\delta = 2r$. For the Euclidean norm, in an m dimensional space, a sphere of radius r has the same volume as a cube of diameter $\delta = c_m^{-1/m} r$, where c_m is the the volume of a m-dimensional unit sphere.

[6] Grassberger [56] has also derived a small n_t correction to this formula, $\log_2 P_t \approx \Psi(\mathrm{n}_t + 1) - \log_2(N) + (-1)^{\mathrm{n}_t+1}/(\mathrm{n}_t + 1)$, which is used in the calculations below.

density estimation is generally considered superior to using a multidimensional histogram (binning), especially for large embedding dimensions. Better results can often be obtained, if one uses a kernel $K(z)$ which decreases with increasing z or even a kernel whose width depends on the local density (see Ref. [58] for details). However, in this paper Eq. (19) will be used, so we can estimate the Shannon entropy from the generalized correlation integral of order one [6, 17, 49]

$$H_1(x_1; x_2; \ldots; x_m, \delta) \approx -\log C_1(x_1, x_2, \ldots, x_m, r). \tag{22}$$

We can now express the redundancies (Eqs. 7 and 9) and K-S entropy (Eqs. 11 and 13) in terms of C_1, and since kernel density estimation is being used instead of binning, we can expect more accurate results with limited data sets.

5 Generalized entropies and redundancies

Instead of using the Shannon entropy to calculate redundancies, we can generalize these statistics by using the Renyi entropies [7]:

$$H_q(\vec{x}, \delta) = \frac{1}{1-q} \log_2 \sum_i [p_i(\vec{x}, \delta)]^q. \tag{23}$$

It is easy to show that the limit as $q \to 1$ leads to the Shannon entropy. Again we can relate probabilities $p(\vec{x}, \delta)$ within boxes of size δ, to probabilities $P(\vec{x}, r)$ in regions of radius $r = \delta/2$ (for maximum norm) about each point:

$$\sum_i [p_i(\vec{x}, \delta)]^q \approx \frac{1}{N} \sum_t [P_t(\vec{x}, r)]^{q-1} \tag{24}$$

so we have[7],

$$H_q(\vec{x}, \delta) \approx \frac{1}{1-q} \log_2 \left(\frac{1}{N} \sum_t \left(\frac{1}{N} \sum_{\substack{s \\ s \neq t}} \Theta(r - \|\vec{x}(t) - \vec{x}(s)\|) \right)^{q-1} \right) = -\log_2 [C_q(\vec{x}, r)] \tag{25}$$

where $C_q(\vec{x}, r)$ is the generalized correlation integral [6, 17]. The idea of relating the Renyi entropies to the generalized correlation integral is by no means new. It was mentioned in the review by Grassberger *et al.* [50], the $q = 1$ case was used by Liebert and Schuster [49] and Buzug *et al.* [53], and Pompe [51] used the correlation integral to calculate H_2. However, by using Eq. (25) we can define generalized redundancies $R_q(\vec{x}, r)$ and $R'_q(\vec{x}, r)$ in terms of $C_q(\vec{x}, r)$. Pompe [51] has also proposed what he calls a generalized mutual information which he expresses in terms of the second order ($q = 2$) correlation integral. Pompe's generalized mutual information is given by:

$$Q_2(x_1, \ldots, x_m) = H_2(x_m) - H_2(x_1, \ldots, x_m) + H_2(x_1, \ldots, x_{m-1}) \tag{26}$$

[7] One can also express the probability in m dimensions as $P = k/c_m r_m(k)^m$ where $r_m(k)$ is the distance to the k^{th} nearest neighbor, and c_m is the volume of a "sphere" of radius r; it depends both on the embedding dimension and the distance norm), so it should be possible to make "fixed mass" versions of the generalized redundancies as well.

which is the same as our second order ($q = 2$) generalized marginal redundancy.

While $q = 1$ leads to the natural definition of the entropy[8] there are reasons to prefer different values of q. For example, for all q except $q = 2$ there are corrections to Eq. (24) for small r. Grassberger has derived the asymptotic form of these corrections [56], but for finite length sets, the best statistics, at small r, are obtained by using $q = 2$ (In contrast, for box counting methods there are small n_i corrections for all q [56]). Another reason for using $q = 2$ is that, because of its symmetry, it is the fastest to compute of all the generalized correlation integrals.

The correlation integral approach is generally more accurate than box-counting methods. For a data set of N points, the correlation integral has a dynamic range of $O(N^2)$, whereas the number of boxes in any box-counting scheme is an integer between 1 and N. The larger dynamic range of the correlation integral implies a larger dynamic range for r, which permits the use of smaller values of r [18]. Another benefit of the correlation integral method is that it is easy to correct for oversampled data, by excluding points close in time to the reference point as suggested in Ref. [59]. We should mention that the speed of box counting methods [60] is not that much of an issue, as there are numerous fast correlation integral algorithms available [61–63] so one can compute the generalized redundancies for small r in $O(N \log N)$ time or faster.

As mentioned above, for data with very non-uniform distributions it can be advantageous to use a kernel whose width depends on the local density (adaptive kernel), an easier approach is to just transform the data to have a uniform distribution as suggested by Pompe [51], that way fixed size bins are equivalent to fixed mass bins in $m = 1$ (this is essentially what is done by the adaptive box counting method of Paluš [64]).

5.1 Generalized linear redundancies

We can also define linear versions of the generalized entropies,

$$\mathcal{H}_q(\vec{x}) = \frac{1}{(1-q)} \log_2 \left(\frac{|\boldsymbol{\xi}|^{q/2}}{(2\pi)^{qm/2}} \int e^{-\frac{q}{2}\sum_{i,j}^{m} \xi_{ij} x_i x_j} d\vec{x} \right) \tag{27}$$

using $|q\boldsymbol{\xi}| = q^m |\boldsymbol{\xi}|$ and the normalization condition of the gaussian we find:

$$\begin{aligned} \mathcal{H}_q(\vec{x}) &= \frac{1}{(1-q)} \log_2 \left[\frac{|\xi_{ij}|^{(q-1)/2}}{q^{m/2}(2\pi)^{(q-1)m/2}} \right] \\ &= \frac{m}{2}\log_2(2\pi) + \frac{1}{2}\log_2|\boldsymbol{\Xi}| + \frac{m \log_2(q)}{2(q-1)}, \end{aligned} \tag{28}$$

and for the generalized linear redundancies:

$$\mathcal{R}(\vec{x}) = \frac{1}{2}\sum_{i}^{m} \log_2(\Xi_{ii}) - \frac{1}{2}\log_2|\boldsymbol{\Xi}| \tag{29}$$

which is the same as Eq. (16); that is, the linear redundancies do not depend on q, but the "linear entropies" (Eq. (28)) do depend on q, as does the linearized version of the generalized

[8] That is, $q = 1$ is the only one of the generalized entropies which is an additive quantity.

correlation integral

$$\mathcal{C}_q(\vec{x}) = \left[\frac{|\boldsymbol{\xi}|^{(q-1)/2}}{q^{m/2}(2\pi)^{(q-1)m/2}}\right]^{1/(q-1)}. \tag{30}$$

Therefore, the idea of comparing a nonlinear statistic to its "linearized" version can be extended to all the statistics based on the correlation integral (*e.g.* Refs. [4, 19, 20, 22, 23]).

6 Applications

6.1 Clean computer generated data

As the first test, the correlation integral based redundancy analysis is applied to 8192 points from the chaotic Rössler equations [65] (with parameters $a = 0.15$, $b = 0.2$, $c = 10$ and a sampling time of $\delta t = 0.314$, these are the same parameters used by Paluš [64]). In Fig. 1 we show the linear, C_2, and C_1 based redundancies and marginal redundancies as a function of the time delay τ, with $r = 0.1\sigma$, where σ is the standard deviation of the data set. The lines are for increasing embedding dimensions ($m = 2, \ldots, 8$) starting at the bottom of the graph. Eq. (12) suggests that as m is increased the marginal redundancy curves as a function of τ should accumulate to a line which has a slope equal to -1 times the K_q entropy. From the slope of the C_2 based marginal redundancies (Fig. 1b) it can be seen that the K_2 entropy is roughly 0.03 bits/timestep (a similar value is found using Eq. (11)). However, a reliable estimate of the K_1 entropy cannot be obtained with this number of points (see Fig. 1d) using either the method of Eq. (11) or Eq. (13). That is, as suggested above, by using the C_2 based redundancies we can either use smaller r or fewer data points. In Fig. 1(e-f) the linear redundancies and marginal redundancies are shown for comparison. Notice the difference between the τ dependence of redundancies and their linearized versions.

6.2 Real data: SFI-A

This method is also applied to 8192 points from a chaotic laser experiment, which exhibits Lorenz-type chaos with a correlation dimension of roughly 2.05 and a positive K_2 entropy [66]. This data set (`A.cont`) was part of the time series competition sponsored by the Santa Fe Institute [67]. In Fig. 2 we show the results of the analysis for the linear, C_1, and C_2 based redundancies as a function of the time delay τ and for embedding dimensions $m = 2, \ldots, 8$. For the C_2 based redundancies we use $r = 0.1\sigma$, however, when using this value of r for the C_1 based redundancies there are positive slopes at large m and τ for the marginal redundancy, as in figure Fig. 1d. The best results are obtained for $r = 0.25\sigma$, which is what is shown in Fig. 2(c-d). The linear redundancies and marginal redundancies are shown in Fig. 2(e-f) for comparison. The difference between the shapes of the redundancies and their linearized versions clearly shows that there is nonlinearity in this data set. From the slopes of the C_2 based marginal redundancies it is seen that the K_2 entropy is roughly 0.025 bits/timestep (we find $K_2 \approx 0.03$ bits/timestep from Eq. (11)). However, we are unable to get an estimate of the K_1 entropy with this number of data points. Paluš [64] also examines this data set and gets very similar results to Fig. 2(c-d) using an adaptive box counting method (compare to figure 8 of Paluš).

7 Relation to other statistics

The connection between the entropy and the correlation integral allows us to relate other correlation integral based statistics to their analogues from information theory. For example, if the sequence is IID then R_q will be zero; this is the idea behind the BDS test [19–21]. Putting R_q in terms of the correlation integral (for delay coordinates) we find:

$$R_q(\vec{x}, r) = \log_2(C_q(\vec{x}, r)) - \log_2([C_q(x_1, r)]^m) \tag{31}$$

which is very similar to the BDS statistic

$$\mathrm{BDS}_q(\vec{x}, r) = C_q(\vec{x}, r) - [C_q(x_1, r)]^m. \tag{32}$$

The linearized version of Eq. (31) is given by Eq. (29), and the linearized BDS statistic can be computed via Eq. (30).

Green and Savit [23] have also proposed a statistic, based on the correlation integral, to quantify the amount of additional information in x_m about x_1 which is not a result the dependence of x_1 on $x_2, \ldots x_{m-1}$. One measure of this is given by the conditional redundancy:

$$\begin{aligned} R_q(x_1; x_m | x_2, \ldots, x_{m-1}, r) &= R'_q(x_1, \ldots, x_{m-1}; x_m, r) - R'_q(x_2, \ldots, x_{m-1}; x_m, r) \\ &= -\log_2 \left(\frac{C_q(x_1, \ldots, x_{m-1}, r) C_q(x_2, \ldots, x_m, r)}{C_q(x_1, \ldots, x_m, r) C_q(x_2, \ldots, x_{m-1}, r)} \right). \end{aligned} \tag{33}$$

For convenience we use $R^*_q(m)$ as an abbreviation for $R_q(x_1; x_m | x_2, \ldots, x_{m-1}, r)$. We also define $R^*_q(1) = -H_q(x_1) = \log_2(C_q(x_1, r))$ and

$$R^*_q(2) = I(x_1, x_2, r) = -\log_2 \left(\frac{C_q(x_1, r) C_q(x_2, r)}{C_q(x_1, x_2, r)} \right). \tag{34}$$

The relationship of this conditional redundancy to the q^{th} order Green-Savit ζ_q statistic [23] is given by $R^*_q(m) = -\log_2(1 - \zeta_q(m-1))$. For a time delay embedding Eq. (33) reduces to:

$$R^*_q(m) = -\log_2 \left(\frac{C_q(x_1, \ldots, x_{m-1}, r)^2}{C_q(x_1, \ldots, x_m, r) C_q(x_1, \ldots, x_{m-2}, r)} \right). \tag{35}$$

which is related to the q^{th} order Savit-Green δ_q statistic [22] by: $R^*_q(m) = -\log_2(1 - \delta_q(m - 1))$. The linearized versions of δ_q and R^*_q are similar to partial autocorrelation coefficients [68] in that if there is no explicit dependence beyond the j^{th} lag then the linearized $\delta_q(m > j)$ will be zero.

Green and Savit also introduce what they call a "predictability index", which also can be related to $R^*_q(m)$ via:

$$S^*_q(m) = \frac{C_q(x_1)}{\prod_{j=1}^{m} \zeta_q(j)} = \exp \left[\sum_{j=1}^{m+1} R^*_q(j) \right]. \tag{36}$$

For a time delay embedding this reduces to the predictability index of Savit and Green, which has a very simple form when expressed in information-theoretic terms:

$$S^*_q(m) = \exp[H_q(m-1) - H_q(m+1)] \tag{37}$$

From this expression it is clear that $\lim_{\substack{m\to\infty \\ r\to 0}} \log_2(S_q^\star(m)) = -2\tau K_q$. The relationship to the generalized K-S entropy makes it clear why this is called the "predictability index".

We are not advocating the use of these information-theoretic statistics in place of the BDS or Green and Savit statistics, but rather, pointing out that there are information theory based analogues to these statistics. The simple connection between δ_q and $R_q^\star$ also means that the results of Wu, Savit and Brock [69] on the asymptotic behavior of δ_2 can be applied to $R_2^\star$. The connection of the δ_q's to the information-theoretic quantities $R_q^\star$ also allows us to address the problem of the negative deltas pointed out by Savit and Green. The negative deltas are an artifact of taking $q \neq 1$, that is they are a result of using the Renyi entropies rather than the Shannon entropy, as pointed out by Pompe [51] this can be overcome either by taking $q = 1$, or by transforming the data to have a uniform distribution. (As mentioned earlier transforming the data to have an uniform distribution is similar to using an adaptive kernel.) The connection of the Green-Savit statistics to information-theoretic quantities means that these statistics can also be estimated by a box counting approach, which could be useful in the case where one has an explicit partition. An important distinction between these statistics, and statistics like the correlation dimension and the K-S entropy, is that they are evaluated at a fixed r as opposed to taking the limit as $r \to 0$. Another measure of this type is the approximate entropy (ApEn) statistic advocated by Pincus [24] (also called ϵ-entropy by Gaspard and Wang [70]), which is just the K_2 entropy [4,5] evaluated at a fixed m and r.

7.1 Cross-redundancies

Estimating the relations between multiple time series is an important problem. Recently statistics have been proposed to estimate nonlinear correlations between variables [23,71]. One simple measure, which we call the cross-redundancy, is given by:

$$I_q(x_1; x_2, l, r) = H_q(x_1(t), r) + H_q(x_2(t-l), r) - H_q(x_1(t), x_2(t-l), r) \tag{38}$$

where l is a lag time, as in a cross-correlation (A similar statistic has been used by Vastano and Swinney [72] for measuring information transport in spatiotemporal systems). That is, the cross-redundancy is just the mutual information between x_1 and a lagged value of x_2. The cross-redundancy can also be expressed in terms of the correlation integral:

$$I_q(x_1; x_2, l, r) = -\log_2 \left(\frac{C_q(x_1(t), r) C_q(x_2(t-l), r)}{C_q(x_1(t), x_2(t-l), r)} \right). \tag{39}$$

It can be seen that this is very similar to the quantity defined in Eq. (34) (the information-theoretic version of the statistic of Green and Savit for $m = 2$), the important difference being the use of the lag time. As with the other statistics we can define a linearized version of the cross-redundancy:

$$\mathcal{I}_q(x_1; x_2, l) = -\frac{1}{2} \log_2(1 - (\Xi_{x_1 x_2}(l))^2) \tag{40}$$

where it is assumed that both of the series x_1 and x_2 have zero mean and unit variance, and $\Xi_{x_1 x_2}(l) = \langle x_1(t) x_2(t-l) \rangle$ is the cross-correlation function between x_1 and x_2 as a function of the lag time l. The cross-redundancy is a very simple example of a statistic to measure the relations between variables, much more complicated measures can be defined in terms of the R_q^* or other information-theoretic statistics.

As an example, the second order ($q = 2$) cross-redundancy and its linearized version are computed for 8192 points of the x and y components of the Lorenz equations [73] (with parameters $\sigma = 10$, $\beta = 5/3$, and $r = 28$, and a sampling time of $\delta t = 0.04$). In Fig. 3 we show the linear cross-redundancy (dashed curve), and the cross-redundancy (solid lines) for r values $0.01\sigma, 0.02\sigma, \ldots, 0.1\sigma$ as a function of lag time (the top curve is $r = 0.01\sigma$). Notice that both the cross-redundancy and its linearized version show a peak at a lag of roughly 2 time steps, but that the cross-redundancy also shows (nonlinear) correlations at longer lags, which are not detected by the linear cross-redundancy.

8 Local information based measures

The relationship between the correlation integral and the entropy makes it easy to define one possible type of "local" information-theoretic measures (local in state space), for example, the local version of the Shannon entropy is:

$$h_1(\vec{x}(t), r) = -\log_2\left(\frac{1}{N}\sum_{\substack{s\\ s\neq t}} \Theta(r - \|\vec{x}(t) - \vec{x}(s)\|)\right) = -\log_2(B(\vec{x}(t), r)). \tag{41}$$

We can also define local redundancies, based on the inner sum of the correlation integral $(B(\vec{x}(t), r))$, as well as a local version of the approximate K-S entropy [24, 70]:

$$\begin{aligned} k_1(\vec{x}(t), r) &= [h_1(x_1(t), \ldots, x_m(t), r) - h_1(x_1(t), \ldots, x_{m-1}(t), r)]/\tau \\ &= \frac{1}{\tau}\left[\log_2\left(\frac{B(x_1(t), \ldots, x_{m-1}(t), r)}{B(x_1(t), \ldots, x_m(t), r)}\right)\right] \end{aligned} \tag{42}$$

where $(x_1(t), x_2(t), \ldots, x_m(t))$ is a time delay embedding: $(x(t), x(t-\tau), \ldots, x(t-(m-1)\tau)$. As with the aproximate entropy of Pincus [24], we evaluate the local K-S entropy at a fixed r and m rather than taking the limits $r \to 0$ and $m \to \infty$. The local K-S entropy gives us a rough measure of the variability of the local predictability, without actually doing nonlinear prediction, or calculating local (finite time) Lyapunov exponents [74–76]. If one is just interested in determining how the degree of predictability changes across the state space, and not in the specific values of the predictability, we suggest that calculating the local K-S entropy may be a numerically easier way to get this information that using nonlinear prediction or local Lyapunov exponents. Like the local Lyapunov exponents, the local K-S entropy is coordinate dependent, therefore it can only be used for qualitative characterization of the predictability.

As an example of the local K-S entropy we generate $N = 65536$ points of the x and z variables of the Rössler equations (with the same parameters as were used in section 6.1) The local K-S entropy is then estimated in embedding dimensions $m = 3, \ldots, 8$ using the x component, a time delay of $\tau = 5$ (the first minimum of the mutual information) and $r = \sigma/4$ for the first 500 points along the trajectory. By examining the Rössler attractor it is clear that most of the stretching and folding occurs when z is large, therefore, we expect the local entropy to be large only when z is large. In Fig. 4 the local K-S entropy for embedding dimensions $m = 3, \ldots, 8$ is shown, as well as the z component of the Rössler equations. Notice that the local entropy has "spikes" when z is large, as was expected. In the figure the curves are shifted to the right by $\tau = 5$ for each increasing embedding dimension, because of the time delay embedding.

Casdagli *et al.* [77] point out that the quality of a reconstructed state space depends on the coupling between the variables, they illustrate this idea using the Lorenz equations [73]:

$$\begin{aligned} \dot{x} &= \sigma(y - x) \\ \dot{y} &= rx - xz - y \\ \dot{z} &= xy - \beta z \end{aligned} \tag{43}$$

they point out that when x is small the coupling between y and z is weak, so in the presence of any noise one expects a poor reconstruction when $x \approx 0$. They quantify this effect using what they call noise amplification, which is related to the conditional variance:

$$\mathrm{Var}(z|\vec{y}) = \int z^2 p(z|\vec{y}) dz - \left(\int z p(z|\vec{y}) dz\right)^2. \tag{44}$$

For delay coordinates, $z = x(t + \tau)$ and $\vec{y} = (x(t), \ldots, x(t - (m-1)\tau))$. Casdagli *et al.* [77] argue that conditional variance is a more appropriate criteria for determining the quality of a embedded than is mutual information. (Čenys *et al.* [71, 78] have also proposed statistics based on the conditional variance.) However, it is also possible to measure the coupling between variables using a local information theoretic measure, rather than one based on the conditional variance. Specifically we can look at the quantity $i_1(z(t); y(t)|x(t))$ along a trajectory in the state space to determine how coupling between y and z depends on the position in state space. We can express $i_1(z(t); y(t)|x(t))$ in terms of the inner sum of the correlation integral ($B(\vec{x}(t), r)$):

$$\begin{aligned} i_1(z(t); y(t)|x(t), r) &= h_1(y(t), z(t), r) + h_1(x(t), y(t), r) - h_1(x(t), y(t), z(t), r) - h_1(x(t), r) \\ &= -\log_2\left(\frac{B(y(t), z(t), r) B(x(t), y(t), r)}{B(x(t), y(t), z(t), r) B(x(t), r)}\right). \end{aligned} \tag{45}$$

As an example, we generate $N = 65536$ points of x, y, and z from the Lorenz equations with parameters $\sigma = 16$, $\beta = 4$, $r = 45.92$, and a time step of $\delta t = 0.01$. Using $r = 3.25$ (roughly 1/4 of the standard deviation) $i_1(z(t); y(t)|x(t))$ is computed for the first 500 points along the trajectory. In Fig. 5 we show $i_1(z(t); y(t)|x(t))$ versus x (*cf.* figure 9 of Casdagli *et al.*). Notice that $i_1(z(t); y(t)|x(t))$ is small near $x = 0$, since when x is small the coupling between y and z is weak.

9 Conclusions

We have established the relationship between various information-theoretic quantities and the correlation integral of order 1, and extensions of these quantities based on the generalized correlation integral. It has been shown that the correlation integral approach has several advantages over box counting methods (especially for $q = 2$), and that the idea of comparing the τ dependence of a statistic to its "linearized" version can be extended to all the statistics based on the correlation integral. We also have shown that the BDS statistic and the statistics proposed by Savit and Green have information-theoretic analogues. Finally, we have introduced some new information-theoretic statistics, including "local" versions of several statistics based on the inner sum of the correlation integral.

Acknowledgments

We thank Milan Paluš for many discussions, for his comments and suggestions on earlier versions of this manuscript, and for pointing out that the difference of marginal redundancies in Eq. (33) could be written as a conditional redundancy. We also thank Thomas Schreiber, Martin Casdagli, Bette Korber, Channon Price, and John Williams for their comments and suggestions. This work was partially supported by NIMH grant 1-R01-MH47184 and the US Department of Energy.

Figure captions

Fig. 1 Redundancy analysis for the Rössler data set, for embedding dimensions $m = 2 - 8$ and time delays $\tau = 1 - 60$ sample times. All curves are for $r = 0.1\sigma$ where σ is the standard deviation of the time series. (a) C_2 based normalized redundancies $(R_2(\tau)/(m-1))$. (b) C_2 based marginal redundancies $(R'_2(\tau))$. (c) C_1 based normalized redundancies $(R_1(\tau)/(m-1))$. (d) C_1 based marginal redundancies $(R'_1(\tau))$. (e) Normalized linear redundancies $(\mathcal{R}(\tau)/(m-1))$. (f) Linear marginal redundancies $(\mathcal{R}'(\tau))$.

Fig. 2 Redundancy analysis for the SFI-A data set, for embedding dimensions $m = 2 - 8$ and time delays $\tau = 1-60$ sample times. Curves (a) and (b) are for $r = 0.1\sigma$, curves (c) and (d) are for $r = 0.25\sigma$, where σ is the standard deviation of the time series. (a) C_2 based normalized redundancies $(R_2(\tau)/(m-1))$. (b) C_2 based marginal redundancies $(R'_2(\tau))$. (c) C_1 based normalized redundancies $(R_1(\tau)/(m-1))$. (d) C_1 based marginal redundancies $(R'_1(\tau))$. (e) Normalized linear redundancies $(\mathcal{R}(\tau)/(m-1))$. (f) Linear marginal redundancies $(\mathcal{R}'(\tau))$.

Fig. 3 Cross-redundancy between the x and y components of the Lorenz equations as a function of time lag (solid curves). Top curve is for $r = 0.01\sigma$, next is for $r = 0.02\sigma$ and so on to $r = 0.1\sigma$. Dashed line is for the linear cross-redundancy.

Fig. 4 Top panel: z component of the Rössler equations. Next 6 panels are the (approximate) local K-S entropy calculated from the x component for embedding dimensions $m = 3 - 8$.

Fig. 5 The local mutual information between z and y given x, $(i_1(z;y|x))$ plotted against x for a short trajectory of the Lorenz equations. When x is small the coupling between y and z is weak, therefore, $i_1(z;y|x)$ is small.

References

1. R. S. Shaw, "Strange attractors, chaotic behavior, and information flow," *Z. Naturforsh.* **36a**, 80–112 (1981).
2. J.-P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," *Rev. Mod. Phys.* **57**, 617–656 (1985).
3. C. Shannon and W. Weaver, *The Mathematical Theory of Communication.* (University of Illinois Press, Chicago, 1949).
4. P. Grassberger and I. Procaccia, "Estimation of the Kolmogorov entropy from a chaotic signal," *Phys. Rev. A* **28**, 2591–2593 (1983).
5. A. Cohen and I. Procaccia, "Computing the Kolmogorov entropy from time signals of dissipative and conservative dyanamical systems," *Phys. Rev. A* **31**, 1872–1882 (1985).
6. K. Pawelzik and H. G. Schuster, "Generalized dimensions and entropies from a measured time series," *Phys. Rev. A* **35**, 481–484 (1987).
7. A. Renyi, *Probability theory.* (North-Holland, Amsterdam, 1970).
8. A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Phys. Rev. A* **33**, 1134–1140 (1986).
9. A. M. Fraser, "Information and entropy in strange attractors," *IEEE Transactions on Information Theory* **IT-35**, 245–262 (1989).
10. M. Paluš, V. Albrecht, and I. Dvořák, "Information theoretic test for nonlinearity in time series," *Phys. Lett. A* **175**, 203–209 (1993).
11. J. D. Farmer, "Information dimension and the probabilistic structure of chaos," *Z. Naturforsh.* **37a**, 1304–1325 (1982).
12. N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a time series," *Phys. Rev. Lett.* **45**, 712–716 (1980).
13. J. D. Farmer, E. Ott, and J. A. Yorke, "The dimension of chaotic attractors," *Physica D* **7**, 153–180 (1983).
14. P. Grassberger and I. Procaccia, "Characterization of strange attractors," *Phys. Rev. Lett.* **50**, 346–349 (1983).
15. P. Grassberger, "Generalizations of the Hausdorff dimension of fractal measures," *Phys. Lett. A* **107**, 101–105 (1985).
16. R. Badii and A. Politi, "Statistical description of chaotic attractors: The dimension function," *J. Stat. Phys.* **40**, 725–751 (1985).
17. J. Kurths and H. Herzel, "An attractor in a solar time series," *Physica D* **25**, 165–172 (1987).
18. J. Theiler, "Estimating fractal dimension," *J. Opt. Soc. Am. A* **7**, 1055–1073 (1990).
19. W. A. Brock, W. D. Dechert, and J. Scheinkman, "A test for independence based on the correlation dimension," Technical Report 8702, Social Systems Research Institute, University of Wisconsin, Madison (1987).
20. W. A. Brock, "Nonlinearity and complex dynamics in economics and finance," in *The Economy as an Evolving Complex System*, P. Anderson, K. Arrow, and D. Pines, eds. (Addison-Wesley, New York, 1988).
21. W. A. Brock, D. A. Hsieh, and B. LeBaron, *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence.* (MIT Press, Cambridge, MA, 1991).
22. R. Savit and M. Green, "Time series and dependent variables," *Physica D* **50**, 95–116

(1991).
23. M. L. Green and R. Savit, "Dependent variables in broad band continuous time series," *Physica D* **50**, 521–544 (1991).
24. S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proc. Natl. Acad. Sci. USA* **88**, 2297–2301 (1991).
25. H. G. Solari and R. Gilmore, "Relative rotation rates for driven dynamical systems," *Phys. Rev. A* **37**, 3096–3109 (1988).
26. G. B. Mindlin, X.-J. Hou, H. G. Solari, R. Gilmore, and N. B. Tufillaro, "Classification of strange attractors by integers," *Phys. Rev. Lett.* **64**, 2350–2353 (1990).
27. N. B. Tufillaro, H. G. Solari, and R. Gilmore, "Relative rotation rates: fingerprints for strange attractors," *Phys. Rev. A* **41**, 5717–5720 (1990).
28. M. R. Muldoon, R. S. MacKay, J. P. Huke, and D. S. Broomhead, "Topology from time series," *Physica D* **65**, 1–16 (1993).
29. A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series," *Physica D* **16**, 285–317 (1985).
30. J.-P. Eckmann, S. O. Kamphorst, D. Ruelle, and S. Ciliberto, "Liapunov exponents from a time series," *Phys. Rev. A* **34**, 4971–4979 (1986).
31. H. D. I. Abarbanel, R. Brown, and M. B. Kennel, "Lyapunov exponents in chaotic systems: their importance and their evaluation using observed data," *Int. J. Mod. Phys. B* **5**, 1347–1375 (1991).
32. J. D. Farmer and J. J. Sidorowich, "Predicting chaotic time series," *Phys. Rev. Lett.* **59**, 845–848 (1987).
33. J. D. Farmer and J. J. Sidorowich, "Exploiting chaos to predict the future and reduce noise," in *Evolution, Learning and Cognition*, Y. C. Lee, ed. (World Scientific, 1988), pp. 277–330.
34. M. Casdagli, "Nonlinear prediction of chaotic time series," *Physica D* **35**, 335–356 (1989).
35. G. Sugihara and R. May, "Nonlinear forecasting as a way of distinguishing chaos from measurement error in forecasting," *Nature* **344**, 734–741 (1990).
36. M. Casdagli, "Chaos and deterministic versus stochastic nonlinear modeling," *J. R. Stat. Soc. B* **54**, 303–328 (1992).
37. D. T. Kaplan and L. Glass, "Direct test for determinism," *Phys. Rev. Lett.* **68**, 427–430 (1992).
38. M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase–space reconstruction using a geometrical construction," *Phys. Rev. A* **45**, 3403–3411 (1992).
39. D. T. Kaplan and L. Glass, "Coarse-grained embedding of time series: random walks, Gaussian random processes, and deterministic chaos," *Physica D* **64**, 431–454 (1993).
40. R. Wayland, D. Bromley, D. Pickett, and A. Passamante, "Recognizing determininism in a time series," *Phys. Rev. Lett.* **70**, no. 5, 580–582 (1993).
41. X. He and H. Asada, "A new method for identifying orders of input-output models for nonlinear dynamical systems," in *Proc. Am. Control Conference*, A. A. C. Council, ed. (IEEE Service Center, 1993), pp. 2520–2523.
42. D. T. Kaplan, "Exceptional events as evidence for determinism," *Physica D* **73**, 38–48 (1994).

43. J. L. Kaplan and J. A. Yorke, "Chaotic behavior of multidimensional difference equations," in *Functional Differential Equations and Approximations of Fixed Points*, H.-O. Peitgen and H.-O. Walther, eds. Vol. 730 of *Springer Lecture Notes in Mathematics*, (Springer-Verlag, Berlin, 1979), pp. 204–?
44. C. Grebogi, E. Ott, and J. A. Yorke, "Unstable periodic orbits and the dimension of chaotic attractors," *Phys. Rev. A* **36**, 3522–3524 (1987).
45. D. Auerbach, P. Cvitanović, J.-P. Eckmann, G. Gunaratne, and I. Procaccia, "Exploring chaotic motion through periodic orbits," *Phys. Rev. Lett.* **58**, 2387–2389 (1987).
46. P. Cvitanović, "Invariant measurement of strange sets in terms of cycles," *Phys. Rev. Lett.* **61**, 2729–2732 (1988).
47. R. Artuso, E. Aurell, and P. Cvitanović, "Recycling of strange sets: I Cycle expansions," *Nonlinearity* **3**, 325–359 (1990).
48. R. Artuso, E. Aurell, and P. Cvitanović, "Recycling of strange sets: II Applications," *Nonlinearity* **3**, 361–386 (1990).
49. W. Liebert and H. G. Schuster, "Proper choice of the time delay for the analysis of chaotic time series," *Phys. Lett. A* **142**, 107–111 (1988).
50. P. Grassberger, T. Schreiber, and C. Schaffrath, "Non-linear time sequence analysis," *Int. J. Bif. Chaos* **1**, 521–547 (1991).
51. B. Pompe, "Measuring statistical dependences in a time series," *J. Stat. Phys.* **73**, 587–610 (1993).
52. F. J. Pineda and J. C. Sommerer, "Estimating generalized dimensions and choosing time delays: a fast algorithm," in *Time Series Prediction: Forecasting the Future and Understanding the Past*, A. S. Weigend and N. A. Gershenfeld, eds. Vol. XV of *SFI Studies in the Sciences of Complexity*, (Addison-Wesley, 1993), pp. 367–385.
53. T. Buzug, K. Pawelzik, L. von Stamm, and G. Pfister, "Mutual information and global strange attractors in Taylor-Couette flow," *Physica D* **72**, 343–350 (1994).
54. M. Paluš, "Testing for nonlinearity using redundancies: quantitative and qualitative aspects," Technical Report SFI-93-12-076, Santa Fe Institute (1993). To Appear in Physica D.
55. J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, "Testing for nonlinearity in time series: the method of surrogate data," *Physica D* **58**, 77–94 (1992).
56. P. Grassberger, "Finite sample correction to entropy and dimension estimates," *Phys. Lett. A* **128**, 369–373 (1988).
57. D. Wolpert and D. Wolf, "Estimating functions of probability distributions from a finite set of samples; Part I: Bayes estimators and the Shannon entropy," Technical Report SFI-93-07-046, Santa Fe Institute (1993).
58. B. W. Silverman, *Kernel Density Estimation Techniques for Statitistics and Data Analysis.* (Chapman Hall, London, 1986).
59. J. Theiler, "Spurious dimension from correlation algorithms applied to limited time series data," *Phys. Rev. A* **34**, 2427–2432 (1986).
60. P. Grassberger, "On efficient box counting algorithms," *Int. J. Mod. Phys. C* **4**, 515–523 (1993).
61. J. Theiler, "Efficient algorithm for estimating the correlation dimension from a set of discrete points," *Phys. Rev. A* **36**, 4456–4462 (1987).
62. S. Bingham and M. Kot, "Multidimensional trees, range searching and a correlation

dimension alogrithm of reduced complexity," *Phys. Lett. A* **140**, 327–330 (1989).
63. P. Grassberger, "An optimized box-assisted algorithm for fractal dimensions," *Phys. Lett. A* **148**, 63–68 (1990).
64. M. Paluš, "Indentifying and quantifying chaos by using infomation theoretic functionals," in *Time Series Prediction: Forecasting the Future and Understanding the Past*, A. S. Weigend and N. A. Gershenfeld, eds. Vol. XV of *SFI Studies in the Sciences of Complexity*, (Addison-Wesley, 1993), pp. 387–413.
65. O. E. Rössler, "An equation for continuous chaos," *Phys. Lett. A* **57**, 397–398 (1976).
66. U. Hübner, C.-O. Weiss, N. B. Abraham, and D. Tang, "Lorenz-like chaos in NH_3-FIR lasers (data set A)," in *Time Series Prediction: Forecasting the Future and Understanding the Past*, A. S. Weigend and N. A. Gershenfeld, eds. Vol. XV of *SFI Studies in the Sciences of Complexity*, (Addison-Wesley, 1993), pp. 73–104.
67. A. S. Weigend and N. A. Gershenfeld, eds., *Time Series Prediction: Forcasting the Future and Understanding the Past*, Vol. XV of *SFI Studies in the Sciences of Complexity*, Addison-Wesley, (1993).
68. G. M. Jenkins and D. G. Watts, *Spectral analysis and its applications.* (Holden-Day, San Francisco, 1968).
69. K. Wu, R. Savit, and W. Brock, "Statistical tests for deterministic effects in broad band time series," *Physica D* **69**, 172–188 (1993).
70. P. Gaspard and X.-J. Wang, "Noise, chaos and (ϵ, τ)-entropy per unit time," *Physics Reports* **235**, 291–343 (1993).
71. A. Čenys, G. Lasiene, and K. Pyragas, "Estimation of interrelation between chaotic observables," *Physica D* **52**, 332–337 (1991).
72. J. A. Vastano and H. L. Swinney, "Information transport in spatiotemporal systems," *Phys. Rev. Lett.* **60**, 1773–1776 (1988).
73. E. N. Lorenz, "Deterministic nonperiodic flow," *J. Atmos. Sci.* **20**, 130–141 (1963).
74. J. N. Nese, "Quantifying local predictability in phase space," *Physica D* **35**, 237–250 (1989).
75. H. D. I. Abarbanel, R. Brown, and M. B. Kennel, "Varability of Lyapunov exponents on a strange attractor," *J. Nonlinear Sci.* **1**, 175–199 (1991).
76. H. D. I. Abarbanel, R. Brown, and M. B. Kennel, "Local Lyapunov exponents from observed data," *J. Nonlinear Sci.* **2**, 345–365 (1992).
77. M. Casdagli, S. Eubank, J. D. Farmer, and J. Gibson, "State space reconstruction in the presence of noise," *Physica D* **51**, 52–98 (1991).
78. A. Čenys and K. Pyragas, "Estimation of the number of degrees of freedom from chaotic time series," *Phys. Lett. A* **129**, 227–230 (1988).

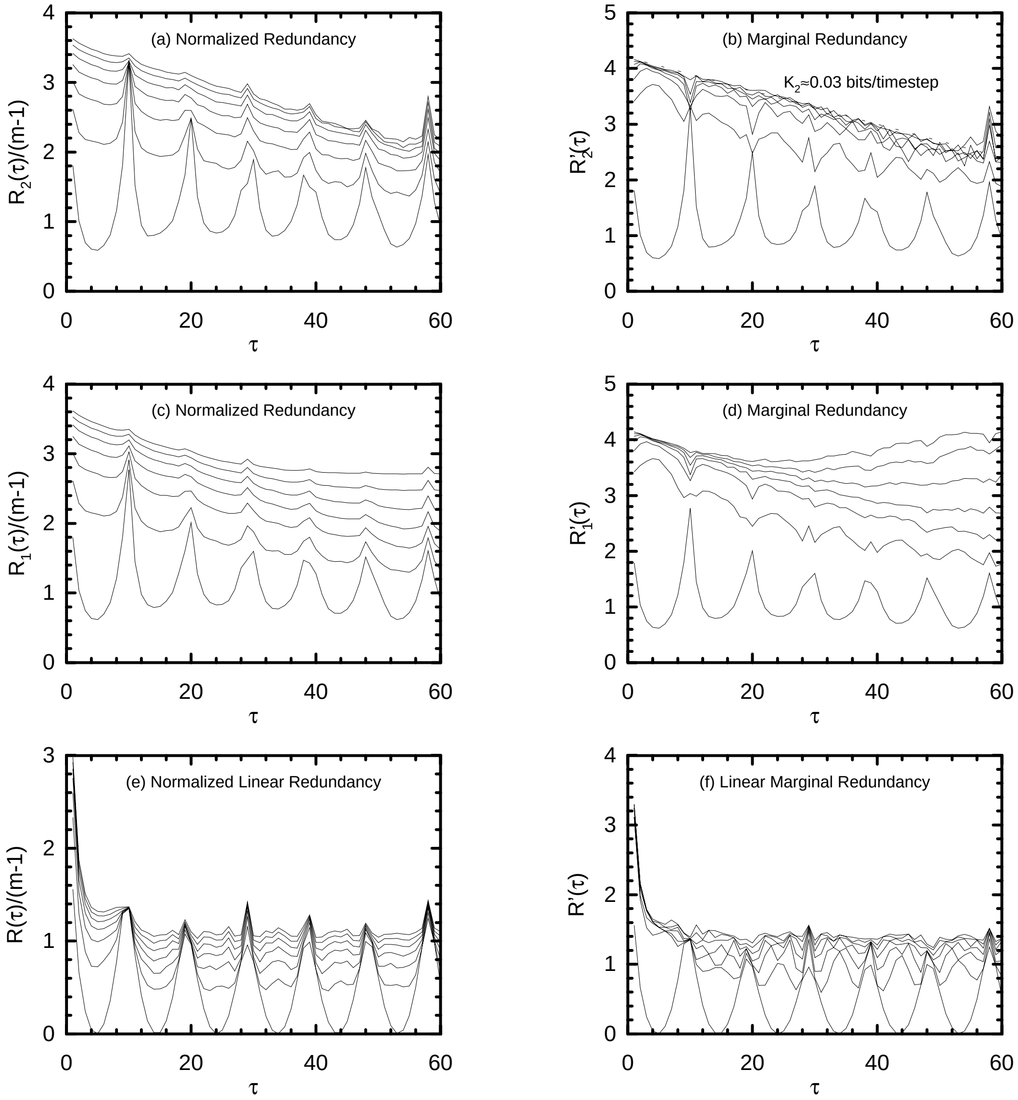

Fig. 1. Redundancy analysis for the Rössler data set, for embedding dimensions $m = 2 - 8$ and time delays $\tau = 1 - 60$ sample times. All curves are for $r = 0.1\sigma$ where σ is the standard deviation of the time series. (a) C_2 based normalized redundancies ($R_2(\tau)/(m-1)$). (b) C_2 based marginal redundancies ($R'_2(\tau)$). (c) C_1 based normalized redundancies ($R_1(\tau)/(m-1)$). (d) C_1 based marginal redundancies ($R'_1(\tau)$). (e) Normalized linear redundancies ($\mathcal{R}(\tau)/(m-1)$). (f) Linear marginal redundancies ($\mathcal{R}'(\tau)$).

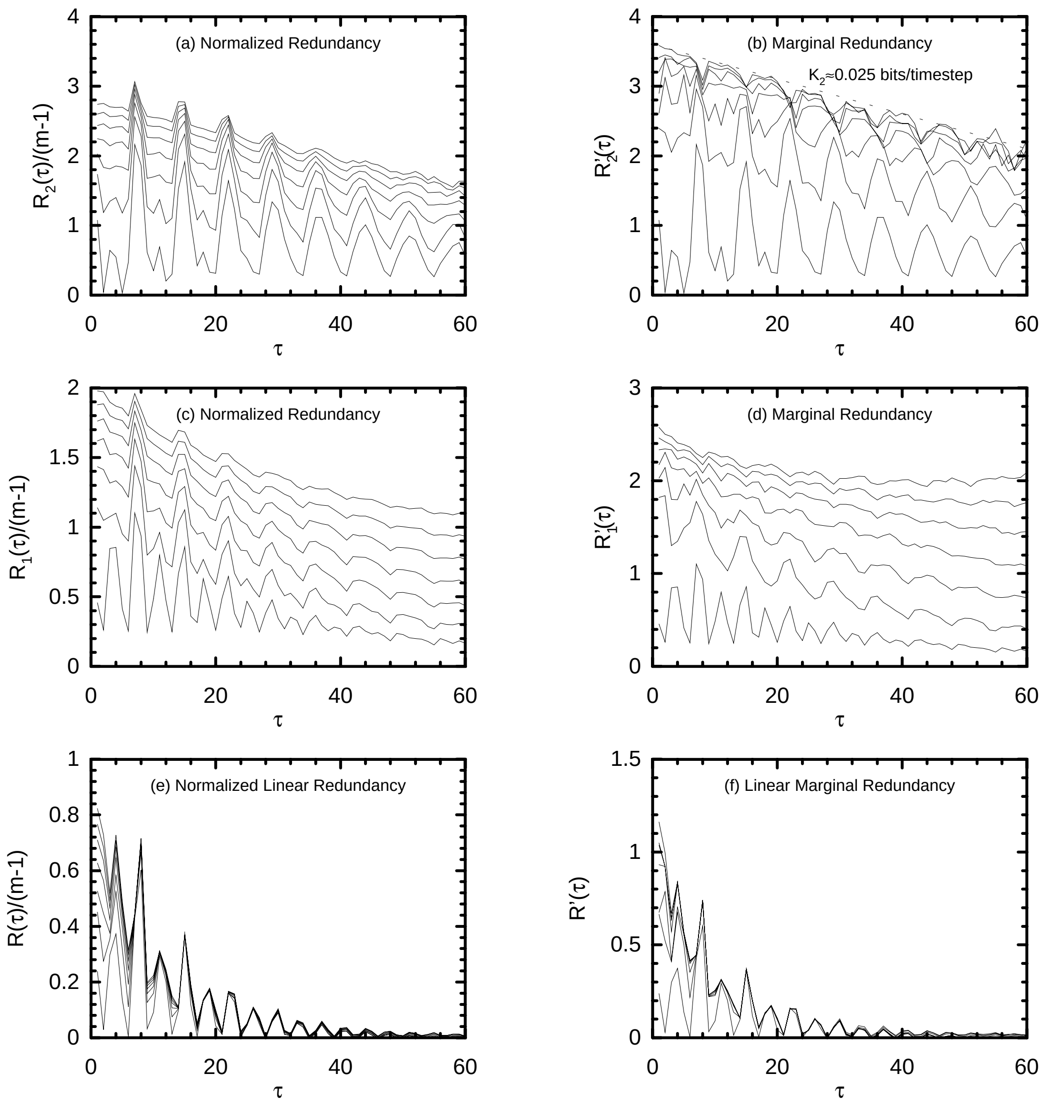

Fig. 2. Redundancy analysis for the SFI-A data set, for embedding dimensions $m = 2 - 8$ and time delays $\tau = 1 - 60$ sample times. Curves (a) and (b) are for $r = 0.1\sigma$, curves (c) and (d) are for $r = 0.25\sigma$, where σ is the standard deviation of the time series. (a) C_2 based normalized redundancies $(R_2(\tau)/(m-1))$. (b) C_2 based marginal redundancies $(R_2'(\tau))$. (c) C_1 based normalized redundancies $(R_1(\tau)/(m-1))$. (d) C_1 based marginal redundancies $(R_1'(\tau))$. (e) Normalized linear redundancies $(\mathcal{R}(\tau)/(m-1))$. (f) Linear marginal redundancies $(\mathcal{R}'(\tau))$.

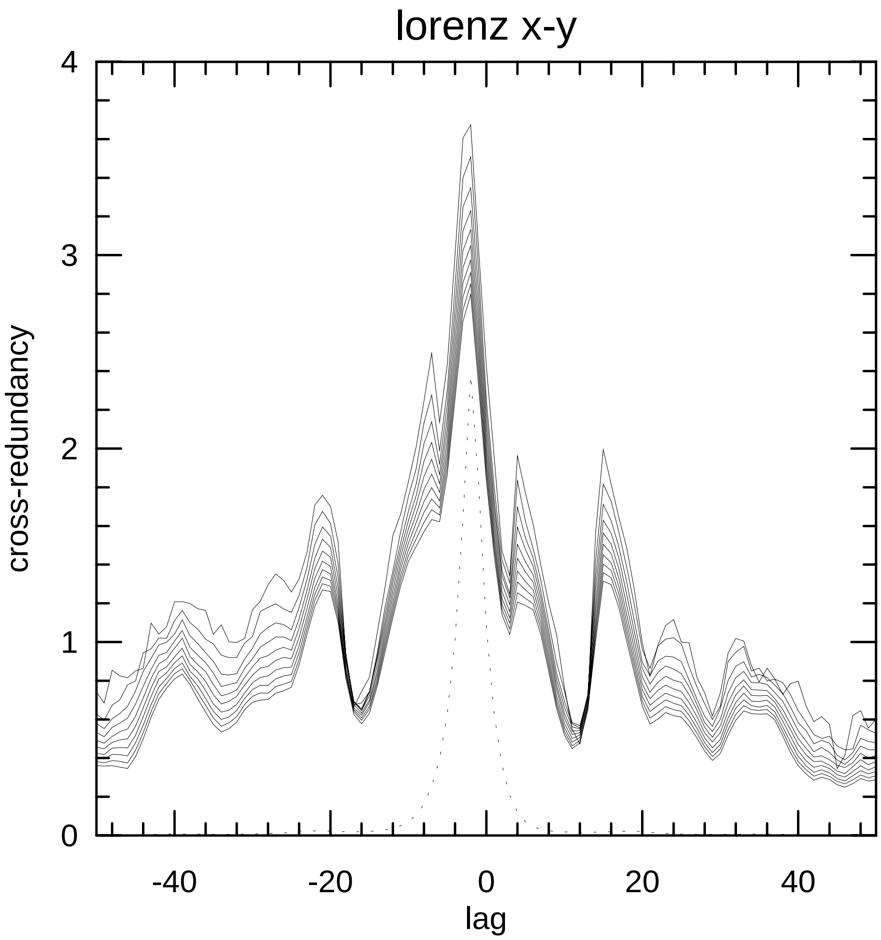

Fig. 3. Cross-redundancy between the x and y components of the Lorenz equations as a function of time lag (solid curves). Top curve is for $r = 0.01\sigma$, next is for $r = 0.02\sigma$ and so on to $r = 0.1\sigma$. Dashed line is for the linear cross-redundancy.

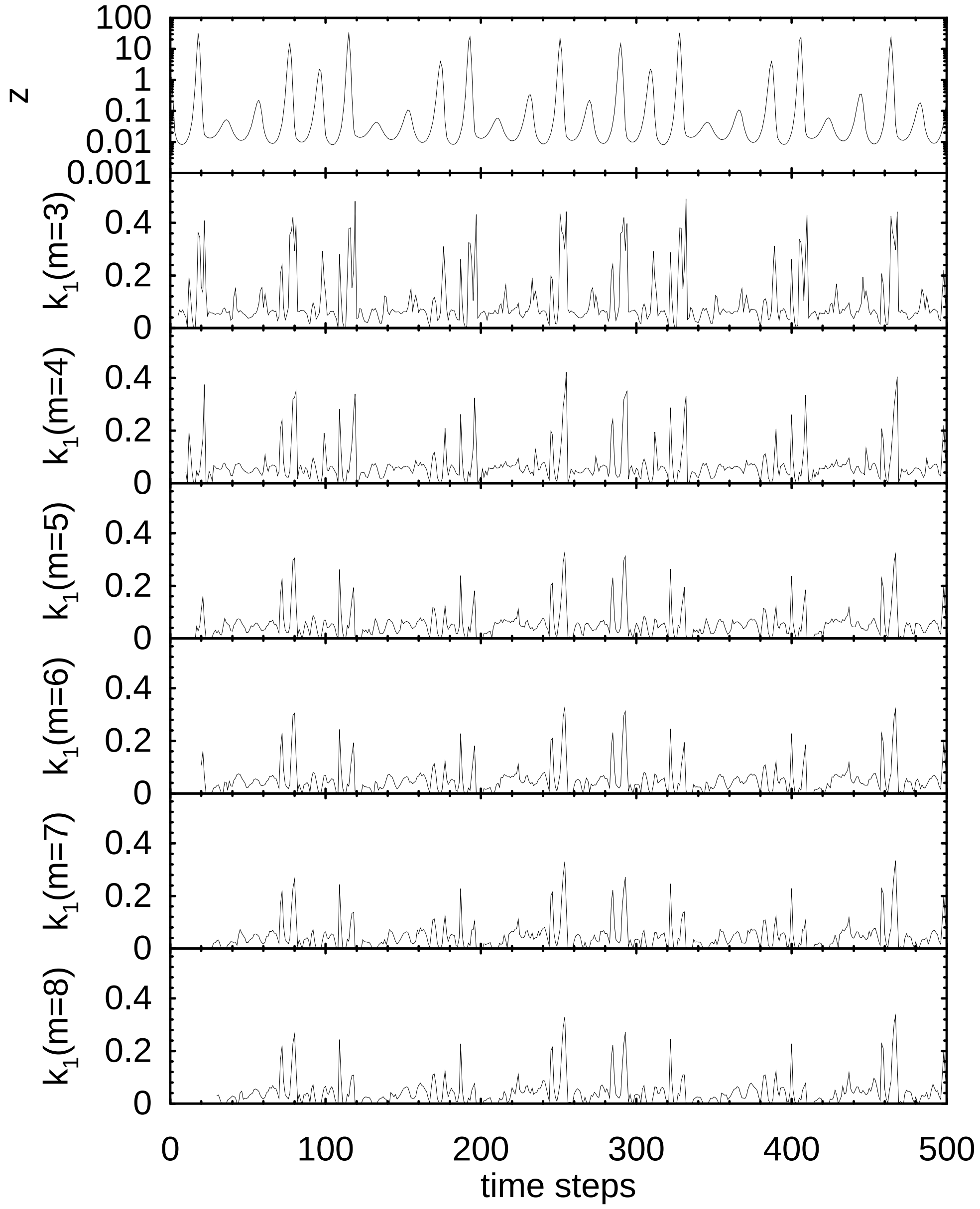

Fig. 4. Top panel: z component of the Rössler equations. Next 6 panels are the (approximate) local K-S entropy calculated from the x component for embedding dimensions $m = 3 - 8$.

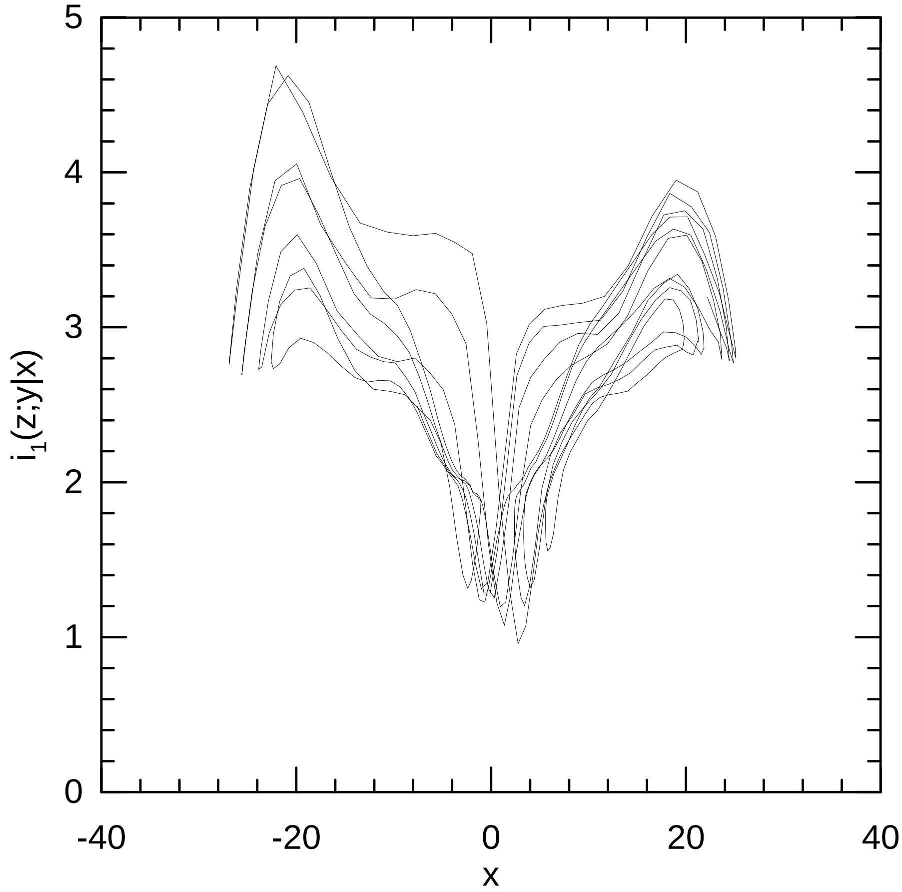

Fig. 5. The local mutual information between z and y given x, $(i_1(z;y|x))$ plotted against x for a short trajectory of the Lorenz equations. When x is small the coupling between y and z is weak, therefore, $i_1(z;y|x)$ is small.