

Dependent variables in broad band continuous time series

Matthew L. Green and Robert Savit

Department of Physics, The University of Michigan, Ann Arbor, MI 48109, USA

Received 24 September 1990

Revised manuscript received 20 February 1991

Accepted 20 February 1991

Communicated by R.M. Westervelt

In this paper we continue our development of new methods for the analysis of broad band time series by deriving quantities which are able to indicate deterministic dependence of an element in one time series on elements in other time series. These methods are very broadly applicable and are particularly well suited to the study of continuous time series, in which the value of the function may depend on derivatives of the function itself, or on other quantities. We apply our methods to a number of mathematical examples including the Lorentz equation, the Hénon–Heiles equations, the forced Brusselator and the Mackey–Glass equation. We show that our methods are very successful at indicating deterministic dependencies in these systems, even if the time series are highly chaotic. Statistical aspects of our procedure are discussed, as are a number of interesting and surprising epistemological implications.

1. Introduction

In a recent paper [1] (hereinafter referred to as SG), we extended the methods of Brock, Dechert, and Scheinkman [2] by developing indicators for functional dependence among elements of a discrete time series. In SG we used conditional probabilities for the repetition of short sequential patterns of values in a time series to determine, quantitatively, the extent to which a term in the series is a function of previous elements of the series. In this paper, we modify the methods developed in SG in order to be able to apply them to discuss conditional dependencies among elements of different time series. One particularly important application of this general idea is to the analysis of continuous time series in which the series may be generated by an underlying differential equation, as in the case of fluid flow. In such a case the series $x(t)$ may have deterministic dependence on (among other things) various derivatives of x . But the functions \dot{x} , \ddot{x} , etc. are time series with different statistical characteristics

than $x(t)$ itself. Thus, a generalization of the methods of SG is necessary to treat this important case. In most of the rest of this paper we shall use continuous time series with deterministic derivative dependence as examples to illustrate and develop our methods. But it should be borne in mind that they are much more generally applicable.

As with the methods in SG, the techniques derived here are particularly useful when applied to time series with broad band characteristics which may be the result of nonlinear processes. The analysis of such series by more traditional (linear) methods may fail to capture the presence of important nonlinear deterministic effects. Our methods are based on the use of certain kinds of conditional probabilities that can be constructed from vector generalizations of the Grassberger–Procaccia correlation integrals [2]. As a result, the methods we derive are often able to capture deterministic effects which become apparent only when viewed in an appropriately defined higher-dimensional space, analogous to

the embedding space of a chaotic attractor. In this context it is important to distinguish between the absence of correlations, as indicated by the value of autocorrelation functions, and statistical independence. It may happen that the autocorrelation functions of a time series are zero, while the elements of the series have nontrivial statistical interdependencies. A simple example is the time series generated by iterations of the tent map. This series has zero autocorrelation functions, but statistical indicators such as those from SG, or from ref. [2] indicate the strong underlying deterministic dynamics.

In the next section we shall derive indicators for dependent derivatives and also explain their applicability to the more general case of dependence among other sets of continuous time series. In section 3, we shall develop a predictability index which characterizes the extent to which a continuous time series is predictable given knowledge of other continuous time series considered as independent variables in the analysis. In section 4 we shall apply our test to several continuous systems including systems with periodic or other simple asymptotic behavior as well as systems which exhibit chaotic behavior. The chaotic systems examined include the Lorentz equations, the Hénon–Heiles equations, and the forced Brussels oscillator, as well as the Mackey–Glass delay equation, which includes non-local effects. Section 5 consists of a summary and conclusions. Some technical issues are relegated to an appendix.

2. Indicators for dependent derivatives

In order to facilitate the discussion of our method, we first derive indicators for dependence of a continuous time series on its derivatives, as might result, for example from an underlying deterministic differential equation. This provides a very familiar model and also a more direct analogy with the derivation of statistics for discrete time series in SG [1]. Keep in mind that

although we consider a special case, the method we derive is applicable to the more general case of dependence among different time series which are not simply derivatives or integrals of each other.

Consider a continuous set of data $x(t) \in \mathbb{R}$, $0 \leq t < \infty$ with continuous first through m th derivatives. We denote these derivatives by $x^{[k]}(t)$ for $k = 1, \dots, m$. The method we develop below will measure, quantitatively, the extent to which the value of x at time t is determined by the values of the low-order derivatives of x at time t , i.e. the extent to which

$$x(t) = F(\{x^{[k]}(t')\}) \quad (1)$$

for some function F and some set $\{x^{[k]}(t')\}$ of derivatives of x of order less than or equal to m . Obviously our method will be most successful in analyzing time series derived from autonomous local differential equations, which satisfy (1) by definition, though it is useful in other circumstances as well. Note, in particular, that the argument of $x^{[k]}$ is t' , by which we mean to include the possibility of non-local equations as well as differential equations that are strictly local in time. In addition, we shall discuss the inclusion of explicit time dependence in (1).

Before describing our procedure, it is useful to review the method described in SG for the analysis of discrete time series. From a discrete time series x_i , $i = 1, \dots, N$ we formed d -dimensional vectors

$$\begin{aligned} v(i) &= (x_i, x_{i-1}, \dots, x_{i-d+1}) \\ &= (v_1(i), v_2(i), \dots, v_d(i)). \end{aligned} \quad (2)$$

In SG we then constructed the Grassberger–Procaccia correlation integrals [3]

$$C_d(\varepsilon) = \frac{1}{\nu} \sum_{\text{pairs}} I_{ij}(\varepsilon), \quad (3a)$$

where

$$I_{ij}(\varepsilon) = \prod_{k=1}^d \Theta(\varepsilon - |v_k(i) - v_k(j)|), \quad (3b)$$

ν is the number of pairs of vectors $(v(i), v(j))$, and ε is a positive real number which controls the tolerance (or uncertainty) with which we observe the dynamics of the discrete system. $C_d(\varepsilon)$ is the probability that all the components of two d -dimensional vectors, $v(i)$ and $v(j)$, are within ε of each other. Using the C_d 's one can construct conditional probabilities which indicate the extent to which an element of the sequence, x_i , depends on previous elements of the sequence, or in other words whether or not $v_1(i)$ depends on $v_2(i)$ through $v_d(i)$.

For continuous time series, $x(t)$, we are interested in dependencies of a function on its derivatives. Therefore, we consider, instead of the vector $v(i)$, the vector whose elements consist of successively higher derivatives of $x(t)$. In particular, we choose a sampling interval, τ , and from the time series $x(t)$ form d -tuples of derivatives at times $t = n\tau$, $1 \leq n \leq N$, which we will consider to be vectors in a d -dimensional space:

$$\begin{aligned} \mathbf{u}(n\tau) &= (x(n\tau), x^{[1]}(n\tau), \dots, x^{[d-1]}(n\tau)) \\ &\equiv (u_0(n\tau), u_1(n\tau), \dots, u_{d-1}(n\tau)). \end{aligned} \quad (4)$$

Next, we have to modify the definition of the indicator function $I_{ij}(\varepsilon)$. In the discrete case, each coordinate of the vectors $v(i)$ samples a large subset of the sequence x_i , $i = 1, \dots, N$, so we use the same parameter ε to measure the separation of all coordinates. In the continuous case we must choose a separate value of ε , ε_k , for each coordinate $u^{[k]}(n\tau)$, $k = 0, \dots, d - 1$. In this paper, we choose

$$\varepsilon_k = \mu \cdot \text{st. dev.} \{x^{[k]}(n\tau), n = 1, \dots, N\} \quad (5)$$

(st. dev. = standard deviation of), where the tolerance with which we examine the data is now described by the positive real number μ . Note

that since we are assuming in general that the coordinate projections of the vector $\mathbf{u}(n\tau)$ are different, the methods we derive in this paper will apply equally to any case in which we are presented with a sequence of vectors $\mathbf{u}(n\tau)$, $n = 1, \dots, N$ and we wish to discover the dependencies among the coordinates of the vectors. Hence, this technique may be used to probe correlations among measurements of several different time dependent quantities of a system, not just among derivatives of a single function. For example the pressure, volume, and temperature of an expanding gas may be combined into the three-dimensional vector

$$\begin{aligned} \mathbf{u}(n\tau) &= (P(n\tau), V(n\tau), T(n\tau)) \\ &= (u_1(n\tau), u_2(n\tau), u_3(n\tau)), \end{aligned}$$

and that vector subjected to the analysis described below.

Using (5) the indicator function (3b) becomes

$$I_{ij}(\mu) = \prod_{k=0}^{d-1} \Theta(\varepsilon_k - |u_k(i\tau) - u_k(j\tau)|). \quad (6)$$

We may then define the correlation integral of dimension d as

$$C_0^{d-1}(\mu) = \frac{1}{\nu} \sum_{\text{pairs}} I_{ij}(\mu), \quad (7)$$

where ν is the number of pairs of vectors $(\mathbf{u}(i\tau), \mathbf{u}(j\tau))$, and where the notation C_0^{d-1} indicates that $C_0^{d-1}(\mu)$ contains information about the first d coordinates (u_0 through u_{d-1}) of $\mathbf{u}(t)$ as shown in (4).

Consider two vectors $\mathbf{u}(i\tau)$ and $\mathbf{u}(j\tau)$. It is easy to see that $C_0^{d-1}(\mu)$ is the probability that $u_k(i\tau)$ and $u_k(j\tau)$ are within ε_k of each other for $k = 0, \dots, d - 1$. That is,

$$C_0^{d-1}(\mu) = P(D_0, D_1, \dots, D_{d-1}), \quad (8)$$

where D_k stands for the statement " $|u_k(i\tau) - u_k(j\tau)| \leq \varepsilon_k$ ".

We now want to test whether the time series $x(t)$ has deterministic dependence on its derivatives, as one would expect if, for example, $x(t)$ had been generated from a differential equation of the form (1) with $x^{[k]}(t) \in \{x^{[k]}(t)\}$. To do so we examine the conditional probability, $P(D_0|D_1)$, that $|x(i\tau) - x(j\tau)| \leq \varepsilon_0$ given that $|x^{[1]}(i\tau) - x^{[1]}(j\tau)| \leq \varepsilon_1$. Since $P(D_0, D_1) = P(D_0|D_1)P(D_1)$, we see that

$$P(D_0|D_1) = \frac{P(D_0D_1)}{P(D_1)}. \tag{9}$$

If knowledge of $x^{[1]}(i\tau)$ to within ε_1 does not increase one's ability to predict the value of $x(i\tau)$ to within ε_0 , then $P(D_0|D_1) = P(D_0)$, or from (9)

$$\frac{P(D_0D_1)}{P(D_1)} = P(D_0). \tag{10}$$

Since, unlike the discrete case (2), the coordinate projections of $u(i\tau)$ in (2) are generically all different, it is not correct to set $P(D_1) = P(D_0) = C_0^0(\mu)$. Instead, we must calculate $P(D_1)$ explicitly. $P(D_1)$ is the probability that $|u^{[1]}(i\tau) - u^{[1]}(j\tau)| \leq \varepsilon_1$, or

$$P(D_1) = \frac{1}{\nu} \sum_{\text{pairs}} \Theta(\varepsilon_1 - |u^{[1]}(i\tau) - u^{[1]}(j\tau)|) \equiv C_1^1(\mu), \tag{11}$$

where the notation $\overset{1}{C}_1^1$ indicates that $C_1^1(\mu)$ is a correlation integral containing information only about the first derivative of $x(t)$. Then (10) becomes

$$\frac{C_0^1(\mu)}{C_1^1(\mu)} = C_0^0(\mu), \tag{12}$$

and in analogy with the discrete case we define

$$\zeta_1(\mu) = 1 - \frac{C_1^1(\mu)C_0^0(\mu)}{C_0^1(\mu)}. \tag{13}$$

In a statistical sense, ζ_1 is zero if $P(D_0|D_1) =$

$P(D_0)$. The size of ζ_1 is thus a measure of the extent to which $P(D_0|D_1) \neq P(D_0)$, i.e. the extent to which knowledge of $x^{[1]}(t)$ to within $\pm\varepsilon_1$ increases our ability to predict $x(t)$ to within $\pm\varepsilon_0$.

Next, we want to know if knowledge of $x^{[2]}(t)$ to within $\pm\varepsilon_2$ increases the predictability of $x(t)$ (to within $\pm\varepsilon_0$) over the predictability already achieved through knowledge of $x^{[1]}(t)$ (to within $\pm\varepsilon_1$). Examine the conditional probability $P(D_0|D_1D_2)$. Since $P(D_0D_1D_2) = P(D_0|D_1D_2) \times P(D_1D_2)$ we have

$$P(D_0|D_1D_2) = \frac{P(D_0D_1D_2)}{P(D_1D_2)} = \frac{C_0^2}{C_1^2}, \tag{14}$$

where for notational convenience, we have dropped the variable μ . If knowledge that D_2 is true does not increase the probability that D_0 is true, then

$$P(D_0|D_1D_2) = P(D_0|D_1) = \frac{C_0^1}{C_1^1}. \tag{15}$$

Using (14) and (15) we define

$$\zeta_2(\mu) = 1 - \frac{C_0^1C_1^2}{C_0^2C_1^1}. \tag{16}$$

The size of $\zeta_2(\mu)$ is a measure of the extent to which $P(D_0|D_1D_2) \neq P(D_0|D_1)$.

In general, we may define

$$\zeta_j(\mu) = 1 - \frac{C_0^{j-1}C_1^j}{C_0^jC_1^{j-1}}. \tag{17}$$

$\zeta_j(\mu)$ is statistically zero if $P(D_0|D_1, \dots, D_j) = P(D_0|D_1, \dots, D_{j-1})$ and so the size of $\zeta_j(\mu)$ is a measure of how much knowledge of $x^{[j]}(t)$ (to within $\pm\varepsilon_j$) increases our ability to predict $x(t)$ (to within $\pm\varepsilon_0$) over the predictability already obtained through knowledge of $x^{[i]}(t)$ to within $\pm\varepsilon_i$ for $i = 1, \dots, j - 1$.

In a practical implementation of this procedure on a digital computer, we must approximate the

derivatives in (4) by finite differences. This requires that in addition to the sampling interval τ we choose an interval Δt to use in forming the finite differences. We must then ask whether aliasing affects our results. Since we assume the derivatives of $x(t)$ exist, we may Taylor expand $x(t)$ and write

$$x(t) = x(t_0) + (t - t_0)x^{[1]}(t_0) + \dots + \frac{1}{m!}(t - t_0)^m x^{[m]}(t_0) + (t - t_0)^m R_m(t_0).$$

Then, using the standard second order differencing scheme [4] (in which odd ordered differences are averaged to fall on a lattice point instead of between two lattice points) it is easy to show that

$$\frac{\delta^k(x(t_0))}{(\Delta t)^k} = x^{[k]}(t_0) + (\Delta t)^2 x^{[k+2]}(t_0) A_k + \mathcal{O}(\Delta t)^3, \tag{18}$$

where $\delta^k(x(t_0))$ indicates the k th difference of $x(t_0)$ on a lattice of spacing Δt . The first ten coefficients, A_k , are listed in table 1. The only place the values of the derivatives enter our calculations is in the indicator function (6), so the only requirement necessary to ensure that our results are unaffected by use of the k th difference (18) instead of the k th derivative $x^{[k]}(t)$ is

Table 1
The first ten coefficients A_k from eq. (18).

k	A_k	k	A_k
1	1/6	6	1/2
2	1/12	7	1/6
3	1/12	8	1/3
4	1/6	9	5/24
5	1/8	10	5/12

that on average

$$\frac{|\delta^k(x(i\tau)) - \delta^k(x(j\tau))|}{(\Delta t)^k} \leq \epsilon_k \Leftrightarrow |x^{[k]}(i\tau) - x^{[k]}(j\tau)| \leq \epsilon_k.$$

Hence, the error due to finite differencing will not affect our results so long as it is smaller, on average, than ϵ_k , the tolerance with which we observe k th derivatives. That is, our results will be accurate so long as

$$(\Delta t)^2 \overline{|x^{[k+2]}(t)|} A_k \leq \epsilon_k, \quad k = 1, 2, \dots,$$

where $\overline{|x^{[k+2]}(t)|}$ is the average magnitude of the second derivative of $x^{[k]}(t)$. Accordingly we define the indicator

$$R_k \equiv \frac{\epsilon_k}{(\Delta t)^2 \overline{|x^{[k+2]}(t)|} A_k}, \quad k = 1, 2, \dots \tag{19}$$

Barring any unusual pathologies, if R_k is greater than one then typically any of our calculations involving $x^{[k]}(t)$ should be accurate, and the larger the value of R_k , the more accurate the results. We will see in section 3 that when the tolerance, μ , is 1/2 we obtain accurate results with surprisingly large values of Δt .

We may further inquire as to how the choice of the sampling interval, τ , affects our results. Clearly, the indicators ζ_j (17) can only indicate dependence among elements of those vectors $\mathbf{u}(n\tau)$ which are included in our sampling (4). This sampling may not include all the information about the system being examined. For example, a small enough portion of a periodic function may appear to be simply linear, or a small enough portion of a chaotic system may in some cases appear to be periodic. In most cases, it is desirable when investigating a periodic system to include at least several periods in the sampling, and many points within each period. Similarly, if we are investigating a chaotic system with a strange attractor, our sampling of vectors $\mathbf{u}(n\tau)$ (4) should

contain enough points to reveal a large portion of the structure present.

Through numerical simulations we have observed that so long as N and τ are chosen consistent with the above requirement, the value of τ does not seem to strongly affect our results, at least on a wide class of representative systems. This effect, though somewhat surprising, can be understood by examining the $C_0^{d-1}(\mu)$ defined in (7). C_0^{d-1} is a sum of the indicator function $I_{ij}(\mu)$ defined in (6) over all possible pairs of d -dimensional vectors $(\mathbf{u}(i\tau), \mathbf{u}(j\tau))$. It follows that one can permute the labels of the vectors $\{\mathbf{u}(i\tau)\}$ in any way (exchanging, for example $u(2\tau)$ for $u(100\tau)$) without changing the value of any of the correlation integrals, since the same pairs of vectors will contribute to the sum (7) with any labeling. Hence, once all of the necessary derivatives are computed (a process involving Δt but not τ), the time ordering of the data points is irrelevant. We are investigating only the relationship between the components of the vectors, not the relationship between two successive vectors. We may conclude that any collection of vectors, $\{\mathbf{u}(t_i) = (x^{[0]}(t_i), x^{[1]}(t_i), \dots, x^{[d-1]}(t_i)), i = 1, \dots, N\}$, which samples sufficiently the structure of the function being examined, may be used to compute the correlation integrals and the indicators ζ . For simplicity, in the examples studied in this paper we have chosen to employ even spacing.

3. Predictability index

In SG we developed a predictability index, S , which indicates quantitatively to what extent a value in a discrete sequence is predictable given the values of previous elements of the sequence. In addition, we showed that

$$S = \lim_{d \rightarrow \infty} P(t_d | t_{d-1}, t_{d-2}, \dots, t_1),$$

where t_k stands for the statement “ $|v_{d-k+1}(i) -$

$v_{d-k+1}(j)| \leq \epsilon$ ” and the $v_k(i)$ are defined in (3). The corresponding expression for our continuous system, under the hypothesis (1), is

$$S = \lim_{d \rightarrow \infty} P(D_0 | D_1, D_2, \dots, D_d). \tag{20}$$

S is 1 if $x(t)$ is completely predictable (to within ϵ_0) given the values of all the derivatives of x at time t (to within their respective tolerances), and $S = P(D_0) = C_0^0$ if knowledge of the derivatives of $x(t)$ does not increase predictability. Note that it is highly unlikely to find S near 1 in cases in which the dynamics of the system are characterized by having one or more positive Lyapunov exponents. In this case nearby trajectories will diverge, thus limiting reproducibility of the system given any small uncertainty in measurement.

In practice, we do not have knowledge of the derivatives of $x(t)$ to all orders. Hence we approximate S by the quantities $S_j \equiv P(D_0 | D_1, \dots, D_j)$, the predictability index taking into account information from the first j derivatives of $x(t)$. S_j is 1 if $x(t)$ is completely predictable (in the sense of being reproducible to within $\pm \epsilon_0$) given the value of the first j derivatives of x at time t (to within $\pm \epsilon_j$), and $S_j = P(D_0) = C_0^0$ if knowledge of these derivatives does not increase predictability. It often happens that there is an integer U for which ζ_j is sensibly negligible for all $j > U$. In that case S_U will be a good approximation to S . To express S_j in terms of the $\zeta_j(\mu)$ we use (13) in the form

$$\frac{C_0^0}{1 - \zeta_1} = \frac{P(D_0)}{1 - \zeta_1} = \frac{C_0^1}{C_1^1} = P(D_0 | D_1)$$

and (17) in the form

$$\begin{aligned} \frac{1}{1 - \zeta_k} &= \frac{C_0^k / C_1^k}{C_0^{k-1} / C_1^{k-1}} \\ &= \frac{P(D_0 | D_1, \dots, D_k)}{P(D_0 | D_1, \dots, D_{k-1})}, \quad k = 2, \dots, j \end{aligned}$$

to get

$$S_j = P(D_0 | D_1, \dots, D_j) = \frac{C_0^0}{(1 - \zeta_1)(1 - \zeta_2) \dots (1 - \zeta_j)} \tag{21}$$

Finally, from (20) and (21) we see that

$$S = \frac{C_0^0}{\prod_{j=1}^{\infty} (1 - \zeta_j)} \tag{22}$$

in analogy with the discrete case [1]. The derivation of the predictability index proceeds in an analogous manner if we consider more general sets of dependent time series. Of course, if we consider only a finite set of independent variables, the upper limit in (22) will be equal to the number of independent variables.

4. Application to mathematical examples

In the following examples, we have used sequences of only 1000 vectors. One might expect significant errors to occur with the use of so few points. However, with 1000 points we have $(1000)^2$ pairs of vectors, and it is the number of pairs that is relevant to the accurate determination of the correlation integrals. Further, because we generate the time sequences from known equations, we know which variables occur in a simple phase space representation of the equations. The choice of different variables inevitably distorts this representation so that a larger number of points and greater accuracy may be required to obtain the same results we achieve here with 1000 vectors and $\mu = 0.5$. This point is discussed further in section 4.6. Finally, note that with $\mu = 0.5$, we do not require great precision in the estimation of derivatives. This allows us to investigate dependencies even among derivatives of order as high as 10. If we required greater precision than $\mu = 0.5$ the accuracy of our approximation of derivatives would have to become correspondingly greater as well.

4.1. Random walk

We first apply our test to a particle executing a random walk in one dimension. For each i from 1 to N , let η_i be a random variable equal to +1 or -1 with equal probability. Then we describe the trajectory of the particle by the function

$$x(t) = \sum_{i=0}^{\text{int}(t)-1} \eta_i + [t - \text{int}(t)] \eta(\text{int}(t)) \tag{23}$$

so that

$$x^{[1]}(t) = \eta(\text{int}(t)), \tag{24}$$

where int is the integer function. Note that $x(t)$ is continuous and $x^{[1]}(t)$ is piecewise continuous. According to these equations, the particle moves with speed 1 in the positive or negative direction for a time interval of length 1, at which point a new direction (positive or negative) is independently and randomly determined.

We formed the vectors

$$u(i) = [x(i + 0.001), x^{[1]}(i + 0.001)], \quad i = 1, \dots, N$$

and applied our test to 1000 independent sequences of length $N = 1000$, choosing a tolerance $\mu = \frac{1}{2}$. Because the velocity of the trajectory is determined independently and randomly, knowledge of $x^{[1]}(t)$ cannot increase our ability to predict $x(t)$. Hence we expect $\zeta_1(\mu)$ to be consistent with zero, indicating no one dependence. Indeed our results give the average value of ζ_1 to be $-4.5 \times 10^{-5} \pm 5.8 \times 10^{-5}$ with a standard deviation of 1.8×10^{-3} . This gives us an indication of what values for ζ_1 are consistent with zero for sequences of 1000 points.

In the following examples, we take a value of ζ_j much greater than $\cong 10^{-3}$ as an indication of j dependence in an underlying differential equation explaining the data. This is a reasonable

thing to do for our illustrative purposes. But statistically a more reliable procedure would be to randomly shuffle the time series in question and repeat the calculations of the ζ_j 's on the shuffled series to determine expected sizes of null values under a hypothesis of lack of statistical dependence. This is a special case of a more general procedure in statistics called bootstrapping (see ref. [5] for a review).

4.2. The sine function

Next, we apply our test to the function $x(t) = \sin(\omega t)$. We choose a frequency $\omega = \pi/2$, a tolerance $\mu = 0.5$, and a sampling interval $\tau = 1/3$. We also require R_k to be greater than or equal to 4.0 for $k = 1, \dots, 5$. According to eq. (19) this restricts Δt to be less than or equal to 0.82. We choose $\Delta t = 0.82$ to illustrate the accuracy of our results with this surprisingly large value. We applied our test to a sequence of length $N = 1000$. The ζ_j and S_j , $j = 1, \dots, 5$ are shown in table 2. Note that ζ_1 and ζ_2 are both large while the ζ_j for $j = 3, \dots, 5$ are quite small indicating that $x(t) = F(x^{[1]}(t), x^{[2]}(t))$. On the other hand, we know that if $x(t) = \sin(\omega t)$,

$$x(t) = -\frac{1}{\omega^2}x^{[2]}(t). \tag{25}$$

We may ask, why is there an indication of dependence on $x^{[1]}(t)$, when clearly information about the second derivative is sufficient to predict the value of $x(t)$? Integrating (25) and using the

initial conditions $x(0) = 0$, and $x^{[1]}(0) = 1$ we have

$$[x(t)]^2 = \frac{1}{\omega^2} (1 - [x^{[1]}(t)]^2). \tag{26}$$

Hence $x(t)$ does depend on its first derivative, but knowledge of $x^{[1]}(t)$ does not predict $x(t)$ completely. Writing (26) in the form of (1), we have

$$x(t) = \pm \frac{1}{\omega} \sqrt{1 - [x^{[1]}(t)]^2}. \tag{27}$$

Knowledge of the value of $x^{[2]}(t)$ allows us to resolve the sign ambiguity.

Also shown in table 2 are the results of applying our test to the same function with the same parameter values except that now $\Delta t = 0.01$, giving a value for R_k greater than or equal to 27000. Note that the values of ζ_j and S_j in the two cases are close, and in particular there is no difference in the eventual predictability S_5 . It may be surprising that we get such accurate results using $\Delta t = 0.82$, when the period of the function is just 4.0. However, according to eqs. (19) the average magnitude of the aliasing errors introduced by using $\Delta t = 0.82$ is still, on average, only about 25% of the tolerance, ϵ , which we set to be one half the standard deviation of the data points.

There are several lessons to learn from our examination of this simple system. First, often the most natural form in which to express a nonlinear differential equation is not in the explicit form

Table 2

ζ_j and predictability index for the sin function with finite differencing interval Δt . $\pi_0 = \left\{ \begin{smallmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 2 & 3 & 4 & 5 \end{smallmatrix} \right\}$ and $\pi_1 = \left\{ \begin{smallmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 0 & 2 & 1 & 3 & 4 & 5 \end{smallmatrix} \right\}$.

j	Permutation = $\pi_0 \Delta t = 0.82$		Permutation = $\pi_0 \Delta t = 0.01$		Permutation = $\pi_1 \Delta t = 0.82$	
	ζ_j	S_j	ζ_j	S_j	ζ_j	S_j
0		0.295		0.295		0.295
1	0.312	0.429	0.311	0.428	0.705	1.0
2	0.571	1.0	0.572	1.0	-2.2×10^{-17}	1.0
3	-3.6×10^{-17}	1.0	-1.1×10^{-16}	1.0	-3.6×10^{-17}	1.0
4	-3.6×10^{-17}	1.0	-1.1×10^{-16}	1.0	-3.6×10^{-17}	1.0
5	-3.6×10^{-17}	1.0	-1.1×10^{-16}	1.0	-3.6×10^{-17}	1.0

(1), but in an implicit form

$$G(t, x(t), x^{[1]}(t), \dots, x^{[n]}(t)) = 0. \tag{28}$$

Since the ζ_j indicate dependences of the form (1), it will often be necessary, at least in principle, to solve for the roots of the implicit equation (28) as a function of $x(t)$. If there are multiple roots, as in (28), and more than one root is present in the data, as in the present case, then any explanation of the data of the form (1) will have to involve derivatives of order higher than n .

Second, differential equations are often integrable. It should be clear from this example that if knowledge of the lowest-order derivatives of $x(t)$ increases the predictability of $x(t)$, the ζ_j will indicate dependence on these derivatives regardless of whether this leads to the simplest or most convenient explanation of the data. This is a consequence of the choice of the form (2) for the vectors $\mathbf{u}(n\tau)$. Since the lowest-order derivatives appear first, dependence on these derivatives is indicated first. Larger values of j correspond to tests for additional *marginal* dependence on successively higher derivatives. If we permute the order of the coordinates in (2) and apply our method to the resulting vectors, we can develop statistics which indicate dependence on, say, the j th order derivative, without first considering dependence on the lower order derivatives.

For example, define the permutation $\pi = 0123\dots \rightarrow 0213\dots$ and the vectors

$$\begin{aligned} \mathbf{u}(n\tau, \pi) &= (x(n\tau), x^{[2]}(n\tau), x^{[1]}(n\tau), \\ &\quad x^{[3]}(n\tau), \dots, x^{[d-1]}(n\tau)) \\ &= (u_0(n\tau, \pi), u_1(n\tau, \pi), \dots, u_{d-1}(n\tau, \pi)). \end{aligned} \tag{29}$$

From these vectors, it is straightforward to con-

struct the statistics

$$\zeta_j(\mu, \pi) = 1 - \frac{C_0^{j-1}(\mu, \pi) C_1^j(\mu, \pi)}{C_0^j(\mu, \pi) C_1^{j-1}(\mu, \pi)}. \tag{30}$$

The $\zeta_j(\mu, \pi)$ and $S_j(\mu, \pi)$ for the sine function are shown in table 2. Note that only $\zeta_1(\mu, \pi)$ is significantly different from zero. From (29) we see that this indication is consistent with eq. (25): Once we know the second derivative of the sine function we get no additional information for the purposes of determining $x(t)$ by knowing, in addition, the first derivative. In general, it may be necessary to apply our test to several different permutations of the vector $\mathbf{u}(n\tau)$ in order to discover the simplest form of differential dependence that explains the data to within the tolerance $\mu^{\#1}$.

4.3. The Lorentz equations

The Lorentz equations [6]

$$\begin{aligned} x^{[1]} &= \sigma(y - x), \\ y^{[1]} &= rx - y - xz, \\ z^{[1]} &= xy - bz \end{aligned} \tag{31}$$

are one of the most commonly studied examples of a continuous system which exhibits chaotic behavior. We chose parameter values $r = 28$, $b = \frac{8}{3}$, and $\sigma = 10$, which are well into the chaotic regime for the Lorentz equations. We integrated these equations and chose to examine the continuous time series $x(t)$. Table 3 shows the results of our test applied to $x(t)$. We chose a sampling interval $\tau = 1.0$, a tolerance (or uncertainty) $\mu = \frac{1}{2}$, and a lattice spacing $\Delta t = 0.06$, which implied $R_k \geq 4.0$ for $k = 1, \dots, 5$. As before, we

^{#1}There are physical systems in which we observe a small value of δ_1 and a large value of δ_2 no matter how the first two coordinates are permuted. This indicates that predictability is increased by knowledge of both variables, while knowledge of either variable alone does not increase predictability. This point will be discussed further in an upcoming publication by the present authors.

Table 3

ζ_j and predictability index for the Lorentz equations. $\pi_0 = \begin{Bmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 2 & 3 & 4 & 5 \end{Bmatrix}$, $\pi_1 = \begin{Bmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 0 & 2 & 1 & 3 & 4 & 5 \end{Bmatrix}$ and $\pi_2 = \begin{Bmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 0 & 3 & 1 & 2 & 4 & 5 \end{Bmatrix}$.

j	Permutation = π_0		Permutation = π_1		Permutation = π_2	
	ζ_j	S_j	ζ_j	S_j	ζ_j	S_j
0		0.268		0.268		0.268
1	0.282	0.374	0.435	0.475	0.330	0.401
2	0.482	0.722	0.323	0.701	0.268	0.547
3	0.157	0.856	0.181	0.856	0.361	0.856
4	2.1×10^{-2}	0.874	2.1×10^{-2}	0.874	2.1×10^{-2}	0.874
5	1.4×10^{-4}	0.874	1.4×10^{-4}	0.874	1.4×10^{-4}	0.874

chose a sequence of 1000 points. Note that there is a strong indication of dependence on the first three derivatives and a much weaker but still significant indication of dependence on the fourth derivative. The value of ζ_5 is less than 1.8×10^{-3} , the value we determined from the random walk which represents no significant dependence.

It is easy to eliminate y and z from eqs. (31) and derive the third order differential equation for x ,

$$\begin{aligned} &\left(1 + \frac{1}{\sigma}\right)\left(\frac{x^{[2]}}{x} - \frac{(x^{[1]})^2}{x^2}\right) \\ &+ \frac{1}{\sigma}\left(\frac{x^{[3]}}{x} - \frac{x^{[2]}x^{[1]}}{x^2}\right) + x\left(\frac{x^{[1]}}{\sigma} + x\right) \\ &- b\left[r - 1 - \left(1 + \frac{1}{\sigma}\right)\frac{x^{[1]}}{x} - \frac{x^{[2]}}{\sigma x}\right] = 0. \end{aligned} \tag{32}$$

Although one might expect from the form of (32) that knowledge of the first three derivatives of x might be sufficient to determine $x(t)$ this does not necessarily mean that only ζ_1 , ζ_2 and ζ_3 will be significantly different from zero, as we see in table 3. We have already seen two examples which argue against this expectation. In SG we showed that, although the logistic map has explicit dependence only on the first lag, δ_2 (and probably also δ_j 's for large j) are non-zero. As explained in SG, this is because of the non-trivial curvature in the logistic function. In addition, we saw in our discussion of the sine function, that multiple solutions to $x(t)$ may be resolved given

the values of higher derivatives which will lead to non-zero values of the ζ_j associated with those derivatives.

These two examples imply that the conditional probability measures of dependences in empirical time series which we are proposing may not be as closely tied to explicit analytic representations of those time series when such representations exist. Indeed, from an observational (or logical positivist) point of view, we believe that our measures of dependence may be more fundamental – they are certainly more general – than an analytic representation. We shall see an even more startling example of this when we study the Mackey–Glass equation, below. In addition, this point will be discussed further in section 5.

Also shown in table 3 are the results of permuting the order of derivatives as discussed in the previous section. Note especially the value of ζ_1 for each of the permutations. The size of ζ_1 is an indication of predictability of $x(t)$ due to knowledge of whichever derivative of x occurs first in the permuted order. Thus we see that x depends most strongly on its second derivative and depends about equally on its first and third derivatives.

Finally, examine the predictability index in table 3. For each permutation, the eventual predictability is equal to about 0.89, much larger than $S_0 \equiv C_0^0 \equiv 0.27$. As we expect, this is a strong indication that the data are explainable by an underlying deterministic process. This should be contrasted with the results from analysis of the

Lorentz equations using the methods of SG. To implement that analysis we form the vectors $v(i) = (x(i\tau), x((i - 1)\tau), \dots, x((i - d + 1)\tau))$, $i = 1, \dots, N$ and we examine the dependence among the elements of these vectors. When the sampling rate τ is small, there is a strong indication of one dependence because $x(t)$ depends on its first derivative. However, as τ increases the pattern of dependence varies in a complicated manner as the sampling interval becomes equal to the periods of the various Fourier components of the solution $x(t)$. It is very difficult to determine the underlying stationary dynamics from such information. Thus, enlarging the space of dependent functions, as we have done in this paper, can lead to very simple results even where the strictly lag-based methods of SG are of limited utility.

4.4. The Hénon–Heiles equations

The Hénon–Heiles equations [7] are a system of two coupled second order nonlinear equations,

$$\frac{d^2q_1}{dt^2} = -q_1 - 2q_1q_2, \tag{33a}$$

$$\frac{d^2q_2}{dt^2} = -q_2 - q_1^2 + q_2^2, \tag{33b}$$

originally proposed by Hénon and Heiles as a simplified model of galactic dynamics. For certain initial conditions these equations exhibit chaotic motion with a strange attractor. We integrated these equations using the initial conditions,

$$q_1^{[1]}(0) = 0.1, \quad q_2^{[1]}(0) = 0.467618, \\ q_1(0) = 0.1, \quad q_2(0) = 0.1,$$

and examined the function $q_1(t)$. Using parameter values $\tau = 1.0$, $\mu = 0.5$, and $\Delta t = 0.20$ so that $R_k \geq 4.0$, we obtained the values for ζ_j and S_j , $j = 1, \dots, 10$ shown in table 4. Note that S_4 through S_{10} are much larger than S_0 , strongly indicating that the data are explainable by a deterministic differential process.

As noted in section 2, our test may be applied to examine the correlations among several variables even if those variables are not successive derivatives of one function. To illustrate this type of application, we form the vectors

$$u(t) = (q_1(t), q_1^{[1]}(t), \dots, q_1^{[5]}(t), q_2(t), q_2^{[1]}(t), \dots, q_2^{[5]}(t)). \tag{34}$$

Table 4 shows the results of applying our method to these vectors using parameters $\tau = 1.0$, $\mu = 0.5$,

Table 4

ζ_j and predictability index for the Hénon–Heiles equations. The first two columns show the dependence of $q_1(t)$ on its first ten derivatives. Rows zero through five of the last two columns are equal to rows zero through five of the first two columns. Rows six through ten of the last two columns show the marginal dependence of $q_1(t)$ on $q_2(t)$ and its first four derivatives. Hence, in the last two columns ζ_{6+j} and S_{6+j} refer to dependence on $q_2^{[j]}$, $j = 0, \dots, 4$.

j	Dependence of $q_1(t)$ on its derivatives		Dependence of $q_1(t)$ on $q_2(t)$ and its derivatives	
	ζ_j	S_j	ζ_j	S_j
0		0.263		0.263
1	6.32×10^{-2}	0.281	6.32×10^{-2}	0.281
2	0.537	0.607	0.537	0.607
3	9.04×10^{-2}	0.667	9.04×10^{-2}	0.667
4	0.122	0.760	0.122	0.760
5	-2.16×10^{-3}	0.758	-2.16×10^{-3}	0.758
6	4.59×10^{-4}	0.758	7.22×10^{-2}	0.794
7	9.94×10^{-3}	0.766	8.76×10^{-2}	0.871
8	9.31×10^{-3}	0.773	6.48×10^{-2}	0.931
9	1.33×10^{-2}	0.783	2.50×10^{-2}	0.955
10	1.02×10^{-4}	0.784	7.93×10^{-3}	0.963

and $\Delta t = 0.20$ ($R_k \geq 4.0$). Note that we obtain a predictability of 0.97, a value much larger than the predictability of 0.78 obtained in table 4 using just one variable and its derivatives. This may be understood through an examination of the defining equations (33). We may rewrite (33) as

$$q_2 = \frac{-q_1^{[2]}}{2q_1} - \frac{1}{2}, \quad (35a)$$

$$q_1^2 = -q_2^{[2]} - q_2 + q_2^2. \quad (35b)$$

Eq. (35b) shows the dependence of q_1 on q_2 and its derivatives. In addition, we may substitute (35a) into (35b) to obtain the fourth order equation for q_1

$$q_1^2 = \frac{q_1^{[4]}}{2q_1} - \frac{q_1^{[3]}q_1^{[1]}}{q_1^2} - \left(\frac{q_1^{[1]}}{q_1} \right)^2 + \frac{q_1^{[2]}(q_1^{[1]})^2}{q_1^3} + \frac{q_1^{[2]}}{q_1} + \frac{3}{4}. \quad (36)$$

So q_1 depends on its first four derivatives, consistent with the results of table 4, and similarly to previous examples we also observe weak dependence on higher order derivatives of q_1 due to ambiguities and nonlinearities in (36). However, it is apparent from (35b) that some of the information contained in the higher order derivatives of q_1 is contained much more succinctly in the function q_2 and its derivatives. On the other hand, it is not obvious from examination of (35) and (36) that knowledge of q_2 and its derivatives increases our ability to predict q_1 above the predictability gained by knowledge of the first four derivatives of q_1 . Nevertheless, that is what our test indicates. We see here again that the dependences indicated empirically by the ζ_j cannot be too closely identified with those implied by an analytic representation of the process generating the data, as we shall discuss below.

4.5. The forced Brussels model

Next we examine an example of a forced nonlinear oscillator, known as the forced Brussels

model [8] or more commonly the forced Brusselator. This model is described by the equations

$$x^{[1]} = x^2y - (B + 1)x - A - a \cos(\omega t), \quad (37a)$$

$$y^{[1]} = Bx - x^2y. \quad (37b)$$

We choose parameters $A = 0.4$, $B = 1.2$, $a = 0.05$, and $\omega = 0.95$ for which solutions to (37) are chaotic. Using parameter values $\tau = 1.0$, $\mu = 0.5$, and $\Delta t = 10^{-3}$ so that $R_k \geq 4.0$ we obtained the values for ζ_j and S_j , $j = 1, \dots, 10$ shown in table 5. Note the strong indication of dependence on the second derivative, consistent with the second order system (37). Note also that as in the previous examples, there is weaker dependence on higher order derivatives. However, the predictability S_{10} is only 0.566. While this is still larger than $S_0 = 0.271$, indicating that $x(t)$ depends strongly on its derivatives, it is smaller than the eventual predictability in the previous examples of sections 4.2 through 4.4. The predictability is smaller in this case because in the forced Brusselator, $x(t)$ depends not only on its derivatives, but also depends explicitly on time through the term $\cos(\omega t)$. The ζ_j indicate only dependence of $x(t)$ on a small number of its derivatives and in many cases may be unable to account for the effects of explicit time dependence in the underlying dynamics. However, we may derive indicators for the dependence of $x(t)$ on functions other than its own derivatives, just as we did in the analysis of the Hénon–Heiles equations.

In order to analyze the explicit time dependence in this equation, we formed the vectors

$$\mathbf{u}(t) = (x(t), \cos(ft)), \quad (38)$$

and plotted the value of ζ_1 for $0 < f \leq 4.0$. In every case we chose parameter values $\tau = 1.0$, $\mu = 0.5$, and $\Delta t = 10^{-3}$ ($R_k \geq 4.0$). The plot of ζ_1 versus f is shown in fig. 1. The five largest peaks occur at $f = \omega$, 2ω , $\frac{5}{2}\omega$, 3ω , and $\frac{7}{2}\omega$. Additional peaks occur at values $f = (n/m)\omega$, where n and m are integers. At first it may be surprising that

Table 5

ζ_j and predictability index for the forced Brussels oscillator. The first two columns show the dependence of $x(t)$ on its first ten derivatives. Row six of the last two columns (marked with an asterisk) shows the marginal dependence of $x(t)$ on the function $\cos(0.95t)$. The last four rows of the last two columns show the marginal dependence of $x(t)$ on $y(t)$ and its first three derivatives. Hence, in the last two columns ζ_{7+j} and S_{7+j} refer to dependence on $y^{[j]}$, $j = 0, \dots, 4$.

j	Dependence of $x(t)$ on its derivatives		Dependence of $x(t)$ on $y(t)$ and its derivatives	
	ζ_j	S_j	ζ_j	S_j
0		0.271		0.271
1	3.72×10^{-2}	0.281	3.72×10^{-2}	0.281
2	0.470	0.530	0.470	0.530
3	4.35×10^{-2}	0.554	4.35×10^{-2}	0.554
4	3.33×10^{-2}	0.573	3.33×10^{-2}	0.573
5	4.38×10^{-3}	0.576	4.38×10^{-3}	0.576
6	-3.19×10^{-2}	0.558	0.147*	0.685*
7	8.07×10^{-3}	0.562	9.24×10^{-2}	0.754
8	9.30×10^{-3}	0.567	0.185	0.925
9	5.18×10^{-3}	0.571	-1.51×10^{-4}	0.925
10	-7.37×10^{-3}	0.566	-7.62×10^{-4}	0.924

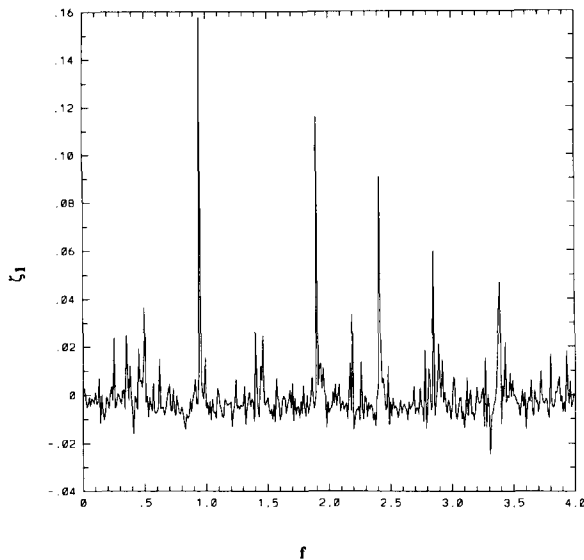


Fig. 1. Dependence of a solution to the forced Brussels equation on $\cos(ft)$.

we get peaks at rational multiples of ω rather than just integer multiples of ω . Recall, though, that our statistics indicate whether one time series is functionally dependent on another time series. This is quite different than the concept of orthogonality of functions indicated by Fourier expansions. $\cos(\omega t)$ and $\cos[(n/m)\omega t]$ are commensurate and therefore functionally dependent

on each other, but they are also, in general, orthogonal so that their relative Fourier amplitude is zero. Another important point is that any family of periodic functions may be used to probe dependence in this map. For example replacing $\cos(ft)$ in (38) by $\text{mod}(t, 2\pi/f)$ does not significantly alter our results. This is a strong indication that, while $x(t)$ is not itself periodic, it depends on a periodic function (not necessarily sinusoidal) of period $\omega = 0.95$.

Finally, we formed the vectors

$$\begin{aligned}
 \mathbf{u}(t) = & (x(t), x^{[1]}(t), \dots, x^{[5]}(t), \\
 & \cos(\omega t), y(t), y^{[1]}(t), \dots, y^{[5]}(t)). \quad (39)
 \end{aligned}$$

Table 5 lists the results of applying our test to these vectors using the same parameters as before. Notice that the predictability increases sharply from 0.576 to 0.685 after inclusion of the function $\cos(\omega t)$. Hence in this case, knowledge of the underlying periodicity in the dynamics increases our ability to predict $x(t)$ above the ability to predict $x(t)$ gained by knowledge of the first five derivatives of x . Notice also, that with inclusion of $y(t)$ and its first five derivatives, the predictability index increases to 0.924, a value much greater than the predictability gained by

knowledge of the derivatives of $x(t)$ alone. Again we see that the conditional probabilistic dependence of a function as measured by the ζ_j is related to, but not identical with, the dependences implied by an analytic representation of the underlying deterministic process. See section 5 for further discussion.

4.6. The Mackey–Glass delay equation

Next we study a system introduced by Mackey and Glass [9] as a model for the regeneration of blood cells in leukemia patients. It is

$$x^{(1)}(t) = \frac{ax(t - \rho)}{1 + [x(t - \rho)]^{10}} - bx(t). \tag{40}$$

This equation is different from those studied thus far because of the presence of the time delay ρ . Usually, the parameters a and b are fixed at the values $a = 0.2$ and $b = 0.1$ and the behavior of the solutions is studied as the delay ρ is varied. For this paper we will study the case $\rho = 17.0$. We performed the integration of this equation following the methods of Grassberger and Procaccia [2]. Table 6 shows the ζ_j and S_j , $j = 1, \dots, 7$ for the variable $x(t)$. We generated these results using parameter values $\tau = 1.0$, $\mu = 0.5$, and $\Delta t = 0.08$, giving $R_k \geq 4.0$, for $k = 1, \dots, 7$. Note the large value of ζ_1 and the smaller but still significant value of ζ_2 . Note also the low value of S_5 compared to the value of S_3 for the systems discussed in subsections 4.2 through 4.4. This value is indicating that the data are not completely explained by dependence of $x(t)$ on its derivatives. The time delay dependence in (40)

Table 6
 ζ_j and predictability index for the Mackey–Glass equation.

j	ζ_j	S_j	j	ζ_j	S_j
0		0.271	4	4.5×10^{-3}	0.390
1	0.121	0.308	5	3.2×10^{-3}	0.392
2	0.183	0.377	6	-3.2×10^{-4}	0.392
3	2.9×10^{-2}	0.389	7	4.4×10^{-3}	0.393

suggests that some ideas developed in SG for discrete time series may be useful here, when combined with the more general methods developed in this paper.

To examine this possibility, we first study the dependence of $x(t)$ only on $x(t - \tau)$. To do this we varied the sampling rate τ from 1.0 through 30.0. For each value of τ we form a discrete series $y_i = x(i\tau)$, and from this series we form d -dimensional vectors as in (3), so that

$$v_i = (x(i\tau), x((i - 1)\tau), x((i - 2)\tau), \dots, x((i - d + 1)\tau)).$$

In SG we derived indicators $\delta_j(\epsilon)$ of dependence among the components of these vectors, just as the indicators $\zeta_j(\mu)$ of this paper signal dependence among components of vectors such as (4). Choosing the tolerance ϵ equal to one half the standard deviation of the sequence $\{y_i\}$, we plot in figs. 2 through 5 δ_1 through δ_4 as a function of τ . Observe in fig. 2, the initial peak of δ_1 at small τ . This peak simply indicates that if we have knowledge of values of $x(t')$ to within ϵ , which are nearby (in time) to $x(t)$, we can accurately predict the value of $x(t)$ to within ϵ , i.e., the function $x(t)$ is relatively slowly varying over time intervals corresponding to small values of τ . This interpretation is supported by fig. 6, in which we plot 10000 pairs of points $(x(t - 3), x(t))$. This figure shows a band of width approximately 0.2 oriented along the line $x(t) = x(t - 3)$. If we choose ϵ greater than about 0.1, then this plot becomes virtually indistinguishable to our statistics from a plot of the function $x(t) = x(t - 3)$. Indeed, when $\epsilon = 0.11$ (one half the standard deviation of the function $x(t)$), the predictability index $S(1)$ is about 0.7 at $\tau = 3$, and it rises steeply to near 1.0 as τ becomes smaller. Hence, this initial peak marks a relatively uninteresting regime of linear dependence.

Before discussing the more interesting non-linear delay dependence of this system, it is worthwhile to comment on the difference be-

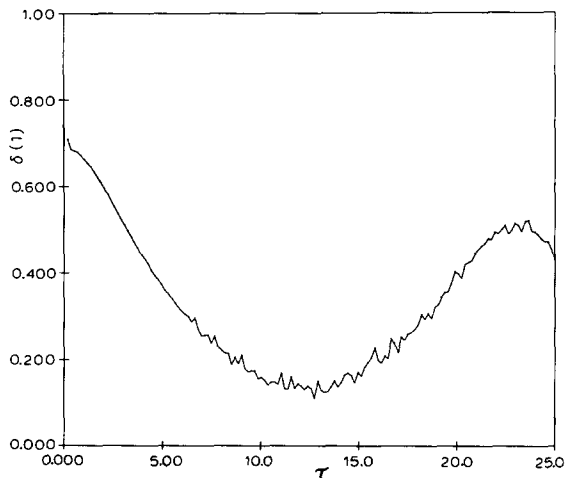


Fig. 2. δ_1 versus sampling rate τ , for the Mackey–Glass equations with delay $\rho = 17.0$.

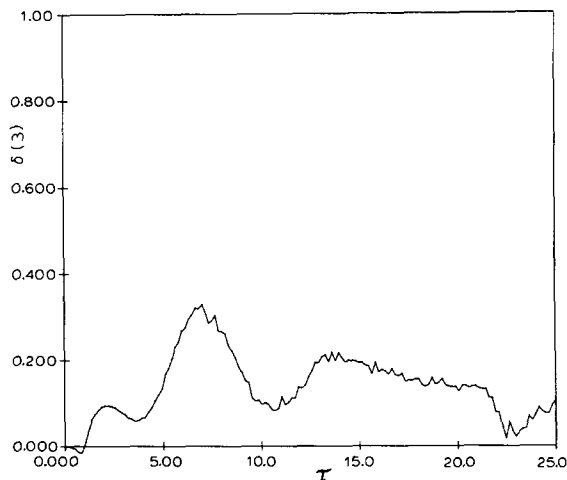


Fig. 4. δ_3 versus sampling rate τ , for the Mackey–Glass equations with delay $\rho = 17.0$.

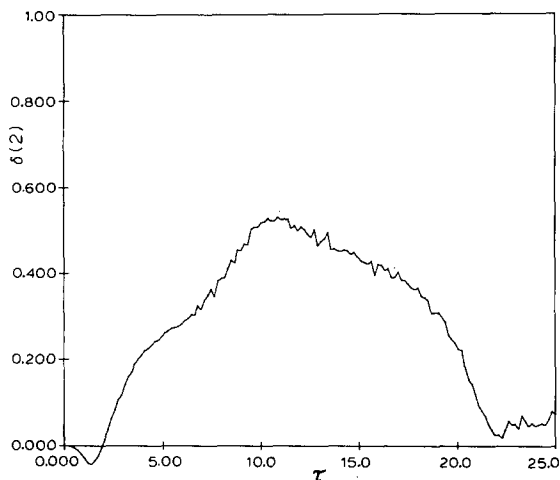


Fig. 3. δ_2 versus sampling rate τ , for the Mackey–Glass equations with delay $\rho = 17.0$.

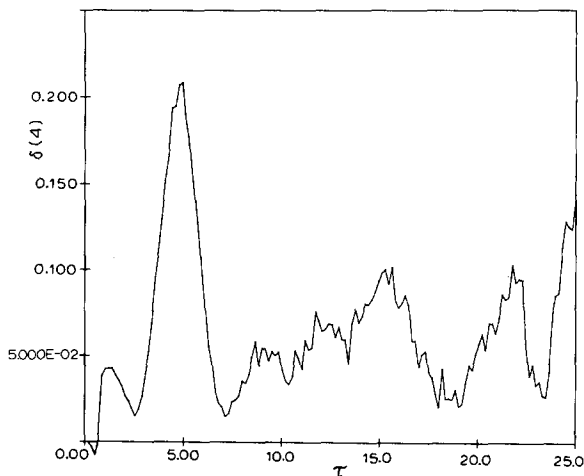


Fig. 5. δ_4 versus sampling rate τ , for the Mackey–Glass equations with delay $\rho = 17.0$.

tween the use of a lagged variable with a small value of τ , and a derivative. These two variables are related by a linear transformation. Nevertheless, the values of the ζ_j computed with these two different choices of dependent variables may be markedly different. The reason is not difficult to see. In the case of a continuous time series nearby values of the series are close in value, so that a plot of $x(t)$ versus $x(t - \tau)$ lies in a narrow diagonal band, as, for example, in fig. 6. On the

other hand, a plot of $x(t)$ versus $x^{[1]}(t)$ is not necessarily so confined, and may range over the entire two-dimensional plot. Although both plots may contain essentially the same information, more details may be easily apparent in the unconstrained plot of $x(t)$ versus $x^{[1]}(t)$. For given values of μ and N , therefore, the values of ζ_j may be quite different for the two choices of independent variables. These comments also apply to other cases in which different choices of

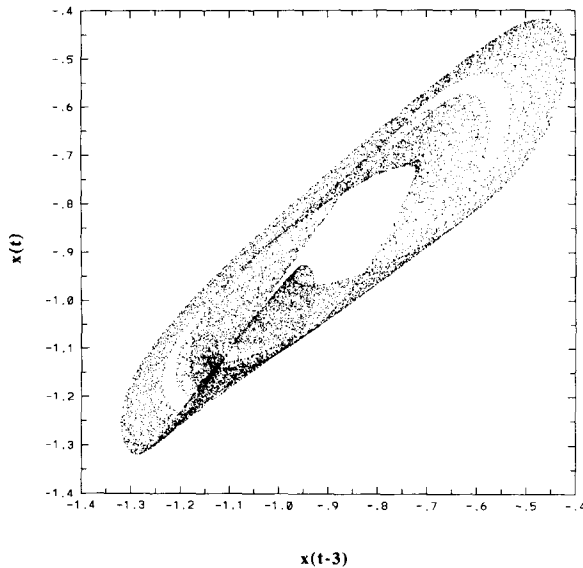


Fig. 6. Phase space plot of a solution to the Mackey-Glass equation, $x(t)$ versus $x(t-3)$.

variables which are apparently simply related may significantly distort the locus of points so that information is easier or harder to obtain.

Turning now to the nonlinear delay dependence, we observe in figs. 2 through 5 that δ_j has a broad peak near $\tau = 22/j$, $j = 1, \dots, 4$ and also that some of the δ_j exhibit more than one peak^{#2}. Since δ_j indicates dependence on $j\tau$, this is, in part, an indication that $x(t)$ depends on $x(t-22)$. To understand further the origin of these peaks, recall that since the δ_j are based on conditional probabilities, δ_j does not indicate dependence on the j th lag that is merely induced by dependencies on previous lags. Such previous lag dependence is already accounted for by one or more of the indicators δ_i , $i = 1, \dots, j-1$. Therefore, in order to observe a peak of δ_j at a value of the sampling rate, τ , we require not only that $x(t)$ depend on $x(t-j\tau)$ but also that this dependence be new information and not simply induced by dependence of $x(t)$ on $x(t-i\tau)$, $i = 1, \dots, j-1$. As an example note that δ_3 (fig. 4) as

^{#2}Note that the precise location of the peak varies from $\tau = 21/j$ to $\tau = 23/j$ for different values of j . We use the value of 22 in the text primarily to simplify notation.

a function of τ has two peaks, one at $\tau = 7 \cong \frac{22}{3}$, and also one at $\tau = 14 \cong 2 \times \frac{22}{3}$. To understand the origin of the first of these peaks, we examine the indicators δ_j , for $j = 1, 2, 3$ at the fixed value $\tau = 7$. Both δ_1 and δ_2 are small at $\tau = 7$ because $x(t)$ has no strong dependence on $x(t-7)$ or $x(t-14)$. In this case there is little dependence on the first or second lag and hence little induced dependence on the third lag, so the dependence of $x(t)$ on $x(t-21)$ is indicated by the large value of δ_3 at $\tau = 7$. Similarly, if we examine the indicators δ_j , for $j = 1, 2, 3$ at the value $\tau = 14$, we see that δ_1 and δ_2 are small because $x(t)$ does not depend strongly on $x(t-14)$ or $x(t-28)$. Accordingly, δ_3 is large at $\tau = 14$ because $x(t)$ does depend on $x(t-42)$ and this dependence is not induced by dependence on the previous lags. By contrast, there is no peak of δ_2 at $\tau = 22$ even though $x(t)$ depends on $x(t-44)$. This is because most of the dependence of $x(t)$ on $x(t-44)$ is induced by the dependence of $x(t)$ on $x(t-22)$, which is already accounted for by the large value of δ_1 at $\tau = 22$. The pattern of peaks of δ_j for $j > 3$ can be understood in a similar manner.

These results indicate that of the class of lag dependencies, $x(t)$ depends most strongly on $x(t-22)$. This is somewhat surprising given the fact that the series was generated with explicit dependence on $x(t-17)$. To make matters even more puzzling, we note that we have studied the lag dependencies of this series for smaller values of μ . Even if we set $\mu = 0.01$, so that the tolerance with which we observe the sequence is only one one-hundredth of the standard deviation, we still have the largest lag dependence on $x(t-22)$.

We can get some intuitive understanding of this result by looking at figs. 7-9. In these figures we have plotted 10000 pairs of points $(x(t-\tau), x(t))$, where the sampling rate is $\tau = 13, 17$ and 22 for figs. 7, 8, and 9, respectively. Roughly speaking, these plots appear to show a (more or less) rigid body rotating as a function of τ in some space. The projection on the $(x(t-\tau), x(t))$ plane is significantly narrower when $\tau = 22$ than when $\tau = 13$ or 17 , so that knowledge of the value

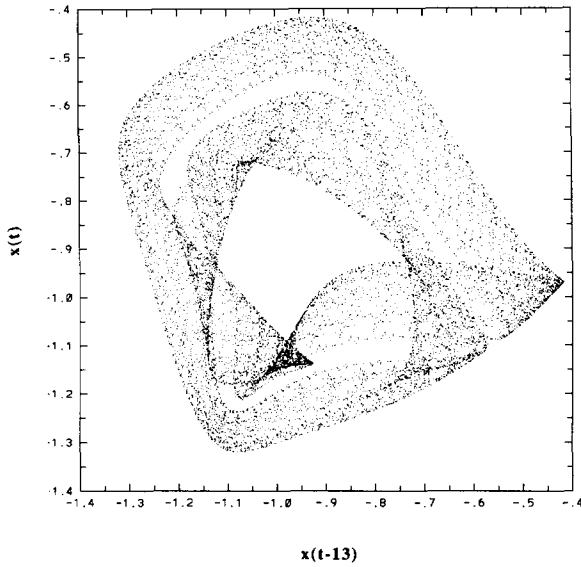


Fig. 7. Phase space plot of a solution to the Mackey–Glass equation, $x(t)$ versus $x(t - 13)$.

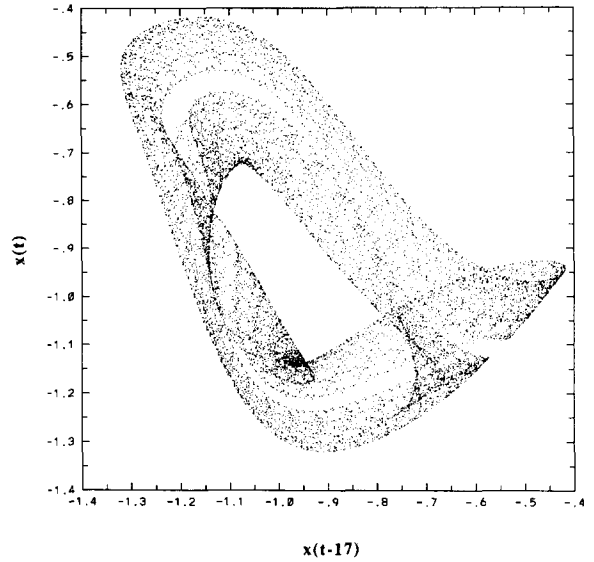


Fig. 8. Phase space plot of a solution to the Mackey–Glass equation, $x(t)$ versus $x(t - 17)$.

of $x(t - 22)$ will lead to a more accurate prediction of $x(t)$.

This visual representation suggests that the apparent paradox may be related to missing information which would be represented by one or more additional dimensions in the phase space of variables. Since the analytic form of the Mackey–Glass equation has explicit dependence on $x^{[1]}(t)$, we are led to study the dependence of $x(t)$ on the two variables, $x^{[1]}(t)$ and $x(t - \tau)$. In particular, let us ask how predictable $x(t)$ is given the information contained in $x^{[1]}(t)$ and $x(t - \tau)$ for different values of τ . To do this, we consider the 3-vector

$$u(t) = (x(t), x^{[1]}(t), x(t - \tau))$$

for τ between 0 and 30. Fig. 10 shows $\zeta(2)$ as a function of τ , and fig. 11 shows the predictability index, $S(2)$ as a function of τ , for the case $\mu = 0.5$. In this case, the maximum predictability is still at $\tau = 22$. However, if we decrease μ , the peak predictability shifts toward 17, as shown in figs. 12 and 13. Note that even when we include informa-

tion about the differential dependence, we still gain predictability through knowledge of values of $x(t')$ for t' close to t , as indicated by the peaks in these figures at $\tau = 0$. Further, observe that the initial peak in the predictability index is much narrower when $\mu = 0.01$ (fig. 13) than when $\mu =$

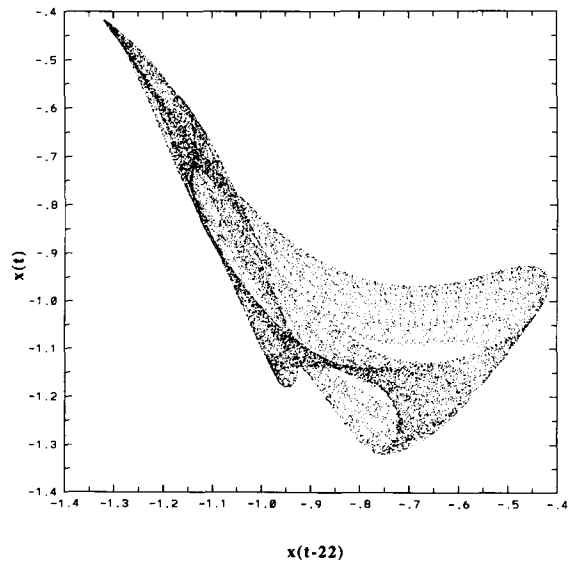


Fig. 9. Phase space plot of a solution to the Mackey–Glass equation, $x(t)$ versus $x(t - 22)$.

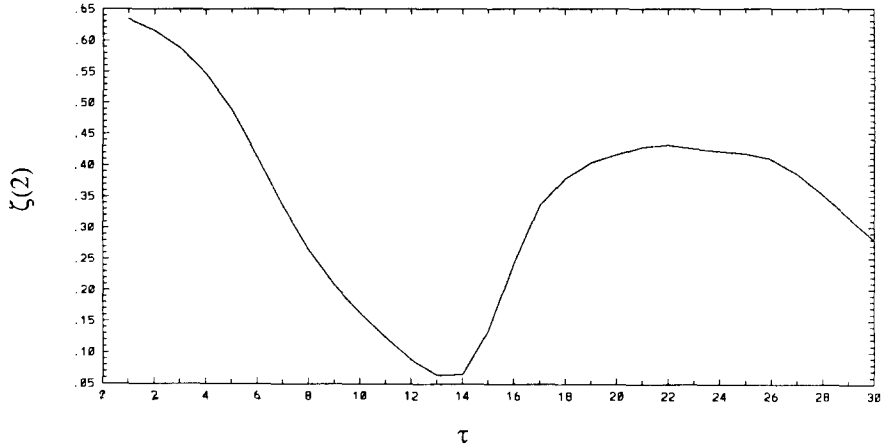


Fig. 10. Dependence of a solution, $x(t)$, of the Mackey–Glass equation on $x^{[1]}(t)$ and on $x(t - \tau)$. This figure shows $\zeta(2)$ versus τ for the case where the tolerance $\mu = 0.5$.

0.5 (fig. 11). This supports our interpretation of this peak as an indicator of continuity of the time series.

It is worthwhile discussing these results at some length. First, it is important to realize that even in the case in which a time series is generated by an analytically expressible process, such as eq. (40), the empirical analysis of the series may not optimally involve the same variables that appear in the analytic form itself. There may be several

reasons why this is so. First, one may not have chosen a set of variables to use to study the dependence which includes all those variables of the underlying analytic form. This was the case in our first analysis of the Mackey–Glass process, in which we did not include dependence on $x^{[1]}(t)$. Second, even if the set of variables includes all those explicitly contained in the analytic form, the maximum information from a given number of those variables may not come from the set that

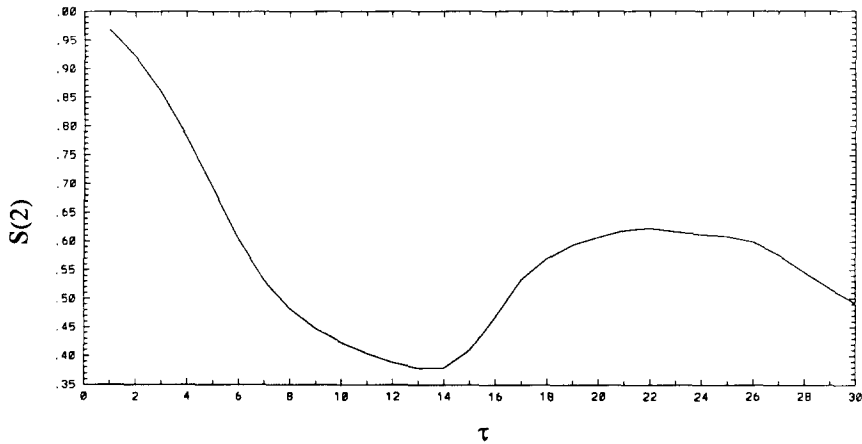


Fig. 11. Dependence of a solution, $x(t)$, of the Mackey–Glass equation on $x^{[1]}(t)$ and on $x(t - \tau)$. This figure shows $S(2)$ versus τ for the case where the tolerance $\mu = 0.5$.

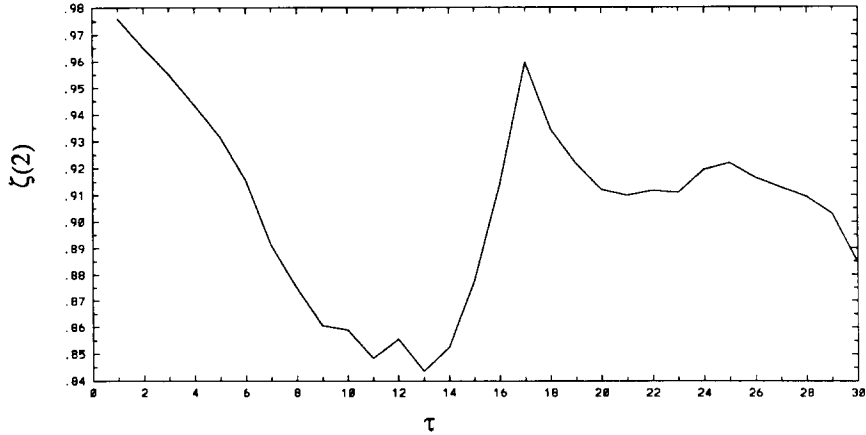


Fig. 12. Dependence of a solution, $x(t)$, of the Mackey–Glass equation on $x^{(1)}(t)$ and on $x(t - \tau)$. This figure shows $\zeta(2)$ versus τ for the case where the tolerance $\mu = 0.01$.

appear in the analytic expression. We see this in fig. 10 for the Mackey–Glass system. In this case, the reason is that the fuzziness in one’s analysis of the process (induced by a nonzero value of μ) may induce a qualitative restructuring of one’s (fuzzy) perception of the attractor. What must be true, though, is that in the limit μ goes to zero, the predictability of a noiseless analytically generated time series cannot exceed the predictability

computed using the set of independent variables in the completely deterministic analytic form. The broader implications of these observations will be discussed in the next section.

Finally, we remark that since a combination of the methods of SG and this paper explain the dependence in the Mackey–Glass equation very well, we are lead to investigate the possibility that there is a way to incorporate both of these meth-

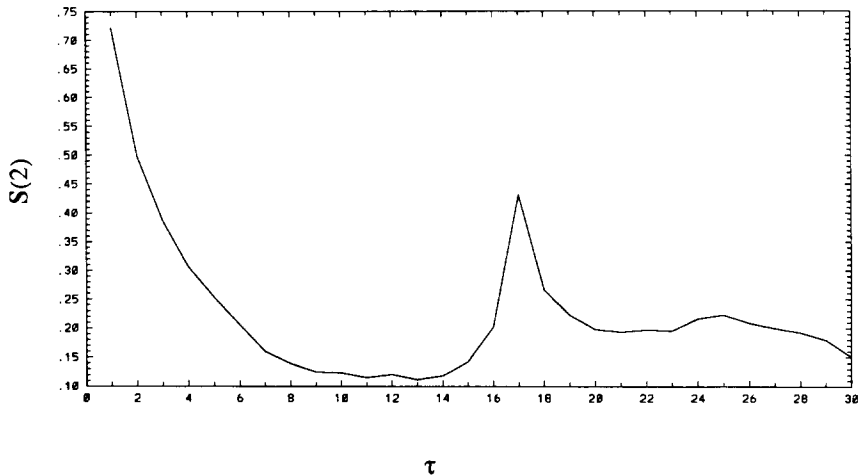


Fig. 13. Dependence of a solution, $x(t)$, of the Mackey–Glass equation on $x^{(1)}(t)$ and on $x(t - \tau)$. This figure shows $S(2)$ versus τ for the case where the tolerance $\mu = 0.01$.

ods into a generalized formalism. This is the subject of the appendix.

5. Summary and conclusions

In this paper we have presented an extension of the methods of SG to cover the case of continuous time series, and more generally, interdependence among different time series. In several examples of both chaotic and periodic functions generated by low order systems of differential equations, we have demonstrated that these methods are capable of identifying dependence of a function on its derivatives. In addition, we have derived a predictability index which indicates quantitatively the extent to which a time series has an underlying deterministic explanation based on the choice of independent functions included in the analysis. We have also explored the flexibility and generality of these methods by demonstrating their usefulness in indicating nonlocal effects and in probing dependence among functions which are not simply derivatives or integrals of each other.

One other very important feature of our statistics is that it is possible to analytically derive their asymptotic behavior, in the limit of a very large data set, under null hypotheses of statistical independence. For example, the asymptotic behavior of the statistics discussed in SG have been derived under the null hypothesis that the time series is IID (independent and identically distributed). This will be discussed in detail elsewhere.

The methods of this paper and SG advance our ability to analyze complicated experimental data which appear random to standard linear statistical methods. It is clear that this new approach has wide potential application in the general area of time series analysis including the study of chaotic physical and biological systems. In addition, we believe that the methods are applicable to noise reduction and signal analysis. We are also using them to study other mathematical ex-

amples of interest including quasi-periodic functions.

It is obvious that our methods provide useful information about the dependencies of broad band systems. But one might also be led to see in this study more profound epistemological implications: Historically, most systems studied by the physical sciences have had simple output and have been simple in description. Indeed, the search for simplicity has been a hallmark of much scientific research for centuries. Whether that point of view was engendered by human limitations, by historical or cultural accident, or by other causes is a question we do not wish to address here. But there is no question that such a view has been a guiding force in much of our most basic scientific inquiry. The recognition that systems may have intrinsically complex behavior, and that that is an interesting thing to study is fairly recent.

Much of the work on systems with complex behavior has focused on those systems which are simple in description, often having an underlying analytic mathematical form. While it is surely of great importance to understand how complexity can arise from intrinsically simple systems, it should be recognized that such systems are probably a set of measure zero in the space of physical, biological and social systems with interesting complex behavior. Given that most systems with this kind of behavior may not have a simple analytic explanation, one is led to ask in what terms such systems should be described. It is our view that the most fundamental description of a process must proceed directly from the analysis of the output of the process itself, and that the description will be conditioned on the precise question one is asking.

Consider the problem of analyzing a general broad band time series. Such a series, if it contains nonrandom effects, may be generated by an underlying process with deterministic elements which may not be representable analytically, and which also contains noise. In general, one's knowledge of the underlying process will be lim-

ited. One might be guided by some general notions about what the important variables might be, but such insights are likely to be only partial. Since, except in rare cases, one cannot know a priori whether the system with which one is dealing has a simple explanation, the most productive course may be to attempt to characterize the system from the data themselves. It is certainly true (as we have shown in our examples) that for such characterizations to be most effective some understanding of the relevant variables in terms of which to describe the system is important. We believe that, generally, this is the role of model building for complex systems. It is usually vain to expect that a model will accurately represent and explain the data generated by a complex system. Instead, the insights gained from studying models can be useful in helping to identify possible relevant variables. This attitude is much different from that traditionally adopted in the physical sciences, in which models built to describe relatively simple physical systems are taken very seriously indeed. There are of course cases in which an underlying simple analytic (or even algorithmic) form is possible. If so, such a form will generally be preferable because of its economy and, often, aesthetic appeal. Nevertheless, such cases are probably quite rare in the space of all broad band systems. Generically, a description of the system based on indicators that encapsulate the degree of dependence of the series on a set of independent variables, supplemented, perhaps with algorithms of predictive power, or methods of control, may be the most fundamental description possible.

We consider the results of our analysis of the Mackey–Glass equation as strong support for this

point of view. In that case, even though the time series was generated by a noiseless analytic process, the variables that contained the most information about the time series were not necessarily the independent variables that appear in the Mackey–Glass equation. Even when the space of variables was enlarged to include the derivative of the time series, the optimum choice of dependent variables for the purposes of empirical analysis depended on the choice of ε . Thus, the most efficient description of even this analytically generated process depends on the precise empirical circumstances and the precise question being asked. If this is true in a case in which the underlying process has an analytic explanation, how much more true is it in the general case in which such a simple explanation is not possible. We cannot emphasize this conclusion too strongly: The most efficient description of a system with complex behavior depends on the precise question being asked and the precise context of that question. Even if the system has an underlying simplicity (as in the Mackey–Glass case), the best description of the system for a given purpose may not involve those variables that generated the complex behavior in the first place.

It is important to realize that this point of view is not a position adopted as the result of philosophical speculation. Instead, it appears forced on us by the recognition that as our questions and tools to analyze the world become more sophisticated, and as the systems we study become more complex, our ability to simplify them in the terms to which we have become used, while at the same time retaining the features which we perceive to be the most important, may be limited.

Appendix

At the end of section 4.6, we noted the possibility of developing a set of generalized indicators of dependence of which the indicators of this paper and the indicators of SG would become special subsets. In this appendix we describe one possibility for such a generalization.

Suppose we have a set of functions, $x^{[j]}(t)$, $j = 1, 2, \dots$. For the purposes of this appendix, in order to simplify notation we will let the set $\{x^{[j]}(t), j = 1, 2, \dots\}$ stand for an arbitrary collection of smooth

functions and derivatives of those functions, instead of simply successive derivatives of one function as defined in the main body of the paper. We may then choose a sampling interval τ , and form the matrices $\mathbf{W}(I, J, n)$, $1 \leq i \leq N$,

$$\mathbf{W}(I, J, n) = \begin{pmatrix} x(n\tau) & x(n\tau + \tau) & x(n\tau + 2\tau) & \dots & x(n\tau + I\tau) \\ x^{[1]}(n\tau) & x^{[1]}(n\tau + \tau) & x^{[1]}(n\tau + 2\tau) & \dots & x^{[1]}(n\tau + I\tau) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x^{[J]}(n\tau) & x^{[J]}(n\tau + \tau) & x^{[J]}(n\tau + 2\tau) & \dots & x^{[J]}(n\tau + I\tau) \end{pmatrix}, \tag{A.1}$$

where I denotes the number of delays included in the matrix and J denotes the number of functions. These matrices replace the vectors v_n of SG and $u(n\tau)$ of this paper. In fact, the vector v_n of dimension d is just $\mathbf{W}(d - 1, 0, n)$, and the vector $u(n\tau)$ of dimension d is $\mathbf{W}(0, d - 1, n)$. Hence, in our generalization, $(I + 1)(J + 1)$ replaces the dimension of the vectors in the two previous formalisms. In the following we will use the notation $W_{ij}(n)$ to denote the (i, j) component of the matrix $\mathbf{W}(I, J, n)$ with $I \geq i$ and $J \geq j$. Note that we do not need to specify the dimensions I, J when we denote the component $W_{ij}(n)$ since for all $I \geq i$, and $J \geq j$, $W_{ij}(n) = x^{[j]}((n + i)\tau)$.

Next, we must develop an indicator function and correlation integrals for the matrices \mathbf{W} . In order to do this we must first define a tolerance ε_{ij} for each element of the matrices. In analogy with the definition of ε_k in this paper, we define ε_{ij} to be a fraction μ of the standard deviation of the set $\{x^{[j]}((n + i)\tau), n = 1, \dots, N\}$. However, following SG note that when $N \gg I$ the standard deviation of this set is approximately independent of the value of i , so that for all values of i , $\varepsilon_{ij} \approx \varepsilon_{1j}$. Accordingly, we define a single value of ε for each row of $\mathbf{W}(I, J, n)$. This allows us to define the indicator function,

$$I_{nm}(\mu) = \prod_{j=0}^J \prod_{i=0}^I \Theta(\varepsilon_j - |W_{ij}(n) - W_{ij}(m)|), \tag{A.2}$$

and the correlation integral,

$$C_{IJ}(\mu) = \frac{1}{\nu} \sum_{\text{pairs}} I_{nm}(\mu), \tag{A.3}$$

where ν is the number of pairs of matrices $(\mathbf{W}(I, J, n), \mathbf{W}(I, J, m))$, and ε_j is the product of μ and the standard deviation of the set $\{x^{[j]}(n\tau), n = 1, \dots, N\}$.

We wish to develop statistics which indicate dependence of $W_{00}(n)$, on the other elements of $\mathbf{W}(I, J, n)$. In particular, if $\mathbf{M}(I, J, n)$ is the set containing those elements of $\mathbf{W}(I + 1, J + 1, n)$ which are not also in $\mathbf{W}(I, J, n)$ (see fig. 14), we wish to investigate the marginal increase in our ability to predict the value of W_{00} due to knowledge of some or all of the elements of $\mathbf{M}(I, J, n)$. Let P be a subset of $\{(I + 1, j)|j = 0, \dots, J + 1\} \cup \{(i, J + 1)|i = 0, \dots, I + 1\}$ and let $\mathbf{M}(P, n)$ stand for the corresponding subset of $\mathbf{M}(I, J, n)$. We may then generalize (A.2) by defining

$$I_{nm}(\mu) = \prod_{T(I, J) \cup P} \Theta(\varepsilon_j - |W_{ij}(n) - W_{ij}(m)|), \tag{A.4}$$

where $T(I, J) = \{(i, j)|0 \leq i \leq I \text{ and } 0 \leq j \leq J\}$. And we may generalize (A.3) by defining the correlation

$$\begin{pmatrix} \mathbf{x}(n\tau) & x(n\tau + \tau) & \dots & x(n\tau + I\tau) & \mathbf{x}(n\tau + (\mathbf{I} + \mathbf{1})\tau) \\ x^{[1]}(n\tau) & x^{[1]}(n\tau + \tau) & \dots & x^{[1]}(n\tau + I\tau) & \mathbf{x}^{[1]}(n\tau + (\mathbf{I} + \mathbf{1})\tau) \\ \vdots & & & & \vdots \\ x^{[J]}(n\tau) & x^{[J]}(n\tau + \tau) & \dots & x^{[J]}(n\tau + I\tau) & \mathbf{x}^{[J]}(n\tau + (\mathbf{I} + \mathbf{1})\tau) \\ \mathbf{x}^{[J+1]}(n\tau) & \mathbf{x}^{[J+1]}(n\tau + \tau) & \dots & \mathbf{x}^{[J+1]}(n\tau + I\tau) & \mathbf{x}^{[J+1]}(n\tau + (\mathbf{I} + \mathbf{1})\tau) \end{pmatrix}$$

Fig. 14. The matrix $\mathbf{W}(I + 1, J + 1, n)$ with the margin set $\mathbf{M}(I, J, n)$ shown in bold face. Our generalized statistics indicate the increase in the ability to predict $x(n\tau)$ due to knowledge of some specific subset of $\mathbf{M}(I, J, n)$ over our ability to predict $x(n\tau)$ due to knowledge of the elements shown in italics above.

integral

$$C_{IJ+P}(\mu) = \frac{1}{\nu} \sum_{\text{pairs}} I_{nm}(\mu). \tag{A.5}$$

Next we wish to define the conditional probability that $|W_{00}(n) - W_{00}(m)| \leq \epsilon_0$ given that $|W_{ij}(n) - W_{ij}(m)| \leq \epsilon_j$ for all $(i, j) \in T(I, J) \cup P - \{(0, 0)\}$. Let Q be any set of ordered pairs (i, j) and define $D(Q)$ to be the statement “ $|W_{ij}(n) - W_{ij}(m)| \leq \epsilon_j$ for all $(i, j) \in Q$ ”. Then in order to write the conditional probabilities in terms of the correlation integrals we may use the identities

$$P(D\{(0, 0)\} | D(T(I, J) \cup P - \{(0, 0)\})) = \frac{P(D(T(I, J) \cup P))}{P(D(T(I, J) \cup P - \{(0, 0)\}))} \tag{A.6}$$

and

$$P(D(T(I, J) \cup P)) = C_{IJ+P}(\mu). \tag{A.7}$$

We must also define new correlation integrals

$$C_{IJ-0+P}(\mu) = P(D(T(I, J) \cup P - \{(0, 0)\})). \tag{A.8}$$

Hence,

$$P(D\{(0, 0)\} | D(T(I, J) \cup P - \{(0, 0)\})) = \frac{C_{IJ+P}(\mu)}{C_{IJ-0+P}(\mu)}. \tag{A.9}$$

If knowledge of $D(P)$ does not increase our ability to predict the value of W_{00} above our ability to predict W_{00} given knowledge of $D(T(I, J))$, then

$$P(D\{(0, 0)\} | D(T(I, J) \cup P - \{(0, 0)\})) = P(D\{(0, 0)\} | D(T(I, J) - \{(0, 0)\})), \tag{A.10}$$

or using (A.9) to write (A.10) in terms of correlation integrals,

$$\frac{C_{IJ+P}(\mu)}{C_{IJ-0+P}(\mu)} = \frac{C_{IJ}(\mu)}{C_{IJ-0}(\mu)}. \tag{A.11}$$

Finally, we may define the generalized indicator

$$\xi_P(\mu) = 1 - \frac{C_{IJ}(\mu) C_{IJ-0+P}(\mu)}{C_{IJ-0}(\mu) C_{IJ+P}(\mu)}. \quad (\text{A.12})$$

The size of $\xi_P(\mu)$ is an indication of the degree to which knowledge of $D(P)$ increases our ability to predict W_{00} above our ability to predict W_{00} gained by knowledge of $D(T(I, J) - \{(0, 0)\})$. Note that these generalized statistics reduce to the statistics of SG in the case $P = \{(i, 0) | 1 \leq i \leq d - 1\}$ and to the statistics of this paper in the case $P = \{(0, j) | 1 \leq j \leq d - 1\}$.

We may also define a generalized predictability index, S . Following the development of the predictability index in section 3, we define

$$S_{IJ} = P(D\{(0, 0)\} | D(T(I, J) - \{(0, 0)\})), \quad (\text{A.13})$$

and

$$S = \lim_{I, J \rightarrow \infty} S_{IJ}. \quad (\text{A.14})$$

Using (A.12) and some algebra, it is easy to write S_{IJ} in terms of the $\zeta_M(\mu)$ as

$$S_{IJ} = \frac{C_{00}}{\prod_{i, j=0}^{I, J} \xi_{P(i, j)}}, \quad (\text{A.15})$$

where $P(i, j) = \{(i + 1, j') | j' = 0, \dots, j + 1\} \cup \{(i', j + 1) | i' = 0, \dots, i + 1\}$.

Clearly, this generalized formalism is cumbersome; however, such a formalism or a suitable variation may be necessary in order to systematically analyze very complicated systems such as the Navier–Stokes equations in regions of turbulent fluid flow.

References

- [1] R. Savit and M. Green, Time series and dependent variables, *Physica D* 50 (1991) 95–116.
- [2] W.A. Brock, W.D. Dechert and J.A. Scheinkman, A test for independence based on the correlation dimension, preprint.
- [3] P. Grassberger and I. Procaccia, Measuring the strangeness of strange attractors, *Physica D* 9 (1983) 189–208.
- [4] W. Press, B. Flannery, S. Teukolsky and W. Vetterling, *Numerical Recipes* (Cambridge Univ. Press, Cambridge, 1986) p. 550.
- [5] B. Efron and R. Tibshirani, *Stat. Sci.* 1 (1966) 54–77.
- [6] E.N. Lorenz, Deterministic nonperiodic flow, *J. Atmos. Sci.* 20 (1963) 130–141.
- [7] M. Henon and C. Heiles, The applicability of the third integral of motion: Some numerical experiments, *Astrophys. J.* 62 (1964) 73.
- [8] I. Prigogine and R. Lefever, Symmetry breaking instabilities in dissipative systems, *J. Chem. Phys.* (1968) 1695–1700.
- [9] M.C. Mackey and L. Glass, *Science* 197 (1977) 287.