PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

A Theory of Correlation Dimension for Stationary Time Series

Colleen D. Cutler

Phil. Trans. R. Soc. Lond. A 1994 **348**, doi: 10.1098/rsta.1994.0095, published 15 September 1994

A theory of correlation dimension for stationary time series†

By Colleen D. Cutler

Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

We develop a formal theory of correlation dimension for a class of stationary time series that includes both deterministic outputs and gaussian processes with continuous paths. This theory enables us to completely analyse correlation dimension in gaussian processes via spectral methods. Our approach plus recent results on the convergence behaviour of the sample correlation integral are then used to re-examine the behaviour of gaussian power-law coloured noise. We show that the finite correlation dimension observed by Osborne & Provenzale (1989) is a local quantity entirely due to the non-ergodicity of the simulation model. We also show that non-ergodic and weakly ergodic finite-dimensional dynamical systems (such as the simple circle map) exhibit the same phenomenon.

1. Introduction

In this paper we propose a formal definition and theory of correlation dimension for a class of strictly stationary real-valued time series with continuous paths. This class encompasses those time series generated by smooth functionals of invariant finite-dimensional dynamical systems (the so-called *deterministic outputs*, in which case our theory produces the usual results) but also includes more general time series generated by other means. For example, we verify that our class also includes all stationary gaussian processes with continuous paths. We examine the gaussian family in detail, and show how our approach clarifies some points in the deterministic *vs* stochastic debate.

The chief motivation for this work was the important paper of Osborne & Provenzale (1989) (hereafter called OP, in the fashion of Theiler (1991)), in which it was demonstrated that, for certain random processes with power-law spectra (often called *coloured noise,* and which can be taken to be gaussian), the sequence of numerical estimates of the correlation dimension (obtained by successively increasing the embedding dimension d) converge to a *finite value* which in fact coincides with the fractal dimension of the embedded sample path. This observation (which we call the *OP phenomenon*) was of particular significance because it contradicted a widely held assumption that correlation dimension estimates obtained from stochastically generated time series would always increase *without bound* as the embedding dimension increased. This unexpected result has

† This paper was produced from the author's disk by using the TEX typesetting system.

Phil. Trans. R. Soc. Lond. A (1994) **348**, 343–355
Printed in Great Britain

spawned a considerable literature devoted to increasingly sophisticated methods of distinguishing stochastic time series from low-dimensional chaotic dynamics (see, for example, Kaplan & Glass 1992; Kennel & Isabelle 1992; Provenzale *et al.* 1992; Smith 1992; Theiler *et al.* 1992). Recent related statistical work can be found in Cheng & Tong (1992).

Correlation dimension was initially introduced and defined (Grassberger & Procaccia 1983, hereafter denoted GP) for $\mathbb{R}^d$-valued data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ as a double limit:

$$\nu_{\mathrm{GP}} = \lim_{r \to 0} \lim_{n \to \infty} \frac{\log C_n(r)}{\log r}, \tag{1.1}$$

where the statistic $C_n(r)$ (which we call the *sample correlation integral*) was defined to be the proportion of pairs of observations no more than distance r apart. That is,

$$C_n(r) = \binom{n}{2}^{-1} \sum_{i=1}^{n} \sum_{j<i}^{n} I_{[\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_d \leqslant r]}. \tag{1.2}$$

In this paper the symbols $\| \ \|_d$ always denote the $\mathcal{L}^\infty$ or 'sup' norm on $\mathbb{R}^d$. The statistic ν_{GP} in (1.1) is actually independent of the choice of norm, but we will find the sup norm most convenient for computation, as well as the most easily extendable to the case of random continuous functions.

In the case that the data takes the form of a real-valued time series $\{X(t_k)\}_k$ (and hence may either be a stochastic process or a functional of a finite-dimensional dynamical system) the standard practice has been to attempt to reconstruct the underlying dynamics by using the time-delay embedding method in Packard *et al.* (1980) and Takens (1981). We assume that the reader is fully familiar with this procedure. Its adaptation to the case of correlation dimension, applied to a single realization of a time series, is generally called the Grassberger–Procaccia (GP) algorithm. The experimentalist constructs a sequence $\{\nu_{\mathrm{GP}}^{(d)}\}$, $d = 1, 2, \ldots$ of correlation dimensions, where $\nu_{\mathrm{GP}}^{(d)}$ is determined via (1.1) after grouping the original observations into vectors $\boldsymbol{X}_k^{(d)} = (X(t_k), X(t_{k+1}), \ldots, X(t_{k+d-1}))$ of length d and computing (1.2). The limit of this sequence

$$\nu_{\mathrm{GP}}^{(\infty)} = \lim_{d \to \infty} \nu_{\mathrm{GP}}^{(d)} \tag{1.3}$$

is often called the *correlation dimension of the time series,* and, prior to OP, was believed to indicate a deterministic output whenever it assumed a finite value. We note that Takens's (1981) embedding theorem is generally regarded as the theoretical basis for the GP algorithm, although Takens' theorem is not a statement about single realizations of time series.

In this paper we depart from the preceding standard approach in two important ways. First, in contrast to the GP method of definition, we adopt the more traditional statistical viewpoint that it is valuable to define parameters of interest in a manner *independent of data,* so that different methods of estimating the parameter (as well as different methods of obtaining data) can be compared for their ability to reproduce the parameter. Hence our definition of correlation dimension, both for random vectors and time series, is formulated in terms of probability distributions rather than data. Second, we treat a time series as a single object (a random continuous function) and make a definition of correlation dimension which seems appropriate for such objects by extending the formal mathematical approach which can be taken for random vectors in $\mathbb{R}^d$ (see next section).

This creates a significant difference in rationale between our approach and the embedding technique – the latter evolved specifically as a way of reconstructing dynamics under the assumption of an underlying deterministic system, whereas our approach makes no such assumption on the time series. Thus our approach leads to a rigorous definition of correlation dimension valid for a larger class of stationary processes. This definition, coupled with recent results on the convergence behaviour of the sample correlation integral (Pesin 1993; Serinko 1993*a*; Aaronson *et al.* 1993), provides a simple explanation of the OP phenomenon, and allows us to complete the discussion initiated by Theiler (1991). Our approach also shows that the behaviour of power-law coloured noise is simply one illustration of a general behaviour common to non-ergodic and weakly ergodic systems. We will demonstrate that uncoupled harmonic oscillators, as well as the simple circle map $T(x) = (x + \alpha) \bmod 1$, exhibit the OP phenomenon.

2. A new approach to correlation dimension in time series

A formal mathematical approach to correlation dimension (and other fractal dimensions) for distributions in $\mathbb{R}^d$ has already been taken by various authors; see Cutler (1993) for an overview, also Pesin (1993) for alternatives. Given a probability measure μ on the Borel sets of $\mathbb{R}^d$, we define the associated *spatial correlation integral* $C_\mu(r)$ by

$$C_\mu(r) = \mu \times \mu(\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^d \times \mathbb{R}^d | \|\boldsymbol{x} - \boldsymbol{y}\|_d \leqslant r\}). \tag{2.1}$$

That is, $C_\mu(r)$ represents the probability that two independent observations from μ are no more than distance r apart. This is portrayed descriptively by the alternate notation

$$C_\mu(r) = P([\|\boldsymbol{X} - \boldsymbol{Y}\|_d \leqslant r]), \tag{2.2}$$

where $\boldsymbol{X}$, $\boldsymbol{Y}$ are independent, identically distributed (IID) random vectors in $\mathbb{R}^d$, each with distribution μ. We note that the sample correlation integral defined in (1.2) might be considered the 'natural estimator' of $C_\mu(r)$, even though the vectors involved in obtaining (1.2) will generally be extracted from the same realization of a time series and hence fail to be independent. The *correlation dimension* of μ is then defined to be

$$\nu = \lim_{r \to 0} \frac{\log C_\mu(r)}{\log r} \tag{2.3}$$

assuming this limit exists. It is well known that $0 \leqslant \nu \leqslant d$.

This approach extends in the obvious way to the case of random continuous functions with compact domain. Let $T > 0$ be fixed, and let $\mathcal{C}([0, T])$ denote the space of continuous real-valued functions on $[0, T]$ equipped with the sup norm:

$$\|x - y\|_T = \sup_{0 \leqslant t \leqslant T} |x(t) - y(t)|. \tag{2.4}$$

Then, given a probability distribution μ on the Borel sets of $\mathcal{C}([0, T])$, we define, in analogy with (2.1) and (2.2), the associated spatial correlation integral by

$$C_T(r) = \mu \times \mu(\{(x, y) \in \mathcal{C}([0, T]) \times \mathcal{C}([0, T]) | \|x - y\|_T \leqslant r\}), \tag{2.5}$$

which has the alternate description

$$C_T(r) = P([\|X - Y\|_T \leqslant r]), \tag{2.6}$$

where X and Y are IID random functions in $\mathcal{C}([0,T])$, each with distribution μ. We have used the notation $C_T(r)$, rather than $C_\mu(r)$, to emphasize the dependence on the length of the time interval. The corresponding correlation dimension ν_T is then obtained via (2.3).

In the case of stationary time series, the natural domain of definition will often be $[0,\infty)$ rather than $[0,T]$. Thus one might consider evaluating the distance between two IID random functions over all $t \geqslant 0$, using a uniform metric of the form

$$\operatorname{dist}(x,y) = \min\left(\sup_{0\leqslant t<\infty} |x(t)-y(t)|,\, 1\right). \tag{2.7}$$

However, this approach is undesirable, as in most instances we would expect to find that the probability of two random functions staying uniformly close for all time is actually zero. (This is in fact the case for chaotic dynamical systems, i.e. systems with a positive Lyapunov exponent.) Instead we study the separation of paths over bounded intervals, and restrict consideration to those processes for which the value of the correlation dimension ν_T is independent of T:

Definition 2.1. A process with stationary distribution μ on $\mathcal{C}([0,\infty))$ is said to be *time-length invariant* with respect to correlation dimension if, for all $T>0$, $T'>0$, we have $\nu_T = \nu_{T'}$. This common value, which we denote by ν_∞, will be called *the correlation dimension of the process.*

We remark that we do not know if there exist stationary processes which are not time-length invariant. There is a second relevant definition we can make which will establish a clear link between our approach and (1.3):

Definition 2.2. Let $\{X(t)\}_t$ be a strictly stationary process with invariant distribution μ on $\mathcal{C}([0,\infty))$. Given a compact interval $[0,T]$ with partition $0 = t_0 \leqslant t_1 < \ldots < t_d \leqslant t_{d+1} = T$, let $\nu_{t_1,\ldots,t_d}$ denote the correlation dimension of the finite-dimensional distribution $\mu_{t_1,\ldots,t_d}$ on $\mathbb{R}^d$ corresponding to the vector $(X(t_1),\ldots,X(t_d))$. The process is said to be *accessible* with respect to correlation dimension if, for every $T>0$, we have $\lim_{\Delta T\to 0} \nu_{t_1,\ldots,t_d} = \nu_T$ where $\Delta T = \max_{1\leqslant j\leqslant d+1} |t_j - t_{j-1}|$ is the mesh of the partition over $[0,T]$.

We note that time-length invariance does not necessarily imply accessibility. If X and Y are two independent realizations of the process, the assumed continuity of paths does imply that

$$\lim_{\Delta T\to 0} \sup_{t_1,\ldots,t_d} |X(t_j) - Y(t_j)| = \|X-Y\|_T \tag{2.8}$$

and hence that the corresponding spatial correlation integrals converge appropriately

$$\lim_{\Delta T\to 0} C_{t_1,\ldots,t_d}(r) = C_T(r) \quad \text{for each} \quad r>0, \tag{2.9}$$

where $C_{t_1,\ldots,t_d}(r)$ is the spatial correlation integral of $\mu_{t_1,\ldots,t_d}$. However, accessibility further requires that we are able to interchange two limits, i.e. we must assume that

$$\lim_{\Delta T\to 0}\lim_{r\to 0} \frac{\log C_{t_1,\ldots,t_d}(r)}{\log r} = \lim_{r\to 0}\lim_{\Delta T\to 0} \frac{\log C_{t_1,\ldots,t_d}(r)}{\log r}. \tag{2.10}$$

The right-hand side of (2.10) is the desired quantity ν_T, while the lim sup of the

left-hand side will certainly always be bounded above by ν_T. Accessibility indicates that the entire path behaviour over $[0, T]$ can be approximated in a very strong way by observations at a finite number of time points $t_1, \ldots, t_d$ for a sufficiently fine partition. However, even in the absence of accessibility, a strong link between the finite-dimensional distributions and the overall process is provided by (2.9). As with time-length invariance, we do not know the size of the class of accessible processes.

The remainder of this section is devoted to showing that both deterministic outputs and gaussian processes are time-length invariant and accessible with respect to correlation dimension; moreover, for deterministic outputs, this new definition of correlation dimension coincides with the standard one (discussed below) usually assigned to the underlying dynamical system.

A stationary finite-dimensional dynamical system on a smooth compact d-dimensional manifold M is represented by a semigroup of mappings $\{\varphi_t\}_t$, $t \geqslant 0$, $\varphi_t : M \to M$, and an invariant distribution μ^* (i.e. $\mu^* \varphi_t^{-1} = \mu^*$ for each t) which describes the distribution of initial conditions on M. Note that, by the Whitney embedding theorem, there exists a minimal embedding dimension $d^* \leqslant 2d + 1$ for which M can be viewed as a subset of $\mathbb{R}^{d^*}$. In this case, the *correlation dimension of the system* is typically defined to be the correlation dimension ν^* of μ^*, computed via (2.2) and (2.3) with d^* in place of d. In practice, however, one rarely observes the actual system evolving on M, but observes instead a real-valued functional $X(t) = h(\varphi_t(\boldsymbol{X}))$ of the system; here h is a smooth mapping of M into $\mathbb{R}$ and $\boldsymbol{X}$ denotes a random initial condition (assumed selected according to μ^*). The resulting process $X(t)$ is strictly stationary with invariant distribution μ on $\mathcal{C}([0, \infty))$ having one-dimensional marginals $\mu_t = \mu^* h^{-1}$. Let ν_T denote the correlation dimension of the functional process over $[0, T]$. Then we have:

Theorem 2.1. *Suppose the semigroup $\{\varphi_t\}_t$ satisfies a Lipschitz condition over compact time intervals, i.e. for each $T > 0$ there exists A_T such that*

$$\sup_{0 \leqslant t \leqslant T} \|\varphi_t(\boldsymbol{x}) - \varphi_t(\boldsymbol{y})\|_{d^*} \leqslant A_T \|\boldsymbol{x} - \boldsymbol{y}\|_{d^*}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in M$, where d^* is the minimal embedding dimension of M.

(a) If h also satisfies a Lipschitz condition over M, then $\nu_T \leqslant \nu^*$ for every $T > 0$.

(b) If h and the semigroup further satisfy the hypotheses of the Takens's embedding theorem, then the functional process $\{X(t)\}_t$ is time-length invariant and accessible with $\nu_\infty = \nu^*$.

Proof. (*a*) Suppose h satisfies a Lipschitz condition $|h(\boldsymbol{x}) - h(\boldsymbol{y})| \leqslant B\|\boldsymbol{x} - \boldsymbol{y}\|_{d^*}$ for all $\boldsymbol{x}, \boldsymbol{y} \in M$. If $X(t) = h(\varphi_t(\boldsymbol{X}))$ and $Y(t) = h(\varphi_t(\boldsymbol{Y}))$ are two realizations of the functional process with initial conditions $\boldsymbol{X}$ and $\boldsymbol{Y}$, then

$$\sup_{0 \leqslant t \leqslant T} |X(t) - Y(t)| \leqslant B \sup_{0 \leqslant t \leqslant T} \|\varphi_t(\boldsymbol{X}) - \varphi_t(\boldsymbol{Y})\|_{d^*} \leqslant A_T B \|\boldsymbol{X} - \boldsymbol{Y}\|_{d^*}.$$

Hence, choosing $\boldsymbol{X}$ and $\boldsymbol{Y}$ to be IID initial conditions, each with distribution μ^*, we get $C_T(r) \geqslant P([\|\boldsymbol{X} - \boldsymbol{Y}\|_{d^*} \leqslant r/(A_T B)]) = C_{\mu^*}(r/(A_T B))$ and the result follows.

(*b*) Suppose h and the semigroup additionally satisfy the hypotheses of Takens's (1981) embedding theorem. Given any $T > 0$ and time points $0 \leqslant t_1 <$

$\ldots < t_k \leqslant T$, it follows that the mapping $\Phi : M \to \mathbb{R}^k$ defined by $\Phi(\boldsymbol{x}) = (h(\varphi_{t_1}(\boldsymbol{x})), \ldots, h(\varphi_{t_k}(\boldsymbol{x})))$ is an embedding for all $k \geqslant d^*$. Since Φ is bi-Lipschitz, the correlation dimensions of μ^* and the image measure $\mu_{t_1,\ldots,t_k}$ are the same, i.e. $\nu^* = \nu_{t_1,\ldots,t_k}$ for all $k \geqslant d^*$. The result now follows from this and (*a*). ■

We now consider stationary gaussian processes. Without loss of generality we can assume that all such processes have mean 0, since (2.7) is unchanged by a location shift. We immediately note that if the covariance matrix of the components $X(t_1), \ldots, X(t_d)$ is *non-singular,* then the joint distribution of $X(t_1), \ldots, X(t_d)$ has a bounded density with respect to d-dimensional Lebesgue measure and hence $\nu_{t_1,\ldots,t_d} = d$. This shows that any gaussian process for which all finite-dimensional joint distributions are non-singular is both time-length invariant and accessible with $\nu_\infty = \infty$. Thus we need only consider those cases where singularities arise in some of the joint distributions; since these distributions are gaussian, a singularity corresponds to the existence of a non-trivial linear relation among the components.

It will be useful here (and later when we look at coloured noise) to consider the spectral decomposition of a general complex-valued second-order stationary process (see, for example, Cramér & Leadbetter 1967). Let $K(t) = E(X(s)\overline{X}(s+t))$ denote the autocovariance function of the process. Under the weak restriction that $K(t)$ be continuous, we have the unique representation

$$K(t) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\lambda t}\, S(d\lambda), \tag{2.11}$$

where S is the associated (non-normalized) spectral measure. The following two lemmas (proofs can be found in Cutler (1994)) provide the needed results:

Lemma 2.1. *Let $X(t)$ be a second-order stationary (complex-valued) process with continuous autocovariance $K(t)$. If there exist time points $t_1, \ldots, t_n$ and complex numbers $a_1, \ldots, a_n$ (not all zero) such that $\sum_{j=1}^n a_j X(t_j) = 0$ a.s., then the associated spectral measure S is purely discrete.*

Now any (complex-valued) second-order stationary gaussian process with mean 0, continuous autocovariance, and purely discrete spectral measure S, has the representation

$$X(t) = \sum_{\lambda_k \in \mathcal{A}} c_k\, \mathrm{e}^{\mathrm{i}\lambda_k t} Z_k \quad \text{a.s.}, \tag{2.12}$$

where $\mathcal{A}$ is the set of atoms of S, $c_k^2 = S(\{\lambda_k\})$, and $Z_1, Z_2, \ldots$ are orthogonal (complex-valued) gaussian variables with mean 0 and variance 1. The sum on the right-hand side of (2.12) is interpreted as a limit in quadratic mean.

Lemma 2.2. *Let $X(t)$ be a complex-valued second-order stationary process with mean 0 and purely discrete spectral measure S. Let $|\mathcal{A}|$ denote the number of atoms of S.*

(a) If $|\mathcal{A}| = n$, then every collection of $k \geqslant n+1$ components $X(t_1), \ldots, X(t_k)$ is linearly dependent.

(b) If $|\mathcal{A}| \geqslant n$, then for every $T > 0$ and every sequence of partitions $\{\mathcal{P}_m\}_m$ of $[0, T]$ for which the mesh $\Delta T \to 0$, there exists m_0 such that, for all $m \geqslant m_0$, $\mathcal{P}_m$ includes n distinct time points $t_1, \ldots, t_n$ (which may vary with m) for which $X(t_1), \ldots, X(t_n)$ are linearly independent.

As every real-valued stationary gaussian process with continuous paths has a continuous autocovariance function, it follows from the preceding lemmas and earlier discussion that:

Theorem 2.2. *Every real-valued stationary gaussian process with continuous paths is time-length invariant and accessible with respect to correlation dimension. If the spectral measure of the process is purely discrete with exactly d atoms, then* $\nu_\infty = d$. *In all other cases* $\nu_\infty = \infty$.

Remarks. (1) Note that spectral theory allowed us to completely determine correlation dimension in the gaussian case. This is partly due to the fact that singularities in gaussian distributions are always reflected in the linear structure. We would not expect such complete success using spectral analysis with general stationary processes.

(2) Since every bounded Borel measure S is the spectral measure of some stationary gaussian process, we may choose S to be continuous with a high-frequency cut-off $S(\{\lambda \,|\, |\lambda| > \lambda_0\}) = 0$. In this case $\nu_\infty = \infty$, but the process has paths which are infinitely differentiable, and the Hausdorff dimension of a graph of a realization of the process is a.s. 1.

(3) A gaussian process comprised solely of d harmonic components has finite correlation dimension d, and any single realization of the process is completely predictable once d (linearly independent) observations have been obtained. However, the solution for one realization is a.s. not valid for any other realization, nor can it be used to predict limiting behaviour and occupation probabilities for other realizations. This situation corresponds to the projection of an (unbounded) d-dimensional dynamical system with uncountably many distinct simple attractors. Note that graphs of the process are periodic or almost periodic and a.s. have Hausdorff dimension 1.

3. Dimension analysis of gaussian power-law coloured noise

In this section we use the real form S^* of the spectral measure S, i.e. $S^*([0, \lambda]) = S([-\lambda, \lambda])$. In this case the autocovariance is expressed as

$$K(t) = \int_0^\infty \cos(\lambda t)\, S^*(\mathrm{d}\lambda). \tag{3.1}$$

The original intent of OP appears to have been to estimate the correlation dimension of a process with (real) spectral density $f^*(\lambda)$ of the form

$$f^*(\lambda) = \mathrm{d}S^*/\mathrm{d}\lambda = \lambda^{-\alpha} \quad \text{for} \quad \lambda > 0, \quad \text{where} \quad \alpha > 1. \tag{3.2}$$

Of course to maintain integrability of f^* we require a low-frequency cut-off λ_0 on the power law; hence the range in (3.2) should be $\lambda > \lambda_0$. Any integrable function can be inserted over $(0, \lambda_0)$.

Since (3.2) corresponds to a continuous spectral measure, our earlier work gives $\nu_\infty = \infty$ for any gaussian process following this model. OP, however, approached the problem numerically through the GP algorithm and (1.3), and consistently obtained the value $\nu_{\mathrm{GP}}^{(\infty)} = 2/(\alpha - 1)$ for $1 < \alpha < 3$. Various explanations of this phenomenon have been attempted, and a particularly insightful discussion can be found in Theiler (1991). Here we examine the problem from a different point of

view, consistent with the approach taken in § 2. We conclude that it is important to distinguish the discrete spectrum simulation method of OP (discussed below) from the continuous spectrum model (3.2); in the latter case, we argue that both low and high frequencies have a role to play in determining the sample size necessary to find a correct scaling region for the sample correlation integral. We also give a new interpretation of the magic number $d(\alpha) = 2/(\alpha - 1)$ as the correlation dimension of a conditional distribution μ^x in the discrete spectrum model.

Any strictly stationary process $X(t)$ with continuous paths (this includes deterministic outputs) and invariant distribution μ can be represented as a functional of a shift dynamical system on $\mathcal{C}([0,\infty))$. For each $t \geqslant 0$, the left-shift operator $L_t : \mathcal{C}([0,\infty)) \longrightarrow \mathcal{C}([0,\infty))$ is defined by

$$L_t(x)(s) = x(s+t). \tag{3.3}$$

The mappings $\{L_t\}_t$, $t \geqslant 0$, form a semigroup on $\mathcal{C}([0,\infty))$, and it follows that $\mu L_t^{-1} = \mu$ for every $t \geqslant 0$ as a consequence of stationarity. Hence $(\{L_t\}_t, \mu)$ defines a stationary dynamical system evolving on $\mathcal{C}([0,\infty))$ with random initial conditions $X \in \mathcal{C}([0,\infty))$ selected according to μ. The original process $X(t)$ is recovered by projecting $L_t(X)$ down to its zeroth coordinate:

$$X(t) = L_t(X)(0). \tag{3.4}$$

Since $(\{L_t\}_t, \mu)$ is a dynamical system, the usual definitions of ergodicity and mixing apply. (Note that if $X(t)$ is a functional of a finite-dimensional dynamical system in $\mathbb{R}^d$, then $(\{L_t\}_t, \mu)$ will be ergodic whenever the original d-dimensional system is ergodic.) The significance of this approach is seen in the following theorem.

Theorem 3.1. *Let $X(t)$ be strictly stationary on $\mathcal{C}([0,\infty))$ with invariant distribution μ.*

(a) Suppose that $(\{L_t\}_t, \mu)$ is ergodic. Let $\tau > 0$ be fixed, and define the discrete-time embedded process $\{\boldsymbol{X}_n\}_n$ in $\mathbb{R}^d$ by $\boldsymbol{X}_n = (X(n\tau), X((n+1)\tau), \ldots, X((n+d-1)\tau))$ for $n = 1, 2, \ldots$. Let $C_n(r)$ be the sample correlation integral of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ as defined in (1.2). Then, for almost all choices of τ, $\lim_{n\to\infty} C_n(r) = C_{\bar{\mu}}(r)$ a.s. where $\bar{\mu} = \mu_{\tau,2\tau,\ldots,\mathrm{d}\tau}$ is the distribution of $\boldsymbol{X}_1$ in $\mathbb{R}^d$.

(b) If $X(t)$ is gaussian, then the dynamical system $(\{L_t\}_t, \mu)$ is ergodic if and only if the real spectral measure S^* of $X(t)$ is purely continuous. (We then say $X(t)$ is ergodic.)

Proof. (*a*) The discrete-time embedded process will be ergodic provided no multiple of τ coincides with a period (should one exist) of the original process. Part (*a*) then follows directly from the work of Pesin (1993), Serinko (1993*a*), and Aaronson *et al.* (1993). Serinko's work is more complete in this context as his method works for all $r > 0$ without any continuity assumptions on the distribution $\mu_{\tau,\ldots,\mathrm{d}\tau}$. Part (*b*) is an old result, and can be found in Maruyama (1949) and Grenander (1950). See also Cramér & Leadbetter (1967) and Cornfeld *et al.* (1982) for discussions. ■

It follows from Theorem 3.1 that the embedding method, applied to a single realization of a continuous-spectrum gaussian process, yields the desired limit $\lim_{n\to\infty} C_n(r) = C_{\bar{\mu}}(r)$ for each $d \geqslant 1$. Moreover, we know from Lemma 2.1 that

$\bar{\mu} = \mu_{\tau,2\tau,\ldots,\mathrm{d}\tau}$ is a non-singular gaussian distribution, and hence $C_{\bar{\mu}}(r) \sim cr^d$ for sufficiently small r. This is consistent with the theoretical calculations of Theiler (1991), who observed a range of small r with the correct scaling $C_n(r) \sim cr^d$ for very large n in a continuous power-law model of the form (3.2). There are two distinct factors influencing the size of n necessary to observe the correct scaling. The *statistical effect* (i.e. the slowing of convergence of $C_n(r)$ to $C_{\bar{\mu}}(r)$ due to the dependence structure between the vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$) is a function of the mixing behaviour of the process. In gaussian processes, a necessary and sufficient condition for ordinary mixing is that $\lim_{t\to\infty} K(t) = 0$. Uniform mixing conditions (see Denker & Keller (1986) and Serinko (1993*b*) for examples and applications) are likely necessary to get rates near optimal. Since mixing in gaussian processes is determined by the tail behaviour of $K(t)$ (hence by the behaviour of S^* near 0) we see that the statistical effect increases as the power at very low frequencies increases. This phenomenon has been noted by Theiler (1986, 1991) who recommends modifications to the sample correlation integral to improve convergence. In the case of dependencies which persist over long time periods, we suggest that repeated independent sampling of the vector $\boldsymbol{X} = (X(\tau), \ldots, X(\mathrm{d}\tau))$ be substituted (when possible) for the GP algorithm. The second factor influencing scaling behaviour is the geometric effect induced by the degree of path smoothness (i.e. by the degree of dependence between components within the vector $\boldsymbol{X}$). The smoothness of gaussian paths depends on $K(t)$ at t near 0 (i.e. on the tail of S^*) so the *geometric effect* is reduced when sufficient power exists at arbitrarily high frequencies. When paths are very smooth, $X(\tau), \ldots, X(\mathrm{d}\tau)$ are strongly statistically correlated for small τ and the distribution $\mu_{\tau,\ldots,\mathrm{d}\tau}$ in $\mathbb{R}^d$ appears low-dimensional from a distance. The correct scaling $C_{\bar{\mu}}(r) \sim cr^d$ is then observed only for small r, and larger n is required for $C_n(r)$ to exhibit correct scaling at small r, even assuming independent $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. (In practice this effect may be lessened somewhat by a good choice of the time delay τ.) Note that the worst-case scenario for an ergodic gaussian process corresponds to a continuous spectral measure with small high-frequency cut-off (very smooth paths) and a lot of power near 0.

OP attempted to simulate a process with spectral density (3.2) by first choosing a small frequency increment $\Delta\lambda$ and approximating (3.2) by the discrete power-law distribution

$$S^*(\{k\,\Delta\lambda\}) = C\,(k\,\Delta\lambda)^{-\alpha}(\Delta\lambda) \quad \text{for} \quad k = 1, 2, 3, \ldots, \tag{3.5}$$

where C is a normalizing constant chosen to yield unit variance $K(0) = 1$. They then simulated a process with spectral distribution (3.5) by generating random uniform phases (but noted that gaussian models could be used as well). The real representation of a gaussian process with general discrete real spectral distribution S^* is given by

$$X(t) = \sum_{\lambda_k \in \mathcal{A}^*} c_k\,(A_k \cos(\lambda_k t) + B_k \sin(\lambda_k t)), \tag{3.6}$$

where $\mathcal{A}^*$ is the set of atoms of S^*, $c_k^2 = S^*(\{\lambda_k\})$, the A_ks, B_ks are IID real gaussian variables with mean 0 and variance 1, and the sum in (3.6) converges in quadratic mean. Substituting (3.5) into (3.6) yields the gaussian power-law

process

$$X_\alpha(t) = C^{1/2}(\Delta\lambda)^{1/2} \sum_{k=1}^{\infty} (k\,\Delta\lambda)^{-\alpha/2}(A_k \cos(k\,\Delta\lambda\, t) + B_k \sin(k\,\Delta\lambda\, t)). \quad (3.7)$$

We note that $X_\alpha(t)$ is also a.s. a continuous periodic function (not just a limit in quadratic mean; see Kahane (1985) for continuity results on random Fourier series) with period $2\pi/(\Delta\lambda)$, and so (3.7) directly defines random elements of $\mathcal{C}([0,\infty))$. OP then sampled the process over the interval $(0, 2\pi/(\Delta\lambda))$ at equally spaced timepoints $t_k = k\,\Delta t$, $k = 1, \ldots, m$. For simulation purposes, and to avoid the aliasing of high frequencies, the sum in (3.7) was truncated at $k = \frac{1}{2}m$. The GP algorithm was subsequently applied to the sampled trajectory $X_\alpha(t_1), \ldots, X_\alpha(t_m)$.

We will need to modify the above set-up slightly so that various convergence issues are not confused. First note that (3.7) is not changed in any fundamental way by taking $C = \Delta\lambda = 1$, and we do so throughout the following. The phase space then becomes $\mathcal{C}_{2\pi}$, the set of all continuous functions with period 2π, equipped with the uniform topology and norm $\| \|_{2\pi}$ of (2.4). We also assume that the sum in (3.7) has not been truncated, that Δt remains fixed (chosen so that no integer multiple of Δt coincides with 2π), and that we let $m \to \infty$ over several periods of the process. Note that the discrete nature of the spectrum would have become apparent had OP continued sampling over several periods.

Since (3.7) is a discrete spectrum gaussian model, the invariant distribution μ is not ergodic and Theorem 3.1(a) does not apply directly. However, we can decompose μ into its ergodic components and apply Theorem 3.1(a) to each component. Note that each initial condition $x \in \mathcal{C}_{2\pi}$ belongs to the basin of attraction (under the left-shift operators L_t) of exactly one ergodic component, say μ^x. Since $C_n(r)$ is computed, for each d, by embedding over left-shifts of the same realization x, it follows that, given x, $\lim_{n\to\infty} C_n(r) = C_{\bar{\mu}^x}(r)$ μ^x-a.s., where $\bar{\mu}^x = \mu^x_{t_1,t_2,\ldots,t_d}$. Thus, given x, we will have

$$\nu^{(d)}_{\mathrm{GP}}(x) = \lim_{r\to 0} \lim_{n\to\infty} \frac{\log C_n(r)}{\log r} = \nu^x_{t_1,\ldots,t_d} \quad \mu^x - \text{a.s.}, \quad (3.8)$$

where $\nu^x_{t_1,\ldots,t_d}$ is the correlation dimension of the conditional distribution $\mu^x_{t_1,\ldots,t_d}$. OP observed $\nu^{(d)}_{\mathrm{GP}}(x) \approx d(\alpha) = 2/(\alpha - 1)$, independent of x, for $1 < \alpha < 3$ and sufficiently large d.

We now show that $d(\alpha)$ is μ-a.s. the correlation dimension of each ergodic component μ^x over $[0, 2\pi]$ (interpreted in the sense of (2.5) and (2.6), with $T = 2\pi$); moreover, it is also μ-a.s. the Hausdorff dimension of the support M^x of μ^x in $\mathcal{C}_{2\pi}$. Define $\psi^x : [0, 2\pi] \longrightarrow \mathcal{C}_{2\pi}$ and $M^x \subset \mathcal{C}_{2\pi}$ by $\psi^x(s) = L_s(x)$ and $M^x = \psi^x([0, 2\pi])$. Since

$$\|\psi^x(s) - \psi^x(t)\|_{2\pi} = \sup_{0\leqslant u\leqslant 2\pi} |x(s+u) - x(t+u)| \quad (3.9)$$

the continuity of x shows that ψ^x is continuous, and so the image M^x is compact, hence closed, in $\mathcal{C}_{2\pi}$ in the uniform topology. This closure property is directly related to the lack of ergodicity in the periodic model; the set of shifts of x is a closed invariant set and therefore not dense in any larger region.

It is quite easy to deduce the form of the conditional distributions μ^x by trans-

ferring from rectangular coordinates (A_k, B_k) to polar coordinates $(R_k, -\phi_k)$ in (3.7). Then $A_k = R_k \cos\phi_k$, $B_k = -R_k \sin\phi_k$, and (3.7), with $C = \Delta\lambda = 1$, becomes

$$X_\alpha(t) = \sum_{k=1}^{\infty} k^{-\alpha/2} R_k \cos(\phi_k + kt). \tag{3.10}$$

The ϕ_ks are IID uniform variables over $[0, 2\pi]$, independent of the radii R_ks. We see from (3.10) that shifting a realization by s units only involves shifting angles

$$L_s(X_\alpha)(t) = \sum_{k=1}^{\infty} k^{-\alpha/2} R_k \cos((\phi_k + ks) + kt). \tag{3.11}$$

Since any set of shifted angles $\phi_1 + s, \phi_2 + 2s, \ldots, \phi_k + ks, \ldots$ (addition mod 2π) is just as likely as the set $\phi_1, \phi_2, \ldots$ under the uniform distribution, we see that all shifts of x are equally likely. Hence μ^x corresponds to a 'lifting' of the uniform distribution m from the circle $[0, 2\pi]$ to M^x, i.e.

$$\mu^x(\psi^x([s,t])) = m([s,t]) = (t-s)/2\pi \quad \text{for} \quad 0 \leqslant s < t \leqslant 2\pi. \tag{3.12}$$

Now from Kahane (1985, p. 199), the paths of (3.7) μ-a.s. satisfy a Lipschitz condition almost of order $1/d(\alpha)$ for $1 < \alpha < 3$, i.e. for μ-almost all x we have

$$\|\psi^x(s) - \psi^x(t)\|_{2\pi} = \sup_{0 \leqslant u \leqslant 2\pi} |x(s+u) - x(t+u)| = O(|t-s|^{(\alpha-1)/2} \sqrt{\log |t-s|^{-1}}). \tag{3.13}$$

A slightly less exact result (a Lipschitz condition of order β for every $\beta < 1/d(\alpha)$) can be deduced from (3.5) and Theorem (ii), p. 181, of Cramér & Leadbetter (1967). This less exact result is sufficient for our purposes. Applying standard dimension techniques (see Proposition 2.1 and Theorem 16.2 of Falconer (1990)), we conclude that $\dim(M^x) = \dim(\psi^x([0, 2\pi])) \leqslant d(\alpha)$ μ-a.s., where $\dim(M^x)$ is the Hausdorff dimension of M^x under $\| \ \|_{2\pi}$. Similarly, if $\psi^x(S)$ and $\psi^x(T)$ are IID observations from μ^x, it follows from (3.12) and (3.13) that, for each $0 < \beta < 1/d(\alpha)$, there exist $0 < C < \infty$ and $0 < c < \infty$ such that

$$P^x([\|\psi^x(S) - \psi^x(T)\|_{2\pi} \leqslant r]) \geqslant m \times m(\{(s,t) \mid C|t-s|^\beta \leqslant r\}) \sim cr^{1/\beta} \tag{3.14}$$

for small r. Hence we must have $\nu^x_{2\pi} \leqslant d(\alpha)$ μ-*a.s.* The reverse inequalities can be obtained by applying the potential theory methods of Frostman; see Cutler (1994).

Our discussion in this section shows that we will observe the OP phenomenon in every non-ergodic system for which the trapping invariant subspaces (or attractors) of the ergodic components have dimension strictly smaller than the dimension of the invariant measure. This behaviour occurs whether the system is finite-dimensional (see examples below) or infinite-dimensional (as in the case of coloured noise). In ergodic systems with poor or no mixing properties (weakly ergodic systems) we may observe a *transient* OP phenomenon; the sample correlation integral exhibits a significant region of low-dimensional scaling in the (n, r)-plane, while the correct scaling is obtained at small r for sufficiently large n (see Example 3.3).

Example 3.2. Consider two harmonic oscillators with commensurate frequencies

$$\mathrm{d}^2y/\mathrm{d}t^2 + \lambda^2 y = 0, \quad \mathrm{d}^2z/\mathrm{d}t^2 + 4\lambda^2 z = 0 \tag{3.15}$$

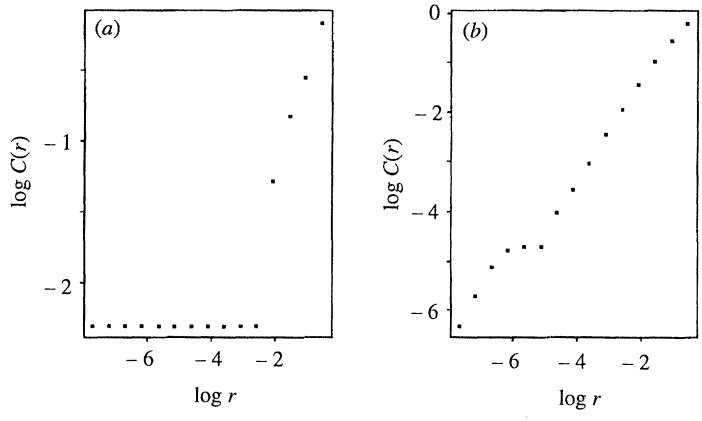

Figure 1. Log–log plots of sample correlation integral for the circle map. (*a*) $\alpha = 0.1$; (*b*) $\alpha = 0.101013$.

with amplitudes $R_1 = R_2 = 1$ and random independent phases ϕ_1, ϕ_2 uniformly distributed over $[0, 2\pi]$. Solving (3.22) in first-order form yields $y_1(t) = y(t) = \cos(\lambda t + \phi_1)$, $y_2(t) = y'(t) = -\lambda \sin(\lambda t + \phi_1)$, and $z_1(t) = z(t) = \cos(2\lambda t + \phi_2)$, $z_2(t) = z'(t) = -2\lambda \sin(2\lambda t + \phi_2)$. Each system describes an ellipse in $\mathbb{R}^2$ and has an invariant distribution absolutely continuous w.r.t. one-dimensional Lebesgue measure on the ellipse. The product system therefore has invariant distribution absolutely continuous w.r.t. two-dimensional Lebesgue measure on a torus. However, each realization $\boldsymbol{x}(t) = (y_1(t), y_2(t), z_1(t), z_2(t))$ generates a simple closed curve in $\mathbb{R}^4$, i.e. the closed support of each ergodic component is one-dimensional. Dimension measurements made on any functional of one realization of such a system will always yield local subspace estimates of 1 or less. Takens's theorem is sometimes said to 'fail' here, but in fact the failure is that of the GP algorithm, not the theorem. Dimension analysis on independent embedded vectors $\boldsymbol{X}_n = (h(\boldsymbol{x}(\tau)), \ldots, h(\boldsymbol{x}(\mathrm{d}\tau)))$ will yield the correct answer for sufficiently rich h and small τ.

Example 3.3. The simple circle map $T(x) = (x + \alpha) \bmod 1$ for $0 \leqslant x \leqslant 1$ is an example of a discrete-time dynamical system which exhibits the OP phenomenon for rational α (the non-ergodic case) and a transient OP phenomenon for irrational α (the weakly ergodic case). This is somewhat similar to the behaviour of power-law coloured noise in the cases of discrete and continuous spectra respectively. Note that the uniform distribution on $[0, 1]$ is always an invariant distribution for T (for any choice of α). In the case that α is rational, the ergodic components are discrete uniform distributions over periodic cycles $(x, x + \alpha, \ldots, x + k\alpha)$ of finite length, and the local dimension is 0. Figure 1*a* shows a plot of $\log C_n(r)$ against $\log r$ (embedding dimension $d = 1$) for 1000 observations from a realization of this system with $\alpha = 0.1$. In the case of irrational α, the system is ergodic but not mixing. Figure 1*b* shows the analogous log–log plot for $n = 1000$ and $\alpha = 0.101013\ldots$. The plateau in the graph of (*b*) represents the transient OP phenomenon.

I am grateful to K. S. Chan, Jiahua Chen, Leonard A. Smith, Tom Salisbury and Rege Serinko for useful discussions. Financial support for this research was provided by an operating grant from the Natural Sciences and Engineering Research Council of Canada.

References

Aaronson, J., Burton, R., Dehling, H., Gilat, D., Hill, T. & Weiss, B. 1993 Law of large numbers for U-statistics of dependent observations. Preprint.

Cramér, H. & Leadbetter, M. 1967 *Stationary and related stochastic processes.* New York: Wiley.

Cheng, B. & Tong, H. 1992 On consistent non-parametric order determination and chaos. *Jl R. statist. Soc.* B **54**, 427–449.

Cornfeld, I. P., Fomin, S. V. & Sinai, Ya. G. 1982 *Ergodic theory.* New York: Springer-Verlag.

Cutler, C. D. 1993 A review of the theory and estimation of fractal dimension. In *Nonlinear time series and chaos*, vol. I. *Dimension estimation and models* (ed. H. Tong), pp. 1–107. Singapore: World Scientific.

Cutler, C. D. 1994 A theory of correlation dimension for stationary time series. Statistics Technical Report, University of Waterloo.

Denker, M. & Keller, G. 1986 Rigorous statistical procedures for data from dynamical systems. *J. statist. Phys.* **44**, 67–93.

Falconer, K. J. 1990 *Fractal geometry: mathematical foundations and applications.* New York: Wiley

Grassberger, P. & Procaccia, I. 1983 Characterization of strange attractors. *Phys. Rev. Lett.* **50**, 346–349.

Grenander, U. 1950 Stochastic processes and statistical inference. *Arkiv. Mat.* **1**, 195–227.

Kahane, J.-P. 1985 *Some random series of functions*, 2nd edn. Cambridge University Press.

Kaplan, D. T. & Glass, L. 1992 Direct test for determinism in a time series. *Phys. Rev. Lett.* **68**, 427–430.

Kennel, M. B. & Isabelle, S. 1992 Method to distinguish possible chaos from colored noise, and to determine embedding parameters. *Phys. Rev.* A **46**, 3111–3118.

Maruyama, G. 1949 The harmonic analysis of stationary stochastic processes. *Mem. Fac. Sci. Kyusyu Univ.* A **4**, 45–106.

Nering, E. D. 1970 *Linear algebra and matrix theory*, 2nd edn. New York: Wiley.

Osborne, A. R. & Provenzale, A. 1989 Finite correlation dimension for stochastic systems with power-law spectra. *Physica* D **35**, 357–381.

Packard, N. H., Crutchfield, J. P., Farmer, J. D. & Shaw, R. S. 1980 Geometry from a time series. *Phys. Rev. Lett.* **45**, 712–716.

Pesin, Ya. B. 1993 On rigorous mathematical definitions of correlation dimension and generalized spectrum for dimensions. *J. statist. Phys.* **71**, 529–547.

Provenzale, A., Smith, L. A., Vio, R. & Murante, G. 1992 Distinguishing between low-dimensional dynamics and randomness in measured time series. *Physica* D **58**, 31–49.

Serinko, R. J. 1993*a* A Glivenko–Cantelli theorem for the Grassberger–Procaccia spatial correlation integral. Preprint.

Serinko, R. J. 1993*b* A consistent approach to least squares estimation of correlation dimension in weak Bernoulli dynamical systems. *Ann. appl. Probab.* (In the press.)

Smith, L. A. 1992 Identification and prediction of low-dimensional dynamics, *Physica* D **58**, 50–76.

Takens, F. 1981 Detecting strange attractors in turbulence. In *Lecture Notes in Mathematics*, no. 898, pp. 366–381. New York: Springer-Verlag.

Theiler, J. 1986 Spurious dimension from correlation algorithms applied to limited time series data. *Phys. Rev.* A **34**, 2427–2432.

Theiler, J. 1991 Some comments on the correlation dimension of $1/f^{\alpha}$ noise. *Phys. Lett.* A **155**, 480–493.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. & Farmer, J. D. 1992 Testing for nonlinearity in time series: the method of surrogate data. *Physica* D **58**, 77–94.