Estimating correlation dimension from a chaotic time series: when does plateau onset occur?

Mingzhou Ding\textsuperscript{a}, Celso Grebogi\textsuperscript{b}, Edward Ott\textsuperscript{c}, Tim Sauer\textsuperscript{d} and James A. Yorke\textsuperscript{e}

\textsuperscript{a}Center for Complex Systems and Department of Mathematics, Florida Atlantic University, Boca Raton, FL 33431, USA
\textsuperscript{b}Laboratory for Plasma Research, Department of Mathematics, and Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA
\textsuperscript{c}Laboratory for Plasma Research, Department of Physics, and Department of Electrical Engineering, University of Maryland, College Park, MD 20742, USA
\textsuperscript{d}Department of Mathematical Sciences, George Mason University, Fairfax, VA 22030, USA
\textsuperscript{e}Institute for Physical Science and Technology and Department of Mathematics, University of Maryland, College Park, MD 20742, USA

Received 22 March 1993
Revised manuscript received 26 May 1993
Accepted 13 June 1993
Communicated by P.E. Rapp

Suppose that a dynamical system has a chaotic attractor $A$ with a correlation dimension $D_2$. A common technique to probe the system is by measuring a single scalar function of the system state and reconstructing the dynamics in an $m$-dimensional space using the delay-coordinate technique. The estimated correlation dimension of the reconstructed attractor typically increases with $m$ and reaches a plateau (on which the dimension estimate is relatively constant) for a range of large enough $m$ values. The plateaued dimension value is then assumed to be an estimate of $D_2$ for the attractor in the original full phase space. In this paper we first present rigorous results which state that, for a long enough data string with low enough noise, the plateau onset occurs at $m = \text{Ceil}(D_2)$, where $\text{Ceil}(D_2)$, standing for ceiling of $D_2$, is the smallest integer greater than or equal to $D_2$. We then show numerical examples illustrating the theoretical prediction. In addition, we discuss new findings showing how practical factors such as a lack of data and observational noise can produce results that may seem to be inconsistent with the theoretically predicted plateau onset at $m = \text{Ceil}(D_2)$.

1. Introduction

The determination of the correlation dimension \cite{1} for a supposed chaotic process directly from experimental time series is an often used means of gaining information about the nature of the underlying dynamics (see, for example, contributions in ref. \cite{2}; for reviews on dimension measurements see refs. \cite{3–5}). In particular, such analyses have been used to support the hypothesis that the time series results from an inherently low-dimensional chaotic process. On the other hand, it is known that noise and the finite length of the data set may, if not properly understood and analyzed, lead to erroneous conclusions. It is the purpose of this paper to present results addressing the above problems. Other relevant work in this area is that of refs. \cite{6–13}.

Our interest in this problem mainly stems from published analyses of experimental time series which plot the estimated correlation dimension for the reconstructed attractor in an $m$-dimensional space as a function of $m$. Let $D_2$ denote the correlation dimension of the attractor in the original full phase space, and $D_2^{(m)}$ the correlation dimension of the projected attractor in an $m$-dimensional space. (The attractor recon-
structured using the delay-coordinate technique is a projected image of the attractor in the original phase space.) It is typically reported that, $\hat{D}_2^{(m)}$, the estimated value of $D_2^{(m)}$ from data, increases with $m$ and reaches a plateau (on which dimension estimates are relatively constant) for a range of large enough $m$ values. The plateaued dimension, denoted $\hat{D}_2$, is then taken to be an estimate of $D_2$. A main objective of this paper is to present new rigorous results which show that, in an ideal case (i.e., zero noise and an infinite length data string), the dimension plateau begins at $m = \text{Ceil}(D_2)$, where $\text{Ceil}(D_2)$ standing for ceiling of $D_2$, is the smallest integer greater than or equal to $D_2$. On the other hand, numerous studies in the literature (see refs. [14-19] for a sample) analyzing experimental data report plateau onsets occurring at $m$ values which are considerably larger than $\text{Ceil}(D_2)$. Our efforts were originally motivated by the question of how noise and finite length data strings might explain this discrepancy. In particular, if $\hat{D}_2^{(m)}$ determined by the analysis of the experimental data reaches a plateau at a value of $m$ substantially greater than $\text{Ceil}(D_2)$, what does that imply regarding the correctness of the assertion that the plateaued dimension value $\hat{D}_2$ is an estimate of the true correlation dimension $D_2$ of the underlying chaotic attractor?

The main conclusions of our paper are as follows:

1. For an infinitely long data set, $D_2^{(m)}$ as a function of $m$ reaches a plateau which begins at $m = \text{Ceil}(D_2)$. This is a rigorous result and is verified by numerical examples with long enough data strings.

2. Short data sets and observational noise can cause the plateau onset of $\hat{D}_2^{(m)}$ versus $m$ curve to occur at an $m$ value which is considerably greater than $\text{Ceil}(D_2)$. We explain this phenomenon based on new findings concerning the behavior of correlation integrals (see next section).

3. Even in cases where 2 applies, there are situations where the plateaued value of the numerically obtained estimates of $\hat{D}_2^{(m)}$ is a good estimate of the correlation dimension $D_2$ of the underlying chaotic attractor.

This paper is organized as follows. In section 2 we present the definition of correlation dimension and relevant background. In section 3 we examine the numerical aspects of determining correlation dimension from a time series, including the influence of noise and the finite length of the data set. In section 4 we give rigorous results regarding the behavior of $D_2^{(m)}$ versus $m$ and discuss the ideas behind these results. In section 5 we conclude this paper.

2. Background

Consider an $n$-dimensional dynamical system that exhibits a chaotic attractor. A probability measure $\rho$ can be defined on the attractor as follows. Let $x \in \mathbb{R}^n$ be a point on the attractor and $B$ be an $n$-dimensional hypercube centered at $x$. The probability measure contained in $B$, $\rho(B)$, is then defined to be the fraction of time a typical trajectory spends in the cube. Since this measure is induced by the dynamics it is also called the natural measure. Research in the past has focussed extensively on the characterization of such probability measures (see, for example, ref. [20]). In particular, a spectrum of dimensions $D_q$, $-\infty < q < \infty$, has been introduced to characterize them. Imagine that we cover the attractor in $\mathbb{R}^n$ with a uniform grid of mesh size $\varepsilon$. Denoting the natural measure in the $i$th cube (or box) of the grid by $P_i$, the dimension spectrum is given by [21,22]

$$D_q = \lim_{\varepsilon \to 0} \frac{1}{q - 1} \log \frac{\sum_{i=1}^{K(\varepsilon)} P_i^q}{\log \varepsilon},$$  \hspace{1cm} (1)$$

where $K(\varepsilon)$ is the total number of boxes with $P_i > 0$. Amongst the infinite number of dimensions, $D_0$, $D_1$ (defined by $\lim_{q \to 1} D_q$ in eq. (1)) and $D_2$ have received the most attention. They are termed box-counting dimension (or capacity), information dimension, and correlation dimension, respectively. From eq. (1) it is
straightforward to show that $D_q$ is a nonincreasing function of $q$; in particular, $D_0 \geq D_1 \geq D_2$. Generally, these dimensions assume noninteger values for a chaotic attractor. Hence chaotic attractors are often cited as a class of fractals in the sense of Mandelbrot [23]. In physical terms, these dimensions give a lower bound on the effective number of degrees of freedom activated in a physical process. The values of $D_1$ and $D_2$ also tell how often one can expect recurrence in the system state; the larger the dimension, the less the recurrence, and the more data is needed for effective short-term prediction of the behavior.

Except for a few special cases, the only avenue to obtain an estimate of attractor dimension is via analysis of an observed finite length orbit on the attractor. In this regard, the correlation dimension $D_2$ offers a substantial advantage over the other dimensions. To see this we turn to a slightly different definition of $D_2$ in terms of correlation integrals [1]. A correlation integral $C(\varepsilon)$ is defined to be the probability that a pair of points chosen randomly with respect to the measure $\rho$ is separated by a distance less than $\varepsilon$ on the attractor.

Assume that for discrete maps, after discarding transients for a typical initial condition, we measure and record the subsequent trajectory, $\{x_i\}_{i=1}^N$, where $x_i \in \mathbb{R}^n$. (In the case of a continuous time system with trajectory $x(t)$, the discrete set of points is obtained by sampling at $N$ discrete times $t_i$ equally spaced in the observation interval of length $L$; $x_i = x(t_i)$. In this paper, we assume that the sampling interval is fixed, thus $L$ and $N$ are linearly proportional to one another.) The correlation integral $C(\varepsilon)$ is then approximated by

$$D_2 = \lim_{\varepsilon \to 0} \frac{\log C(\varepsilon)}{\log \varepsilon}.$$  

(2)

where $\Theta(\cdot)$ is the Heaviside step function defined as $\Theta(x) = 0$ for $x \leq 0$ and $\Theta(x) = 1$ for $x > 0$, and $|\cdot|$ denotes the norm used in the space. Equation (3) can be interpreted as the ratio of the number of pairs of trajectory points with distances less than $\varepsilon$ to the total number of distinct pairs formed by $N$ points. In the limit $N \to \infty$, we have $C(N, \varepsilon) \to C(\varepsilon)$. See ref. [24] for a precise study of $C(\varepsilon)$.

To see the equivalence of eq. (2) and eq. (1) with $q = 2$ we rewrite the quantity $I_2(\varepsilon) = \sum_{i=1}^{N} P_i^2$ in eq. (1) as

$$I_2(\varepsilon) = \langle P(x) \rangle,$$  

(4)

where $P(x)$ denotes the probability measure in the grid cube containing the point $x$, and the angle brackets $\langle \cdots \rangle$ denote an average over $x$ with respect to the natural measure $\rho$, $\langle \cdots \rangle = \int \cdots \rho(dx)$. We now replace the average over $x$ by a time average following a chaotic trajectory. The result is

$$I_2(\varepsilon) \sim I_2(N, \varepsilon) = \frac{1}{N} \sum_{j=1}^{N} P(x_j),$$  

(5)

for large $N$. Since $P(x_j)$ is the probability that a point chosen randomly with respect to the natural measure falls within $\varepsilon$ of the point $x_j$, we expect that it can be replaced by the probability $P'(x_j)$ that a point on the trajectory falls in the $\varepsilon$ neighborhood of the point $x_j$. This probability can be approximated as

$$P'(x_j) \approx \frac{1}{N} \sum_{i \neq j} \Theta(\varepsilon - |x_i - x_j|).$$  

(6)

Hence

$$I_2(N, \varepsilon) \approx \frac{1}{N} \sum_{j=1}^{N} P'(x_j) = C(N, \varepsilon).$$  

(7)

From eqs. (5) and (7) we are led to the estimate

$$I_2(\varepsilon) \sim C(\varepsilon)$$  

(8)

in the limit $N \to \infty$. The difference between $\log I_2(\varepsilon)/\log \varepsilon$ and $\log C(\varepsilon)/\log \varepsilon$ becomes insignificant as $\varepsilon \to 0$. To illustrate this, we consider
the example where the probability measure is uniformly distribution in an \( n \)-dimensional hypercube of edge length one. In this case, \( C(\epsilon) \) and \( I_2(\epsilon) \) are calculated to be
\[
C(\epsilon) = [\epsilon(2 - \epsilon)]^n \tag{9}
\]
(see next section) and
\[
I_2(\epsilon) = \epsilon^n. \tag{10}
\]
In agreement with the previous discussion, eqs. (10) and (1), and eqs. (9) and (2) both yield
\[
D_2 = n.
\]
In practice the range of \( \epsilon \) over which \( C(N, \epsilon) \) can be studied is limited by the finite value of \( N \).
The commonly used procedure of obtaining an estimate of \( D_2 \) starts with plotting \( \log C(N, \epsilon) \) versus \( \log \epsilon \). This graph will typically show an approximately linear dependence in some ranges of \( \epsilon \) values. The slope of the graph in the linear scaling range is then the estimated dimension \( D_2 \). Assuming that the system variables are normalized so that the attractor size is of order unity, the upper \( \epsilon \) value, \( \epsilon_u \), at the border of the scaling range is also of order one, meaning that \( C(N, \epsilon_u) \approx 1 \). We now estimate the \( \epsilon \) value at the lower \( \epsilon \) border of the scaling range which we denote \( \epsilon_c \). Suppose that \( \epsilon_c \) is on the order of the smallest distance between any pair of the \( N \) orbit points, then
\[
C(N, \epsilon_c) = 2/N(N - 1) \approx 2/N^2.
\]
Roughly linear dependence of the slope \( D_2 \) down to the scale \( \epsilon_c \) then implies
\[
D_2 \sim \frac{\log C(N, \epsilon_c)}{\log \epsilon_c} \approx -2 \frac{\log N}{\log \epsilon_c}.
\]
Thus the smallest scale that can be probed by a data set of length \( N \) is
\[
\epsilon_c \sim N^{-2/D_2}. \tag{11}
\]
To extract a reasonable dimension estimate from the data for \( \log C(N, \epsilon) \) versus \( \log \epsilon \), we require the existence of a reasonably large scaling range, \( 0 \leq \log(1/\epsilon) \leq \log(1/\epsilon_c) \), where the inequality on the left results from the normalization of the attractor size to one. Arbitrarily taking \( \epsilon_c < 0.1 \), corresponding to a scaling range of at least a factor of 10 in \( \epsilon \), yields the result of Eckmann and Ruelle [9]
\[
N \geq 10^{D_2/2}. \tag{12}
\]
For comparison, we estimate the relationship between \( N \) and \( \epsilon \) for the box-counting dimension. Setting \( q = 0 \) in eq. (1) and letting \( K(N, \epsilon) \) be the number of occupied boxes, \( D_0 \) is calculated by
\[
D_0 = \lim_{\epsilon \to 0} \frac{\log K(N, \epsilon)}{\log(1/\epsilon)}.
\]
Assume \( K(N, \epsilon) \approx \text{constant} \ (1/\epsilon)^{D_0} \) where we may suppose constant \( \approx 1 \) since \( K(N, 1) = 1 \). If we are to determine \( D_0 \) from a trajectory of \( N \) points, we clearly can only examine \( \epsilon \) for which \( \epsilon \geq \epsilon_{\text{box}} \) where \( K(N, \epsilon_{\text{box}}) = N \). Hence the required number of points \( N \) satisfies
\[
N \geq \left( \frac{1}{\epsilon_{\text{box}}} \right)^{D_0},
\]
or
\[
\epsilon_{\text{box}} \geq N^{-1/D_0}. \tag{14}
\]
From eqs. (11) and (14) we get
\[
\epsilon_c \leq \epsilon_{\text{box}}^{2/D_2}. \tag{15}
\]
Since \( D_0 \geq D_2 \) the exponent in eq. (15) is at least 2. Thus, \( \epsilon_c \approx \epsilon_{\text{box}}^{2/D_2} \), and we see that there is a much larger expected linear scaling range for a plot of \( \log C(N, \epsilon_{\text{box}}) \) versus \( \log \epsilon \) than for a plot of \( \log K(N, \epsilon) \) versus \( \log \epsilon \).
In the case of information dimension a similar argument yields
\[
\epsilon_c \leq \epsilon_{\text{inf}}^{2D_1/D_2}, \tag{16}
\]
where we again note that \( D_1 \geq D_2 \). The significance of eqs. (15) and (16) is that one can explore much more refined structure of the attractor with \( D_2 \) than with \( D_0 \) or \( D_1 \). This can be a decisive advantage in situations where only
a limited number of points is available to approximate a chaotic attractor.

In many situations arising in practice one does not have a priori knowledge concerning the dimensionality of the dynamics or the equations of motion. Instead it is often the case that the only information available is a time series \( y(t) \) obtained by measuring a single scalar quantity \( y \). The system state variable \( x \) in the original phase space \( \mathbb{R}^n \) is related to \( y \) by a scalar function \( y = h(x) \) where \( h: \mathbb{R}^n \to \mathbb{R} \) is differentiable. In practice \( h \) corresponds to physical devices one uses to perform measurements in an experiment; \( x \) is the state variable of the experimental system. From \( y(t) \) one reconstructs the high-dimensional dynamics using the delay-coordinate method \([25,26]\). It proceeds by generating an \( m \)-dimensional vector \( y(t) \) from the time-delayed copies of \( y(t) \):

\[
y(t) = \{y(t), y(t-T), y(t-2T), \ldots, y(t-(m-1)T)\} \\
= \{h(x(t)), h(x(t-T)), h(x(t-2T)), \ldots, h(x(t-(m-1)T))\}, \quad (17)
\]

where \( T > 0 \) is the delay time and \( m \) is the dimension of the reconstruction space. The mapping from the trajectory \( x(t) \) in the original phase space \( \mathbb{R}^n \) to the trajectory \( y(t) \) in the reconstruction space \( \mathbb{R}^n \) is called the delay coordinate map. Results in ref. [27] show that, for typical \( T > 0 \), typical measurement function \( h(\cdot) \), and

\[
m > 2D_0, \quad (18)
\]

this delay coordinate map is one-to-one on the attractor; that is, distinct state on the attractor yields distinct vector \( y \). Here \( D_0 \) is the box-counting dimension of the chaotic attractor in the original phase space.

In this paper our aim is to estimate the correlation dimension from a time series. The result to be presented below shows that, for this purpose, \( m \geq D_2 \) suffices. Note that this result holds true irrespective of whether the delay coordinate map is one-to-one or not. This is contrary to the commonly accepted notion that an embedding (one-to-one and differentiable) is needed to estimate dimension, leading to the false surmise that \( m \) needs to be at least \( 2D_0 \) to guarantee a correct dimension estimation \(^*\). An additional note is that this value of \( m \) for dimension estimation is less than half that required by eq. (18), which is a sufficient condition for the delay coordinate map to be one-to-one.

To further stress the point that mappings which are not one-to-one can preserve dimension (as shown in figs. 1 and 2), while mappings which are one-to-one may not preserve dimension, we consider the example of the well-known Weierstrass function \( u = F_w(V, \alpha, \beta) \) where

\[
F_w(v, \alpha, \beta) = \sum_{n=0}^{\infty} \frac{1}{n^n} \cos(2\pi n^\beta v).
\]

The plot, \( u \) versus \( v \), is fractal with a capacity dimension between 1 and 2 for typical \( \alpha \) and \( \beta \) satisfying \( 1 < \alpha < \beta \). On the other hand, we can regard the relation \([v, u = F_w(v, \alpha, \beta)] \to v \) as specifying a continuous (but not differentiable) one-to-one projection mapping points on the

\(^*\) There is a widespread misconception about the relevance of embedding to dimension estimation. This is reflected to a large extent by the frequent reference, when computing correlation dimension, to the result of Takens [26] which states that \( 2D + 1 \) is a sufficient condition for an embedding, where \( D \) is the dimension of a smooth compact manifold in the original phase space. To get a rough idea of how widespread this misconception is, we conducted a literature search using the Science Citation Index for the years from 1987 to 1992 by looking for papers citing both ref. [1] and ref. [26]. We found 183 such papers. We then randomly selected a sample of 22 of these papers for closer examination. The following is what we found. Among the 22 papers there are 15 of them that calculate correlation dimension from time series. 5 of these 15 papers make explicit connections between \( 2D + 1 \) and dimension estimation. The rest of these papers ignore this issue entirely. Based on this information we estimate that, during the period from 1987 to 1992, there are at least 42 papers (probably many more) where the authors implicitly or explicitly assumed that a one-to-one embedding is needed for dimension calculation. In addition, among the papers we researched for this work, only refs. [3,4] imply that \( m \geq D_2 \) is sufficient for estimating \( D_2 \), although no justification is given.
Weierstrass curve to points on the $v$-axis. Thus we have a one-to-one projection of a set with dimension greater than one to a set with dimension equal to one. An example of an attractor arising from a dynamical system which has this character is given in ref. [28].

A measure $\rho$ is called a probability measure on a set $A$ if $\rho(A) = 1$ and $\rho(A') = 0$, where $A'$ is the complement of $A$. We say that $A$ is a set “with” a probability measure $\rho$ if this is so. We are interested in the dimension of the image of $A$ under a map $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$. Given a probability measure $\rho$ on $A$, we define the induced probability measure $F(\rho)$ as follows. For $E \in \mathbb{R}^m$, define $F^{-1}(E)$ to be the set of points in $\mathbb{R}^n$ that are mapped to $E$ under $F$, then $F(\rho)(E) = \rho(F^{-1}(E))$.

**Theorem 2.1.** Let $A$ be a set in $\mathbb{R}^n$ with a probability measure $\rho$. If $m \geq D_2(\rho)$, then for almost every smooth map $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$, the correlation dimension is preserved, that is, $D_2(F(\rho)) = D_2(\rho)$.

The set $A$ in theorem 2.1 can be any set with a probability measure $\rho$. Chaotic attractors are examples of such sets. Other examples include sets that need not to be closed. “Almost every” in the statement of the theorem is understood in the sense of prevalence to be discussed later in this paper (see also ref. [27]). Roughly speaking, “almost every” means that the functions $h$ that do not give the stated result are very scarce and are not expected to occur in practice. A pictorial illustration of theorem 2.1 appears in figs. 1 and 2. The set $A$ in fig. 1a is a folded 2-dimensional surface in $\mathbb{R}^3$ for which $D_2 = 2$. The projected image of $A$ in $\mathbb{R}^2$ (fig. 1b) is again a two-dimensional set with $D_2 = 2$ even though the projection is clearly not one-to-one. Further projection of $A$ into the real line $\mathbb{R}$ yields a one-dimensional interval with $D_2 = 1$. Thus the correct dimension is not obtained if $m < D_2$. The second example is shown in fig. 2. The set $A$ is a closed curve in $\mathbb{R}^3$ (fig. 2a). $D_2 = 1$ in this case.

The projected image of $A$ in $\mathbb{R}^2$ via general projections $u = u(x, y, z)$ and $v = v(x, y, z)$ is again a one-dimensional set with $D_2 = 1$ (fig. 2b).
Figure 2c shows the image of A projected into the real line under the projection \( w = w(x, y, z) \). The dimension information is preserved in both projections although the projected set in \( \mathbb{R} \) (fig. 2c) bears little resemblance to the original set in \( \mathbb{R}^3 \) (fig. 2a).

Theorem 2.1 can also be applied directly to physical situations where the dynamics is reconstructed in an \( m \)-dimensional Euclidean space \( \mathbb{R}^m \) by drawing data from an apparatus that makes \( m \) simultaneous measurements. In this case, the \( m \)-dimensional vector is \( (h_1(x), h_2(x), \ldots, h_m(x)) \), where \( h_1, \ldots, h_m \) are \( m \) measurement functions, and we retrieve the correct dimension from the observed time records for almost any set of measurement functions if \( m \) has a value greater than or equal to the correlation dimension of the original chaotic attractor.

Let our dynamical system be described by iterating an \( n \)-dimensional map \( G: \mathbb{R}^n \rightarrow \mathbb{R}^n \). Let \( A \) be a set in \( \mathbb{R}^n \) with a probability measure \( \rho \), and let \( A \) and \( \rho \) be invariant under the map \( G \). That is, \( A = G(A) \) and \( \rho = G(\rho) \). \( \rho \) is invariant if \( \rho(E) = \rho(G^{-1}(E)) \) for all measurable sets \( E \). This holds if \( \rho \) is the natural measure of a chaotic attractor. Assume that there is only a finite number of periodic points of period \( \leq m \) of \( G \) in \( A \). For any function \( h: \mathbb{R}^n \rightarrow \mathbb{R} \), the delay coordinate map \( F_h: \mathbb{R}^n \rightarrow \mathbb{R}^m \) is defined by

\[
F_h(x) = [h(x), h(G^{-1}(x)), \ldots, h(G^{-(m-1)}(x))].
\]

The issue of dimension preservation under delay coordinate maps is then addressed by the next theorem.

**Theorem 2.2.** If \( m = D_2(\rho) \), then for almost every \( h \), \( D_2(F_h(\rho)) = D_2(\rho) \).

(A sharper version of this theorem will be given in section 4. Similar results can be stated for flows generated by ordinary differential equations.)

The delay coordinate map in the theorem corresponds to taking \( T = 1 \) and \( t \) as integers in eq. (17). The condition assuming that there exist only finitely many periodic points of period less than or equal to \( m \) in \( A \) is needed to avoid situations such as the case where the invariant set \( A \) is a two-dimensional set of fixed points. In this case, under a delay coordinate map, all the fixed points are projected to the diagonal \( y_1 = y_2 = \ldots = y_m \), which has a correlation dimension \( D_2 = 1 \). This condition is expected to be always fulfilled for chaotic attractors encountered in practice.

Similar dimension preserving results also hold for the Hausdorff dimension \( D_H \), but do not hold for the box-counting dimension \( D_0 \) (see section 4).

In applications, theorem 2.2 has important predictive value. This can be seen in the way in which \( D_2 \) is extracted from a time series. It proceeds as follows. First, we reconstruct the attractor from the data set in an \( m \)-dimensional space by the delay-coordinate technique. Second, we compute the correlation integral \( C_m(N, \varepsilon) \) according to eq. (3). (Here a subscript \( m \) is added to indicate the dimensionality of the reconstruction space.) From the curve \( \log C_m(N, \varepsilon) \) versus \( \log \varepsilon \) we locate a linear region (also called a scaling region) for small \( \varepsilon \) and estimate the slope of the curve over the linear region. This slope, denoted \( D_2^{(m)} \), is then taken as an estimate of \( D_2^{(m)} \), the correlation dimension of the projection of the attractor to the \( m \)-dimensional reconstruction space. If these estimated values \( D_2^{(m)} \), plotted as a function of \( m \), appear to reach a plateau for a range of large enough \( m \), then the plateaued dimension value, denoted \( D_2 \), is taken to be an estimate of the correlation dimension \( D_2 \) for the underlying system. From theorem 2.2 it is clear that the plateau onset should ideally take place at \( m = \text{Ceil}(D_2) \), where \( \text{Ceil} \) (\( D_2 \)) standing for ceiling of \( D_2 \), is the smallest integer greater than or equal to \( D_2 \). In the next section of this paper we show numerical examples illustrating this statement. In the same section we also present new results explaining the systematic behavior of correlation integrals.
C_m(N, ε). From the behavior of these correlation integrals we discuss how practical factors such as a lack of data and observational noise can produce results that may appear to be inconsistent with the predicted plateau onset at \( m = \text{Ceil}(D_2) \).

3. Numerical evaluations of \( D_2 \)

This section is devoted to the numerical aspects of estimating the correlation dimension \( D_2 \) from a time series. The systems we use to produce such time series are the tent map,

\[
y_{t+1} = T(y_t) = \begin{cases} 2y_t, & 0 \leq y_t \leq \frac{1}{2} \\ 2(1 - y_t), & \frac{1}{2} < y_t \leq 1 \end{cases}
\]

(19)

the Hénon map,

\[
(y_{t+1}, x_{t+1}) = (1 + 0.3x_t - 1.4y_t, y_t),
\]

(20)

and the Mackey–Glass equation [29]. The Mackey–Glass equation is a time-delay differential equation originally proposed to model the blood regeneration process in patients with leukemia. It is

\[
\frac{dy(t)}{dt} = \frac{ay(t - \tau)}{1 + y(t - \tau)^c} - by(t),
\]

(21)

where \( a, b, c, \) and \( \tau \) are parameters. In this study we fix \( a = 0.2, b = 0.1, c = 10.0, \) and \( \tau = 100.0 \). The numerical integration of eq. (21) is done by the following iterative scheme [1]:

\[
y(t + 8t) = \frac{2 - b \delta t}{2 + b \delta t} y(t)
+ \frac{\delta t}{2 + b \delta t} \left\{ \frac{ay(t - \tau)}{1 + [y(t - \tau)]^10}
+ \frac{ay(t - \tau + \delta t)}{1 + [y(t - \tau + \delta t)]^10} \right\},
\]

(22)

where \( \delta t \) is the integration step size. For \( \delta t = 0.1, \) meaning that the interval \([t - \tau, t]\) is divided into 1000 subintervals of equal length, eq. (22) gives rise to a 1000-dimensional map. This map, aside from being an approximation to eq. (21), is itself a dynamical system.

As a convention, the time series generated by eqs. (19), (20) and (22) are normalized to the unit interval so that the reconstructed attractor lies in the unit hypercube in the reconstruction space. The norm we use to calculate distances in eq. (3) is the max-norm in which the distance between two points is the largest of all the component differences. Under the max-norm the largest distance between two points in the unit cube is one (in contrast, for the Euclidean norm, this quantity is \( \sqrt{m} \), where \( m \) is the dimension of the space). This feature has the advantage of allowing us to compare results with different \( m \). We stress, however, that the results presented in this paper are independent of the particular norm selected.

3.1. Numerical examples illustrating the prediction of theorem 2.2

The example we use to illustrate the prediction by theorem 2.2 (see the last paragraph of section 2) is a time series of 50 000 points generated by eq. (22). Specifically, we choose the initial function over the interval \([-\tau, 0]\) to be a constant \( y = 0.5 \) and discard the first 5000 iterations to eliminate transients. The time series is then formed by using a sampling time \( t_s = 10.0 \).

To reconstruct the attractor, we follow eq. (17) and take the delay time to be \( T = 20.0 \). The dimension of the reconstruction space is varied from \( m = 2 \) to \( m = 25 \). For each reconstructed attractor at a given \( m \) we calculate the correlation integral \( C_m(N, \varepsilon) \) according to eq. (3). In fig. 3 we display \( \log_2 C_m(N, \varepsilon) \) versus \( \log_2 \varepsilon \) for \( m = 2-8, 11, 15, 19, 23 \). For each \( m \) we identify a scaling region for small \( \varepsilon \) and fit a straight line through the region. The slope of the straight line gives the value of \( D_2(m) \) which is an estimate of the true correlation dimension \( D_2(m) \) for the time series reconstructed in the \( m \)-dimensional reconstruction space. Figure 4 shows the values of \( D_2(m) \) so estimated, plotted as open circles, as a
include the Hénon attractor, the Lorenz attractor, and the Mackey–Glass equation with various other parameter settings.

In applications, however, factors such as a lack of sufficient data and observational noise can produce results that may seem to be inconsistent with theorem 2.2. It is the purpose of the remainder of this section to study these issues which, we believe, are partly responsible for some of the reports in the literature [14–19] where the $D_2(m)$ versus $m$ curve reaches a plateau whose value $D_2$ is markedly smaller than the space dimension $m$ at which the plateau onset takes place.

### 3.2. The behavior of correlation integrals $C_m(N, \varepsilon)$

We begin by examining how the correlation integrals $C_m(N, \varepsilon)$ behave as a function of $m$. Figure 5 is a schematic diagram on log–log scale of a set of correlation integrals for $m = 2$ to $m = 13$. This figure is drawn to represent the systematic behavior of typical correlation integrals and is based roughly on a data set generated by eq. (22) with a different set of parameter values than that used in fig. 3. A dashed line is fit through the scaling region for each $m$. For $m \leq 5$ the slope of the dashed line is around $m$. For $m > 5$ the scaling regions are roughly parallel.

![Fig. 3. Log–log plots of the correlation integrals $C_m(N, \varepsilon)$ for the data set of 50000 points generated by eq. (22) $\log_2 C_m(N, \varepsilon)$ versus $\log_2 \varepsilon$ for $m = 2, 11, 15, 19, 23$ are displayed.](image)

![Fig. 4. $D_2(m)$ versus $m$, plotted as open circles, for the data set of 50000 points used in fig. 3. $D_2(m)$ versus $m$, plotted as triangles, for the data set of 2000 points used in fig. 10.](image)

function of $m$. For $m \leq 7$, $D_2(m) \approx m$. For $m \geq 8$, $D_2(m)$ reaches a plateau at a value of about 7.1. Identifying the plateau value with the correlation dimension $D_2$ of the attractor in the original phase space, this result is consistent with the prediction by theorem 2.2 that the plateau onset occurs at $m = \text{Ceil}(D_2)$, where $\text{Ceil}(D_2)$ is the smallest integer greater than $D_2 = 7.1$.

It is worth noting that, in our studies, numerous other attractors were also found to yield results conforming with theorem 2.2. Examples
with a slope of about 5.7. This slope is an estimate of $D_2$ for the attractor underlying the data set. The correlation integral corresponding to $m = \text{Ceil}(5.7) = 6$ is denoted by a thicker line in fig. 5.

Figure 5 exhibits several features that are typical of correlation integrals for chaotic systems (see fig. 3 for a comparison). The first feature we note is that the horizontal distance between $\log_2 C_m(N, \epsilon)$ and $\log_2 C_{m+1}(N, \epsilon)$ for $m \geq 6$ in the scaling regions is roughly a constant. This constant is predicted [30], for large $m$ and small $\epsilon$, to be

$$\Delta h = \Delta v / D_2,$$

where

$$\Delta v = TK_2$$

(24)

with $K_2$ the correlation entropy defined in ref. [30] and $T$ the delay time in eq. (17). See also ref. [26] for discussions involving increasing $m$.

Two other features exhibited by fig. 5 are also significant. These features concern the dependence of $\log_2 C_m(N, \epsilon)$ on $\log_2 \epsilon$, particularly, the patterns of deviation from linear behavior as $\epsilon$ increases through the upper boundary of the linear scaling region. For $m \leq 9$, $\log_2 C_m(N, \epsilon)$ increases with a gradually diminishing slope; while for $m \geq 11$, after exiting the linear region, the log-log plots in fig. 5 first increase with steeper slope compared to that in the scaling region, and then level off to meet the point $(0,0)$. These two different trends give rise to an uneven distribution in the extent of the scaling regions for different $m$ with the most extended scaling region occurring at $m = 10$. In what follows we discuss the origin of this behavior.

Consider an uncorrelated random sequence \( \{y_i\}_{i=1}^N \) with \( y_i \) uniformly distributed in the interval [0,1]. Taking $T=1$ and $t$ as integers in eq. (17) we reconstruct this time series in an $m$-dimensional Euclidean space. In the limit $N \rightarrow \infty$, we obtain a uniform probability measure supported on the $m$-dimensional unit cube. The correlation integral for such a probability measure can be calculated exactly [3,31]. To do so we first consider $C_1(\epsilon)$. By definition, $C_1(\epsilon)$ is the probability that two points $y'$ and $y''$ chosen randomly with respect to the uniform measure are separated by a distance less than $\epsilon$ in the unit interval. On the plane spanned by $y'$ and $y''$, $C_1(\epsilon)$ is equal to the area of the shaded region in fig. 6. This area is easily calculated to be $2\epsilon - \epsilon^2$. Thus

$$C_1(\epsilon) = 2\epsilon - \epsilon^2.$$  (25)

The slope of the plot of $\log C_1(\epsilon)$ versus $\log \epsilon$ is

$$\mu_1 = \frac{d \log C_1(\epsilon)}{d \log \epsilon} = 1 - \frac{\epsilon}{2 - \epsilon},$$  (26)

which is a monotonically decreasing function of $\epsilon$. When $\epsilon$ is small $\mu_1 \approx 1$. As $\epsilon \rightarrow 1$, $\mu_1 \rightarrow 0$. This behavior is caused by the presence of the $\epsilon^2$ term in eq. (25) which, in turn, can be attributed to the presence of edges in the attractor, since only points lying within $\epsilon$ to the edge of the set contribute to that behavior. To further illustrate this edge effect, consider a point $P_1$ in the interval $(\epsilon, 1-\epsilon)$ and a point $P_2$ in the interval $(1-\epsilon, 1)$ or the interval $(0, \epsilon)$. The probability for finding a randomly chosen point separated by a distance less than $\epsilon$ from $P_1$ is greater than that from $P_2$. This is because an interval of length $2\epsilon$ centered at $P_2$, which has nonempty intersection

![Fig. 6. Illustration for the calculation of the correlation integral $C_1(\epsilon)$ from a random data set.](image-url)
with the complement of the unit interval, contains “fewer” points than that centered at $P_1$. This fact underlies the subtraction of $\epsilon^2$ from $2\epsilon$ in eq. (25).

In higher dimensions, the edge effect is more severe. Heuristically, this can be seen as follows. Consider an $m$-dimensional unit cube. The volume of the set of points separated from the boundary by a distance less than $\epsilon$ is $1 - (1 - 2\epsilon)^m$ which approaches 1 as $m \to \infty$. This means that, proportionally, more points stay within $\epsilon$ distance from the boundary as $m$ increases.

Since the time series is uncorrelated, we have under the max-norm

$$C_m(\epsilon) = C_1(\epsilon)^m = (2\epsilon - \epsilon^2)^m.$$  \hfill (27)

The corresponding derivative is

$$\mu_m = \frac{d \log C_m(\epsilon)}{d \log \epsilon} = m - m \frac{\epsilon}{2 - \epsilon},$$  \hfill (28)

which is again a monotonically decreasing function of $\epsilon$, approximating the correct answer $\mu_m = m$ for small $\epsilon$.

In fig. 7 we plot $\log_2 C_m(\epsilon)$ versus $\log_2 \epsilon$ from eq. (27) for $m = 2 - 6$. The trend in these plots is qualitatively the same as that we observe in fig. 5 for relatively small $m$. Thus the particular pattern of deviation from linear behavior in plots of $\log_2 C_m(N, \epsilon)$ versus $\log_2 \epsilon$ in fig. 5 for relatively small $m$ is attributable to the edge effect of the reconstructed attractor. For more general situations this effect can be seen as arising in the following way. Consider two points on the attractor, one of which is deep in the interior and the other is near the boundary. Let $\epsilon$ be smaller than the distance from the interior point to the attractor boundary and larger than the distance from the point close to the boundary to the boundary. Then the number of points separated from the interior point by a distance less than $\epsilon$ is greater than that separated from the point close to the boundary by a distance less than $\epsilon$. As $\epsilon$ increases, more and more points can be considered as being near the boundary. This causes the growth of the number of pairs separated by $\epsilon$ or less to slow down for large $\epsilon$, thus furnishing an intuitive explanation of the observed behavior in fig. 5.

For relatively large $m$ we observe a very different systematic behavior in fig. 5. In particular, as $\epsilon$ increases past the upper border of the scaling region, the value of $\log_2 C_m(N, \epsilon)$ tends to rise with steeper slope than that in the scaling region, and then it levels off as the curve approaches the point $(0, 0)$. We remark that this leveling behaviour is again caused by the edge effect discussed above. We now argue that the steeper slope appears to be a result of foldings in the original attractor. To illustrate this, consider the tent map eq. (19) shown in fig. 8. The attractor in this case is the interval $[0, 1]$ with a uniform natural invariant probability measure. The folding occurs at $y = \frac{1}{2}$. Reconstructing the time series from eq. (19) by using $T = 1$ and $t$ as integers in eq. (17), we calculate the correlation integrals for $m = 1$ and $m = 2$. For $m = 1$, $C_1(\epsilon) = 2\epsilon - \epsilon^2$, which is the same as that for the random data set in eq. (25). $C_2(\epsilon)$ can be written as

$$C_2(\epsilon) = C_1(\frac{1}{2} \epsilon) + R(\epsilon).$$  \hfill (29)

To see how the first term arises consider two
points $y_j$ and $y_i$ in the time series. If $|y_j - y_i| < \frac{1}{2} \epsilon$, then $|y_{j+1} - y_{i+1}| < \epsilon$. This means that, in the two-dimensional reconstruction space, the two points, $y_{i+1} = \{y_{j+1}, y_j\}$ and $y_{i+1} = \{y_{i+1}, y_i\}$, are separated by a distance less than $\epsilon$. Thus pairs separated by a distance less than $\frac{1}{2} \epsilon$ in the unit interval give rise to pairs separated by a distance less than $\epsilon$ in the two-dimensional reconstruction space. Because of the folding at $y = \frac{1}{2}$, we may also have situations in which $|y_j - y_i| > \frac{1}{2} \epsilon$, but $|y_{j+1} - y_{i+1}| < \epsilon$. An example is a pair of points located symmetrically with respect to the point $y = \frac{1}{2}$. After one iteration the distance between the images is zero. Thus the folding in the attractor underlies the correction term $R(\epsilon)$. Let $y'$ and $y''$ be two points randomly chosen in the unit interval. $R(\epsilon)$ is defined to be the probability such that $\frac{1}{2} \epsilon < |y' - y''| < \epsilon$ and $|T(y') - T(y'')| < \epsilon$, where $T(y)$ denotes the tent map.

In the appendix we calculate $C_2(\epsilon)$. The result is

$$C_2(\epsilon) = C_1(\frac{1}{2} \epsilon) + R(\epsilon)$$

$$= \epsilon - \frac{1}{4} \epsilon^2 + \left\{ \begin{array}{ll}
\frac{1}{2} \epsilon^2 , & 0 < \epsilon < \frac{3}{2} , \\
3 \epsilon - \frac{1}{2} \epsilon^2 - 1 , & \frac{3}{2} \leq \epsilon \leq 1 .
\end{array} \right. \quad (30)$$

For $0 \leq \epsilon < \frac{3}{2}$, $d \log_2 C_2(\epsilon)/d \log_2 \epsilon = 1 + \epsilon/(2 + \epsilon)$. This derivative is 1 when $\epsilon = 0$ ($D_2 = 1$ for the attractor) and increases due to the term $\epsilon/(2 + \epsilon)$ whose presence reflects the influence of $R(\epsilon)$, which, in turn, is caused by the folding on the attractor.

Figure 9a shows plots of $\log_2 C_m(\epsilon)$ versus
log₂ₑ for m = 1 and 2. The linear dependence obtained in the scaling regions are indicated by the dashed lines. The systematic deviation from the linear behavior is qualitatively the same as that seen in fig. 5. The reason underlying the nonlinear dependence for m = 1 is due to the edge effect resulting from the finite extent of the attractor. The nonlinear deviation for m = 2 is a consequence of the folding in the attractor and the edge effect. Figure 9b shows the plots in fig. 9a for m = 1 and m = 2 (solid lines) superposed on numerically calculated correlation integrals for a time series of 50 000 points generated by eq. (19). The dimension of the reconstruction space ranges from m = 1 to m = 6. The agreement between the numerical results and the analytical results for m = 1 and m = 2 is excellent.

3.3. Dimension estimate with short data set and in the presence of noise

It frequently occurs in applications that the measured time series is short and is contaminated by noise. We consider the issue of dimension measurement under these circumstances in what follows.

Consider a time series of 2000 points generated by eq. (22). The sampling time used is tₛ = 10.0 for the attractor reconstruction. The dimension of the reconstruction space ranges from m = 2 to m = 25. A representative plot of log₂Cₘ(N, e) versus log₂ₑ is shown in fig. 10 for m = 2-6, 8, 11, 15, 19, 23. For each m, we identify an apparent linear region in the plot log₂Cₘ(N, e) versus log₂ₑ and fit a straight line through the region. The slope of this straight line, denoted Dₘₑ, is presumed to be an estimate of Dₘₑ, which is then plotted using triangles as a function of the space dimension m in fig. 4. This function attains an approximate plateau beginning at m = 16 and extending beyond m = 25. The slope averaged over the plateau is about 7.05 which is consistent with the value of 7.1 using the long data set (N = 50 000). Evidently the dimension estimates for both the short and the long data sets give similar results for m = 16-25. But the dimension estimates for the short data set fall systematically under that for the long data set for 5 ≤ m ≤ 13. This constitutes an apparent inconsistency with the prediction by theorem 2.2. In what follows we explain the origin of this inconsistency from the systematic behavior of correlation integrals discussed in section 3.2.

Refer to fig. 5. Imagine that, for a given m, we label the values of log₂ₑ and log₂Cₘ(N, e) at the upper end point of the scaling region by (LE₁(m), LC₁(m)) and that at the lower end point of the scaling region by (LE₂(m), LC₂(m)). Then D₂ₑ = (LC₁(m) - LC₂(m))/(LE₁(m) - LE₂(m)). With regard to the situation in fig. 5, LC₁(6) < LC₁(7) < LC₁(8) < LC₁(9) < LC₁(10) and LC₁(13) < LC₁(12) < LC₁(11) < LC₁(10). LC₂(m) = -30 for m = 6-13. LE₁(m) and LE₂(m) are both monotonically increasing functions of m.

The smallest value that a correlation integral assumes is 2/N². Thus the vertical range in fig. 5 is indicated by log₂2/N². For log₂2/N² = -30, N ≈ 20 000, which is on the same order of mag-
magnitude as the number of points used in fig. 3. Consider a time series that contains \( N = 2000 \) points. The plots of \( \log_2 C_m(N, \epsilon) \) versus \( \log_2 \epsilon \) for this data set roughly correspond to the portion of fig. 5 exposed above the horizontal line at \( \log_2 C_m(N, \epsilon) = \log_2[2/(2000)^2] = -20 \). Since \( L_C(6) < -20 \) and \( L_C(7) < -20 \), the correct dimension is not obtained for \( m = 6 \) and \( m = 7 \). In fact, if we fit a straight line to an apparent linear region above \( \log_2 C_m(N, \epsilon) = -20 \) for \( m = 6 \) or \( m = 7 \) the slope of this straight line will be markedly smaller than the actual dimension. However, since \( L_C(m) > -20 \) for \( m \geq 8 \), we can still obtain the correct dimension of the underlying dynamics for \( m \geq 8 \). If we plot the curve \( C(m) \) versus \( m \) for this short data set corresponding to the situation in fig. 5, the plateau onset occurs at \( m = 8 \) while \( C(6) = 6 \).

Apply the above considerations to the data set generated by eq. (22). Refer to fig. 3. Imagine that we draw a line at about \( \log_2 C_m(N, \epsilon) = -20 \). If we fit a line through an apparent linear region above the line for \( m = 8 \), the slope of this straight line is about 5.9, which is roughly the same as that of 5.8 estimated using 2000 points. Thus, by knowing the correlation integrals for a larger data set, we can roughly predict the outcome of dimension measurement for a subset of this data.

We remark that, in this paper we only consider the systematic effects in dimension estimation. For a small data set the statistical effects are also likely to be important. In applications a thorough understanding of the result from a dimension measurement procedure may require the analysis of both systematic and statistical aspects of the problem. We refer the interested reader to ref. [32–34] for some relevant information.

A much discussed problem is how many points are required for estimating a correlation dimension of value \( D_2 \) [9,31,35]. A recent result gives the following inequality [9]:

\[
N \geq 10^{D_2/2},
\]

where \( N \) is the length of the data set. This inequality first appeared in this paper as eq. (12). For an attractor with \( D_2 = 7.1 \), the above equation says that the maximum size of the data set required is on the order of 3500 points. The numerical example shown above uses only \( N = 2000 \) points. In this case a good estimate of the correlation dimension is still obtained although the extents of the scaling regions are smaller than a decade. For other examples using relatively small data sets to estimate relatively large dimensions see refs. [10,11].

The size of a data set referred to in this paper is with respect to a fixed sampling interval. Thus the length of the data string is proportional to the length of the observation interval, which, in turn, is proportional to the extent of the coverage of the chaotic attractor by the observed trajectory. For a given observation interval, arbitrary increase of the sampling frequency will not result in much improved results. Ref. [8] addresses the systematic effects introduced by high sampling frequencies.

A further observation is that, if one extends the range of \( m \) values beyond what is shown in fig. 4, at large enough \( m \), \( C(m) \) will start to deviate from the plateau behavior and increase monotonically with \( m \). This is caused by the finite length of the data set and can be understood from the systematic behavior of correlation integrals seen in fig. 5. Specifically, imagine that as the systematic trend exhibited by the curves of \( \log_2 C(N, \epsilon) \) versus \( \log_2 \epsilon \) continues, the correct scaling regions become smaller and smaller with each increment of \( m \), and eventually vanish for large enough \( m \). The lower end portion of the correlation integral for such large values of \( m \) is the segment with steeper slope caused by foldings on the original attractor. An apparent linear fit to this segment will yield a slope that is markedly larger than that of the true correlation dimension, and increases with \( m \) monotonically. A lack of sufficient data will not only delay the plateau onset, but also make the deviation from the plateau behavior occur at smaller values of
of \( m \), thus shortening the plateau length from both sides. The reason for this is that a shorter data set gives rise to shorter scaling regions (see fig. 5), thus causing the disappearance of scaling regions to occur at smaller values of \( m \).

The effect of observational noise is simulated by adding a random number \( e_i \) to each measurement \( y_i \) in the time series. The new time series is \( \{y'_i\} = \{y_i + e_i\} \). In this study we take \( e \) to be uniformly distributed in a finite interval which is denoted by \((-\epsilon_{\text{noise}}, \epsilon_{\text{noise}})\) after normalization.

The first result we show here is for a time series of 20,000 points generated by the Hénon map eq. (20). The log–log plots of the correlation integrals for both the clean time series and the contaminated time series are shown in figs. 11a and 11b, respectively. Here the dimension of the reconstruction space ranges from \( m = 2 \) to \( m = 8 \). The impact of observational noise is clearly seen in fig. 11b where the noise amplitude is indicated by a vertical line at \( \log_2 \epsilon = \log_2 \epsilon_{\text{noise}} \). For \( \epsilon < \epsilon_{\text{noise}} \) the correlation integral \( C_m(N, \epsilon) \) behaves as if the data set consisted of a sequence of uncorrelated random noise. This gives rise to the monotonic increase with \( m \) of the slopes of the linear fits in the region \( \epsilon < \epsilon_{\text{noise}} \). For \( \epsilon > \epsilon_{\text{noise}} \) the result is only slightly affected by the presence of noise. One can still roughly estimate the correct dimension by fitting a straight line through an apparent scaling region on the right side of the vertical line.

Our second result is for the time series of 50,000 points used in obtaining fig. 3. The ratio of the noise amplitude to the signal amplitude is about 10%. To reconstruct the attractor we use \( T = 20.0 \) as the delay time. The dimension of the reconstruction space ranges from \( m = 2 \) to \( m = 25 \). Representative plots of \( \log_2 C_m(N, \epsilon) \) versus \( \log_2 \epsilon \) are shown in fig. 12 for \( m = 2–8, 11, 15, 19, 23 \). The contaminated region and the uncontaminated region are divided by the vertical line in the figure. Identify an apparent linear region for a given \( m \) for \( \epsilon < \epsilon_{\text{noise}} \) and an apparent linear region for \( \epsilon > \epsilon_{\text{noise}} \). The slopes of the straight line fits in these two regions as functions of \( m \) are shown in fig. 13a. The slopes of the linear regions for \( \epsilon > \epsilon_{\text{noise}} \) attain an approximate plateau beginning at \( m = 12 \) and extending beyond \( m = 25 \); while the slopes of the scaling regions for \( \epsilon < \epsilon_{\text{noise}} \) increase monotonically with \( m \). The slope averaged across the plateau is about 7.5, which is consistent with that of 7.1 in fig. 4. Assume that we only use the portion with \( \epsilon > \epsilon_{\text{noise}} \) for dimension estimation. Then the
plateau onset for $\tilde{D}_2^m$ versus $m$ is delayed by the presence of observational noise. But the plateaued value $\tilde{D}_2$ gives a good estimate of the correlation dimension of the underlying dynamics. We contrast the situation with noise with the situation without noise in fig. 13b where the crosses in fig. 13a are superposed on the open-circle data in fig. 4.

To explain fig. 13 we again refer to the systematic behavior of correlation integrals. As discussed earlier, the scaling region for a given $m$ is defined by an upper point $(\text{LE}_1(m), \text{LC}_1(m))$ and a lower point $(\text{LE}_2(m), \text{LC}_2(m))$. Specifically, let $m = \text{Ceil}(D_2)$. If $\text{LE}_1(m) < \log_2 \epsilon_{\text{noise}}$, the scaling region for this particular $m$ is contaminated by the noise. The slope for an apparent linear region above the noise level is markedly smaller than the correct estimate of $D_2$. On the other hand, if $\text{LE}_1(m) > \log_2 \epsilon_{\text{noise}}$, we can still expect to measure a reasonable slope of the curve over the horizontal interval $(\log_2 \epsilon_{\text{noise}}, \text{LE}_1(m))$, although this slope represents a slight overestimate of the correct dimension.

An interesting implication of fig. 13 is that, for a time series which is contaminated by observational noise, it is desirable to go to higher reconstruction dimensions to minimize the impact of noise.

4. Dimension preservation results

4.1. Correlation dimension

To discuss the ideas behind the dimension preservation results we first present an alternative definition of the correlation dimension. Consider an $n$-dimensional Euclidean space and let $\rho$ denote a probability measure supported on a compact set $A$. If $A$ is a chaotic attractor then $\rho$ is the natural measure induced by the dynamics. We define the correlation dimension of $\rho$, $D_2(\rho)$, to be the supremum of all real numbers $s$ such that the average value
\[ \langle |x-y|^{-s} \rangle_\rho \] (31) is finite, where the average is taken with respect to pairs \((x, y)\) chosen independently from \(\rho\). Here the quantity \(|x-y|^{-s}\) is called the Riesz kernel and the average in eq. (31) is called the \(s\)-energy of the measure \(\rho\). Thus, in words, the correlation dimension of a measure is the supremum of \(s\) for which the \(s\)-energy of the measure is finite.

A virtue of this definition of correlation dimension is that because the average in eq. (31) is a nondecreasing function of \(s\), the supremum is always defined. In the case where the correlation integral \(C(\varepsilon)\) scales with \(\varepsilon\) as \(\varepsilon\) raised to some power, this definition yields the same value of \(D_2\). To see this, let \(C(\varepsilon)\) be the correlation integral for the measure \(\rho\). In the limit \(\varepsilon \to 0\), \(C(\varepsilon)\) scales with \(\varepsilon\) as

\[ C(\varepsilon) \sim \varepsilon^{D_2}. \] (32)

By definition \(C(\varepsilon)\) is a monotonically increasing positive function of \(\varepsilon\). Consider the supremum of all the real numbers \(s\) such that the average in eq. (31) converges. This translates to

\[ \int_0^1 \frac{dC(\varepsilon)}{\varepsilon^s} \, d\varepsilon \sim \int_0^1 \varepsilon^{D_2-s} \, d\varepsilon < \infty. \] (33)

Clearly the integral converges for all \(s < D_2\). Thus the supremum is \(D_2\). Equation (33) shows that when eq. (32) holds, the definition eq. (31) gives the same correlation dimension.

Now we state and discuss our main results on correlation dimension preservation under projections and under delay coordinate maps.

**Theorem 4.1.** Let \(\rho\) be a compactly-supported probability measure on \(\mathbb{R}^n\) which satisfies \(D_2(\rho) \leq m\). For almost every \(C^1\) function \(F:\mathbb{R}^n \to \mathbb{R}^m\), \(D_2(F(\rho)) = D_2(\rho)\), where \(F(\rho)\) denotes the measure induced on \(\mathbb{R}^m\) by \(F\).

The induced measure on \(\mathbb{R}^m\) is defined to be

\[ F(\rho)(S) = \rho(F^{-1}(S)) \] (34)

where \(S\) is a subset of \(\mathbb{R}^m\). If we consider the measure \(\rho\) to be the natural measure of a chaotic attractor, a corollary immediately follows:

**Corollary 4.2.** Let \(A\) be a compact attractor in \(\mathbb{R}^n\) with natural measure \(\rho\) which satisfies \(D_2(\rho) \leq m\). Then for almost every \(C^1\) function \(F: \mathbb{R}^n \to \mathbb{R}^m\), \(D_2(F(\rho)) = D_2(\rho)\).

The term “almost every” in the above statements is understood in the sense of prevalence. More precisely, let \(\alpha_1, \alpha_2, \ldots, \alpha_k\) be \(k\) scalar parameters, and let \(F_1, F_2, \ldots, F_k\) be the basis functions of a \(k\)-dimensional subspace of the infinite-dimensional space spanned by all \(C^1\)-functions from \(\mathbb{R}^n\) to \(\mathbb{R}^m\), then there is a finite-dimensional cube in this subspace \(\sum_{i=1}^k \alpha_i F_i : 0 \leq |\alpha_i| \leq 1\) of \(C^1\)-functions from \(\mathbb{R}^n\) to \(\mathbb{R}^m\) such that for any \(C^1\)-function \(F_o: \mathbb{R}^n \to \mathbb{R}^m\), the perturbed function \(F_\alpha = F_0 + \sum_{i=1}^k \alpha_i F_i\) gives rise to \(D_2(F_\alpha(\rho)) = D_2(\rho)\) for almost every \(\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_k\}\) in the \(k\)-dimensional cube. In particular, prevalent implies dense, so the theorem and corollary also hold in the weaker sense of a dense subset of \(C^1\)-functions \(F\). If the basis function set is enlarged by adding a new set of functions, \(F_{k+1}, \ldots, F_K\), and the parameter set is increased by adding \(\alpha_{k+1}, \ldots, \alpha_K\), the result is still true.

We sketch a proof of theorem 4.1, and refer the interested reader to ref. [36] for more details. Dimension cannot increase under a function \(F\) with bounded stretching (e.g., differentiable functions), but it can decrease. We argue that the set of parameters \(\alpha\) for which \(F_\alpha\) decreases correlation dimension is a measure zero set. Here we choose \(\alpha\) from a unit cube of dimension \(k = mn\), and assume that \(\{F_i\}\) are a basis for the linear functions from \(\mathbb{R}^n\) to \(\mathbb{R}^m\).

We will show that for each pair \(x \neq y\) in \(\mathbb{R}^n\), and for \(s < m\), the inequality

\[ \int_C \frac{d\alpha}{|F_\alpha(x) - F_\alpha(y)|^s} \leq \frac{B}{|x-y|^s} \] (35)

This completes the proof of theorem 4.1.
holds, where $B$ is a finite constant and $C$ is the $k$-dimensional unit cube of the parameters $\alpha$. Given this inequality, we can average both sides with respect to pairs $(x, y)$ chosen from the measure $\rho$. The average is finite for $s < D_2$ by assumption. Reversing the order of integration of the left-hand side, we find that the average over the unit cube $C$ of the $s$-energy of the measure $F_\alpha(\rho)$ is finite for $s < D_2$. Therefore the $s$-energy must be finite for all but a measure zero set of $\alpha$.

This completes the proof. Now we proceed to argue the validity of the inequality in eq. (35). Assume for simplicity that $x - y$ lies along the $e_1$ coordinate axis in $\mathbb{R}^n$, and $F_1(e_1), \ldots, F_m(e_1)$ give the unit vectors along the $m$ coordinate axes in $\mathbb{R}^m$. Then

$$F_\alpha(x) - F_\alpha(y) = h + \sum_{i=1}^{m} \alpha_i F_i(x - y),$$

where $h = F_0(x) - F_0(y) + \sum_{i=m+1}^{m} \alpha_i F_i(x - y)$.

Considering the unit cube as the product $C = C_m \times C_{m-n}$, where $C_m$ corresponds to the unit cube of $\mathbb{R}^m$ spanned by $F_1, \ldots, F_m$, results in the inequality

$$\int_{C_m} \frac{d\alpha_1 \cdots d\alpha_m}{|F_\alpha(x) - F_\alpha(y)|^s} \leq \frac{1}{|x - y|^s} \int_{C_m} \frac{dr}{r^s}.$$ (37)

This inequality relies on the fact that the upper bound of the integrand on the left is given by the worst case possibility $h = 0$. The integration on the right-hand side converges if and only if $s < m$. Finally, integrating both sides over $C_{m-n}$ yields eq. (35).

The following results are proved for the special class of maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ which are delay coordinate maps.

**Theorem 4.3.** Let $A$ be a compact set in $\mathbb{R}^n$ with probability measure $\rho$. Let $A$ and $\rho$ be invariant under the diffeomorphism $G: \mathbb{R}^n \rightarrow \mathbb{R}^n$ with no periodic points of period $\leq m$ in $A$. For a function $h: \mathbb{R}^n \rightarrow \mathbb{R}$, the delay coordinate map $G_h: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is defined by

$$F_h(x) = [h(x), h(G^{-1}x), \ldots, h(G^{-(m-1)}x)].$$

(38)

If $D_2(\rho) \leq m$, then for almost every $h$, $D_2(F_h(\rho)) = D_2(\rho)$.

**Corollary 4.4.** Assume $A$ is an invariant set of $G$ whose periodic points with period $\leq m$ comprise a set of correlation dimension strictly less than $D_2(\rho)$. If $D_2(\rho) \leq m$, then for almost every $h$, $D_2(F_h(\rho)) = D_2(\rho)$.

The version stated in theorem 2.2 in section 2 is weaker than corollary 4.4 but stronger than theorem 4.3.

The proof of theorem 4.3 is similar to that of theorem 4.1, except that for a given delay coordinate map $F_h$, we want to know that almost every perturbation within the subspace of delay coordinate maps preserves dimension. In the absence of periodic points of period at most $m$, we can assume that the points $x, G^{-1}(x), \ldots, G^{-(m-1)}(x)$ are distinct. Then the role played by linear functions in the proof of theorem 4.1 can be filled by interpolating polynomials. Again see ref. [36] for more details. The corollary follows from applying the theorem to the complement of the set of low period orbits.

### 4.2. Other dimensions

The above dimension preservation results also hold for the Hausdorff dimension $D_H$. This can be seen from the following definition of $D_H$. Let $A$ be a Borel set of $\mathbb{R}^n$, that is, a set that can be obtained by taking countable unions, intersections, and complements of open balls in $\mathbb{R}^n$. Assume that $A$ is bounded, then the Hausdorff dimension $D_H(A)$ is the supremum of the real numbers $s$ satisfying

$$\langle |x - y|^{-s} \rangle_\mu < \infty,$$ (39)

where the supremum is taken over not only $s$ but also all probability measures compactly support-
ed on $A$ (see ref. [37]). Thus the Hausdorff dimension of a set can be defined in terms of its correlation dimension. The Hausdorff dimension of $A$ is the largest correlation dimension for probability measures compactly supported on $A$. Namely,

$$D_H(A) = \sup \{ D_2(\rho) : \rho(A) = 1 \text{ and } \rho(A^c) = 0 \} ,$$

(40)

where $A^c$ denotes the complement of $A$. The procedure of proving the dimension preservation for $D_H$ is based on eq. (40).

Finally, we mention that the above dimension preservation results do not hold for the box counting dimension $D_0$. In ref. [36] an example is given of a compact subset $Q$ of $\mathbb{R}^{10}$ with $D_0(Q) = 5$, such that for each $C^1$ function $F: \mathbb{R}^{10} \to \mathbb{R}^6$, $D_0(F(Q)) < 4$.

5. Conclusions

The main contributions of this paper are as follows.

1. We present rigorous results concerning the preservation of dimensions under projections and under delay coordinate maps. In particular, we show that, for a large enough data set, the correlation dimension $D_2^{(m)}$ as a function of the reconstruction space dimension $m$ reaches a plateau beginning at $m = \text{Ceil}(D_2)$, where $\text{Ceil}(D_2)$, standing for ceiling of $D_2$, is the smallest integer greater than or equal to $D_2$.

2. We show numerical examples which illustrate the prediction in 1.

3. We report new results on the behavior of correlation integrals. In particular, we relate the systematic behavior exhibited by these correlation integrals to the finite extent of the attractor and to the folding taking place on the attractor.

4. From the behavior of correlation integrals we discuss how factors such as a lack of data and observational noise can produce results that may seem to be inconsistent with point 1 above.

Acknowledgement

This work was supported by the National Science Foundation (Divisions of Mathematical and Physical Sciences) and the US Department of Energy (Scientific Computing Staff, Office of Energy Research). M. Ding’s research is also supported in part by a grant from the National Institute of Mental Health and a grant from the Office of Naval Research.

Appendix. Derivation of $C_2(\epsilon)$ for the tent map

In what follows we calculate $C_2(\epsilon)$ for the tent map. Consider the plane spanned by the two points $y'$ and $y''$. The region in this plane corresponding to $|y' - y''| < \epsilon$ is the shaded region shown in fig. 6. The area of this region is $C_1(\epsilon) = 2\epsilon - \epsilon^2$. The region in the $y'$-$y''$ plane such that $y' > \frac{1}{2}$, $y'' < \frac{1}{2}$, and $|T(y') - T(y'')| < \epsilon$ is given by

$$|2(1 - y') - 2y''| < \epsilon$$

or

$$|1 - y' - y''| < \frac{1}{2} \epsilon ,$$

which is shown shaded in fig. 14a. Similarly, the region such that $y' < \frac{1}{2}$, $y'' > \frac{1}{2}$, and $|T(y') - T(y'')| < \epsilon$ is shown in fig. 14b. The shaded regions in figs. 15a and 15b correspond to the region defined by $y' < \frac{1}{2}$, $y'' < \frac{1}{2}$, and $|T(y') - T(y'')| < \epsilon$, and the region defined by $y' > \frac{1}{2}$, $y'' > \frac{1}{2}$, and $|T(y') - T(y'')| < \epsilon$, respectively. Figure 16 shows the entire region in the $y'$-$y''$ plane such that $|T(y') - T(y'')| < \epsilon$. By definition, $C_2(\epsilon)$ is the area of the region in the $y'$-$y''$ plane such that both $|T(y') - T(y'')| < \epsilon$ and $|y' - y''| < \epsilon$. This region is the intersection of the two shaded regions in figs. 6 and 16.
There are two cases. In case 1 the intersection is as shown in fig. 17a. $a < b$, where $a = \varepsilon$ and $b = 1 - \frac{1}{2}\varepsilon$, in this case. Clearly, $C_2(\varepsilon)$ is the sum of the area of the vertically hatched region and the area of the horizontally hatched region. The first area is $C_1(\frac{1}{2}\varepsilon) = \varepsilon - \frac{1}{2}\varepsilon^2$. The second area is $R(\varepsilon)$ which is calculated to be $R(\varepsilon) = \frac{1}{2}\varepsilon^2$. In case 2 the intersection is as shown in fig. 17b. In this case $a \geq b$ (or $\varepsilon > \frac{1}{2}$). The area of the vertically
hatched region is $C_1(\frac{1}{2} \epsilon)$ as before. The area of the horizontally hatched region denoted by $R(\epsilon)$ is the previously horizontally hatched area, $\frac{1}{2} \epsilon^2$, less the total area of the four darkened triangles. That is,

$$R(\epsilon) = \frac{1}{2} \epsilon^2 - (a - b)^2 = 3 \epsilon - \frac{7}{4} \epsilon^2 - 1.$$

Thus

$$C_2(\epsilon) = C_1(\frac{1}{2} \epsilon) + R(\epsilon)$$

$$= \epsilon - \frac{1}{4} \epsilon^2 + \begin{cases} \frac{1}{2} \epsilon^2, & 0 < \epsilon < \frac{1}{3} \\ 3 \epsilon - \frac{7}{4} \epsilon^2 - 1, & \frac{1}{3} \leq \epsilon \leq 1. \end{cases}$$

References