Practical Considerations in Estimating Dimension from Time Series Data

Eric J. Kostelich and Harry L. Swinney

Center for Nonlinear Dynamics and Department of Physics, University of Texas, Austin, Texas 78712, U.S.A.

Abstract
A numerical comparison is made of algorithms for computing the dimension of attractors using the Grassberger-Procaccia correlation dimension and the Badii-Politi nearest neighbor approach. Using experimentally realizable data sets, the nearest neighbor method appears to have the better power law behavior when the attractor dimension is between 3 and 7.

1. Introduction
The estimation of the dimension of chaotic attractors reconstructed from time series data has become an important tool in the analysis of data in many fields, ranging from optics [1, 2] and neurobiology [3] to economics [4]. There are many possible characterizations of dimension; a comprehensive survey is given by Farmer, Ott and Yorke [5]. Grassberger and Procaccia [6] have introduced the notion of correlation dimension, which has become a popular method for estimating dimension from time series data. In this case, the dimension is given from a power law, namely, the fraction of point pairs closer than \( \epsilon \) scales as \( \epsilon^d \), where the exponent \( d \) is the correlation dimension. The algorithm is easily programmed on a computer and gives accurate estimates of attractor dimension using modest amounts of experimental data, as long as the dimension of the attractor is less than 3 or so [7]. However, it is difficult to find a power law from experimental data if the attractor dimension is larger than 4. For this reason, the correlation dimension has been used primarily in situations near the onset of chaos, e.g., where the experimental system has bifurcated from a periodic or quasiperiodic regime to low-dimensional chaos. The purpose of this paper is to compare the correlation dimension method with an approach based on nearest neighbor distances suggested by Badii and Politi [8]. An important question is whether reasonable estimates of dimension can be deduced from laboratory data for attractors whose dimension is larger than 4. In the next section we give a brief review of the two definitions of dimension. We then compare them using some experimental data sets obtained from a Couette-Taylor fluid flow experiment and discuss a numerical experiment with a 6.8-dimensional attractor.

2. Definitions of dimension
Most definitions of dimension use the notion of a covering of the attractor by boxes of side length \( \epsilon \), i.e., the smallest number of \( \epsilon \)-boxes needed so that each point of the attractor is contained in some box. Let \( p_i \) be the probability that a point chosen “at random” on the attractor lies in the \( i \)th box. Let \( I(\epsilon) = \sum_i p_i \log p_i \). (The quantity \( I(\epsilon) \) can be considered as the amount of information necessary to specify the state of the system within an accuracy \( \epsilon \); the physical significance of this sum is discussed in [10] and [11].) The information dimension is defined as

\[
D_1 = \lim_{\epsilon \to 0} \frac{- I(\epsilon)}{\log \epsilon}.
\]

(1)

This definition can be modified slightly to give the correlation dimension \( D_2 \) introduced by Grassberger and Procaccia, viz.,

\[
D_2 = \lim_{\epsilon \to 0} \frac{\log \sum_i p_i^2}{\log \epsilon}.
\]

(2)

Other generalizations of dimension are discussed in [12]. It can be shown [12] that \( D_2 \leq D_1 \), but the numerical values of the two dimensions are usually equal within the measurement precision for attractors reconstructed from experimental data; therefore, we will use the same symbol, \( d \), for both types of dimension.

The correlation dimension can also be defined in the following manner, which gives an algorithm for its computation. Suppose one is given \( N \) points on the attractor. Define the correlation integral [6]

\[
C(N, \epsilon) = \frac{1}{N^2} \times \{ \text{number of point pairs } x_i, x_j \text{ which are closer than } \epsilon \text{ apart} \}.
\]

(Any convenient metric can be used, e.g., the Euclidean metric.) Under suitable assumptions, it can be shown [6] that

\[ C(N, \epsilon) \sim \epsilon^d \]

for large \( N \) and small \( \epsilon \), where the exponent \( d \) is the correlation dimension. Because it is prohibitively time-consuming to compute the distances between every pair of points when \( N \) is large, \( C(N, \epsilon) \) typically is estimated using a small number of reference points where the normalization factor is adjusted accordingly. Hence this method counts the number of points in a ball about each reference point.

A different approach, which tallies nearest neighbor distances, has been examined by Badii and Politi [8]. As before, we consider a set of \( N \) points on the attractor and let \( x \) be a reference point. For \( k < N \), let \( \{ y_j \} \) be a set of \( k \) distinct points (different from \( x \)), chosen at random from the original set of \( N \) points. Let

\[
\delta_k = \min_{1 \leq j \leq k} \| x - y_j \|.
\]

(3)

The distance \( \delta_k \) is computed for a sequence of \( k \) values up to \( N \); the calculation is repeated for many reference points \( x \). It can be argued [8] that

\[
\left< \delta_k \right> \sim k^{-1/d}
\]

(4)
for large $N$, where $d$ is the information dimension and the angle brackets represent the average value of $\delta_0$ over all the reference points. Various generalizations of this idea are possible; see [8] for details.

The power law in both dimension methods is a limit as the number of data points becomes infinite, reflecting the scaling behavior of the attractor as distances between points becomes infinitesimal; this has important implications when the methods are applied to experimental data. For example, in practice the correlation integral $C(N, \varepsilon)$ obeys a power law only for a certain range of values of $\varepsilon$. Large values of $\varepsilon$ are unsuitable because the limit is poorly approximated. Very small values of $\varepsilon$ must also be avoided because the scaling behavior at that level is governed by noise.

Likewise, one must modify the nearest-neighbor method slightly for experimental data. Because of noise, distances between reference points and their nearest neighbors do not reflect the fractal structure of the attractor. A power law for this method can be discerned only if one considers nearest neighbors that are sufficiently far away. That is, the definition of $\delta_0$ becomes the distance from the reference point to its $n$th nearest neighbor, where $n$ is the order of the nearest neighbor and varies from 10 to 300 in the examples discussed below.

The determination of whether an approximate power law holds for some range of the fitting parameters is an essential problem in calculating dimension from experimental data. The criteria are subjective, and a rule which works for one attractor may not be applicable for another. In the next section, we present preliminary results using experimental and numerical data which suggest that in many cases, especially when the attractor dimension is larger than 3, a power law can be more apparent in the nearest-neighbor approach than in the usual correlation method.

3. Application to experimental data

Attractors are reconstructed from time series by the method of time delays [13], which by now is a standard technique. The $j$th point on the attractor is given by

$$x_j = (x(j), x(j + \tau), \ldots, x(j + (m - 1)\tau))$$

where $x(j)$ is the $j$th value in the time series, $\tau$ is the time delay, and $m$ is the embedding dimension. The dimension of the attractor can be recovered if $m$ is large enough.

Theoretically, the choice of the time delay is arbitrary, but in practice it strongly affects the dimension estimates. The effect of the time delay can be understood heuristically as follows. If the signal is sampled relatively quickly, then the $(k + 1)$st measurement is nearly the same as the $k$th. Thus, the reconstructed attractor lies nearly on a straight line if $\tau = 1$, i.e., if the coordinates consist of consecutive measurements. Conversely, if $\tau$ is too large, the coordinates are completely independent and the structure of the reconstructed attractor is hard to discern. Ideally, $\tau$ should be chosen so that the coordinates are reasonably independent, which makes the fractal features of the attractor apparent at scales above the noise level. This issue is discussed in [14], where an information-theoretic criterion, called the mutual information, is given for choosing time delays. The method outlined in that paper has been used to choose $\tau$ in each of the examples discussed here.

In this section we consider some of the problems involved with both the correlation and nearest-neighbor algorithms as they are applied to measurements of fluid velocity in a Couette–Taylor experiment. The fluid is contained between concentric cylinders with the inner cylinder rotating and the outer cylinder at rest. The Reynolds number for the Couette–Taylor system can be defined as $R = a\Omega(b - a)/\nu$, where $a$ and $b$ are, respectively, the radii of the inner and outer cylinders, $\Omega$ is the angular velocity of the inner cylinder, and $\nu$ is the kinematic viscosity. For our system, $a = 5.205\text{cm}$ and $b = 5.947\text{cm}$, which gives a radius ratio of 0.875. At this radius ratio, the critical Reynolds number for the onset of Taylor vortex flow in an infinite system is $R_c = 118.4$ [15]. In this experiment, the ratio of the fluid height to the gap between the cylinders is 20.0.

The experimental data consist of time series of laser Doppler measurements of the radial component of velocity at points midway in the gap between the inner and outer cylinders. The velocity data files each contain 32768 points. More details of the experiment are given in [16].

Figure 1 shows the results of applying the correlation dimension approach to these experimental data. Figure 1(a) is for a modulated wavy vortex flow regime, which is quasiperiodic, for a Reynolds number of $R/R_c = 10.1$. Figure 1(b) shows the results for a weakly turbulent flow at $R/R_c = 12.9$, and Fig. 1(c) is for a moderately turbulent flow at $R/R_c = 15.0$. Each attractor has been reconstructed in 8 dimensions. The correlation integral $C(N, \varepsilon)$ is approximated with 1000 reference points and 40 values of $\varepsilon$ which are equally spaced logarithmically from $2^{-1} = 1/2$ to $2^{-6} = 1/64$. (Here $\varepsilon$ is normalized as a fraction of the time series extent.)

The log–log plot in Fig. 1(a) shows a straight line, clearly indicating a power law. However, there is some deviation from the power law for the largest and the smallest $\varepsilon$s. The small-$\varepsilon$ deviation is due to noise and a lack of data; the large-$\varepsilon$ deviation reflects the gross shape of the attractor.

This deviation is quantified by performing a least-squares fit through each set of six consecutive ($\log_2 \varepsilon$, $\log_2 C(N, \varepsilon)$) pairs. The slope of the regression line is plotted as a function of the smallest $\varepsilon$ used in each regression in the second plot in each row. This "local slope" varies from 2.05 to 2.4, with a nearly constant value of 2.07 for 1/32 $\leq \varepsilon \leq 1/8$.

Ideally, such a "plateau" in the slope is evident for a wide range of $\varepsilon$ values, but often this plateau is not very wide. The analysis is repeated for some reasonable range of embedding dimensions ($4 \leq m \leq 10$ in this example), in the hope that the plateau value will become constant once $m$ is large enough. (One looks for a zero second derivative of the curve in Fig. 1(a).) Figure 1(c) shows a constant value of the attractor dimension for all values of the embedding dimension. The plateau value is computed from the local slope plot by finding the six consecutive $\varepsilon$ pairs which were most nearly horizontal, based on a linear regression. The values of the slope at those $\varepsilon$ values are averaged together, and that value is plotted as a function of embedding dimension in Fig. 1(c).

The computed dimension of 2.07 differs slightly from the "correct" dimension value of 2 for a quasiperiodic attractor. This discrepancy presumably can be attributed to noise in the data (which is unfiltered) and to numerical roundoff. This example suggests that dimension estimates accurate to 3–5% can be achieved with low dimensional laboratory data.

As the graphs in Fig. 1(b) and (c) show, the plateau region for the slope of the log–log plot shrinks rapidly as the dimen-
ension of the attractor increases (and the number of data points $\sim 32$ K here) remains the same). For weakly chaotic Couette–Taylor flow in Fig. 1(b), a plateau in the slope is still evident for $1/10 \lesssim \epsilon \lesssim 1/4$, yielding a dimension of about 2.4. The criterion given above for finding the plateau value shows relatively constant results for embedding dimensions larger than 5.

The more turbulent case in Fig. 1(c) has almost no plateau for any embedding dimension. The usual cause of this lack of convergence is the presence of a relatively high dimensional attractor and a small data set. (This will be discussed in more detail in the next section.) Large noise levels have a similar effect, because the noise “fills out” the attractor at small enough scales. Assuming that the experimental noise is about the same between the different runs, we attribute the lack of convergence to a high-dimensional attractor (say $d > 3.5$). However, only a qualitative estimate of the dimension is possible in this example.

Figure 2 shows the results of the nearest-neighbor approach on the same experimental data used above. Using 1000 reference points, the average distance $\langle \delta_k \rangle$ from each reference point to selected nearest neighbors is computed using 32 equally-spaced values of $k$ from 1024 to approximately 32000. In the following analysis, nearest neighbors of orders 10 to 300 are computed. That is, a sequence of averages $\langle \delta_k \rangle$ is computed using all the reference points; the first value in the sequence is the average distance to the 300th nearest neighbor.

Figure 2(a) shows the results for the attractor reconstructed from the quasiperiodic flow time series. The embedding dimension is 8 and the distances are for the 100th nearest neighbor. As with the correlation dimension, a power law is apparent. The dimension is taken as the negative reciprocal of the regression line through all of the $(\log_2 k, \log_2 \langle \delta_k \rangle)$ pairs. A similar calculation is done for each nearest neighbor order starting at 10 and proceeding from 25 in steps of 25 up to order 300. The results are shown in the middle plot, where the top line is for the 10th nearest neighbor, the next for the 25th, and so on down to the 300th nearest neighbor. Accurate dimension estimates are possible only with high-order nearest neighbors, presumably because the average distance between the reference points and the first 25 nearest neighbors is in the noise scale of the data. The rightmost plot shows the estimated dimension as a function of nearest neighbor order, using the 8-dimensional reconstruction. The nearest-neighbor dimension values agree with the correlation dimension, about 2.07 for this data set. As before, we attribute the error to noise in the data and numerical artifacts.

Figures 2(b)–(c) are similar plots for the weakly and moderately turbulent Couette–Taylor data, respectively. The dependence of the dimension estimates on nearest neighbor order is especially evident in Fig. 2(b). The estimated dimen-
Fig. 2. Nearest neighbor dimension for Couette-Taylor data. Each row of graphs shows, from left to right: a log-log plot of the average distance \( \langle \delta_k \rangle \) from each reference point to its 100th nearest neighbor as a function of \( k \) for an 8-dimensional reconstruction, the dimension computed from the slope of this curve for embedding dimensions from 4 to 18, using nearest neighbors of order 10, 25, 50, \ldots, 300; the dimension as a function of nearest neighbor order for the 8-dimensional reconstruction (10-dimensional reconstruction in (c)). The Reynolds numbers are (a) \( R/Re = 10.1 \); (b) \( R/Re = 12.9 \); (c) \( R/Re = 15.0 \).

sion agrees with the correlation dimension only for nearest neighbors of order 100 or more. Lower-order nearest neighbors give inaccurate dimension estimates whose values change little as the embedding dimension increases. This example shows that careful variation of parameters is essential for good results!

The convergence of the nearest-neighbor algorithm is not as good for the moderately turbulent data, because the attractor is higher dimensional and the data set is relatively short. Nevertheless, the results suggest a dimension between 4 and 5, an estimate which is not possible for the correlation dimension.

In order to understand these dimension results better, we applied both methods to a synthetic data set from an attractor which is about 6.8 dimensional. The next section discusses the numerical experiment, and the final section presents some conclusions.

### 4. Numerical experiments

The results above suggest that the attractor dimension is larger than 4 for the moderately turbulent Couette-Taylor flow. In this section we consider some numerical tests to determine whether it is possible to get reasonable estimates of attractor dimension when \( d > 4 \) for reasonable amounts of laboratory data. For this purpose, we tested the performance of the two dimension algorithms on a relatively high dimensional data set, obtained by integrating the Mackey–Glass time delay equation [17].

The Mackey–Glass delay equation is

\[
\dot{x}(t) = \frac{ax(t - s)}{1 + (xt - s)^c} - bx(t).
\]  

(6)

Equation 6 was approximated by a set of 1272 coupled ordinary differential equations using a fixed-step, fourth-order Runge–Kutta method with a time step of 0.05. The parameters were fixed at \( a = 0.2 \), \( b = 0.1 \), and \( c = 10 \); the time step corresponds to a delay \( s = 63.6 \). The first seven Lyapunov exponents were computed according to a numerical scheme described in [18]; the values are given in Table I.

<table>
<thead>
<tr>
<th>Table I. Numerically computed Lyapunov exponents for the Mackey–Glass equation. Parameters are given in the test following eq. (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_1 )</td>
</tr>
<tr>
<td>( \lambda_2 )</td>
</tr>
<tr>
<td>( \lambda_3 )</td>
</tr>
<tr>
<td>( \lambda_4 )</td>
</tr>
<tr>
<td>( \lambda_5 )</td>
</tr>
<tr>
<td>( \lambda_6 )</td>
</tr>
<tr>
<td>( \lambda_7 )</td>
</tr>
</tbody>
</table>
The Lyapunov dimension [5] is \( d_l = 6.8 \), assuming that a conjecture by Kaplan and Yorke holds for this attractor, then the information dimension should be 6.8 also [5].

A time series of \( 2^{18} = 262,144 \) values was obtained by storing the \( x_t \) coordinate of the numerically computed solution every 40 integration steps. Gaussian noise with a variance equal to 0.5% of the numerical range was added, and the result was converted to 16-bit integers. The numerical integration began with a constant initial condition and continued for 50,000 time steps before the \( x_t \) coordinates were output.

Figure 3 shows the results for the correlation dimension applied to this data set. It is difficult to determine whether or over what range of \( \varepsilon \) a power law behavior occurs in Fig. 3(a). The plot is curvilinear, and the slope varies continuously from 5 to 10, with no apparent plateau, as Fig. 3(b) indicates.

Evidently, the convergence of the correlation dimension is poor, which can be attributed to a lack of data. Even though there are more than 250,000 points on the attractor, its dimension is so large that accurate statistics about its small-scale structure are not possible. The same results are seen in Fig. 1(c) above for the moderately turbulent Couette data, where 32 K points are distributed on a 3.5 or greater dimensional set. In general, the number of points needed to estimate dimension grows exponentially with the dimension [5, 16].

Although they are not a comprehensive survey of data requirements, these results suggest that the best results from the correlation dimension are obtained in cases where \( d < 4 \) for data sets of the size considered here.

The power law shown in Fig. 4(a) for the nearest-neighbor algorithm should be compared with Fig. 3(a), which shows the power law from the correlation dimension approach. (The dimension here was estimated from the nearest-neighbor distances by regressing over the 40 equally-spaced values of \( k \) from 100,000 to 262,000, which represents the right-hand third of the curve; however, results for values of \( k \) down to 1024 are plotted.) The computed value of the information dimension is relatively constant for \( m > 10 \) and nearest neighbors of sufficiently high order. The dimension estimates decrease monotonically as the order increases as Fig. 4(c) indicates, but all values are accurate within 10% of the

\[ d_0 = 6 + (\lambda_1 + \cdots + \lambda_N)/\lambda_1 \approx 6.8 \] for the values of \( \lambda_i \) given in Table I.

Fig. 3. Correlation dimension for Mackey-Glass attractor. (a) A log-log plot of the correlation integral \( C(N, \varepsilon) \) as a function of \( \varepsilon \) for a 10-dimensional attractor reconstruction. The distance \( \varepsilon \) is normalized as a fraction of the time series extent. (b) The slope of the curve in (a) computed from a linear regression through six points at each \( \varepsilon \).

Lyapunov dimension (6.8) when the order is larger than 50 or so.

The estimate of the dimension varies according to the order of the nearest neighbor, particularly in the higher-dimensional attractors, as the results for Mackey–Glass attractor in Fig. 4(b) and the moderately turbulent Couette data in Fig. 2(c) show. Presumably the convergence would be better with more data points. Nevertheless, the nearest-neighbor algorithm seems to be more efficient (the data requirements are less) than the correlation dimension approach when the attractor dimension is relatively high.

5. Discussion

A new method of computing information dimension, based on the computation of nearest neighbor distances, has been applied to experimental data and compared to another method for computing dimension, the so-called correlation dimension. Both methods work well when applied to low-dimensional (3 or less) chaotic attractors using modest data sets (32 K points). However, the convergence of these dimension algorithms deteriorates as the dimension of the attractor increases and the number of data points remains the same. A more detailed study of the data requirements of the two methods will be presented elsewhere [19]. However, the convergence of the nearest neighbor method seems better for high dimensional attractors, since power law behavior is clearer than with the correlation dimension. Nevertheless, both methods require large data sets, especially when the attractor dimension is larger than 3. Dimension estimates accurate to within 10% for the 6.8-dimensional Mackey–Glass attractor require about 250,000 data points using the nearest-neighbor approach. Such large data sets may not be obtainable in many experimental situations.

Both methods have advantages and drawbacks. The correlation dimension in principle is probably the easiest to compute. It works especially well when the dimension \( d < 3 \) using modest data sets (32 K points). The primary disadvantage is that the power law behavior deteriorates rapidly as the dimension increases and one is left to decide whether and where a power law dependence exists as a function of the ball size \( a \). If \( a \) is too small, the behavior is governed mainly by noise, so the dimension estimates are too large; but \( a \) cannot be too large because the limit in eq. 2 is poorly approximated.

The nearest neighbor approach has a nice power law behavior in higher-dimensional cases where the correlation method breaks down, but it too has difficulties. In particular, the dimension is the negative reciprocal of the slope of the \( (\log_k \varepsilon, \log C(\varepsilon)) \) curve. As the dimension \( d \) increases, small absolute errors in determining the slope cause large errors in the estimate of \( d \). Moreover, the nearest-neighbor approach is especially sensitive to noise and requires the calculation of high-order nearest neighbors to get accurate dimension estimates. The question of which order of nearest neighbors gives the best estimates for a given attractor is an open one; if the order is too low, the computed dimension can be in error by a factor of 2, but this will not be apparent from the \( (\log_k \varepsilon, \log C(\varepsilon)) \) curves, nor from a regression analysis as a function of embedding parameters. In both methods, a careful variation of the fitting parameters is essential to obtain good results.

The computer time required by the correlation dimension
algorithm varies considerably, depending on the attractor. Dimension computations with nearest-neighbors can be arranged so that the time required grows logarithmically with the length of the data file and as $n \log n$ with the maximum order $n$ of the nearest neighbors. Both algorithms are vectorizable; data files of 32 K points can be processed for a wide range of parameter values in a few minutes on a CRAY-XMP.

Acknowledgments

We thank R. Badii for suggesting this method and for helpful discussions and correspondence. This research is supported by the Department of Energy, Office of Basic Energy Sciences.

References