

PHYSICAL REVIEW A

STATISTICAL PHYSICS, PLASMAS, FLUIDS, AND RELATED INTERDISCIPLINARY TOPICS

THIRD SERIES, VOLUME 41, NUMBER 12

15 JUNE 1990

Comparing a nearest-neighbor estimator of local attractor dimensions for noisy data to the correlation dimension

Mary Eileen Farrell, A. Passamante, and T. Hediger

Naval Air Development Center, Warminster, Pennsylvania 18974

(Received 11 September 1989; revised manuscript received 15 February 1990)

One technique of estimating the local intrinsic dimensionality of attractors is to apply the nearest-neighbor (NN) approach to local regions of attractor data in the phase-space domain. In noise-free, infinite signal-to-noise ratio data, the NN method is shown to produce results in reasonable agreement with the correlation dimensions for some known examples. All NN results are compared to the correlation dimension, as computed by the standard Grassberger-Procaccia algorithm. However, it is shown here that even a small amount of corrupting noise will severely degrade the performance of NN-type dimension estimators. NN methods are therefore of limited use for these calculations if data are noisy.

I. INTRODUCTION

Recent advances in theoretical physics have generated a new theory of deterministic chaos in which many signals may now be interpreted as being deterministic (rather than stochastic) in origin. Perhaps the most important feature of these chaotic signals is that they commonly require a small number of parameters to model them. Therefore, these chaotic signals are said to be low-dimensional. This is in marked contrast to the stochastic signal, which is by definition high-dimensional. In this paper we discuss a measure of attractor dimensionality called the local intrinsic dimension (LID). The LID is an upper bound on the correlation dimension (d_2) as computed using the Grassberger-Procaccia algorithm¹ (GPA) and this in turn is a lower bound on both the information dimension (d_1) and the capacity or Hausdorff dimension (d_0).² The LID is a measure of the number of significant local orthogonal directions along which the data points are distributed. We describe a method of estimating the LID using a nearest-neighbor approach by Pettis and co-workers.³⁻⁵ This method is then applied to chaotic attractor data such as the Lorenz, Hénon, Duffing, and Rössler systems in the phase space. For a recent review of dimension estimation methods see Ref. 6.

II. THEORY

The zero-mean time-varying data samples (x_i), assumed to consist of a true deterministic signal and white

Gaussian noise, are first embedded in a high-dimension (r) space by forming vectors:

$$\begin{aligned} \mathbf{x}_1 &= (x_1, x_2, \dots, x_r), \\ \mathbf{x}_2 &= (x_2, x_3, \dots, x_{r+1}), \\ &\vdots \\ \mathbf{x}_n &= (x_n, x_{n+1}, \dots, x_{n+r-1}), \end{aligned} \quad (2.1)$$

where typically $r=10$ or 20 and $n=20\,000$. Using more data points (e.g., $n=60\,000$) does not significantly alter the local intrinsic dimension results.

Here an accurate phase-space portrait of the attractor is reconstructed from the time data using the method of delays,⁷ such that a given reconstructed phase-space point is equated with r successive points of the time series, each separated by the interpoint sampling time t_s . Choice of r should satisfy the embedding criterion set forth by Whitney,⁸ and rigorously justified by Takens,⁷ namely, $r \geq 2d + 1$, where d is the topological dimension of the manifold. Here we set up the embedding dimension on the basis of the true signal's anticipated characteristics, since that is our main focus. This has yielded reasonable results, at least down to the moderate values of signal-to-noise ratios (SNR's) under consideration. If the corresponding time data window, which is defined as $t_w = (r-1)t_s$, is too large, the components of the embedding vectors may be decorrelated, appearing as random

noise with high dimension. On the other hand, if t_s or the window is too small, then the points will all lie near the main diagonal in phase space. In choosing an optimum time data window, or equivalently, an optimum r , we follow the work of Albano *et al.*⁹ in which an appropriate t_w is taken to be 2 to 3 times the correlation time. We determine the correlation time from the first zero of the autocorrelation function R , where R is defined as the expectation of the product of any two data samples separated in time by m samples:

$$R(m) = E[x(t)x(t+mt_s)]. \quad (2.2)$$

Next, the algorithm randomly selects arbitrary points on the attractor to serve as local centers around which a fixed number, q , of nearest neighbors will be processed (after subtracting out the center coordinate of each local region). The local regions, each of which has its own characteristic spatial distribution of data points extending into various orthogonal dimensions,¹⁰⁻¹⁴ are assumed to cover the entire attractor. The number of significant orthogonal dimensions in each local region is the local intrinsic dimension.

In this paper we apply the k -nearest-neighbor (NN) technique to the problem of estimating the LID. This approach, which is based on nonparametric estimation of the density functions, relies solely on the local properties, and not on the global properties, of the distribution of the data. An initial estimate of the probability density function $p(\mathbf{X})$ may be formulated as follows:

$$p(\mathbf{X})V = \mathbf{k}(\mathbf{X})/N, \quad (2.3)$$

where N is a large number of samples, \mathbf{k} is the number of samples falling in a small local region $L(\mathbf{X})$ around \mathbf{X} , and V is the corresponding volume of $L(\mathbf{X})$. This is known as the Parzen density estimate, in which case V is fixed and \mathbf{k} is allowed to be a random variable that is dependent on \mathbf{X} .⁵ An alternate approach is the NN technique discussed in this paper in which k is fixed and \mathbf{V} is allowed to be a random variable, $\mathbf{V}(\mathbf{X})$. The local region $L(\mathbf{X})$ may then be extended around \mathbf{X} until the k th nearest neighbor is located. We may interpret the NN method as being equivalent to the Parzen density estimate with a uniform kernel function whose size is adjusted automatically, depending on the location. In this way, if k is fixed, \mathbf{V} becomes larger in regions of low density and smaller in regions of high density. Assuming a uniform kernel function, such that

$$\kappa(Y) = \begin{cases} 1/V & \text{inside } L(\mathbf{X}) \\ 0 & \text{outside } L(\mathbf{X}), \end{cases} \quad (2.4)$$

the distance to the k th nearest neighbor may be expressed in terms of the corresponding volume as

$$\mathbf{d}_{k_{\text{NN}}}(\mathbf{X}) = \frac{\Gamma^{1/n}(n/2+1)}{\pi^{1/2}|\mathbf{A}|^{1/2n}} \mathbf{V}^{1/n}(\mathbf{X}), \quad (2.5)$$

where Γ is the gamma function, n is the dimension of the local region, and, with a defined as the radius of the local region, $a^2 \mathbf{A}$ gives the covariance matrix, Σ , of the kernel density.⁵

Next the first-order approximation $u(\mathbf{X}) = p(\mathbf{X})V(\mathbf{X})$

is used, where $u(\mathbf{X})$ is the coverage of $L(\mathbf{X})$ whose boundary is determined by the k th nearest neighbor, and the density function for u is taken to be

$$p_u(u) = \frac{N!}{(k-1)!(N-k)!} u^{k-1}(1-u)^{N-k}, \quad 0 \leq u \leq 1 \quad (2.6)$$

which is a beta distribution $B(k, N-k+1)$; the beta distribution can describe a large class of local data distributions. Then the m th-order moments of $\mathbf{d}_{k_{\text{NN}}}(\mathbf{X})$ can be determined by¹⁵

$$E[\mathbf{d}_{k_{\text{NN}}}^m(\mathbf{X})] \cong \int_0^1 \mathbf{d}_{k_{\text{NN}}}^m(\mathbf{X}) p_u(u) du \cong \nu p^{-m/n}(\mathbf{X}), \quad (2.7a)$$

where

$$\nu = \frac{\Gamma^{m/n}((n+2)/2)}{\pi^{m/2}|\Sigma|^{m/2n}} \frac{\Gamma(k+m/n)}{\Gamma(k)} \frac{\Gamma(N+1)}{\Gamma(N+1+m/n)}. \quad (2.7b)$$

The overall average of the distance in Eq. (2.7a) in the entire space is then

$$E_x E[\mathbf{d}_{k_{\text{NN}}}^m(\mathbf{X})] \cong \nu E_x [p^{-m/n}(\mathbf{X})]. \quad (2.8)$$

For the case $m=1$ and k large in Eq. (2.7b), we may use the Stirling approximation of $\Gamma(x)$ for large positive x , that is,

$$\Gamma(x) \cong x^x e^{-x} (2\pi/x)^{1/2} \left[1 + \frac{1}{12x} + \frac{1}{288x^2} + \dots \right], \quad (2.9)$$

to show that $\Gamma(k+1/n)/\Gamma(k) \simeq k^{1/n}$ for the typical values of n considered in this paper (i.e., $n=2-8$). Then from Eqs. (2.7) and (2.8) the ratio of two NN distances is found to be

$$\frac{E_x E[\mathbf{d}_{k_{\text{NN}+1}(\mathbf{X})}]}{E_x E[\mathbf{d}_{k_{\text{NN}}}(\mathbf{X})]} \cong 1 + \frac{1}{k_{\text{NN}} n} \quad (2.10)$$

to first order in $1/k_{\text{NN}}$. (The second-order term dropped here is approximately two orders of magnitude smaller than the first-order term retained.)

From computations of the left-hand side of Eq. (2.10) one can solve for n . Since the spatial averaging is performed over many small local regions, conservatively 500, this dimension n must be the average local intrinsic dimension of the data, which we denote as ALID.

III. NUMERICAL RESULTS

To demonstrate the method presented here we have applied it to several standard cases of interest, in particular, to the Lorenz system, the Hénon map, the Duffing and Rössler systems, and to cases of multiple sinusoids with additive noise over a range of SNR's.

Lorenz data were generated from the Lorenz system of equations:

$$\dot{x} = a(y - x), \quad (3.1a)$$

$$\dot{y} = cx - y - xz, \quad (3.1b)$$

$$\dot{z} = -bz + xy, \quad (3.1c)$$

where $a=10$, $b=\frac{8}{3}$, and $c=28$. In generating the Lorenz data we used initial conditions $x(0)=y(0)=z(0)=1$ and a Runge-Kutta integration step size of 0.003. Using a decimation factor of 10 the data used for this analysis were taken every 0.03 time unit. Hénon data were generated via the map:

$$x_{i+1} = bx_{i-1} + 1 - ax_i^2, \quad (3.2)$$

where $a=1.4$ and $b=0.3$. Duffing data were generated from a special case of the Duffing equation:¹⁶

$$\ddot{x} + b\dot{x} + x^3 = B \cos(\omega t). \quad (3.3)$$

This equation can be used to describe an electrical circuit with a nonlinear inductance and linear resistance, driven by a sinusoidal voltage. We used the parameters $b=0.05$, $B=7.5$, and $\omega=1.0$. Rössler data, which are best known for their use in describing the dynamics of chemical reactions in a stirred tank, were generated by

the following system of equations:

$$\dot{x} = -(y + z), \quad (3.4a)$$

$$\dot{y} = x + ay, \quad (3.4b)$$

$$\dot{z} = b + z(x - c), \quad (3.4c)$$

where the parameters used are $a=0.15$, $b=0.20$, and $c=10.0$. In order to obtain high resolution over the full spectrum, a decimation factor of 10 was also used for the Duffing and Rössler data. Data for the two-torus were generated using sine waves of frequencies $2^{1/2}$ Hz and $7^{1/2}$ Hz and a sampling time equal to 0.05 sec, which more than satisfies the Nyquist condition that the sampling rate should be at least twice the highest frequency.

Table I contains results for the ALID for different values of k_{NN} for the data sets studied. We find that the NN approach produces results close to the actual dimension in the absence of noise (i.e., infinite SNR). Empirical studies⁴ indicate that $k_{NN} \approx (q)^{1/2}$ is a reasonable choice. In Fig. 1 ALID results for the Duffing attractor for a typical value of $k_{NN}=6$ are plotted against the SNR. For the data in Table I the average radius $\langle a \rangle$ of the local regions is approximately 0.1 or 0.2 of the full attractor radius A . Typical values of $\langle a \rangle / A$ for different SNR's are plotted in Fig. 2. In general, sufficiently small local regions may be obtained if we restrict the number of samples q to a maximum of three, four, or five times the

TABLE I. ALID results determined via Eq. (2.10) for the Lorenz, Hénon, two-torus, Duffing, and Rössler attractors. Averaging of NN distances here was performed over 500 randomly positioned local regions, with each local region containing $q=40$ total points. $\langle a \rangle / A$ gives the ratio of the mean local hypersphere radius to the full attractor radius. Δ represents the average standard deviation of the ALID results. The actual correlation dimension values d_2 , determined by applying the GPA algorithm for large a and infinite SNR are listed in the last column.

| | SNR | $\frac{\langle a \rangle}{A}$ | $k_{NN}=4$ | $k_{NN}=8$ | $k_{NN}=12$ | $k_{NN}=16$ | Δ^a | d_2 |
|-----------|----------|-------------------------------|------------|------------|-------------|-------------|------------|------------------|
| Lorenz | 10 | 0.2 | 7.4 | 8.3 | 7.2 | 7.6 | 0.2 | |
| | 20 | 0.1 | 7.3 | 7.5 | 7.1 | 6.7 | 0.3 | |
| | 30 | 0.03 | 6.3 | 5.1 | 4.4 | 4.4 | 0.2 | |
| | ∞ | 0.02 | 2.1 | 2.1 | 2.3 | 2.3 | | 2.0 |
| Hénon | 10 | 0.2 | 6.9 | 7.3 | 6.9 | 7.4 | 0.1 | |
| | 20 | 0.1 | 6.3 | 5.8 | 5.7 | 4.3 | 0.5 | |
| | 30 | 0.1 | 3.7 | 3.0 | 2.1 | 2.1 | 0.1 | |
| | ∞ | 0.1 | 1.3 | 1.4 | 1.3 | 1.3 | | 1.2 |
| Two-torus | 10 | 0.2 | 7.6 | 7.3 | 7.3 | 7.1 | 0.4 | |
| | 20 | 0.05 | 7.5 | 7.4 | 6.7 | 6.3 | 0.2 | |
| | 30 | 0.03 | 2.5 | 2.3 | 2.7 | 2.9 | 0.1 | |
| | ∞ | 0.08 | 1.9 | 1.3 | 2.2 | 2.4 | | 2.0 ^b |
| Duffing | 10 | 0.2 | 8.4 | 8.5 | 8.1 | 8.8 | 0.3 | |
| | 20 | 0.1 | 6.5 | 6.3 | 5.4 | 6.1 | 0.3 | |
| | 30 | 0.1 | 3.9 | 3.7 | 3.4 | 3.4 | 0.1 | |
| | ∞ | 0.1 | 2.7 | 2.4 | 2.4 | 2.5 | | 2.6 |
| Rössler | 10 | 0.1 | 9.9 | 9.6 | 9.5 | 9.1 | 0.3 | |
| | 20 | 0.05 | 8.8 | 9.0 | 8.8 | 8.6 | 0.2 | |
| | 30 | 0.02 | 6.8 | 6.4 | 6.3 | 6.1 | 0.4 | |
| | ∞ | 0.01 | 2.1 | 1.9 | 2.0 | 2.0 | | 2.0 |

^aAverage standard deviation for ALID determined using nearly 40 values of k_{NN} .

^bActual dimension determined here by the number of sinusoids used to generate the data.

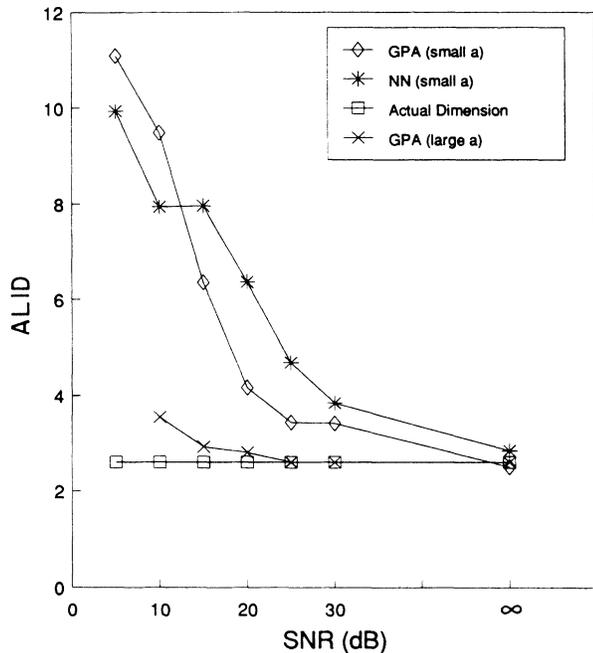


FIG. 1. Plot of ALID for the Duffing attractor determined from spatial averaging of nearest-neighbor (NN) distances in accordance with Eq. (2.10) for a typical value of $k_{NN}=6$. The number of points within each local region here is $q=40$. Also shown for comparison are the correlation dimension values computed using the Grassberger-Procaccia algorithm (GPA). Small- a GPA results are determined from Eq. (3.5), but large- a GPA results constitute only rough estimates from slope plateaus having no clear convergence for finite SNR. The horizontal line represents the large- a GPA result, or correlation dimension d_2 , for infinite SNR, which we refer to as the actual dimension.

embedding dimension. We find that fixing the number of samples in each local region offers more control than fixing the radius because the latter approach may lead to an insufficient number of data points in a given local region.

One may inappropriately attempt to apply the Grassberger-Procaccia algorithm using the smaller (LID) radius region, producing high values of correlation dimension, indicating the dominance of noise. We include these smaller radius GPA results as a reference demonstrating this effect. It should be realized, however, that GPA may be applied at any radius and, in fact, for the noisy case, the larger values of the radius are more useful. However, we caution that, for some attractors, even at larger radius values, for SNR's below approximately 15–20 dB the correlation dimension d_2 , as produced by the GPA can be quite variable. For larger radii the actual SNR limit for the GPA will depend on the attractor under consideration. As a reference, results are given in Fig. 1 for the correlation dimension given by the Grassberger-Procaccia algorithm, which is defined by the slope of the $\log_{10}c$ versus $\log_{10}a$ plot, where c is the correlation summation,¹ and which is calculated by

$$d_2^{(ij)} = \frac{\log_{10}q_i - \log_{10}q_j}{\log_{10}a_i - \log_{10}a_j}, \quad (3.5)$$

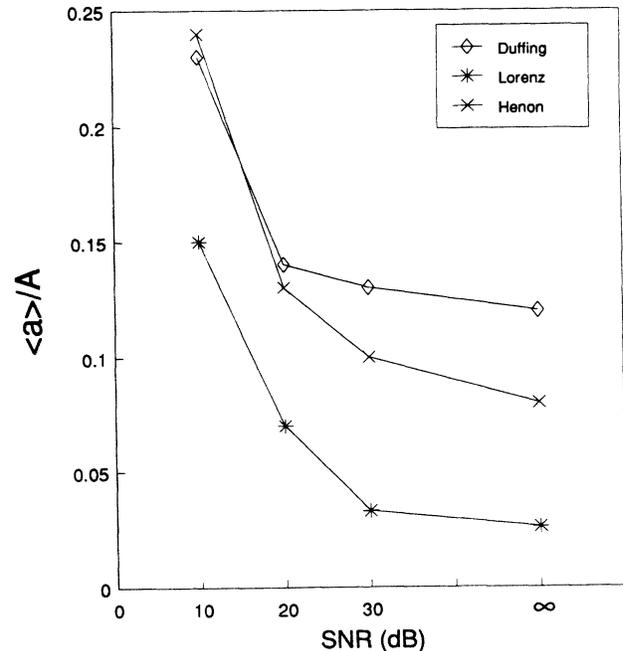


FIG. 2. Selected typical values of $\langle a \rangle / A$ for small length scales where $\langle a \rangle$ is the mean local hypersphere radius and A is the full attractor radius. Usually, larger values of $\langle a \rangle / A$ correspond to smaller SNR's because of point spreading in the phase space due to the presence of noise.

where q is the number of data points within each local region and a_i is the corresponding mean radius of the local regions containing q_i data points. Note that in this analysis the length scales are restricted to those characteristic of the local regions. Results for both small and large radius are shown in Fig. 1. However, even at 20 dB the results for d_2 obtained using the GPA algorithm over the whole range of attractor radius are better but still marginally useful.

IV. CONCLUSIONS

We have shown that the nearest-neighbor method can produce reasonable estimates of the dimension of low-dimensional attractors only in the case of little or no additive noise. In previous work Passamante and co-workers^{13,14,17} applied singular-value decomposition and information-theoretic criteria to determine the ALID. Those results were found to be close to d_2 for SNR's in the range of 5–12 dB. The NN method, on the other hand, breaks down at SNR's even as high as 30 dB.

ACKNOWLEDGMENTS

The authors would like to thank the Naval Air Development Center internal Independent Research office and the Office of Naval Research, Code 12, J. G. Smith, for the support and opportunity to do this work.

- ¹P. Grassberger and I. Procaccia, *Physica* **9D**, 189 (1983).
- ²F. Hausdorff, *Math. Ann.* **79**, 157 (1919).
- ³K. W. Pettis, T. A. Bailey, A. F. Jain, and R. C. Dubes, *Trans. IEEE Pattern Anal. Machine Intell.* **1**, 25 (1979).
- ⁴R. L. Somorjai, in *Dimensions and Entropies in Chaotic Systems*, edited by G. Mayer-Kress (Springer-Verlag, New York, 1986).
- ⁵K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. (Academic, New York, 1990).
- ⁶J. Theiler, *J. Opt. Soc. Am. A* (to be published).
- ⁷F. Takens, *Dynamical Systems and Turbulence*, Vol. 898 of *Lecture Notes in Mathematics*, edited by D. A. Rand and L. S. Young (Springer-Verlag, Berlin, 1981).
- ⁸H. Whitney, *Ann. Maths.* **37**, 645 (1936).
- ⁹A. M. Albano, J. Muench, C. Schwartz, A. J. Mees, and P. E. Rapp, *Phys. Rev. A* **38**, 3017 (1988).
- ¹⁰D. S. Broomhead, R. Jones, and Gregory P. King, *J. Phys. A* **20**, L563 (1987).
- ¹¹E. R. Pike, in *Computational Systems: Natural and Artificial*, edited by H. Haken (Springer-Verlag, Berlin, 1986).
- ¹²K. Fukunaga and D. R. Olsen, *IEEE Trans. Comput.* **20**, 176 (1971).
- ¹³A. Passamante, T. Hediger, and M. Gollub, *Phys. Rev. A* **39**, 3640 (1989).
- ¹⁴A. Passamante, T. Hediger, and Mary Eileen Farrell, in *Measures of Complexity and Chaos*, edited by N. B. Abraham, A. M. Albano, A. Passamante, and P. E. Rapp (Plenum, New York, 1990).
- ¹⁵K. Fukunaga and T. E. Flick, *IEEE Trans. Pattern Anal. Machine Intell.* **6**, 779 (1984).
- ¹⁶Y. Ueda, *J. Stat. Phys.* **20**, 181 (1979).
- ¹⁷T. Hediger, A. Passamante, and Mary Eileen Farrell, *Phys. Rev. A* **41**, 5325 (1990).