

1992

Mutual Information, Strange Attractors, and the Optimal Estimation of Dimension

J. M. Martinerie

Alfonso M. Albano

Bryn Mawr College, aalbano@brynmawr.edu

A. I. Mees

P. E. Rapp

[Let us know how access to this document benefits you.](#)

Follow this and additional works at: http://repository.brynmawr.edu/physics_pubs



Part of the [Physics Commons](#)

Custom Citation

J.M. Martinerie, A.M. Albano, A.I. Mees and P.E. Rapp, *Phys. Rev. A* **45**, 7058 (1992).

This paper is posted at Scholarship, Research, and Creative Work at Bryn Mawr College. http://repository.brynmawr.edu/physics_pubs/38

For more information, please contact repository@brynmawr.edu.

Mutual information, strange attractors, and the optimal estimation of dimension

J. M. Martinerie

Laboratoire d'Electrophysiologie et de Neurophysiologie Appliquée, Centre National de la Recherche Scientifique, Hôpital de la Salpêtrière, 47 boulevard de l'Hôpital, 75651 Paris CEDEX 13, France

A. M. Albano

Department of Physics, Bryn Mawr College, Bryn Mawr, Pennsylvania 19010

A. I. Mees

Department of Mathematics, University of Western Australia, Nedlands, Western Australia 6009, Australia

P. E. Rapp

Department of Physiology and Biochemistry, The Medical College of Pennsylvania, 3300 Henry Avenue, Philadelphia, Pennsylvania 19129

(Received 12 August 1991)

It has been shown that the appropriate setting of data windows is crucial to a successful estimation of a time-series correlation dimension using the Grassberger-Procaccia algorithm [Physica **9D**, 189 (1983); Phys. Rev. Lett. **50**, 346 (1983)], and it has been proposed that the first minimum of the corresponding mutual-information function may be an appropriate window value. We have tested this hypothesis against data generated by the Rössler equations, the Lorenz equations, and a three-dimensional irrational torus. We conclude that mutual information is not consistently successful in identifying the optimal window.

PACS number(s): 05.45.+b, 02.60.+y

I. INTRODUCTION: OPTIMAL SETTING OF WINDOWS AND DIMENSION ESTIMATES

This paper considers the question of optimal setting of windows of a time series when estimating dimension. Therefore, we must address three issues. (i) Why is the dimension of a time series potentially important? (ii) Why is the setting of windows a crucial issue? (iii) How can an optimal window be identified? Estimating a signal's dimension is often an important step in its dynamical characterization. In ideal cases when large sets of high-quality data are available, the dimension estimate can be used to infer something about the dynamical structure of the system generating the signal. However, even in circumstances where this is impossible, dimension estimates can be an important empirical characterization of a signal's complexity. In many instances where data limitations preclude reliable calculations of an absolute value of dimension, the variation of the estimate in response to changing conditions can be a valuable quantitative measure of system behavior. A variety of procedures for estimating dimension is available. The most popular is to estimate the correlation dimension using the Grassberger-Procaccia algorithm [1,2]. In this procedure measurements v_1, v_2, \dots, v_M are obtained from a single dynamical variable. The time interval between each measurement, the sampling interval, will be denoted by T_s . (Generalizations to multichannel recordings are straightforward.) The measured values are used to create points in an N -dimensional space,

$$X_1 = (v_1, v_2, \dots, v_N),$$

$$X_2 = (v_2, v_3, \dots, v_{N+1}),$$

⋮

$$X_K = (v_K, v_{K+1}, \dots, v_M),$$

where $K = M - N + 1$; N is the embedding dimension, and $\{X\} = \{X_1, \dots, X_K\}$ is the embedding set. (There are other equally valid embedding protocols.) Takens [3] and Mañé [4] have shown that if certain conditions are met, the dimension of set $\{X\}$ is the same as the dimension of the attractor. It should be noted that the embedding dimension N is an arbitrary positive integer. While the examples used in subsequent calculations are low dimensional, the procedure, in principle, generalizes to higher-dimensional attractors. There are, however, severe practical difficulties associated with extensions to higher-dimensional objects [5,6].

The Grassberger-Procaccia algorithm estimates the dimension by first calculating the correlation integral

$$C_N(r) = (1/N_p) \sum_{i=1}^K \sum_{j=i+1}^K \Theta(r - |X_i - X_j|)$$

where Θ is the Heaviside function, $||$ is the Euclidean distance operator, and N_p is the number of distinct pairs of points X_i, X_j in the N -dimensional space. Grassberger and Procaccia [1,2] demonstrated that if the embedding dimension and the number of data points are large enough and if the data are sufficiently noise free, then the

function $\ln C_N(r)$ vs $\ln(r)$ has a linear region, called the scaling region. The slope of the function in that linear region is D_2 , the correlation dimension. Therefore, graphs of $d\ln C_N(r)/d\ln(r)$ vs $\ln(r)$ should have a horizontal region called the plateau. The value of the derivative in the plateau region is the dimension.

Unless care is exercised, spurious estimates of dimension can result [7–10]. It is necessary to impose a series of convergence conditions on these calculations.

(i) The plateau must be flat. The degree of variation of the derivative in the scaling region must not exceed some specified standard. Typically a limit of 15% of D_2 is set.

(ii) The scaling region must meet a minimum-length requirement. For example, we can require that $\ln(r)$ varies over a factor of at least 1.5. This corresponds to a factor of 5 in r .

(iii) The estimated value of D_2 must be stable as the embedding dimension is increased. We usually impose a limit of 15% deviation over four consecutive embeddings.

If these criteria are imposed, it is frequently impossible to produce a dimension estimate. This is especially true of small sets of noisy experimental data.

When analyzing experimental data, methodological considerations become crucial to the success of dimension estimation. Specifically, the choice of sample interval T_s and embedding dimension N can dramatically effect the plateau. This is shown in Fig. 1, which uses data generated by a three-torus.

$$v(t) = \sin\left[\frac{3t}{500}\right] + \sin\left[\frac{3\sqrt{2}t}{250}\right] + \sin\left[\frac{9\sqrt{3}t}{500}\right],$$

where the time interval between each sample, T_s , is 1.

Figure 1 shows the difference in plateau that can result when N and T_s are varied. [In this and later figures $d\ln C_N(r)/d\ln(r)$ is plotted against $\ln(r/r_{\max})$, where r_{\max} is the largest interpoint distance in the N -dimensional embedding set $\{X\}$; r_{\max} will vary with embedding dimension. As will be seen presently, this normalization facilitates comparison of computations with different embeddings.] Figure 1 shows that the choice of T_s and N can have a dramatic effect on the plateau. An inappropriate choice can result in the unnecessary failure to resolve the dimension of the data set.

The results summarized in Fig. 1 suggest that the selection of N and T_s are crucial in estimating dimension. To cast more light on this, we begin with the definition of window as the time interval spanned by a point in $\{X\}$: if

$$X_j = (v_j, v_{j+1}, \dots, v_{j+N-1})$$

and the time interval between the measurement of each value of v is T_s . The window is the time interval spanned by X_j , namely, $T_s(N-1)$. Results in Fig. 2 indicate that the window, rather than T_s and N separately, determine the characteristics of the plateau. In that diagram, computed with the same three-torus data used in Fig. 1, the values of N and T_s are different for each curve. However, the windows are approximately equal. When plotted against the normalized variable $\ln(r/r_{\max})$, the deriva-

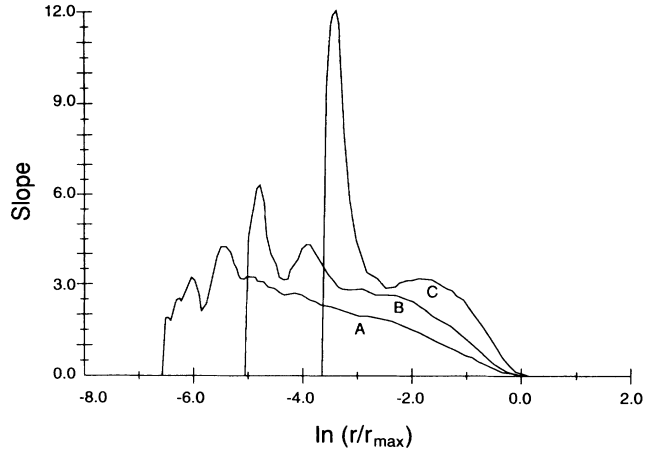


FIG. 1. Plot of $d\ln C_N(r)/d\ln(r)$ vs $\ln(r/r_{\max})$ for a three-torus. 1000 vectors were used in each calculation. Curve A: $N=9$, $T_s=5$, $W=40$; curve B: $N=11$, $T_s=11$, $W=100$; curve C: $N=10$, $T_s=20$, $W=180$.

tives superimpose. This has been investigated for other systems [11]. The same result is obtained; the superposition shown in Fig. 2 is not limited to the three-torus.

These results provide the answer to the second question stated in the opening paragraph: why is the setting of windows a crucial issue? The window is important because, first, as indicated by Fig. 2, the product $T_s(N-1)$ determines the characteristics of the correlation integral. Second, as indicated by Fig. 1, correlation integrals are extremely sensitive to window values. For some values of window, a measurable plateau is completely absent even when high-quality data are examined. Thus, an inappropriate value of window can result in an unnecessary failure to resolve a time-series' dimension. We operationally define an optimal window as one that gives the largest plateau. This leads us to the third question: how can an optimal window be identified? The rest of this paper

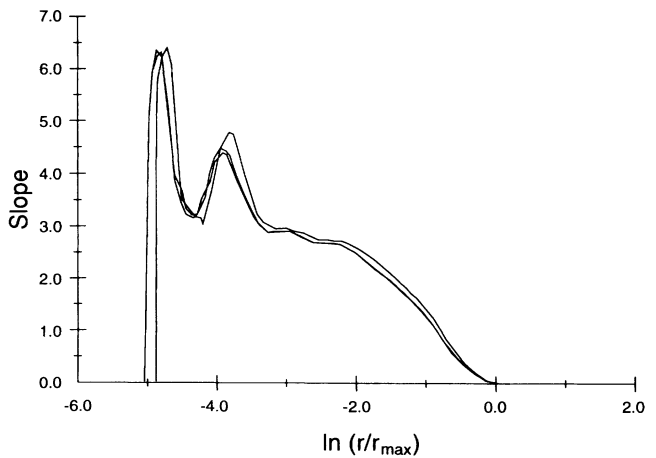


FIG. 2. Plot of $d\ln C_N(r)/d\ln(r)$ vs $\ln(r/r_{\max})$ for a three-torus. 1000 vectors were used in each calculation. The three curves correspond to $(N, T_s, \text{window}) = (6, 20, 100)$, $(8, 14, 98)$, $(11, 10, 100)$.

addresses this question.

Albano *et al.* [11] performed calculations to determine if a signal's autocorrelation function could be used to predict the optimal window. Three dynamical systems (the Lorenz attractor, the Rössler attractor, and the three-torus) were examined. "Optimal" windows for each system were determined empirically by calculating $d \ln C_N(r) / d \ln(r)$ for different windows. The optimal windows were compared with the signal's autocorrelation time (the time required for the autocorrelation function to drop to $1/e$ of its original value), the time at the first minimum of the autocorrelation function, and the time at the first zero of the autocorrelation function. It was found that none of these measures provided a reliable prediction of the optimal window.

The Wiener-Khinchin theorem establishes that the Fourier transform of the autocorrelation function is equal to the power spectrum. The power spectrum in turn, under the z transform, can be expressed as a Laurent series. Thus the autocorrelation function can be examined in terms of the poles of its corresponding Laurent series [12–14]. An alternative approach to the problem of window estimation would be an examination of the success of these poles as estimators. Two arguments have encouraged the investigation of the autocorrelation function in the time domain presented in Albano *et al.* [11]. First, the time-domain characterization of the autocorrelation function is more commonly used in signal analysis. This makes it possible to relate the results obtained here with a substantial body of literature and with our own investigations of the autocorrelation times of biological signals. Second, the window of an embedded data set as defined here, $(N-1)T_s$, is itself expressed in units of time. Thus an examination in the time domain seems more natural for this problem. However, this alternative approach raises an attractive possibility for further investigations. Another possible approach to the problem of window estimation is the examination of higher-order correlation functions. This possibility has been explored by Albano *et al.* [15]. Their preliminary results suggest that this method is a successful window estimator, at least in cases using high-quality, numerically generated data obtained from low-dimensional strange attractors.

Fraser and Swinney [16(a)] suggested that mutual information may identify an optimal window. This possibility is explored in the following sections of this paper.

II. MUTUAL INFORMATION

The development and notation given here follows that in Fraser and Swinney [16(a)]. For simplicity, these definitions are stated in terms of discrete systems. They can be generalized to continuous variables.

Consider systems S and Q consisting of discrete sets of possible messages $\{s_1, s_2, \dots, s_n\}$ and $\{q_1, q_2, \dots, q_m\}$ with associated probabilities $\{P_s(s_1), \dots, P_s(s_n)\}$ and $\{P_q(q_1), \dots, P_q(q_m)\}$. The entropy $H(S)$ is the average amount of information gained from a measurement of S

$$H(S) = - \sum_{i=1}^n P_s(s_i) \log_2 P_s(s_i).$$

$H(Q)$ is defined analogously.

The mutual information of the S and Q systems is denoted by $I(Q, S)$. Given a measurement of s , $I(Q, S)$ is the number of bits of q , on average, that can be predicted,

$$I(Q, S) = H(Q) + H(S) - H(S, Q),$$

where $H(Q)$ and $H(S)$ are the entropies of system Q and system S , respectively. $H(S, Q)$ is the joint entropy function. $H(S, Q)$ is the average amount of information gained from measuring (s, q) pairs, where the joint-probability distribution, denoted $P_{sq}(s_i, q_j)$, is the probability that $s = s_i$ and $q = q_j$,

$$H(S, Q) = - \sum_i^n \sum_j^m P_{sq}(s_i, q_j) \log_2 P_{sq}(s_i, q_j).$$

Because $H(S, Q) = H(Q, S)$, we see that $I(Q, S) = I(S, Q)$.

Suppose a variable v is being investigated by being digitized with a sample interval T_s . To put this process into the context of system S and system Q , let s be the measurement of v at time t , and let q be the measurement at time $t + T_s$. Using these measurements to define systems S and Q , mutual information $I(S, Q)$ can be calculated. Thus, mutual information becomes a function of T_s . For this problem, mutual information will be the number of bits of $v(t + T_s)$ that can be predicted, on average, when $v(t)$ is known. One wants to pick T_s so that the maximum amount of new information is obtained from each measurement. Therefore, T_s should be chosen so that $v(t + T_s)$ is as unpredictable as possible. Maximum unpredictability occurs at a minimum of predictability; that is, at a minimum in the mutual information. Because of the exponential divergence of chaotic trajectories, the first minimum of $I(Q, S)$, rather than some subsequent minimum, should probably be chosen for the sampling interval.

An extension of this argument is applicable to the question of window selection. In the case of window selection, one is selecting the temporal length of trajectory fragment to be used in a dimension calculation. Each fragment should provide the greatest amount of new information to the algorithm. Therefore, according to this argument, the window length, $(N-1)T_s$ (not the sampling rate T_s alone), should be related to the first minimum of the mutual information.

It can be shown that

$$I(Q, S) = \sum_i^n \sum_j^m P_{sq}(s_i, q_j) \log_2 \left[\frac{P_{sq}(s_i, q_j)}{P_s(s_i)P_q(q_j)} \right].$$

It is seen that the central issue in estimating mutual information is estimating probability densities, particularly P_{sq} . Estimation of probability densities from data is an active research area [17].

It is necessary to estimate the joint-probability density P_{sq} on the S - Q plane. The S - Q plane is partitioned into elements, and P_{sq} on any given element is estimated by dividing the number of points in that element by the total number of points. We have implemented the Fraser-Swinney algorithm [16(a)]. In this algorithm the partition is nonuniform and adaptive. A finer partition is con-

structed in regions of the plane where P_{sq} has a detailed structure. A sequence of partitions $G_0, G_1, G_2, \dots, G_m$ is constructed. Each partition is a grid of 4^m elements generated by dividing each axis into 2^m equiprobable segments. Let $R_m(K_m)$ denote an element of partition G_m ($K_m=0, 1, \dots, 2^m-1$). Let $N_m(R_m(K_m))$ be the number of points in this element and let N_0 denote the total number of elements. Then

$$P_{sq}(R_m(K_m)) = N(R_m(K_m)) / N_0 .$$

Fraser and Swinney demonstrate that i_m , the corresponding approximation of mutual information, is

$$i_m = 2m + \sum_{K_m} P_{sq}(R_m(K_m)) \log_2 P_{sq}(R_m(K_m)) .$$

Partitioning the S and Q axes into equiprobable elements is essential to obtaining this relationship for i_m . The partitioning continues element by element until P_{sq} is uniform on each element. P_{sq} is said to be uniform on element $R_m(K_m)$ if P_{sq} values on its subdivisions $R_{m+1}(K_m, j)$ are approximately equal as specified by some statistical criterion. This elegant approach to estimating P_{sq} results in the calculation of mutual information as a recursive function. Details are given in Fraser and Swinney [16(a)].

We have performed the calculations for the examples, given later, using both our program and a version of Fraser's program. The χ^2 criterion used to define "detailed structure" differ slightly between the two programs. Our criterion is the one described in Fraser and Swinney [16(a)], while Fraser's current program [16(b)] uses a simpler procedure in an attempt to speed calculations. Therefore, there is occasionally a difference in the subdivision depth. This results in different numerical values for mutual information, but does not significantly alter the positions of minima and maxima, or other details required for the conclusions we wish to draw. It is clear that some research is required into the sensitivity of the calculations to structure criteria, but this does not affect the work we are describing here. The graphs we show were all calculated using Fraser's program.

III. MUTUAL-INFORMATION CALCULATIONS WITH SMALL DATA SETS

Three systems were examined. The first, a three-torus, was specified in Sec. II. The second is the Lorenz attractor

$$\begin{aligned} \frac{dx}{dt} &= -10(x - y) , \\ \frac{dy}{dt} &= x(28 - z) - y , \\ \frac{dz}{dt} &= xy - (\frac{8}{3})z , \end{aligned}$$

where $T_s=0.01$. The third is the Rössler attractor

$$\frac{dx}{dt} = -y - z ,$$

$$\begin{aligned} \frac{dy}{dt} &= x + 0.2y , \\ \frac{dz}{dt} &= 4 + xz - 5.7z , \end{aligned}$$

where $T_s=0.05$.

It is seen in Fig. 3 that success in locating the extrema of mutual information varies with the system examined.

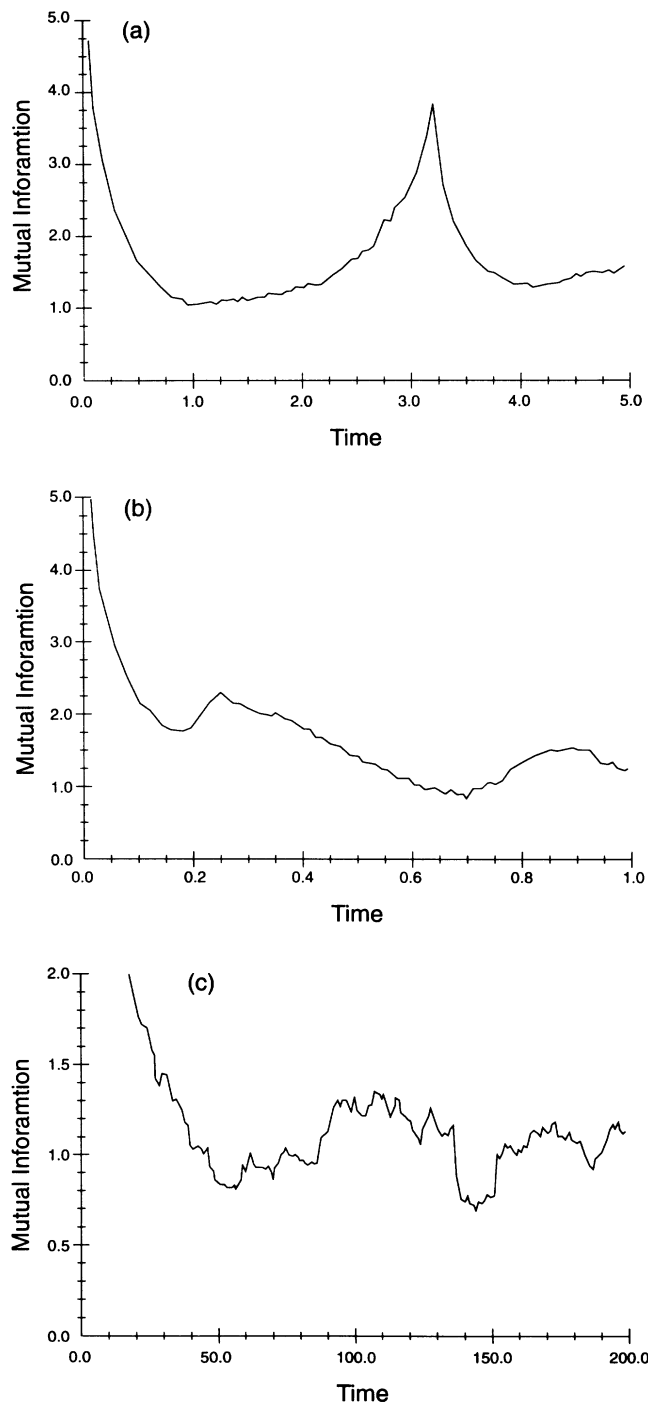


FIG. 3. Mutual-information function for (a) the Rössler attractor, (b) the Lorenz attractor, and (c) the three-torus. Calculations were performed with 8192 data points. Essentially identical functions result from 2000, 4000, and 8000 data points.

The Rössler data produce a mutual information function that is essentially identical for 2000, 4000, and 8000 data points. The location of the first maximum and minimum agrees with that obtained using 65 000 data points [16(a)].

While the mutual-information function obtained with Lorenz data is reasonably stable in response to changes in data set size, the minima and maxima are not as well defined as for the Rössler system. These difficulties are even more severe with the three-torus data. Irrational tori are unstable objects. The difficulties encountered in estimating extrema of the mutual information of a three-torus is consistent with the difficulty in numerically estimating its correlation dimension.

A helpful perspective can be gained by comparing the mutual-information functions for these three systems with those obtained using random numbers and a sine function corrupted with low-level noise. The pseudorandom numbers were generated using the random number utility on a VAX 11/750. The sine function is $x(t) = \sin[0.062(t-1)] + \epsilon_t$; $T_s = 1$. Here ϵ_t is equal to 0.0001 multiplied by a pseudorandom number in $[0,1]$, generated independently for each value of t . The reason for adding noise is that the mutual information is not defined for a pure sine function, and neither our nor

Fraser's program gives useful results in such a case. In the case of a corrupted sine wave, we see, Fig. 4(a), that there are sharp peaks at half-period intervals, with essentially zero values at other lag values. In the case of the random numbers, Fig. 4(b), the very low ordinate values should be noted. As anticipated, chaotic attractors are thus seen to be intermediate to pseudorandom and periodic time series.

Given the results of Fig. 4(b), it was speculated that mutual-information calculations might be helpful in assessing the quality of random-number generators. This, evidently, is not the case. Five programs from Press *et al.* [18] for generating uniform deviates were tested. The pseudorandom numbers used in the calculation shown in Fig. 5(a) were produced by an intentionally bad linear congruential generator. The data set of Fig. 5(b) was generated by combining the generator of Fig. 5(a) with a randomizing shuffle described in Knuth [19], to give output that is effectively free of sequential correlation. Though the second program is far more successful as a random number generator, comparison of Figs. 5(a) and 5(b) does not reveal a dramatic difference in their

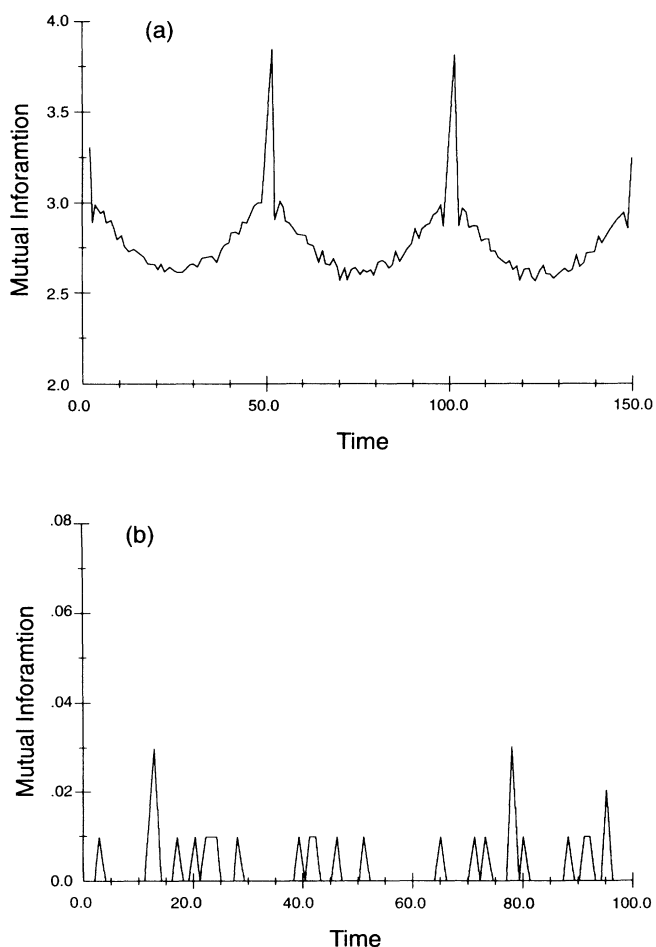


FIG. 4. Mutual-information functions for (a) the sine function plus noise and (b) numbers generated by a random-number generator; 8192 data points were used in each calculation.

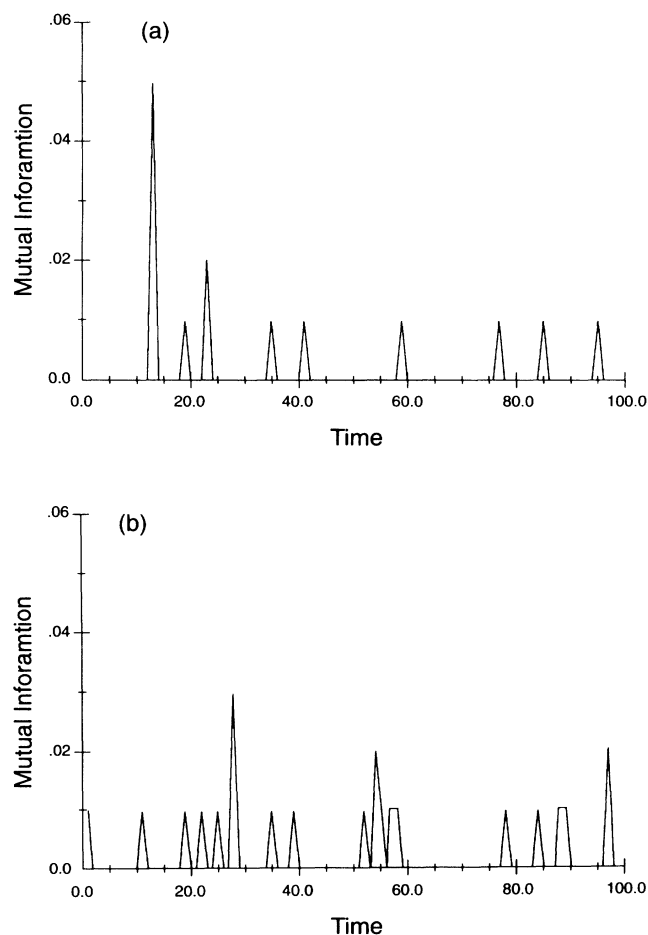


FIG. 5. Mutual-information functions for (a) the output of a linear congruential random number generator, and (b) the output of a random number generator that combines a linear congruential generator with a randomizing shuffle; 8192 data points were used in each calculation.

TABLE I. Autocorrelation and mutual-information functions, and the resulting window estimation.

| | Autocorrelation function | | | Mutual-information function | | Window |
|-------------|--------------------------|-----------|------------------|-----------------------------|------------------|--------|
| | T_{AC} | T_z | T_{min-AC} | T_{min-MI} | T_{max-MI} | T_w |
| Rössler | 1.15 | 1.60 | 3.00 | 1.00 | 3.20 | 3.30 |
| Lorenz | 0.30 | 1.22 | 0.72 | 0.17 | 0.25 | 1.00 |
| Three torus | 62.0 | 90.0 | 122.0 | 54.0 | 108.0 | 100.0 |
| | T_w/T_{AC} | T_w/T_z | T_w/T_{min-AC} | T_w/T_{min-MI} | T_w/T_{max-MI} | |
| Rössler | 2.87 | 2.06 | 1.10 | 3.30 | 1.03 | |
| Lorenz | 3.33 | 0.82 | 1.39 | 5.88 | 4.00 | |
| Three torus | 1.61 | 1.11 | 0.82 | 1.85 | 0.93 | |

mutual-information functions. The mutual-information function calculated with data from the other programs were qualitatively similar.

For the results reported in Table I, extrema of the mutual-information function were located by visual inspection of their graphs and by inspecting a numerical tabulation of all local extrema. Because mutual-information functions calculated with small data sets are noisy, it is difficult to construct a numerical procedure for

locating extrema that operates directly on the mutual-information function. One possible way to automate the process is to filter the mutual-information function and locate the extrema of the filtered function. A variety of filters was tested, including single-pass two-, five-, ten-, and twenty-point moving-average filters, and a double-pass ten-point moving-average filter. Examples are shown in Fig. 6. Signals in these examples were made symmetric to zero prior to filtering. Comparison calcula-

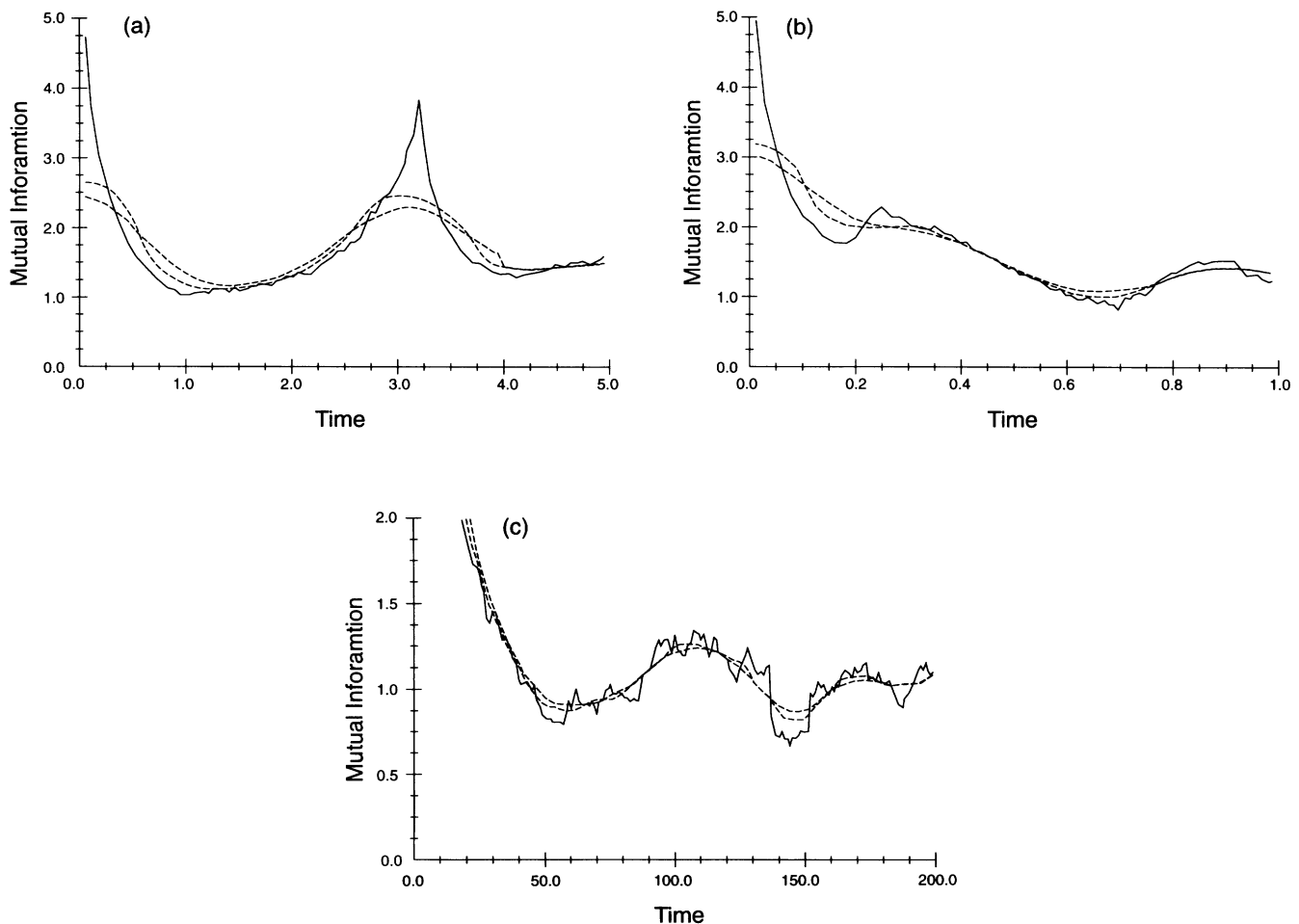


FIG. 6. Mutual-information functions filtered by a double-pass ten-point moving-average filter. Output from each pass is plotted as a dashed line. Input functions (solid line) were calculated using 8192 data points. (a) Rössler data, (b) Lorenz data, and (c) three-torus data.

tions indicate that there is no one single filter that is best for all mutual-information functions. Thus, to the list of problems associated with calculating mutual information from small data sets, a further problem must be added: what filter or other extremum estimator should be used?

IV. EXTREMA OF MUTUAL INFORMATION AND OPTIMAL WINDOWS

The results of using mutual-information and autocorrelation functions to predict an optimal window are summarized in Table I. T_w is the value obtained empirically by determining the window length that gave the largest scaling region. The details of the empirical process are given in Albano *et al.* [11]. T_{AC} is the autocorrelation time, which is the time required for the autocorrelation function to reach $1/e$ of its original value. T_z is the time of the autocorrelation function's first zero crossing. $T_{\min-AC}$ is the time of the autocorrelation function's first minimum. $T_{\min-MI}$ is the time of the first minimum of the mutual-information function. $T_{\max-MI}$ is the time of the first maximum of mutual information.

It is seen that neither mutual information nor autocorrelation is consistently successful in identifying the optimal window. Their use as adjuncts to estimating correlation dimension using the Grassberger-Procaccia algorithm is therefore uncertain. However, the importance of mutual information in estimating metric entropy

remains undisputed [20]. We remark in passing that subsequent calculations have suggested that mutual information can be used for pattern recognition of noisy time series.

ACKNOWLEDGMENTS

This work was supported in part by a contract from the Central Intelligence Agency (P.E.R.), Contract No. 22919-PH with the U.S. Army Research Office (A.M.A.), Grant No. NS19716 from the National Institutes of Health (P.E.R.), and grants from the Allegheny Singer Research Foundation (P.E.R.) and the Whitaker Foundation (P.E.R.). J.M.M. thanks INRIA, Paris, for its support. We would like to thank A. M. Fraser for a copy of his mutual-information program [16(b)] which enabled comparisons with our program and for several very helpful discussions. A.I.M. thanks Kings College, Cambridge University, and the Department of Mathematics, University of Arizona, for generous hospitality. P.E.R. would like to thank the Department of Mathematics, University of Western Australia, and Kingswood College for their support. The continuing support of the College Computer Center of the Medical College of Pennsylvania is gratefully acknowledged. We wish to express our thanks to UniSys Corporation for the generous donation of computer equipment to this project. The technical assistance of Joseph Waldron is gratefully acknowledged.

-
- [1] P. Grassberger and I. Procaccia, *Physica D* **9**, 189 (1983).
 - [2] P. Grassberger and I. Procaccia, *Phys. Rev. Lett.* **50**, 346 (1983).
 - [3] F. Takens, in *Dynamical Systems and Turbulence*, edited by D. A. Rand and L.-S. Young, *Lecture Notes in Mathematics* Vol. 898 (Springer, Berlin, 1981), p. 366.
 - [4] R. Mañé, in *Dynamical Systems and Turbulence*, edited by D. A. Rand and L.-S. Young, *Lecture Notes in Mathematics* Vol. 898 (Springer, Berlin, 1981), p. 230.
 - [5] L. A. Smith, *Phys. Lett. A* **133**, 283 (1988).
 - [6] D. Ruelle, *Proc. R. Soc. London Ser. A* **427**, 241 (1990).
 - [7] P. Grassberger, *Nature (London)* **323**, 609 (1986).
 - [8] J. Theiler, *Phys. Rev. A* **34**, 2427 (1986).
 - [9] A. M. Albano, A. I. Mees, G. C. deGuzman, and P. E. Rapp, in *Chaos in Biological Systems*, edited by H. Degn, A. V. Holden, and L. F. Olsen (Plenum, New York, 1987), p. 207.
 - [10] P. E. Rapp, A. M. Albano, and A. I. Mees, in *Dynamic Patterns in Complex Systems*, edited by J. S. A. Kelso, A. J. Mandell, and M. F. Schlesinger (World Scientific, Singapore, 1988), p. 191.
 - [11] A. M. Albano, J. Muench, C. Schwartz, A. I. Mees, and P. E. Rapp, *Phys. Rev. A* **38**, 3017 (1988).
 - [12] A. van den Bos, *IEEE Trans. Inform. Theory* **IT-17**, 493 (1971).
 - [13] N. Andersen, *Geophys.* **39**, 69 (1974).
 - [14] D. Ruelle, *Phys. Rev. Lett.* **56**, 406 (1986).
 - [15] A. M. Albano, A. Passamante, and M. E. Farrell, *Physica D* **54**, 85 (1991).
 - [16] (a) A. M. Fraser and H. L. Swinney, *Phys. Rev. A* **33**, 1134 (1986); (b) A. M. Fraser (unpublished).
 - [17] B. W. Silverman, *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, London, 1986).
 - [18] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes. The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1986).
 - [19] D. E. Knuth, *Seminumerical Algorithms*, Vol. 2 of *The Art of Computer Programming* (Addison-Wesley, Reading, MA, 1981).
 - [20] A. M. Fraser, in *Dimension and Entropies of Chaotic Systems*, edited by G. Mayer-Kress (Springer, Berlin, 1986), p. 82.