

## An upper bound for the proper delay time in chaotic time-series analysis

Ying-Cheng Lai<sup>1</sup>, David Lerner<sup>2</sup>, Robert Hayden

*Department of Mathematics, The University of Kansas, Lawrence, KS 66045, USA*

Received 22 February 1996; revised manuscript received 12 April 1996; accepted for publication 8 May 1996

Communicated by A.P. Fordy

### Abstract

We establish an upper bound for the proper delay time in chaotic time series analysis using delay-coordinate embeddings. The derivation is based on analyzing the effective scaling regime in the computation of the correlation dimension using the Grassberger–Procaccia algorithm. Numerical results agree with the theoretical prediction.

PACS: 05.45.+b

Analyzing chaotic time series has been a field of intensive research for the past fifteen years [1] since the pioneering work of Packard et al. [2] and Takens [3] who suggested the use of the delay-coordinate embedding technique to reconstruct a smooth copy of the underlying dynamics. Specifically, let  $z(t)$  be a trajectory of the  $d$ -dimensional dynamical system  $dz/dt = f(z)$ , with  $z(t_0) = z_0$ , and let  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  be a physical observable. If the measuring process samples  $g$  every  $T$  units of time beginning at  $t = t_0$ , the observations corresponding to the trajectory constitute a scalar time series  $\{x_n\}$  with  $x_n \equiv x(t_0 + nT) = g(z(t_0 + nT))$ . An image of the original trajectory is obtained using the map  $H : \mathbb{R}^d \rightarrow \mathbb{R}^m$  given by

$x_n = H(z_n) = (x_n, x_{n+\tau}, \dots, x_{n+(m-1)\tau})$ , for positive integers  $m$  and  $\tau$  [2,3], where  $m$  is the embedding dimension, and  $\tau$  is the delay time (the true time lag is  $\tau T$  for continuous dynamical systems). The proper choice of the two parameters  $m$  and  $\tau$  is critical for an accurate reconstruction of the underlying dynamics.

For a chaotic time series, the underlying dynamics is usually governed by an attractor. A theorem of Takens [3] shows that if  $m \geq 2D + 1$ , where  $D$  is the dimension of the manifold in which the attractor lies, then  $H$  is an embedding. More recently, Sauer, Yorke and Casdagli proved that [4] an embedding is also obtained if  $m > 2D_0$ , where  $D_0$  is the box-counting dimension of the attractor. If a one-to-one correspondence is not required, this constraint can be relaxed. For example, Ding et al. [8] have shown that the correct value of the correlation dimension  $D_2$  can be obtained with the the Grassberger–Procaccia (GP) algorithm [5–7] if  $m$  is any integer larger than  $D_2$  provided that the data set is both infinite and uncontaminated by noise. For finite or noisy time series,  $m$  must

<sup>1</sup> Also at Department of Physics and Astronomy, and Kansas Institute for Theoretical and Computational Science, The University of Kansas, Lawrence, KS 66045, USA. E-mail: lai@poincare.math.ukans.edu.

<sup>2</sup> Also at Kansas Institute for Theoretical and Computational Science, The University of Kansas, Lawrence, KS 66045, USA. E-mail: lerner@poincare.math.ukans.edu.

be larger, but can still be considerably smaller than  $2D_0 + 1$  [8]. The criterion for choosing the embedding dimension is therefore relatively clear.

The proper choice of the delay time  $\tau$ , on the other hand, is to some extent still an open question. The basic idea is to choose  $\tau$  so that the coordinates  $x_n$  and  $x_{n+\tau}$  are independent of each other but not completely uncorrelated. If the time series is derived from a continuous dynamical system and  $\tau$  is too small,  $x_n$  and  $x_{n+\tau}$  are too correlated to serve as independent variables in the reconstructed phase space. If  $\tau$  is too large, then for a chaotic time series,  $x_n$  and  $x_{n+\tau}$  are effectively uncorrelated and hence the underlying deterministic dynamics may be lost. A series of previous works [9–14] addresses the choice of the proper delay time, and the criteria presented there work well in practice.

In this paper, we present both theoretical and numerical results establishing an upper bound for the proper delay time for a chaotic time series of finite length. By analyzing the effective scaling regime used to compute the correlation dimension using the GP algorithm, we derive an inequality relating the maximum allowed delay time to quantities such as the embedding dimension, the length of the time series, and the dynamical properties of the underlying chaotic set. Our main result is that for a given embedding dimension  $m$ , there exists a maximum allowed delay time  $\tau_{\max}$  beyond which the computation gives misleading results. As  $m$  increases,  $\tau_{\max}$  decreases substantially.

Our starting point is the GP algorithm which allows us to compute  $D_2$  in terms of a correlation integral  $C(\epsilon, \tau, m)$ , defined as the probability that a pair of points chosen randomly with respect to the natural measure on the attractor is separated by a distance smaller than  $\epsilon$  [5]. For a trajectory of length  $N$  in the embedding space, this integral can be approximated by

$$C_N(\epsilon, \tau, m) = \frac{2}{N(N-1)} \sum_{j=1}^N \sum_{i=j+1}^N \Theta(\epsilon - |\mathbf{x}_i - \mathbf{x}_j|). \quad (1)$$

Here,  $\Theta(\cdot)$  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 maximum norm of distance from  $\mathbf{x}_i$

to  $\mathbf{x}_j$  in the reconstructed phase space<sup>3</sup>. For large  $N$ ,  $C_N(\epsilon, \tau, m) \approx C(\epsilon, \tau, m)$ , and the correlation dimension  $D_2$  is given by

$$D_2 = \lim_{\epsilon \rightarrow 0} \frac{\log C(\epsilon, \tau, m)}{\log \epsilon}. \quad (2)$$

In practice,  $D_2$  is estimated by extracting the slope in the plot of  $\log C_N(\epsilon, \tau, m)$  versus  $\log \epsilon$  in an apparently linear scaling region. For a proper choice of  $\tau$ , the computed  $D_2$  at first increases with  $m$  and then stabilizes for sufficiently large values of  $m$ . This “plateau” value is then taken as an estimate of the attractor’s correlation dimension. The accuracy of the estimate is evidently determined by the size of the effective linear scaling region. For a given  $m$ , as the delay time  $\tau$  changes, so does the size of the linear scaling region; this gives us a way to estimate  $\tau_{\max}$ .

Our purpose now is to establish conservative upper and lower bounds for  $\log \epsilon_{\min}$  and  $\log \epsilon_{\max}$ . This will allow us to fix an upper limit for the delay time  $\tau$  (for fixed embedding dimension  $m$  and number of data points  $N$ ).

Inside the scaling region, the plots of  $\log(C_N(\epsilon, \tau, m))$  against  $\log \epsilon$  give a family of parallel lines with slope  $D_2$  whose equations have been determined by Grassberger and Procaccia [16],

$$\log_2 C_N(\epsilon, \tau, m) = D_2 \log_2 \epsilon - m\tau TK_2 \log_2 e. \quad (3)$$

Here  $K_2$  is the order-2 entropy. Equivalently, if  $J$  is the number of distinct pairs of points on the trajectory less than  $\epsilon$  units apart, then in the scaling region

$$J = \frac{N^2}{2} \epsilon^{D_2} e^{-m\tau TK_2}.$$

Imposing the modest requirement that  $J$  be  $> 1$  throughout the good scaling region, we obtain a constraint which, when solved for  $\epsilon$ , establishes the lower bound of the region as

$$\frac{1 - 2 \log_2 N + \tau m TK_2 \log e}{D_2} < \log_2 \epsilon_{\min}(\tau, m). \quad (4)$$

Note that  $\epsilon_{\min}(\tau, m)$  is an increasing function of both its arguments.

<sup>3</sup> Historically, the possibility of estimating  $D_2$  and other dynamical invariants (with maximum norm) from delay-coordinate embedding was proposed by Takens [3].

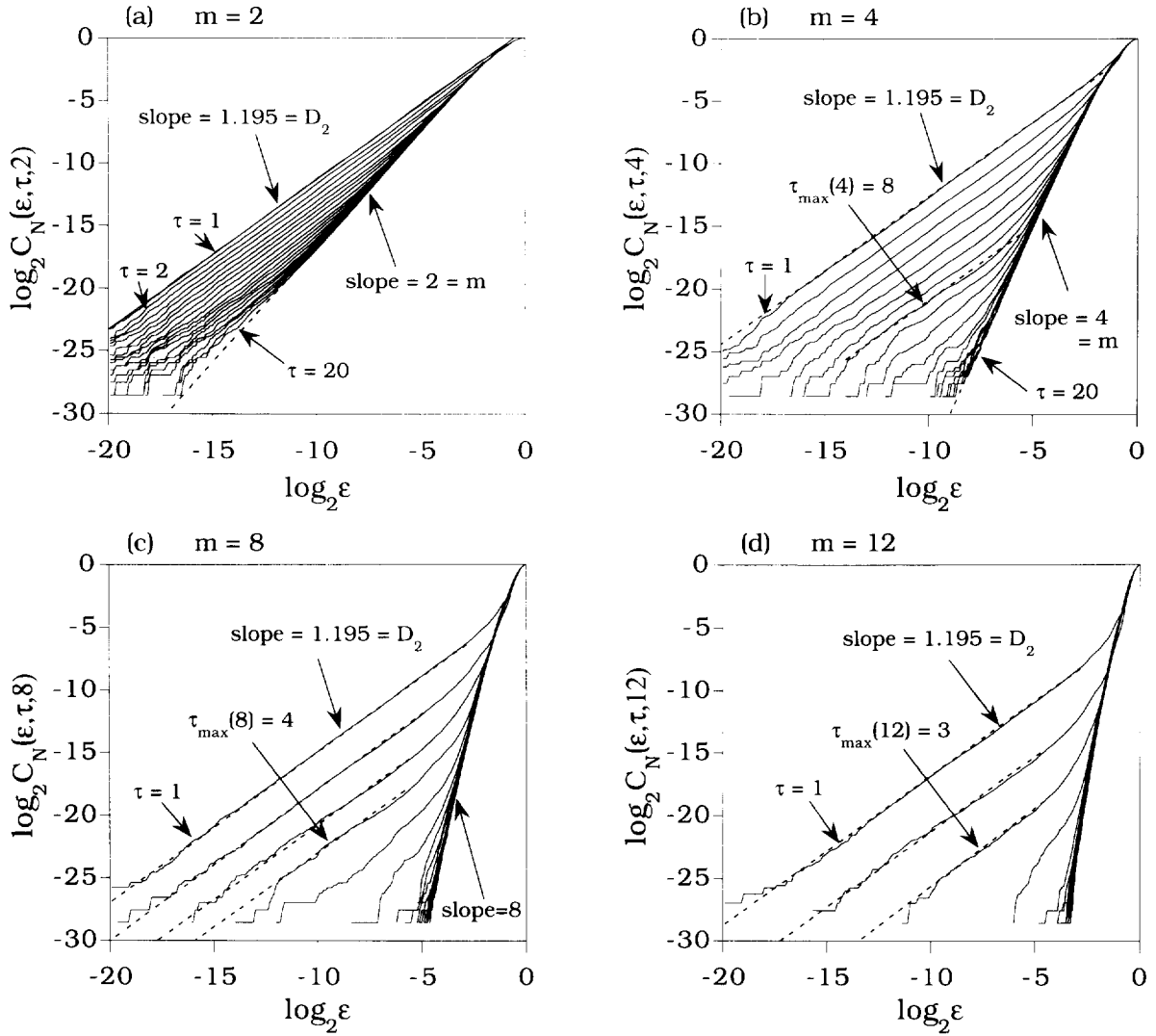


Fig. 1. For a time series of length  $N = 28000$  from the Hénon map at  $a = 1.4$  and  $b = 0.3$ ,  $\log_2 C_N(\epsilon, \tau, m)$  versus  $\log_2 \epsilon$  for  $m = 2$  (a), 4 (b), 8 (c), and 12 (d). In each figure, there are 20 plots corresponding to values of the delay time  $\tau$  ranging from 1 to 20.

Next, we consider  $\epsilon_{\max}(\tau, m)$ , which determines the upper end point of the linear scaling region. For large values of  $\tau$ , each delay coordinate is essentially independent of the others. Consequently, at large distances  $\epsilon$ , the behavior of  $C_N(\epsilon, \tau, m)$  will be similar to that obtained from a sequence of random vectors in the  $m$ -dimensional embedding space. Thus, we expect the family of curves to have a straight line with slope  $m$  as its envelope. This behavior is indeed observed in numerical computations as shown in Fig. 1. The line

given by Eq. (3) intersects the line of slope  $m$  through the origin at a value of  $\epsilon$  which satisfies

$$D_2 \log_2 \epsilon - m\tau TK_2 \log_2 e = m \log_2 \epsilon.$$

Taking this for the upper bound of the scaling region gives

$$\log_2 \epsilon_{\max} < \frac{-m\tau TK_2 \log_2 e}{m - D_2}. \quad (5)$$

Using Eqs. (4) and (5), and letting  $\Delta$  represent the size of the linear scaling region, we find

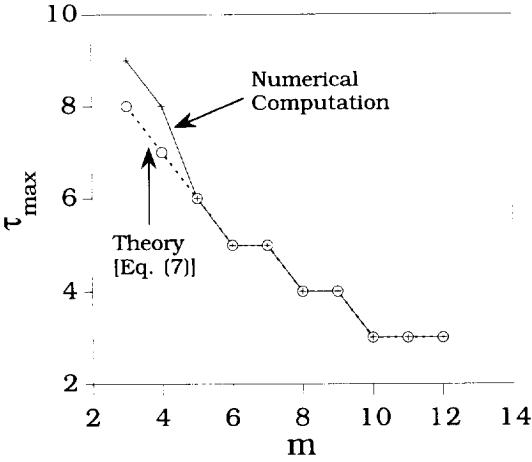


Fig. 2. Theoretical and numerical results for the upper bound of the delay time at embedding dimensions ranging from  $m = 2$  to  $m = 12$  for the standard Hénon map.

$$\begin{aligned} \Delta &= \log_2 \epsilon_{\max}(\tau, m) - \log_2 \epsilon_{\min}(\tau, m) \\ &< \frac{mK_2\tau T \log_2 e}{m - D_2} \\ &+ \frac{2 \log_2 N - 1 - m\tau TK_2 \log_2 e}{D_2}. \end{aligned} \quad (6)$$

Note that  $\Delta$  depends on the quantities  $m$  and  $\tau$  through their product, which is consistent with the results of Albano et al. [17]. As a practical matter, in order to estimate  $D_2$ , the linear scaling region must span at least an order of magnitude in  $\epsilon$ . In terms of the base-2 logarithmic scale, we shall require  $\Delta \geq 4$ . Then, solving Eq. (6), for  $\tau$ , we obtain the desired upper bound for the delay time

$$\tau_{\max} < \frac{(2 \log_2 N - 1 - 4D_2)(m - D_2)}{m^2 K_2 T \log_2 e}, \quad (7)$$

which is our main result.

To verify the accuracy of Eq. (7), we have undertaken a series of numerical experiments [18]. Here we just report the results obtained with the standard Hénon map  $(x_{n+1}, y_{n+1}) = (1.4 - x_n^2 + 0.3y_n, x_n)$ , which is believed to possess a chaotic attractor with one positive Lyapunov exponent estimated at  $\lambda_1 \approx 0.42$ . Since there is only one positive Lyapunov exponent, we have  $K_2 \approx \lambda_1$  [16]. The correlation dimension of this attractor is  $D_2 \approx 1.195$ . Thus, for sufficiently long time series, an embedding dimension of  $m = 2$  should suffice to extract the correct value of  $D_2$

[8]. To generate such a series, we take a random initial condition, iterate it 2000 times to get rid of the initial transient, and then record the next 28000 values of  $x_n$ . The correlation integral  $C_N(\epsilon, \tau, m)$  is evaluated at 200 values of  $\epsilon$  for  $\log_2 \epsilon \in [-20, 0]$ , for embedding dimensions ranging from  $m = 2$  to  $m = 12$ , and for delay times  $\tau$  ranging from 1 to 20. Since  $T = 1$  and  $N = 28000$ , Eq. (7) reduces to

$$\tau_{\max}(m) < 39.22 \left( \frac{m - D_2}{m^2} \right). \quad (8)$$

Figs. 1a–1d show  $\log_2 C_N(\epsilon, \tau, m)$  versus  $\log_2 \epsilon$  for  $m = 2, 4, 8, 12$ , respectively, where in each figure, curves correspond to different values of the delay times ranging from  $\tau = 1$  to  $\tau = 20$ . For large values of the delay  $\tau$ , delay coordinates are essentially independent of each other and, hence, we see that the curves  $\log_2 C_N(\epsilon, \tau, m)$  versus  $\log_2 \epsilon$  have a slope  $m$  for large values of  $\epsilon$ . The curves follow this envelope for  $\epsilon > \epsilon_{\max}(\tau, m)$  and then level off to obey the asymptotic law  $\log_2 C_N(\epsilon, \tau, m) \sim D_2 \log_2 \epsilon - K_2 \tau T m \log_2 e$ . For  $m = 4$ , we have  $\tau_{\max}(4) \approx 7$  from Eq. (8), while numerical computation gives  $\tau_{\max}(4) \approx 8$  (Fig. 1b). For  $m = 8$ , we have  $\tau_{\max}(4) \approx 4$  from Eq. (8), and numerics gives  $\tau_{\max}(4) \approx 4$  (Fig. 1c). Fig. 2 shows both the theoretical prediction for  $\tau_{\max}(m)$  and the corresponding numerical results for embedding dimension ranging from  $m = 3$  to  $m = 12$ . The agreement between Eq. (8) and the numerics for  $m \geq 5$  is good.

We briefly mention two additional issues related to the bound on the delay time. The first is the influence of environmental noise, whose presence can significantly reduce the range from which  $\tau$  can be chosen. Thus the choice of a proper delay time is more limited in practical applications where random noise is inevitable. The second issue is the appearance of a “knee” behavior in the plot of correlation integrals when a somewhat large delay time, but still less than the upper bound, is used. This behavior is qualitatively different from that due to excessive autocorrelation between nearby points in the reconstructed space [19], and therefore it cannot be removed by excluding the contribution from nearby points to the correlation integral [18].

As the field of chaotic dynamics matures, it becomes important to establish and test theoretical relations between the various system parameters such

as  $D_2$  and  $\lambda_1$  and those such as  $m$  and  $\tau$  which are to some extent under the control of the observer. A well-known example is the result of Eckmann and Ruelle [15]. In this spirit, we have shown the existence of such a relationship for the size of the effective linear scaling region which incorporates an upper bound for the delay time. Our result agrees well with the numerical experiments.

Y.C. Lai acknowledges partial support from AFOSR through grant no. F49620-96-1-0066 and a New Faculty Research Fund at the University of Kansas. This work was also supported by the Kansas Institute for Theoretical and Computational Science through the NSF/K\*STAR EPSCoR Program.

## References

- [1] H.D.I. Abarbanel, R. Brown, J.J. Sidorowich and L.S. Tsimring, *Rev. Mod. Phys.* 65 (1993) 1331.
- [2] N.H. Packard, J.P. Crutchfield, J.D. Farmer and R.S. Shaw, *Phys. Rev. Lett.* 45 (1980) 712.
- [3] F. Takens, in: *Lecture notes in mathematics*, Vol. 898. *Dynamical systems and turbulence*, eds. D. Rand and L.S. Young (Springer, Berlin, 1981) p. 336.
- [4] T. Sauer, J.A. Yorke and M. Casdagli, *J. Stat. Phys.* 65 (1991) 579.
- [5] P. Grassberger and I. Procaccia, *Phys. Rev. Lett.* 50 (1983) 346; *Physica D* 9 (1983) 189; A. Ben-Mizrachi, I. Procaccia and P. Grassberger, *Phys. Rev. A* 29 (1984) 975.
- [6] G. Mayer-Kress, ed., *Dimensions and entropies in chaotic systems* (Springer, Berlin, 1986).
- [7] P. Grassberger, in: *Chaos*, ed. A.V. Holden (Manchester Univ. Press, Manchester, 1986).
- [8] M. Ding, C. Grebogi, E. Ott, T. Sauer and J.A. Yorke, *Phys. Rev. Lett.* 70 (1993) 3872; *Physica D* 69 (1993) 404.
- [9] A.M. Fraser and H.L. Swinney, *Phys. Rev. A* 33 (1986) 1134; A.M. Fraser, *IEEE Trans. Info. Theory* 35 (1989) 245.
- [10] W. Liebert and H.G. Schuster, *Phys. Lett. A* 142 (1989) 107; W. Liebert, K. Pawelzik and H.G. Schuster, *Europhys. Lett.* 14 (1991) 521.
- [11] J. Theiler, *J. Opt. Soc. Am. A* 7 (1990) 1055; *Phys. Rev. A* 41 (1990) 3038.
- [12] T. Buzug and G. Pfister, *Phys. Rev. A* 45 (1992) 7073.
- [13] A. Provenzale, L.A. Smith, R. Vio and G. Murante, *Physica D* 58 (1992) 31.
- [14] G. Kember and A.C. Fowler, *Phys. Lett. A* 179 (1993) 72.
- [15] J.-P. Eckmann and D. Ruelle, *Physica D* 56 (1992) 185.
- [16] P. Grassberger and I. Procaccia, *Physica D* 13 (1984) 34.
- [17] A.M. Albano, J. Muench, C. Schwartz, A.I. Mees and P.E. Rapp, *Phys. Rev. A* 38 (1988) 3017; P.E. Rapp, A.M. Albano, T.I. Schmah and L.A. Farwell, *Phys. Rev. E* 47 (1993) 2289.
- [18] Y.-C. Lai, D. Lerner and R. Hayden, to be published.
- [19] J. Theiler, *Phys. Rev. A* 34 (1986) 2427.