ELSEVIER

Physica D 110 (1997) 43–50

PHYSICA D

Practical method for determining the minimum embedding dimension of a scalar time series

Liangyue Cao

Department of Mathematics, University of Western Australia, Nedlands, WA 6907, Australia

Received 11 September 1996; received in revised form 20 April 1997; accepted 6 May 1997
Communicated by A.M. Albano

Abstract

A practical method is proposed to determine the minimum embedding dimension from a scalar time series. It has the following advantages: (1) does not contain any subjective parameters except for the time-delay for the embedding; (2) does not strongly depend on how many data points are available; (3) can clearly distinguish deterministic signals from stochastic signals; (4) works well for time series from high-dimensional attractors; (5) is computationally efficient. Several time series are tested to show the above advantages of the method.

There have been many discussions on how to determine the optimal embedding dimension from a scalar time series based on Takens' theorem [1] or its extensions [2]; for a survey, see e.g., Ott et al. [3]. Following are the three basic methods which are usually used to choose the minimum embedding dimension:

(1) computing some invariant on the attractor [4]. By increasing the embedding dimension used for the computation one notes when the value of the invariant stop changing. The typical problem with this approach is that it is often very data intensive, certainly subjective, and time-consuming for computation.

(2) singular value decomposition [5]. The procedure identifies orthogonal directions in the embedding space which may be ordered according to the magnitude of the variance of the trajectory's projection on them. The ordering is done using the singular values of the embedding. The number of these directions visited by the reconstructed trajectory, and indicated by large singular values, is an estimate of the dimension of the smallest space that contains the trajectory. This approach is also subjective to some extent. As Mees et al. [6] pointed out, the number of large singular values may depend on the details of the embedding and the accuracy of the data as much as they do on the dynamics of the system.

(3) the method of false neighbors [7]. It was developed based on the fact that choosing too low an embedding dimension results in points that are far apart in the original phase space being moved closer together in the reconstruction space. Certainly this method is a good approach. But the criterion in [7] is subjective in some sense for saying that a neighbor is false, where different values of the parameters R_{tol} and A_{tol} may lead to different results (see example (v) later). Therefore, for realistic time series, different optimal embedding dimensions are probably obtained if we use

PII S0167-2789(97)00118-8

different values of the parameters. In addition, one will see some other weaknesses of the false neighbor method through our examples in this paper.

There are other methods and some modified methods developed, based on the above methods, e.g., [8], but they are more or less subjective as well in determining the minimum embedding dimension, or contain the other shortcomings we mentioned above.

The method presented in this paper overcomes the shortcomings of the above methods. Suppose that we have a time series $x_1, x_2, \ldots, x_N$. The time-delay vectors can be reconstructed as follows:

$$y_i(d) = (x_i, x_{i+\tau}, \ldots, x_{i+(d-1)\tau}),$$
$$i = 1, 2, \ldots, N - (d-1)\tau,$$

where d is the embedding dimension and τ is the time-delay. Note that $y_i(d)$ means the ith reconstructed vector with embedding dimension d. Similar to the idea of the false neighbor method [7], we define

$$a(i,d) = \frac{||y_i(d+1) - y_{n(i,d)}(d+1)||}{||y_i(d) - y_{n(i,d)}(d)||},$$
$$i = 1, 2, \ldots, N - d\tau, \tag{1}$$

where $||\cdot||$ is some measurement of Euclidian distance and is given in this paper by the maximum norm, i.e.,

$$||y_k(m) - y_l(m)|| = \max_{0 \le j \le m-1} |x_{k+j\tau} - x_{l+j\tau}|;$$

$y_i(d+1)$ is the ith reconstructed vector with embedding dimension $d+1$, i.e., $y_i(d+1) = (x_i, x_{i+\tau}, \ldots, x_{i+d\tau})$; $n(i,d)$ $(1 \le n(i,d) \le N - d\tau)$ is an integer such that $y_{n(i,d)}(d)$ is the nearest neighbor of $y_i(d)$ in the *d-dimensional* reconstructed phase space in the sense of distance $||\cdot||$ we defined above. $n(i,d)$ depends on i and d.

Notes. (1) The $n(i,d)$ in the numerator of Eq. (1) is the same as that in the denominator. (2) If $y_{n(i,d)}(d)$ equals $y_i(d)$, we take the second nearest neighbor instead of it.

If d is qualified as an embedding dimension by the embedding theorems [1,2], then any two points which stay close in the d-dimensional reconstructed space will be still close in the $(d+1)$-dimensional reconstructed space. Such a pair of points are called true neighbors, otherwise, they are called false neighbors. Perfect embedding means that no false neighbors exist. This is the idea of the false neighbor method in [7], where the authors diagnosed a false neighbor by seeing whether the $a(i,d)$ (note: in [7], $a(i,d) = |x_{i+d\tau} - x_{n(i,d)+d\tau}|/||y_i(d) - y_{n(i,d)}(d)||$, which is a little bit different from Eq. (1)) is larger than some given threshold value. The problem is how to choose this threshold value. From the definition of $a(i,d)$ in [7] or Eq. (1), one can see that the threshold value should be determined by the derivative of the underlying signal, therefore, for different phase points i, $a(i,d)$ should have different threshold values at least in principle. Furthermore, different time series data may have different threshold values. These imply that it is very difficult and even impossible to give an appropriate and reasonable threshold value which is independent of the dimension d and each trajectory's point, as well as the considered time series data.

To avoid the above problem, we instead define the following quantity, i.e., the mean value of all $a(i,d)$'s,

$$E(d) = \frac{1}{N - d\tau} \sum_{i=1}^{N-d\tau} a(i,d). \tag{2}$$

$E(d)$ is dependent only on the dimension d and the lag τ. To investigate its variation from d to $d+1$, we define

$$E1(d) = E(d+1)/E(d). \tag{3}$$

We found that $E1(d)$ stops changing when d is greater than some value d_0 if the time series comes from an attractor. Then $d_0 + 1$ is the minimum embedding dimension we look for.

Remarks. The parameter τ is a necessary parameter which must be given before the minimum embedding dimension is determined numerically no matter what methods are used. Although in principle the embedding dimension is independent of the time delay τ, the *minimum* embedding dimension is dependent on τ in practice. Different values of τ may lead to different minimum embedding dimensions, especially for time series from continuous time systems, see example (vi) later (for time series from discrete time maps, the best choice of τ is 1). It can be easily understood that a good choice of τ may decrease the minimum

embedding dimension which is necessary for phase space reconstruction. This is also one of the reasons why we usually choose a time delay such that each component of the reconstructed vectors is (linearly or nonlinearly) independent as far as possible. The methods for choosing such a τ are, for example, autocorrelation function method and mutual information method [9]. We will use the method of mutual information to choose the parameter τ in our examples later.

Before we do numerical tests, it is necessary to define another quantity which is useful to distinguish deterministic signals from stochastic signals. Let

$$E^*(d) = \frac{1}{N - d\tau} \sum_{i=1}^{N-d\tau} |x_{i+d\tau} - x_{n(i,d)+d\tau}|, \qquad (4)$$

where the meaning of $n(i, d)$ is the same as above, i.e., it is the integer such that $y_{n(i,d)}(d)$ is the nearest neighbor of $y_i(d)$. We define

$$E2(d) = E^*(d+1)/E^*(d). \qquad (5)$$

For time series data from a random set of numbers, $E1(d)$, in principle, will never attain a saturation value as d increases. But in practical computations, it is difficult to resolve whether the $E1(d)$ is slowly increasing or has stopped changing if d is sufficiently large. In fact, since available observed data samples are limited, it may happen that the $E1(d)$ stops changing at some d although the time series is random. To solve this problem, we can consider the quantity $E2(d)$. For random data, since the future values are independent of the past values, $E2(d)$ will be equal to 1 for any d in this case. However, for deterministic data, $E2(d)$ is certainly related to d, as a result, it cannot be a constant for all d; in other words, there must exist some d's such that $E2(d) \neq 1$.

We recommend calculating both $E1(d)$ and $E2(d)$ for determining the minimum embedding dimension of a scalar time series, and to distinguish deterministic data from random data. Now we turn to test several time series.

(i) data from the x-component values of Hénon attractor [10] with the usual parameters ($a = 1.4$, $b = 0.3$). To investigate whether our method depends strongly on the length of the time series, we calculate the $E1(d)$ and the $E2(d)$ using 1000 data points and 10 000 data points, respectively, where we let the time delay τ equal to 1. Shown in Fig. 1 are our results. Very clearly the minimum embedding dimension is 2, and the result does not strongly depend on how many data points are used.

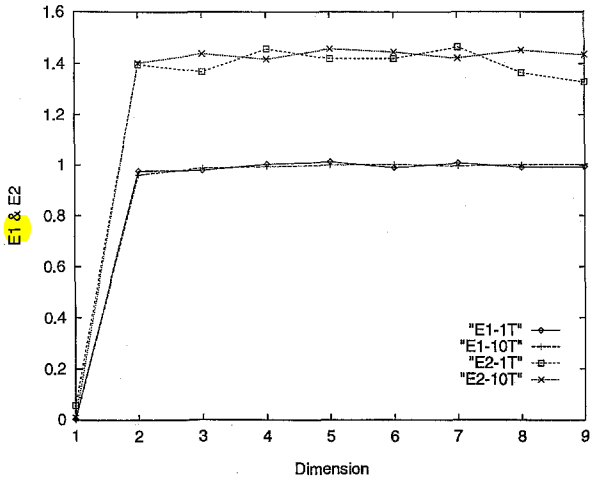

Fig. 1. The values $E1$ and $E2$ for time series data from the chaotic Hénon attractor. "$E1$-1T" and "$E1$-10T" represent $E1$ values obtained using 1000 data points and 10 000 data points, respectively, and the similar meaning for the "$E2$-1T" and "$E2$-10T".

(ii) data from the x-component values of Ikeda attractor [11]. The Ikeda map is as follows:

$$\begin{cases} x_{n+1} = p + \mu(x_n \cos(t) - y_n \sin(t)), \\ y_{n+1} = \mu(x_n \sin(t) + y_n \cos(t)), \end{cases}$$

where $p = 1.0$, $\mu = 0.9$, $t = k - \alpha/(1 + x_n^2 + y_n^2)$ with $k = 0.4$ and $\alpha = 6.0$. The map is the same as that considered in [7]. In the same way as above, we also calculate the $E1(d)$ and the $E2(d)$ using 1000 data points and 10 000 data points, respectively, where $\tau = 1$. Shown in Fig. 2 are our results. Very clearly the minimum embedding dimension is 4, which is the same as the result in [7]. Our result does not strongly depend on how many data points are used. Here we also used the false neighbor method, and found that if we use 1000 data points, the result is not stable, i.e., the percentage of false neighbors increases after dimension 4. This implies that more data points are

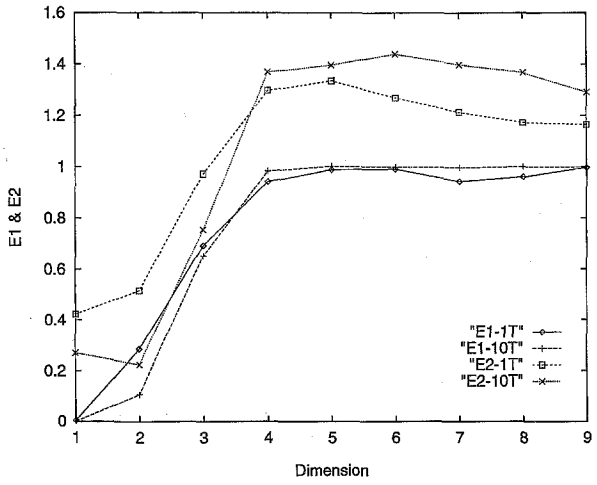

Fig. 2. The same as Fig. 1 but the data come from the Ikeda attractor.

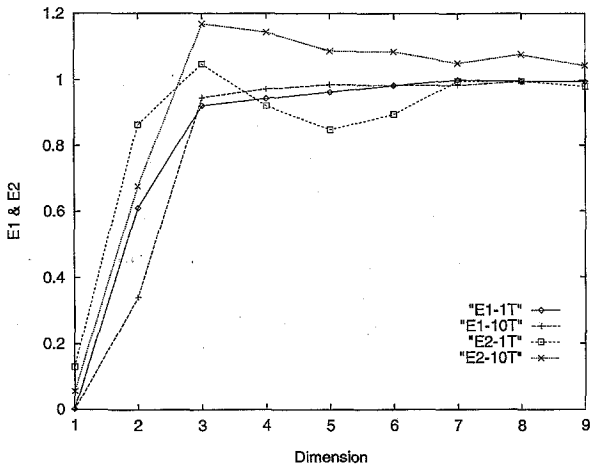

Fig. 3. The values $E1$ and $E2$ for the data from the chaotic Lorenz attractor.

needed using the false neighbor method than using our method.

(iii) data from the x-component values of Lorenz attractor [12] with the usual parameters ($\sigma = 10$, $r = 28$, $b = 8/3$). We numerically integrate the equations with integral step 0.01, and record the time series data from the numerical solution (x-component values) with sampling time 0.01 after all transients have diminished. We test this time series as above and show the results in Fig. 3, where $\tau = 15$ sampling time (i.e., $\tau = 15 \times 0.01 = 0.15$). One can see that 3 is the minimum embedding dimension.

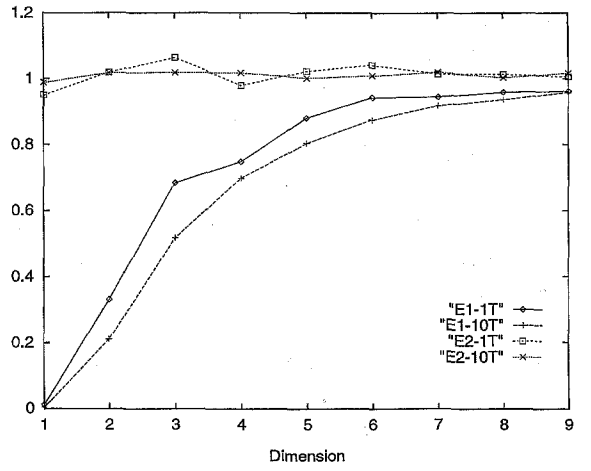

Fig. 4. The same as the above figures but the data come from random colored noise.

(iv) data from random colored noise. We generate time series data $\{x_n\}$ through:

$$x_{n+1} = 0.95x_n + y_n,$$

where $\{y_n\}$ is a white Gaussian time series. Such a time series can fool the false neighbor method. But our method can clearly distinguish it from deterministic chaos. Fig. 4 shows the results on this time series by our method, where $\tau = 1$. Here the $E2$ values approximately equal 1 for any d and have certainly no relation to the $E1$ values. As a comparison, we show the results by the false neighbor method in Fig. 5, where the values of two parameters in the method are: $A_{\text{tol}} = 2$ and $R_{\text{tol}} = 10$. One can see that the false neighbor method cannot distinguish this time series from chaotic time series.

(v) data from Mackey–Glass delay-differential equation [13],

$$\frac{dx(t)}{dt} = -0.1x(t) + \frac{0.2x(t-\Delta)}{1 + x(t-\Delta)^{10}}, \quad \Delta = 100. \tag{6}$$

Here our aim to choose $\Delta = 100$ in the above equation is to generate a time series from high-dimensional attractor [14]. The information dimension of the attractor generated by Eq. (6) is about 10 estimated by the

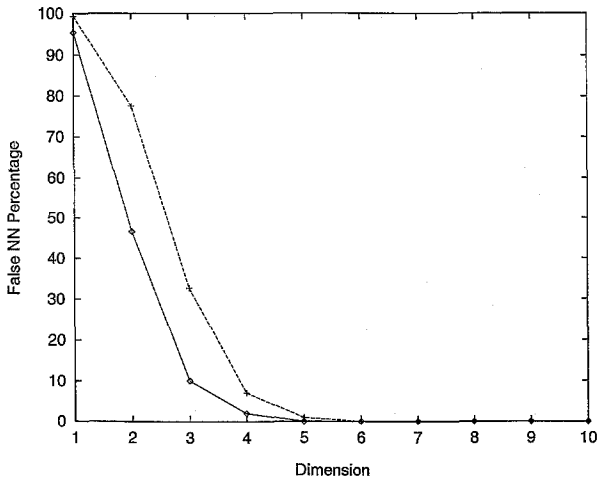

Fig. 5. The percentage of false nearest neighbors for the same time series data as used in Fig. 4, i.e., the random colored noise. The line with the diamonds corresponds to the case of using 1000 data points and the line with the crosses to the case of using 10 000 data points.

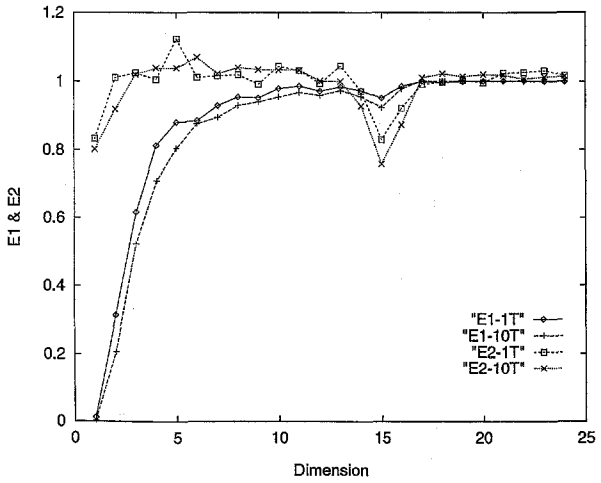

Fig. 6. The values $E1$ and $E2$ for the data from the Mackey–Glass delay-differential equation with delay equal 100.

Kaplan–Yorke formula [15], see also [16]. For a more accurate estimate of the attractor dimension, one can use the method proposed in [17]. As we know, most recent discussions of time series analysis are only concerned with low-dimensional chaotic systems. This is quite unsatisfactory. Through this example, we show that our method works very well for time series from high-dimensional attractors. We numerically integrate Eq. (6) with integral step 0.01, and record the time series data from the numerical solution with sampling time 6 after all transients have diminished. The results for this time series are shown in Fig. 6 by our method, where $\tau = 1$ sampling time (i.e., $\tau = 1 \times 6$). One can see that the $E1(d)$ attains its saturation value at $d = 17$ and so does the $E2(d)$, therefore, 17 should be the minimum embedding dimension for the time series in this example. Obviously the result does not strongly depend on how many data points are used. A curious phenomenon in Fig. 6 is that there is a sudden dip at $d = 15$. We do not fully understand it. The reason may be related to the fact that the sampling time used to generate the time series was too large so that the data is best modeled by a discrete map rather than a differential equation. We will be investigating it in our future work.

We now test the above Mackey–Glass time series using the false neighbor method. We fix the parameter $A_{\text{tol}} = 2$ and let the parameter $R_{\text{tol}} = 10$, 20, 30 and 40, respectively. The results are shown in Fig. 7 (10 000 data points are used). One can see that the minimum embedding dimension is about 6 by the method of false neighbors, which is too small since the information dimension of the attractor is about 10. In addition, we can also see that different values of R_{tol} can lead to different minimum embedding dimensions (for example, if we think 0.5% is small enough, then the minimum embedding dimension is, 6 for $R_{\text{tol}} = 10$, 5 for $R_{\text{tol}} = 20$, and 4 for $R_{\text{tol}} = 30$ and 40).

In the following we would like to test two unusual time series. The first one is usually considered to be an example for which it is difficult to determine the embedding dimension. For the second one, we know its exact embedding dimension, we just want to test if our method can give the correct result.

(vi) data from torus. The data are generated from the flow:

$$x(t) = \sin(t) + 0.3\sin(\pi t + 1), \quad t \in R \tag{7}$$

with the sampling time 0.1. Note that Eq. (7) is a flow, not a map. Shown in Fig. 8 are our results, where time delay $\tau = 10$ sampling time (i.e., $\tau = 10 \times 0.1 = 1$).

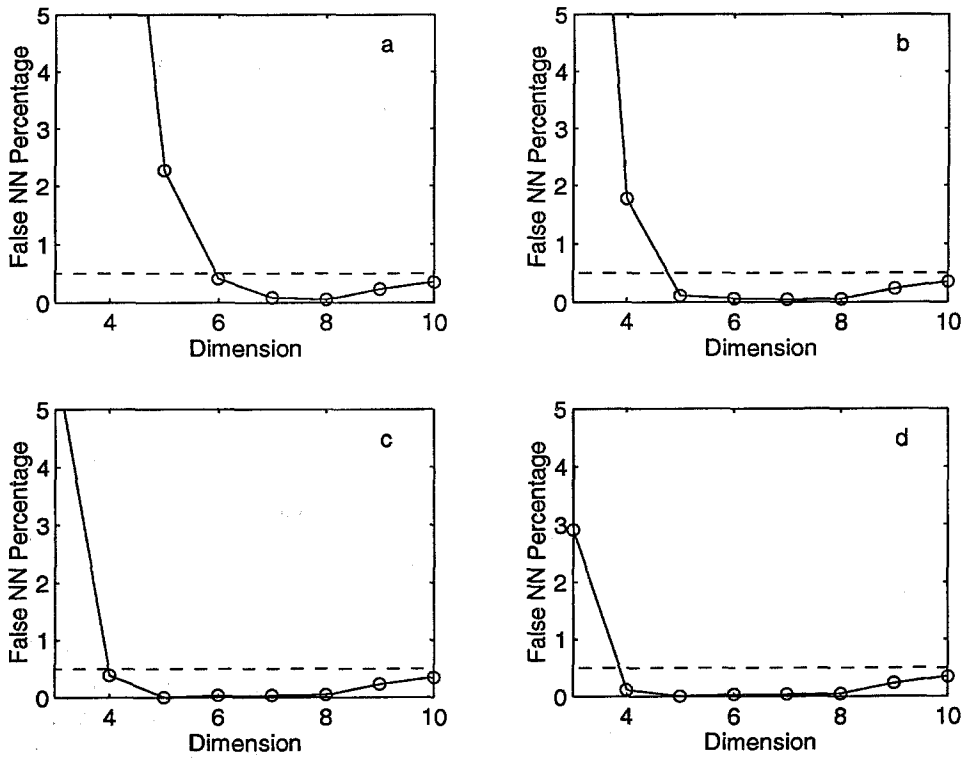

Fig. 7. The percentage of false nearest neighbors for the same time series data as used in Fig. 6. Here 10 000 data points are used. We only plot the points with the percentage of false nearest neighbors less than 5. The dashed straight line corresponds to the percentage equal to 0.5. (a) $R_{\text{tol}} = 10$. (b) $R_{\text{tol}} = 20$, (c) $R_{\text{tol}} = 30$, (d) $R_{\text{tol}} = 40$.

Very clearly the minimum embedding dimension is 3, and the result does not strongly depend on how many data points are used. Here we also used the false neighbor method and found the same results. We now take this example to test how the minimum embedding dimension is dependent on τ. Fig. 9 shows the results with $\tau = 2$ sampling time by our method, where one can obviously see that the minimum embedding dimension becomes equal to 4. In fact, $\tau = 2$ is too small and therefore the components of the reconstructed vectors with this τ are strongly correlated, as a result, the dimension which is needed to frame the attractor is larger than that with larger τ such as $\tau = 10$.

(vii) data from the following map:

$$x_{n+4} = \sin(x_n + 5) + \sin(2x_{n+1} + 5) + \sin(3x_{n+2} + 5) + \sin(4x_{n+3} + 5). \quad (8)$$

Certainly the minimum embedding dimension is 4 for time series from this map because we can exactly write down the following predictive model with embedding dimension 4 for this time series,

$$x_{n+4} = F(x_n, x_{n+1}, x_{n+2}, x_{n+3}) = \sin(x_n + 5) + \sin(2x_{n+1} + 5) + \sin(3x_{n+2} + 5) + \sin(4x_{n+3} + 5).$$

We check our method by testing this time series. Shown in Fig. 10 are our results, where $\tau = 1$. Obviously we obtain the minimum embedding dimension 4 which is the same as the theoretical result. We also test this time series using the false neighbor method, and show the results in Fig. 11, where $A_{\text{tol}} = 2$, $R_{\text{tol}} = 10$. From Fig. 11, we can see that the false neighbor method cannot distinguish this time series from random data if using 1000 data points, and the

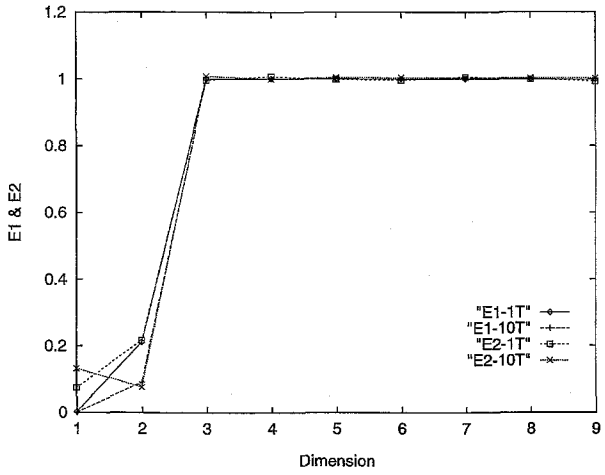

Fig. 8. The values $E1$ and $E2$ for the data from the torus Eq. (7), with $\tau = 10$ sampling time.

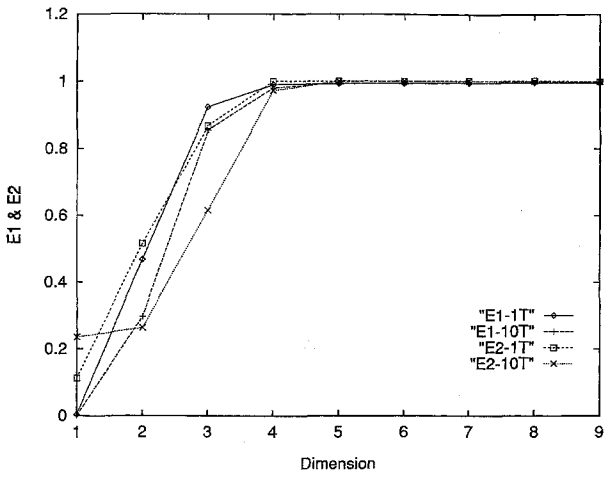

Fig. 9. Same as Fig. 8, but with $\tau = 2$ sampling time.

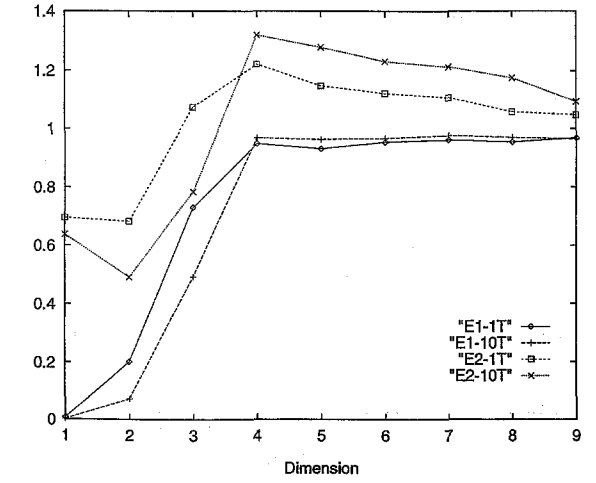

Fig. 10. The values $E1$ and $E2$ for the data from the map Eq. (8).

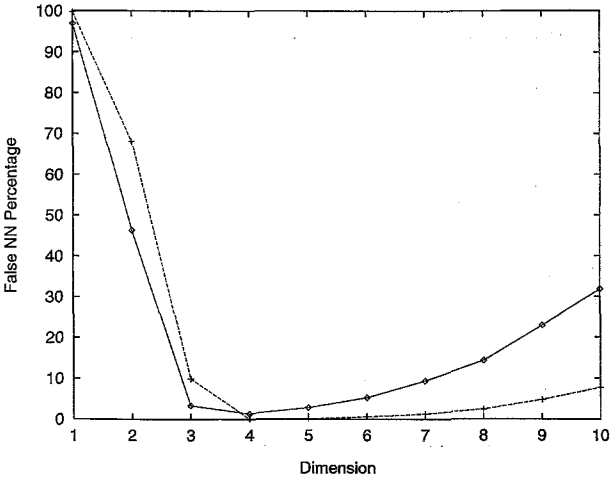

Fig. 11. The percentage of false nearest neighbors for the same data as used in Fig. 10, i.e., generated from the map Eq. (8). The line with the diamonds corresponds to the case of using 1000 data points and the line with the crosses to the case of using 10 000 data points.

result is not stable if using 10 000 data points, where the percentage of false neighbors increases after dimension 4.

The above examples have shown us the effectiveness of our method in applying to determine the minimum embedding dimension of a scalar time series. Since it does not contain any free parameters except the time delay τ, we could expect that our method would be more useful than the previous methods in practical time series analysis where the underlying dynamics is unknown. We would like to test a realistic time series as our final example.

(viii) data from the experimental laser data – Santa Fe Data A [18]. The time series contains 1000 samples. Fig. 12 shows the results by our method, and one can see that the minimum embedding dimension is 7, where $\tau = 1$. As a comparison, we show the result by the false neighbor method in Fig. 13, where

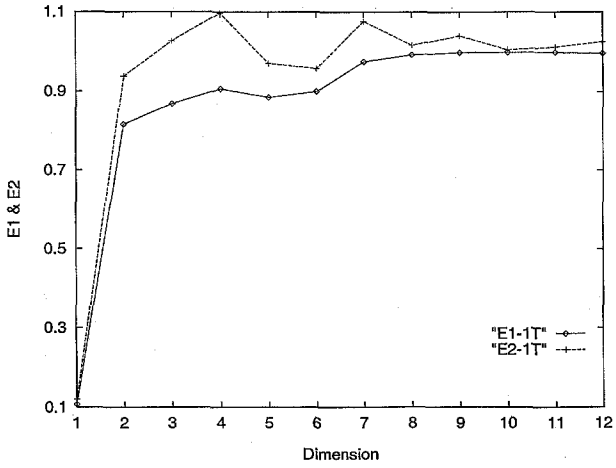

Fig. 12. The values $E1$ and $E2$ for the experimental laser data – Santa Fe Data A, where $\tau = 1$ sampling time, and the time series has only 1000 samples.

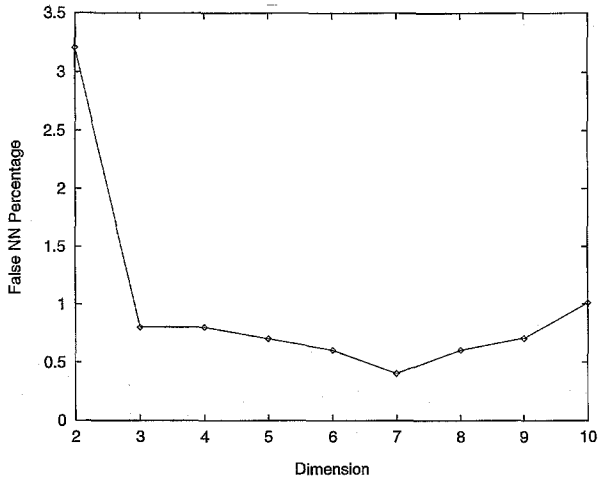

Fig. 13. The percentage of false nearest neighbors for the Santa Fe Data A, where $\tau = 1$ sampling time, $A_{\text{tol}} = 2$, $R_{\text{tol}} = 10$.

$A_{\text{tol}} = 2$, $R_{\text{tol}} = 10$. One can see that the result in Fig. 13 is not stable, where the percentage of false neighbors increases after dimension 7.

In summary, we have proposed a method for determining the minimum embedding dimension from a scalar time series, which has the advantages as mentioned in the abstract. Several time series were tested, and the numerical results showed that the method presented in this paper is very effective. We hope that our method will be useful in applications of nonlinear techniques to analyze realistic time series as well as artificial time series.

Acknowledgements

The author would like to thank Prof. Alistair Mees and Dr. Kevin Judd for useful discussions; and he wishes to thank the referees, whose valuable comments enhanced the clarity of this paper.

References

[1] F. Takens, Lecture Notes in Mathematics, Vol. 898 (Springer, Berlin, 1981) p. 366.
[2] T. Sauer, J.A. Yorke and M. Casdagli, J. Stat. Phys. 65 (1991) 579.
[3] E. Ott, T. Sauer and J.A. Yorke, Coping with Chaos (Wiley, New York, 1994).
[4] P. Grassberger and I. Procaccia, Phys. Rev. Lett. 50 (1983) 346.
[5] D.S. Broomhead and G.P. King, Physica D 20 (1986) 217.
[6] A.I. Mees, P.E. Rapp and L.S. Jennings, Phys. Rev. A 36 (1987) 340.
[7] M. Kennel, R. Brown and H. Abarbanel, Phys. Rev. A 45 (1992) 3403.
[8] D. Kaplan and L. Glass, Phys. Rev. Lett. 68 (1992) 427; A.M. Albano, J. Muench, C. Schwartz, A.I. Mees and P.E. Rapp, Phys. Rev. A 38 (1988) 3017; For a review; see H.D. Abarbanel, R. Brown, J. Sidorowich and L. Tsimring, Rev. Mod. Phys. 65 (1993) 1331.
[9] A.M. Fraser and H.L. Swinney, Phys. Rev. A 33 (1986) 1134; See also W. Liebert and H.G. Schuster, Phys. Lett. A 142 (1989) 107.
[10] M. Hénon, Commun. Math. Phys. 50 (1976) 69.
[11] K. Ikeda, Opt. Commun. 30 (1979) 257; S.M. Hammel, C.K.R.T. Jones and J.V. Moloney, J. Opt. Soc. Amer. 2 B (1985) 552.
[12] E.N. Lorenz, J. Atmospheric Sci. 20 (1963) 130.
[13] M.C. Mackey and L. Glass, Science 197 (1977) 287.
[14] J.D. Farmer, Physica D 4 (1982) 366.
[15] J.P. Eckmann and D. Ruelle, Rev. Modern Phys. 57 (1985) 617.
[16] M. Casdagli, Physica D 35 (1989) 335.
[17] M. Ding, C. Grebogi, E. Ott, T. Sauer and J.A. Yorke, Physica D 69 (1993) 404.
[18] A.S. Weigend and N.A. Gershenfeld, Time Series Prediction: Forecasting the Future and Understanding the Past (Addison-Wesley, Reading, MA, 1994).