Vol 16 No 5, May 2007
1009-1963/2007/16(05)/1252-06
Chinese Physics
© 2007 Chin. Phys. Soc.
and IOP Publishing Ltd

A new method of determining the optimal embedding dimension based on nonlinear prediction*

Meng Qing-Fang(孟庆芳)[a)†], Peng Yu-Hua(彭玉华)[a)], and Xue Pei-Jun(薛佩军)[b)]

[a)] *School of Information Science and Engineering, Shandong University, Jinan* 250100, *China*
[b)] *Graduate School of Shandong University, Jinan* 250100, *China*

(Received 15 August 2006; revised manuscript received 18 December 2006)

A new method is proposed to determine the optimal embedding dimension from a scalar time series in this paper. This method determines the optimal embedding dimension by optimizing the nonlinear autoregressive prediction model parameterized by the embedding dimension and the nonlinear degree. Simulation results show the effectiveness of this method. And this method is applicable to a short time series, stable to noise, computationally efficient, and without any purposely introduced parameters.

Keywords: embedding dimension, nonlinear autoregressive prediction model, nonlinear time series
PACC: 0545

1. Introduction

The reconstruction of a phase space which is equivalent to the original phase space of a system from a scale time series is the basis of nonlinear time series analysis.[1] Based on Whitney's topological embedding theorem, Takens[2] has proved the delay embedding theorem which offers the possibility of reconstructing m-dimensional dynamics from a scale time series in the absence of noise. However, the delay embedding theorem does not directly answer how to choose its parameters-embedding dimension and delay time. So the key to the phase space reconstruction from a real time series is to determine its optimal parameters. References [3–18] are devoted to the study of phase space reconstruction and methods of determining its parameters.

The optimal embedding dimension is an important parameter not only for the phase space reconstruction but also for the nonlinear prediction.[19,20] There have been many discussions on how to determine the optimal embedding dimension from a scalar time series. There are four basic methods which are usually used to determine the embedding dimension. The first kind of method is based on computing some invariant on the attractor.[5,6] By increasing the embedding dimension used for the computation one notes the minimum embedding dimension appearing when the value of the invariant stops changing. The typical problems of this method are time-consuming for computation, unsuited for a short time series, and sensitive to noise. The second kind of method is based on a singular value decomposition, which is proposed by Broomhead and King.[7] But the singular value decomposition is essentially a linear method based on the covariance matrix which reflects the linear dependence.[8] Numerical experience has led several researchers[9,10] to suspect the authenticity of this method in the analysis of nonlinear time series. The method in Ref.[11] is measure preserving and can reflect the nonlinear structure, but the results are considerably larger than those from the minimum embedding dimension. In Refs.[12,13] used is a four-order cumulant function instead of correlation function to improve this method. The third is the method of false neighbours,[14,15] which is based on the fact that choosing too low an embedding dimension results in points that are far apart in the original phase space moving closer together in the reconstruction phase space. But the criterion in Ref.[14] is subjective in some sense, and is improved in Ref.[15]. The fourth kind of method is based on nonlinear prediction errors,[16] in which the short term predictability of nonlinear time series is utilized, and the method used in this paper belongs in this kind.

There are some other methods and their modified

*Project supported by the Scientific Research Foundation for the Returned Overseas Chinese Scholars of China (Grant No 2004.176.4) and the Natural Science Foundation of Shandong Province of China (Grant No Z2004G01).
†E-mail: qfmeng@sohu.com

methods developed based on the above methods,[17,18] but they also have some shortcomings like inapplicability to a short time series, sensitivity to noise, etc.

In this paper we propose a new method to determine the optimal embedding dimension from a scalar time series by optimizing the nonlinear autoregressive prediction model parametrized by the embedding dimension and the nonlinear degree. Our method is applicable to a short time series, stable to noise, computationally efficient, without any purposely introduced parameters.

2. Method of determining the optimal embedding dimension

For nonlinear time series y_n, where $n = 1, 2, \ldots, N$, we use a closed-loop version of the dynamics, in which the output y_n feeds back as a delayed input. Within this framework, we analyse a nonlinear time series by using a discrete Volterra autoregressive series of degree d, and memory m, as a model to calculate the prediction time series $\hat{y}_n$, i.e.

$$\begin{aligned}\hat{y}_n = \; & a_0 + a_1 y_{n-1} + a_2 y_{n-2} + \cdots + a_m y_{n-m} \\ & + a_{m+1} y_{n-1}^2 + a_{m+2} y_{n-1} y_{n-2} + \cdots \\ & + a_{M-1} y_{n-m}^{\mathrm{d}} \\ = \; & \sum_{k=0}^{M-1} a_k z_k(n), \end{aligned} \tag{1}$$

where the functional basis $\{z_k(n)\}$ is composed of all the distinct combinations of the embedding space coordinates $(y_{n-1}, y_{n-2}, y_{n-3}, \ldots, y_{n-m},)$ up to degree d, with a total dimension $M = \mathrm{C}_{m+d}^{d} = (m + d)!/(m!d!)$. Thus, each model is parameterized by m and d which correspond to the embedding dimension and the degree of the nonlinearity of the model, respectively. The coefficients a_k can be calculated through the least square curve fitting.

The short term prediction power of a model can be measured by the standard one-step prediction error, i.e.

$$\varepsilon(m,d)^2 = \frac{\sum_{n=M+1}^{N} \left[\hat{y}_n(m,d) - y_n\right]^2}{\sum_{n=M+1}^{N} \left[y_n - \bar{y}\right]^2}, \tag{2}$$

where $\bar{y} = \dfrac{1}{N-M} \sum_{n=M+1}^{N} y_n$ and $\varepsilon(m,d)^2$ is in effect a normalized variance of the prediction error. We now search for the optimal model parameterized by $\{m_{\mathrm{opt}}, d_{\mathrm{opt}}\}$ that minimizes the following Akaike information criterion:[1,21]

$$C(m,d) = \ln \varepsilon(m,d) + 2M/N. \tag{3}$$

The numerical procedure is as follows: for each time series, obtain the best linear model by searching for m^{lin} which minimizes $C(m,d)$ with $d = 1$. Repeat with increasing $d > 1$, to obtain the model's optimal degree d_{opt} with $m = m^{\mathrm{lin}}$. Then repeat with increasing $m \geq 1$, to obtain the model's optimal embedding dimension with $d = d_{\mathrm{opt}}$.

The delay embedding theorem states that when the embedding dimension is larger than the minimum embedding dimension the original phase space of a system can be embedded in the reconstructed phase space. However, for a practical time series, if too large an embedding dimension is taken, the other nonlinear analysis methods are time-consuming for computation and influenced by noises. By increasing the embedding dimension ($m \geq 1$) of the nonlinear model parametrized by m and d, with the parameter d fixed, one notes the optimal embedding dimension when $C(m,d)$ stops drastically declining and begins to slowly decline. When $m = m_{\mathrm{opt}}$, the value of $C(m,d)$ is small enough, and with m increasing the value of $C(m,d)$ declines very little.

3. Simulation and results

To investigate whether our method is applicable to a short time series and whether our method depends strongly on the length of time series, we respectively use 500 data points and 6000 data points to determine the optimal embedding dimension.

Analyse the data from the x-component values of Henon attractor with the usual parameters $a = 1.4$ and $b = 0.3$. Our results are shown in Fig.1. We can see that when the embedding dimension $m = 2$, the value of $C(m,d)$ is small enough. So the optimal embedding dimension is 2, and the method is applicable to a short time series and does not strongly depend on how many data points are used.

Analyse the data from the x-component values of Lorenz attractor with the usual parameters $\sigma = 10$, $r = 28$, and $b = 8/3$. The sampling time is 0.01. Shown in Fig.2 are our results. We can see that when the embedding dimension $m = 3$, the value of $C(m,d)$ begins to slowly decline and becomes small enough.

So the optimal embedding dimension is 3, and the result is applicable to a short time series and does not strongly depend on how many data points are used.

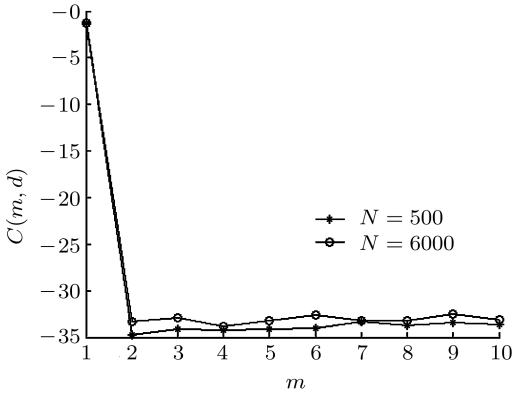


Fig.1. $C(m,d)-m$ for a time series from Henon attractor ($d=2$).

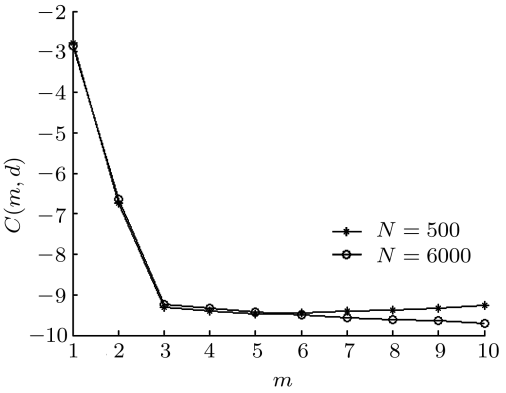


Fig.2. $C(m,d)-m$ for a time series from Lorenz attractor ($d=3$).

Analyse the data from the x-component values of Duffing attractor with the usual parameters $\varepsilon = 1$, $c = 0.4$, $A = 0.4$, and $\omega = 1$. The sampling time is 0.25. Our results are shown in Fig.3, from which we can see that the optimal embedding dimension is 4 and our method is applicable to a short time series and does not strongly depend on how many data points are used.

Analyse the data from the x-component values of Chen attractor with the usual parameters $a = 35$, $b = 3$, and $c = 28$. The sampling time is 0.01. Our results are shown in Fig.4, from which we can see that the optimal embedding dimension is 4 and our method is applicable to a short time series and does not strongly depend on how many data points are used.

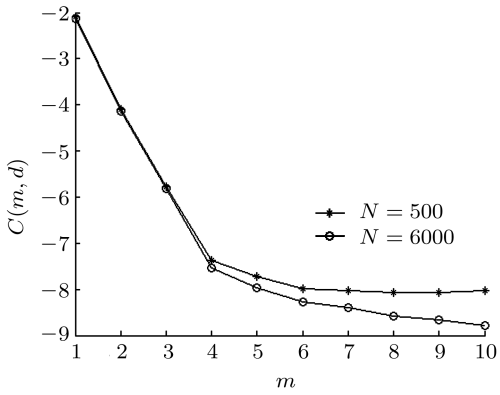


Fig.3. $C(m,d)-m$ for a time series from Duffing attractor ($d=3$).

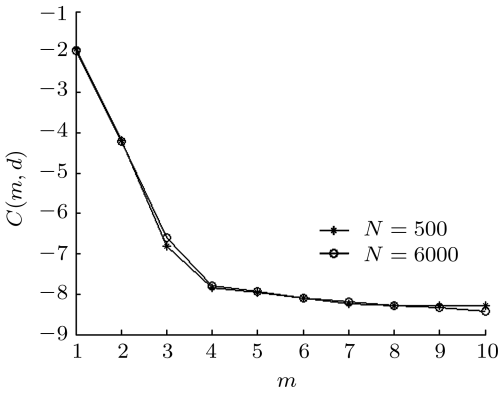


Fig.4. $C(m,d)-m$ for a time series from Chen attractor ($d=3$).

In the following we test two unusual time series. The first one is usually considered to be an example in which it is difficult to determine the embedding dimension. For the second one, we know its exact embedding dimension, and we just test whether our method can give the correct result or not.

Analyse the data from a torus. The data are generated from the flow:

$$x(t) = \sin(t) + 0.3\sin(\pi t + 1), \quad t \in R \tag{4}$$

with a sampling time of 0.1. Note that Eq.(4) is a flow, not a map. The results obtained by using our method for this time series are shown in Fig.5. One can see that the value of $C(m,d)$ is small enough at $m = 4$, therefore, the optimal embedding dimension for the time series should be 4. Obviously, the result is applicable to a short time series and does not strongly depend on how many data points are used.

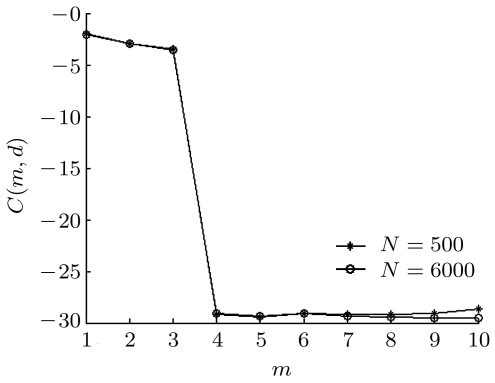


Fig.5. $C(m,d) - m$ for a time series from Torus attractor ($d = 3$).

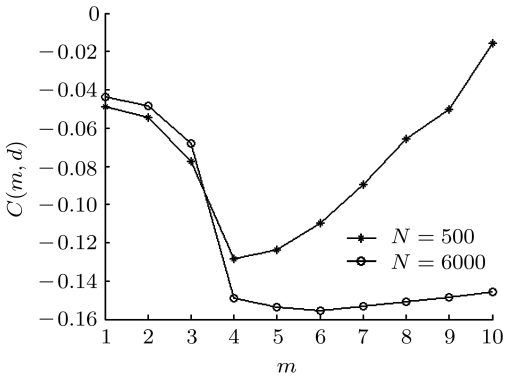


Fig.6. $C(m,d) - m$ for a time series from the map Eq.(5) ($d = 2$).

Analyse the data from the following map:

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

We can see clearly that the minimum embedding dimension is 4 for a time series from this map. We check our method by testing this time series, and the results are shown in Fig.6. We can see that when the embedding dimension $m = 4$, the value of $C(m,d)$ is small enough. So the optimal embedding dimension is 4 which is the same as the theoretical result.

The above examples have shown the effectiveness of our method in applying to the determination of the optimal embedding dimension of six time series from six chaotic attractors, and our method is applicable to a short time series and does not strongly depend on how many data points are used.

To investigate the robustness of our method under noises we analyse the time series from Henon attractor and Lorenz attractor with additive white Gaussian noise with different signal-to-noise ratios (SNRs). The results obtained by using our method are shown in Fig.7 and Fig.8. One can see that the optimal embedding dimensions for the time series from Henon attractor with SNR = 30 dB and SNR = 20 dB are 2, and the optimal embedding dimensions for the time series from Lorenz attractor with SNR = 30 dB and SNR = 20 dB are 5. So our method is robust to the additive white noise.

In the following we apply our method to a practical time series where the underlying dynamics is unknown.

Analyse the data from the experimental laser data- Santa Fe Data A. The time series contains 1000 samples. The results obtained by using our method are shown in Fig.9. We can see that the value of $C(m,d)$ begins to slowly decline and becomes small enough at $m = 8$, therefore, the optimal embedding dimension for this time series should be 8. Our result is in accord with that in Ref.[15].

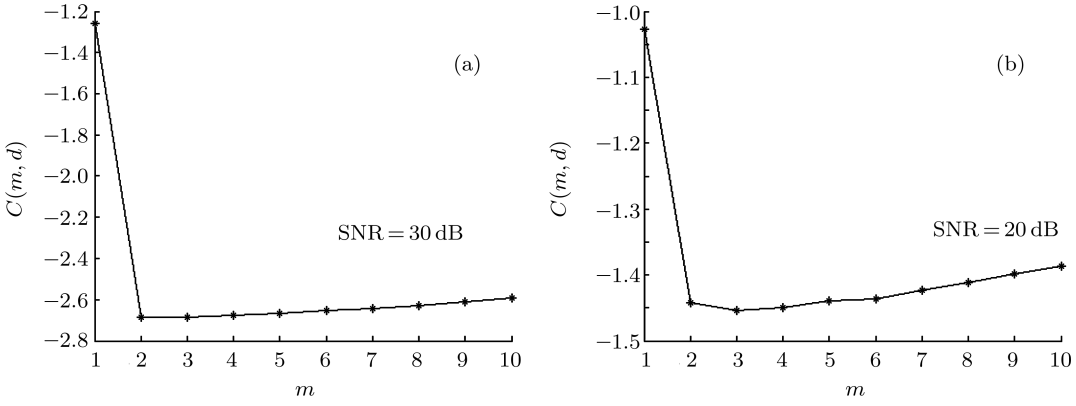


Fig.7. $C(m,d) - m$ for a time series from Henon attractor, with white noise added ($d = 2$) for (a) SNR = 30 dB; (b) SNR = 20 dB.

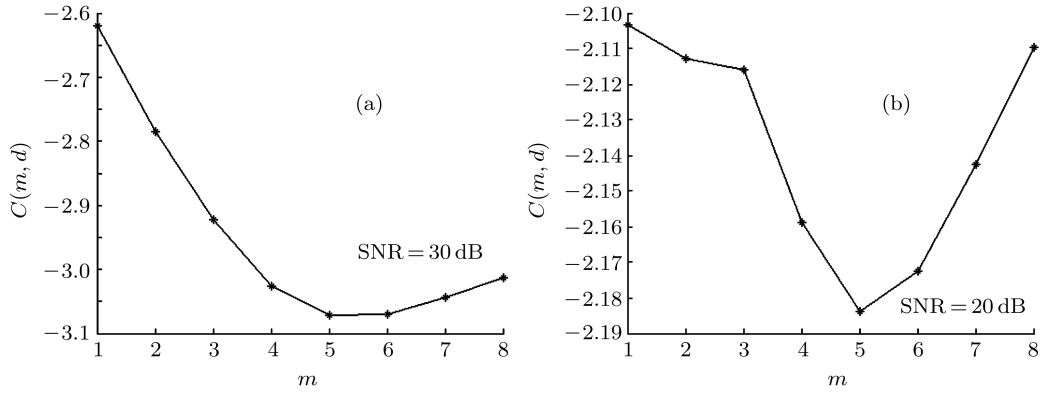


Fig.8. $C(m,d) - m$ for a time series from Lorenz attractor, with white noise added ($d = 3$) for (a) SNR = 30 dB; (b) SNR = 20 dB.

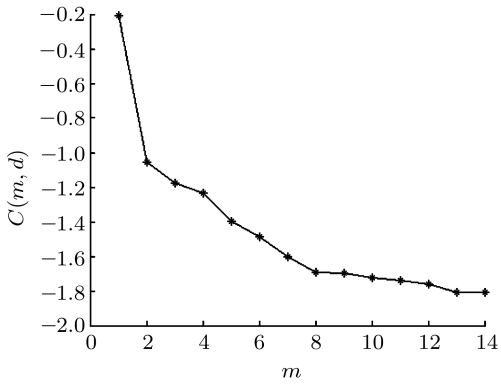


Fig.9. $C(m,d) - m$ for the experimental laser data-Santa Fe Data A, where the time series has only 1000 samples ($d = 2$).

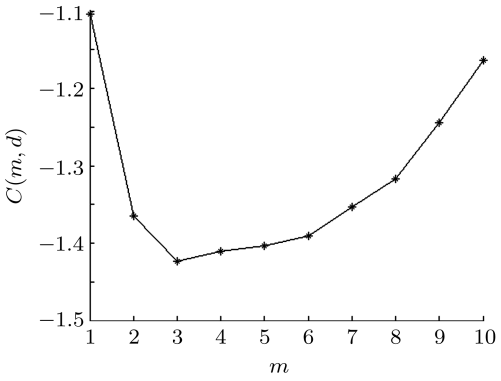


Fig.10. $C(m,d) - m$ for the heart rate variability data-Santa Fe Data B, where using only 1000 data points ($d = 3$).

Analyse the data from the heart rate variability signal- Santa Fe Data B, using 1000 data points. Figure 10 shows the results obtained by using our method. One can see that the value of $C(m,d)$ begins to slowly decline and becomes small enough at $m = 3$, therefore, the optimal embedding dimension for this time series should be 3. So the heart rate of the patient during sleep is mainly influenced by three factors.

4. Conclusion

A new method is proposed to determine the optimal embedding dimension from a scalar time series. We test six time series from six chaotic attractors, and the numerical results show the effectiveness of our method in applying to the determination of the optimal embedding dimension from a scalar time series. The method presented in this paper has the following advantages, it is (1) applicable to a short time series, (2) stable to noise, (3) computationally efficient (typically, the analysis of a 500-point time series takes just a few seconds on a desktop computer), and (4) without any purposely introduced parameters.

References

[1] Kantz H and Schreiber T 2003 *Nonlinear Time Series Analysis* (Second Edition) (Cambridge: Cambridge University Press)

[2] Takens F 1981 *Dynamical Systems and Turbulence* (Berlin: Springer) p366

[3] You R Y, Chen Z, Xu S C and Wu B X 2004 *Acta Phys. Sin.* **53** 2882 (in Chinese)

[4] Yang S Q and Jia C Y 2002 *Acta Phys. Sin.* **51** 2452 (in Chinese)

[5] Xiao F H, Yan G R and HanY H 2005 *Acta Phys. Sin.* **54** 550 (in Chinese)

[6] Grassberger P and Procaccia I 1983 *Phys. Rev. Lett.* **50** 346

[7] Broomhead D S and King G P 1986 *Physica* D **20** 217
[8] Palus M and Dvorak I 1992 *Physica* D **55** 221
[9] Mess A I, Rapp P E and Jennings L S 1987 *Phys. Rev.* A **36** 340
[10] Fraser A M 1989 *Physica* D **34** 391
[11] Lei M, Wang Z and Feng Z 2002 *Phys. Lett.* A **303** 179
[12] Yuan J and Xiao X C 1998 *Acta Phys. Sin.* **47** 897 (in Chinese)
[13] Meng Q F, Zhang Q and Pan J F 2005 *Syst. Eng. Theo. and Prac.* **25** 83 (in Chinese)
[14] Kennel M, Brown R and Abarbanel H D I 1992 *Phys. Rev.* A **45** 3403
[15] Cao L 1997 *Physica* D **110** 43
[16] Sugihara G and May R M 1990 *Nature* **344** 734
[17] Galka A, Maaβ T and Pfister G 1998 *Physica* D **121** 237
[18] Judd K and Mees A 1998 *Physica* D **120** 273
[19] Zhang J S, Li H C and Xiao X C 2005 *Chin. Phys.* **14** 49
[20] Li H C, Zhang J S and Xiao X C 2005 *Chin. Phys.* **14** 2181
[21] Zhang X D 2003 *Modern Signal Processing* (Second Edition) (Beijing: Tsinghua University Press) (in Chinese)