

A novel method to identify the scaling region for chaotic time series correlation dimension calculation

Ji Cui¹, ZHU Hua^{1*} & JIANG Wei²

¹ School of Mechanical & Electrical Engineering, China University of Mining and Technology, Xuzhou 221116, China;

² Institute of Medical and Biological Engineering, School of Mechanical Engineering, University of Leeds, Leeds LS2 9JT, UK

Received July 6, 2009; accepted September 2, 2010; published online January 17, 2011

To obtain more accurate correlation dimension estimations for chaotic time series, a novel scaling region identification method is developed. First, points that obviously do not belong to the scaling region associated with the whole double logarithm correlation integral curve are removed using the K-means algorithm. Second, a point-slope-error algorithm is developed to recognize a possible scaling region. Third, the K-means algorithm is used again to further remove a small interval of interfering points in the possible scaling region to obtain a more precise scaling region. The correlation dimension of four typical chaotic attractors and five curves generated by the Weierstrass-Mandelbrot fractal function were calculated using the proposed method. These calculated values were very close to the respective theoretical fractal dimensions. Moreover, the effectiveness of our method in identifying the scaling region was compared with existing methods. Results show that our method can distinguish the scaling region objectively, accurately, automatically and quickly, making estimations of the correlation dimension more precise and affording significant improvements in nonlinear analysis.

scaling region, correlation dimension, chaotic time series, attractor

Citation: Ji C C, Zhu H, Jiang W. A novel method to identify the scaling region for chaotic time series correlation dimension calculation. Chinese Sci Bull, 2011, 56: 925–932, doi: 10.1007/s11434-010-4180-6

Developments in chaos and fractal theories have offered new ideas in solving nonlinear problems. The correlation dimension D_2 is mainly used in chaotic time series processing and analysis. Together with the Lyapunov exponent and K-entropy, this parameter is one of the more critical characteristic parameters used to measure chaotic properties of nonlinear time sequences [1]. Based on the Phase-space Reconstruction [2–4] and Takens Embedding Theorem [5,6], the G-P algorithm for calculating the correlation dimension was proposed in 1983 by Grassberger and Procaccia [7,8]. This opened up a path for the study of chaotic time series appearing in practical applications.

The scaling region is a domain with measurement invariability in which the object exhibits self-similarity over a range of scales, a property that is characteristic of fractals.

When calculating the correlation dimension, the scaling region can be simply considered as the straight line portion of the entire double logarithm correlation integral curve, the slope of which defines the correlation dimension. Thus, the identification of the scaling region plays a vital role in determining the calculational accuracy of the correlation dimension.

However, the G-P algorithm provides no criteria that would lead to an identification of the scaling region. Many researchers have dedicated themselves to such a search [9]. A quick and easy estimate of the correlation dimension is afforded by visual inspection [10], namely choosing by eye a relatively straight portion from the whole curve. Obviously, this method relies heavily on human judgment and subjective criteria because different people can obtain different results for the same curve. Therefore, researchers seek a more systematic and objective solution for recognizing

*Corresponding author (email: hzhu@cumt.edu.cn)

ing the scaling region to obtain better correlation dimension estimates.

Various methods have been developed and these include the following. Yokoza et al. [11] proposed their largest linearity method which calculated the upper limit of the scaling region through an empirical formula. This method is short on providing objectivity and has some limitations in its operation. An alternative method using a three-segment polyline approximation [12] that aimed at minimizing the overall error was developed. It achieves the best approximation for the piecewise-linear fit of the curve with the slope of the middle polyline taken to be the correlation dimension. The computational cost of this approach is great and the method lacks a theoretical basis. Maragos et al. [13] also tried to calculate the upper limit of the scaling region using an empirical formula, which again lacks a certain objectivity. Hong et al. [14] thought that the compressibility of the logarithm data could weaken the differences between the scaling region and other regions, but suggested that the similar ratio of the image could be recovered by taking out the logarithmic curve. They presented the self-similar ratio algorithm which had a complicated computational process that is not useful in practical applications. A self-adaptive method [15] was proposed to give a more steady, reliable and systematic identification of the scaling region based on the notion of standard deviation. Meanwhile, the genetic algorithm was also used to reduce the computation time. This method offers more objectivity in identifying the scaling region, but has not been universally adopted. Dang et al. [16] proposed a method involving grouped recursive computer-recognition based on two so-called appraisal indices, confidence level and correlation. However, detailed techniques for choosing the initial scale value for the specific system are unavailable. Founded on clustering analysis theory, Yang et al. [17] presented a solution to identify automatically the non-scale region. Although this technique can deal more effectively with irregular correlation integral curves, there still exists some shortcomings; for instance, it cannot identify the scaling region accurately. Lai et al. [18] considered that a good scaling region did not only depend on the window length alone, but was associated with the embedding dimension m and delay time τ .

To improve the calculational accuracy of the correlation dimension, a novel method based on the K-means algorithm and the point-slope-error algorithm is studied to identify the scaling region. The outline of this method is as follows. First, a 2-means algorithm is performed to rule out points that clearly do not belong to the scaling region. Second, using the point-slope-error algorithm, a possible scaling region is recognized from the points remaining in the first step. Third, this 2-means algorithm is implemented again to identify the actual scaling region from the possible scaling region. Points from the actual scaling region are utilized in a least-square method to obtain an accurate estimate of the correlation dimension. The effectiveness of the proposed

new identification method is verified by calculating the correlation dimension of attractors from various well-known chaotic systems, including the Lorenz attractor and Hénon attractor, and the correlation dimension of signals generated from the Weierstrass-Mandelbrot (W-M) fractal function.

1 Correlation dimension and scaling region

The correlation dimension D_2 is a type of fractal dimension that helps to characterize fractals such as strange attractors of chaotic systems [19–21]. The fractal dimension is a parameter describing the space-filling degree of a fractal object, while the correlation dimension is an important characteristic parameter that quantitatively describes the “strangeness” of fractal attractors; it also provides the main criterion to distinguish a chaotic system from others.

The phase-space reconstruction technique should be used initially when calculating the correlation dimension with the G-P algorithm. The intention here is to recover the strange attractor of the chaotic system through mapping the 1-dimensional time series into a higher-dimensional phase space. Specifically, given a suitable embedding dimension $m \geq 2D+1$ (D is the dimension of the strange attractor) and time delay τ , the time series x_1, x_2, \dots, x_n will be restructured as a series of vectors in m -dimensional phase space:

$$X_i = [x_i, x_{i+\tau}, x_{i+2\tau}, \dots, x_{i+(m-1)\tau}], \quad i=1, 2, \dots, N, \quad (1)$$

where X_i is the reconstructed phase space vector, and N is the number of vectors given by $N = n - (m-1) \times \tau$.

The Euclidean distances between each vector point X_i and the other $(N-1)$ points are calculated in m -dimensional phase space. Then, the frequency of each distance falling within the volume element with hyperspherical radius r is obtained. Thus, the correlation integral function $C_m(r)$ is established, as shown in eq. (2).

$$C_m(r) = \sum_{i \neq j} H(r - \|X_i - X_j\|) / N(N-1), \quad (2)$$

where $H(\cdot)$ is the Heaviside function defined as $H(x)=0$ if $x \leq 0$ and $H(x)=1$ if $x > 0$.

Set r_{\min} and r_{\max} as the respective minimum and maximum distance between points. If the radius $r \leq r_{\min}$, none of points will fall in the volume element and $C_m(r)=0$; in contradistinction, if $r \geq r_{\max}$, all points will fall in the volume element and $C_m(r)=1$. It can be seen that $C_m(r)$ is a cumulative distribution function reflecting the distribution probability of the distance less than r between pairs of points on the chaotic attractor. The cumulative correlation function is related to the correlation dimension D_2 by the power law

$$C_m(r) \propto (r)^{D_2}, \quad (r \rightarrow 0). \quad (3)$$

This results in D_2 being defined as

$$D_2 = \lim_{r \rightarrow 0} d \ln C_m(r) / d \ln r. \quad (4)$$

For deterministic systems, graphs of $\ln C_m(r)$ against $\ln r$ should be plotted first by increasing the value of r until $\ln C_m(r)$ does not vary with $\ln r$. At this stage, the slope of the $\ln r - \ln C_m(r)$ curve over the linear region can be calculated by the least-square method, giving a numerical estimation of the correlation dimension.

The above-mentioned linear region is considered to be the scaling region. Visual inspection is generally adopted to identify the scaling region as the process is simple and fast albeit subjective and haphazard. For the same curve, the scale or the upper/lower limit of the scaling region as determined by different people can differ, making a variation in calculation results of the correlation dimension. Therefore, an accurate and objective identification of the scaling region is essential and preferable. Figure 1 displays an example of the scaling region as judged by the naked eye.

In addition, the selection of the embedding dimension m and delay time τ is significant in estimations of D_2 . However it is difficult to predetermine these two parameters. In this study, the embedding dimension m begins from value 3 and proceeds successively higher dimensions until the estimated value of D_2 stabilizes or fluctuates within a small error range (the relative error $r_D = (D_2(m_2) - D_2(m_1)) / D_2(m_1) \leq 5\%$ is a reasonable range) [22]. It can be seen from Figure 1(a) that by increasing m two linear regions of the adjacent $\ln r - \ln C_m(r)$ curves begin to align more closely which means that D_2 begins to converge gradually. The time delay τ is calculated by the autocorrelation function which is the more widely used and established method.

2 The novel scaling region identification method

The K-means algorithm, first introduced by MacQueen [23] and then developed by Hartigan and Wong [24], is a classic clustering algorithm based on distance. The aim of the K-means algorithm is to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. The algorithm comprises the following

steps: (1) Place k points into the space represented by the objects that are being clustered. These points represent initial group centroids. (2) Assign each object to the group that has the closest centroid. (3) Recalculate the positions of the centroids. Repeat steps 2 and 3 until the centroids no longer move. This algorithm produces a separation of the objects into groups from which the metric that is to be minimized can be calculated.

The first derivative curve $\ln r - \ln C_m(r)'$ curve is composed of points presenting slopes between neighboring points of the $\ln r - \ln C_m(r)$ curve. The scaling region in this curve is a linear region, and the slope of a line is a constant. Thus, points $(\ln r, \ln C_m(r)')$ in the scaling region fluctuate slightly around a constant value, while points outside the scaling region are accompanied by larger fluctuations as illustrated in Figure 1(b). Taking this fact and the feature of the K-means algorithm into consideration, points on the $\ln r - \ln C_m(r)$ curve can be classified according to whether they lie in the scaling region. The distribution of these two kinds of points in the $\ln r - \ln C_m(r)$ curve is crossed, which divides the curve into several intervals. The interval containing the most points is retained, the others are ignored.

Note that the new method states that all $\ln r - \ln C_m(r)$ curves are drawn from r_{\min} to r_{\max} with incremental step d . In this case, the number of points in a curve is $n = \text{fix}((r_{\max} - r_{\min})/d)$. The curve satisfies $\ln C_m(r) = -\infty$ if $r \leq r_{\min}$ and $\ln C_m(r) = 0$ if $r \geq r_{\max}$.

As an illustration, consider the $\ln r - \ln C_m(r)$ curve of the Hénon attractor with an embedding dimension of $m=10$; the 2-means algorithm was utilized to recognize the scaling region. It can be seen that the $\ln r - \ln C_m(r)$ curve separates into several regions by two kinds of points. The region, shown in Figure 2(a) between two vertical dashed lines, corresponds to the larger cluster (including the most consecutive points) and was retained, while other regions (shown in scaled-up view in Figure 2(b)) were extracted. This latter region is short in appearance compared with the former region but includes quite a number of points. This is due to the fact that the logarithm coordinate compresses

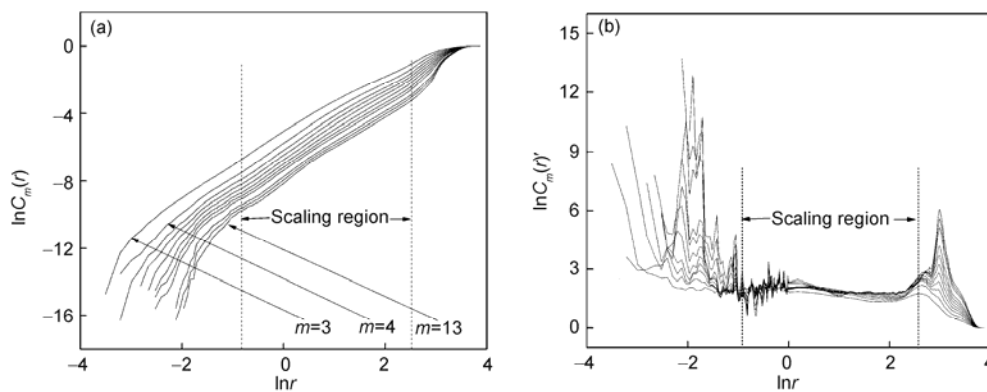


Figure 1 One example of the computation of the correlation dimension for a chaotic time series of length $N=5000$ with m ranging from 3 to 13. (a) $\ln C_m(r)$ versus $\ln r$; (b) $\ln C_m(r)'$ versus $\ln r$. The region between the two dashed lines recognized by naked eyes is considered as the scaling region.

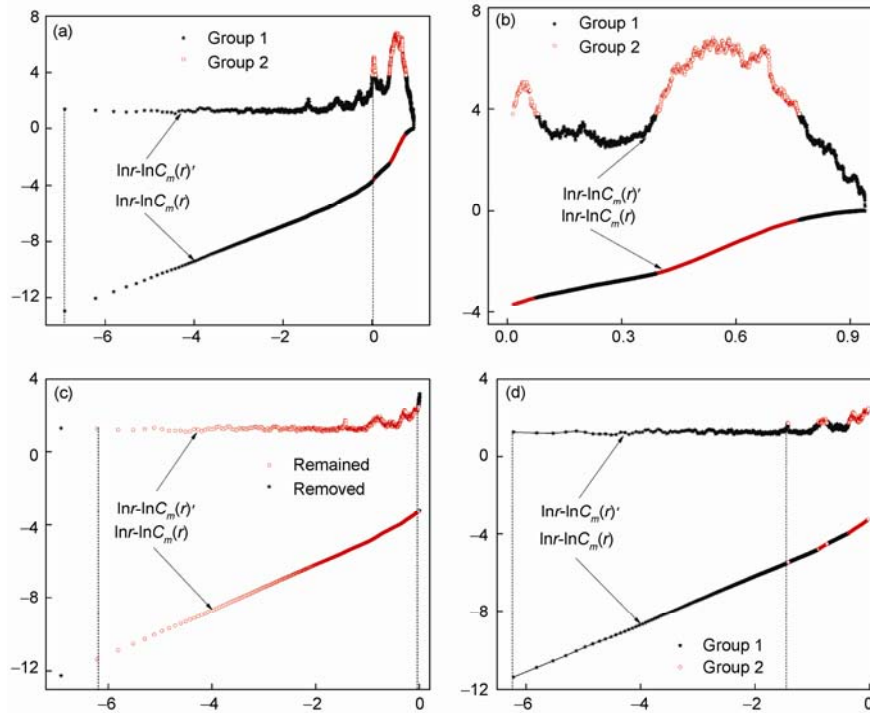


Figure 2 Scaling region identification of the $\ln r - \ln C_m(r)$ and $\ln r - \ln C_m(r)'$ curves for the Hénon attractor with embedding dimension $m = 10$ using the proposed method. (a) The first step with 2-means algorithm; (b) the extracted region by 2-means algorithm; (c) the second step with point-slope-error algorithm; (d) the third step with 2-means algorithm again and the region between two dashed lines considered as the actual scaling region and yielding the accurate correlation dimension D_2 .

scale making the x -coordinate uniformly-spaced points (r , $C_m(r)$) appear densely-distributed as $\ln r$ goes backward but becomes sparser as $\ln r$ goes forward. Therefore, such points will impart a negative bias on the correlation dimension calculation if these were not removed. A little fluctuation is seen still to remain in the retained portion of the $\ln r - \ln C_m(r)'$ curve in Figure 2(a). This signifies that the actual scaling region cannot be completely identified by solely the K-means algorithm.

This letter presents a new identification method for situations where the scaling region as recognized only by the K-means algorithm is non-ideal. The fact that the points in the $\ln r - \ln C_m(r)'$ curve represent slopes between adjacent points in the $\ln r - \ln C_m(r)$ curve can be considered as a body of measurements to be used to estimate the correlation dimension. We intend to identify the scaling region based on a comprehensive application of the K-means and principles used in statistics. First, the 2-means algorithm is used to classify points over the scale $[\ln r_{\min}, \ln r_{\max}]$. The “parasitic error” of the measured value can be eliminated from this step, as seen clearly in Figure 2(a) and (b). Second, the point-slope-error algorithm is developed. The arithmetic mean value S_m and standard deviation σ of the points in the $\ln r - \ln C_m(r)'$ curve retained from the former step are computed, and the error range is set at $(S_m - k\sigma, S_m + k\sigma)$, where k is the control parameter for the error range and can be calculated from the formula $k = \text{fix}(2(S_{\max} - S_{\min})/\sigma)$. Points in the error range are not necessarily consecutive; therefore, a scale region containing the most points is retained as a pos-

sible scaling region. In addition, based on numerous experimental data, the scaling region for practical time series is sometimes relatively restricted, in which case, the error range can be suitably widened. Figure 2(c) demonstrates the point that the point-slope-error algorithm recognizes a region considered as the possible scaling region (the red lines) from the portion between the two vertical dashed lines in Figure 2(a), as well as that the lines within this region fluctuate gently. However, this scaling region is not sufficiently precise because there are some points that still do not belong to the actual scaling region. Third, the 2-means algorithm is applied again to extract that part corresponding to interfering points to get a more precise scaling region shown between the two dashed lines in Figure 2(d). It is this portion from which points of the $\ln r - \ln C_m(r)$ curve are used to accurately estimate D_2 using the least-square method.

The above identification process was programmed using Matlab software to automatically recognize the scaling region. The proposed method provides an accurate, effective, automatic and rapid identification of the scaling region that takes only a few seconds to reach 5000 points.

3 Verification of the method

3.1 Classical chaotic signals

Chaotic signals are very sensitive to initial conditions and have random transformation characteristics, the long-term

behavior of which cannot be forecaste [25,26]. In this section, the well known Lorenz attractor (Figure 3(a)) and the Hénon attractor (Figure 3(b)) are chosen mainly to verify the effectiveness of the proposed method. Grassberger and Procaccia [7,8] estimated the Hausdorff dimension to be 2.06 ± 0.01 and the correlation dimension to be 2.05 ± 0.01 for the Lorenz attractor, and a correlation dimension of 1.25 ± 0.02 for the Hénon map. Russel [27] estimated the Hausdorff dimension to be 1.261 ± 0.003 for the Hénon map.

The Lorenz system, named after Lorenz in 1963, is a three-dimensional autonomous system of differential equations:

$$\dot{x} = \sigma(-x + y), \dot{y} = -xz + rx - y, \dot{z} = xy - bz, \quad (5)$$

where σ , r and b are positive real parameters, which were developed to demonstrate the unpredictable behavior of weather. For a typical set of parameter values ($\sigma=10$, $r=8/3$, and $b=28$), the system is chaotic and unpredictable but not completely random as evidenced by the well known butterfly pattern called the Lorenz attractor displayed in Figure 3(a). The system was solved by the Runge-Kutta method with a time step of 0.01s and initial values $x_0=12$, $y_0=2$, $z_0=9$. A 50000×3 matrix was obtained by taking a sampling interval of 1s over a total sampling time length of 500s. The transient data were removed and a time series including 10000 data points of y variables was taken for the study.

The Hénon map, introduced by Michel Hénon in 1976, is a discrete-time dynamical system defined by the equations:

$$x_{n+1} = y_n + 1 - ax_n^2, y_{n+1} = bx_n. \quad (6)$$

The map depends on two parameters, a and b , and for the particular values $a = 1.4$ and $b = 0.3$, the Hénon map is chaotic. With randomly generated initial values of x_0 and y_0 , 15000 iterates of Hénon map were obtained to plot the Hénon attractor shown in Figure 3(b). Here, the transient data were also removed and a time series including 10000 points for the y variable was taken to identify the scaling region and to perform the correlation dimension calculation.

For the Lorenz attractor, the embedding dimension m ranged from 3 to 12 and the time delay τ was 14 computed by the autocorrelation function, while for the Hénon map m ranged from 3 to 14 and τ was 1. Figure 4(a) and (b) show

the $\ln r - \ln C_m(r)$ curve and $\ln r - \ln C_m(r)'$ curve respectively for the Lorenz attractor; Figure 4(c) and (d) show the $\ln r - \ln C_m(r)$ curve and $\ln r - \ln C_m(r)'$ curve for the Hénon map. It can be seen from Figure 4(a) and (c) that the curves in the two correlation integral logarithm plots become more aligned and the distance between each successive lines become narrower in the scaling regions with increasing embedding dimension m , illustrating the correlation dimension obtained for the scaling region tend to a single value for sufficiently large m .

The scaling region identification and correlation dimension computation were conducted for both attractors. Figure 5(a) and (b) demonstrate a sequence of $D_2(m)$ obtained from slopes of the linear scaling regions in the double logarithm correlation integral plots, and a sequence of relative errors $r_D(m)$ by increasing m for the corresponding attractors. From these figures, we observe that $D_2(m)$ tends to a constant with the exception of minor fluctuations within a small error bound as m increases. Estimated correlation dimensions computed based on our new scaling region identification method and theoretical correlation dimensions, as well as Hausdorff dimensions for the Lorenz and Hénon maps, are listed in Table 1.

In addition, correlation dimensions of the Rössler attractor and logistic map were also estimated to further verify the effectiveness of the proposed method. The Rössler system is described by three coupled non-linear differential equations (eq. (7)). Given parameters $a=0.2$, $b=0.2$, $c=5.7$ and initial values $x_0=-1$, $y_0=0$, $z_0=1$, a time series of 10000 points for the y variable was obtained by the same data processing technique for the Lorenz system. The correlation dimension of Rössler attractor as estimated in this paper is compared with reported fractal dimensions, which have been listed in Table 1. The logistic map is known for its modeling of how bug populations achieve dynamic equilibrium, as well as being a very simple model yielding chaotic dynamics (eq. (8)). With parameter values of $\lambda=3.6$ and $\lambda=3.9$, where the system is extremely chaotic, calculations of the correlation dimension for the logistic map were performed from their respective time series with lengths of 10000; these were also compared with reported correlation dimensions and are shown in Table 1.

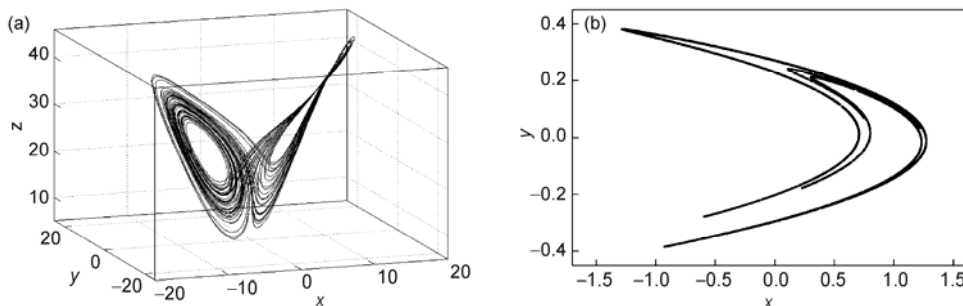


Figure 3 Attractor plots for (a) Lorenz attractor in the phase space of the variables (x, y, z) and (b) Hénon attractor in the phase space of the variables (x, y).

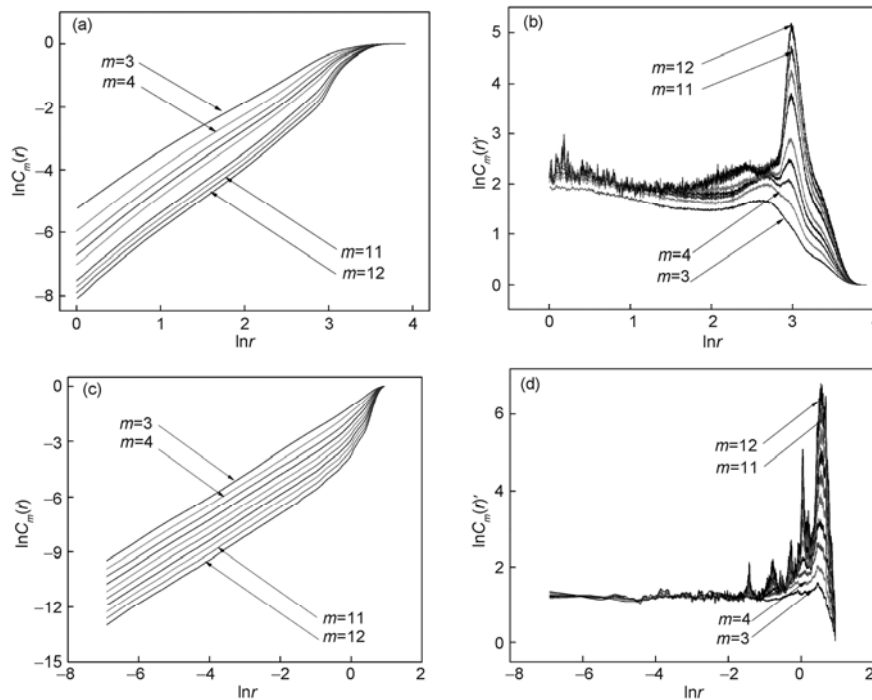


Figure 4 Correlation integral $C(r)$ as a function of sphere radius r with embedding dimension m varying from 3 to 12; (a) $\ln C_m(r)$ versus $\ln r$ for the Lorenz attractor; (b) $\ln C_m(r)'$ versus $\ln r$ for the Lorenz attractor; (c) $\ln C_m(r)$ versus $\ln r$ for the Hénon attractor; (d) $\ln C_m(r)'$ versus $\ln r$ for the Hénon attractor.

$$\dot{x} = -(y + z), \dot{y} = x + ay, \dot{z} = b + z(x - c). \tag{7}$$

$$x_{n+1} = \lambda x_n (1 - x_n). \tag{8}$$

From Table 1, it is clear that the correlation dimensions of the attractors stabilize in acceptable error ranges after the embedding dimension reaches a certain value. The estimated correlation dimension based on the proposed scaling region identification method is sufficiently close to theoretical values or other reported fractal dimensions, thereby illustrating the reasonable effectiveness of the new scaling region identification method.

3.2 Weierstrass-Mandelbrot (W-M) fractal function curves

The W-M function, which is continuous everywhere but differentiable nowhere, is a well known fractal function for

which the Hausdorff-Besicovitch dimension exceeds unity [30]. The function has been widely applied in the study of those profiles that appear to have self-affinity and self-similarity. It can be expressed in the following form

$$X(t) = G^{(D-1)} \sum_{n=n_l}^{\infty} r^{-(2-D)n} \cos(2\pi r^n t), \tag{9}$$

where D is the fractal dimension of the profile and $1 < D < 2$; G is the scaling constant; r is a constant with $r > 1$ ($r = 1.5$ is suitable and practicable for general fractal cases); r^n is the frequency mode corresponding to the reciprocal of the wavelength λ ; L is the profile length; n_l corresponds to the low cut-off frequency of the profile under measurement and $r^{n_l} = 1/L$.

The correlation dimension D_2 of the fractal curve generated by the W-M function was calculated using the G-P algorithm and compared with the corresponding given fractal dimension D as a further test of the validity of the proposed method. Fractal curves with $D = 1.1, 1.3, 1.5, 1.7$ and

Table 1 Comparison between experimental D_2 values obtained by the proposed new method and theoretical D_2 as well as other fractal dimensions for the Lorenz attractor, Hénon map, Rössler attractor and logistic map

Attractor	Experimental D_2				Theoretical D_2	Other reported fractal dimension D
	$m=9$	$m=10$	$m=11$	$m=12$		
Lorenz	2.046	2.056	2.064	2.064	2.05 ± 0.01 [7]	2.06 ± 0.01 [7] (Hausdorff dimension)
Hénon	1.233	1.241	1.241	1.240	1.25 ± 0.02 [8]	1.261 ± 0.003 [27] (Hausdorff dimension)
Rössler	2.009	2.070	2.072	2.071		2.01–2.02 [28]
Logistic	$\lambda = 3.6$	1.043	1.046	1.037	1.042	0.98 [29]
	$\lambda = 3.9$	0.982	0.979	0.976	0.980	0.99 [29]

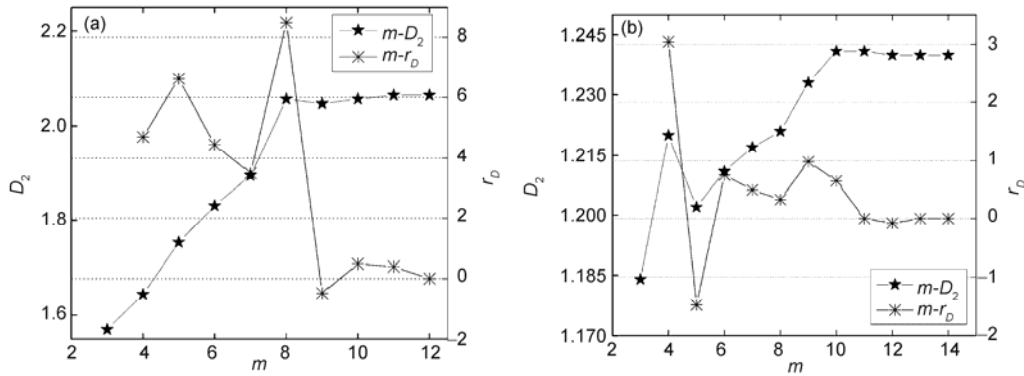


Figure 5 Correlation dimension $D_2(m)$ and relative error $r_D(m)$ as functions of embedding dimension m for (a) Lorenz attractor and (b) Hénon attractor.

1.9 after 10000 iterations were generated. Due to space limitations, Figure 6 illustrates three approximated curves with $D=1.1$, $D=1.5$ and $D=1.9$, from which it can be seen that the curves become more and more complicated as D increases because the fractal dimension reflects the degree of variational complexity of the profile. The results of the correlation dimension computation and five profiles given set fractal dimensions are listed in Table 2, along with their absolute errors. It is not difficult to discern that the two dimensions are very close. Thus, the scaling region can be recognized objectively and accurately using the proposed method to derive more precise values of the fractal dimension. This further supports the view that the proposed method is effective.

3.3 Comparisons with existing methods

Two commonly-used scaling region identification methods, viz. the visual inspection method and the K-means algorithm method, were chosen to illustrate the accuracy and objectivity

Table 2 Comparison between the estimated D_2 value obtained by the proposed new method and the given fractal dimension D as well as the absolute error between the two dimensions

Given fractal dimension D	Estimated D_2	Absolute error
1.1	1.125	0.025
1.3	1.324	0.024
1.5	1.498	-0.002
1.7	1.706	0.006
1.9	1.900	0.000

of the present methodology. The correlation dimensions of the Lorenz and Hénon attractors with embedding dimension $m=12$ were calculated based on the above three methods. Using the visual inspection method, five volunteers were invited to scrutinize by eye the double logarithmic correlation integral curve and select a region with good linearity. The least square method was applied to fit the selected five regions to obtain estimated values of the correlation dimension, the average value of which was considered as the actual correlation dimension. Table 3 demonstrates the calculation

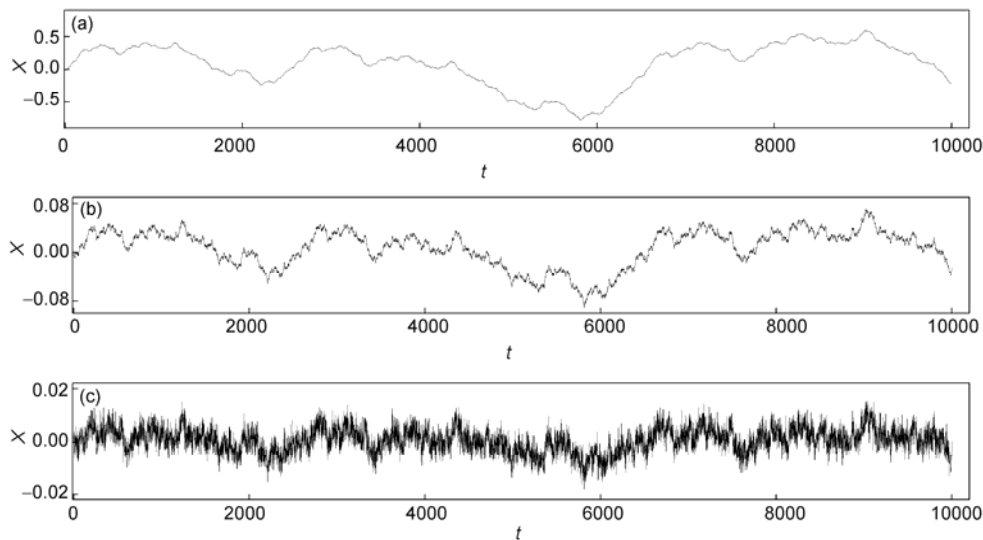


Figure 6 Examples of the Weierstrass-Mandelbrot function (eq. (9)), each graph displays 10000 points of a Weierstrass-Mandelbrot function with $r=1.5$, $G=10^{-3}$ and $L=1$. t presents the sequence number of the points. The given fractal dimensions are (a) $D=1.1$; (b) $D=1.5$ and (c) $D=1.9$.

Table 3 Comparison between correlation dimensions calculated from three scaling region identification methods

Attractor	Visual inspection method	K-means algorithm method	Proposed new method
Lorenz	2.100±0.061	2.381	2.064
Hénon	1.334±0.056	1.407	1.240

results of the three methods.

Clearly, as compared with the theoretical fractal dimension in Table 1, the new method affords higher precision than either the visual inspection method or the K-means algorithm method.

4 Discussion

(1) A new method for scaling region identification was proposed by formulating a point-slope-error algorithm and combining it with the K-means algorithm. This can recognize the scaling region objectively and accurately making the estimation of the correlation dimension more precise.

(2) Attractors from four well known chaotic systems as well as five fractal curves generated from the W-M function were chosen to estimate their correlation dimensions using our method. The results of the correlation dimension computation were close to theoretical fractal dimensions, thereby verifying the effectiveness of the proposed method.

(3) By comparing the correlation dimension calculation of two of the well-known attractors, the calculation accuracy of the new method was found to be superior to visual inspection or K-means algorithm methods.

(4) The new scaling region identification method was easily implemented and performed by a computer; the procedure recognized the scaling region objectively, accurately, automatically and quickly. This development has important significance for chaotic analyses within nonlinear time series studies.

This work was supported by the National Natural Science Foundation of China (50975276 and 50475164) and the Ph.D. Programs Foundation of Ministry of Education of China (200802900513).

- 1 Yang X D, He A J, Zhou Y, et al. Multifractal mass exponent spectrum of complex physiological time series. *Chinese Sci Bull*, 2010, 55: 1996–2003
- 2 Gibson J F, Farmer J D, Casdagli M, et al. An analytic approach to practical state space reconstruction. *Phys D*, 1992, 57: 1–30
- 3 Kennel M B, Brown R, Abarbanel H D I. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys Rev A*, 1992, 45: 3403–3411
- 4 King G P, Stewart I. Phase space reconstruction for symmetric dynamical systems. *Phys D*, 1992, 58: 216–228

- 5 Takens F. Detecting strange attractors in turbulence. *Lect N Math*, 1981, 898: 366–381
- 6 Packard N, Crutchfield J, Farmer D, et al. Geometry from a time series. *Phys Rev Lett*, 1980, 45: 712–716
- 7 Grassberger P, Procaccia I. Characterization of strange attractors. *Phys Rev Lett*, 1983, 50: 346–349
- 8 Grassberger P, Procaccia I. Measuring the strangeness of strange attractors. *Phys D*, 1983, 9: 189–208
- 9 Tang G J, Du B Q, Wang S L. Scaleless band automatic identification for fractal fault diagnosis of rotor system (in Chinese). *J Power Eng*, 2009, 29: 440–444
- 10 Liu J Q. Chaotic phenomena in observation series of water consumption (in Chinese). *J Zhejiang Univ*, 2004, 31: 236–240
- 11 Yokoya N, Yamamoto K, Funakubo N. Fractal-based analysis and interpolation of 3D natural surfaces and their applications to terrain modeling. *CVGIP*, 1989, 46: 284–302
- 12 Wang F Q, Luo C. An improvement of G-P algorithm and its application. *Chin J Comput Phys*, 1993, 10: 345–351
- 13 Maragos P, Sun F K. Measuring the fractal dimension of signals. *IEEE Trans SP*, 1993, 41: 108–121
- 14 Hong S Z, Hong S M. A new method to determine scaling range of fractals self-similar ratio algorithm(in Chinese). *Dis Nat*, 1993, 12: 53–57
- 15 Fei B, Jiang Z D, Wang H R. Genetic algorithm for determining scale independent fractals associated with grinding surface(in Chinese). *J Xi'an Jiaotong Univ*, 1998, 32: 72–75
- 16 Dang J W, Shi Y, Huang J G. The identification of fractal scaleless band in the study of fractal with computers (in Chinese). *Comput Eng Apps*, 2003, 39: 25–27
- 17 Yang H Y, Ye H, Wang G Z, et al. Identification of scaling regime in chaotic correlation dimension calculation. In: 3rd IEEE Conference on Industrial Electronics and Applications, ICIEA.2008. 1383–1387
- 18 Lai Y C, Lerner D. Effective scaling regime for computing the correlation dimension from chaotic time series. *Phys D*, 1998, 115: 1–18
- 19 Wang Z Z, Ning X B, Zhang Y, et al. Distribution of correlation dimensions of synchronous 12-lead ECG signals. *Chinese Sci Bull*, 2000, 45: 1628–1632
- 20 Zhao D H, Ruan J, Cai Z J. Combination prediction method of chaotic time series. *Chinese Sci Bull*, 2006, 52: 570–573
- 21 Wang X Y. Relation of chaos activity characteristics of the cardiac system with the evolution of species. *Chinese Sci Bull*, 2002, 47: 2042–2048
- 22 Buzug T, Pfister G. Optimal delay time and embedding dimension for delay-time coordinate by analysis of the global static and local dynamical behavior of strange attractors. *Phys Rev A*, 1992, 45: 7073–7084
- 23 MacQueen J. Some methods for classification and analysis of multivariate observations. In: *Proc 5th Berkeley Symp. Mathemat Statist Prob I*, 1967, 281–297
- 24 Hartigan J A, Wong M A. A K-Means clustering algorithm. *Appl Stat*, 1979, 28: 100–108
- 25 Yan S L. Enhancement of chaotic carrier bandwidth in a semiconductor laser transmitter using self-phase modulation in an optical fiber external round cavity. *Chinese Sci Bull*, 2010, 55: 1007–1012
- 26 Yan S L. Control of chaos in an external-cavity multi-quantum-well laser subjected to dual-wedges and optical dual-feedback. *Chinese Sci Bull*, 2009, 54: 1158–1163
- 27 Russel D A, Hanson J D, Ott E. Dimension of strange attractors. *Phys Rev Lett*, 1980, 45: 1175
- 28 Peitgen H, Jurgens H, Saupe D. *Chaos and Fractals: New Frontiers of Science*. 2nd ed. New York: Springer-Verlag, 2004
- 29 Sprott J C. *Chaos and Time-Series Analysis*. Oxford: Oxford University Press, 2003
- 30 Liao X, Wang G. Employing fractals and FEM for detailed variation analysis of non-rigid assemblies. *Int J Mach*, 2005, 45: 445–454

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.