Direct dynamical test for deterministic chaos and optimal embedding of a chaotic time series

Jianbo Gao and Zhemin Zheng
Laboratory for Nonlinear Mechanics of Continuous Media, Institute of Mechanics, Chinese Academy of Sciences, Beijing, 100080, People's Republic of China
(Received 7 September 1993)

We propose here a local exponential divergence plot which is capable of providing an alternative means of characterizing a complex time series. The suggested plot defines a time-dependent exponent and a "plus" exponent. Based on their changes with the embedding dimension and delay time, a criterion for estimating simultaneously the minimal acceptable embedding dimension, the proper delay time, and the largest Lyapunov exponent has been obtained. When redefining the time-dependent exponent $\Lambda(k)$ curves on a series of shells, we have found that whether a linear envelope to the $\Lambda(k)$ curves exists can serve as a direct dynamical method of distinguishing chaos from noise.

PACS number(s): 02.50.−r, 05.45.+b, 02.70.−c

I. INTRODUCTION

Complex time series are ubiquitous in nature and in man-made systems, and a variety of measures have been proposed to characterize them. Among the most widely used approaches today are state space reconstruction by the time delay embedding [1,2], calculation of the correlation dimension and of the K_2 entropy [3,4], and estimation of the Lyapunov exponents [5–7], for characterizing strange attractors. Since low-dimensional strange attractors produce a small and usually noninteger value of the dimension and a converging entropy, and a positive largest Lyapunov exponent, in practice these have often been taken as "proof" of the presence of a strange attractor. However, there exist some stochastic processes which generate time series with finite correlation dimension and converging K_2 entropy estimates [8–11]. These signify that under certain circumstances one may not understand very much by routine calculations. Therefore it is very important to develop a method for distinguishing chaos from noise in an observed time series and develop a means of characterizing chaotic time series to gain an insight into the system under investigation.

In this paper we present a means of characterizing chaotic time series. We define a local exponential divergence plot [12] which enables one to view the dynamics on an attractor constructed from a time series. If the time series is indeed chaotic, then the plot provides a criterion for the selection of the minimal acceptable embedding dimension and an optimal delay time. When the unstable motion on the chaotic attractor only is extracted, a proper estimation of the largest positive Lyapunov exponent can also be obtained. The unstable motion on the attractor enables one to develop a direct dynamical method of distinguishing low-dimensional deterministic chaos from stochastic processes.

II. LOCAL EXPONENTIAL DIVERGENCE PLOT AND OPTIMAL EMBEDDING

Assume we have a time series $x_1, x_2, \ldots$, with sampling time δt and construct vectors of this form: $X_i=(x_i, x_{i+L}, \ldots, x_{i+(m-1)L})$, with m the embedding dimension and L the delay time. Hence a dynamics F: $X_i \rightarrow X_{i+1}$ is defined, which is assumed to be representative of the original system. The distance between X_i and X_j, $\|X_i-X_j\|$, is mapped to $\|X_{i+k}-X_{j+k}\|$ after k iterations of F. The local exponential divergence plot is defined by plotting $\ln(\|X_{i+k}-X_{j+k}\|/\|X_i-X_j\|)$ vs $\ln(\|X_i-X_j\|)$ when $\|X_i-X_j\|$ is smaller than a prescribed small distance r^*. As typically done, we assume that most of these sufficiently small distances $\|X_i-X_j\|$ can be regarded as distances between orbits, then if the motion is truly chaotic, points with $\|X_{i+k}-X_{j+k}\| > \|X_i-X_j\|$ will dominate and lie above the zero level line in the plot.

Figures 1(a)–1(e) show divergence plots with different m and L for the Rössler attractor constructed from the x component of the flow. For comparison, a divergence plot constructed from the original (x,y,z) system is also plotted in Fig. 1(f). The zero level line is added to Figs. 1(c), 1(e), and 1(f) for a clear view of the divergence-dominated behavior. We notice that Fig. 1(c) is representative of that constructed from the original system [Fig. 1(f)] and does not change much when m is further increased [Fig. 1(e)]. We will show below that the difference between these plots gives a hint to proper embedding, and $m=3$, $L=8$ correspond to optimal parameter values.

A problem of significant practical importance is to determine the minimum acceptable embedding dimension m_e. A basic idea is that in the passage from dimension m to $m+1$ one can differentiate between points on the orbit $X(n)$ that are "true" neighbors and points which are "false" neighbors—points which appear to be neighbors because the orbit is being viewed in too small an embedding space. Based on this basic idea, several methods are now available [13–16], which differ in implementations either by way of graphic display or by defining some appropriate statistical quantity. When the embedding dimension is increased from m_e-1 to m_e, the structure of the graphic representation and/or the value of the statistical quantity will undergo a radical change while further increasing m causes little change. Our divergence plot implements this basic idea dynamically. When the

embedding space is too small, the ill-defined dynamics F and the false neighbors will generate many points of $\ln(\|X_{i+k}-X_{j+k}\|/\|X_i-X_j\|)$ with excessively large positive values in the plot. This is clearly shown by the difference between Figs. 1(a) and 2(c).

Figure 1 also points out how to select a proper delay time L. Dynamically, when L is either too small or too large the dynamics F will not be very well defined, in the sense that excessively large values of $\ln(\|X_{i+k}-X_{j+k}\|/\|X_i-X_j\|)$ frequently appear. The key to selecting a proper delay time L is that the orbital motion should be as uniform as possible, and distortion be small. This can be achieved by requiring that the number of points $\ln(\|X_{i+k}-X_{j+k}\|/\|X_i-X_j\|)$ with excessively large positive values in the plot be as small as possible and the structure of the plot be as compact as possible. This is the reason that the structure of Fig. 1(c) is preferred to that of Fig. 1(b) or 1(d).

For a quantitative description, we define the time-dependent exponent Λ by

$$\Lambda(k,m,L)=\langle \ln(\|X_{i+k}-X_{j+k}\|/\|X_i-X_j\|)\rangle , \quad (1)$$

with $\|X_i-X_j\|\le r^*$. The evolution time corresponding to k is $k\delta t$ and the angle brackets denote ensemble average of all possible pairs of (X_i,X_j). Since we are more concerned with points $\ln(\|X_{i+k}-X_{j+k}\|/\|X_i-X_j\|)$ with excessively large positive values, we also define the "plus" exponent Λ_+ by

$$\Lambda_+(k,m,L)=\langle \ln(\|X_{i+k}-X_{j+k}\|/\|X_i-X_j\|)_+\rangle , \quad (2)$$

where + simply denotes that points with positive values of $\ln(\|X_{i+k}-X_{j+k}\|/\|X_i-X_j\|)$ only are averaged. In the following discussion, when only one variable is considered, we will simply write $\Lambda(k)$, $\Lambda(L)$, and $\Lambda_+(L)$ for convenience.

Now the problem of properly reconstructing the state space can be stated as follows. It is required that F be a continuous mapping preserving neighborhood relations. The minimal acceptable embedding dimension m_e is determined by requiring that the structure of the divergence plot no longer changes radically, or equivalently, that Λ and Λ_+ do not decrease significantly by further increasing m. When m is thus selected, for a series of L, the minima of $\Lambda_+(L)$ and $\Lambda(L)$ determine an optimal delay time.

The physical significance of the quantity Λ is obvious. When the evolution time $k\delta t$ is very small, $\Lambda/k\delta t$ is the mean value of the largest local Lyapunov exponent [17,18]. After several iterations, the separation vector between X_i and X_j will align with the eigendirection for the largest positive Lyapunov exponent, and $\Lambda/k\delta t$ is equivalent to the standard estimation of this exponent. Hence when the proper reconstruction of the state space has been achieved, we can expect that the $\Lambda(k)$ curve for $k\delta t$ not very small will be a straight line which passes through the origin when extrapolated. This should be, otherwise, different values of the largest Lyapunov ex-

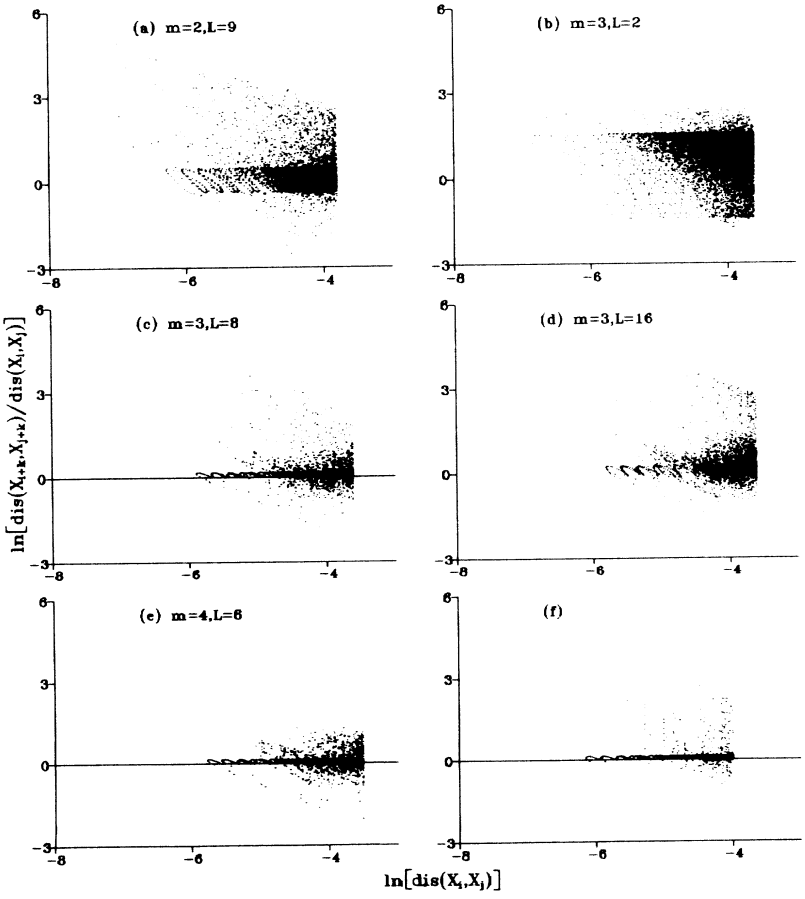

FIG. 1. Local exponential divergence plots for the Rössler attractor, $k=9$. $\|X_i-X_j\|$ has been rewritten as $\mathrm{dis}(X_i-X_j)$.

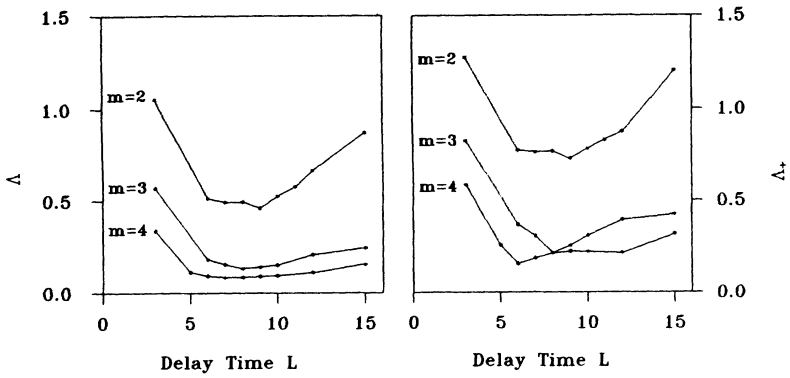

FIG. 2. The $\Lambda(L)$ and $\Lambda_+(L)$ curves with different m for the Rössler attractor, $k=9$.

ponent will be obtained with different selections of k. The largest Lyapunov exponent can be objectively estimated by the slope of the linear $\Lambda(k)$ curve, i.e., $[\Lambda(k_1)-\Lambda(k_2)]/(k_1-k_2)\delta t$, with $k_1,k_2\in(k_{\min},k_{\max})$.

Let us continue to discuss the Rössler system. Figure 2 shows the $\Lambda(L)$ and $\Lambda_+(L)$ curves, and we see that the combination of $m=3$ and $L=8$ is an optimal choice. Note that improper embedding (underestimated m or improper L) always results in overestimated positive exponent. Figure 3 gives the $\Lambda(k)$ curves for different m and L. We see that when $m=3$, the $\Lambda(k)$ curve for $L=8$ shows a clear linear dependence, while curves for the smaller value of $L=6$ or the larger value of $L=12$ (approximately the optimal value suggested by [19]) are less satisfactory. Figure 3 also shows two curves for $m=4$, $L=6$ and $m=8$, $L=4$, with improved linearity. Note that the degree of the linearity is "saturated" when m is increased to 4, while the minimum acceptable value of m is 3. The estimated value of the largest Lyapunov exponent from the slope of the $\Lambda(k)$ curve of $m=3$ and

TABLE I. Simultaneous minimal embedding dimension m and optical delay time L, and the largest Lyapunov exponent λ. The total number of points N is given in the table for different model systems.

Systems	Optimal embedding parameters: Others	Optimal embedding parameters: Ours ($N=2000$)	λ value: Others	λ value: Ours
Hénon [20] $(a=1.4,b=0.3)$ $X_{n+1}=1-aX_n^2+Y_n$ $Y_{n+1}=bX_n$	$m=2$ [14] $N=5000$	$m=2$, $L=1$	0.418 [5]	0.421 ±0.003 $N=2000$
Rössler [21] $(\delta t=\pi/25,a=0.15,$ $b=0.20,c=10.0)$ $\dot{X}=-(Y+Z)$ $\dot{Y}=X+aY$ $\dot{Z}=b+Z(X-c)$	$m=3,L=7$ [15] $N=10000$	$m=3$, $L=8$	0.09 [5]	0.067 ±0.006 $N=2000$
Lorenz [22] $(\delta t=0.03,\sigma=10,$ $b=8/3,r=45.92)$ $\dot{X}=\sigma(Y-X)$ $\dot{Y}=X(r-Z)-Y$ $\dot{Z}=XY-bZ$		$m=3$, $L=3$	1.497 [5]	1.48 ±0.03 $N=3000$
Mackey-Glass [23] $(\delta t=1.5,a=0.2,$ $b=0.1,c=10,\Gamma=30)$ $\dot{X}=\frac{aX(t+\Gamma)}{1+[X(t+\Gamma)]^c}-bX(t)$	$m=4$, $L\delta t=0.3\Gamma$ [15] $N=9000$	$m=4$ $L\delta t=0.3\Gamma$	0.0071 [7]	0.0072 ±0.0006 $N=4000$

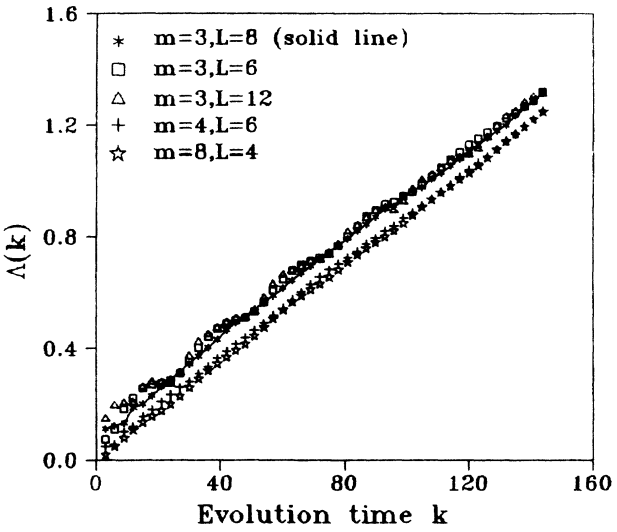

FIG. 3. The $\Lambda(k)$ curves with different m and L for the Rössler attractor.

$L=8$ is 0.067. These results, together with results for other model chaotic systems, are summarized in Table I.

We note that, with regard to the proper reconstruction of the state space, results of [14,15] can be easily obtained by our method with a very small data set. Our method has the additional advantage that the approach is simpler, more natural, and easier to understand and implement, and capable of providing more information.

III. TANGENTIAL MOTION AND UNSTABLE MOTION

In this section we examine the assumption that most sufficiently small distances $\|X_i - X_j\|$ can be regarded as distances between orbits. A problem related to this is whether and how the divergence plot for a continuous system changes with the sampling time δt. Let us discuss the Lorenz system with a much smaller δt of 0.003 than the one taken in Table I. Figure 4(a) shows the divergence plot for $m=3$, $L=27$. We see that the plot consists of two parts, a dotted curvelike part and a heavy black part. Do they have the same origin? Figure 5(a) shows the $\Lambda(k)$ curve, which is by no means linear. Hence we should conclude that each part has its distinct origin, and that the above assumption does not hold in this case.

The answer is rather simple, however. There are certain small distances like $\|X_i - X_{i+w}\|$, with w very small, and their kth iterations can also be very small. These points obviously correspond to the orbital motion, and cannot be regarded as small distances between orbits. Points in the divergence plot corresponding to these points merely reflect changes of the phase velocity along the orbit, and such points will increase if δt is decreased. Let us call this part of the motion tangential motion. The tangential motion contributes a dimension nearly one, corresponding to a Lyapunov exponent equal to zero [24]. Hence this motion should be excluded when calculating the fractal dimension and estimating the largest Lyapunov exponent. A possible way of doing this is to add an additional condition to expression (1), namely,

$$j - i \geq w \ . \tag{3}$$

The expression (1) corresponds to $w=1$. Theiler [25] has proposed a similar improvement to the calculation of the correlation dimension, and suggests that w be selected as the autocorrelation time. In our case, theoretically, the larger the w, the safer the assumption holds. However, we have tested numerically that putting w equal to the embedding window $(m-1)L$ already does the job.

Figure 4(b) is a modified divergence plot. We see that the curvelike part of Fig. 4(a), especially the part corresponding to very small distances where statistics is poor, is largely suppressed. Figure 5(b) is the modified $\Lambda(k)$ curve. It is now nearly linear. Actually the slope of the modified $\Lambda(k)$ curve gives the correct Lyapunov exponent. Thus we know that the heavy black part in the divergence plot originates from the unstable motion, and a linear $\Lambda(k)$ curve is a property of the unstable motion and characteristic of chaotic motions.

Let us summarize. Two kinds of motion, tangential motion and unstable motion, can be discerned from the structure of the divergence plot. The former is irrelevant in the calculation of the fractal dimension and estimation of the largest positive Lyapunov exponent. Hence the former should be removed as much as possible in these calculations. Since the damaging effect of the tangential motion is enhanced when δt is decreased, too small a sampling time is not recommended. Also δt is suggested not to exceed the optimal delay time.

A note on the Lorenz system needs to be made. This system is very complicated in that, even when δt is not small (for example, $\delta t=0.009$), the tangential motion still occupies a large fraction in the divergence plot, and the $\Lambda(k)$ curve is not linear by expression (1), if condition

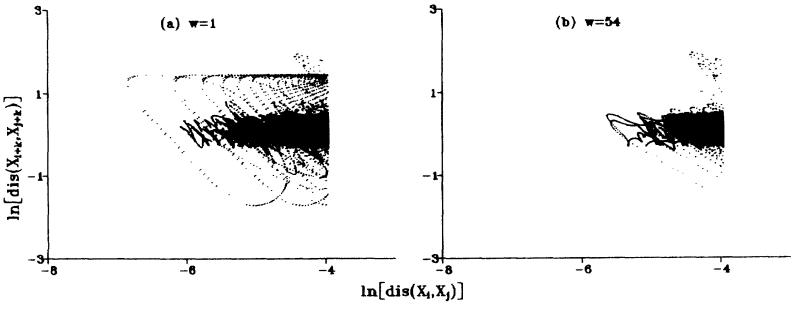

FIG. 4. Original ($w=1$) and modified ($w=54$) local exponential divergence plots for the Lorenz system constructed from the x component of the flow, $m=3, L=27$, $k=30, \delta t=0.003$.

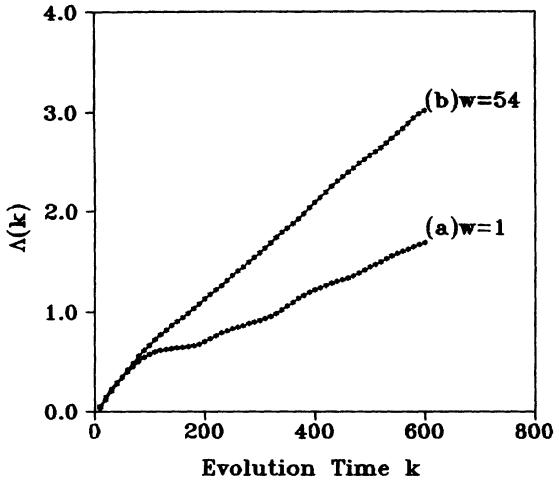

FIG. 5. $\Lambda(k)$ curves corresponding to Fig. 4.

(3) is not imposed. This was probably the reason that Wolf *et al.* [5] used a very large δt to estimate the largest Lyapunov exponent for this attractor. Hence, whenever one calculates the $\Lambda(k)$ curves for a continuous system, it is recommended that condition (3) always be imposed.

Another important note should also be made. The optimal values of m and $L\delta t$ do not change with the sampling time δt. Though condition (3) is suggested when calculating the $\Lambda(k)$ curve to estimate the Lyapunov exponent, it is not needed when reconstructing a state space, since the tangential motion is an integral part of the motion on the attractor.

IV. DYNAMICAL TEST FOR DETERMINISTIC CHAOS

In the last section, we have identified two kinds of motion, the tangential motion and the unstable motion, from the structure of the divergence plot. The former corresponds to the vector field, and is a characteristic of deterministic processes. By employing this property, Kaplan and Glass [26] have used a coarse-grained directional vector to determine whether the motion is deterministic or not. The unstable motion is a characteristic of chaotic motions, which permits short-term prediction but denies long-term forecasting. Sugihara and May [27], and recently Kennel and Isabelle [28], have used prediction as a means of distinguishing chaos from noise. In this section we extend the above formulations to develop a direct and dynamical test for deterministic chaos.

As can be imagined easily, a quantity like $\Lambda(k)/k\delta t$ calculated from expression (1) for a noise may also be positive. However, a stochastic process has no such motions as tangential or unstable, hence, a linear $\Lambda(k)$ curve cannot be expected. Figure 6 shows two divergence plots for a Gaussian white noise. We see that the smaller the r^*, the larger the Λ. In this respect, divergence plots for a chaotic system with underestimated m or too small and too large L also look noisy [Figs. 1(a), 1(b), 1(d)].

Let us first discuss how the $\Lambda(k)$ curve of an IID (independent with identical distribution) random variable series behaves. Formally we construct vectors $X_i=(x_i,x_{i+1},\ldots,x_{i+m-1})$, with m as the embedding dimension (the delay time L is taken to be 1 in accordance with the independency assumption). Take the norm $\|X-Y\|=\max_i|x_i-y_i|$ and assume the probability distribution function of $|x_i-x_j|$ to be $P(x)$. Rewriting expression (1) as

$$\Lambda(k)=\langle\ln\|X_{i+k}-X_{j+k}\|\rangle-\langle\ln\|X_i-X_j\|\rangle\ , \quad \|X_i-X_j\|\le r^* \tag{4}$$

we see that $\Lambda(k)$ is well defined if we have enough pairs of (i,j). Denoting $P(\le y\,|\,\|X_i-X_j\|\le r^*)$ as the conditional probability, we have

$$P(\ln\|X_i-X_j\|\le y\,|\,\|X_i-X_j\|\le r^*)=\begin{cases}[P(e^y)/P(r^*)]^m, & y\le\ln r^*\\ 1, & y>\ln r^*\ ,\end{cases} \tag{5}$$

and when $1\le k\le m-1$,

$$P(\ln\|X_{i+k}-X_{j+k}\|\le y\,|\,\|X_i-X_j\|\le r^*)=\begin{cases}[P(e^y)]^m/[P(r^*)]^k, & y\le\ln r^*\\ [P(e^y)]^{m-k}, & y>\ln r^*\ ,\end{cases} \tag{6}$$

when $k\ge m$,

$$P(\ln\|X_{i+k}-X_{j+k}\|\le y\,|\,\|X_i-X_j\|\le r^*)=[P(e^y)]^m\ . \tag{7}$$

Further considerations lead to the following functional forms:

$$\Lambda=\begin{cases}f(k,m,r^*), & 1\le k\le m-1\\ g(m,r^*), & k\ge m\ .\end{cases} \tag{8}$$

Hence we see that $\Lambda(k)$ as a whole cannot be linear in

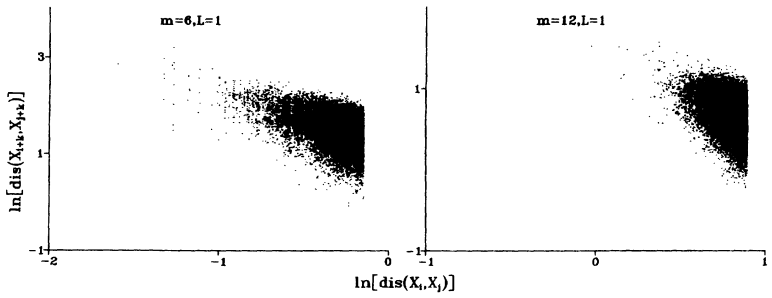

FIG. 6. Divergence plots for Gaussian white noise.

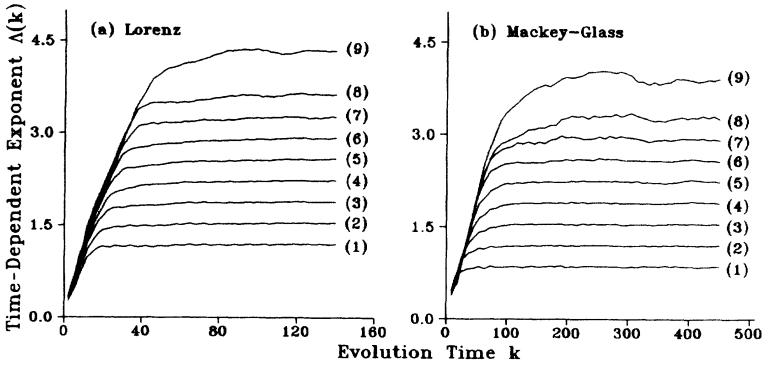

FIG. 7. The $\Lambda(k)$ curves for (a) the Lorenz system ($m=4, L=2$), and (b) Mackey-Glass equation ($m=5, L=1$). Time series is normalized to (0.1). 5000 data points are used. Curves (1)–(9) correspond to shells $(2^{-i-1}, 2^{-i})$, with (a) Lorenz system, $i=5,6,\ldots,13$, and (b) Mackey-Glass equation, $i=3,4,\ldots,11$.

k. Rather, Λ depends on r^* and m. We also note that $\Lambda(k)$ is always positive if r^* is very small, because $\|X_{i+k}-X_{j+k}\|$ has greater probability to be larger than $\|X_i-X_j\|$. By the same reasoning we also know that $\Lambda(k)$ increases with increasing of k when $1\le k\le m-1$. When $k\ge m$, however, we conclude that the $\Lambda(k)$ curve is a horizontal line when there are many pairs of (X_i, X_j) to well define the probabilities.

For a white noise, the K_2 entropy also depends on r^*. Actually, we have the following expression [4]:

$$K_2 \sim \frac{1}{\delta t}\ln\frac{C_m(r)}{C_{m+1}(r)} = -\frac{1}{\delta t}\ln P(r) . \tag{9}$$

Since the correlation integral $C_m(r)$ is equal to $[P(r)]^m$, the probability that two randomly chosen points, X_i and X_j, is less than r apart when the size of the data set is big enough [3].

An implication is that if we calculate the largest Lyapunov exponent or the K_2 entropy from a white noise, we would always have a positive number. However, this number does not imply chaos since it depends on r^* and m, and probably also on L for a colored noise. Actually this number can become as large as one desires provided that the data size is so large that r^* can be very small.

To gain an insight into the dependence of the $\Lambda(k)$ curve on r^* for a noise and to deal with noise-contaminated data with unknown noise level, we extend the formulation (1) by defining $\Lambda(k)$ on a series of shells, $r_{l+1}\le\|X_i-X_j\|\le r_l$, and calculate the corresponding $\Lambda(k)$ curves. We take the Lorenz equations ($\delta t=0.06$) and Mackey-Glass equation ($\delta t=6$) as two examples to illustrate some typical results. We notice from Fig. 7 that there exists a linear envelope to the $\Lambda(k)$ curves. Actually the slope of the envelope estimates the largest Lyapunov exponent. Also we see that a time scale of dynamical correlation [corresponding to the linear increasing segment of the $\Lambda(k)$ curves] is associated with each of the shells, which is important for prediction. Beyond that time scale, the chaotic motion is indistinguishable from a stochastic process.

We now turn to a discussion of stochastic processes. Figure 8 gives a typical result for a white noise of uniform distribution, which shows all the features obtained by the qualitative analysis given above. Figure 9 shows the result for the surrogate data of the Lorenz system, i.e., a data set with the same spectrum but randomized Fourier phases. The qualitative features of Fig. 9 are similar to those of Fig. 8. This is in accordance with the fact that a divergence plot for surrogate data looks quite similar to that of a white noise (Fig. 6). The new characteristics are that $\Lambda(k)$ depends on L and a time scale corresponding to the increasing of $\Lambda(k)$ is slightly larger than the embedding window $(m-1)L$ due to the conditional probability caused by the embedding procedure and the color of the noise. However, this time scale is of the magnitude of the embedding window. The most important fact is that due to the dependence of the $\Lambda(k)$ curves on the radii of the shells, an envelope to the $\Lambda(k)$ curves no longer exists, and the largest positive Lyapunov exponent cannot be defined. Another important fact is that if the time scale corresponding to the increasing of $\Lambda(k)$ is taken as the prediction time, then it is significantly smaller than the time scale of dynamical correlation given by Fig. 7.

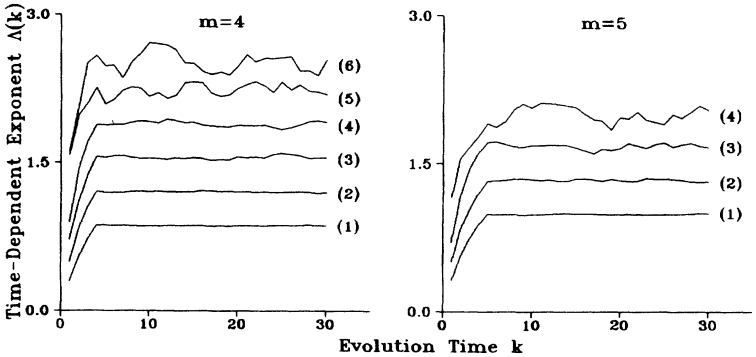

FIG. 8. The $\Lambda(k)$ curves for a normalized uniformly distributed white noise. 6000 data points are used. Curves (1)–(6) correspond to shells $(2^{-i-1}, 2^{-i})$, $i=4,5,\ldots,9$.

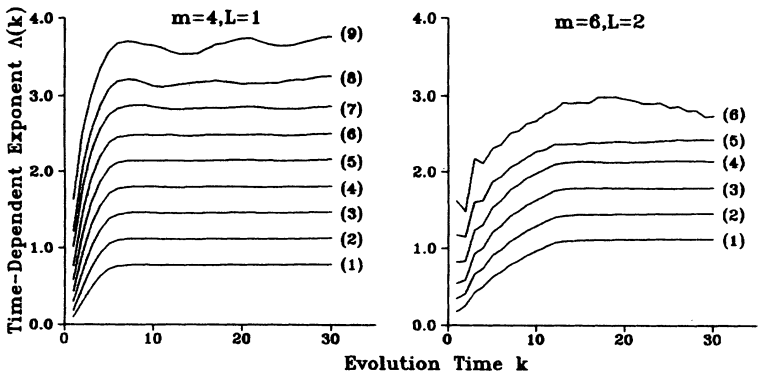

FIG. 9. The $\Lambda(k)$ curves for the surrogate data of the Lorenz system. Time series is normalized to (0,1) and 6000 data points are used. Curves (1)–(9) correspond to shells $(2^{-i-1},2^{-i})$, $i=4,5,\ldots,12$.

Having distinguished clean chaotic signal from pure noise, we now discuss time series of this type, $\{x_i\}+a\{\eta_i\}$. $\{x_i\}$ is a clean chaotic signal and $\{\eta_i\}$ is a pure noise, both of which are normalized to (0,1), and a is the noise level. Intuitively one can imagine that the $\Lambda(k)$ curves for shells of smaller radii will take the characteristics of Figs. 8 and 9, while for shells of larger radii, the characteristics of Fig. 7 will be preserved, i.e., there will only exist a kind of envelope to $\Lambda(k)$ curves of larger radii shells. The higher the noise level, the more the envelope is destroyed. When the noise level is too high, it may be difficult to extract the characteristic of the chaotic motion, since the largest acceptable radius of the shell is bounded by the upper bound $r_{\max}$, which is meaningful for the calculation of the fractal dimension, as pointed out by Eckmann and Ruelle [6]. Figure 10 shows typical results for the Lorenz system supplemented with its surrogate data, which confirm the qualitative features described above. We also note that when $a \geq 0.2$, the characteristics of the chaotic motion are already very difficult to identify. The characteristics of Fig. 10 remain similar when a white noise of Gaussian distribution or uniform distribution is added.

V. DISCUSSION AND CONCLUSION

In this work, we have proposed a local exponential divergence plot to characterize complex time series. Two quantities, Λ and Λ_+, are defined from the divergence plot. Based on their changes with m and L, a criterion for the selection of the minimal acceptable embedding dimension and optimal delay time has been obtained. A chaotic motion is characterized by a linear $\Lambda(k)$ curve, the slope of the curve yields an estimate of the largest Lyapunov exponent. Hence there exists an envelope, which is linear or nearly linear, to the $\Lambda(k)$ curves defined on a series of shells. For stochastic processes, $\Lambda(k)$ cannot be linear in k, and the value of Λ depends on the radii of the shells. Therefore there no longer exists an envelope to the $\Lambda(k)$ curves. This clear difference provides a direct and dynamical method of distinguishing chaos from stochastic processes. When a noise is added to a chaotic signal, the envelope to the $\Lambda(k)$ curves of smaller radii for the underlying chaotic system is destroyed. The higher the noise level, the more the envelope is destroyed.

A point on what a minimal acceptable embedding dimension means should be made clear. Mathematically, the reconstructed dynamics F should have no self-intersections, or physically, the "false neighbors" should drop to zero. However, if only statistical quantities such as Lyapunov exponents are concerned, we may permit very few "false neighbors" to exist. In the case of the Rössler attractor, we conclude that when m reaches 3, some, though very few, false neighbors still exist. This is reflected in the divergence plot of Fig. 1(c) that there exist some excessively large positive points in the plot when compared with Figs. 1(e) and 1(f), and that the $\Lambda(k)$ curves for $m=3$ do not pass through the origin when extrapolated. However, when m reaches 4, the false neighbors have probably dropped to zero (Fig. 3). Liebert, Pawelzik, and Schuster [15] have pointed out a particular difficulty related to this attractor, that most methods before theirs only obtain a minimal embedding dimension

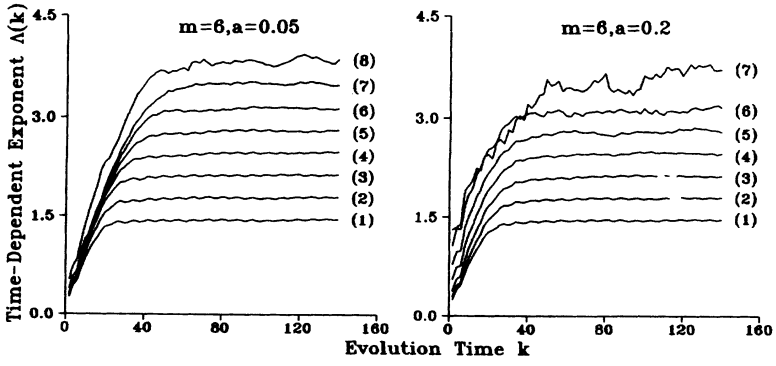

FIG. 10. The $\Lambda(k)$ curves for the Lorenz system added with its surrogate data. Time series is normalized to (0,1) and 6000 data points are used. Curves (1)–(8) correspond to shells $(2^{-i-1},2^{-i})$, $i=5,6,\ldots,12$, respectively.

of 2 or 4, while their method obtains 3. It is quite clear that $m=2$ should be rejected. However, to select m to be 3 or 4 may depend on whether one permits very few false neighbors to exist and what one's purpose is. We regard $m=3$ as acceptable, since the $\Lambda(k)$ curve for $m=3$ and $L=8$ is quite linear, and the estimation of the largest Lyapuov exponent is objective. We may regard a linear $\Lambda(k)$ curve as a practical "safety" test for proper embedding.

However, what Liebert, Pawelzik, and Schuster [15] call a particular problem to the Rössler attractor is not so unique. We have numerically found that when m is selected as the minimal value listed in Table I, the $\Lambda(k)$ curves for the Lorenz system and the Mackey-Glass equation also do not pass through the origin, though this is less serious than in the case of the Rössler attractor.

The issue of determining an optimal delay time is intimately related to the determination of the minimal acceptable embedding dimension. Theoretically [2], when m is large enough, for a large data set, a range of time delays ought to work well. This is reflected in $\Lambda(k)$ curves such as are shown in Fig. 2 where the $\Lambda(L)$ curve around the minima is quite flat. Since the minimal acceptable embedding dimension may frequently be less than the theoretically required value of the embedding dimension, the determination of the delay time may turn out to be quite critical. For example, in the case of the Rössler attractor, when $m=3$, only the $\Lambda(k)$ curve of $L=8$ has a satisfactory linearity, while when m is larger, more values of L are workable. Anyway, the selection of the minima of $\Lambda(L)$ and $\Lambda_+(L)$ curves as an optimal delay time will always guarantee us that the selection is safe.

Sato, Sano, and Sawada [29] have proposed a quantity similar to our expression (1) based on the ensemble average of nearest neighbors. However, the concept of nearest neighbors will not work when there is noise. The idea of introducing r_*, which is determined by the noise level, to form a shell (r_*, r^*) and calculate the Lyapunov exponents has been proposed by Zeng, Eykholt, and Picke [30]. However, they have reported a situation equivalent to that of the $\Lambda(k)$ curve being not linear. We should emphasize the importance of a linear $\Lambda(k)$ curve or a linear envelope to the $\Lambda(k)$ curves. Only when this is the case can we calculate the Lyapunov exponents in a shell objectively.

We have not explored the potentiality of using the significant difference between the dynamical time scale of Fig. 7 and that of the embedding time scale of Figs. 8 and 9. When the noise level is high, up to 20%, the envelope to the $\Lambda(k)$ curves is destroyed (Fig. 10); however, the time scale corresponding to the dynamical one of Fig. 7 is nearly preserved. We expect that an appropriate statistic incorporating this can be worked out to deal with noise-contaminated data with much higher noise level.

ACKNOWLEDGMENTS

This work is jointly supported by the National Basic Research Project "Nonlinear Science" and the Youth Foundation of Institute of Mechanics, Chinese Academy of Sciences.

[1] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Phys. Rev. Lett. **45**, 712 (1980).
[2] F. Takens, in *Dynamical Systems and Turbulence,* edited by D. A. Rand and L. S. Young, Lecture Notes in Mathematics Vol. 898 (Springer-Verlag, Berlin, 1981), p. 366.
[3] P. Grassberger and I. Procaccia, Physica D **9**, 189 (1983).
[4] P. Grassberger and I. Procaccia, Phys. Rev. A **28**, 2591 (1983).
[5] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, Physica D **16**, 285 (1985).
[6] J.-P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).
[7] M. Sano and Y. Sawada, Phys. Rev. Lett. **55**, 1082 (1985).
[8] A. R. Osborne and A. Provenzale, Physica D **35**, 357 (1989).
[9] A. Provenzale, A. R. Osborne, and R. Soj, Physica D **47**, 361 (1991).
[10] A. Provenzale, L. A. Smith, R.Vio, and G. Murante, Physica D **58**, 31 (1992).
[11] J. Theiler, Phys. Lett. A **155**, 480 (1991).
[12] Jianbo Gao and Zhenmin Zheng, Phys. Lett. A (to be published).
[13] A. Cenys and K. Pyragus, Phys. Lett. A **129**, 227 (1988).
[14] Z. Aleksic, Physica D **52**, 362 (1991).
[15] W. Liebert, K. Pawelzik, and H. G. Schuster, Europhys. Lett. **14**, 521 (1991).
[16] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, Phys. Rev. A **45**, 3403 (1992).
[17] J. M. Nese, Physica D **35**, 237 (1989).
[18] H. D. I. Abarbenel, R. Brown, and M. B. Kennel, J. Nonlinear Sci. **1**, 175 (1991); **2**, 343 (1992).
[19] A. M. Fraser and H. L. Swinney, Phys. Rev. A **33**, 1134 (1986).
[20] M. Hénon, Commun. Math. Phys. **50**, 69 (1979).
[21] O. E. Rössler, Phys. Lett. **71A**, 155 (1979).
[22] E. N. Lorenz, J. Atmos. Sci. **30**, 130 (1963).
[23] M. C. Mackey and L. Glass, Science **197**, 287 (1977).
[24] H. Haken, Phys. Lett. **94A**, 71 (1983).
[25] J. Theiler, Phys. Rev. A **34**, 2427 (1986); J. Opt. Soc. Am. A **7**, 1055 (1990).
[26] D. T. Kaplan and L. Glass, Phys. Rev. Lett. **68**, 427 (1992).
[27] G. Sugihara and R. M. May, Nature (London) **344**, 311 (1990).
[28] M. B. Kennel and S. Isabelle, Phys. Rev. A **46**, 3111 (1992).
[29] S. Sato, M. Sano, and Y. Sawada, Prog. Theor. Phys. **77**, 1 (1987).
[30] X. Zeng, R. Eykholt, and R. A. Pieke, Phys. Rev. Lett. **66**, 3229 (1991).