Predicting and characterizing data sequences from structure-variable systems

H. P. Fang[1,2,*] and L. Y. Cao[2]

[1] China Center of Advanced Science and Technology (World Laboratory), P.O. Box 8730, Beijing 100080, China

[2] Institute of Theoretical Physics, P.O. Box 2735, Beijing 100080, China

(Received 9 August 1994)

In principle, all the natural systems such as biological, ecological, and economical systems are structure-variable systems (in which some environment parameters are not fixed). In this paper we show that data sequences from many structure-variable systems are short-term predictable. We also argue regarding how to characterize the data sequences from structure-variable systems.

PACS number(s): 05.45.+b

In the past two decades, there has been rapid progress in understanding deterministic chaos, not only in theoretical modeling [1], but also in experimental testing [2]. Since the pioneering paper [3] and an embedding theorem due to Takens [4], it has become widely known that the current state of the finite-dimensional dynamical systems can be identified using a vector of time series measurements. In the proxy state space consisting of the delay-coordinate vector, analysis of the topological properties of the chaotic attractor underlying the time series is performed. The application of this idea includes noise filtering [5], control of unstable periodic orbits solely from a time series record [6], and prediction of the chaotic time series [7,8].

Up to now, all the studies and the applications have been restricted to the chaotic dynamics of structure-invariable systems (SIS's) (in which all the parameters are fixed) or assumed structure-invariable systems [9–11]. However, many natural systems such as biological, ecological, and economical systems belong to structure-variable systems (SVS's) (in which some parameters are not fixed), they have developed into other ones before they reach equilibrium states. The stock system, the financial expenditure system for one country, the seismic system, and the climatic system are examples since the environments of these systems change rather rapidly before they settle down at some asymptotic states. Remarkably, it was found recently [12] that the Chinese national financial expenditure from 1973 to 1992 can be well predicted based on the records in 1953–1972 by a nonlinear prediction algorithm [8]. It is well known that the Chinese economical environment had been changed greatly from 1953 to 1992. One may ask a question, why does a SIS prediction algorithm work so well? The first motivation of this paper is to present an approach to answer this question, and then the numerical results in [12] can be understood. We will use the concept of the developing diagram [13] to illustrate our idea. It is found that every SVS can be transferred into a corresponding higher-dimensional SIS by adding some new variables so that the nonlinear SIS prediction algorithms can work well for SVS's. Moreover, in the last decade, there has been much work trying to apply the chaotic dynamics to many natural systems [e.g., 9–11]. The correlation dimensions (D_2) and the largest Lyapunov exponents (λ) are calculated with the familiar algorithms for chaotic attractors in SIS's, and the existence of those quantities has been considered as evidence of chaotic dynamics in those natural systems. In principle, many of these natural systems are SVS's. Only when the environment parameters change sufficiently slowly can a theory of SIS's be a good approximate tool to analyze the data sequences from them. However, there are no asymptotic limit sets in the SVS's. Consequently, it is inconvenient to take the dynamics of the SVS's as chaotic dynamics although the real dynamics still reveals sensitivity to initial conditions. Even in these SIS's, transferred from SVS's, there are still no chaotic attractors at all (discussed below). An argument about the physical meaning for these calculated D_2 and λ is the second motivation of this paper.

Let us begin with a typical developing digram [13] shown in Fig. 1 for the logistic map

$$x_{n+1} = f(\mu, x_n) = 1 - \mu x_n^2, \tag{1}$$

in the parameter interval $\mu \in [1.6, 1.8]$. In the figure, we cover the parameter range $\mu \in [1.6, 1.8]$ by small steps $\delta\mu = 10^{-5}$ (as that usually used in bifurcation diagrams). At $\mu = 1.6$, an initial value of x_0 is chosen, which may be a point on the attractor. We iterate Eq. (1) only once with $\mu = 1.6$ and obtain a point x_1. In the successive parameter $\mu = 1.6 + \delta\mu$, an iterate x_2 of Eq. (1) is obtained starting from x_1 by considering the continuously evolutionary process of natural systems. We repeat it until we reach $\mu = 1.8$ and draw all the pairs (1.6, x_0), (1.6 + $\delta\mu$, x_1), (1.6 + $2\delta\mu$, x_2), ... in a figure. In this way we get the *developing diagram* in $x - \mu$ coordinates as shown in Fig. 1. In the process, no point is thrown away as transient. This developing diagram reflects the evolutionary process of a SVS modeled by the logistic map with a definite rapidity (the parameter step $\delta\mu$ reflects the rapidity of evolutionary process of the dynamical system). It can be taken as the simplest example of SVS's. In [13], the basic properties of this developing diagram such as the discontinuity of period-doubling bifurcations, the time arrow [14], and a brief comparison with those of

*Present address and address for correspondence: Physics Department, Fudan University, Shanghai, 200433, People's Republic of China. Electronic address: zlin@fudan.ihep.ac.cn

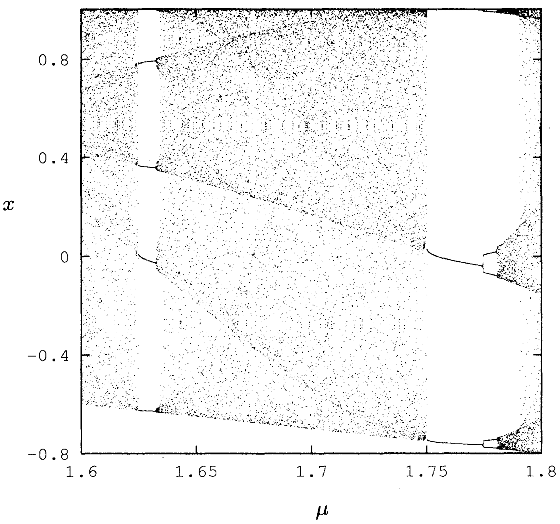

FIG. 1. The developing diagram of the logistic map in the parameter interval $\mu \in [1.6, 1.8]$ with $\delta\mu = 10^{-5}$.

the bifurcation diagram had been discussed.

Now we test the short-term predictability for the above developing diagram. Considering that there is a data sequence $x_1, x_2, \ldots$ from a SIS, according to Takens' theorem [4], there exists a function F such that

$$x_n = F(x_{n-m\tau}, x_{n-(m-1)\tau}, \cdots, x_{n-\tau}), \tag{2}$$

where m is the embedding dimension and τ is the time delay. The problem of predictability is how to find a good estimate of F based on the past history of x_n, on which various techniques of nonlinear deterministic prediction have been developed. In this paper, we will use the prediction algorithm [8] which is based on the wavelet analysis and the neural networks [15]. This prediction algorithm has succeeded in testing the time series from many dynamical models such as the Ikeda map, the Lorenz equations, the Ushiki map, the Mackey-Glass differential-delay equations, etc., for fixed parameters [8].

Suppose $\psi(x)$ $(x \in R^n)$ is a wavelet function, which can be thought of as a band-pass function. Define

$$g(x) = \sum_{i=1}^{N} w_i \psi(D_i R_i (x - b_i)) + \overline{g}, \quad x \in R^n,$$

where the b_i's are arbitrary translation vectors, the D_i's are diagonal matrices built from arbitrary dilation vectors [i.e., $D_i = \mathrm{diag}(d_i)$, $d_i = (d_{1i}, d_{2i}, \ldots, d_{ni})$ is the dilation vector], the R_i's are rotation matrices, which are used to compensate for the orientation selective nature of the dilations, the parameter $\overline{g}$ is introduced to help dealing with nonzero mean functions on finite domains [since the wavelet $\psi(x)$ is zero mean], an, the w_i are weight coefficients. It had been proved [15] that for any function in $L^2(R^n)$, there exist some $g(x)$ with proper parameters such that the function can be approximated by $g(x)$. Our prediction algorithm [8] was proposed by determining the free parameters of $g(x)$ such that $g(x)$ can be a good approximation to the function F in Eq. (2). And the wavelet function $\psi(y)$ was chosen explicitly as

$$\psi(y) = \omega(y_1)\omega(y_2)\cdots\omega(y_m)$$

with [15]

$$\omega(y_i) = (1 - y_i^2)e^{-\frac{y_i^2}{2}},$$

where $y_i, i = 1, 2, \ldots, m$ are m components of the reconstructed vectors.

For the data sequences $x_0, x_1, \ldots$ from the developing diagram in Fig. 1, we fix the embedding dimension $m = 2$ (will be discussed later), and predict one step ahead. We take arbitrary 2000 data points in succession from the developing diagram in Fig. 1. The first 500 points are taken as input to determine the parameters in the prediction algorithm; then we use the next 1500 points as the test points of prediction. All of our numerical results are very satisfactory. In Figs. 2 and 3 we show typical prediction results and the absolute prediction errors for the parameter range [1.74, 1.76]. It is remarkable to find that the algorithm works so well that it can even predict the period 3 motion. Similar observations have been obtained for different selected $\delta\mu$ as well as in other ranges of parameters. We emphasize that the developing structure with variable parameters of the logistic map (which is a SVS) including its periodic motion and its period-doubling counterparts can be well predicted with a sophisticated prediction algorithm for SIS's.

In fact, the developing diagram shown in Fig. 1 is a trajectory in the phase portrait of the following two-dimensional (2D) map

$$\begin{aligned} x_{n+1} &= f(\mu_n, x_n) = 1 - \mu_n x_n^2, \\ \mu_{n+1} &= \mu_n + \delta\mu, \end{aligned} \tag{3}$$

from a 2D initial point $(x, \mu) = (x_0, 1.6)$. The only parameter in this 2D map is $\delta\mu$. Once the parameter $\delta\mu$ is fixed, the developing structure of the logistic map shown in Fig. 1 can be described by a 2D SIS (2). As has been

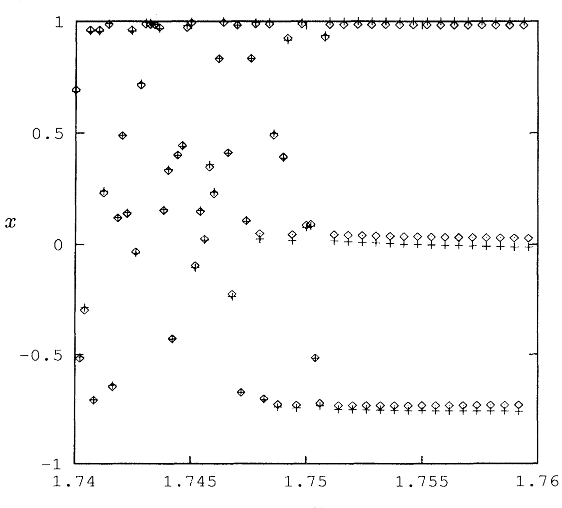

FIG. 2. One-step-ahead prediction results for the developing structure of the logistic map shown in Fig. 1 with $\mu \in [1.74, 1.76]$. The crosses are the true data points and the diamonds are the corresponding predictive ones. Only one point in each 20 data points is plotted.

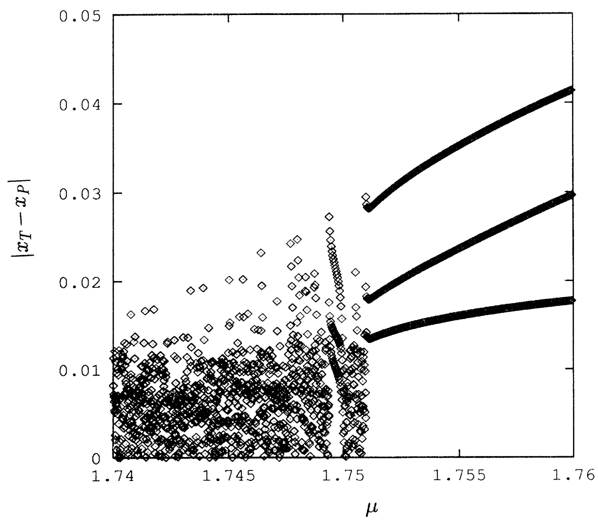

FIG. 3. Prediction error for Fig. 2. x_T and x_P are the true and the corresponding predictive values, respectively.

shown above, we have succeeded in predicting the developing structure of the logistic map shown in Fig. 1 with a 2D prediction algorithm (m=2) for SIS's.

Now we arrive at the first conclusion of this paper. A SVS can be discussed in its variable-parameter space in which the SVS is transferred into a higher-dimensional SIS so that the data sequence is short-term predictable. We have checked this idea for many other dynamical systems. In Fig. 4 we show our prediction results for the developing structure of the Hénon map

$$x_{n+1} = 1 - a x_n^2 + y_n, \qquad y_{n+1} = b x_n, \tag{4}$$

with parameters (a, b) change simultaneously as

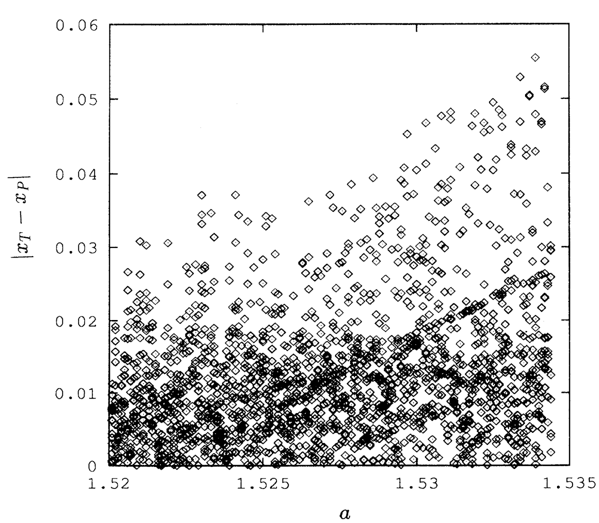

FIG. 4. Prediction error from the developing structure of the Hénon map with its parameters changes nonlinearly shown in Eq. (5).

$$a_{n+1} = a_n + a_n^3 \delta a, \qquad b_{n+1} = b_n + a_n \delta b, \tag{5}$$

where a_0 =1.52, b_0=0.1, $\delta a = 2\times10^{-6}$, and $\delta b = 3\times10^{-6}$. Our prediction algorithm still works very well.

From the above discussion we can understand the numerical result in [12] for the Chinese national financial expenditure from 1953 to 1992 and other economic data series. Since the natural environment and the political structure had not changed much from 1953 to 1992, the evolutionary process of the economical environment in this period might be governed by some equations similar to Eq. (5) (though we might not know exactly what they are). In this variable-parameter space the system is a SIS so that the data sequence might be well predicted. With this idea, we have successfully tested some seismological data in China recently [16].

Now we come to the second point in this article. How could a data sequence from a SVS be characterized? So far, there are many sophisticated ways of characterizing chaotic attractors, among which the Lyapunov exponents and the correlation dimensions are widely used. In the past decade, these techniques have been applied to the data sequences from many natural systems. D_2 and λ are calculated and their existence has been considered as evidence of chaotic dynamics in those natural systems [e.g., 9–11]. In fact, many of these natural systems are SVS's. Only when the environment parameters change sufficiently slowly, can these techniques be approximately used to analyze the data sequences from these natural systems. A SVS does not exhibit chaotic attractor since there is no asymptotic state. For these SIS's from SVS in the variable-parameter space as Eq. (3), there is no unstable periodic orbit in their asymptotic state since the parameter μ always increases. As a result, it seems inconvenient to apply the ways of characterizing chaotic attractors to analyze the data sequences from SVS's, although these data sequences might be predictable.

However, λ and D_2 have been calculated for the data sequences for some natural systems [e.g., 9–11] by using the usual algorithms for chaotic attractors. What do these calculated λ and D_2 mean? Are they the evidence that the environment parameters for these natural systems change so slowly that they can be approximately taken as SIS's? Could the existence of those quantities be taken as evidence of chaotic dynamics in those natural systems? In order to answer these questions we have calculated the values of λ and D_2 for 10 000 data sequences for the developing structure of the Hénon map in Eqs. (4) and (5) with the algorithms from [17,18]. They exist and D_2=1.25±0.1 and λ=0.294±0.002. We also have computed these quantities for other developing parameter ranges of the Hénon map and the logistic map. From these results, it is clear that the existence of the calculated λ and D_2 does not represent that the system evolves so slowly that it can be approximated by a SIS. However, the positive value of λ is consistent with the sensitivity to initial conditions of these nonlinear dynamical systems in our calculations and a little smaller than the largest Lyapunov exponents for the dynamical systems with fixed parameters in the developing parameter range except for periodic orbits, and the correlation dimensions are greater than those for the fixed parameters of which the above developing diagram sweeps out [For the Hénon map, D_2=1.03±0.02, λ=0.353±0.003 for (a,b)=(1.52, 0.1) and D_2=1.07±0.01, λ=0.321±0.001 for (a,b)=(1.5955, 0.14671), corresponding to the first and the last datum, respectively]. The physical meaning for these calculated quantities and how to characterize the data sequences from SVS still need to be investigated, which will be presented in an extended paper.

It should be noted that a data sequence from SVS can be easily confused with a time series of a chaotic attractor with noise. Figure 5 shows the plotting for successive x_{n+1} to x_n for the same data sequence used above for calculating D_2 and λ. This plotting looks very similar to what one would get for a time series of a chaotic attractor with noise. How to distinguish a data sequence between that from a SVS and from a chaotic attractor with or without noise is still undertaken.

To conclude, we will say that the data sequences from some SVS's are predictable with a nonlinear SIS prediction algorithm although the SVS's do not exhibit chaotic attractors. This result encourages us to apply the prediction algorithm to predict the data sequences from natural systems no matter their environment parameters change slowly or rapidly. However, there are still some theoretical problems needed to be clarified if one wants to apply the usual ways of characterizing the chaotic attractors to characterize the data sequences from SVS's otherwise *misleading* results might be obtained. Finally, we note that all the discussions in this article can be extended to discuss the dynamical systems described by differential or difference-differential equations.

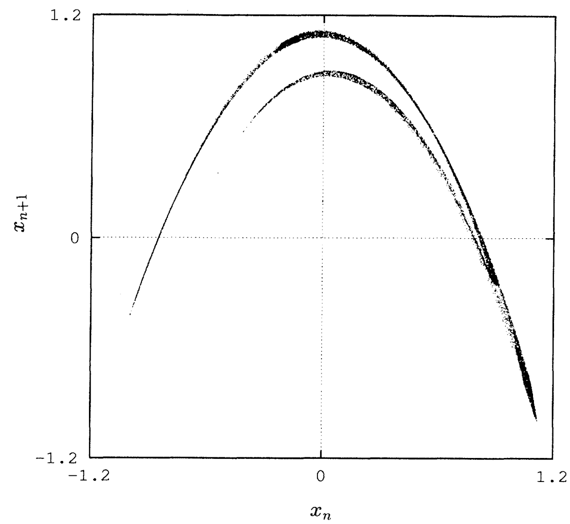

FIG. 5. The plotting for successive x_{n+1} to x_n for the first 10 000 data from Eq. (4).

H.P.F. thanks Professor Hao Bailin for his encouragement.

[1] E. Ott, Rev. Mod. Phys. **53**, 655 (1981).
[2] N.B. Abraham, J.P. Gollub, and H.L. Swinney, Physica D **11**, 252 (1984).
[3] N. Packard, J. Crutchfield, J.D. Farmer, and R. Shaw, Phys. Rev. Lett. **45**, 712 (1980).
[4] F. Takens, in *Dynamical Systems and Turbulence*, edited by D.A. Rand and L.-L. Young (Springer-Verlag, Berlin, 1981); T. Sauer, J.A. Yorke, and M. Casdagli, J. Stat. Phys. **65**, 579 (1991).
[5] P. Grassberger *et al.*, Chaos **3**, 127 (1993); E.J. Kostelich and T. Schreiber, Phys. Rev. E **48**, 1752 (1993).
[6] T. Shinbrot, C. Grebogi, E. Ott, and J.A. Yorke, Nature (London) **363**, 411 (1993).
[7] J.D. Farmer and J.J. Sidorowich, Phys. Rev. Lett. **59**, 845 (1987); M. Casdagli, Physica D **35**, 335 (1989); H.D.I. Abarbanel, R. Brown, and J.B. Kadtke, Phys. Rev. A **41**, 1782 (1990); J. Stark, J. Nonlin. Sci. **3**, 197 (1993).
[8] L.Y. Cao, Y.G. Hong, and H.P. Fang, Physica D (to be published).
[9] C. Nicolis and G. Nicolis, Proc. Natl. Acad. Sci. USA **83**, 536 (1986).
[10] P. Chen, Syst. Dyn. Rev. **4**, 605 (1988).
[11] G.P. Pavlos *et al.*, Int. J. Bifurc. Chaos **4**, 87 (1994).
[12] L.Y. Cao *et al.* (unpublished). In this paper, several Chinese macroeconomic data sequences, e.g., the national financial expenditure, total value of retail sales, gross output value of retail, and monetary supply are found to be short-term predictable.
[13] H.P. Fang (unpublished).
[14] I. Prigogine, *From Being to Becoming* (W.H. Freeman, San Francisco, 1980).
[15] Q.-H. Zhang and A. Benveniste, IEEE Trans. Neural Networks **3**, 889 (1992); C. Chui, *Wavelet: An Introduction* (Academic, New York, 1992).
[16] L.Y. Cao and H.P. Fang, Int. J. Bifurc. Chaos (to be published). In this paper, the b value for the seismic events in Tangshan region in China is well tested with our prediction algorithm.
[17] J.-P. Eckmann *et al.*, Phys. Rev. A **34**, 4971 (1986).
[18] P. Grassberger and I. Procaccia, Physica D **9**, 189 (1983).