Chaotic Analysis on Monthly Precipitation on Hills Region in Middle Sichuan of China

Baohui Men [1, 2] , Xiejing Zhao [1] , Chuan Liang [2]

(1.Agricultural Department, Key Open Laboratory of Agriculture Resource and Environment in Yangtze River Upper Reaches, Chengdu, Sichuan 610066, China, bhmen1973@sohu.com;
2.College of Hydraulic Engineering, Sichuan University, Chengdu, Sichuan 610065, China)

Abstract: The chaotic behavior of monthly precipitation time series is investigated using the phase-space reconstruction technique and the principal component analysis method. A hydrological time series, monthly precipitation series of 50 years (with a total of 600 values) observed on hills in middle region of Sichuan, China, is studied. Relationship between embedding dimension m and correlation dimension D_2 is discussed and saturation correlation dimension, minimum embedding dimension and Kolmogorov entropy are calculated, that is, $D_2 = 4.02$, $m = 19$ and $k = 0.25$. Meanwhile, primary component analytic method (PCA) is applied to validate its chaotic character and result shows forecasting length for this precipitation time series should be less than 4.0 months. Thus, chaotic analysis on precipitation time series provides a scientific gist for precipitation forecasting. [Nature and Science, 2004,2(2):45-51]

Key words: correlation dimension; hills region in middle of Sichuan; principal component analysis (PCA) method; Kolmogorov entropy

1 Introduction

The science of chaos is a burgeoning field, and the available methods to investigate the existence of chaos in time series are still in a state of infancy. However, the considerable attention that the theory has received in almost all fields of natural and physical sciences has motivated improvements in existing methods for the diagnosis of chaos and the proposal of new ones. The methods available thus far are the correlation dimension method (Grassberger, 1983a, 1983b), the nonlinear prediction method (Farmer, 1987; Casdagli, 1989; Sugihara, 1990) including deterministic versus stochastic diagram (Casdagli, 1991), the Lyapunov exponent method (Wolf, 1985), the Kolmogorov entropy method (Grassberger, 1983c), the surrogate data method (Theiler, 1992), and the linear and nonlinear redundancies (Palus, 1995; Prichard, 1995). Among these the correlation dimension method has been the most widely used one for the investigation of deterministic chaos in hydrological phenomena (Hense, 1987; Rodriguez-Iturbe, 1989; Sharifi, 1990; Berndtsson, 1994; Jayawardena, 1994; Puente, 1996; Sangoyomi, 1996; Porporato, 1996, 1997; Sivakumar, 1998, 1999a; Sivakumar, 2000). In the present study, the correlation dimension method is employed, and the presence of a low-dimensional attractor (a geometric object which characterizes the long-term behavior of a system in the phase space) is taken as an indication of chaos.

It is relevant to note that the application of chaos identification methods, particularly the correlation dimension method, to hydrological time series and the reported results have very often been questioned because of the fundamental assumptions with which the methods have been developed, that is, that the time series is infinite and noise-free. Important issues, in the application of chaos identification methods to hydrological data, for example, data size, noise, delay time, etc., and the validity of chaos theory in hydrology have been discussed in detail by Sivakumar (2000) and therefore are not reported herein. It is relevant to note, however, that the studies by Sivakumar (1999, 2000) reveal that the presence of noise in the data does not significantly influence the correlation dimension estimates (though it significantly influence the prediction accuracy estimates). This suggests that the correlation dimension may be used as a preliminary indicator to identify the existence of chaos in the monthly precipitation time series. In this thesis, according to precipitation time series of Sichuan middle part in upper regions of Yangtze, from 1953 to 2002 (Figure 1), from the view of correlation dimension D_2 and Kolmogorov entropy, regulation of precipitation formation and evolution is discussed, and then PCA is applied to validate chaotic feature of precipitation time series.

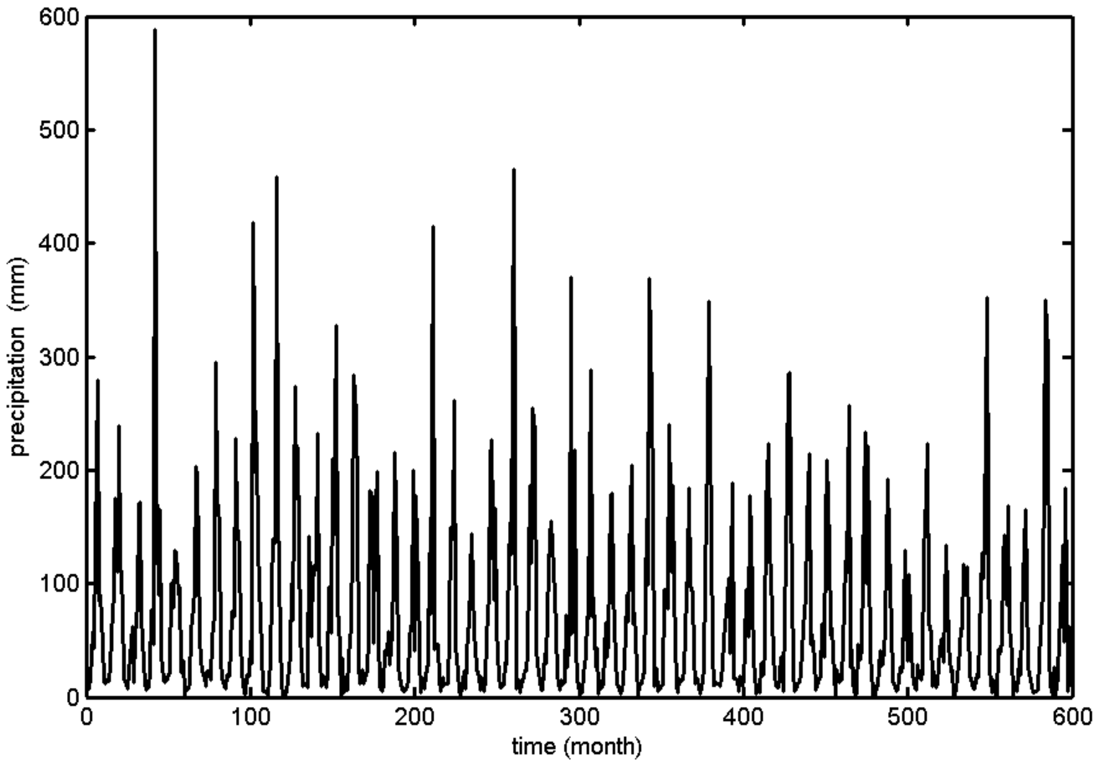

Figure 1 Precipitation Series of Hills Region in Middle Sichuan of China

2 Analysis of Chaotic Time Series

2.1 Reconstruction of the phase space

For a scalar time series x_t, where $t = 1,2,\cdots,N$, the phase space can be reconstructed using the method of delays (Takens, 1980). The basic idea in the method of delays is that the evolution of any single variable of a system is determined by the other variables with which it interacts. Information about the relevant variables is thus implicitly contained in the history of any single variable. On the basis of this an "equivalent" phase space can be reconstructed by assigning an element of the time series x_t and its successive delays as coordinates of a new vector time series

$$Y_t = \{x_t, x_{t-\tau}, x_{t-2\tau}, \cdots, x_{t-(m-1)\tau}\} \quad (1)$$

where $t = 1,2,\cdots,N-(m-1)\tau/\Delta t$, m is the dimension of the vector Y_t, also called the embedding dimension, and τ is a delay time taken to be some suitable multiple of the sampling time Δt (Packard, 1980; Takens, 1980). Take a scalar time series $x_1, x_2, \cdots, x_n$ in system phase-space as an example. Supposing its dimension d is 1, its dimension of embedding phase-space should be 3. If here $m = 4$, x_1, x_2, x_3, x_4 forms the first vector Y_1 of a four-dimensional state space and then moving right one step, x_2, x_3, x_4, x_5 forms the second vector Y_2. Just do in the same way, $Y_1, Y_2, Y_3, \cdots, Y_l$ forms the time series of reconstruction phase-space.

2.2 Correlation dimension method

The goal of determining the dimension of an attractor is that the dimensionality of an attractor furnishes information on the number of dominant variables present in the evolution of the corresponding dynamical system. Dimension analysis will also reveal the extent to which the variations in the time series are concentrated on a subset of the space of all possible variations. The central idea behind the application of the dimension approach is that systems whose dynamics are governed by stochastic processes are thought to have an infinite value for the dimension. A finite, noninteger value of the dimension is considered to be an indication of the presence of chaos.

Correlation dimension is a measure of the extent to which the presence of a data point affects the position of the other point lying on the attractor. The correlation dimension method uses the correlation integral (or function) for distinguishing between chaotic and stochastic behaviors. The concept of the correlation integral is that an irregular-looking process arising from deterministic dynamics will have a limited number of degrees of freedom equal to the smallest number of first-order differential equations that capture the most important features of the dynamics. Thus, when one constructs phase spaces of increasing dimension for an infinite data set, a point will be reached where the dimension equals the number of degrees of freedom and beyond which increasing the dimension of the representation will not have any significant effect on the correlation dimension.

According to the embedding theorem of Takens (1980), to characterize a dynamic system with an attractor dimension d, an $(m=2d+1)$-dimensional phase space is required. However, Abarbanel (1990) suggested that $m>d$ would be sufficient. For an m-dimensional phase space the correlation function $C(r)$ is given by

$$C(r)=\lim_{N\to\infty}\frac{2}{N(N-1)}\sum_{i,j}H\left(r-\left|Y_i-Y_j\right|\right)_{1\le i\le j\le N} \quad (2)$$

where H is the Heaviside step function, with $H(u)=1$ for $u>0$, and $H(u)=0$ for $u\le 0$, where $u=r-\left|Y_i-Y_j\right|$, N is the number of point on the reconstructed attractor, r is the radius of the sphere centered on Y_i or Y_j.

If the time series is characterized by an attractor, then for positive values of r the correlation function $C(r)$ is related to the radius r by the following relation:

$$C(r)\underset{\substack{r\to 0\\ N\to\infty}}{\approx}\alpha r^{D_2} \quad (3)$$

where α is a constant; and D_2 is the correlation exponent or the slope of the $\log C(r)$ versus $\log r$ plot is given by:

$$D_2=\lim_{\substack{r\to 0\\ N\to\infty}}\frac{\log C(r)}{\log r} \quad (4)$$

The slope is generally estimated by a least squares fit of a straight line over a certain range of r, called the scaling region. For a finite data set, such as the precipitation data series, it is clear that there is a separation r below which there are no pairs of point; that is, it is "depopulated." At the other extreme, when the value of r exceeds the set diameter, the correlation function increases no further; that is "saturated." Therefore, for a finite data set, the region sandwiched between the depopulation region and the saturation region is considered as the scaling region. A somewhat better way to identify the scaling region is to estimate the local slope given by $d[\log C(r)]/d[\log r]$.

To observe whether chaos exists, the correlation exponent (or local slope) values are plotted against the corresponding embedding dimension values. If the value of the correlation exponent is finite, low, and noninteger, the system is considered to exhibit low-dimension chaos. The saturation value of the correlation exponent is defined as the correlation dimension of the attractor. The nearest integer above the saturation value provides the minimum number of phase spaces or variables necessary to model the dynamics of the attractor. On the contrary, if the correlation exponent increases without bound with increase in the embedding dimension, the system under investigation is considered as stochastic (Fraedrich, 1986).

2.3 Principal component analysis (PCA)

PCA is the most widely used method of multi-variate data analysis owing to the simplicity of its algebra and its straightforward interpretation (Cerón, 1999). However, it is a newly proposed method (Lv, 2002) in organizing noise and chaos. The step of this method is as follows:

Supposing a scalar time series is $x_1, x_2, \cdots, x_n$, after reconstructing phase-space (embedding dimension is m, and delay time is τ), matrix $Y_{l\times m}$ $(l=n-(d-1))$ is formed:

$$Y_{l\times m} = \frac{1}{l^{1/2}}\begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ x_2 & x_3 & \cdots & x_{m+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_l & x_{l+1} & \cdots & x_n \end{bmatrix} = \frac{1}{l^{1/2}}\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_l \end{bmatrix}$$

Calculate covariance matrix $A_{d\times m} = \frac{1}{l} Y_{l\times m}^T Y_{l\times m}$ and its eigenvalue λ_i $(i = 1,2,3,\cdots,m)$ and eigenvector U_i $(i = 1,2,3,\cdots,m)$, then order them $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ in descending sequence. Eigenvalue and eigenvector is called primary sector. Sum of all eigenvalue γ is $\gamma = \sum_{i=1}^{d} \lambda_i$. Chart of i and $\ln(\lambda_i / \gamma)$ is called primary spectrum. Primary spectrum of noise, which is parallel to x axis, is quite different from that of chaotic serial, which is a line across fixed dot with negative slope.

2.4 Kolmogorov entropy

Another important index of chaotic feature is Kolmogorov entropy, which provides upper and lower range of average amount of information in unit time. Generally, for a sequential system, $K = 0$; for a stochastic system, $K = \infty$. When $0 < K < \infty$, system is a chaotic system, and the bigger K is, the more serious the degree of chaos is. Formula proposed by Grassberger-Procaccia algorithm is:

$$K_2 = \frac{1}{\tau} \ln \frac{C_m(r)^2}{C_{m+1}(r)^2} \qquad (5)$$

Where: τ is delay time, $C_m(r)$ is the value of $C(r)$ when embedding dimension of phase-space is m, $C_{m+1}(r)$ is the value of $C(r)$ when embedding dimension of phase-space is $m+1$.

Choice of τ and m is key to calculation of dimension, index and entropy. In application, we need to consider dimension of embedding phase-space as well as τ which has better simulating effect.

In theory, when $m \to \infty$, $K_2 \to K$. In fact, when m is somewhat value, K_2 tends to be stable and this stable value can be used as estimating value of K.

3 Result and Analysis

3.1 Correlation dimension calculation

LnC(r) versus lnr is shown in Figure 2. Figure 2 is obtained by using formula (2) and (4). Form the Figure 2, the scaling region is existed; herein the precipitation series has chaotic character. The slope of the line in scaling region is the correlation dimension. The relationship between the correlation dimension values and the embedding dimension values is shown in Figure 3. It can be seen that the correlation dimension value increases with the embedding dimension up to a certain value and then saturates beyond that value. The saturation of the correlation dimension beyond a certain embedding dimension value is an indication of the existence of deterministic dynamics. The saturated correlation dimension is about 4.02 ($D_2 = 4.02$), and the embedding dimension $m = 19$. The finite and low correlation dimension is an indication that the precipitation series exhibit chaotic behavior.

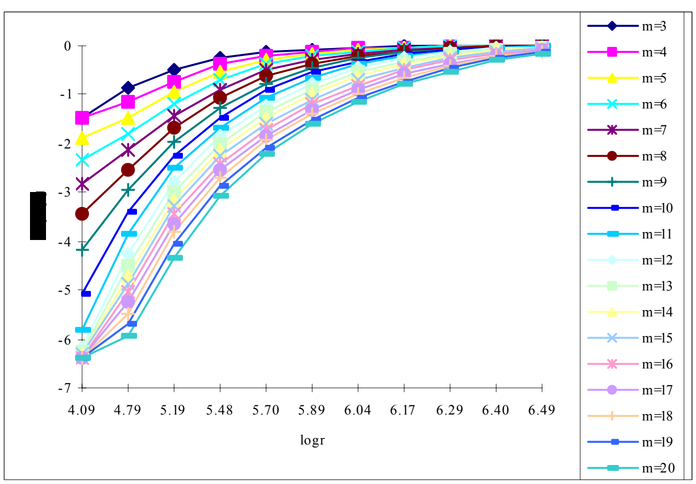

Figure 2 LnC(r) versus lnr plot

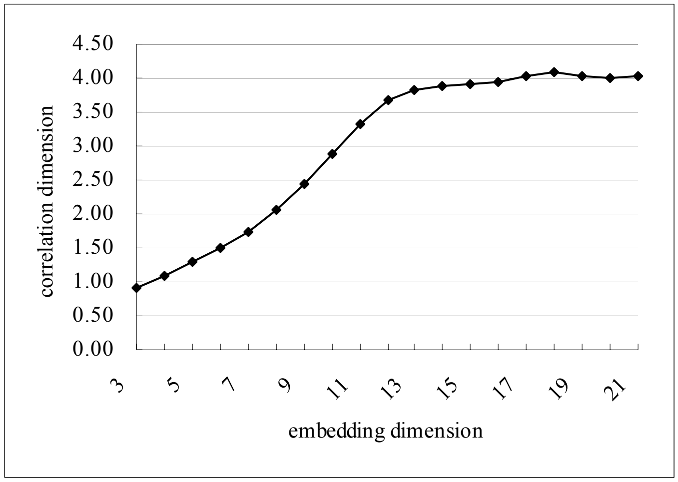

Figure 3 Relationship between embedding dimension and correlation dimension

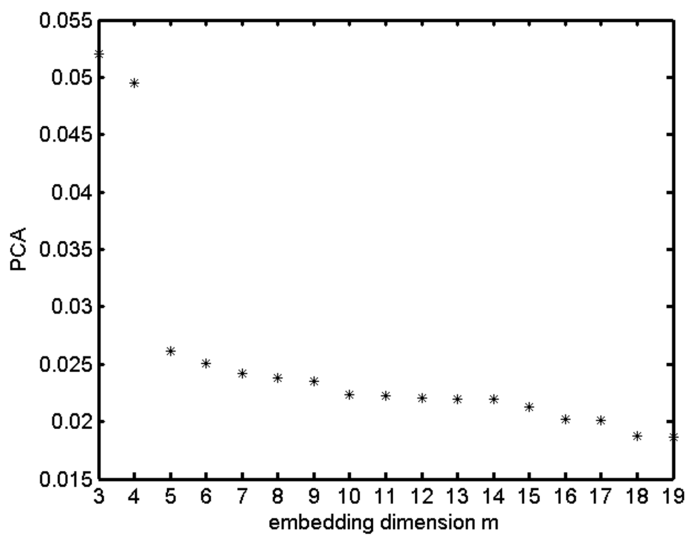

Figure 4 Relationship between embedding dimension and PCA

3.2 Principal component analysis method

The relationship between embedding dimension value and the PCA value is shown in Figure 4. Figure 4 is plotted by the method of principal component analysis. Principal spectrum of precipitation series is a line across fixed dot with negative slope. It further validates that precipitation time series has chaotic character.

3.3 Kolmogorov entropy calculation

The relationship between the embedding dimension value and the Kolmogorov entropy value is shown in Figure 5. The Kolmogorov entropy value is calculated by formula (5). With the increasing of embedding dimension value, Kolmogorov entropy value k tends to be stable and when the embedding dimension value is about 20, that is $(m+1)=20$, Kolmogorov entropy comes to saturation, that is, $K=0.25(>0)$. This data also indicates the chaotic feature of precipitation time series of hills region in middle Sichuan of China. $1/K$ shows that predictable length of this system is 4.0 months.

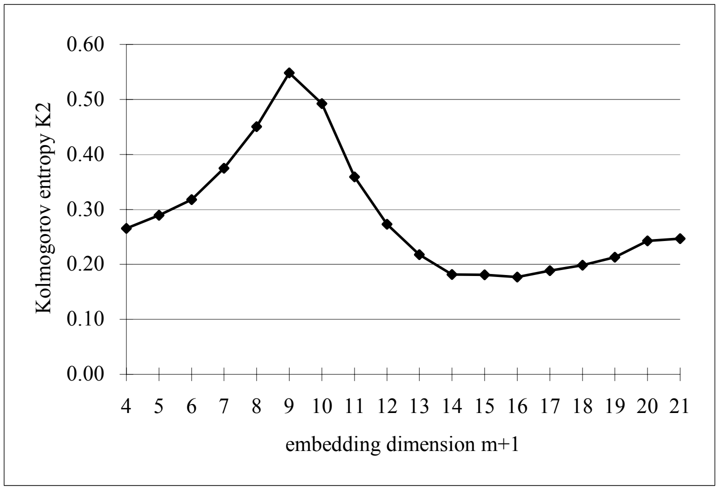

Figure 5 Relationship between embedding dimension and kolmogorov entropy

4 Conclusion

Correlation dimension $D_2 = 4.02$ and Kolmogorov entropy $k = 0.25$ are achieved by reconstructing phase-space. Primary component analysis further validates the chaotic feature of precipitation time series of Sichuan middle part in upper regions of Yangtze and the reciprocal of Kolmogorov entropy tells us predicting length of precipitation time series should be 2 to 3 years instead of long-term prediction, which provides scientific gist for determining length of predicting period.

Acknowledge

This project is supported by HI-TECH Research and Development Program of China （2002AA6Z3263）.

References

[1] Abarbanel HDI, Brown R, Kadtke JB. Prediction in chaotic nonlinear systems: Methods for time series with broadband Fourier spectra. Phys Rev A 1990;41:1783-807.

[2] Berndtsson RK, Jinnon A, Kawamura J, Olsson, Xu S, Dynamical systems theory applied to long-term temperature and precipitation time series. Trends Hydrol 1994;1:291-7.

[3] Casdagli M. Nonlinear prediction of chaotic time series. Physica D 1989;35:335-6.

[4] Casdagli M. Chaos and deterministic versus stochastic non-linear modeling. J R Stat Soc, Ser B 1991;54(2):303-28.

[5] Cerón JC, Pulido-Bosch A, Bakalowicz M. Application of principal component analysis to the study of CO_2-rich thermomineral waters in the aquifer system of alto guadelentín (Spain). Hydrol Sci J 1999;44(6):929-42.

[6] Fraedrich K. Estimating the dimensions of weather and climate attractors. J Atmos Sci 1986;43(5):419-32.

[7] Grassberger P, Procaccia I. Measuring the strangeness of strange attractors. Physica D 1983a;9:189-208.

[8] Grassberger P, Procaccia I. Characterization of strange attractors. Phys Rev Lett 1983b;50:346-9.

[9] Grassberger P, Procaccia I. Estimation of the Kolmogorov entropy from a chaotic signal. Phys Rev A 1983c;28:2591-3.

[10] Hense A. On the possible existence of a strange attractor for the southern oscillation. Beitr Phys Atmos 1987;60(1):34-47.

[11] Jayawardena AW, Lai F. Analysis and prediction of chaos in rainfall and stream flow time series. J Hydrol 1994;153:3-52.

[12] Lv JH, Lu JA, Chen SH. Chaotic time series analysis and its application. Wuhan University Press, Wuhan, China. 2002. Farmer DJ, Sidorowich JJ. Predicting chaotic time series. Phys Rev Lett 1987;59:845-8.

[13] Packard NH, Crutchfield JP, Farmer JD, Shaw RS. Geometry from a time series. Phys Rev Lett 1980;45: 712-6.

[14] Palus M. Testing for nonlinearity using redundancies: Quantitative and qualitative aspects. Physica D 1995;80:186-205.

[15] Porporato A, Ridolfi L. Clues to the existence of deterministic chaos in river flow. Int J Mod Phys B 1996;10:1821-62.

[16] Porporato A, Ridolfi L. Nonlinear analysis of river flow time sequences. Water Resour Res 1997;33(6):1353-67.

[17] Prichard D, Theiler J. Generalized redundancies for time series analysis. Physica D 1995;84:476-93.

[18] Puente CE, Obregon N. A deterministic geometric representation of temporal rainfall: Results for a storm in Boston. Water Resour Res 1996;32(9):2825-39.

[19] Rodriguez-Iturbe I, De Power FB, Sharifi MB, Georgakakos

KP. Chaos in rainfall, Water Resour Res 1989;25(7):1667-75.

[20] Sangoyomi TB, Lall U, Abarbanel HDI. Nonlinear dynamics of the Great Salt Lake: Dimension estimation. Water Resour Res 1996;32(1):149-59.

[21] Sharifi MB, Georgakakos KP, Rodriguez-Iturbe I. Evidence of deterministic chaos in the pulse of storm rainfall, J Atmos Sci 1990;47:888-93.

[22] Sivakumar B. Chaos theory in hydrology: Important issues and interpretations. J Hydrol 2000;227:1-20.

[23] Sivakumar B, Liong SY, Liaw CY. Evidence of chaotic behavior in Singapore rainfall. J Am Water Resour Assoc 1998;34(2):301-10.

[24] Sivakumar B, Liong SY, Liaw CY, Phoon KK. Singapore rainfall behavior: chaotic? J Hydrol Eng 1999a; 4(1):38-48.

[25] Sugihara G, May RM. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature 1990;344:734-41.

[26] Takens F. Detecting strange attractors in turbulence, in Dynamical Systems and Turbulence, Lecture Notes in Mathematics 898, edited by Rand DA, Young LS. Springer-Verlag, New York. 1980:366-81.

[27] Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD. Testing for nonlinearity in time series: The method of surrogate data. Physica D 1992;58:77-94.

[28] Wolf A, Swift JB, Swinney HL, Vastao A. Determining Lyapunov exponents from d time series. Physica D 1985;16:285-317.