

# 基于复杂度分析 logistic 映射和 Lorenz 模型的研究\*

侯 威<sup>1)†</sup> 封国林<sup>2)‡</sup> 董文杰<sup>2)</sup>

<sup>1)</sup> 扬州大学物理科学与技术学院, 扬州 225009)

<sup>2)</sup> 国家气候中心气候研究开放实验室, 北京 100081)

<sup>3)</sup> 中国科学院大气物理研究所, 北京 100029)

(2004 年 10 月 27 日收到, 2005 年 2 月 28 日收到修改稿)

采用三次粗粒化方法得到了 logistic 映射和 Lorenz 模型的符号序列, 运用动态非线性时间序列分析方法——Lemper-Ziv 复杂度, 分别对两组符号序列进行了对比分析. 对于 logistic 映射, 其复杂度反映了时间序列的演化; Lorenz 模型三个分量的复杂度序列都具有混沌性质, 即由许多振幅非常接近而长度完全不同的循环所组成, 反映了 Lorenz 模型内在的准周期特性. 进一步研究发现, 当取不同的窗口长度时, 复杂度序列的特征基本相同, 并且复杂度反映了时间序列的时空特性. 因此, 可以借助复杂度的计算来反演观测资料的动力学结构.

关键词: 三次粗粒化, Lemper-Ziv 复杂度, logistic 映射, Lorenz 模型

PACC: 9260X

## 1. 引 言

近年来对各种复杂现象的研究已成为非线性科学的前沿课题之一<sup>[1-4]</sup>, 在地球科学中对复杂性也进行了大量的研究<sup>[5-9]</sup>. 自 20 世纪 70 年代末以来对系统复杂度的研究都是用分维数、Lyapunov 指数、近似熵等作为度量, 但这些指数要求具有大量数据才能保证其计算结果的可靠性. 然而在气候、地震等应用领域观测资料的长度却是非常有限的. Lemper-Ziv 复杂度<sup>[8]</sup>是对任意给定的有限长度符号序列复杂性的度量, 计算方便, 适用于短序列的研究.

本文分别对 logistic 映射和 Lorenz 模型的三个分量进行三次粗粒化, 得到它们的符号序列. 利用 Lemper-Ziv 复杂度方法来研究这些符号序列. 计算并分析了两个目前典型的非线性动力系统的复杂度, 作为混沌时间序列复杂度研究的一个尝试.

## 2. 粗粒化方法

对于非线性时间序列复杂度的研究, 目前绝大多数文献都采用二次粗粒化方法, 即对于某一时间序列  $H(i) (i = 1, 2, \dots, n)$ , 令

$$R_{ave} = \left( \sum_{i=1}^n H(i) \right) / n \quad (1)$$

为这一时间序列的平均值,  $L(i) (i = 1, 2, \dots, n)$  是一个与时间序列  $H(i)$  长度相同的空符号串. 当时间序列中的某一元素  $H(j) > R_{ave}$ , 则符号  $L(j)$  为“+”, 否则为“-”, 由此建立符号序列  $L(i)$ . 由于只考虑了大于平均值和小于等于平均值这两种状态, 用二次粗粒化方法所产生的符号序列很可能会丢失原动力系统的一些重要信息. 由此, 本文提出了多次粗粒化的概念.

假设有一时间序列为  $H(i) (i = 1, 2, \dots, n)$ ,  $R_{max}$  和  $R_{min}$  分别为这一时间序列的最大值和最小值,  $L(i) (i = 1, 2, \dots, n)$  是一个与时间序列  $H(i)$  长度相同的空符号串, 令

$$p = (R_{max} - R_{min}) / q, \quad (2)$$

式中  $q$  为粗粒化的次数. 此时可以确定分割点  $a(j) (j = 1, 2, \dots, q-1)$  为

$$a(j) = R_{max} - p \times j \quad (j = 1, 2, \dots, q-1), \quad (3)$$

则数据点  $R_{max}, a(1), \dots, a(j)$  和  $R_{min}$  就将时间序列  $H(i)$  分成了  $q$  个层次,  $R_{max} \rightarrow a(1)$  为第一层次,  $a(1) \rightarrow a(2)$  为第二层次, 依此类推,  $a(q-1) \rightarrow R_{min}$

\* 国家重点基础研究发展规划 (批准号 2004CB418300) 和国家自然科学基金 (批准号 90411008, A0231006) 资助的课题.

† E-mail: hou-w@126.com

即为第  $q$  个层次,当  $q = 1$  时即为原时间序列.自定义一组符号序列  $g(k) \chi k = 1, 2, \dots, q$ ,当  $m \neq n$  时,

$$g(m) \neq g(n) \quad (m, n = 1, 2, \dots, q). \quad (4)$$

当时间序列中的某个元素  $H(j)$  处于第  $k$  个层次中,符号  $L(j)$  就用  $g(k)$  来表示,由此建立相应的符号序列  $L(i)$ .本文采用三次粗粒化的方法,即  $q = 3$ .

### 3. 复杂度的定义及其物理意义

下面对 Lempert-Ziv 复杂度的算法则进行扼要介绍<sup>[5]</sup>.

设由有限个符号组成的长度为  $n$  的符号序列为  $s_1 s_2 s_3 \dots s_n$ ,向一个空符号串  $\epsilon$  中添加  $s_1$ .假设已经向  $\epsilon$  中加入  $s_1 s_2 s_3 \dots s_{r-1}$ ,并且下一个符号  $s_r$  是用添加操作加入到  $\epsilon$  中的,记为  $s_1 \cdot s_2 s_3 \dots s_{r-1} s_r \cdot$ ,这里  $s_1$  和  $s_r$  后面的符号“ $\cdot$ ”表示  $s_1$  和  $s_r$  是由添加操作生成的.

具体的演化法则是令  $Q = s_{r+1} \cdot \epsilon = s_1 s_2 s_3 \dots s_r \cdot$ ,观察  $Q$  是否为  $\epsilon Q \pi$  的子串. $\pi$  表示的操作为:将  $\pi$  前面符号串中的最后一个操作去掉,即  $\epsilon Q \pi = s_1 s_2 s_3 \dots s_r \cdot$ .如果  $Q$  不能从  $\epsilon Q \pi$  中的某个子串复制得到,就向  $\epsilon$  中用添加操作加上  $s_{r+1}$ ,并加上一个符号“ $\cdot$ ”,再令  $Q = s_{r+2} \cdot \epsilon = s_1 s_2 s_3 \dots s_{r+1} \cdot$ 并重复上述操作.如果  $Q$  可以从  $\epsilon Q \pi$  中的某个子串复制得到,则令  $Q = s_{r+1} s_{r+2} \cdot$ ,这时  $\epsilon Q \pi = s_1 s_2 s_3 \dots s_r s_{r+1} \cdot$ .如果  $Q$  仍然可以从  $\epsilon Q \pi$  的某个子串复制得到,则再考虑  $Q = s_{r+1} s_{r+2} s_{r+3} \cdot$ 并重复上述操作.这样下去有两种结果,第一种结果是  $Q$  已包含了原来给定的长度为  $n$  的符号序列的最后一个符号,则分析结束.第二种结果是对于某个  $Q$ ,它再也不能从  $\epsilon Q \pi$  中的任何一个子串复制得到,这时就采取添加操作,将这个  $Q$  的最后一个符号添加上,并在它后面加上符号“ $\cdot$ ”.可见,符号“ $\cdot$ ”的个数即为采取添加操作的次数.如果符号串  $\epsilon$  在上述分析结束时以符号“ $\cdot$ ”结束,则“ $\cdot$ ”的个数就等于符号串的复杂度,否则将符号“ $\cdot$ ”的个数加 1 就得到复杂度.亦即对于给定的符号串被符号“ $\cdot$ ”分成段的数目就定义为复杂度.这个复杂度实际上就是为了描述给定的符号序列所需要最少的、互不相同的“子串”个数,这些子串一般称为“词”.显然所需“词”越多,则符号序列越复杂,描述它就需要越多的子串.

一个经过了复制和添加操作的符号序列  $s_1 s_2 s_3$

$\dots s_n \rightarrow s_1 \cdot s_2 s_3 \dots s_{r-1} s_r \cdot s_{r+1} \dots \cdot$ ,此序列的复杂度为 3.可以解释为:两个“ $\cdot$ ”号之间的符号串,例如  $s_2 \dots s_r$  是这个符号序列的一个新模式.这里的“模式”就是上述的“词”;“新模式”就是向给定符号序列中采取添加操作所加入的“新词”,它不可以从符号串  $s_1 \dots s_{r-1}$  中复制得到,而  $s_2 \dots s_k (k = 3, 4, \dots, r-1)$  可以从符号串  $s_1 \dots s_{r-1}$  中复制得到,是以前已经存在的某种模式的“翻版”或“在当前的周期性再现”.一个符号序列的复杂度越大,说明它的添加操作越多,新模式也越多,描述给定符号序列所需最少的、互不相同的“子串”个数越多,给定的符号序列周期性越弱,出现新模式的速率越快.反之则复制操作越多,新模式越少,周期性越强,出现新模式的速率越慢.关于复杂度的物理机制可以参见文献 [8—9].现在普遍认为复杂度的物理意义在于复杂度反映了一个时间序列随着序列长度的增加出现新模式的速率<sup>[9]</sup>,即复杂度代表了数据在窗口长度时期内随时间出现新变化(新模式)的速率,两个“ $\cdot$ ”之间的时段表明这一时段的数据变化形式在此之前是没有的.复杂度越大,说明数据在窗口长度时期内随时间出现的新变化越多,发生新变化的速率越快,表明这一时期的数据变化是无序而复杂的.反之,复杂度越小,则说明发生新变化的速率越慢,数据变化是规则的,周期性越强.所以,计算一个时间序列的复杂度能够描述出系统的状态发生变化的速率,这一点是非常重要的.

### 4. 复杂度的演变特征

本文中复杂度的计算采用滑动窗口的方法:对一个长度为  $N$  的时间序列,从资料的第一个数据开始,首先取一长度为  $L$  的矩形窗口的数据并计算窗口内数据的复杂度,然后将窗口向后滑动,滑动步长为  $M$  个点,得到下一组数据并计算其复杂度,重复上述步骤直至全部数据的最后一点;每个窗口计算得到的复杂度赋值给窗口内最后一点,将各个窗口计算得到的复杂度依次连接,由此建立复杂度序列.

#### 4.1. logistic 映射的复杂度演变特征

logistic 映射随时间的演化是人们最熟知的非线性动力学系统,logistic 映射可以写为  $x_{n+1} = ux_n(1 - x_n) \chi n = 0, 1, 2, \dots, x_n \in [0, 1], u$  为控制参数.如图 1(a)所示,当  $u \in [0, 1)$  时,存在  $x_n \rightarrow 0$  的稳定不动

点 当  $x_n \in [1, 3)$  时, 存在  $x_n \rightarrow 0$  的不稳定不动点和  $x_n \rightarrow 1 - 1/u$  的稳定不动点, 周期 1 解; 当  $u = 3$  时, 周期 1 解分叉为周期 2 解; 当  $u \in [3, u_\infty)$  时, 此时的  $u_\infty = 3.569945672$ , 依次出现轨道周期变化  $1 \rightarrow 2 \rightarrow 4 \rightarrow 8 \rightarrow 16 \rightarrow \dots$ , 即倍周期分叉现象; 当  $u \in (u_\infty, 4)$  时出现混沌现象. 从图 1(b) 中可以看到, 当时间序列依次经历稳定不动点  $\rightarrow$  不稳定不动点  $\rightarrow$  周期  $\rightarrow$  混沌四个不同的演化阶段时, 复杂度明显地依次变大, 而且混沌区域中的周期窗口在复杂度序列中也有明显的反映, 并且复杂度的变化与时间序列复杂程度的变化步调一致. 由此可见, 复杂度可以鉴别出

一个时间序列中不同的演化特征.

将窗口长度和滑动步长取不同的值进行组合以及在原时间序列中加入噪声, 此时计算得到的复杂度的演变特征均类似于图 1(b).

### 4.2. Lorenz 模型的复杂度演变特征

运用复杂度分析方法分别计算了 Lorenz 模型的  $x, y, z$  三个分量的复杂度. 图 2 是 Lorenz 模型的  $x$  分量在滑动步长为 5, 窗口长度分别取 1100, 851 和 750 时的复杂度演化曲线. 图 3 是 Lorenz 模型的  $x$  分量在窗口长度为 851, 滑动步长分别取 9, 6 和 3 时

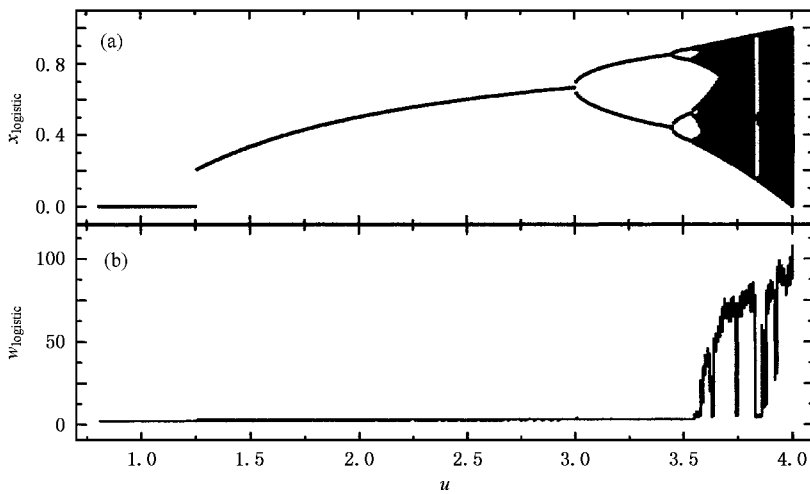


图 1 logistic 映射的图像 (a) 和 logistic 映射复杂度 (b)  $x_{\text{logistic}}$  为 logistic 映射的值,  $w_{\text{logistic}}$  为复杂度的值,  $u$  为控制参数. 窗口长度 1000, 滑动步长 5

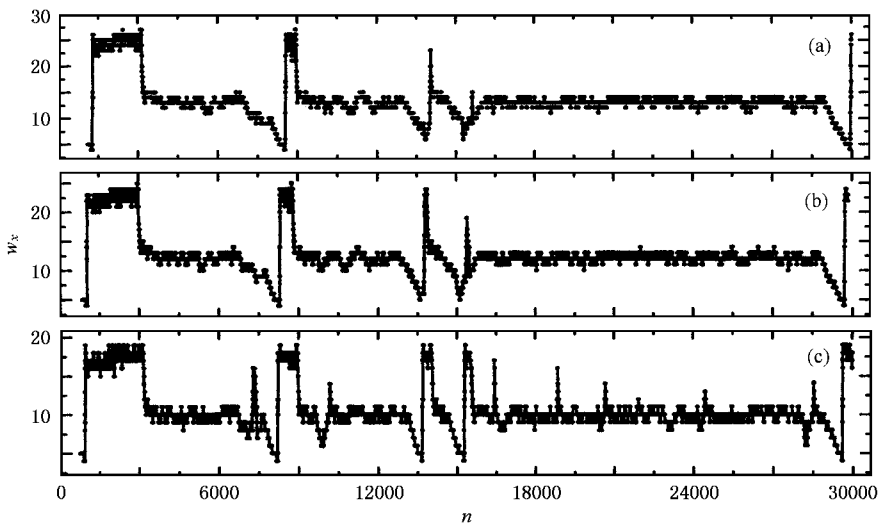


图 2  $x$  分量在滑动步长取 5 时不同窗口长度的复杂度演化曲线  $w_x$  表示复杂度,  $n$  表示数据的样本量. (a) 窗口长度 1100 (b) 窗口长度 851 (c) 窗口长度 750

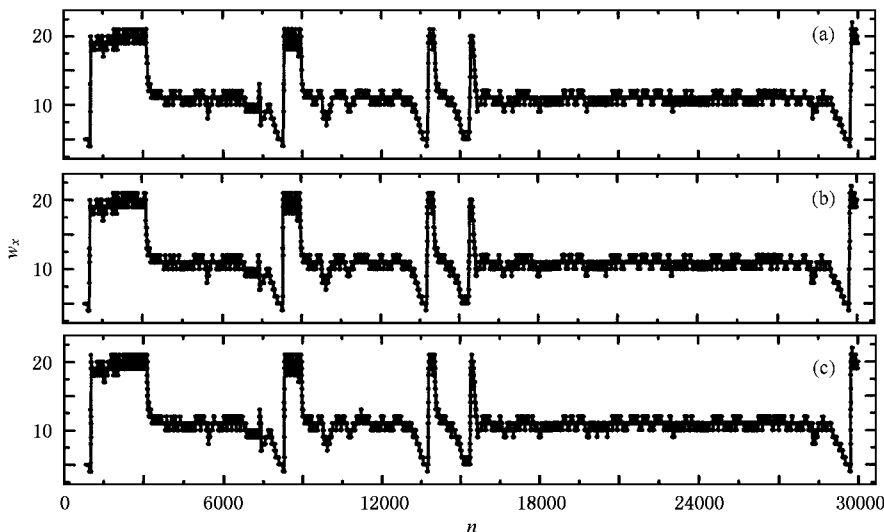


图 3  $x$  分量在窗口长度取 851, 不同滑动步长时的复杂度演化曲线  $w_x$  表示复杂度,  $n$  表示数据的样本量. (a) 滑动步长 9 (b) 滑动步长 6 (c) 滑动步长 3

的情况. 从图 2 和图 3 可以看出, 无论是取不同的窗口长度还是取不同的滑动步长, 复杂度的特征均是相似的, 这充分说明了复杂度分析方法及其计算结果是十分可靠的. Lorenz 模型的  $y$  分量和  $z$  分量的复杂度在取不同的窗口长度和滑动步长时的特征与  $x$  分量的情形一致. 鉴于此, 本文在下面的分析中统一采用窗口长度为 1000, 滑动步长为 5.

在图 4 (b) 中可以看到  $x$  分量复杂度的演化特征. 经过对比研究发现: 当  $x$  分量长时间连续围绕零点左侧的吸引子旋转时, 复杂度出现极大值  $B, D, F, H, J$ ; 当  $x$  分量长时间连续围绕零点右侧的

吸引子旋转时, 复杂度出现极小值  $A, C, E, G, I$ ; 当  $x$  分量从右侧吸引子跳到左侧吸引子时, 复杂度的变化迅速从极小值跃迁到极大值, 例如  $A$  到  $B, C$  到  $D$  等; 当  $x$  分量从左侧吸引子跳到右侧时, 复杂度的变化迅速从极大值降到 13.5, 并围绕该值上下振荡, 振荡的时间有长有短, 然后降到极小值, 例如  $B$  到  $C, D$  到  $E$  等. 这充分说明时间序列的演化可以通过复杂度的特征来进一步研究其动力结构特征. 图 5 为  $y$  分量时间序列及其复杂度的演化特征, 与  $x$  分量类似.

图 4 (a) 为向 Lorenz 模型的  $x$  分量中加入噪声

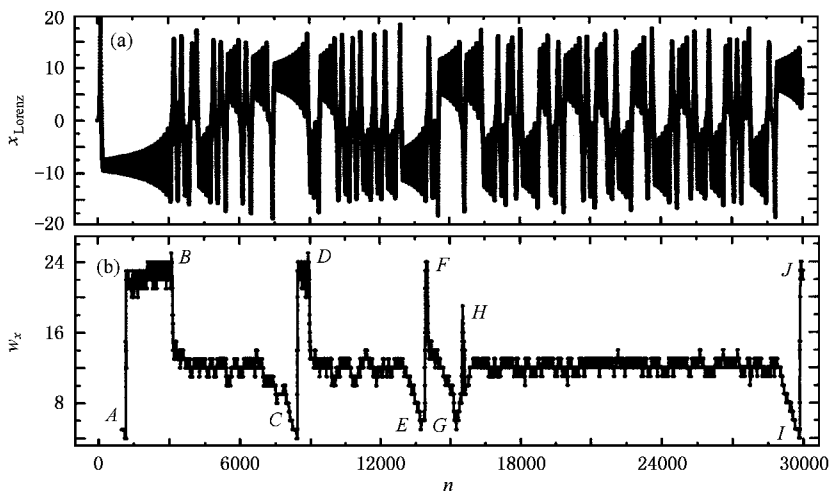


图 4 取 30000 个点时 Lorenz 模型的  $x$  分量时间序列 (a) 和  $x$  分量时间序列的复杂度 (b)  $x_{Lorenz}, w_x$  分别表示 Lorenz 模型的  $x$  分量值及其复杂度,  $n$  表示数据的样本量

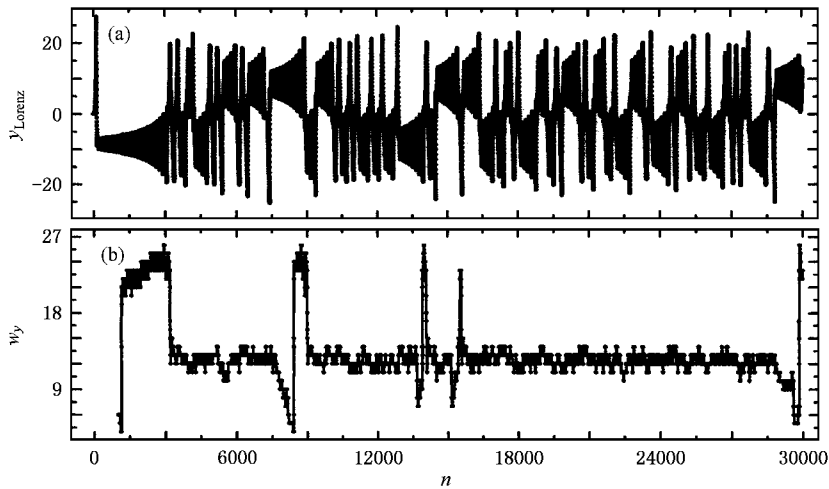


图 5 取 30000 点时 Lorenz 模型的  $y$  分量时间序列 (a) 和  $y$  分量时间序列的复杂度 (b)  $y_{\text{Lorenz}}$ 、 $w_y$  分别表示 Lorenz 模型的  $y$  分量值及其复杂度,  $n$  表示数据的样本量

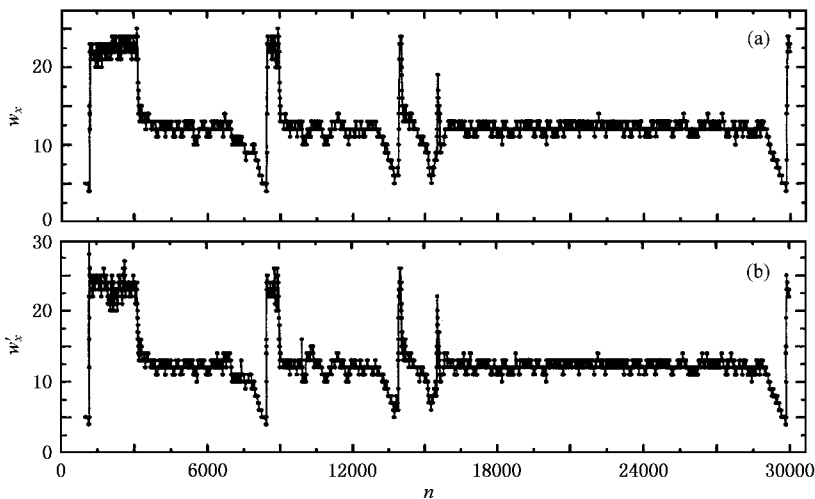


图 6 没有向 Lorenz 模型的  $x$  分量中加入噪声时 (a) 以及加入噪声时 (b) 的复杂度  $w_x$  为不加入噪声时的复杂度,  $w'_x$  为加入噪声后的复杂度,  $n$  表示数据的样本量

时计算得到的复杂度 (图 6 (b)) 为不加入噪声时的情况. 从图 6 可以看出, 两者的基本特征几乎完全一样. 向  $y$  分量中加入噪声时得到的复杂度的特征与不加入噪声时的情况同样也几乎一样 (图略). 计算 Lorenz 模型的  $x$  分量、 $y$  分量更长数据的复杂度, 此时复杂度所表现出来的性质与图 4 和图 5 中的情形一致.

## 5. 结 论

本文根据三次粗粒化方法得到了 logistic 映射和 Lorenz 模型的符号序列, 运用 Lempel-Ziv 复杂度

算法, 计算并分析了这两个时间序列的复杂度. 发现 logistic 映射的复杂度反映了时间序列的演化特征, 并且混沌区域中的周期窗口在复杂度序列中有很好的对应关系. 对于 Lorenz 模型, 复杂度同样反映了时间序列的演化特征.  $x$ 、 $y$ 、 $z$  三个分量的复杂度序列均具有混沌性质, 即由许多振幅非常接近而长度完全不同的循环所组成, 反映了 Lorenz 模型内在的准周期特性. 发现各分量在两个奇怪吸引子之间跃迁时, 当各分量长时间连续围绕零点左侧的吸引子旋转时, 复杂度出现极大值; 当各分量长时间连续围绕零点右侧的吸引子旋转时, 复杂度出现极小值; 当各分量从右侧吸引子跳到左侧吸引子时, 复杂度的变

化迅速从极小值跃迁到极大值 ;当各分量从左侧吸引子跳到右侧时 ,复杂度的变化迅速从极大值降到 13.5 左右 ,并围绕该值上下振荡 ,振荡的时间有长有短 ,然后降到极小值 .这充分说明了时间序列的演化可以通过复杂度的特性来进一步研究其动力结构

特征 .

事实上 ,许多观测数据的动力学方程是未知的 ,可以通过计算其复杂度初步判断时间序列的特性 ,这有利于其动力学结构的反演<sup>[10-13]</sup> .

- 
- [ 1 ] Paolo G , Patti H , Jim R *et al* 2004 *Chaos Solitons and Fractals* **20** 1
- [ 2 ] Hildegard M O 2004 *Physica A* **337** 679
- [ 3 ] Martin M T , Plastino A , Rosso O A 2003 *Phys. Lett. A* **311** 126
- [ 4 ] Wang P , Dai X G 2004 *Chin. Phys.* **13** 1770
- [ 5 ] Deng Z W , Min J Z , Zhang Y 2001 *Journal of Nanjing Institute of Meteorology* **24** 186 ( in Chinese ) [ 邓自旺、闵锦忠、张 勇 2001 南京气象学院学报 **24** 186 ]
- [ 6 ] Hou W , Feng G L , Gao X Q *et al* 2005 *Acta Phys. Sin.* **54** 2441 ( in Chinese ) [ 侯 威、封国林、高新全等 2005 物理学报 **54** 2441 ]
- [ 7 ] Feng G L , Dong W J , Li J P 2004 *Chin. Phys.* **13** 1582
- [ 8 ] Kaspar F , Schuster H G 1987 *Phys. Rev. A* **36** 842
- [ 9 ] Zheng J Y , Ge Q S , Zhang P Y 1999 *Adv. Earth Sci.* **14** 177 ( in Chinese ) [ 郑景云、葛全胜、张丕远 1999 地球科学进展 **14** 177 ]
- [ 10 ] Feng G L , Dong W J , Jia X J 2004 *Chin. Phys.* **13** 413
- [ 11 ] Feng G L , Dong W J 2003 *Acta Phys. Sin.* **52** 2347 ( in Chinese ) [ 封国林、董文杰 2003 物理学报 **52** 2347 ]
- [ 12 ] Feng G L , Dong W J , Li J P *et al* 2004 *Acta Phys. Sin.* **53** 2389 ( in Chinese ) [ 封国林、董文杰、李建平等 2004 物理学报 **53** 2389 ]
- [ 13 ] He W P , Feng G L , Dong W J *et al* 2005 *Chin. Phys.* **14** 21

# Investigation about the Lorenz model and logistic equation based on the complexity<sup>\*</sup>

Hou Wei<sup>1)</sup> Feng Guo-Lin<sup>2)B)</sup> Dong Wen-Jie<sup>2)</sup>

<sup>1)</sup>*College of Physics Science and Technology, Yangzhou University, Yangzhou 225009, China*

<sup>2)</sup>*Key Laboratory of National Meteorological Bureau, Climate Research Center, Beijing 100081, China*

<sup>3)</sup>*Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China*

(Received 27 October 2004 ; revised manuscript received 28 February 2005)

## Abstract

The complexity series of the logistic equation and the Lorenz model were respectively calculated using a dynamical nonlinear analysis method for time series——Lemper-Ziv complexity algorithm, and the physical implication of Lemper-Ziv complexity is also discussed. Results show that for the logistic equation, the complexity is obviously different when the complicated degree of the time series is variational; and for Lorenz model, i. e. its  $x$ -,  $y$ -, and  $z$ -portions, their complexities are all chaotic and are all composed of many cycles whose swings are almost the same and the lengths are different. The result reflects the internal quasi-periodicity. Further investigations indicate that when different window lengths are selected, the characters of the complexity series for a given time series are basically the same, and there exists a coherency between the jumps of the complexity series and the jumps of the time series. Thus one can judge the characteristic of a time series by calculating its complexity. This may be useful to predict the kinetic behavior of a time series.

**Keywords** : thrice coarse-graining, Lemper-Ziv complexity, logistic equation, Lorenz model

**PACC** : 9260X

<sup>\*</sup> Project supported by the State Key Development Program for Basic Research of China ( Grant No. 2004CB418300 ) and the National Natural Science Foundation of China ( Grant Nos. 90411008 , 40231006 ).