

从时间序列中提取维数信息

罗 勇

(国家气候中心, 北京 100081)

提 要

作者介绍了分形和分维的概念、性质和意义, 重点讨论了从一维时间序列中提取维数信息的方法, 并给出了一些在天气和气候中的应用。

关键词: 分形与分维 ENSO 系统 时间序列

1 何谓分形与分维

这方面的研究是由一个著名的问题开始的。1977 年美国科学家 B. B. Mandelbrot 在他的一部经典性著作中提出: “英国的海岸线有多长”的问题。这个问题很难回答, 因为答案显然依赖于测量时所使用的尺度。这里我们不妨不考虑涨潮落潮这些使海岸线无法严格定义的因素。如果用公里作测量单位, 从几米到几十米的弯曲就会被忽略, 若改用米作测量单位, 测出的总长度会增加, 但是一些厘米量级的弯曲还是不能准确地反映出来。象海岸线、闪电、树木、山等自然界许多不规则的复杂现象, 我们在对它们进行描述时遇到了困难。Mandelbrot 提出可用分数维(或称分维, Fractal Dimension)来描述。

我们知道, 规整的几何对象具有整数维数, 例如零维的点, 一维的线, 二维的面, 三维的体积。维数和测量有着密切的关系。例如把单位长度的线段放大 l 倍, 则得到 $l^1 = l$ 个原来的线段; 同样, 对于二维的面, 如果把每个方向的单位长度都放大 l 倍, 则得到 l^2 个原来的面积; 类似地, 对于 D 维尺度的几何对象, 放大 l 倍则得到 N 个原来的对象, 即 $l^D = N$ 。不难验证, 对于一切规整的几何对象, 上式都成立。现将上式两边取对数, 写成:

$$D = \ln(N)/\ln(l) \quad (1)$$

这样, 我们就完成了一次飞跃, D 不必再

为整数。以后把这样推广定义的维数称为分维。具有分数维的几何对象称为分形(Fractal, 或称分维几何, 碎形)。Fractal 源自拉丁文 Fractus, 意思是不规则或破碎。

对于不规则的分形, 我们不是把对象放大, 而是将测量单位缩小为原来的 ϵ 分之一, 即 $l = \frac{1}{\epsilon}$ 。这时分维定义要修改成:

$$D_0 = \lim_{\epsilon \rightarrow 0} \frac{\ln N(\epsilon)}{\ln(\frac{1}{\epsilon})} \quad (2)$$

这里, D_0 称为 Kolmogorov 容量或 Hausdorff 维数。即要看测量单位不断缩小时, 测得的结果有无极限存在。

应该指出, l 不能取得太小; 若太小就不能判断出分维结构, 只能得出是整数维。 l 也不能取得太大; 若太大就会得出对象是零维的物体。实际上上式定义的过程是 $l \rightarrow \infty$ 的一种极限过程。但对于实际存在的对象, 总存在一个最小的特征长度 ϵ , 再比 ϵ 小时, 该对象就不能看成是分形了。

2 一些典型分形的分维

首先是一个 Cantor 集合的例子。它是由一单位长度的线段, 分成 3 段而去掉中间一段, 然后在剩下的两段中间仍按同样的办法处理下去而不断形成的点集(图 1 左)。Cantor 集合最后由无数个点组成, 显然其维数应比零维(一个点)大; 但因为点数虽多, 却又未

充满一条直线,中间还有很多空隙,故维数又

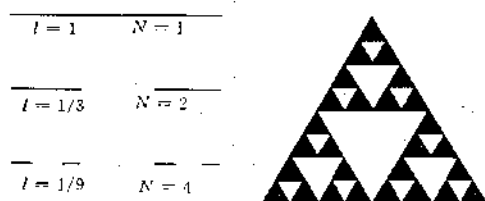


图1 Cantor 集合(左)与 Sierpinski 垫片(右)

应比1小。按(2)式的定义,其维数为:

$$D_0 = \frac{\ln 4}{\ln 9} = 0.6309$$

再看一个称为 Sierpinski 垫片的例子(图1右)。它是由一个正三角形分成相等的4块后,挖去中间一块而不断形成的。因此用 $l = \frac{1}{2}$ 的小三角形去量它,得到的数目 $N=3$,故其维数为

$$D_0 = \frac{\ln 3}{\ln 2} = 1.5850$$

现在回到最初的问题,看看海岸线的维数是多少。图2是几种海岸线形成的模型。它们的形态当 l 很小时,变得非常复杂,长度趋向很大,因而维数应大于1,但又未能充满平面,所以维数又低于2。实际的海岸线维数在1.2附近^[1]。这时,我们知道“海岸线有多长”的提法不严格,而应问“海岸线有多复杂”。而分维就是这种复杂性的量度。

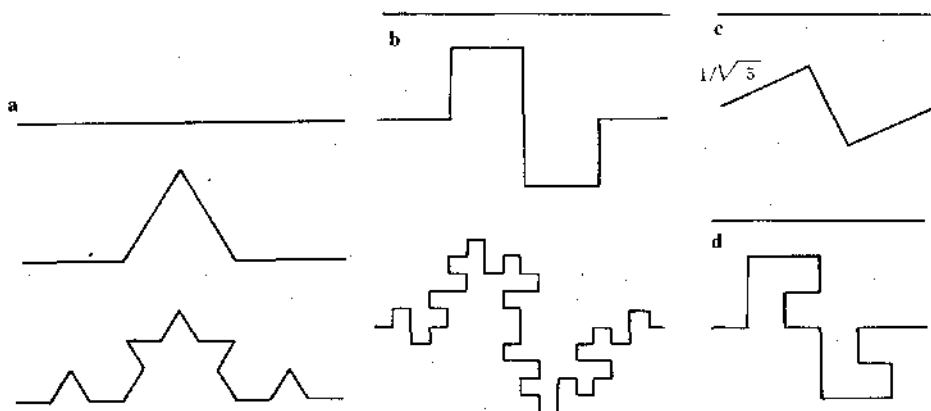


图2 几种海岸线形成的模型及其维数

$$a. D_0 = \frac{\ln 4}{\ln 3} = 1.2619 \quad b. D_0 = \frac{\ln 8}{\ln 4} = 1.5 \quad c. D_0 = \frac{\ln 3}{\ln \sqrt{5}} = 1.3652 \quad d. D_0 = \frac{\ln 18}{\ln 6} = 1.613$$

上面介绍的是几个由简单构造规则形成的分形。它们的分维也很容易计算。自然界更为常见的分形,是具有自相似分布特征的随机对象,其维数计算也较困难,一般用“数盒子”的方法。比如布朗运动的维数是2,云彩边界的分维大约是1.35。

3 分形的性质及其维数的意义

由上面的讨论可以将分形的性质概括为:(1)具有自相似性;(2)具有无穷多内部层

次;(3)有些是“有皮无肉”(多孔),对应维数降低。如 Sierpinski 垫片;(4)有些是多皱、弯曲(不光滑),对应维数升高。如各种海岸线模型。

现已知道,混沌吸引子正具有这些特点。因此,我们不妨将混沌吸引子也看作是一种分形。而如果我们计算出吸引子的维数是分数,则可说明其为混沌吸引子。

计算分维的意义在于,一个吸引子的维

数是说明该吸引子上点位置所必需的信息量,也是模拟一个系统的动力学性态所必需的基本变量的最低限数目。例如,连续流的 Lorenz 混沌吸引子的维数为 $D_0 = 2.07$ (当 $\sigma = 10, r = 45.92, b = 4.0$ 时),这是由于它是未充满三维空间的有间隙的流。

4 从时间序列中提取维数信息

大多数情形下,人们往往希望从实验数据中得到关于系统维数的估计。但通常的实验并不记录一个系统的所有方面,而仅仅是把它投影到更低维的平面或直线上来观测,即只测量一两个变量的数据序列。此时,就需要在不知道背景相空间维数的情况下,根据有限的数据来取得关于维数的信息,重现系统的动态特性。

近年来发展了一种非常简便实用的统计方法。我们知道象 ENSO 这样的复杂系统,其基本变量数目很多,而且彼此之间是耦合

的。系统任一变量的演化是由与之相互作用的其它变量所决定的。因此,可以认为这些相关变量的信息就隐含在任一变量的发展过程中,故在考察这一系统时,只需研究其中一个变量随时间变化的序列。即为了重构出一个“等价”的相空间,只需考察一个变量,并将在某些适当延迟点上的观测值作为新维处理(也就是将延迟点上的值作为新的坐标),它确定了某个多维状态空间中的一点。对不同的时间点重复这一步骤,可以得到重构相空间(或称嵌入空间)中一系列这样的点。然后通过计算嵌入空间的维数,就能获得这一系统的主要性态。F. Takens 从数学上证明了只要嵌入空间维数足够大,重建的系统与原来的动力系统的几何性质是等价的。

首先,我们由原始时间序列(记作 $X(t)$)重构相空间 R^p 。 R^p 中的点集 $Y(t)$ 与 $X(t)$ 的关系为

$$Y(t_1) = \{X(t_1), X(t_1 + \tau), X(t_1 + 2\tau), \dots, X(t_1 + (p-1)\tau)\}$$

$$Y(t_2) = \{X(t_2), X(t_2 + \tau), X(t_2 + 2\tau), \dots, X(t_2 + (p-1)\tau)\}$$

⋮

$$Y(t_l) = \{X(t_l), X(t_l + \tau), X(t_l + 2\tau), \dots, X(t_l + (p-1)\tau)\} \quad (3)$$

式中角标 l 为序列 $Y(t)$ 在 R^p 空间中的点数, τ 为适当的滞后时间, p 为嵌入空间的维数。三者之间的关系为

$$l \leq N - (p-1)\tau \quad (4)$$

其中, N 为原序列 $X(t)$ 的点数。

然后,定义关联函数 $C(r)$

$$C(r) = \lim_{l \rightarrow \infty} \frac{1}{l^2} \sum_{i,j=1}^l H(r - |Y_i - Y_j|) \quad (5)$$

式中, $z = r - |Y_i - Y_j|$ 表示 R^p 空间中两个点 Y_i 和 Y_j 的距离与函数自变量 r 的差, $H(z)$ 是 Heaviside 函数

$$H(z) = \begin{cases} 1 & z \geq 0 \\ 0 & z < 0 \end{cases} \quad (6)$$

对于适当的 r , 可能在 r 的一段区间内, $C(r)$ 随 r 的变化满足

$$C(r) \propto r^{D_2} \quad (7)$$

即 $C(r)$ 与 r 的指数幂成正比, 指数 D_2 是双

对数 $\ln C(r) - \ln(r)$ 图上直线部分的斜率, 即

$$D_2 = \frac{\ln C(r)}{\ln(r)} \quad (8)$$

不妨定义满足式(7)、(8)的 D_2 为关联维数。可以看出关联维数与最初定义的测度学上的维数是一致的。变化的 r 相当于测量的尺度 l , 而关联函数 $C(r)$ 则相当于测量的结果 $N(l)$ 。应该指出, 对于过大的 r , 当然所有点对的距离都不会超过它, 故有 $C(r) = 1$, $\ln C(r) = 0$ 。太大的 r 反映不出系统的内部性质。如果 r 取得太小, 实验中一切偶然的噪声都会表现出来, 由于噪声在任何一维上都起作用, p 维空间中就测得 $D_2 = p$ 。 D_2 具有描述一个吸引子结构特征的能力。对均匀分布在一条线上的资料点, 有 $C(r)$ 和 r 的一次方成正比, 即 $D_2 = 1$; 而均匀分布在一个面上的资料点, 有 $C(r)$ 和 r 的二次方成正比, 故 D_2

=2。

嵌入空间维数 p 增加时,其斜率 D_2 也会增大。最后当达某一个 p 值时, D_2 达到饱和值。这时 D_2 就反映了该吸引子的结构特征,而 p 则是描述该混沌吸引子的最小状态空间变数的个数,称为饱和嵌入空间维数。根据 Whitney 嵌入定理^[2],嵌入空间维数 p 必须很大,它和实际的 D_2 维吸引子之间要满足

$$p = 2D_2 + 1 \quad (9)$$

综上所述,由时间序列计算分维的步骤可以归纳为:(1)相空间重整;(2)计算关联函数和关联维数。

5 ENSO 系统的维数估计

由于 ENSO 系统在全球气候变化研究中的重要地位,近年来计算其维数的工作较多。考虑到 ENSO 的主要特征之一是海表面温度异常,故本文选择 1949—1979 年逐月的 SST 资料,计算 ENSO 的分维。共选择了中东太平洋上 5 个点,即 0°N 、 165°W 、 10°N 、

150°W 、 0° 、 130°W 、 5°S 、 130°W 和 10°S 、 130°W 。

实际计算过程中,嵌入空间 R^p 的维数 p 从 2 变到 12。图 3 是 0° 、 165°W 的计算结果。由式(5)、(6)可知,当 $r \geq \max_{i,j=1}^l |Y_i - Y_j|$ 时, $C(r) = 1$; 当 $r \leq \min_{i,j=1}^l |Y_i - Y_j|$ 时, $C(r) = 0$, 即 $0 \leq C(r) \leq 1$ 。因此,图中各曲线 $\ln C(r) \leq 0$ 。

从图 3 中看到,随着 $\ln(r)$ 增大, $\ln C(r)$ 值由急升转缓升,而后趋向 0。这说明关联函数 $C(r)$ 随 r 的增大呈指数增长且趋近 1。注意到图中曲线中间一段近于直线,在它对应的 r 值区间内直线的斜率更符合式(8),即其斜率就是关联维数 D_2 。随着嵌入空间维数 p 的增大, D_2 也不断增大。当 $p=9$ 以后,相对应于 $p=10, 11, 12$ 的 3 条曲线大体是平行的,斜率 D_2 趋于饱和值 4.9。图 4 示出这一趋近过程。我们把此时的饱和值作为 ENSO 系统分维的近似值。

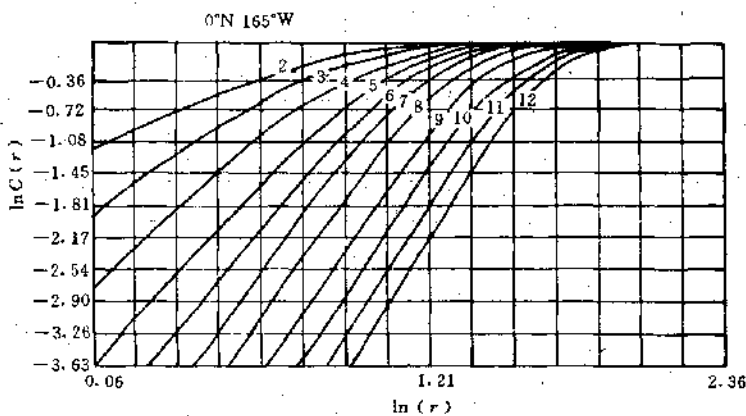


图 3 0° 、 165°W SST 的 $\ln C(r) - \ln(r)$ 双对数图

对其它网格点的计算,得到类似的结果(图略),见表 1。计算结果表明,ENSO 系统的维数在 4.9—5.7 之间。由于现有的 SST 资料对 FD 的计算并不足够长,本文不可能做更精细的计算,不同网格点的 FD 值也稍

有差别。但即使如此,对 ENSO 系统的主要性态还是有所反映。

(1) ENSO 是一个混沌系统。比较图 3 和图 5、6,可以看出 ENSO 系统与确定性系统、随机系统具有显著的不同。图 5 和图 6 分别

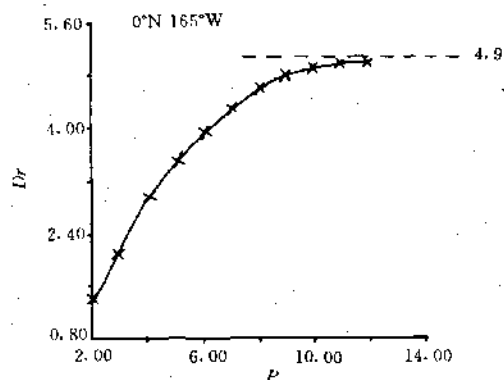


图4 0°N、165°W SST 的关联维数 D_2 随嵌入空间维数 p 的变化

表1 各网格点关于 ENSO 系统 FD 的估计结果

网格点	0°, 165°W	10°N, 150°W	0°, 130°W	5°S, 130°W	10°S, 130°W
FD	4.9	5.7	5.1	5.4	5.3

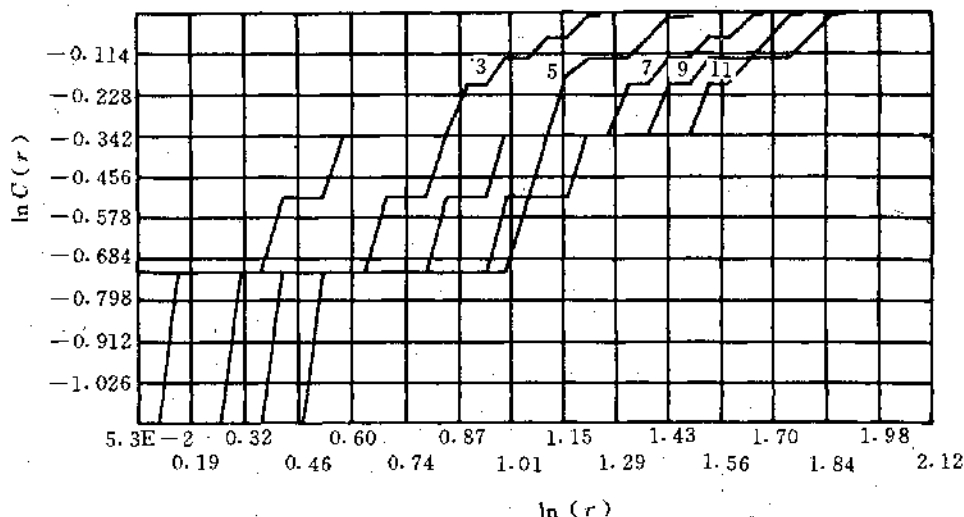


图5 正弦函数的 $\ln C(r) - \ln(r)$ 双对数图

6 分形与分维在天气和气候中的应用

由时间序列中提取维数信息的方法,在很多领域得到广泛应用。如一些研究表明,地震前的维数降低;人在闭目静坐时脑电波的维数为 2.4,而数数时上升到 3.0 左右;经济学领域的维数小于 2^[1]。

近年来,分形与分维在天气和气候研究

是正弦函数序列与白噪音的 $\ln C(r) - \ln(r)$ 双对数图,按相同的步骤计算得到。对于确定性过程,曲线全部平行,斜率不随 p 变化,曲线呈阶梯状;对于随机过程,曲线斜率不断增大,随 p 大致呈线性变化。因此,可以认为 ENSO 系统中含有浑沌吸引子。

(2) 由分维的意义,客观描述 ENSO 系统需要 5—6 个基本变量。这对构造 ENSO 的动力学模型具有指导意义。

(3) 通过分形和分维方法,可以依靠系统某一方面的观测资料对它的动力学性态作出判断。在很多情形下,分形与分维都是一种既简单又实用的方法。

中的应用也得到迅速开展。杨培才和陈烈庭(1990)利用近百年达尔文站月平均气压距平值(DP)的时间序列及近 50 年达尔文与塔希堤两站月平均气压距平差值(DTP)的时间序列,计算出 ENSO 的维数,如表 2 所示^[3]。其中,DP3、DP6 和 DP9 表示对 DP 序列作 3、6 和 9 个月的加权滑动平均。表 3 还给出了其它一些例子。

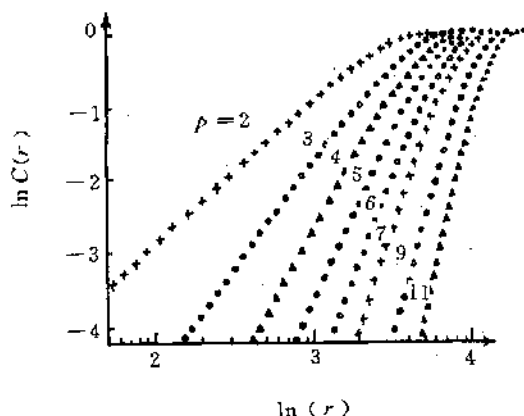
图6 白噪音的 $\ln C(r) - \ln(r)$ 双对数图

表2 由近百年达尔文站月平均气压距平与近50年达尔文和塔希提两站月平均气压距平差值时间序列计算出的 ENSO 的维数

系统	关联维数	饱和嵌入维数
DP	6.8 ± 0.1	15
DP3	6.2 ± 0.1	15
DP6	5.7 ± 0.1	14
DP9	5.5 ± 0.1	14
DTP	6.6 ± 0.1	15

表3 其它一些分形与分维的计算在天气与气候中的应用

资料及作者	样本容量	采样间隔	关联维数	饱和嵌入维数
深海岩芯 V_{23-238} (太平洋) C. Nicolis 等 ^[4]	约 500	2000 年	3.1	4
氧同位素 $\delta^{18}O$ (大西洋) K. Fraedrich ^[5]	182	2000—4000 年	4.4	约 12
逐日地面气压 (柏林) K. Fraedrich ^[5]	1680—1800	1 天	3.2—3.9	8—10
大气边界层风速和气温 (北京) 杨培才等 ^[6]	4096×4	1 秒	5.5—7.2	12—14
中层大气平均水平风速 (京都) 杨培才等 ^[7]	1536×2	150 秒	6.8	14
Sonic Anemometer 大气边界层水平风速 (北京) ^[8]	4096	1 秒	2.3	3

参考文献

- 1 陈顺等. 分形与分维在地球科学中的应用. 北京: 学术期刊出版社, 1988, 163.
- 2 刘式达, 刘式适. 非线性动力学和复杂现象. 北京: 气象出版社, 1989, 249.
- 3 杨培才, 陈烈庭. 埃尔尼诺/南方涛动的可预报性. 大气科学, 1990, 14: 64—71.
- 4 Nicolis C., Nicolis G. Is there a climatic attractor? Nature, 1984, 311: 529—532.
- 5 Fraedrich K. Estimating weather and climate predictability on attractors. J. Atmos. Sci., 1987, 44: 722—728.
- 6 Yang Peicai, Liu Jinli, Yang Shuwen, Some features of Chaotic motions in the atmospheric boundary layer. Presented at the International Conference on Bifurcation theory and Its Numerical Analysis, Xi'an (China), July 2—10, 1988.
- 7 Yang Peicai, Lu Daren, Wu Beiyang, et al. Chaotic features of the motion of upper troposphere and lower stratosphere revealed by the radar observation—case study. Presented at the Fourth MST Radar Workshop Kyoto (Japan), Nov. 29—Dec. 2, 1988.
- 8 郑祖光, 刘式达. 用大气湍流资料计算 Lyapunov 指数和分数维. 气象学报, 1988, 46: 41—48.

Computation of Fractal Dimension from Time Series

Luo Yong

(National Climate Center, Beijing 100081)

Abstract

The concepts and characteristics of fractals and fractal dimension are firstly introduced. And the computation method of fractal dimension of dynamical system from one variable time series $X(t)$ is mainly discussed. The computation technique of FD can be separated into two steps, reconstruction of phase space $R(p)$ from $X(t)$ and computation of relation integral and dimension for approaching to FD.

The FD of ENSO is calculated by SST data at some grid points in central and eastern equatorial Pacific during the period from 1951 to 1979. The results show that FD of ENSO is 4.9—5.7 and ENSO is therefore proved to be a complicated chaotic system according to which five or six independent variables are enough to describe. Finally FD of some weather and climatic attractors is given.

Key Words: fractals and fractal dimension ENSO time series