

文章编号: 1007-6301 (2001) 增刊-0014-15

基于扩展分形的地图信息自动综合研究

毋河海

(武汉大学资源与环境科学学院, 武汉 430079)

摘要: 源自地学的分形学不仅广泛地应用于多得难以想象的学科, 已成为一种新的方法论。然而分形学到底能为人们提供多少信息、有什么样的局限、如何与其它科学方法结合应用等, 是本文研究的主线。文中在作者研究扩展分维的基础上, 对扩展分维在地图信息自动综合中的应用问题进行了原理性的探讨。

关键词: 分数维; 扩展分数维; 自动综合; 多尺度空间数据库

中图分类号: P283 **文献标识码:** A

分数维的发现, 改变了人们观察世界的方式, 与此相关的分形思想已经成为一种新的方法论。目前, 大多数分形模拟方法都是用来生成各种逼真显示和生成复杂精美的、令人陶醉的艺术作品或景观图画。对于空间数据处理来说, 分形原理既与空间信息可视化有密切联系, 而且与空间数据建模和空间信息压缩也密切相关。对于这些领域来说, 一个共同的也是最基本的问题是利用分形这一有力工具来揭示自然现象或过程的结构特征, 对空间信息进行科学的描述与建模, 以达到控制与预测的目的。

1 分形学方法

尽管分形学的思想可追溯到 1919 年的 Hausdorff 著作 (甚至更早的 1875 年的 Weierstrass 等的著作), 但真正成为一门新兴学科, 应是 20 世纪 70 年代的事了。分形学诞生的触发事件应归属于地图量测问题“英国海岸线有多长”。有的学者称“分形学源于地学”, 笔者认为地图量测学扮演了“分形学之母”的角色。众多事实表明, 地学与地图中的多种地物与现象均是典型的分形对象。分形学源于地学, 但其应用领域迅速遍及各个学科, 现在, 应该是恰如其份地把分形原理应用于地学信息处理、特别是地图信息处理的时候了。

分形几何是用来描述难以用欧氏几何来描述的具有多层嵌套的自相似结构。严格的自相似只存在于数学演算中, 客观世界中不存在严格的、无限精细的自相似, 而存在的是统计性的、有限层次的自相似。地图内容要素, 特别是其中的自然要素如河流、海岸和地貌等, 具有嵌套性的复杂结构, 是典型分形体, 用一般数学方法难以恰当描述。而分形方法恰是用来描述这种不规则形体的科学方法。

收稿日期: 2001-07; **修订日期:** 2001-08

基金项目: 国家自然科学基金资助项目 (49971068)

作者简介: 毋河海, 教授, 博士生导师。主要从事地图学特别是自动制图综合的理论、方法和技术研究, 已发表论文 50 余篇, 专著近 10 部。

分形学的概念与尺度变化中的不变性有密切的联系。地图中的自然要素的轮廓图形,呈现出在一定尺度范围内的分形趋势。在这种情况下,就可利用表达不变性的分形参数——分数维来考察自然要素的图形特征。

分数维是度量自相似的特征量。但分数维的描述能力如何、究竟能为我们提供多少信息,是一个应该冷静考虑的问题。

2 分数维的扩展问题

2.1 单一分维数描述能力的局限性

基于分数维的分形学方法,其应用犹如雨后春笋,遍及难以想像的众多学科,表征着五彩缤纷的结构与现象。但与此相关的另一个问题——这种表征是否具有唯一性和完备性?正如不少学者所指出,不同的分形体会有相同的分维数,即存在“一值多形”的问题。分形学的创始人 Mandelbrot 曾说过, Himalays 与 JFK 的机场跑道可能有相同的分数维。我国学者敖力布、林鸿溢等也指出: Mandelbrot 虽然揭示了分形结构的自相似本质,并且用分维数来表示,可是这种描述并没证明它是唯一的和充分的。不同分布方式的两种结构可以具有相同分维数。……这说明要确定结构的特征,只有一个分维数是不行的,还需要其它参量。日本学者高安秀树在其名著“分数维”一书中写道:“分数维的维数是定量地表示自相似的随机形状和现象的最基本的量。因为仅利用分数维维数这样一个数字去描述所有的复杂形状和现象,无论如何也是不可能的。因此,就产生了扩大分数维维数的必要性。扩大分数维的考虑方法大致要两种。一种是不把分数维维数仅看作是一个常数,使其能有赖于观测的尺度,即使在自相似性不成立的那种范围内也能使用。另一种考虑方法是,在相似性成立的情况下,为了弥补只用分数维不能描述的信息,要重新引进另外的量”。张济忠、王桥也都论述了类似的问题。所以,仅利用分维数这样一个数值难以充分地描述各种具有不同复杂程度的形体和现象是学者们的共识。

2.2 观测信息的充分利用

客观世界不存在纯数学的理想分形,存在的只是统计意义下的随机分形。因此,自相似仅出现在一个确定的范围,这个范围叫无标度区,在无标度区内存在着明显的自相似。在通常的分析与应用中,首先要测定各种分形体的无标度区的上下界。为此目的,在测定过程中,所采用的量测尺度显然要超出无标度区的范围,获得了相应的无标度区之外的信息。显然,仅应用无标度区内部的信息是不够的,无标度区之外的观测信息有其相应的作用和价值,因而不能简单地丢掉。要想合理地应用无标度区之外的观测信息,就引起分维的扩充问题。分维的扩充旨在对无标度区内部与外部的大范围信息进行整体描述与分析应用。因此,就产生了扩展分数维的必需性。

本文作者曾用数值方法对分维的扩充进行了初步研究,并把它用于地图图形信息的分形描述与自动综合研究。

2.3 分维函数的理论表达

在分数维的最基本定义

$$D(r) = - \frac{\log N(r)}{\log r} \quad (1)$$

式中, 在函数 $N(r)$ 不是非常特殊的函数型(幂) 情况下, 由于该式的右端不能成不常数, 所以不能定义通常的分数维数。若把上式中的 $N(r)$ 的内涵扩大, 使其在非幂型的情况下也能定义分数维数, 这时, 若在双对数图上描出 r 和 $N(r)$ 时, $D(r)$ 就表示其斜率。由于斜率或导数的局部或瞬时极限特性, 即不再把分维数 $D(r)$ 看作是独立于观测尺度 r 的常量, 而把它看作是依赖于观测尺度 r 的函数, 即

$$D(r) = - \frac{d \log N(r)}{d \log r} \quad (2)$$

对此式求解, 得

$$N(r_2) = N(r_1) \cdot \exp \left(- \int_{r_1}^{r_2} \frac{D(r)}{r} dr \right) \quad (3)$$

式(3)是从理论上表达了两个尺度 r_2 和 r_1 的观测值 $N(r_2)$ 和 $N(r_1)$ 与分维数 $D(r)$ 之间的函数关系。它的适应范围不受无标度区的限制, 同时也可充分地、平等地使用全部观测数据, 建立更具有普遍意义的信息变动规律。

3 建立分维函数的数值途径

客观世界过于复杂, 不便于作全局观察与理解, 需要作适当的简化与去粗取精。即以“模型”来理解、描述、解释与预测客观世界。数学模型是我们对客观现实的一种理想化(简化)的抽象。模型的好坏表现在它对客观现实描述的符合程度和对未来的预测程度。同时, 模型要尽可能地简单。

3.1 模型的机理分析

对于自然界的不规则的(随机)分形, 不具有严格的自相似, 只有在一定的尺度范围内的自相似或具有统计自相似。以地图上河流、海岸线等线状目标为例, 其最小量测尺度 d_{\min} 显然不能小于 0.1 mm, 更不能趋于零。同样, 最大量测尺度 d_{\max} 不仅不能取无穷大, 甚至也不能大于线状物体的首末点连线的直线长度。同时, 其量测结果(长度)不能也不允许为无穷大。因此, 对于观测尺度和观测值都有一个逻辑界定问题, 即观测值与观测尺度均是有界的。就曲线长度与其量测步长之间的关系来说, 当在 $(d_{\min} - d_{\max})$ 范围内取所有可能的步长时, 观测尺度与观测结果之间一般呈非线性关系: 当量测尺度接近但尚未等于 d_{\min} 时, 量测长度的增长在逐步减弱进而趋于常值 L_{\max} ; 反之, 当量测尺度接近但尚未等于 d_{\max} 时, 量测长度已不再作明显减少而趋于常值 L_{\min} 。因此, 从逻辑机理上看, 此种关系曲线随着量测尺趋于 d_{\min} 与 d_{\max} 时, 应分别渐近于两个观测值的极值 L_{\max} 与 L_{\min} , 即呈反 S 形曲线。以往的应用研究大都没有顾及这个非线性问题, 其主要原因是量测尺度的取值范围太有限, 或把无标度区之外的观测信息视作无用信息从而予以抛弃。当尺度取值范围顾及到给定线状目标的尺度上下限时, 可以得到相应的反 S 型分布。后者可称为扩充分维, 此处的维数已不是常数而是变量。在确定分维数时, 非线性现象的出现, 说明自然界实际存在的统计分形的出现, 对它的研究具有更为普遍的意义。要从新的角度作进一步的研究, 充分使用所获得的各种不同信息。

此处需要对尺度 d 和观测值 L 进行规格化处理: 即对最小尺度 d_{\min} 与最大尺度 d_{\max} 的范围进行等比划分(表现为对数的等差划分)。由于对地图数据来说, 最小步长 d_{\min} 可认为

是常量，而最大步长 d_{max} 是取决于具体物体的图形的，因此，这样所确定的量规步长序列是取决于 d_{max} ，因而也就是取决于物体本身。显然，这种步长序列是因物而异的。此处要求在数据观测开始就顾及到物体本身的特征，采用规格化步长，从而是一种具有自适应性能的步长确定方法。为了使反 S 曲线图形能比较完美地表达，在试验中对每一个图形用 20 个（从最小到最大）步长去量测该图形的曲线长度。量测数据的规格化处理可使不同的分形体的特征便于进行比较。此处采用 CAD 中的常用的规格化方法：将各个长度量测结果转换为 0~1 之间的规格化数值。

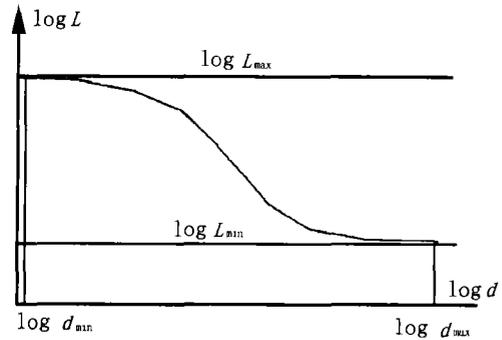


图 1 观测尺度 d 与观测结果 L 之间的反 S 关系曲线

Fig. 1 The Contrary-Relation Curve of Observatory Dimension d and Observatory Results L

3.2 反 S 形关系的分布的数学表达

此处的问题是一个数学建模的问题。数学建模的目的就是要给出实际问题的一个数学结构，具体表现为经验方程的建立。对于经验方程的建立可有两种主要选择：确定型方程与类型待定方程。对于前者，通过机理分析确定函数的类型，获得问题的数学结构，然后对结构的参数进行估计。对于后者，通常可用代数多项式或傅里叶级数来实现。显然，反 S 形模型已经是通过机理分析得出的确定型模型，可利用它与 S 形模型的简单关系通过求解 S 形模型来实现。

用来表达 S 形分布的数学模型有多种，如最为常见的 Logistic 模型、Log-Logistic 模型、Gompertz 模型等。此处我们主要使用 Logistic 模型。

3.2.1 Logistic 模型（阻滞增长模型、密度制约（density dependence）模型）

Logistic 模型是一种增长模型，该模型又称为 Verhulst Pearl 模型。有着极其广泛的应用领域：生物学、生态学、化学以及政治和经济领域等。增长模型一般是机理模型而不是经验模型。

增生长曲线的函数式为：

$$y = \frac{L}{1 + ae^{-bt}} \quad (4)$$

式中 L 、 a 、 b 是参数，其中 L 是函数 y 增长的上限。

曲线在点 $(t = \ln a/b, y = L/2)$ 处有拐点，故呈现为 S 形（图 2）。

该方程可改写为

$$L/Y = 1 + ae^{-bt}$$

进而有 $\ln(L/Y - 1) = \ln a - bt$

它变为直线方程，且它的取值不受限制。可在对数坐标系中用最小二乘法对观测数据进行拟合。

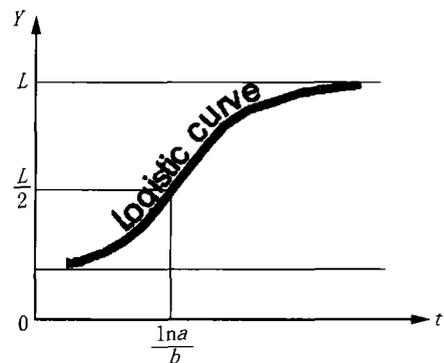


图 2 Logistic 曲线

Fig. 2 Logistic Curve

下面介绍 Logistic 模型的分段三和值近似解法。

设有时间序列: $(t_1, y_1), (t_2, y_2), \dots, (t_n, y_n)$, 对它作皮尔曲线拟合。首先将模型函数改写如下:

$$y^{-1} = 1/y = 1/L + a/L * e^{-bt}$$

然后将整个时间分成相等的三个间距, 对序列数据分 3 组求合:

$$S_1 = \sum_{t=1}^r Y_t^{-1} = \frac{r}{L} + \frac{a}{L} \sum_{t=1}^r e^{-bt} = \frac{r}{L} + \frac{a}{L} \frac{e^{-b}(1 - e^{-rb})}{1 - e^{-b}}$$

$$S_2 = \sum_{t=r+1}^{2r} Y_t^{-1} = \frac{r}{L} + \frac{a}{L} \sum_{t=r+1}^{2r} e^{-bt} = \frac{r}{L} + \frac{a}{L} \frac{e^{-(r+1)b}(1 - e^{-rb})}{1 - e^{-b}}$$

$$S_3 = \sum_{t=2r+1}^{3r} Y_t^{-1} = \frac{r}{L} + \frac{a}{L} \sum_{t=2r+1}^{3r} e^{-bt} = \frac{r}{L} + \frac{a}{L} \frac{e^{-(2r+1)b}(1 - e^{-rb})}{1 - e^{-b}}$$

令

$$D_1 = S_1 - S_2 = \frac{a}{L} \frac{e^{-b}(1 - e^{-rb})^2}{1 - e^{-b}}$$

$$D_2 = S_2 - S_3 = \frac{a}{L} \frac{e^{-(r+1)b}(1 - e^{-rb})^2}{1 - e^{-b}}$$

于是

$$D_1/D_2 = \frac{1}{e^{-rb}} = e^{rb}$$

两边取对数

$$b = 1/r * \ln(D_1/D_2)$$

又有

$$\frac{D_1^2}{D_1 - D_2} = \frac{\frac{a^2}{L^2} \frac{e^{-2b}(1 - e^{-rb})^4}{(1 - e^{-b})^2}}{\frac{a}{L} \frac{e^{-b}(1 - e^{-rb})^3}{1 - e^{-b}}} = \frac{a}{L} \frac{e^{-b}(1 - e^{-rb})}{1 - e^{-b}} = S_1 - \frac{r}{L}$$

由此得

$$L = \frac{r}{S_1 - D_1^2/(D_1 - D_2)}$$

同时

$$a = \frac{L}{c} \frac{D_1^2}{D_1 - D_2}$$

式中 $c = \frac{e^{-b}(1 - e^{-rb})}{1 - e^{-b}}$

该模型的应用都有一个前提: 其右端都是以无穷大为渐近条件。而在应用过程中, 自变量 t 只能取有限值, 因而使 S 形模型在其右端难以较为准确地拟合观测数据。因此, 也可使用符合 S 形特性的更易于实现的数学模型: 分段 Hermite 带导数插值模型和带导数三次多项式等。

3.2.2 Hermite 带导数插值模型

由前述机理分析可知, 在图 3 中 D_1 和 D_3 分别为最大观测值与最小观测值处的渐近线的斜率, 它们应该均等于零。顾及到实际观测数据的分布特征不一定对称, 故增加中段处的观测值 P_2 及该处的趋势方向 D_2 。这样就构成了分段带导数 Hermite 插值模型。

3.2.3 带导数三次多项式

其建模原理与前者相似, 只是在处理曲线

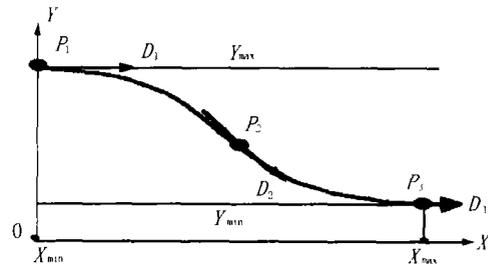


图 3 分段带导数 Hermite 插值
Fig. 3 Segmental Hermite Interpolation Value

中段部分不同。前者是借助 (P_2, D_2) 来控制，分两段进行；此处是通过最小二乘原理最佳地制中段部分。

设 三次多项式为

$$y = a + bx + cx^2 + dx^3$$

根据 x 的两个极值 x_{\min} , x_{\max} 将 x 的序列值规格化, 使其取值范围为 $0 \leq x \leq 1$ 。根据前述机理分析可知, 当 $x=0$ 和 $x=1$ 时, 函数的一阶导数等于零, 即

$$y' = b + 2cx + 3dx^2 = 0$$

为满足首点处的条件, 有 $x=0$, 这时, 由上式得 $b=0$; 为满足末点处的条件, 有 $x=1$, 这时并顾及 $b=0$ 的已知条件, 由上式得 $c=1.5d$ 。最后, 三次多项式变为:

$$y = a + (x^3 - 1.5x^2)d$$

利用中部其它观测点的值通过最小二乘原求出增未知参数 a 、 d 。

上式确保函数曲线在其首末端点 $x_{\min}=0$, $x_{\max}=1$ 处分别逼近两条水平渐近线。

4 扩展分维函数的结构分析

由于扩展分维表达了包括无标度区以外的信息, 从而揭示了分形体的更多层次的特征 (图 4): 织构分形 (textural fractal) —— $(d_{\min} - d_1)$ 所对应的部分, 它描述分形体的精细结构与纹理特征; 结构分形 (structure fractal) —— 无标度区 $(d_1 - d_2)$ 对应的部分, 它描述分形体的严格自相似特征; 态势分形 (state fractal) —— $(d_2 - d_{\max})$ 所对应的部分, 它描述着分形体的总体变化特征。

这种从全程上对分维函数的结构性划分, 从本质上对分形的概念进行了扩充: 把原来位于于无标度区以外的观测成果给予了新的理解与定义, 使得扩展分维获得了新的活力, 进而使它具有从多个层次上描述复杂现象的能力。从图 4 可以看出, 通常的分形与扩展分形有着一个近似关系: 通常分形的无标度区的上下界大体上与结构分形重合。

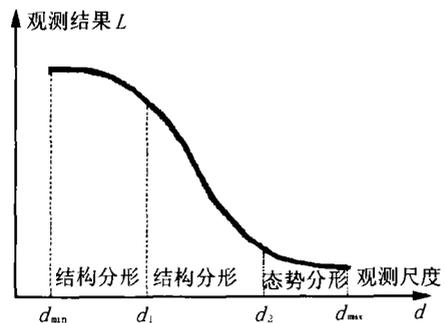


图 4 扩展分维的多层分形体特征描述
Fig. 4 The Relationship Between The Observation Measures And The Observed Results

5 分形体的多参数描述

前面已就分数维扩充的第一个任务 (不把分数维维数仅看作是一个常数, 使其能有赖于观测的尺度, 即使在自相似性不成立的那种范围内也能使用) 进行了论述。此处再就第二个任务 (在多段相似性成立的情况下, 为了弥补只用分数维不能描述的信息, 要重新引进另外的量) 作初步探讨。

关于形状的描述与分析是一个极其复杂的问题, 涉及地图量测和模式识别等问题。由于自然现象的复杂性, 对其图形形状的数学描述总是一个逼近与优化问题。

以河流为例, 当其总体趋势呈直线状时, 其弯曲程度可用目标曲线相对于目标首末端

点连线的偏移量来度量(图5)。由此得出一个重要概念:弯曲程度是一个相对比较量。但是,当目标的总体趋势呈明显弯曲时(图6),采用前述方法就不合理了。这时就要分两步来作:先求目标的总体趋势线,我们把它叫做总体基准,后者不一定是一条简单的曲线,只是其迂回曲折相对较少,必要时也可对它估算其曲折程度;然后以它为基线来度量目标曲线的偏移程度。这种方法我们把它叫作“双模型描述”法:总体曲折模型与相对弯曲估值。

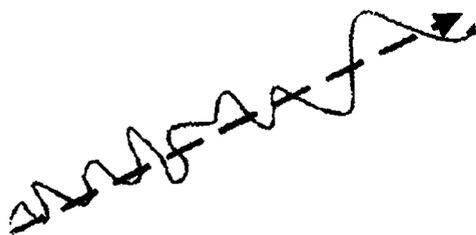


图5 直线型总体基准
Fig. 5 Overall Linear Standards

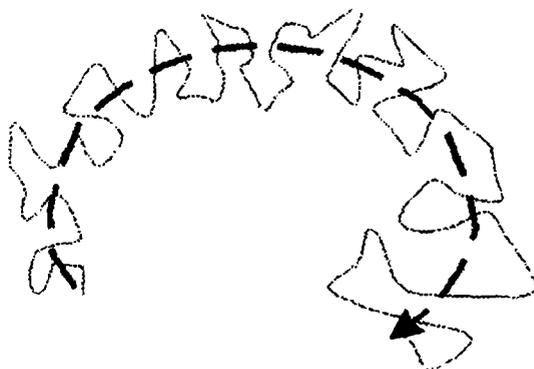


图6 曲线型总体基准
Fig. 6 Overall Curve Standards

5.1 总体曲折模型

一般曲线的总趋势线可称为该复杂曲线的总体曲折。曲线的总体曲折(如河流的山岳曲折,山体的总体走向等)表达复杂目标的总体走向,是评价曲线弯曲的参考基线或基准。由于总体曲折也可能是在较小的程度上迂回的,因此,总体曲折也是相对的。总体曲折的确定方法可有两种。

5.1.1 拐点连线法

迂回曲线拐点的查找在于连续地对每相邻4点判断拐点的存在与否。拐点存在的条件是至少有相邻4点。设相邻4点的坐标为 (Y_i, X_i) 、 (Y_{i+1}, X_{i+1}) 、 (Y_{i+2}, X_{i+2}) 和 (Y_{i+3}, X_{i+3}) ,过中间两点引一直线,其方程为

$$\Phi(y, x) = (Y_{i+2} - Y_{i+1})(X - X_{i+1}) - (X_{i+2} - X_{i+1})(Y - Y_{i+1}) = 0$$

用四点中的第一点和第四点的坐标,即 (Y_i, X_i) 和 (Y_{i+3}, X_{i+3}) 分别代入上式中的 Y 和 X ,如果此时 Φ 取相反的符号值,则在2,3点之间存在着拐点。否则,没有拐点。

5.2.2 条带树法

条带树的建立过程可描述如下:作为树根的是一个矩形,它的长轴与取向就是曲线首末点的连线,该矩形宽是过两个距长轴的最远距离点所作的平行线。如果只有一个最远距离点,则长轴算作另一条平行线(图7)。所得的最远距离点将原始曲线分成两个或四个段落,接着再对每个新的段落进行同样的操作。

条带树可借助 DOUGLAS 法反复地对曲线段的左右最远距离点来进行。

图7中所示的条带树用图表格式表示见图8。

在具备条带树的条件下,线状物体的曲折程度可形式化为子结点相对父结点或祖父结点的图形偏移。

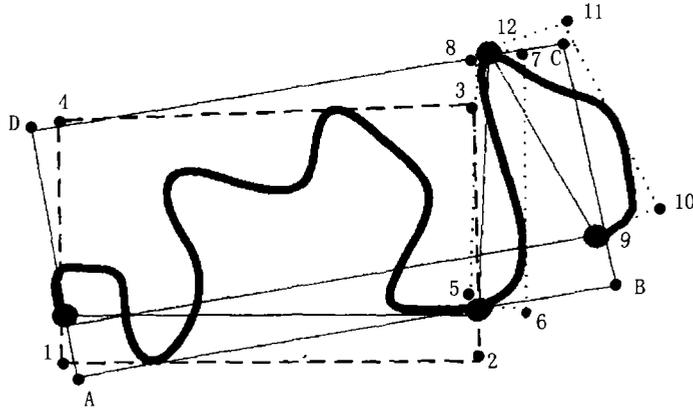


图 7 线状地物条带树的生成

Fig. 7 Generation of Trip-tree for Linear Objects

5.2 相对曲折的确定

复杂曲线相对其总体曲折的随机摆动可称为复杂曲线的相对曲折。但仅用通常的简单比值一个数是难以区分不同的曲线目标。采用两个参数是合理的：一个参数表示单位长度上

弯曲的多少，另一个参数表示弯曲的大小。从分形学的角度看，可利用分维布朗运动 (fB_m) 来描述曲线的相对曲折，它有两个参数 H 、 σ ，分别描述曲线的弯曲多少与弯曲大小。通常的分维布朗运动都是处理单值曲线的，而地图上的曲线显然是迂回多值的。除了使用极化（区域形心剖面法、周边展开图法等）处理以外，按基准线逐段展开用乃是一个可行的转换方法（图 9、图 10 和图 11）。

设曲线原始坐标点列为 $(x_i, y_i), i = 0, 1, 2, \dots, n$ 。在求出曲线的总体趋势线之后，以总趋势线的每个线段作为横坐标轴，构成一个局部坐标系，并把属于该总趋势线线段的原始曲线段分割出来，进而变换到当前的局部坐标系。待整条曲线变换完毕之后，得到一组新的点列，仍记为 $(x_i, y_i), i = 0, 1, 2, \dots, n$ 。

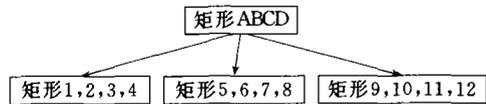


图 8 线状地物的条带树结构图

Fig. 8 Structure of Trip-tree for linear Objects

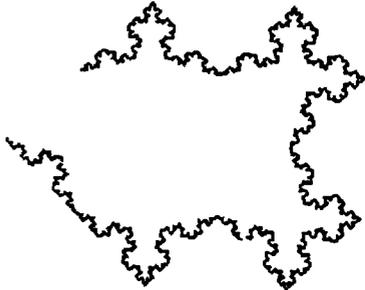


图 9 多值分形曲线

Fig. 9 Multi-Value Fractal Curve

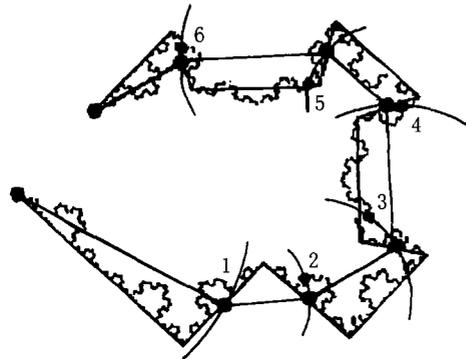


图 10 基准线的形成与分段展开准备

Fig. 10 Generation and Segment Unfolding of Baselines

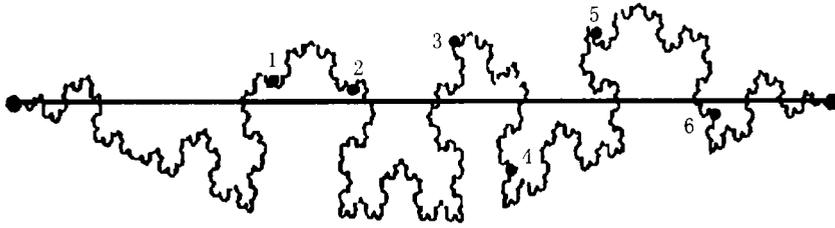


图 11 多值曲线近似地展开为单值曲线

Fig. 11 Multi-value Curve Unfolded to Linear Curve

通过这样的展开处理,为按分维布朗运动方法来确定原始曲线相对总体曲折的相对曲折的两个参数作了适应性准备。

分维布朗函数被认为是描述统计分形的最为适宜的模型。用于描述分维布朗运动有两个参数 H 和 σ 。 H 描述曲线的粗糙程度, H 越大,图形越平滑; H 越小,图形变化越不规则,越复杂。 H 可认为是复杂程度的指标。 σ 参数决定着弯曲的垂直方向的比例,称为比例因子。 σ 越大,目标弯曲的幅度就越大。 σ 越小,形状的起伏就越小。分形特征值 H 与 σ 一起,决定着图形的复杂程度与起伏程度。

此处提出一种 H, σ 的估值方法:

由于
$$Pr \left\{ \frac{f(x + \Delta x) - f(x)}{(\Delta x)^H} < y \right\} = F(y)$$

$F(y)$ 是正态分布 $N(0,1)$ 的累积分布函数,所以有

$$E|f(x + \Delta x) - f(x)| = \frac{2}{\sqrt{2\pi}} \sigma (\Delta x)^H = C (\Delta x)^H$$

此处
$$C = \frac{2}{\sqrt{2\pi}} \sigma$$

E 表示随机变量的数学期望,即多个样本的平均值。

两端取对数

$$\ln[E|f(x + \Delta x) - f(x)|] = \ln C + H \ln(\Delta x)$$

由于 H, C 是常数,故以 Δx 和 $[E|f(x + \Delta x) - f(x)|]$ 在双对数 (\ln, \ln) 坐标系绘图,将得近似直线。直线的斜率即为 H ,截距为 $\ln C$ 。

对于所得到一组新的被展开了的点列 $(x_i, y_i), i = 0, 1, 2, \dots, n$,以弧长 h 作基本单位步长将曲线参数化。于是可记为:

$$x_{i+1} - x_i = h$$

式中 h 为长量,使点的自变量参数呈等间距)。这样就使得:

$$x_0 < x_1 < x_2 < \dots < x_n$$

取 $dx = kh, (k = 1, 2, 3, \dots, n)$, 计算相应的 $E|f(x + \Delta x) - f(x)| = E|f(x + kh) - f(x)| = \frac{1}{n - k + 1} \sum_{i=0}^{n-k} |y_{i+k} - y_i|, (k = 1, 2, \dots, n)$ 。

对点集 $\{\ln kh, \ln(\frac{1}{n - k + 1} \sum_{i=0}^{n-k} |y_{i+k} - y_i|), (k = 1, 2, \dots, n)\}$ 作回归直线拟合,得出 H 和 $\ln C$ 的须估值式:

$$H = \left\{ \sum_{k=1}^n (\ln kh) \left[\ln \left(\frac{1}{n-k+1} \sum_{i=1}^{n-k} |y_{i+k} - y_i| \right) \right] - \frac{1}{n} \left(\sum_{k=1}^n \ln kh \right) \left[\sum_{k=1}^n \ln \left(\frac{1}{n-k+1} \sum_{i=1}^{n-k} |y_{i+k} - y_i| \right) \right] \right\} / \left\{ \sum_{k=1}^n (\ln kh)^2 - \frac{1}{n} \left[\sum_{k=1}^n (\ln kh)^2 \right] \right\}$$

$$\ln C = \frac{1}{n} \sum_{k=1}^n \ln \left(\frac{1}{n-k+1} \sum_{i=1}^{n-k} |y_{i+k} - y_i| \right) - H \frac{1}{n} \sum_{k=1}^n (\ln kh)$$

从而
$$\sigma = \frac{\sqrt{2\pi}C}{2}$$

5.3 分形体的形态塑造参数

以曲线形状为例，双参数描述了曲线的弯曲密度与弯曲幅度，而弯曲的形状尚未得到描述。为此引入第三个参数 S (Shape)，我们称它为弯曲形状描述子。弯曲形状描述子在此已不是具有分形意义的参数，而是曲线特征的补充描述参数。形状描述子要借助图形座标计算分析，对弯曲进行曲率分析与形态分类。

根据解析几何学原理，一个弯曲可定义为位于两个相邻拐点之间的曲线段。在此区间，曲线的一阶导数（切线斜率）单调地变化。即拐点是弯曲的分界点，也是一阶导数（切线斜率）变号的地方。在图 12 中， A 、 D 为相邻两个拐点， B 为曲率分析所求得的弯曲顶点。 ABD 为一个弯曲。 A 、 B 、 D 处的一阶导数（切线斜率）很容易求得。在三条切线与拐点连线所形成的四边形 $ACBD$ 与三角形 ABD 之间可构造无数单峰曲线段，因此，四边形 $ACBD$ 与三角形 ABD 就分别构成弯曲族的外边界与内边界。在这两个边界范围内，可作无数条带导数插值曲线。但是，对于地学应用来说，区分与构造四种弯曲就足够了：

- ① V 型弯曲，其形状贴近于内边界三角形 ABD ；
- ② V-U 过渡型弯曲，其形状接近弯曲 A_1B ；
- ③ U 型弯曲，其形状接近弯曲 A_2B ；
- ④ 超 U 型弯曲，其特征是弯曲顶部更

平更宽，如弯曲段 A_3B 。

概念与形状的分类要求相应的量化措施，根据上述情况，形状描述子可用下述方法近似确定：作角 C 与角 E 的角平分线并分别交 AB 边与 BD 边于 F 和 G ，并将各条角平分线 4 等分，得出 3 个中间点 1、2、3 和 4、5、6，这些点可作为所需弯曲类型前半支与后半支的顶点，可使用“广义二次曲线（孙家昶，1982）”或“三个型值点和首末点斜率法（电子计算机和数控技术在飞机制造中的应用，国防工业出版社，1978，p132）”来构造弯曲的前半支与后半支。若在这些中间点处把其所依附的底边的斜率作为它们的切线斜率，则可用前述 Hermite 分段插值来构

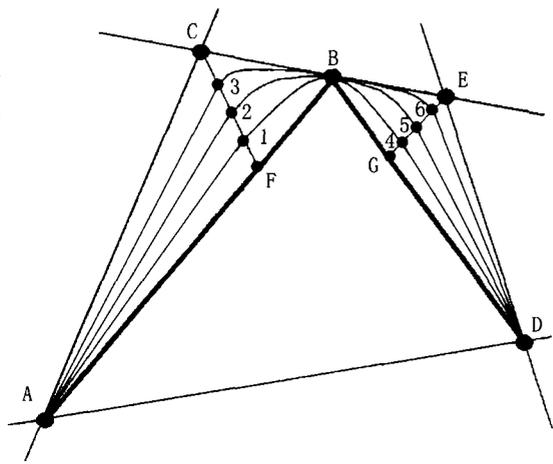


图 12 弯曲形状描述子的构造
Fig. 12 Construction of descriptive algorithm for bend shape

造各类弯曲的前、后半支。用计算所得的弯曲支段与实际弯曲支段进行匹配比较（如二者之间的面积等），实现弯曲形状的近似分类，生成相应的形状描述子。

6 地图信息的自动分形估值与基本综合途径

6.1 多尺度信息容量的自动演绎

对于一幅新图或一个新的空间数据库来说，其总容量“多少”是一个要首先确定的问题。是关于“总貌”的“总体”构思（CONCEPTION）问题。这个问题可以说是通过数量来表达质量的问题。

地图容量体现为地理实体的类别多少和地理实体个数的多少。前者属语义信息范畴的设计问题，要由人工进行确定；后者属于实体数量关系问题，是此处要予以研究的问题。实体数量的估值可通过对系列地图的量测或依据规范或行业标准来解决。显然，前者极不可取，后者属于指导性文件，需要人在实际情况下再灵活运用，一适用于具体情况下的问题的解决。在拥有空间数据库的情况下，可用分形原理对地理信息的图形内容进行自动演绎，即根据扩展分形的粗视化原理，对不同尺度下的空间数据库容量进行估值：动态地建立图形的容量与地图比例尺的近似关系。可分要素和按目标类型的不同分别实施。

这种估值是针对当前具体的空间数据库的，而不是引自专业文献，也不是参照其它国家、其它地区或其它地图量测与经验数据。其针对性与自适应性均是显而易见的。

可采用如下途径实现分形演绎：用国家基本图式将起始地图或初始空间数据库作精密栅格化，用栅格单元统计的方法获取初始信息容量。然后，利用分形原理进行信息容量的分形演绎。

分形法的主要优越性在于描述与处理地物或其集合时，其固有特征（分数维）均起着主导作用。

地图内容的总容量确定之后，需要进一步确定每类物体各有多少且由哪些实体构成。作为宏观控制信息来说，一方面可根据原始数据库中各要素的比例关系得出与其相似的分布比例。另一方面，当采用分要素作分形演绎时，则首先得到的是相应要素在新尺下的近似容量。

这里区分两个概念是有益的：选取与概括。前者是针对目标整体，后者是针对目标局部的。选取是决定目标的去留，概括是对保留目标局部细节的处理。目标整体具有数据库关键字或标识码，从而是可以直接地识别。目标的细节则不然。

地图的内容可看作是点、线、面 3 类地物的总和，因而可通过对这 3 类地物的粗视化估值来逼近地图内容的随比例尺的动态变化规律。若顾及扩展分形的概念，则尺度的变化不仅可以是连续的，而且其变化范围是很大的，实际上是不受限制的。

“多少”问题解决之后，“哪些”的问题立即提上日程。显然，这需要借助相应的结构化手段来解决。如对于点状目标集合来说可有：点集凸壳层嵌套、点集的 Delaunay 三角网和 Voronoi 图等；对于线状地物集合来说，有河网的树结构和基于线状目标集合的 Voronoi 图等。

6.2 点状地物集合的分形描述与综合

6.2.1 点状物体集合的分形描述

属于点状地物的有：纯点目标和群斑目标。前者包括：各种测量控制点、高程点、水位点等。若仅考虑图形信息描述，则这类点状物体仅有平面位置和/或高程信息 (x, y, z) ，没有本身的轮廓图形和进一步的内部结构。点状物体集合的分布特征成为唯一的描述对象。后者包括：风蚀残丘群、湖群与岛群等，它们虽有着大小不等的面积，但在研究其分布特征时，实体的面积和轮廓特征可忽略不计。对于点状目标，作为分形描述的对象是它们的分布结构特征。分数维在此描述了集合中的点数随观测尺度变化的规律。

对于平面上分布的无轮廓形状与面积信息的离散点集，可用盒维数方法计算其分维值，即用间隔为 r 的格子把平面分割成 $(M \times M)r$ 个正方形，数出其中至少包含一个点的正方形的个数 $N_i(r)$ ($i = 1, 2, \dots, M$)，如果当 r 改变时， $N(r) \propto r^{-d}$ 的关系得到满足，可称 d 为这个点集的维数，它描述了集合中点的数目随观察尺度变化的速率，即与点集特征有关的数量变化规律。可用线性回归模型拟合点对 $(\ln r_i, \ln N_i)$ ，由回归直线的斜率得出点集的分数维 d 。

6.2.2 点状目标的分形选取

由于纯点目标无本身形状信息，故其综合仅涉及目标的合理选取问题。对于图斑群来说，尽管各个图斑有其大小不等的面积与轮廓图形，但表达其结构信息仍是第一位的。因此，它的选取是首要的综合措施。

这里介绍基于分维的图斑数量选取规律：

设所考虑的点群目标综合前后的比例尺分别为 M_1 和 M_2 ，目标的选取数量分别为 N_1 和 N_2 ，根据目标数量与观测尺度（分辨距离） S 的关系，有：

$$N_1 = C_1 * S_1^{-d_1} \quad N_2 = C_1 * S_2^{-d_1}$$

此处 N_1 为点数， C_1 为某一常数， S 为分辨距离， d_1 是点集的分数维，它通常可用网格覆盖法来确定。顾及到地图符号满足开方根规律 $S_2/S_1 = (M_1/M_2)^{1/2}$ ，则可得出基于分形原理的点状地物选取规律：

$$N_2 = N_1 * (M_1/M_2)^{d_1/2}$$

这说明目标选取结果不仅与 N_1 和前后比例尺有关，而且也关联于表征点集分布特征的分数维有关。但是，这量仅解决了“多少”的问题，还有“哪些”的问题要接着解决。建立点集目标的 Voronoi 图或 Delaunay 三角网是解决“哪些”问题的好办法（见图版 2 中的图 13~15）。

6.3 线状地物集合的结构化描述与综合（此处以河网的综合为例）

6.3.1 河网结构的结构化描述

河网作为一种地图图形要素，在地图图形数据处理中，有着非常重要的地位。它的重要性不仅表现在河网自身方面，而且还表现在它对其他要素所具有的制约作用。由于河网的形状和结构方面所表现出的复杂性和不均匀性，它的形状结构描述就有其特殊的重要性。

尽管在文献中有关河网的多种分维估值方法，如 Horton 的有关定律、基于河网总长的分维估值、基于覆盖法的分维估值等。它们显得过于统计平均，难以适应地理数据处理中的“区域性”特征的顾及。分形方法是很好的方法，但不是万能的方法。笔者主张多种方法集成使用。在对河网进行结构特征描述时，显然河系的树结构有着极其重要的优越性：在 GIS 环境下，数字化河网是现成的，构造河系的树结构只是一个专用软件模块问题，其关键功能是具有鲁棒性缓冲区检索。在河系树结构中可以自动获取很多信息：主流左右两侧各

有几条支流注入, 以及这些支流的等级信息; 对于每条支流又可进一步获得其两侧支流以及它们的等级信息, 以此递推下去, 直至叶结点河流出现为止。河网树结构所自动提供的河流级别信息对于河流的选取有着直接的重要意义。若同时生成河系的 Voronoi 图(见图版 2 中图 16), 则河流的选取原则上就有了科学的保证。根据河系树结构信息, 可以得出一种独特的分数维: 河系树结构分数维持 D_r 。

6.3.2 线状目标选取数额的分形估值

对于具有分形特征的线状物体集合(如河系结构), 存在着随比例尺缩小河流条数的选取问题, 其基于分形学原理的选取公式为:

$$N_2 = N_1 * (M_1/M_2)^{(-D_r/2)}$$

式中 D_r 为河系的分数维。

这里仍然是仅解决了“多少”的问题, 同样地还有“哪些”的问题。河网的树结构加河网的 Voronoi 图可以从基本原理上解决“哪些”的问题。

6.3.3 线状目标图形的分形概括

(1) 曲线的自动分段算法。由于曲线特别是较长的曲线, 在其不同区段上有其不同的曲折度或复杂程度, 因此, 对曲线进行分段处理显然会提高处理质量。问题是如何对曲线进行分段的问题。目前已提出的有: 统计有序聚类法(郭庆胜, 1995)、分形法(王桥, 1996)等。

(2) 用分形布朗运动参数量化图形结构特征及复杂程度。该部分内容已在前面详述过, 此处不再述及。

6.4 面状地物集合的分形综合

6.4.1 离散图斑群(面状物体集合)的分形描述

对于具有轮廓图形及面积信息的图斑群来说, 随着观察尺度或比例尺的变化, 图斑群的数量也会相应地发生变化。Korcak (1940) 指出, 湖泊群面积的概率分布呈幂型:

$$P_r(A > a) = ka - b$$

式中 a, k, b 为常数; $P_r(A > a)$ 表示面积大于 a 的湖泊的概率。显然该式对于岛群也是成立的。Bruce、Milne 在对某种植物类型所组成的图斑群进行研究中发现了更进一步的规律:

$$N_r(A > \lambda) = C\lambda - D/2$$

式中 A 表示图斑的直径, λ, C 为常数, $N_r(A > \lambda)$ 表示直径大于 λ 的图斑个数。实际上此式可改写为与前式等价的形式:

$$N_r(A > a) = Ca - D$$

此处 D 为图斑群的分数维, 它描述了图斑数量随观察尺度变化的速率, 即描述了图斑数量随比例尺变化的规律。

对上式两端取对数, 得

$$\log N_r(A > a) = \log C + (-D)\log a$$

对于 a 的不同取值 $a_i (i = 1, 2, \dots, m)$ 求出相应的 $N_i = N_r(A > a_i)$, 在双对数坐标系中建立回归直线, 其斜率值即为 D 。

6.4.2 离散面状地物集合中地物选取的分形方法

对于面状地物, 其轮廓线的综合已按线状地物处理, 而面状地物的选取则主要是针对小而离散的图斑群, 如湖群、岛群和风蚀残丘群等, 其基于分形学原理的选取公式为:

$$N_2 = N_1 * (M_1/M_2)D_3$$

式中 D_3 为图斑群的分数维，它描述图斑数量随观测尺度而变化的速率。为确定 D_3 ，可先求出图斑集中面积最大的图斑 M 和最小的图斑 m ，在 $[m, M]$ 范围内对图斑群进行划分，然后用线性回归求出 D_3 。

总之，用分形学方法描述自然要素的固有特征，借此对其信息进行多比例或多层次处理，其主要优越性在于描述与处理地物或其集合时，其固有特征（分数维）均起主导作用。

6.4.3 离散面状地物轮廓图形综合的分形方法

原则上这是一个线状地物图形概括问题。故此处不再赘述。

参考文献：

- [1] 敖力布, 林鸿溢. 分形学导论[M]. 内蒙古人民出版社, 1996.
- [2] Barmesley M, Devaney R et al. 分形图形学[M]. 北京: 海洋出版社, 1995.
- [3] 陈其明. 在自动制图中对曲线信息量的数学描述与压缩[J]. 测绘学报, 1981, 10(3): 239-254.
- [4] 电子计算机和数控技术在飞机制造中的应用[M]. 国防工业出版社, 1978.
- [5] 高安秀树. 分数维[M]. 地震出版社, 1994.
- [6] 郭庆胜. 曲线特征的自动分段及其在自动综合中的应用[J]. 武汉测绘科技大学学报, 1995, 20(增刊): 64-68.
- [7] Kaye B H. 分形漫步(A random walk through fractal dimensions)[M]. 东北大学出版社, 1994.
- [8] Laplante P. 分形图形基础与编程技巧(Fractal mania)[M]. 学苑出版社, 1994.
- [9] 李后强, 张国祺, 汪富泉等. 分形学理论的哲学发轫[M]. 四川大学出版社, 1993.
- [10] 林鸿溢, 李映雪. 分形论——奇异性探索[M]. 北京理工大学出版社, 1992.
- [11] 刘式达, 刘式适. 分形和分维引论[M]. 气象出版社, 1993.
- [12] 刘文予等. 金属断口纵剖面的分形分析[A]. 分形理论用其应用[C]. 武汉: 华中理工大学, 1991.
- [13] Preparata F P, Shamos M J. 计算几何导论[M]. 北京: 科学出版社, 1992.
- [14] 王桥, 毋河海. 地图信息的分形描述与自动综合研究[M]. 武汉测绘科技大学出版社, 1998.
- [15] 毋河海. 河系树结构的自动建立[J]. 武汉测绘科技大学学报, 1995, 20(增刊): 7-14.
- [16] 毋河海. 自动综合的结构化实现[J]. 武汉测绘科技大学学报, 1996(3): 277-285.
- [17] 毋河海. 等高线树的自动建立及其应用[J]. 测绘科技动态, 1996(1): 2-7.
- [18] 毋河海. 分维扩展的数值试验研究[J]. 武汉测绘科技大学学报, 1998(4): 329-336.
- [19] 毋河海. 地图信息自动综合基本问题研究[J]. 武汉测绘科技大学学报, 2000(5): 377-386.
- [20] 毋河海. 地图信息自动综合基础理论与方法体系研究[A]. 第三届两岸测绘发展研讨会论文集, 香港, 2000c. 611-632.
- [21] 孙家昶. 样条函数与计算几何[M]. 北京: 科学出版社, 1982.
- [22] 仪垂祥. 非线性科学及其在地学中的应用[M]. 气象出版社, 1995.
- [23] 张济忠. 分形[M]. 北京: 清华大学出版社, 1995.
- [24] Bittner H R, Sernetz M. Selfsimilarity within limits: Description with the Log-Logistic function[A]. In: H-O Feigen, J M Henriques, Penedo L F (ed.). Fractals in the Fundamental and Applied Sciences[C]. Elsevier Science Publishers B. V. (North-Hoolland), 1991.
- [25] Волков Н М. Принципы и методы картометрии[M]. АН СССР. Москва, 1950.
- [26] Wu Hehai. Prinzip und Methode der automatischen Generalisierung der Reliefformen[J]. Nachrichten aus dem Karten- und Vermessungswesen, 1981, 85: 163-174.
- [27] Wu Hehai. Structured approach to implementing automatic cartographic generalization[J]. Proceedings of the 18th ICA International Cartographic Conference, Stockholm, Sweden, 1997, 1: 349-356.
- [28] Zhu Q. Fractal-based analysis and realistic display the theory and application to DEM[M]. ISPRS V. XXXI, Part B4, Vienna, 1996. 1007-1010.

Research on Extended—Fractal Based Automatic Map Generalization

WU He-hai

(College of Resources And Environment, Wuhan University, Wuhan 430079, China)

Abstract: The Fractal method, which is originated from geo-sciences, has been widely applied in various sciences and disciplines. This paper, however, concentrates on the essence of its application, i. e., how much information it can provide for us, what limit it has, how it can be combined with other methods, in map generalization. Several issues related to the application of Extended Fractal in automatic map generalization is expounded.

Key words: Fractal Dimension; Extended Fractal Dimension; Automatic Map Generalization; Multiple Dimension Database