

用多重分形研究元素的共生组合性

郭科, 施泽进, 唐菊兴, 陈聆, 刘二永, 魏友华

(成都理工大学 成都 610059)

【摘要】应用分形理论来研究测区内元素在土壤中富集组合关系以及相关关系,并在此基础上用多重分形算法,对数据进行处理,作出了元素的趋势图,判定了元素的共生组合性,为进一步揭露、控制、评价该测区的元素异常打下了良好的基础。与传统方法相比,该处理结果具有较好的效果。

关键词 分形; 分维; 最小二乘法; 多重分形

中图分类号 O29; TD913 文献标识码 A

Study Chemical Elements' Intergrowth Combination with Multi-Fractals

Guo Ke, Shi Zejin, Tang Juxing, Chen Ling, Liu Eryong, Wei Youhua

(Chengdu University of Technology Chengdu 610059)

Abstract Fractal theory is applied to study the combination relations with chemical elements' enrichment in the soil of the testing area in the article, and then the data are dealt with by adopting the arithmetic of multi-fractals so as to draw a trend line to show how the elements are distributed in the soil. As a result the intergrowth combination of the elements is determined. Thus it provided a good basis on which more work can be done to expose; control and evaluate elements' abnormality in the area. Compared with the traditional methods, this method can work quite well in this research.

Key words fractal; fractal dimension; least square method; multi-fractals

地学中的许多事物都十分复杂,是非线性和不规则的,运用非线性科学(包括分形、混沌和非线性模型等)才有可能得以解决^[1]。传统的理论只能是简化或定性地刻画自然界中广泛存在的无序、混乱、不规则和不光滑的复杂现象。非线性理论和方法的提出为揭示隐藏于混乱复杂现象中的精细结构和定量地刻画描述它们提供了理论方法基础。

分形理论研究的主题是多重分形(multi-fractals),因为简单分形只需要一个维数来描述其整体的特征,不能完整地刻画大自然的复杂性。比如,对湍流、混沌和分形生长类型的非均匀复杂几何体,必须用多个维数来描述,才能全面的刻画其特征。多重分形就是针对这类情况而提出的新概念。

多重分形也称为分形测度。它是研究一种物理量在一个支撑(support)上的分布状况,换句话说,多重分形理论是定义在分形上的多个标度指数的奇异测度所组成的无限集合。多重分形理论定量刻画了分形测度在支撑上的分布状况。

本文所给出的分形模型,指出了齐波夫定律(Zipf law)是多重分形模型的数学基础,论证了齐波夫定律在低端截尾条件下具有尺度不变的分形性质;指出 q 次广义分维数 D_q 满足不等式 $D_q > D_{q'}$ (当 $q > q'$ 时),并指出了用线性回归方法处理实际数据并计算出的 D_q 不一定都符合该不等式的内在原因。通过在计算机上产生随

收稿日期: 2002-09-04

基金资助: 西藏地矿厅藏东基金资助(040015413)

作者简介: 郭科(1959-),男,博士,教授,主要从事数学地质方面的研究。

机数对多重分形模型进行模拟研究和应用实例说明多重分形模型在实际问题中应用的方法及步骤,并解释了分维数的实际意义。

1 多重分形模型

从信息论角度选出一套描述多重分形的参量 q 和 D_q 。当 $r \rightarrow 0$ 时,得到

$$I_q(r) = Cr^{-(1-q)D_q} \quad (1)$$

式中 $I_q(r) = \sum_{i=1}^N P_i^q$ 为 q 次几率矩, D_q 为 q 次广义分维数(或 q 次扩展分维数), q 次表征多重分形不均匀程度的量。 $C > 0$, 称为比例常数, $\sum_{i=1}^N P_i = 1$ 。式(1)称为多重分形模型^[2]。

通过Legendre变换可得

$$D_q = \begin{cases} \lim_{r \rightarrow 0} \frac{1}{q-1} \ln \sum_{i=1}^N P_i^q(r) / \ln r & q \neq 1 \\ \lim_{r \rightarrow 0} \sum_{i=1}^N P_i(r) \ln P_i(r) / \ln r & q = 1 \end{cases} \quad (2)$$

定理 q 次广义分维数 D_q 满足不等式

$$D_q \geq D_{q'}, \quad q > q'$$

根据定理的结论,可推知: D_0 (相似维) D_1 (信息维) D_2 (关联维)。

在用线性回归方法处理实际数据并计算出广义分维数的 D_q (即线性回归方程的斜率)时,不一定都满足不等式 $D_q \geq D_{q'}$ (当 $q > q'$ 时)。这是因为用统计上的线性回归方法得出的结果是整体上的结果(取决于所有的数据)^[3],它与用取极限方法得出的结果是不一样的。

引入齐波夫定律(John.N.L.and Kotz.s.),如果随机变量 Z 分布律

$$P(Z=r) = ar^{-p} \quad p > 1, r = 1, 2, \dots \quad (3)$$

则随机变量 Z 服从齐波夫定律。式中 $a = (\sum_{r=1}^{\infty} r^{-p})^{-1}$ 为常数, p 为指数参数。

随机变量 Z 的分布函数、数学期望和方差分别为

$$H(z) = P(Z \leq z) = \sum_{r=1}^z ar^{-p} \quad p > 1$$

$$E(z) = (\sum_{r=1}^{\infty} r^{-p+1}) / (\sum_{r=1}^{\infty} r^{-p}) \quad p > 1$$

$$V(z) = (\sum_{r=1}^{\infty} r^{-p+2}) / (\sum_{r=1}^{\infty} r^{-p}) - E^2(Z) \quad p > 1$$

在低端截尾条件下, $r \geq k > 0$,推得公式

$$P(Z=r|Z \geq k) = (ar^{-p}) / (\sum_{r=k}^{\infty} ar^{-p}) \quad r \geq k > 0$$

说明齐波夫定律具有分形性质

$$P(Z=r|Z \geq k) = P(Z=cr|Z \geq ck) \quad r \geq k > 0$$

式中 k 为尺度参数(正整数), c 为任意正的常数。该式表明齐波夫定律具有尺度不变的性质。因此,齐波夫定律具有尺度不变的概率分布(在低端截尾条件下),是多重分形模型式(1)的数学基础。

2 多重分形算法

2.1 将原始数据进行标准化变换

变换公式: $x'_i = (x_i - \bar{x})/S$, 其中 $x_i (i=1, 2, \dots, N)$ 为原始数据

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

变换后的数据 x'_i 的平均数为0,方差为1,且各元素数据的量纲一致,且两元素数据在标准化变换前后的相关程度不变。

2.2 判定数据具有分形特征

将标准化变换后的各元素数据，按从小到大的次序排列，并把该元素数据分布的总区间分成 r 个子区间，计算进入第 i 个子区间内的随机数的频率 $P_i (i=1, 2, \dots, r)$ ，令

$$I_q(r) = \sum_{i=1}^r p_i^q(r)$$

这样得到了数据 $I_q(r_1), I_q(r_2), \dots, I_q(r_n)$ 和 (r_1, r_2, \dots, r_n) ，然后将该数据绘在双对数坐标系统中(即 $\ln I_q(r) - \ln r$)，连接各点，曲线存在明显的直线段，则存在无标度区 ($q > 0$)，说明数据具有分形特征。

2.3 计算 q 次广义分维数

将数据 $I_q(r_1), I_q(r_2), \dots, I_q(r_n)$ 和 (r_1, r_2, \dots, r_n) 代入式(1)中，然后两边取对数，应用最小二乘法求出斜率的估计量 $\hat{D}_q^{[4]}$ ，即 q 次广义分维数。

2.4 作拟合图

作出元素数据的 \hat{D}_q 拟合图，通过元素的变化趋势来判定元素间关系是否密切。

3 应用实例

下面所用到的数据是来自某地区2000年的化探异常查证工作。以该地区的Pb、Zn、Sb、Ag、Cu元素为例，应用上面给出的多重分形算法对原始数据进行处理，所得到的拟合图如图1~5所示。

从拟合图中可知道该地区的Pb、Zn元素的广义分维数 D_q 变化趋势基本一致，说明了Pb、Zn元素的关系密切，Ag、Sb之间的关系次之，而Cu元素的分布似乎与以上元素关系不密切。

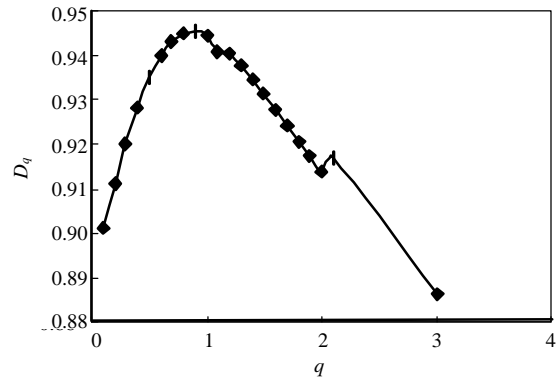


图1 Pb元素的 D_q 拟合图

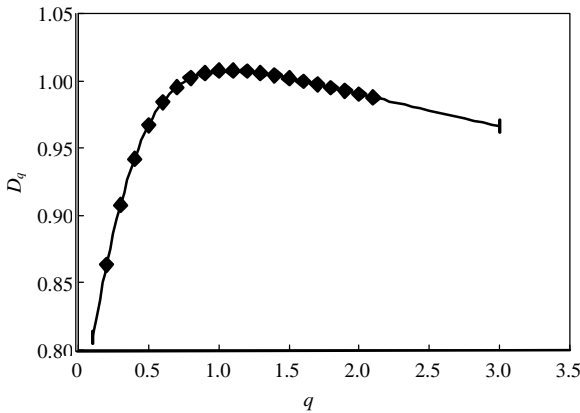


图2 Zn元素的 D_q 拟合图

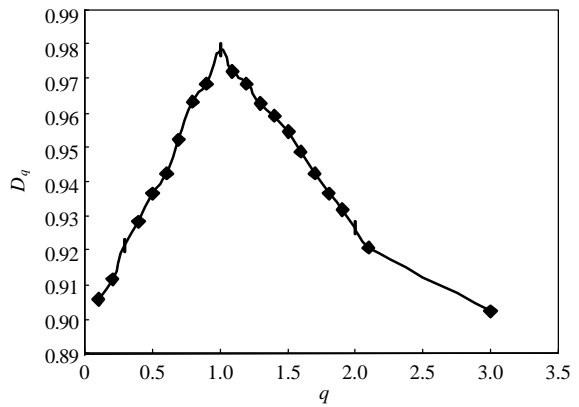


图3 Sb元素的 D_q 拟合图

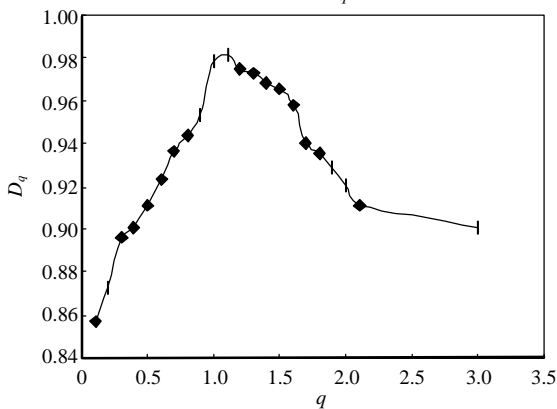


图4 Ag元素 D_q 拟合图

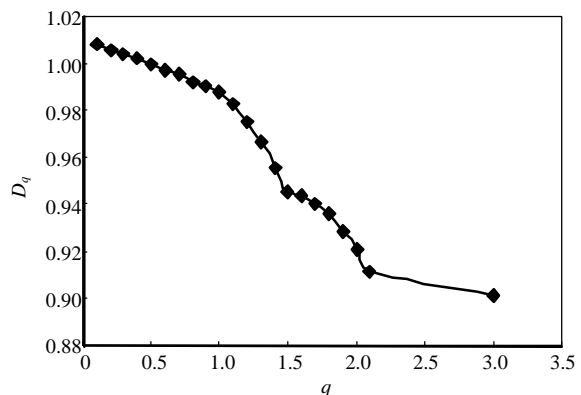


图5 Cu元素 D_q 拟合图

4 结束语

用传统方法所判定出元素之间的关系为：Pb、Zn之间的关系最为密切，Ag、Sb之间的关系次之，而Cu元素的分布与以上元素关系不密切，表明Cu元素和其他元素具有不同的地球化学背景和地球化学异常；Pb、Zn、Ag、Sb具有相似的成矿作用和成矿背景，而铜的成矿作用与成矿背景与前者显然是不同的。

从用多重分形方法处理出来的 D_q 拟合图也可以看出，Pb、Zn之间的关系最为密切，Ag、Sb之间的关系次之，而Cu元素的分布似乎与以上元素关系不密切。本文的结论和传统方法的结论非常接近。

1) 结果和用传统的逐步回归、正交因子分析和聚类分析等常用的统计分析所得到的结果非常相近。

2) 通过作 D_q 拟合图，只是定性的判定元素之间的关系，完全可以根据 D_q 寻找一种定量的方法来刻画元素之间的关系，并判定它们的共生组合性，这是可以改进的地方。

参 考 文 献

- [1] 赵鹏大, 陈永清, 刘吉平, 等. 地质异常成矿预测理论与实践[M]. 北京: 中国地质大学出版社, 1999
- [2] 胥泽银, 郭 科. 多元统计及其程序设计[M]. 成都: 四川大学出版社, 1999
- [3] 袁慰平, 孙志忠, 吴宏伟, 等. 计算方法与实习[M]. 南京: 东南大学出版社, 1999
- [4] 张济忠. 分形[M]. 北京: 清华大学出版社, 1995

编 辑 漆 蓉

(上接第183页)

参 考 文 献

- [1] Copeland G, Alexander W, Boughter E, *et al.* Data placement in Bubba[C]. In Proceedings of ACM SIGMOD Conference, 1988
- [2] Hua K, Lee C, Young H. An efficient load balancing strategy for shared-nothing database systems[C]. In Proceedings of DEXA' 92 Conference. Valencia, Spain, 1992, 69-474
- [3] Lakshmi M, Yu P. Limiting factors of join performance on parallel processors[C]. In Proceedings of the 5th International Conference on Data Engineering. Los Angeles, California, 1989, 488-496
- [4] Kitsuregawa M, Yasushi O. Bucket spreading parallel hash: a new, robust, parallel hash join method for data skew in the super database computer(sdc)[C]. In Proceedings of the 16th VLDB Conference. Brisbane, Australia, 1990, 210-221
- [5] Hua K, Lee C. Handling data skew in multiprocessor database computers using partition tuning[C]. In Proceedings of the 17th VLDB Conference. Barcelona, Spain, 1991, 525-535
- [6] Hua K, Lee C. An adaptive data placement scheme for parallel database computer systems[C]. In Proceedings of the 16th VLDB Conference. Brisbane, Australia, 1990, 493-506
- [7] Boral H. Parallelism and data management[C]. In Proceedings of 3rd International Conference on Data and Knowledge Bases. Jerusalem, 1988, 362-373
- [8] 杨 峰, 刘心松, 左朝树, 唐 续. 分布式并行服务器透明性及任务调度研究[J]. 计算机研究与发展, 2003, 40(9): 1 319-1 325
- [9] Hua K A, Su J X W. Dynamic load balancing in very large shared-nothing hypercube database computers[J]. IEEE Transactions on Computer, 1993, 42 (12): 1 425-1 439

编 辑 徐培红