



数学地质与因子分析法

范守志 李长珍

(中国科学院海洋研究所)

从莱伊尔的“地质学原理”问世(1830)到现在,地质学的发展到现在已有150多年了。地质学家们根据野外观察的丰富资料,采用新旧对比、将今论古的逻辑推理法,解释了许多地质现象,发现了许多规律,找到了许多地下矿床。近几十年来,由于大量使用先进的物理、化学、机械设备,获取的资料急剧增多。这就为地质研究从以定性地逻辑辩证为主转向定量地进行数据处理准备了条件。

1968年,国际数学地质协会(ISMG)在布拉格成立,随即就出版了定期刊物——国际数学地质会志,就是一个新的里程碑。

如今,数学地质还是一个年轻的分枝。它着眼于各类地质现象中的数量特征,并力求把所研究的地质规律定量化,力求把因果关系表达成相应的指标量、特征值,把地质过程用方程式表现出来。这些就像是物理学家们用各种物理量描述物性和运动状态,用方程式表达运动过程中的因果关系那样。由此可见,要说如今的地质学正像牛顿时代的力学一样,面临着重大的突破,并不是言过其词。

数学地质方法常与传统地质学的各种方法结合起来使用,两者并无相左之处。从方法的角度来看,数学地质又分成两枝——统计地质学和模式计算地质学。前者直接从观测资料出发,运用各种数学方法解剖其中的数量关系,得到相应的结论;后者则首先把地质现象看成是某种地质过程,即某种物理—化学—生物的综合过程的产物,从而利用有关的物理定律、化学定律建立描写该地质过程的微分方程式,再求出它的公式解或数值解,并与实测结果相对比,得出有关的地质结论。从逻辑角度看,

前者以定量归纳为主,后者以模式演绎为主。两者都要使用电子计算机。

因子分析法是统计地质学的重要方法之一。它渊源于精神病起因的分析。1964年,美国著名的数学地质家英布瑞(Imbrie, J.)加以系统地发展,把它变成了一个有严格数学根基和完全是定量形式的多元统计方法。他取名为向量分析法(Vector analysis method),用来研究加利福尼亚海湾中的泥砂运动^[1]。而在中文出版物中,都称为因子分析法。

1975年,美国芝加哥大学为数学地质专业的教师们举办了讲习班,由许多著名的数学地质专家分专题进行讲授。英布瑞的文章^[1]被推崇为“因子分析法的经典性文献”。各专题讲授的内容已由麦卡蒙(McCamm, R.)主编出版^[2]。

但国内应用因子分析法研究地质问题,正式发表者却不算太多。本文作者已为在我所的108-Z机上实施因子分析法编制了源程序。

下面就来简略地评介一下因子分析法。

设想在某海区的N个站位上进行了测量,现场测量或是采样后室内分析测定,在每个站位上都获得了M种参数的数据。于是全部测量结果可表示成一个N行M列的资料阵X。

$$X = (x_{ik})_{N \times M}$$

其中 x_{ik} 是第i站的第K种参数的值。因子分析法以 $\cos\theta$ 因子为基础,因此各 x_{ik} 可用绝对数值也可用百分含量;对同一个站来讲,也就是矩阵X中的同一行的M个数据,必须或者都是绝对数值或者同为百分含量。

因子分析法的目的之一就是要把这N个站位分群为若干个小组,每一组代表一个区,区

内各站就这M个参数而言具有一致的特征。

孤立地就某一站位来说,要表达它的特征就要用M个参数。从数学上说,矩阵的每一行都是一个M维向量。例如第i行代表第i站,

$$\begin{aligned} \mathbf{X}_i &= (x_{i1}, x_{i2}, \dots, x_{iM}) \\ &= x_{i1}e_1 + x_{i2}e_2 + \dots + x_{iM}e_M, \end{aligned}$$

这里, $e_1 = (1, 0, 0, \dots, 0)$, $e_2 = (0, 1, 0, \dots, 0)$,

$$\dots, e_M = (0, 0, \dots, 1).$$

即 $\mathbf{X}_i$ 须表达为M个基本向量 $e_1, e_2, \dots, e_M$ 的线性组合。因此,如果不分析资料阵X的数学结构,即使不把N个站分成N群(每站自成一区),而仅由M个数据的大小排列情况来分类,也得有 $M! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot M$ 种可能情形。何况这其中的每一种情形中实际上又包含着许多差异。注意到 $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$, $8! = 40320$,而通常的矿物资料, M达到几十,生物资料中物种数M能上百,可知这种“优势种先后顺序对比法”之艰难。实际上,人们被迫捨去许多参数的鉴定结果,只取少数几项数据进行对比。这不仅是一种浪费,而且也有其局限性。

另一方面,研究分区时,并不希望分出数目庞大的支离破碎的许多小区。在一定的比例尺的成果图上,倒是希望得出数目不太大的一级分区。例如在沉积物来源的分区研究中,若某海区有5个物质源,那么尽管做了许多站位(N很大),做了M种矿物成份的分析, M多达几十,但分出的区系数m不应比5大许多。

英布瑞的因子分析法捉住了这一点。他认为,如果在上述意义之下N个站位可分成m个区系($m \ll N$, $m \ll M$)的话,那么这N个M维向量中的每一个均应能表达成某m个M维向量的线性组合:

$$\mathbf{X}_i \approx a_{i1}\mathbf{A}_1 + a_{i2}\mathbf{A}_2 + \dots + a_{im}\mathbf{A}_m (*)$$

其中诸 $\mathbf{A}$ 称为端元向量,代表m个区系的典型站位,它们构成m维空间的“骨架”,可以是非正交的。如果系数 $a_{i1}, a_{i2}, \dots, a_{im}$ 中最大的一个是 a_{ik} ,并且 a_{ik} 比其余的各 a_{ij} 大许多,站位 $\mathbf{X}_i$ 即可划入以 $\mathbf{A}_k$ 为代表的那个区。各区的特征,即其学科涵义可由 $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_m$ 来阐明。

由此可见,因子分析法的本质是同时对N

个M维向量进行m元线性回归。

因子分析法的另一个优点是并不预先指定m和 $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_m$,它先设想式(*)都是些精确的等式,于是资料阵X的秩是m。X的各行归一化后得到的矩阵Y,它的转置阵 Y^* 的秩因而也是m,从而乘积 $Q = YY^*$ 是秩为m的N阶方阵。Q的不为零的特征根应有m个,对应了m个非零的特征向量 $\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_m$,称为m个数学端元。

由于(*)只是近似式,因而Q阵的秩,从严格的数学意义上说,当 $N > M$ 时,有可能大到等于M。但是当(*)能较好地成立时,这M个特征值中的最大的m个之和就占了全部特征值总和的绝大部分 δ 。通常取 $\delta = 0.97$ 或 0.95 ,并依此确定m。相应地,线性组合式

$$\mathbf{X}_i \approx b_{i1}\mathbf{B}_1 + b_{i2}\mathbf{B}_2 + \dots + b_{im}\mathbf{B}_m \quad (**)$$

在 δ 的精度水平上说明了各 $\mathbf{X}_i$ 的组成。

但是,正如英布瑞自己在文章中指出的那样,各系数 $b_{i1}, b_{i2}, \dots$ 并不总是大小悬殊,因而 $\mathbf{X}_i$ 很难判属哪个 $\mathbf{B}_k$ 群;并且最大的特征根对应的 $\mathbf{B}_1$ 往往只是一个“对称位置”,并不属于最大的一群。

因子分析法的第二步就是由 $\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_m$ 产生(*)中的实际端元 $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_m$ 来,这是通过一系列的方差计算、调整、迭代而在计算终结时产生的。

因子分析法也有其缺点。一是计算量很大;二是只适用于模式(*)成立的情况;三是不像枝状图法适用于许多相似性指标,因为Q的各元素 Q_{ij} 正是 $\mathbf{X}_i$ 和 $\mathbf{X}_j$ 间的Q因子($\cos\theta$ 系数)。

在使用时,由于(*)是要求M个参数均服从比例混和的模式,就要预先对资料阵X进行筛选。例如,在用矿物含量研究物质源分区时,要求参加统计的M种矿物的颗粒具有相近的水力学特性,否则(*)必定与实际情况不符。

因子分析法的最主要的特色就是引入了端元 $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_m$ 。

(下转第54页)

厚度。阿维斯 (Aves) 海岭上有晚中生代花岗岩赋存, 其波速为6.0—6.4公里/秒。

斯科歇 (Scotia) 海的海沟-岛弧-弧后盆地系统的形成与南美、非洲、南极大陆之间的分裂密切相关。海盆周围受安第斯造山运动影响形成古生代地层, 各岛仍为大陆性基底。

澳大利亚北部之巴布亚-新几内亚等均系中生代以前的陆壳基底, 它们是新生代从澳大利亚分裂开来的次生岛弧。深海钻探在海底采得的流纹岩说明这些陆壳岛弧已部分地沉没。

亚洲东南部古海沟-岛弧-弧后盆地系统在中新世澳大利亚与之碰撞后加速了大洋化进程。

陆壳岛屿塞舌尔岛已部分地沉没。陆壳海台的莫霍面深度一般达30公里, 其下地壳层由6.8—7.0公里/秒岩层组成, 厚度约14—18公里, 其上覆盖着一层速度为5.8—6.4公里/秒而厚度不等的陆壳层。

不同大地构造环境中的陆壳大洋化过程具有各自的特点。地中海型的大洋化过程受周围大陆的强烈影响, 表现出陆壳大洋化程度的巨大差异性、不彻底性和多次重复性。地中海周围山脉出露的多处蛇绿岩套表明侏罗-白垩纪地中海是个面积更为广大的洋盆, 即古特提斯海, 其范围包括黑海和死海在内。黑海的洋壳面积目前只占全海盆面积的12.2%。在死海与地中海之间进行的折射地震观测表明, 以色列南部地壳厚度超过30公里, 而北部地壳厚度剧降为15公里, 由死海向地中海方向地壳厚度迅速减小。这说明以色列北面地下原来是特提斯海洋壳海盆之所在。深海钻探发现, 距今500—700万年前地中海曾一度同大西洋分隔而干涸, 形成了很厚的石膏、硬石膏及岩盐等蒸发岩类。中新世晚期地壳下沉导致新的海侵, 乃再度出现今日之地中海。黑海也是一个年轻海盆, 仅200万年的历史。这次新生代开始的地中海陆壳大洋化范围尚未达到中生代的巨大规模。

大陆边缘的大洋化过程受海、陆地壳相互作用的强烈影响。这种类型的大洋化的本质是多次岩石圈断裂造成陆壳向大洋方向的蠕散及

残留陆壳在大洋地壳上的仰冲。

孤立型陆壳大洋化进程在大地构造方面处于有利环境, 其速度当较快。

不论大洋化陆壳处于何种大地构造环境, 陆壳大洋化都必须具备三个条件, 即陆壳的沉降、断裂及玄武岩浆活动。这些条件的逐步实现使陆壳大洋化进入不同阶段; 这些条件的同时实现将加速陆壳的大洋化进程。各不同条件对陆壳的大洋化具有不同的作用。陆壳的沉降改变陆壳原来的热状态, 主要是温度上升引起地壳结构变化。断裂使陆壳更容易受大洋环境的影响, 从而促进大洋化过程, 特别是岩石圈断裂诱发玄武岩上侵与喷发, 形成新洋壳将原来的陆壳取而代之, 完成大洋化过程。这些条件的实现在菲律宾海西北部陆块上表现的十分清楚。这个陆块自始新世以来下沉4000—5500米。始新世与渐新世的岩石圈断裂将这个陆块切割, 使之破碎。同时发生玄武岩浆喷发, 在破碎陆块周围形成新的大洋型玄武岩地壳, 部分地取代较老的陆壳。西南印度洋洋底的阿格哈斯 (Agulhas) 海台也出现了玄武岩侵蚀陆壳的明显迹象。(参考文献略)

(上接第59页)

笔者认为, 对因子分析法的改进可沿二个方向进行: 在统计地质学的范围内, 保留模式(*), 但不从Q阵的特征值及 $B_1, B_2, \dots, B_m$ 入手, 寻求新的简捷的确定 $A_1, A_2, \dots, A_m$ 的途径; 在模式计算地质学的范畴中, 则应更改模式(*)本身。

在第一个改进方向上, 目前已有些进展, 将另文介绍。

参 考 文 献

- [1] Imbrie, J. and T.H. Van Andel, 1964. Geol. Soc. Amer. Bull. 75 (11): 1131—1156.
- [2] McCamm, R., 1975. Concepts in Geostatistics. Springer-Verlag, Berlin-Heidelberg-New York.