

一种化探异常评价方法——多变量场

张丕富

(河北地质学院)

用多变量场 (Multivariate Field) 评价化探异常, 是在多维变量或主轴空间, 利用超球面把具有相同或相似地球化学性质的样品点分别圈出来, 从而达到对异常进行分类、评价的目的。本文就多变量场的原理与方法作了简介, 并对其计算机算法作了说明。

关键词: 异常评价; 多变量场; 超球面

问题的提出

化探是一种行之有效的找矿方法, 已得到广泛应用。目前, 除边远地区外, 全国1:20万的化探扫面工作已经完成, 重点地区大比例尺的化探详查也在进行。大批化探数据的输入, 如何快速、有效地圈定化探异常并给以正确的评价, 就成了亟待解决的问题。

众所周知, 化探工作不仅所取的样本大, 而且同时跨越众多的地质单元 (如不同时代的地层、岩体、构造蚀变带等) 或地球化学单元 (如背景区、无矿异常、有矿异常、不同元素组合异常等)。就是说样品来源复杂, 源于多重母体。目前较常采用的化探异常评价方法是多元统计分析方法, 如聚类分析、因子分析等。由于其要求参加统计的样品服从正态分布或对数正态分布, 即要求样品源于均一母体。因此, 用多元统计方法评价化探异常是有缺陷的。本文介绍的化探异常评价方法——多变量场法则可对样本中不同来源的样品进行自动归类, 即圈出异常, 然后再给出异常的评价指标。该方法原理明了, 方法简单并能达到快速、准确地评价化探异常的目的。

多变量场的原理

尽管化探样本可能取自于若干个不同的地球化学单元 (母体), 然而, 取自于同一地球化学单元的样品, 其同一化学元素在含量上是彼此接近的。不难理解, 在画异常等值线时, 按元素含量高低分界, 等值线周围的样品含量彼此接近。就这些样品来说, 各元素间应存在一种基本的比例关系, 即元素的组合关系。这一点是多变量场的特色, 它不是简单的异常勾划过程, 而是综合考虑各元素异常在空间上的内在联系。也就是说, 均一地球化学单元有自身的样品组合与元素组合。这就是本方法对异常分类与评价的地球化学准则。

用数学语言来讲, 上述思想可以简单地表达为: 在多维变量或主轴空间中, 属于同一类的样品点, 其在空间的分布应处于彼此接近的地方 (图1); 然而, 在空间的分布彼此接近的样品点不一定是同一类, 还要通过均一性的检验才能确定是否为同一类。所以, 对样品点还有一个不断筛选的过程。而利用在多维变量或主轴空间中, 能不断移动又具伸缩性的超球面就可以把不同类型的样

品点分别圈出来,从而达到异常分类与评价的目的。

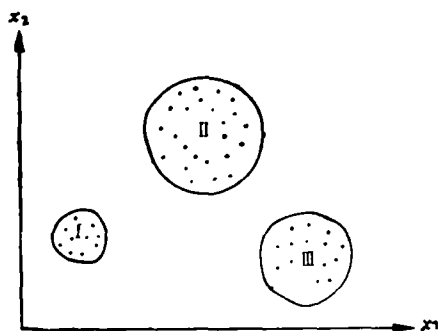


图 1 不同类样品分布散点图
 x_1, x_2 —变量或主轴; I~III—样品类编号

多变量场的方法步骤

1. 原始数据的预处理

设原始数据矩阵 $Z = [z_{ij}]_{n \times m}$ 。其中, z_{ij} 为第 i 个样品在第 j 个变量上的观测值。

为了消除各元素在地壳中丰度不同而造成样品分类的误差,在对样品进行分类之前,对矩阵 Z 进行适当的变换是必要的。本方法通常采用如下几种变换方法:

(1) 主轴变换 用因子分析、主成分分析或其他线性、非线性变换将多维变量空间转换成低维主轴空间,然后在主轴空间中给样品进行分类。

(2) 对数转换

$$z'_{ij} = \log z_{ij}, \quad i=1, 2, \dots, n; \quad j=1, 2, \dots, m$$

(3) 衬度转换

$$z'_{ij} = \begin{cases} z_{ij}/a_j - 1, & z_{ij} \geq a_j \\ 1 - a_j/z_{ij}, & z_{ij} < a_j \end{cases}$$

或 $z'_{ij} = z_{ij}/a_j$

$$i=1, 2, \dots, n; \quad j=1, 2, \dots, m$$

其中, a_j 为第 j 个元素的背景值或克拉克值。

(4) 标准化转换

$$z'_{ij} = (z_{ij} - \bar{z}_j) / z_{j \min}$$

其中, $\bar{z}_j = \frac{1}{n} \sum_{i=1}^n z_{ij}$, $z_{j \min}$ 为第 j 个变量的最小值。 $i=1, 2, \dots, n, j=1, 2, \dots, m$ 。式中除以最小值的目的是消除背景的影响而突出表现高含量的异常点。

2. 样品的分类

化探异常是由取自异常的样品点所表现出来的,对样品的分类实际上就是对异常的分类。

经过预处理后的矩阵 Z 用超球面对样品进行分类,具体方法如下:

在多维变量或主轴空间中,选择一个点 Q_0 作为超球面计算起点。经过预处理后,样品点的重心以及源于背景区的样品点皆在坐标原点附近。也就是说原点附近聚集大量的样品点,因此,赋 $Q_{0j} = 0 (j=1, 2, \dots, m)$, 并以其作为初始超球面的球心。同时,以 $R_0 = 5 \ln R$ 为初始超球面的半径,其中 R 为全部样品重心 $\bar{Y}_j = \frac{1}{n} \sum_{i=1}^n z_{ij}$ 到最远一个样品点的距离。于是初始超球面为 $S_0(Q_0, R_0)$ 。算出 S_0 中各样品点的重心 Q_{1j} , 若 Q_{1j} 与 Q_0 重合,即 $Q_{1j} = Q_0 (j=1, 2, \dots, m)$, 则认为 S_0 中的样品为一组合;否则,以 Q_{1j} 为新球心,计算 $S_1(Q_{1j}, R_0)$ 中样品点的重心 Q_{2j} , 再拿 Q_{2j} 与 Q_{1j} 比较,如此反复,直至下一个超球面 S_{r+1} 的重心 Q_{r+1j} 与其球心 Q_{rj} 重合为止。

上述圈出的样品组合只能说其样品点在空间上是彼此接近的。前已述及,彼此接近的样品点还不能说它们就是同一样品组合,还必须满足样品点具有均一性的要求。用主成分分析来检验组中样品点的均一性具有其独到之处,因为主成分分析能最佳地逼近样品点的总体。它使组内的样品点以原坐标系转换到一个新的坐标系中去,新坐标系中的每个坐标(主成分)是原坐标系特征的线性组合。

假设上述分组计算得出的一组数据矩阵

为:

$$Z_1 = [z_{1j}]_{n_1 \times m}$$

其中 n_1 为组内的样品数。

主成分分析第一步计算 z_1 的协差阵为:

$$s_{ij} = \frac{1}{n_1} \sum_{k=1}^{n_1} (z_{1k_i} - \bar{z}_{1i})(z_{1k_j} - \bar{z}_{1j})$$

其中, $\bar{z}_{1j} = \frac{1}{n_1} \sum_{k=1}^{n_1} z_{1k_j}$, $i, j = 1, 2, \dots, m$

第二步计算 $[s_{ij}]_{m \times m}$ 矩阵的特征值及特征向量:

特征值为: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$

特征向量为: $U_j = (u_{1j}, u_{2j}, \dots, u_{mj})'$, $j = 1, 2, \dots, m$

特征值是研究组中样品均一性的特征标志。各特征值的大小分布愈均匀,表明组中样品愈不均一。因为均一的地球化学单元是有其特定的元素组合的(主成分),因而其第一特征值往往远大于其余特征值。所以,用第一特征值 λ_1 在所有特征值总和中所占的比例来评价组中样品的均一性是可行的。其表达式为:

$$P_1 = \lambda_1 / \sum_{i=1}^m \lambda_i$$

经验表明, $P_1 \in (0.7, 0.95)$ 时, 分类效果较为显著。还可根据所要求的分组数目来适当拓宽和缩小 P_1 的范围。

上述分组如不能满足 P_1 的要求, 说明组中样品还未达到均一性, 需要重新改变球心与半径再进行计算, 其基本计算过程见图2, 直至满足 P_1 的要求才进行下一步运算。

3. 异常评价参数的输出

(1) 变量或主轴对分组的贡献: 既然第一特征值是评价组中样品均一性的决定因素, 那么, 其对应的特征向量应是对分组贡献的特征指标了。用 $I_j = (u_{j1})^2 / \sum_{k=1}^m (u_{kj})^2$, $j = 1, 2, \dots, m$, 来表示第 j 个变量或主轴对分组的贡献。 I_j 越大, 表示第 j 个元素(组

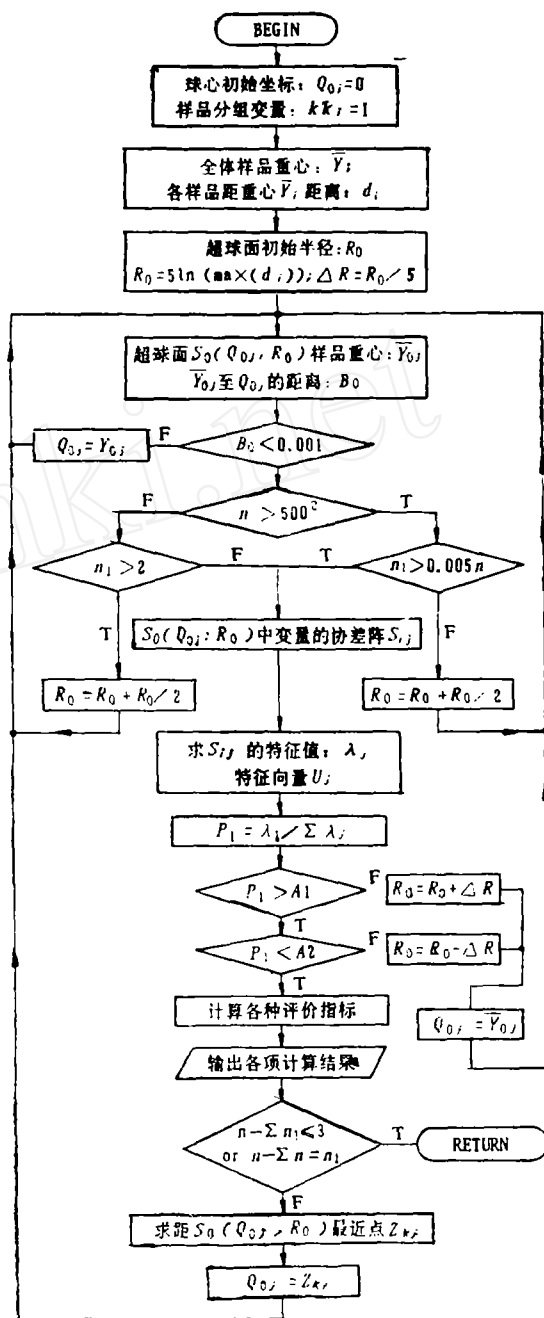


图2 超球面算法框图

合)对异常的贡献越大。在某种程度上 I_j 类似于因子分析中的因子载荷。

(2) 超球面的半径和均一性指数。

(3) 分组样品的类别号。

(4) 均值(球心)均、方差和变化系数。

化探异常多变量场分析实例

类 别	因 子	贡献百分比	均 值 (球心)	均方差	变化系数	超球面 半 径	均一性 指 数	异常元 素组合	异常性质
4	F_1	0.33	-0.26	0.64	-2.46	3.97	0.71	Pb、Zn、 Sn、Mo	铅锌矿、黄 铁矿化异常 区
	F_2	98.45	-4.63	0.97	-0.12				
	F_3	0.33	0.35	0.39	1.12				
	F_4	0.11	-0.11	1.00	-9.35				
	F_5	0.78	0.33	0.99	3.04				

由上述输出的参数可以对各异常进行有效的评价。用样品类别号可以对不同异常进行圈定。从分组贡献可以了解异常属性,即异常的元素组合及各元素在异常形成中的贡献。从超球面半径和均一性指数中了解各样品在空间聚集的紧密程度和异常均一性度量。从均值、方差和变化系数上可以确定异常的规模及各元素分布的连续性。由此可见,本方法对地球化学异常的分类与评价是有其独到之处的。

如浙江某地化探详查工作,原始数据经主轴变换,采用因子分析后,使原来的11个元素组合成为5个因子,然后进行分类。这里只列举7类异常中的第4类供参考(见上表)。

由表可知,第4类异常其实是 F_2 因子的异常。 F_2 因子的元素组合为Pb、Zn、Sn、Mo,代表本区铅锌矿、黄铁矿化。其均值(因子得分均值)明显偏离坐标原点(均值为0),说明异常强度大。均一性指数为0.71,表示本类中的样品基本达到均一性的要求,其较低的均方差和变化系数表明该异常区的矿化比较均匀。如能和野外的地质、地球化学特征有机地结合起来,对异常的解释与推断将更为准确,更能显示本方法的优越性。

多变量场的实现

为了使多变量场方法能在化探中发挥作用,作者配备了该方法的计算程序。源程序MUTVFD用FORTRAN77语言编写,并在IBM-PC系列微机上通过。下面将MUTVFD运行时所要输入几项参数作如下说明。

1. A_1 , A_2 ——均一性指数下、上界,一般赋 $A_1=0.7$, $A_2=0.95$ 。还可根据所需结果进一步调整。

2. N, M, FIL——原始数据的样品数,变量数及其文件名。

3. UNIT——选择输出设备。UNIT=1时,屏幕输出;UNIT=2时,以ZPF.DAT为文件名磁盘输出。

4. IOP——选择原始数据的预处理方法。当IOP分别赋0, 1, 2, 3, 4时,原始数据为不变换,主轴变换,对数转换,衬度转换和标准化转换。

在本文的撰写过程中得到了长春地质学院吴锡生研究员的支持和帮助,在此谨表谢意!

参 考 文 献

- [1] 张丕富,河北地质学院学报,1988,第4期。
- [2] 地矿部情报所,化探资料选编,1984,第6期。

A Method for Geochemical Anomaly Evaluation— The Multivariate Field Method

Zhang Pifu

In the evaluation of geochemical anomalies, multivariate field method is used to delineate the sample points of equal or similar geochemical characteristics respectively with the help of superspherical surface in multivariate space or principal axis space, in order to attain the goal for the classification and evaluation of geochemical anomalies. The basic theory and concrete practice of this method are briefly introduced in this paper. An explanation of the computational algorithms for this method is also given.