数学地质漫谈

侯景儒

一 数学地质的出现是地质科学发展的必然产物

革命导师马克思曾经说过："**科学仅当它成功地利用数学时才达到完善程度**。"逐步发展并开始推广应用的数学地质，正是这一伟大预见的具体实践和必然产物。

地质工作的迅速发展，仪器分析和记录自动化程度的提高，使各种地质观测数据和实验数据大量增加。这就向地质工作者提出了如何对大量的地质数据及时而系统地进行整理、综合分析，从而更深刻地揭示地质现象和理解地质过程的新课题。随着国家对矿产资源需求的日益增长，地质工作的难度也愈来愈大，同样要求对各种地质体在空间和时间上的变化提出较为精确的定量评价，从而为普查勘探和矿山开采提供可靠的工作依据。为了有效地指导找矿，也要求我们从经过漫长地质历史时期的地质体的各种复杂变化中抽出规律性的东西来。可见，仅用直观地、经验地定性描述是无法做到的。在这种情况下，好象本世纪20年代传统地质学与化学、物理学结合产生了地球化学及地球物理学（相应的出现了化探、物探）一样，60年代初，地质学开始与数学及现代计算工具——电子计算机——相结合产生了数学地质。

数学地质目前已涉及到诸如地层、构造、矿床、地球化学、地球物理、地貌、水文、古生物以及成矿预测等各个方面，并开始在国内外应用。尽管数学地质工作方法本身还存在着各种问题，但它在地质找矿的生产、科研实践中已显出应有的生命力，并收到了一定的地质效果。

从数学地质的发展历史中更能体会到马克思预言的无比正确，也能更好地说明在地质科学发展的进程中，地质工作者是怎样努力把地质学与数学结合起来，以便将地质现象从传统的定性解释中解脱出来使其定量化。

十九世纪中叶，英国的莱伊尔（C.Lyeell，1840年）首先应用统计方法把第三纪地层中的生物种属与现代生物种属相比较，依此划分第三纪地层。此后，地质工作者开始注意把地质现象的定性观测解释逐步定量化。十九世纪末，在地层学、矿物学、岩石学、沉积学等领域开始应用频数分布、方差分析等单变量统计方法。二十世纪40年代，开始引入简单的回归分析、相关分析等双变量方法，用它来研究矿床及地球化学的一些问题。到了50年代，电子计算机诞生了，它使整个科学技术的发展飞跃到一个新阶段，也给地质学和数学进一步结合，并使地质解释定量化提供了有利条件。

在酝酿了一个世纪之后，本世纪60年代，地质学与数学终于有机地结合起来产生了"数学地质"这一新的学科。至此，一些多变量统计方法开始成功地应用于地质学，解决那些地质工作中长期不易解决的问题。在矿床成因、地层与矿化期的划分、地层与矿层的对比、异常评价、岩石与矿物分类等问题的定量解释方面，迈进了一大步。目前，国内外已开始研究

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

和应用地质过程的模拟模型和概率模型。可以预计，在毛主席伟大旗帜指引下，有华主席的英明领导，有社会主义制度的优越条件，我国的数学地质一定会在短期内走在世界的前列，为在本世纪把我国建设成为具有四个现代化的强大的社会主义国家贡献出力量。

二 数学地质的研究方法

数学地质内容广泛，本文所涉及到的、即一般常用的，确切地说是“多元分析方法”。多元分析是相对一元分析而言，前者是后者的推广，从一个变量推广到多元随机变量相互关系的研究。当然，许多多元分析方法并非由一元分析推广而来，它有自己的特殊理论。

从地质、地球化学观点来研究，多元分析方法大致有以下几方面（于崇文，1976）：

1.地质体属性的研究

1）相关分析与回归分析，
2）点群分析，
3）判别分析，
4）典型相关分析，
5）因子分析，
6）非线性映射分析（Nonlinear mapping technique），
7）对应分析（Correspondence analysis）；

2.地质体特征的空间研究

1）趋势面分析，
2）调和趋势面分析，
3）典型（Canonical）趋势面分析，
4）单位向量场分析（Unit vecter fields）；

3.变量的空间相关，地质统计分析

4.参数的时间变化，马尔科夫概型分析。

下面仅就常用的判别分析、点群分析、因子分析和趋势面分析作一简单介绍，而一些基本的方法如频率分布、回归分析、相关分析已有许多专著，不再赘叙；对于目前正处于研究阶段的方法，待以后再作讨论。

三 判 别 分 析

判别分析是一种分类方法。分类学的共同问题是根据两类（或多类）个体的某些属性来分辨或判别两类（或多类）个体。

当两类事物具有一个很不相同的属性时，就很容易将其分开，但当两类事物，例如两种矿物或岩石的几种特征都非常相似时（如阳起石与透闪石），只靠一种特征就很难正确地将两者分开，因为这些特征的观测值有很大的重叠性。假若我们测定了几种特征（即几个变量或称“多元”），并将其结合起来，用数学方法找出它们的一种适当的线性组合，就可使重叠程度大大减小，从而正确地分辨两类事物。这个线性组合，可以作为一种标志，借以区分两类（或多类）事物。这种方法就称为“判别分析”。因此可以说，判别分析是把几个变量构成一个或几个线性组合（称为“判别函数”），利用该线性组合（判别函数）将多变量的观测值区分成两组或多组的多元分析方法。

解决这个问题的基本途径有二：

1.判别函数必须满足费歇（Fisher）准则。即组间均值差与组内方差之比要最大，也就是要求组之间的均值差愈大愈好，而组内方差要小。

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

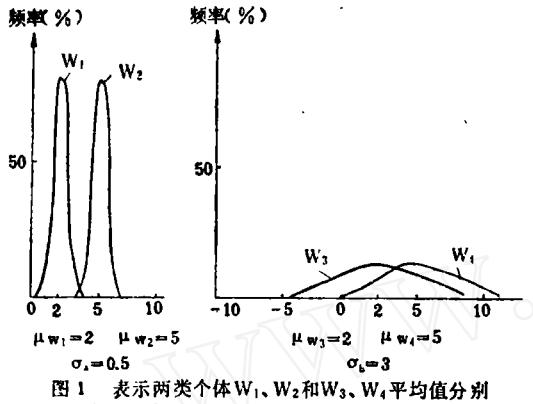

图1 表示两类个体W_1、W_2和W_3、W_4平均值分别相同而标准离差不同时必须使$|\mu_{w_1}-\mu_{w_2}|/\sigma_a$ ⇒ Max时才能有效地将两类个体分开

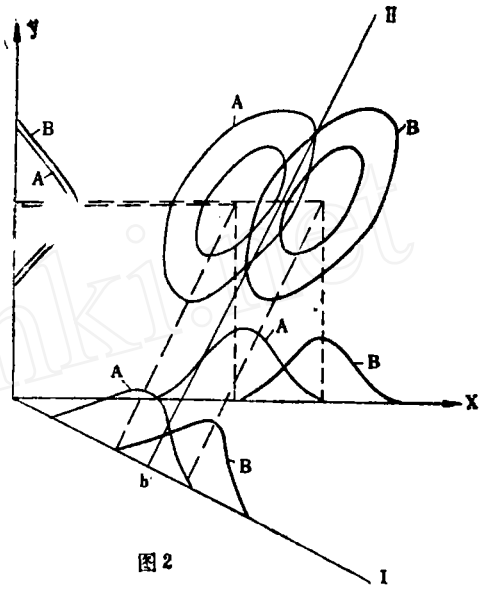

图2

2.判别函数必须满足贝叶斯（Bayes）准则。即以归属某类的概率最大或错分损失最小为原则。

现就费歇准则进一步解释如下：

设有两个个体为W_1，W_2；W_1的平均值$\mu w_1=2$，W_2的平均值$\mu w_2=5$，两者组内离差$\sigma_a=0.5$，两者组间均差值$\phi a=3$；

另有两个个体W_3，W_4；$\mu w_3=2$，$\mu w_4=5$，$\sigma_b=6$，$\phi_b=3$。

如图1所示，W_1，W_2和W_3，W_4的平均值分别相同，组间均差值均为3，但前者组内离差$\sigma_a=0.5$，后者$\sigma_b=6$，按照费歇准则：W_1，W_2的$\phi_a/\sigma_a=3/0.5=6$，W_3，W_4的$\phi_b/\sigma_b=3/6=0.5$。

很明显，W_1，W_2较W_3，W_4容易分开。因此，为使两者有效地分辨开来，就必须在两类个体间均值差与两类个体内标准离差之比为极大。

下面再以两个变量为例，进一步说明这个问题。

图2中两个椭圆表示A类或B类个体等密度点的轨迹，即A类（或B类）90%的个体均落在A类（或B类）个体外圈椭圆范围之内。A、B两类个体外圈椭圆之交点确定一直线Ⅱ，作直线Ⅰ垂直直线Ⅱ，并将A、B两类在二维空间的图点投影在直线Ⅰ上。由此可见，在直线Ⅰ上，两类个体之间的重叠小于任何可能直线（如x轴或y轴）上的重叠。直线Ⅰ即代表了多个变量线性组合的线性判别函数。因此，从图2可知，仅用x或y一个变量不能很好地将A、B两类母体分开，而x、y的线性组合则能有效地判别两类个体。

基于上述主导思想，判别分析的步骤如下。

（1）求判别函数 设A、B分别代表已知两类事物，它们的样品数分别为n_A、n_B，每一类事物均有P个变量。先按已知样品构成判别函数：

$$y=\sum_{i=1}^{P} C_iX_i$$

其中X_i为参加判别的变量数值（如元素Cu、Pb、Zn的含量）；

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

Ci为特定的判别系数。

为了求出Ci，假设$\bar{y}_A$、$\bar{y}_B$分别表示A、B两类的平均值，则两类之间的均值差为$\bar{y}_A - \bar{y}_B$。再设y_{AK}、y_{BK}分别表示A、B两类第K个样品的平均值，则A、B两类内判别函数y值的离差为：

$$\left[\sum_{K=1}^{nA}(y_{AK}-\bar{y}_A)^2+\sum_{K=1}^{nB}(y_{BK}-\bar{y}_B)^2\right]$$

根据费歇准则

$$\frac{\bar{y}_A-\bar{y}_B}{\left[\sum_{K=1}^{nA}(y_{AK}-\bar{y}_A)^2+\sum_{K=1}^{nB}(y_{BK}-\bar{y}_B)^2\right]}=I\Rightarrow极大$$

经过一系列数学运算，最后得出一线性方程组，据此求出判别系数Ci，并得出线性判别函数：$y=\sum_{i=1}^{P} CiXi$。例如，变量为Cu、Mo、Ag、Zn元素时，则$y = C_1Cu + C_2Mo + C_3Ag + C_4Zn$。

（2）求判别值（判别标准） 先求A、B两类事物判别函数的平均值

$$\bar{y}_A=\sum_{i=1}^{P} CiX_{Ai},\qquad \bar{y}_B=\sum_{i=1}^{P} CiX_{Bi}$$

据$\bar{y}_A$、$\bar{y}_B$求判别值$\bar{y}_c$。

求$\bar{y}_c$的方法有二，这里只介绍两类样品数的加权方法：

$$y_C=\frac{n_A\bar{y}_A-n_B\bar{y}_B}{n_A+n_B}$$

（3）求待判样品的判别计量 将未知样品用判别函数y求出判别计量ye，看ye大于或小于y_c来判别它的归属。

（4）检验判别效果。

以上所述只涉及到两类个体的判别，叫做“二级判别分析”。在实际工作中往往要对多类个体进行分辨，例如，矿化期的划分、岩体属性及含矿性、铁帽成因等，这种多个母体的分辨称为“多级判别分析”。它的基本思想和解释问题的途径与二级判别分析相似，只不过更为复杂些。

在“二级判别”或“多级判别”分析中都是利用所引入的全部P个变量$x_1, x_2, \cdots\cdots x_p$来判别两类或多类母体。然而，实际工作发现：已被引入判别函数的变量往往由于引入新的变量而失去原有的分辨能力，以致必须将其从判别函数中剔除。即不是变量愈多，判别效果越好，有时适得其反，会降低判别效果。所以，在考虑种种因素之后，在进行判别分析时，要从全部P个变量中选择几个合适的变量，使其能“最优地”判别多个母体。这种对变量进行的筛选，以期得出最优判别函数的方法称为“逐步判别分析”（Stepwise discriminant analysis）。“多级判别分析”和“逐步判别分析”由于计算繁杂，这里不作叙述。

地质及化探工作中的大量实践证明，判别分析是一种较为客观而可靠的分类方法。它可用于岩体含矿性、矿点露头、铁帽及异常的评价，地层及岩相的划分，成矿标志、矿物及古

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

生物的分类等研究方面。下面举两个实例以说明其应用效果。

例一 鞍本地区辽河群和鞍山群的区分问题

众所周知，鞍本地区前震旦系地层由于混合岩化而使其对比困难。有人认为该区前震旦纪地层是连续沉积的，中间无大的间断；许多人认为前震旦系地层可分为两大沉积旋回 即太古代的鞍山群和元古代的辽河群，中间以炮台山砾岩分开。产生这种分歧的原因之一是两群地层交界处之千枚岩肉眼及镜下分辨都有困难。如果能找出两群地层的地球化学标志及其数学模式，则将有益于其划分。为此，我们在四花岭采了较为可靠的鞍山群樱桃园组顶部千枚岩，在胡家庙子采了较为可靠的辽河群浪子山组底部千枚岩，将其作为已知样品进行两类判别。参加判别的元素是TFe、Fe_2O_3、FeO、SiO_2、Al_2O_3、MgO、CaO、K_2O、Na_2O、MnO、TiO_2。经变量筛选后，仅引入FeO、CaO、MgO、MnO参加判别函数。各类判别函数的常数项及系数列如表1。

由表1可以写出浪子山组底部千枚岩的判别函数为：

$y_1 = -19.1880 + 6.6029FeO - 0.7459CaO + 5.9735MgO - 11.5155MnO$

樱桃园组顶部千枚岩的判别函数为：

$y_2 = -27.2780 + 25.3260FeO - 21.5900CaO - 11.9730MgO - 1.5350MnO$根据上述判别函数对这两组千枚岩所采全部样品进行判别，结果如表2。

经统计检验，判别效果良好。对炮台山砾岩进行判别时，全部划入樱桃园组，这是因为砾岩的成分与樱桃园组地层相同所致。因此，上述两群地层交界处的千枚岩尽管非常相似，但应用判别分析却能很好地将二者分开。这说明两者具有明显地地球化学差异，从而引起数学模型上的不同。再考虑到两者岩相上的区别及因子分析的成果（详见第五节因子分析），可以考虑鞍山群和辽河群是在两种不同的地球化学和地质环境下沉积的。

例二 桂东南铁帽的判别分析（据苏树钧、张先容等，1975）

广西铁帽多，变化大，给判定其成因带来很大困难。为了评价铁帽成因，先选择了已知成因的10个铁帽（89个样品）进行工作。它们分属于5个类型（表3）

这10个铁帽中参加判别分析的变量共有7个元素：Cu、Pb、Zn、Sn、Co、As和Ni。计算后各类判别函数的常数项及各变量的系数如表4，判别结果如表5。

表1

系数 / 类别	常数项 C_0	FeO	CaO	MgO	MnO
1	−19.1880	6.6029	−0.7459	5.9735	−11.5155
2	−27.2780	25.3260	−21.5900	−11.9730	−1.5350

表2

类别	样品数	划入1类样	划入2类样	划入各类样，%
1类（辽河群浪子山组底部千枚岩）	25	25	0	100
2类（鞍山群樱桃园组顶部千枚岩）	15	0	15	100

表3

类别	1	2	3	4	5
成因类型	夕卡岩型	多金属硫化物型	黄铁矿型	菱铁矿型	赤铁矿型
样品数（个）	20	23	14	9	23

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

判别函数的系数及常数项（七个变量） 表4

类 别	常数项C_0	Cu的系数	Pb的系数	Zn的系数	Sn的系数	As的系数	Co的系数	Ni的系数
1	−77.2250	5.6574	3.7043	11.9931	16.6099	4.0704	19.5034	3.5301
2	−70.2504	18.5083	3.6359	7.6712	6.7683	4.5124	27.6936	9.5080
3	−35.0578	21.8017	−1.6532	4.2364	−0.5388	4.7330	7.2353	2.0908
4	−41.3491	11.3200	1.1598	3.2940	6.5860	3.7218	31.9734	9.8747
5	−14.8916	9.5628	0.3911	3.5667	2.6066	2.3449	11.7267	4.0659

从表5得知，89个样品只判错3个，可见判别效果良好。在此基础上，对9个未知（待判）铁帽（164个样）进行了五类判别，从而给这些铁帽的评价提出了地球化学依据。

知已标准样的判别分类（七个变量） 表5

原分类 \ 判别分类	样数（个）	划入类别（个数）					注
		1类	2类	3类	4类	5类	
1 类	20	19			1		错判3个，占3.4%
2 类	23		23				
3 类	14			12		2	
4 类	9				9		
5 类	23					21	
总 计	89	19	23	12	10	23	

四 点群分析（又称群分析，簇群分析）

如前所述，判别分析需首先在研究区找出标准样品（如含矿和不含矿岩体），藉以建立判别函数，然而，在一个研究程度较低的地区这种要求很难达到，因而使判别分析的应用受到阻碍；点群分析却能弥补其不足。

点群分析是根据母体的多种地质或地球化学标志的相似性，对母体进行分类的一种方法。它应用数学方法对样品进行数字分类，定量地确定样品相互间的亲疏关系，并以其亲疏差异程度，归入不同的分类单位。

点群分析的基本思想，是采用聚合的方法；即选择一个统计量（如距离式相关性）作为相似的程度，把最相似的样品（如图3中距离最近者）联在一起，构成一个新样品。如图3中第4、第6样构成4—6样品，第3、第5样构成3—5样品，再找出与新样品最相似的（距离最近者）样品，再形成一个新样品，如此下去，最后作出一张谱系图（图4）。按其相似程度将样品分成几类。图4中，按0.2的相似性可分为两类：4，6为一类，1，2，3，5为另一类。

根据被分类的对象不同，点群分析可分为R型分析和Q型分析是按样品对变量进行分类，Q型分析是按变量对样品进行分类。下面就两种分群作一简单介绍。

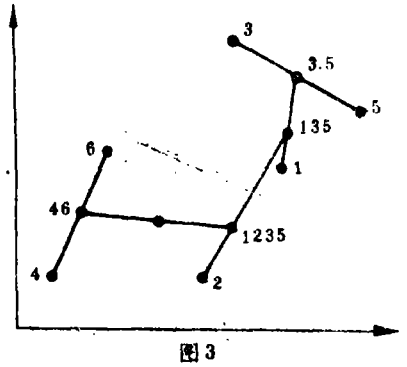

图3

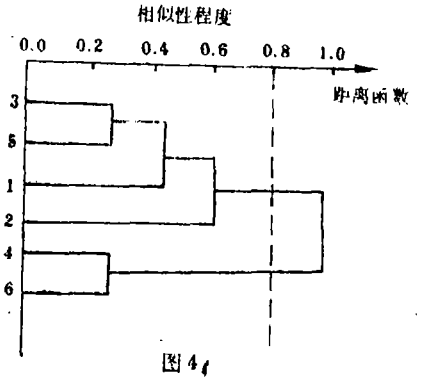

图4

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

1.R型分析 一般用的统计量为相关系数R，变量i和K之间的相关系数Rik为：

$$R_{iK}=\frac{\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)(x_{Kj}-\bar{x}_K)}{\sqrt{\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)^2\sum_{j=1}^{n}(x_{Kj}-\bar{x}_K)^2}}$$

其中：x_{ij}为第j个样品第i个变量的观测值；

x_{Kj}为第j个样品第k个变量的观测值；

$\bar{x}_i$为第i个变量的平均值；

$\bar{x}_K$为第k个变量的平均值

相关系数R_{iK}实际上是变量i和k之间夹角的余弦$\cos\theta$（图5）。从图5可知，当$\theta=0°$时，R=1；当$\theta=90°$时，R=0；R变化在0与1之间。

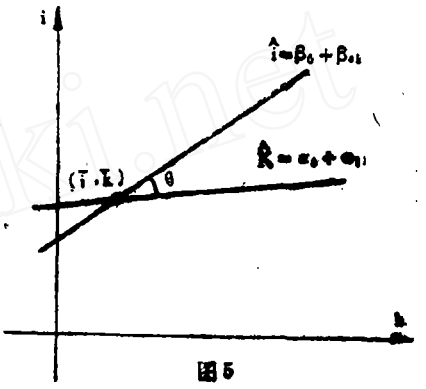

图5

2.Q型分析 度量相似性程度的统计量常用的有相似性系数和距离函数。

（1）相似性系数S 若把两个样品j，k当作向量，它们之间的相似性程度可用两向量之间的夹角余弦$\cos\theta$来度量（图6），并用S表示，θ越小越相似，越大越不相似。样品j，k之间相似性系数为：

$$S=\cos\theta_{j,k}=\frac{\sum_{i=1}^{m}x_{ij}\,x_{ik}}{\sqrt{\sum_{i=1}^{m}(x_{ij})^2\sum_{i=1}^{m}(x_{iK})^2}}$$

其中：i为变量，共有m个。

（2）距离函数D 如图7所示，样品j，k之间的距离D_{jk}为：

$$D_{kj}=\sqrt{\frac{\sum_{i=1}^{m}(x_{ij}-x_{ik})^2}{m}}$$

从而可知，距离函数越小越相似。

此外，统计量还有误差平方和、斜交距离等，这里不作介绍。

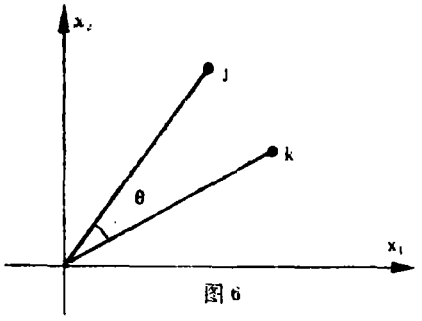

图6

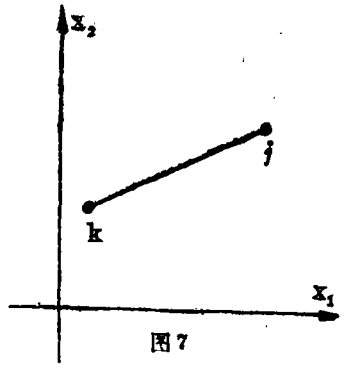

图7

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

点群分析在实际工作中应用成功的实例很多，而且应用范围也日趋广泛（如异常分类，矿床类型、岩相及地层的划分等）。现举例如下：

例一 为了研究我国北方几个风化淋滤型铁矿床（点），我们将其与国外几个风化淋滤型铁矿作了点群分析。参加点群分析的矿床（点）有：西鞍山M15坑，晋北花子、硃砂洞、袁家村，舞阳的岗庙刘，冀东司家营—马城一带；苏联库尔斯克磁异常区的雅克夫列夫矿床、克里沃罗格矿区的阿诺夫斯克及萨克萨甘矿床，美国的米萨比矿床。此外，特别将鞍本地区的弓长岭富矿和樱桃园富矿也一并引入，参加分类，目的在于鉴别上述各矿床（点）中是否有这类富铁矿床。参加分群的变量有TFe、Fe_2O_3、FeO、SiO_2、Al_2O_3、CaO、MgO等七个氧化物，其谱系图如图8。

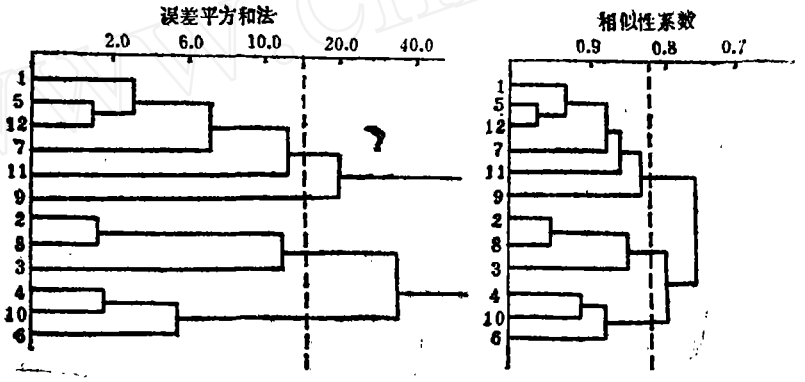

图8 国内外几种风化淋滤型铁矿点（床）谱系图

I—西鞍山M 15坑；2—弓长岭富矿；3—樱桃园富矿；4—岗庙刘；5—花子；6—袁家村；7—硃砂洞；8—司家营—马城；9—雅克夫列夫；10—阿诺夫斯克；11—萨克萨甘；12—米萨比

由图8可见，若取相似性系数为0.82时，结合各矿区地质、矿床及地球化学特征，可将上述矿床（点）大致分为三个典型：

（1）西鞍山M15坑和晋北花子、硃砂洞可归为一类，它们可与美国的米萨比、苏联的雅克夫列夫、萨克萨甘矿床类比。

（2）舞阳岗庙刘和晋北袁家村为一类型，可与阿诺夫斯克矿床类比。

（3）司家营—马城一带的富矿可与鞍本地区弓长岭、樱桃园富矿床类比。

前二者均属风化淋滤型富铁矿，而第三类为深变质热液矿床。

例二 红旗岭铜镍矿床周围发现基性—超基性岩体30多个，其中已查明有大、中、小型矿床6个，为了对其余20多个岩体进行含矿性评价，作了点群分析。选用的变量为Cu、Co、Ni、S四个元素，图9为其谱系图（据郑闲石等，1973）。

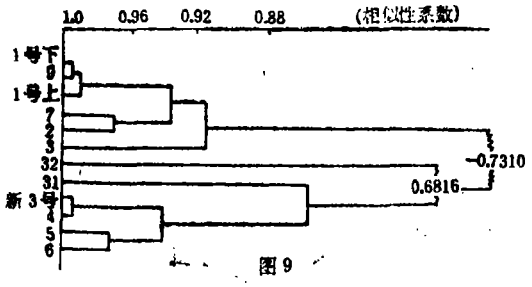

图9

由图9可知，相似性系数取0.7时，可将12个岩体分成两大类：即1号下，9号，1号上，7号，2号，3号岩体为一类，为含矿岩体。而32号，31号，新3号，4号，5号，6号岩体为另一类，属于矿化或无矿岩体。若将相似性系数提到0.86时，又可将第二类岩体划分为矿化岩体（31，32号岩体）及无矿岩体（新3号，4号，5号，6号岩体）。

（未完待续）

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net