目　录

原书前言
计算普查网度的图解法…… 1
岩石和矿石物质快速分析的综合方法……11
地球化学资料统计处理时分析误差的估计和测量
　误差的计算……24
客体分类、地球化学场图示及解释时效益率的计
　算和等级近似法……28
地球化学图的编制及地质解释实例……53
利用内插法和趋势分析编制地球化学图和剖面图……82
论残留地球化学晕的解释……96
利用最大似然法对地球化学异常进行分类的方法…… 105
矿体及矿体晕分带性经验模型的建立…… 119
主分量分析法在解释地球化学资料中的应用…… 138
地球化学和地球物理资料的综合解释…… 147
论选择、论证和建立数据数学处理方法的理论问题…… 166

计算普查网度的图解法

Г.А.巴扎诺夫　Г.С.瓦赫罗麦耶夫

首先利用概率-统计方法来论证勘探地球物理和应用地球化学普查网度的是A. П. 索洛沃夫(1959)。后来，这一方向在И. Д. 萨文斯基（1964），A. 库拉拉（1964）、Д. И. 萨弗拉索夫(1965)、О. Г. 谢苗诺娃和Ю. Д. 叶夫多基莫夫（1970）、А. М. 舒雷金（1972）等人的文章中进一步得到发展，他们提出了各种解析公式、表格和计算机程序。上述方法都很精确，但由于计算方法不一定都能搞成表格和列线图，故其中大多数使用起来是很费事的，既使有这样的表格和列线图，也会由于要填入的数据太多而感到困难，何况往往是需要简单而迅速地估算出最合适的观测网度。为此，可以利用И. В. 弗兰茨基和Г. А. 巴扎诺夫（1969）为确定沿走向延伸的客体起码被一个探矿工程所穿切的概率而制作的列线图。但在勘查地球化学中，只有被一条剖面上的两个或两个以上测点查明的异常才被认为是可靠的。因此，又提出了计算满足这些条件的新列线图的任务。

我们来探讨事件C——最低限度用两个点来揭示异常。这个事件包括另外两个须协同完成的事件：事件A——最低限度有一条普查线穿切普查客体；事件B——每条普查线上最低限度有两个点落在普查客体上。因为这些事件是互相依赖的，事件C的概率可以写成：

$$P(C)=P(A)\cdot P(B|A) \qquad (1)$$

事件A的概率取决于客体长度L、剖面间距a和客体走向与剖面线之间的夹角φ（图1）。事件B的概率则取决于客体宽度m、沿剖面的测点间距σ和夹角φ。

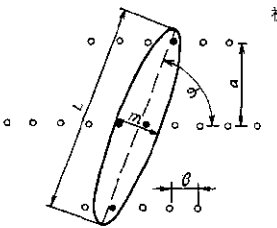


图 1　普查网示意图

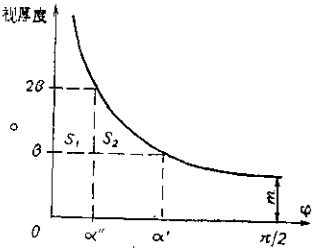


图 2　视厚度变化曲线图

我们用一些符号来表示这些概率

$$P(A)=P_L,\quad P(B/A)=P_m,\quad P(C)=P_{n2}$$

如果不考虑异常的方向，就可按下列以解比尤方(Бюффон)题为基础的公式（弗兰茨基，巴扎诺夫，1969）计算概率P_L:

$$P_L=[L(1-\sqrt{1-a^2/L^2})+a(\pi-2y)]/a\pi \qquad (2)$$

在这里，$y=\arcsin\ a/L$

在异常方位处于$\varphi_1\leqslant\varphi\leqslant\varphi_2$范围内的条件下，公式（2）可变换成：

$$P_L=\int_{\phi'}^{\varphi^2} L\sin\varphi d\varphi/a(\varphi_2-\varphi_1)=L(\cos\varphi_1-\cos\varphi_2)/a(\varphi_2-\varphi_1) \qquad (3)$$

在这里，φ_1，φ_2——弧度角。

许多文章（萨文斯基，1964；舒雷金，1972）指出，在一个方向伸长的客体，走向偏转30°对穿切它的概率不会有重大改变。因此，编制夹角φ变化范围为30°的三种（即0—30°，30—60°，60—90°）列线图便足够了。

我们再来探讨概率P_m的确定问题。图2是客体的视宽度(在剖面线方向上的宽度）随夹角φ而变化的曲线。视厚度值可由公

式$m_{极}=m/\sin\varphi$来确定。

如果$2\sigma<m$，则在异常被剖面穿切（以任意角度φ）时，至少都有两个点落在异常上。在这种情况下，$P_{\geqslant 2}=1$。

在φ值中间（见图2）存在着两个可使下列命题成立的值α'和α''。这些命题是：

如果$2\sigma>m$，那么，当夹角$\varphi<\alpha''$时，$2\sigma<m_{极}$，$P_{\geqslant 2}=1$；当$\varphi>\alpha''$时，$P_{\geqslant 2}<1$；当$\varphi>\alpha'$时，$\sigma<m_{极}$，$P_{\geqslant 2}=0$，也就是说，异常内落有两个点是不可能的事件。

概率P_m可以根据通过上述夹角确定的矩形及与其毗邻的曲线三角形的面积S_1和S_2（见图2）来计算：$P_m=(S_1+S_2)/S_0$，这里，$S_0=\pi\sigma/2$。

概率P_m的解析表达式是：

$$P_m=[\sigma(\alpha''-\varphi_1)+m\ \mathrm{In}\ \mathrm{tg}\alpha'/\mathrm{tg}\alpha' -(\alpha'-\alpha'')\sigma]/\sigma(\varphi_2-\varphi_1) \qquad (4)$$

在这里，$\sin\alpha''=m/2\sigma$，$\sin\alpha'=m/\sigma$，φ_1和φ_2—夹角φ的变化范围。

根据上面的表达式，可将公式（1）写成：

$$P_{\geqslant 2}=\sum_{i=1}^{n}\left\{L\frac{\cos\varphi_{i-1}-2\cos\varphi_i-\cos\varphi_{i+1}}{a(\varphi_2-\varphi_1)}\rightarrow \rightarrow\frac{+[(i-1)\varphi_{i-1}+(i+1)\varphi_{i+1}-2i\varphi_i]}{}\right\}\left\{iP_m-C_{i,2}P_{m,2} +C_{i,3}P_{m,3}-\cdots+(-1)^n P_{m,n}\right\}, \qquad (5)$$

在这里，i——穿切异常的剖面线数目；φ_i——剖面线与异常间的夹角；$\sin\varphi_1=ia/L$；n——当夹角由φ_1变至φ_2时，剖面线穿切异常的最大可能次数。

在解公式（4）和（5）时可作如下替换：

当$\alpha''<\varphi_1$时，以φ_1代α''，

当$\alpha'>\varphi_2$时，以φ_2代α'，

当$\varphi_{i-1}<\varphi_1$时，以φ_1代φ_{i-1}，

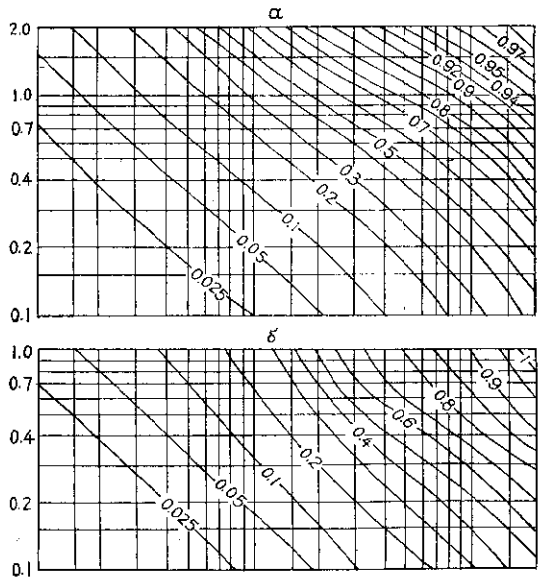
当$\varphi_{i+1}>\varphi_2$时，以φ_2代φ_{i+1}，

当$\varphi_i<\varphi_1$时，以φ_1代φ_i，

当$\varphi_i>\varphi_2$时，以φ_2代φ_1。

根据公式（5）编制出了夹角分别为0—90°（图3，*a*），0—30°，30—60°和60—90°（图3，*б-в*）的列线图。这些列线图编制成概率曲面形式，座标轴相应为按普查网的参数规格化的异常长度和宽度，即$L_a=L/a$，$m_s=m/s$.

为了检验所编制的列线图，做了下述试验。准备了几套有机玻璃的客体模型，往绘有给定的观测网度的图案板上，进行1000次任



意投放，模型的参数和实验结果列于表 1 。可以看出，计算的概率和试验的概率是很接近的。

列线图的编制和模拟是由大学生 A. B. 斯米尔诺夫和 B. A. 马雷舍夫完成的。

下面以实例说明用列线图进行工作的方法。

例1 在普查与硫化物-黄铁矿矿化带有关的钻矿床时，按200

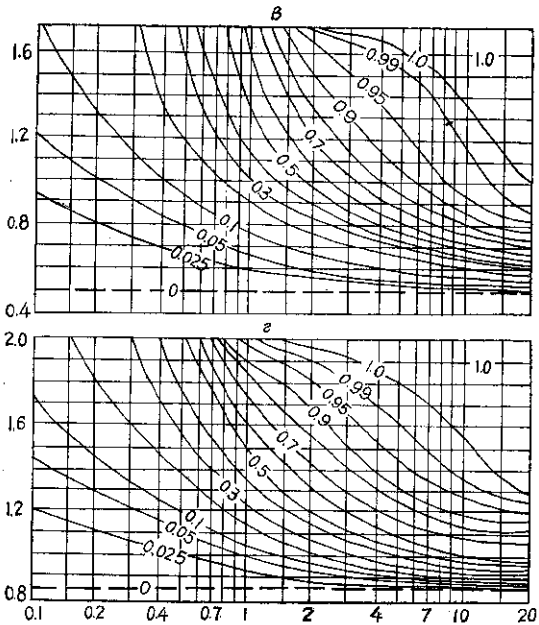


图 3 确定$P_{\geqslant 3}$的列线图

夹角φ分别为0—90°（а）、0—30°（б）、30—60°（в）和60—90°

×20米和100×10米的网度进行了次生分散晕岩石地球化学测量（瓦赫罗麦耶夫，福科夫，1972）。试估计当以钻❶的地球化学异常下限圈定的异常的平均大小为400×30米和交角φ=60—90°时，这种测量的可靠程度。

А. 网度200×20米，$L_a=L:a=400:200=2.0$，$m_a=m:a=30:20=1.5$。根据60—90°的列线图，求得其概率$P_{\geqslant 2}=0.75$。

Б. 网度100×10米，$L_a=400:100=4.0$，$m_a=30:10=3.0$。概率$P_{\geqslant 2}=1$。

因此，100×10米的网度可保证查明所有的钻晕，而200×20米的网度只能查明75%的钻晕。

表 1　根据模拟结果对0—90°列线图所作的检验

（网度5×10毫米）

普查岩体的规模		概　率　$P_{\geqslant 2}$	
以客体的宽度和长度表示，毫米 $m\times L$	以规格化后的客体的宽度和长度表示 $m_b\times L_a$	根据列线图查出的概率	由实验取得的概率
7.5×15	1.5×1.5	0.620	0.612
7.5×30	1.5×3.0	0.819	0.814
7.5×70	1.5×7.0	0.925	0.934
7.5×150	1.5×15.0	0.966	0.964
4.0×20	0.8×2.0	0.410	0.397
4.0×70	0.8×7.0	0.750	0.738
10×100	2.0×10.0	0.965	0.993

例 2　列线图不仅可在L，m和φ为固定值的情况下使用，亦可在知道这些参数的统计分布的情况下使用。我们以乌里克斯科-伊伊斯克地堑（近萨彦岭东部）的一个地段为例，来估计通过地面磁测、并配合重砂和岩石地球化学取样对该地段的金伯利岩进行普查的可靠程度。使用的测量网度为100×20米，剖面线的方位角等于55°。表2分别按φ、L和m列出了区内已知的磁场垂

❶ 此句中原文是“钙”，似应为“钻”——译注。

直分量ΔZ异常的频率分布。并已查明，L和m之间不存在相关关系，异常参数和异常方向之间也没有关系。

拟求概率$P_{\geqslant 2}$可以根据下列全概率公式来计算：

$$P_{\geqslant 2} = \sum_{k=1}^{q} P_k \cdot P_{\varphi k} \qquad (6)$$

在这里，$P_{\varphi k}$——交角为φ_k的异常出现的频率；P_k——含有交角φ_k的各个交角组与异常相交切的概率。

表 2　金伯利岩脉上的ΔZ异常按交角φ，长度L和密度m的分布情况

（近萨颜岭东部）

按交角		按长度			按宽度		
角度 φ_k	频率 P_φ	长度分类	平均长度 L_i	频率 P_{Li}	宽度分类	平均宽度 m_i	频率 P_{mi}
0—30	0.03	50—150	100	0.40	<25	20	0.20
30—60	0.26	150—250	200	0.37	25—35	30	0.55
60—90	0.71	250—350	300	0.15	35—45	40	0.16
		>350	400	0.08	>45	50	0.09

概率P_k由下述公式确定：

$$P_k = \sum_{i=1}^{n} P_{L,m} \cdot P_{Li} \cdot P_{mi} \qquad (7)$$

在这里，$P_{L,m}$——根据与给定的夹角φ_k相应的列线图确定的发现长度为L_i、宽度为m_i的异常的概率；P_{Li}——平均长度为L_i的异常的出现频率；P_{mi}——平均宽度为m_i的异常的出现频率。

根据公式（7），通过专门的表格（表3）来求解是较方便的。该表左侧列的是P_{mi}值；表头中所列的P_{Li}值取自表2。将P_{mi}的数据和P_{Li}的数据写成分数，每个分数的分子是根据列线图查出的概率$P_{L,m}$，分母则是$P_{L,m}$与左侧的P_{mi}和表头的P_{Li}的乘积。例如，对左上端第一格来说，在$L=100$（即$La=1.0$）和$m=10$（即$m_a=1.0$）时，计算出$P_{mi}=0.20$，$P_{Li}=0.40$。根据列线图查得概

率$P_{L,m}=0.05$。这一数据就相应于给定的P_{mi}和P_{Li}写在左上端第一格分数的分子上，而将$0.05\times0.20\times0.40=0.004$写在分母上。

在表的最右边，列出了分母的总和，写在相应那一行的位置上。该表底部一行写出了相应纵列分母的总和，右侧最下面一格则是所有分母的总和。在本例中，它等于$P_k=P_{k1}=0.663$。

利用相似的方法找出另外两交角组的概率，在$\varphi=30$—60°时，$P_{k2}=0.76$；在$\varphi=0$—30°时，$P_{k3}=0.85$。

根据公式（6）算出拟求的有限概率：

$$P_{\geqslant 3}=0.66\times0.71+0.76\times0.26+0.85\times0.03=0.70$$

从这里，不应当得出所研究的磁法测量不太可靠的结论，这是因为，普查工作一开始就是以至少根据一个点来发现金伯利岩脉上的磁异常为目标的，其工作的详细程度是通过将剖面上的测量间距缩小到10—15米，再每隔50米增加一条中间剖面来保证的（瓦赫罗麦耶夫，罗季奥诺夫，日特科夫，罗斯托娃，1970）。根据标准曲线计算出的概率$P_{\geqslant1}=0.97$（弗兰茨基，巴扎诺夫，1969），也就是说，磁侧是足够可靠的。

表 3　根据公式(7)针对交角$\varphi=60$—90°的异常组得出的解

P_{mi}	P_{Li}				$\sum P_{L,m}\cdot P_{Li}\cdot P_{mi}$
	0.40	0.37	0.15	0.08	
0.20	$\frac{0.05}{0.004}$	$\frac{0.20}{0.015}$	$\frac{0.30}{0.009}$	$\frac{0.35}{0.006}$	0.034
0.55	$\frac{0.53}{0.117}$	$\frac{0.75}{0.153}$	$\frac{0.90}{0.074}$	$\frac{0.94}{0.041}$	0.385
0.16	$\frac{0.90}{0.058}$	$\frac{1.00}{0.059}$	$\frac{1.00}{0.024}$	$\frac{1.00}{0.013}$	0.154
0.09	$\frac{1.00}{0.036}$	$\frac{1.00}{0.033}$	$\frac{1.00}{0.014}$	$\frac{1.00}{0.007}$	0.090
$\sum P_{L,m}\cdot P_{Li}\cdot P_{mi}$	0.215	0.260	0.121	0.067	0.663

我们用普查工作的实际结果对求得的数据$P_{\geqslant3}=0.70$作检验。所查明的27条金伯利岩脉中，有20条脉是由一个以上的异常

点发现的。

例 3 我们预计某一地区经统计查明的地球化学异常密度为20米，并假设异常长度为100米，这时，我们计算：当客体与测线基本上垂直（60—90°）分布时，若要以不小于 0.95 的概率发现拟寻客体，测量网度应多大。显然，位于列线图右上侧的0.95等值线区就符合这种条件，因而可以采用与座标为 $L_o=1.0$，$m_o=2.0$的点相对应的网度100×10米。

计算出的列线图与诸如被广泛采用的И. Д. 萨文斯基等的一些表格相比，具有许多优点。第一，使用起来要比表格简便而快速；第二，列入计算的参数较少，只有两个（L_o和m_o），而不是表格所需的三个；第三，列线图能够在较宽的L_o范围内和各种观测网方位角的情况下估计概率。但是，列线图也存在一些缺点，尤其是它们只适用于沿走向延伸的客体。通过列线图和И. Д. 萨文斯基的表格的对比表明，在普查客体与测网的方向任意交切时，用这两种方法估计出的概率$P_{\geqslant}$，的相对差值 δ 是：当客体的收缩系数$\theta'=m/L\leqslant 0.25$时，$\delta$ 不超过5%；当 $\theta'=0.4$时，δ 增至10—12%。但如果剖面垂直于拟寻异常的长轴（$\varphi=60—90$），那么，这两种方法的估计值是相近的（$\delta\leqslant 1—2\%$），而与异常的延伸程度无关。

参 考 文 献

Вахромеев Г. С., Родионов Н. И., Житков А. Н., Ростова Л. С. Опыт применения магниторазведки при поисках кимберлитов в Присаянье. —В кн.: Геофизические исследования при решении геологических задач в Восточной Сибири, вып 4. М., "Недра", 1970, с. 155—163.

Вахромеев Г. С., фоков Н. М. Применение геохимических методов при поисках кобальт-пиритового оруденения. —В кн.: Геохимические поиски рудных

месторождений. М., "Недра", 1972, с. 83—87.

Савинский И. Д. Таблицы вероятностей подсечения эллиптических объектов прямоугольной сетью наблюдений. М., "Недра", 1964. 86с.

Саврасов Д. И. Оценка вероятностей обнаружения аномалий геофизическими съемками. —"Геол. и геофиз. ", 1965, № 4, с. 134—137.

Семенова О. Г., Евдокимов Ю. Д. Оценка эффективности поисковой сети наблюдений с помощью ЭЦВМ. —"Разведка и охрана недр", 1970, № 10, с. 21—24.

Соловов А. П. Основы теории и практики металлометрических съемок. Алма-Ата, Изд-во АН КазССР, 1959. 266с.

Францкий И. В., Вазанов Г. А. Вероятностный метод определения плотности поисковой сети. —В кн.: Маркшейдерское дело в социалистических странах, вып. 4. Л., 1969, с. 261—269. (ВНИМИ. LXXV).

Шурыгин А. М. Расчет оптимальных сетей для поисков эллиптических залежей. М.,"Недра", 1972. 95 с.

Kyrala A. Basic probabilistic methods in Geological Search. —"Geophysics", 1964. 29, № 1 p. 105—108.

王家枢 译

岩石和矿石物质快速分析的综合方法

B. E. 卡普兰

浸染状钛磁铁矿工业矿体的界线是根据主要有益组分的含量确定的。矿石（可分为铁钒矿石和铜铁钒矿石）和矿化程度不够工业要求的围岩（辉长岩和辉岩）用肉眼区分不开。因此在普查、勘探和开采这类矿床时，有用组分含量的快速估测方法具有重要意义。

从1961年起就开始研究用物理方法快速分析铁钒矿床岩石和矿石的可能性（米克舍维奇，特鲁鲍夫，1963；米克舍维奇，特鲁鲍夫，萨莫依洛夫，维谢洛夫，1966；波诺马廖夫，格卢希赫，1963）。通过这些工作，确定了铁含量同岩石及矿石的磁化率、某些极化特征及其他物理参数的配对相关关系，研制出了测定仪器和方法，但它们在生产条件下的应用遇到了一系列困难，由于快速估测的精度不够而未被地质勘探和采矿企业所采用。

本文拟讨论在铁钒矿床以及以前没有研究过的铜铁钒矿床中，为了达到必需的快速估测精度和扩大所分析组分的数目，在对结果作适当的数学处理的情况下综合应用物理测量的经验。

快速估测的随机误差及减小方法

快速测定的随机误差是由物理测量的误差和由所使用的物理参数同所研究的含量之间的随机特征引起的随机偏差造成的。正如一些著作（杜鲍夫，1969；杜鲍夫，拉比诺维奇，顿科皮，富尔索夫，1970，等等）所指出的那样，通过改进仪器和测量方

法，特别是进行独立的重复测量，可以减小物理测量的误差。然而即便物理参数是绝对精确地测定出来的，成矿物质的含量与某种物理参数之间的配对对应关系也是不稳定的，因为所研究的这种关系只是一种局部的对应关系，并未全面考虑岩石和矿石的物质成分与其物理性质之间多种多样的复杂关系。

例如矿石磁化率值与矿石铁含量的对应关系，取决于钛磁铁矿中铁的数量、钛磁铁矿的总含量，以及在一定结晶条件和析出形式下铁所特有的性质。铜的二次X射线强度，通常不仅取决于所测元素的含量，而且也与原子序数跟它接近的元素的含量有关。钛磁铁矿的成分和磁性的不稳定性，与所测元素原子序数相近的元素的含量变化，分别造成了铁含量与磁化率之间和铜含量与二次X射线强度之间对应关系的随机特征。如若事先将被分析样品分成若干个干扰因素数值相近的抽样，我们就可以对每一抽样建立较为稳定的对应关系。通过对测区划分地段和对每个地段建立训练（标准）抽样的办法，上述原则是能够在快速分析的实践中在某种程度上实现的。然而，划分区的详细程度是有限的，这就限制了这种方法解决问题的可能性。

为了解决这些问题，更有发展前途的办法是，利用某种哪怕只能近似地估计干扰因素影响程度的辅助的物理方法。例如，在上述用X射线分析法测定铜含量时，原子序数相近的铁族元素的影响就能够根据磁性参数值估计出来。当矿石的磁化率介于100至$30000\cdot10^{-6}$厘米克秒制单位时，二次X射线的强度与铜含量间的相关系数大约等于0.6。若把这些样品划分为磁性变化区间更小的抽样，相关系数便增至0.90—0.98。对于每一个这样的抽样，就能够确定出铜含量与二次X射线强度间的比较稳定的关系，或者根据磁性参数值计算出相应的修正值。

下面讨论干扰因素影响的计算方法，其基础是对所研究组分的含量与两个变量——专门选定的主要和辅助物理参量的测定结果——的相关性进行最佳拟合。

岩石和矿石的物理性质与其物质成分经验相关性的最佳拟合

成矿组分的含量（C）与矿石物理性质（如x和σ）间相依关系的形式和参数事先是不知道的，因此，要通过对训练抽样作相应的数学处理来确定函数关系及其参数。近似计算法的一般算法，人们是了解的（杜鲍夫，1969）。

拟求的相关性

$$C = f(x, \sigma) \qquad (1)$$

可用一个多重幂级数以任何预定的精度来近似：

$$C = f(x, \sigma) \approx a_{00} + a_{10}x + a_{01}\sigma + a_{20}x^2 + a_{11}x\sigma + a_{02}\sigma^2 + \cdots \approx \widetilde{C}_{lm} \qquad (2)$$

于是，问题就变成了根据训练抽样的测量结果来确定多项式$\widetilde{C}_{lm}$的必要的阶和它的系数。

多项式的系数可根据所谓剩余平方和的极小化条件用最小二乘法确定：

$$\widetilde{R} = \sum_{i=1}^{n} [\widetilde{C}_i - (\widetilde{C}_{lm})_i]^2 = \text{极小值} \qquad (3)$$

式中，$\widetilde{C}_i$——训练抽样的第i个样品中组分的已知含量；$(\widetilde{C}_{lm})_i$——由这个样品的x和σ值求得的多项式$\widetilde{C}_{lm}$的值；n——是训练抽样中的样品数。

多项式的最佳阶可根据P. И. 杜鲍夫的第二级近似法求得。这种算法要求选择的多项式（2）具有这样的阶，即采用它可以使（3）式中$\widetilde{C}_i$与$(\widetilde{C}_{lm})_i$的均方差跟测量值$\widetilde{C}_i$的均方差δ最接近。也就是说，应当有$\frac{1}{n}\widetilde{R} \approx \delta^2$，或

$$\widetilde{R} \approx n\delta^2 \qquad (4)$$

由于n值的有限性，用这种办法算出的$\widetilde{R}$的估计值将偏倚于自己的数学期望。利用众所周知的求剩余平方和的数学期望的无

偏估计值的规则，可得出较精确的估计值以代替式（4）的规定：

$$\widetilde{R}=(n-S)\delta^2 \qquad (5)$$

式中S为（2）式中根据训练抽样求得的参数a_{im}的数目。

δ常常是未知数。在这种情况下可利用下述规律。

当多项式（2）的阶次增高直到它不再描述$\widetilde{C}_i$变化的基本规律之前，$\widetilde{R}$值减小得较快。阶次进一步增高，$\widetilde{R}$就减小得比较缓慢。因此，根据费歇离差准则估计出$\widetilde{R}$值减小的统计显著性，便可求出多项式（2）的近似最佳阶次。

利用作者编的ALGOL程序，可在电子计算机上实现这种演算。通过下面讨论的实例，即可了解计算结果。

铁族元素的快速分析

在所研究的矿石中，铁族元素是铁、钛、钒。主要矿物——这些元素的富集者——是钛磁铁矿。必须强调的是，矿石的质量取决于钛磁铁矿中铁族元素的含量，因为只有钛磁铁矿才进入磁选精矿。所以把磁选精矿的储量、它的回收率和成分作为评价矿床的主要指标，就是很自然的了。这些参数用物理方法（样品的湿磁选，用卡帕仪对精矿作快速分析）是容易确定的。

然而，由于把国家储委的规定机械地从富矿床搬用于浸染状钛磁铁矿矿床，因此只好按铁族元素的总含量来作评价。矿床开采的实践表明（萨莫依洛夫、马克西莫夫、乌斯科夫，1969），元素总量的估计值不能用来制定按预定质量生产最终产品（烧结矿和球团矿）的计划。因此，除了元素总含量，看来还必须考虑矿石的工艺类型，工艺类型可通过大量复杂的岩石学测定来查明（萨莫依洛夫、马克西莫夫、乌斯科夫，1969）。

在求铁族元素总含量的估计值时，正如前面已经谈到的，磁法是不够精确的。图1，Ⅰ—Ⅲ，*a*，*б*表示出利用磁法解这一课题的可能性。

矿石的磁性是用粉末样品测定的，这在很大程度上消除了钛磁铁矿的晶出形式、颗粒方位及其他因素的影响（米克含维奇、特鲁鲍夫，1963；波诺马廖夫、格卢希赫，1963）。同时，粉末样品的测量还需要考虑粉末的密度，同一样品的密度会变化于1.7克/厘米^{3}到2.4克/厘米^{3}之间。将载样盘的容积固定，把不同密实程度的样品装满载样盘，就不难证实，所测的磁化率值κ与粉末的密度有关，或者说与器皿中粉末的重量p有关：

$$\kappa = \varkappa p \qquad (6)$$

比例系数$\varkappa$具有比磁化率的意义，等于用固定体积（即载样盘容积）的样品测得的磁化率。在一般情况下，关系式（6）可写成如下形式：

$$\kappa = \varkappa(p - p_0) \qquad (7)$$

式中，p_0是由于器皿中样品的几何形态的关系未能进入卡帕仪送样器有效体积之内的那部分样品的重量。对于苏联产ИМВ型感应卡帕仪，根据感应磁通的几何形态，可以凭经验选择载样盘的最佳形状，使全部样品都能在送样器的有效体积之内。

由关系式（6）确定的$\varkappa$值，又直接与样品中的钛磁铁矿含量有关。如果认为全铁含量C_{Fe}包括钛磁铁矿铁和一些硅酸盐铁（假定它是固定的）C_0，则可写出等式：

$$C_{Fe} = k\varkappa + C_0 \qquad (8)$$

系数k和C_0可由训练抽样的测量值根据经验确定。

在快速分析实践中，有时建议利用铁含量与直接测出的κ值之间的经验关系。由（6）和（8）式可以看出，这种相关性可写成下列方程式：

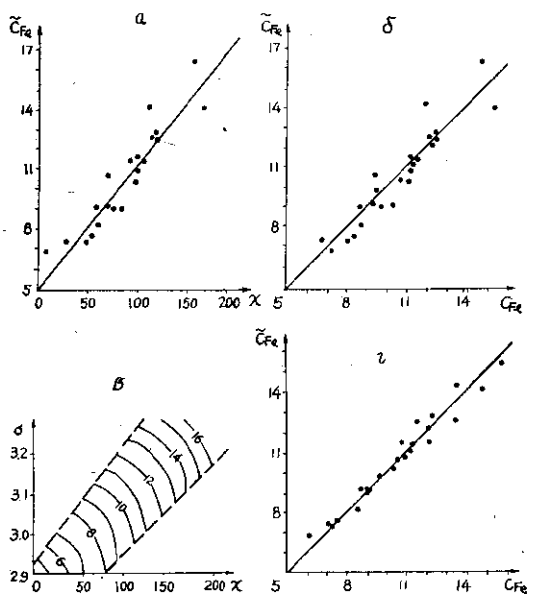
$$C_{Fe} = \left(\frac{k}{p}\right)\kappa + C_0 \qquad (9)$$

如果操作人员有一定熟练程度，要做到使某一批粉末样品的密实程度大体相同，从而$\frac{k}{p}$近似地等于常数，是完全可能的。但某些操作人员未必能保证经常遵守这项规定。

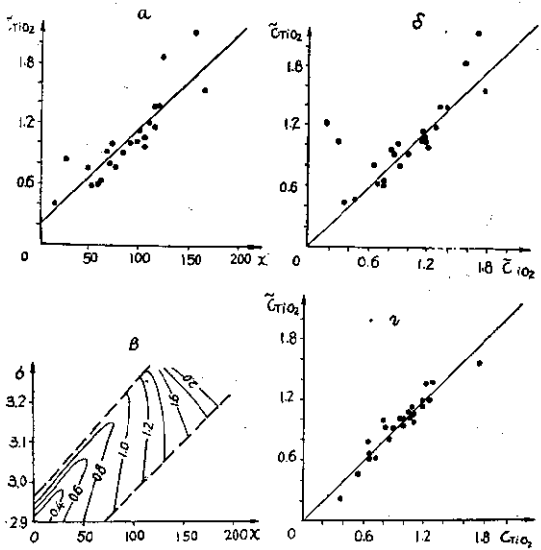
钛和钒的含量也存在类似于公式（8）的关系。各元素的全含量与比磁化率的相关系数，铁是0.93，钛是0.85，钒是0.80（图1，Ⅰ—Ⅲ，*а*）。这些全量浓度快速估计值的平均相对误差，分别是6，4.15和15%（图1，Ⅰ—Ⅲ，*б*）。在所有情况下，x值都是用下列方法求得的，即对同一个样品在几种压实条件下分别进行测量，再求其加权平均值：

$$x=\left(\sum_{i=1}^{m}\kappa_i p_i\right)/\sum_{i=1}^{m}p_i^2 \qquad (10)$$

Ⅰ



II


а
б
в
г
$\tilde{C}_{TiO_2}$
1.8
1.2
0.6
0
50
100
150
200 X
0.6
1.2
1.8
3.2
3.1
3.0
2.9
0.4
0.6
0.8
1.0
1.2
1.6
2.0

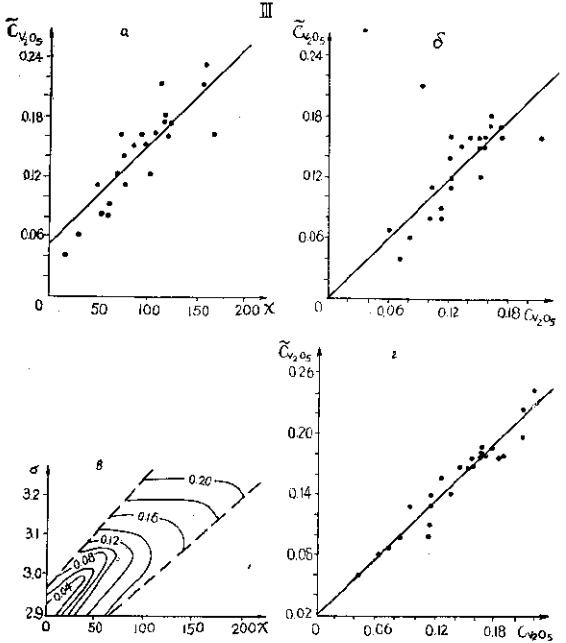


图 1　铁(Ⅰ)、TiO_2(Ⅱ)和V_2O_5(Ⅲ)含量快速分析结果

а—成矿组分含量$\widetilde{C}$（含量%）与比磁化率$\chi\cdot10^{-6}$（厘米克秒制单位）的相关性，根据已知成矿组分含量的样品的χ值求得；*б*—$\widetilde{C}$值与成矿组分含量估计值C（重量%）的相关性，C值是在$\widetilde{C}$的真值未知的情况下，利用*а*图的相关性根据χ值求出来的；*в*—用数学近似法求出的$\widetilde{C}$、密度σ（克/厘米^3）和χ值之间的相关性（等值线表示$\widetilde{C}$值）；*г*—$\widetilde{C}$值与成矿组分含量估计值C的相关性，C是利用*а*图的相关性根据χ和σ（也就是根据所提出的快速分析的数据）求出的。网点表示测量结果

正如在关于光谱分析的文章（杜鲍夫、拉比诺维奇、顿科皮、富尔索夫，1970）中所推荐的，必需的测量次数m取决于对x值作估计（10）时的精度要求。已得到的结果看来证明了以卡帕仪为基础的快速分析解决问题的可能性。

进一步提高快速估计的精度，靠增加x的测量精度是办不到的，需要采取专门措施稳定所采用的关系式（8）中的系数。为此，利用了辅助的物理参数——矿石的密度，因为密度既与矿化

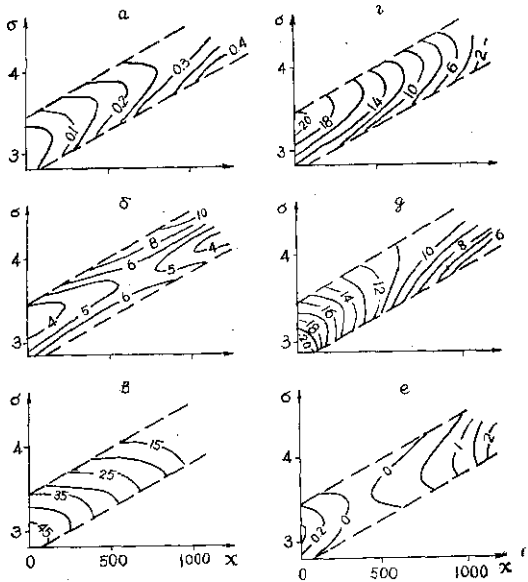


图 2 用数学近似法求出的化学组分 MnO(*а*)、Al_2O_3(*б*)、SiO_2(*в*)、CaO(*г*)、MgO(*д*)、CrO(*е*) 含量与岩石和矿石的密度和比磁化率之间的相关性。等值线是上述成分的含量（重量%）

强度有关，也与围岩成分，特别是与基性程度和暗色程度有关。例如，在铜铁钒矿床中，在矿化强度相同的情况下，从拉长辉长岩到培长辉长岩和钙长辉长岩，密度大约增加0.10—0.20克/厘米^3。同一种辉长岩，其特点首先是造岩矿物的成分相同和所含的铁的数量一致，岩石中所含的钛磁铁矿的磁性和铁、钛、钒含量相近。由此可见，在矿化强度相同的情况下，岩石密度上的差异与其成分上的差异是对应的，因而就能使等式（8）的系数稳定。

实际上整个这个过程可利用上述的建立拟求的元素含量与两个物理参数的最佳经验关系式的方法进行。

在图1，Ⅰ—Ⅲ，σ和图2中，在两个坐标轴上标的是物理参数的值，在坐标面中画的是拟求成矿组分和造岩组分的等含量线，它们是相应训练样组的最佳经验函数的曲线。利用这些图，根据比磁化率和密度值，用图解法不难求出拟求的含量。

图1，Ⅰ—Ⅲ中的α图说明了这种方法解决问题的能力。从图中所见的偏差，按平均值计，铁是3%，TiO_2是10%，V_2O_5是7%。相关系数分别提高到0.95、0.93和0.93。因此，成矿组分含量的快速估计值没有超出储委对化学分析规定的允许偏差范围。估计非矿组分含量的相对误差为3—50%。

铜含量的快速分析

在铜铁钒矿床中，含铜硫化物矿化与钛磁铁矿矿化伴生。在分析铜的样品中，铁的含量变化范围是5到25%或更多。铁族元素的原子序数与铜的原子序数相近。所以尽管采取一些专门措施调整X射线摄谱仪，总是不能完全避免铁、钛和钒原子的干扰。图3的α图表示出用X射线光谱分析法快速分析铜的可能性，测定铜浓度的平均相对误差是27%。为了近似地计算铁族元素的干扰影响，正如前面指出的，还利用了一个物理参数——比磁化率。在计算过程中，利用的是训练抽样的铜含量与两个物理参数

间的最佳经验关系式。在图 3 的 *б* 图中，两个坐标轴分别是二次X射线强度和比磁化率的值，坐标面中画出了铜的等含量线。求快速估计值，只要把这两个物理测量的数据投到坐标轴上，然后在坐标面找出这两个坐标所确定的点，在距该点最近的两条等含量线之间内插，即可求出相应样品的铜含量值。图 3 的 *в* 图示出快速估计的结果，并同相应的化学分析结果作了比较。平均相对误差是 6 %左右。

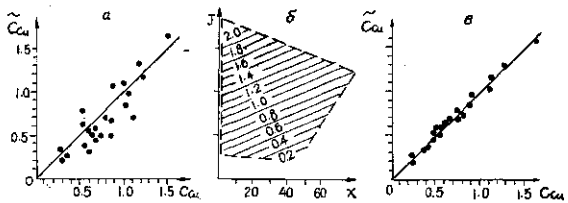


图 3　铜含量的快速分析结果

а—铜含量的化学分析数据 $\tilde{C}_{Cu}$（重量%）与X射线分析数据 C_{Cu}（重量%）的相关性；*б*—利用数学近似法求出的铜含量、二次X射线强度 J（假定单位）和X之间的相关性（等值线上的数值表示铜的重量百分含量）；*в*—根据 *б* 图的相关性用所提出的快速分析法求出的铜含量估计值 C_{Cu} 与化学分析数据 $\tilde{C}_{Cu}$ 的相关性

对于采自铁含量相近的地段的某些样品来说，只用X射线这一种物理方法，即可得到满意的分析结果，但这只有在样品铁含量稳定并接近标准样品时才有可能。

结　论

如对训练（标准）抽样的数据进行相应的数学处理，综合运用物理方法可以大大提高用快速分析法估测浸染状钛磁铁矿矿床成矿组分的能力。

参 考 文 献

Дубов Р. И. Физико-математический анализ гипергенного поля рассеяния. —В кн.: Литогеохимические поиски рудных месторождений по их гипергенным ореолам и потокам рассеяния. Алма-Ата, 1968, с. 50—52.

Дубов Р. И. Количественные исследования геохимических полей для поисков рудных Месторождений. Дис. на соиск. учен. степени д-ра геол.-мин. наук. Т. 1. Иркутск, 1969. 253 с.

Дубов Р. И., Рабинович И. М., Тонкопий М. С., Фурсов В. З. Оценка погрешностей измерений при анализе геохимических проб. —"Разведка и охрана недр", 1970, №7, с. 21—25.

Каплан В.Е., Гутин Н.Д., Егорченков Н.Е., устинов В. А. Экспресс-анализ рудных компонентов с помощью геофизических методов на месторождении медно-железо-ванадиевых руд. —В кн.: Геология месторождений полезных ископаемых урала и новые методы их изучения. М., 1974, с. 174—176.

Микшевич В. Н., Трудов Э. А. Опыт применения магнитометрического способа опробования качканарских железных руд.—В кн.: Вопросы разведочной геофизики. Свердловск, 1963, с. 113—119. (Труды Свердл. горного ин-та им. В.В.Вахрушева, вып. 44).

Микшевич В. Н., Трудов Э. А., Самойлов П. И., Веселов Ю. Н. Определение объемных весов руд и концентрата по методу ослабления гаммалучей.

—"Изв. вузов. Горный журн.", 1969 № 9, с. 3—7.

Пономарев В. Н., Глухих Н. И. К вопросу определения содержания железа в магнетитовых рудах по величине их магнитной восприимчивости. — "Изв. АН СССР. Сер. геофиз.", 1963, № 8, 1225—1229.

Самойлов П. И., Максимов Г. С., Усков Е. Д. Изучение вещественного состава и тектонических свойств титаномагнетитовых руд Качканарского месторождения для оперативного планирования добычи руды и производства железо-ванадиевого концентрата. —В кн.: Изучение вещественного состава и технол огических свойств железных руд. М., "Недра", 1969, с.148—159. (Труды ин-та Механобрчермет).

纪忠元译

地球化学资料统计处理时分析误差的估计和测量误差的计算

H. H. 茹科夫　M. И. 托尔斯泰

地球化学方法解决地质问题的效果主要取决于测量结果的可靠性、物质分析的准确度和重现性。

测量误差可以达到被研究量的自然变差那样大，有时甚至超过它，这时，测量结果可能使人对所研究地质体和地质作用的地球化学特征产生严重歪曲的概念。但是，懂得并计算实际误差—测量误差的统计表述—就能够对被歪曲的概念作重要的订正。

早在十多年以前，在一系列文章中就指出了这种计算的必要性和可能性，并提供了一些必要的基本公式（后来这些文章被收入了《金属矿床普查中地球化学场的定量研究》一书（杜鲍夫，1974），但那些对测量误差未加校正的地球化学研究成果仍继续在发表。本文以乌克兰地盾的花岗岩类为例来探讨这一问题。

众所周知，地球化学场最重要的统计特征之一是它们的方差。

我们提醒一下，在任一岩体中，被研究元素的浓度测量结果的方差σ_η^2与这些浓度的自然方差σ_ξ^2和由测量误差造成的分析方差$\sigma_{\Delta\eta}^2$有关：

$$\sigma_\eta^2=\sigma_\xi^2+\sigma_{\Delta\eta}^2+2K_{\xi,\Delta\eta} \qquad (1)$$

式中：$K_{\xi,\Delta\eta}$——相关矩，它反映着测量误差$\Delta\eta$和被测量量（元素浓度ξ）之间的关系。

σ_η，$\sigma_{\Delta\eta}$，$K_{\xi,\Delta\eta}$的值可通过将测量公式代入已知公式而计算出来，因此，根据公式（1），就可求出我们所要研究的量：

$$\sigma_\xi^2=\sigma_\eta^2-\sigma_{\Delta\eta}^2-2K_{\xi,\Delta\eta} \qquad (2)$$

地球化学场的另一个最重要的特征是不同元素（ξ_1,ξ_2）之间的相关关系 r：

$$r=\frac{K_{\xi_1,\xi_2}}{\sigma_{\xi_1}\cdot\sigma_{\xi_2}} \qquad (3)$$

式中：K_{ξ_1,ξ_2}——元素浓度ξ_1和ξ_2之间联系的相关矩；σ_{ξ_1}和σ_{ξ_2}——分别是元素浓度ξ_1和ξ_2的均方差。

当测量误差的影响忽略不计时，可利用σ_η代替σ_ξ，用下式的估计值R代替r，即：

$$R=\frac{K_{\eta_1,\eta_2}}{\sigma_{\eta_1}\times\sigma_{\eta_2}}$$

式中：K_{η_1,η_2}——元素ξ_1和ξ_2测量结果联系之间的相关矩；σ_{η_1}和σ_{η_2}——相应为元素ξ_1和ξ_2的浓度测量结果的均方差。显然，被用来估计r的量R与量r在原则上是有差别的。

为了求得r的估计值，应当把σ_{ξ_1}和σ_{ξ_2}代入公式（3）中，而σ_{ξ_1}和σ_{ξ_2}则是分别针对元素ξ_1和ξ_2根据公式（2）估计出来的；如果这些结果和相关的量都是按对称律分布的，公式（3）所需要的K_{ξ_1,ξ_2}与根据测量结果直接计算出来的K_{η_1,η_2}就是相当的(哈尔德，1956；杜鲍夫，1974)。如果它们的分布不是对称的，就必须采用对称变换。例如，如果ξ_1，ξ_2为对数正态分布的元素的浓度，那么就根据下述公式计算（杜鲍夫，1974）：

$$K_{\xi_1,\xi_2}=\tilde{\xi}_1\tilde{\xi}_2\times\mathrm{antlg}[(\sigma^2_{\lg\xi_1}+\sigma^2_{\lg\xi_2})/2\lg e]\times$$
$$\times[\mathrm{antlg}(K_{\lg\eta_1,\lg\eta_2}/\lg e)-1] \qquad (4)$$

式中：$\tilde{\xi}_1$，$\tilde{\xi}_2$——相应元素浓度的几何平均值；ant lg——反对数符号；$\sigma^2_{\lg\xi_1}$，$\sigma^2_{\lg\xi_2}$——相应为元素浓度ξ_1和ξ_2的对数的方差，它们根据下式计算：

$$\sigma^2_{\lg\xi}=\sigma^2_{\lg\eta}-\sigma^2_{\Delta\lg\eta}-2K_{\lg\xi,\Delta\lg\eta} \qquad (5)$$

如果浓度ξ和误差$\Delta\eta$的分布是对称的（例如是正态分布的），那么，$K_{\xi,\Delta\eta}=0$，这就使公式（2）的使用很为简便；如果被测元素的浓度对数值$\lg\xi$和它们的误差$\Delta\lg\eta$也是正态分布，那么，$\tilde{\xi}_1=\tilde{\eta}_1$，$\tilde{\xi}_2=\eta_2$，$K_{\lg\xi\Delta\lg\eta}=0$，公式（4）和（5）的利用就很

简便了。

表 1　针对被研究元素求得的 σ_ξ, $\sigma_{\Delta\eta}$ 和 $\sigma_\xi^2/\sigma_\eta^2$ 的估计值

被估计值	Ti, 10^{-2} %	Cr, 10^{-4} %	Sn, 10^{-4} %	Pb, 10^{-4} %	Mn, 10^{-3}%	Ga, 10^{-3}%
σ_ξ	2.23	1.65	0.173	5.75	0.10	0.06
$\sigma_{\Delta\eta}$	2.65	4.60	0.70	4.50	6.61	0.33
$\sigma_\xi^2/\sigma_\eta^2$	0.58	0.89		0.38		
被估计值	Ni, 10^{-4} %	Nb, 10^{-3} %	Mo, 10^{-3} %	Zr, 10^{-2} %	Ag, 10^{-5} %	Y, 10^{-3}%
σ_ξ	0.127	0.187	3.26	1.105	0	0.31
$\sigma_{\Delta\eta}$	0.26	0.33	7.00	0.65	0.17	0.94
$\sigma_\xi^2/\sigma_\eta^2$	0.84	0.76	0.82	0.25		

表 1 中列出了针对我们研究过的乌克兰地盾上30个花岗岩类样品求得的 σ_ξ 和 $\sigma_{\Delta\eta}$ 的估计值。我们使用的是近似定量分析数据，它们是用目前在地球化学研究中广泛使用的方法测得的。不难看出，在上述样品量的情况下，可以足够可靠地反映出许多元素的 $\sigma_{\Delta\eta}$ 值要比 σ_ξ 值大得多。再加上 $K_{\xi,\Delta\eta}$ 的影响，便导致 σ_η 与研究者感兴趣的 σ_ξ 值有重大差别，这一点根据所列出的几个 $\frac{\sigma_\xi^2}{\sigma_\eta^2}$ 比值就可以看出来。

表 2 列出了对所研究的不同元素对求得的 R 和 r 的估计值，正如我们所见到的，这些数值也是各不相同的。

根据测量结果直接计算出的 σ_η 和 R 值与研究者感兴趣的相应的 σ_ξ 和 r 值有差别，并不意味着所采用的物质分析方法不适用。问题在于不管如何仔细测量，这种差别总会存在，因为完全排除测量误差是不可能的，所以应考虑到这种情况，以必要的准确度去求 σ_ξ 和 r 的估计值。为此，不仅要提高样品的单个物质的分析准确度，而且要增加样品数量和测量次数，测量的结果要按一定

表 2 相关系数R和r的估计值

相关系数	Ni—Cr	Nb—Ti	Mo—Pb	Mo—Nb	Zr—Ti	Zr—Nb	Ti—Pb	Pb—Zr
R	0.51	0.53	0.66	0.69	0.33	0.89	0.51	0.47
r	0.64	0.43	0.90	0.84	0.12	0.72	0.42	0.44

注：R是按空间法来估计的，r则通过计算测量误差的方差和相关性来估计的。

的系统记录，而平均值则要根据与误差分布律相应的公式来计算（杜鲍夫，1974）。除此之外，在所有情况下，最好对剩余误差作前已述及的统计校正。

参 考 文 献

Дубов Р. И. Количественные исследования геохимических полей для поисков рудных месторождений. Новосибирск, "Наука", 1974. 278 с.

Хальд А. Математическая статистика с техническими приложениями. М., ИЛ, 1956. 664 с.

王家枢译

客体分类、地球化学场图示及解释时效益率的计算和等级近似法

P. И. 杜鲍夫

引　言

地球化学普查时发现的许多异常不具有经济价值。异常分类和有实际意义异常的划分，一般是以工作经验为依据的，但并非所有地质人员都有这种经验。如果异常存在一些相互矛盾的标志，那么这种划分就特别困难。这时，唯有将这些标志进行定量对比，也就是借助数学方法，才能得出正确的结论。

但是，现在最为流行的数学方法是用于解决异常归类问题的，同时也对异常分类出错的可能性作估计，而并不考虑经济条件。自然，若能作出绝对正确的唯一的分类，就有可能得出在经济方面也是正确的解。但是，这种解实际上不会是绝对正确的，也不会是唯一的，因为同一个异常被归入不同类别中去的概率不等于零。在这种情况下，虽然异常分类出错的可能性减至最低程度，并不意味着经济上的风险亦降至最小。下面就来说明这个问题。

我们假定，一定类型的异常与工业和非工业客体有关的概率估计值分别等于0.1和0.9，如果我们把这些异常中的每个异常都称为无工业意义的异常，那么，在90%的情况下将会是正确的。但是，如果必需的勘探工作成本不高，并且在那十分之一的机会中发现了工业矿床，它所得到的效益又可补偿全部花费，那么，对这些异常不做进一步的勘探工作，在经济上讲是不正确的。

现在问题已十分明显，就是说在异常分类时，需要对所采取的解的效益率（Полезность）进行估算。在我们讨论的这种情况下，就是要计算勘探效益的数学期望的估计值。所谓勘探效益，是指在考虑到这些异常属于不同类别的概率的条件下，用一定方法来勘探它们所能获得的赢利。

本文论述的就是关于计算这种估计值的理论，以及与其有关的解所谓反演题的理论。所谓反演题，就是根据地球化学观测结果，对地质客体和作用于其上的地质过程作描述。

基本原理

1. 假定q为被观测异常或异常观测特征（化学元素浓度、异常位置的地质条件等）的向量；q也可视为某种标志空间Q中的一个点。用符号r^*表示与异常有关的客体，或表示该客体拟求特征（与异常有关的矿体的规模、矿体中的矿石储量等）的向量。用符号r表示客体的类别，在客体空间R中，它可通过某个与被研究客体等同的点反映出来。对被研究客体引入一个与客体空间中的点不同的专门符号，是因为这个客体的特征不能直接观测到，并且在解题时也是未知的，也就是说，不知道这个客体究竟可与那一个点r等同，这种等同性只能以某种概率来予以确定。

此外，还引入下述符号：A_J——用方法J（探槽、钻探等）勘探异常所得效益的数学期望；$a_J(r)$——在完成工作J并查明客体r^*属于类别r时，应当实际得到的效益。如果r是不具经济价值的客体类别，那么，$a_J(r)$就为零或负值，其绝对值等于用方法J工作的成本。

在最简单的情况下，客体r^*可列为有限集合$\{R_j\}$的类别之一，$j=\overline{1,N}$（例如，在$N=2$时，可能有类别R_1和R_2，它们分别为矿体和非矿客体），假定$P(R_j|q)$是在客体r^*与异常q有关的条件下，$r^*=r\in R_j$这种情况的条件概率。这时，根据定义。

$$A_J|q=\sum_{j=1}^{N} a_J(R_j)\cdot P(R_j|q) \qquad (1)$$

最好是利用类别 r 的连续空间 R（例如，在根据与异常有关的矿石的预测储量值进行异常分类时）。在这种情况下，根据定义，

$$A_J|q=\int_R a_J(r)\cdot p_\varphi(r|q)\cdot dR \qquad (2)$$

在这里，dR——包含点 r 在内的空间 R 的一个基元空间；$p_\varphi(r|q)$——在异常 q 的条件下，出现事件 $r^*=r$ 的条件概率密度。

$P(R_j|q)$ 和 $p_\varphi(r|q)$ 值可以根据理论概念和对 q 型异常的现有勘探经验估算出来，也就是说，可根据已知（训练）资料来估计。当根据公式（1）或（2）计算 $A_J|q$ 值时，就会找出对何种方法 J 来说这个值是最大的，也就是选出了进一步研究异常的最佳解。这个解也可能是不必再对异常进行勘探。

对这类估计值的计算是一个重大的课题。甚至较为简单的公式（1）也不会总能直接使用，这是由于训练资料不足以按应有的估计精度分别取得公式中每个 $P(R_j|q)$ 的值。问题在于，在地球化学中经常要遇到 q 值连续分布的情况，在这种情况下，不同客体 r^* 上观测到相同 q 值的概率等于零，而不管这些岩体属于什么样的类别 R_j。因此，常常没有可用来估计 $P(R_j,q)$ 值的严格满足要求的训练资料，这时，就无法使用具体的 q 值，而是使用 q 值的区间，这显然会使所得到的估计值更加失真。如果空间 R 是连续空间或者复合空间（在一个地区是连续的，而在另一个地区是离散的），那么情况就更为复杂。

因此，我们就产生了一个对公式（1）和（2）中的分布函数进行近似的问题。在这种情况下，显然，最好是利用在形式上一致的方法。为此，我们根据下述原则将公式（1）变成分式（2）的形式。

假定 V 是空间 Q 和 R 的乘积空间。我们先讨论 Q 和 R 都是一维空间时的一种直观的情况。这时，V 就为具有 q 轴（横座标）

和 r 轴（纵座标）的二维空间。

我们再假设，上述座标轴上的点 q 和 r 只能分别属于集合 $\{Q_i\},(i=\overline{1,M})$ 和 $\{R_j\},(j=\overline{1,N})$ 中的某一类。对于这些类中的每一类，我们分别都用与其相应的轴上的区间 ΔQ 和 ΔR 表示。这些区间的中心在点 Q_i 和 R_j 处，长度任意规定。这时，当 $q\in Q_i$，$r^*=r\in R_j$ 时，事件 $q\cap r^*$ 即可归属于类别 $Q_i\cap R_j$，后者在空间 V 内表示为中心在点（Q_i，R_j）上、边长为 ΔQ 和 ΔR 的矩形。这种事件将称为事件 $Q_i\cap R_j$ 或者点（Q_i，R_j）。

根据已知的概率论原理，

$$P(Q_i\cap R_j)=P(Q_i)\cdot P(R_j|Q_i) \qquad (3)$$

在这里，

$$P(Q_i)=\sum_{l=1}^{N}P(Q_i\cap R_l) \qquad (4)$$

因而，根据公式（3）得出，

$$P(R_j|Q_i)=\frac{P(Q_i\cap R_j)}{P(Q_i)}=\frac{P(Q_i\cap R_j)}{\sum\limits_{l=1}^{N}P(Q_i\cap R_l)} \qquad (5)$$

如果不是利用量 q、r^* 和 $q\cap r^*$ 沿点 Q_i、R_j 和（Q_i，R_j）的离散分布，而是利用这些值沿相应的具有相同概率的长度区间 ΔQ、ΔR 和平方面积 $\Delta Q\cdot\Delta R$ 的均匀连续分布，那么，根据公式（5）计算出的结果将不会变化。但在这种情况下，还要确定条件概率密度

$$\begin{gathered}p_\varphi(Q_i)=P(Q_i)/\Delta Q;\quad p_\varphi(R_j)=P(R_j)/\Delta R;\\ p_\varphi(Q_i\cap R_j)=P(Q_i\cap R_j)/(\Delta Q\cdot\Delta R)\end{gathered} \qquad (6)$$

在代表一个类的点区内，所有的点在这里均是等效的。

如果考虑到

$$p_\varphi(R_j|Q_i)=P(R_j|Q_i)/\Delta R \qquad (7)$$

并且利用公式（6），那么可从公式（5）中得出

$$p_\varphi(P_j|Q_i)\cdot\Delta R=\frac{p_\varphi(Q_i\cap R_j)\cdot\Delta Q\cdot\Delta R}{\sum_{l=1}^{N}p_\varphi(Q_i\cap R_l)\cdot\Delta Q\cdot\Delta R}$$

$$=\frac{p_\varphi(Q_i\cap R_j)}{\sum_{l=1}^{N}p_\varphi(Q_i\cap R_l)} \qquad (8)$$

或者，将公式（8）两边除以ΔR，得出

$$p_\varphi(R_j|Q_i)=\frac{p_\varphi(Q_i\cap R_j)}{\sum_{l=1}^{N}p_\varphi(Q_i\cap R_l)\cdot\Delta R} \qquad (9)$$

由于每个区间ΔR内所利用的分布是连续的，所以可写成下述公式：

$$\sum_{l=1}^{N}p_\varphi(Q_i\cap R_l)\cdot\Delta R=\int_{V_{Q_i}}p_\varphi(Q_i\cap r)\cdot dR \qquad (10)$$

在这里，V_{Q_i}——空间V的截线Q_i，也就是在空间V内具横座标Q_i的点的集合；r——这些点的纵座标（与R_l类相对应的滑动值，$l=\overline{1,N}$）。如果将公式(10)代入公式(9)，就得到：

$$p_\varphi(R_j|Q_i)=\frac{p_\varphi(Q_i\cap R_j)}{\int_{V_{Q_i}}p_\varphi(Q_i\cap r)\cdot dR} \qquad (11)$$

现在，就$q\in Q_i$的情况而言，根据公式（7），将量$P(R_j/Q_i)=p_\varphi(R_j/Q_i)\cdot\Delta R$代入公式(1)，并像导出公式(10)那样，以积分代替求和，这时，参照公式(11)，就可对$q\in Q$得出

$$A_J|q=\int_{V_Q}a_J(r)\cdot p_\varphi(r|q)\cdot dR$$

$$=\int_{V_Q}a_J(r)\cdot p_\varphi(q\cap r)\cdot dR:\int_{V_Q}p_\varphi(q\cap r)\cdot dR \qquad (12)$$

按结构来说，这些公式和公式（1）是相当的。同时，公式(12)的

前一部分类似于公式(2)(在这里，空间V_Q所反映着的正是公式(2)中的空间R)。因而，公式(12)既可在r和q的数值呈离散分布的情况下使用，亦可在它们为连续分布时使用。在这两个值或其中一个值为离散分布的情况下，相应的区间ΔR和(或者)ΔQ最好取成1；这时，就会得到实际概率和引入的条件概率密度之间最为简单的关系(如果两个量r和q都是离散分布，那么按此原理在数量上可取$P(Q_i \cap R_j)=P_\varphi(Q_i \cap R_i)$。

我们利用空间V的点(q，r)来表示事件$q \cap r$，并以便于进一步利用的形式将公式(12)写出来：

$$A_J|q=\int_{V_Q} a_J(r)\cdot p_\varphi(q,r)\cdot dR:\int_{V_Q} p_\varphi(q,r)\cdot dR \tag{13}$$

在这里，$a_J(r)$值是基于经济考虑而提出来的，为了估计出被研究客体r^*的A_J/q值，只要根据与该客体相应的训练资料近似出函数$P_\varphi(q,r)$就足够了，然后再把研究客体r^*时求得的数据q代入公式(13)。

过去，我们制定过一些原则，用于对由研究空间的点得出的函数进行近似(杜鲍夫，1974)。这些原则也可用于对函数$P_\varphi(q,r)$进行近似。作为对一般原则的补充，应当考虑到的只是这样一点，用来近似函数$P_\varphi(q,r)$的函数$P^*(q,r)$，应服从于与概率密度的特征有关的一些限制：这个函数在空间V内处处均不应为负数，并满足于下述公式：

$$\int_V p^*(q,r)\cdot dV=1 \tag{14}$$

十分明显，公式(13)和与其有关的上述所有公式，在特征r和q为多维的情况下也是正确的。这时，表示点(q，r)的符号和r是相应空间内的该点座标的组合(这些空间的乘积就是空间V)。下面，即本节的第三点和下一节，将探讨本文宗旨所必需的那些近似原则。

2．每个地球化学异常都具有这样的特征：在一定数量的样

品中（这个样品数量不是完全确定的，因为异常的界限不是截然的），一系列元素的浓度一般是不同的。因此，要同时直接利用训练资料中关于异常的全部信息，就需要建立函数 $p^*(q,r)$，而这种函数甚至对现代的电子计算机来说也是极为复杂的课题。

用现在广泛采用的方法对异常分类，是根据仅仅在一个样品（例如，从异常最大值中取的样品）中所观测到的各种化学元素的浓度比值来进行的。用这种简单作法，大部分有用信息显然都未被利用。

为了利用有关异常的更多有用信息，就必须对它进行压缩，也就是说，对每个异常来说，要取得为数相对少的特征，但这些特征本身却包含着容纳了有用信息的大量原始数字资料。这类主要特征之一是晕的金属量，而它与残坡积层之下矿体露头的金属量大体呈比例关系（索洛沃夫，1959）。这个特征可由晕中被研究元素异常含量的积分来确定。后来又发现，可以在残余晕与一些参量间建立起对应关系。这些参量是：与晕有关的客体的可能规模，该客体中被研究元素的平均浓度，以及表征成晕过程的分散参数（杜鲍夫，1964）。这种把信息压缩成一些描述被研究客体和过程的少而精的特征的工作，实质上就是对异常的一种定量解释。这个问题具有独特意义，我们将较详细地探讨一下。

十分明显，对异常形成过程的自然本质考虑得越全面，在其它条件相同时，解释的精度就越高。在相邻学科——地球物理学中，十分重视通过对拟寻客体的参数作定量评价来进行解释的工作，这种解释已变成了反演课题的一个主要类别。在地球化学中，这种解释理论较晚才开始发展起来，这在某种程度上与地球化学现象的复杂性有关。但现在已能提出一些最一般的原则，从而有可能克服这样的困难。

我们将建立一些原则，作为解决上述问题的理论基础。

“浓度分布”一词，通常应理解为统计分布，也就是浓度的概率函数或概率密度函数。在解决地球化学反演题时，被研究客体就是浓度场或者浓度的空间分布，也就是浓度关于座标的函

数，它决定着矿体规模、矿石储量等。在下文中，如果被研究分布的统计特点或空间特点由上下文或公式即可一目了然，有时就不再说明了。

假若地球化学观测不被干扰因素（误差、样品中的浓度波动等）所复杂化，那么，我们就可观测到样品浓度沿各个点（x，y，z）的空间分布，每一元素浓度的空间分布就可以写成下述形式：

$$C_\varphi(x,y,z)=F\{C_0(\xi,\eta,\zeta)\} \qquad (15)$$

在这里，$\{C_0(\xi,\eta,\zeta)\}$——被研究浓度的空间分布函数在各个点(ξ,η,ζ)上的数值的集合；F——反映被研究过程的作用的算子。例如，在疏松岩石的岩石地球化学测量中，函数$C_\varphi(x,y,z)$可以反映出被研究元素的浓度在残坡积物上层（取样层）中的分布。在这种情况下，函数$C_0(\xi,\eta,\zeta)$可以是拟求的元素在基岩中的分布。这时，算子F反映内生作用所造成的物质再分布。

假如已经知道表达式(15)右边的确切公式形式，并且也有了一定数目（n个）可以确切决定$C_\varphi(x,y,z)$的数值的参数，那么，在具有同样数目（n个）的数值$C_\varphi(x,y,z)$的条件下，我们就可把这些数值代入公式（15），得到一个由n个方程构成的方程组。针对未知参数解此方程组，我们就算是解出了所提出的反演题，也就是解释了所观测的浓度场。

但实际上，观测结果$C(x,y,z)$会被一些随机干扰（噪音）因素所复杂化。此外，有关公式(15)右边的具体形式的资料一般是不完全的，因此，就不能探讨公式(15)，而要探讨公式(16)：

$$C^*(x,y,z)=F^*\{C_0^*(\xi,\eta,\zeta)\} \qquad (16)$$

在这里，$C_0^*(\xi,\eta,\zeta)$和F^*——分别近似于函数$C_0(\xi,\eta,\zeta)$和算子F的函数和算子；$C^*(x,y,z)$——为了近似出函数$C_\varphi(x,y,z)$，根据观测值$C(x,y,z)$计算出来的函数。如果公式(16)中的F^*和C_0^*被选取出来，而且所得的函数C^*能以最好的方式满足近似的要求，那么，提出的反演题就算解决了。这样一来，解决所探讨形式的反演题，如同估计函数p一样，也是个近似值问题。因此，我们

就转而论证这种近似法所必不可少的一些原则。

3．在地球化学课题中，往往必须在常用的已知方法未加规定的条件下采用近似法。由于先验性的限制不足；会遇到颇大困难。在研究如何解决这些课题的方法时，首先应当考虑以下原则。

由于自然现象的复杂性，在被观测到的地球化学异常特征和与其对应的地质体及地质作用之间一般没有十分确定的关系。因此，在标志 q 已确定的情况下，对不同类别的 r 来说，值 $p_\varphi(r|q)$ 都不会是零。但是，与任何一些客体 r^* 相应的标志的 q 的差别越小，客体有区别的数学期望也就越小，不管这种差别取何种定义都是如此。正因为这样，在其它条件相同时（及有其它资料，或者其它观测标志的作用尚不清楚），首先投入勘探的是主要见于已知大型矿床上的那种类型的异常，而比较常见的、大多数是没有意义的客体上的那种异常，我们就忽略之。

这种方法常常会取得显著的成功。在个别情况下也可能发生错误，但这时见矿的平均机会要比勘探那些主要见于小型客体上的异常要大，也就是说，所见矿床不是大型矿床的数学期望比勘探主要见于小型客体上的异常要小。

如果条件复杂、资料不全，在选择解决问题的方法时，也不得不建立一种只是平均说来最佳的算法，在这个时候，得出的解与未知的精确解的差别的数学期望要低于我们现已掌握的其它算法。

如果利用函数分析的术语，就可用下述形式表达出相应的普遍原理。

假定：w_0、w_1、w_2——分别为某种被研究的客体（地质体，地质过程、函数，算法等，在部分情况下，它们可能是前一节所讨论的客体 r^*）；$u_0=u(w_0)$，$u_1=u(w_1)$、$u_2=u(w_2)$——相应为这些客体的标志（在部分情况下，它们可能是关于客体 r^* 的地球化学资料 q）；$\mathscr{L}_W$ 和 $\mathscr{L}_U$——相应为客体空间 W 中和标志空间 U 中的距离算子；E——数学期望算子。于是，在其它条件相同时：

如果，$\mathscr{L}_U(u_1,u_0)<\mathscr{L}_U(u_2,u_0)$，

那么，$E(\mathscr{L}_W(w_1,w_0))<E(\mathscr{L}_W(w_2,w_0))$ (17)

也就是说，在空间U中反映任何客体的标志的点的间距越小，在其它条件相同时，在空间W中反映这些客体的点的间距的数学期望也就越小，而这些空间中的距离可以用不同的方法来确定。

在本文作者的文章（杜鲍夫，1972）中，曾根据较严格的推论，论证过原理(17)。

现在，我们假定，w_0——未知函数，而w_1、w_2、…，w_n——现已为我们掌握的（已知的）函数，并且必须从其中选出一个对函数w_0进行近似。同样，我们假定，存在与这些函数对应的某个量（标志）u的值u_1、u_2、…，u_n，以及作为函数w_0的已知标志的值u_0。那么，根据原理(17)，在其它条件相同时，就应当从上述函数中选择出使得量$\mathscr{L}_U(u_i,u_0)$为最小的函数作为最佳近似函数。这就意味着，量$\mathscr{L}_U$在这种情况下就是近似准则。

这种利用各种标志的可能性，以后将用于选择函数P^*和C^*。

可以看到，标志u可以包含着关于客体w的任何信息，同时也包含着可肯定地确定客体w_0的那样一些信息。在后一种情况下，表达式(17)中的符号E可以简化掉，因为固定的量的数学期望等于这一量的本身。因而，在表达式(17)中就包含了这样一种方案，在此方案中，已有的（包含在标志u中的）信息可足以保证选择出每个具体情况下的最佳算法，而不仅仅是选出一种平均的最佳算法。

原理(17)，并综合关于选择近似函数和近似准则的结论，可以认为是便于以后利用的基本近似原则。

计 算 方 法

1．假定有事件$q_k \cap r_k$作为训练资料，在这里，$k=\overline{1,N}$——事件的编号。例如，这些事件可能是在研究地区内侥倖地选出的N个晶的勘探结果。我们提醒一下，空间V中的点(q_k,r_k)是与这

些事件相对应的。

我们先假定，已经选出函数形式 $p^*(q, r)$，并且该函数包含一定数目（L个）自由的（不固定的）应根据训练资料来估计的参数。在这里和下面均假定，$L \ll N$。

函数 $p^*(q, r)$ 中自由参数的值，可以根据已知的最大似然原则和上述近似概率密度的特点求出，也就是说，可考虑补充条件(14)，根据下面条件求出：

$$\prod_{k=1}^{N} p^*(q_k, r_k) = \text{最大值} \tag{18}$$

例如，我们假定，$p^*(q, r)$为下面形式的指数抛物线

$$p^*(q,r) = \exp(A_{00} + A_{10}q + A_{01}r + A_{11}qr + A_{20}q^2 + A_{02}r^2) \tag{19}$$

在这里，q和r——一维的量，A_{ij}——自由参数。

如果假定 $q=q_k$和$r=r_k$，$k=\overline{1,N}$，就可从公式(19)得到由N个$p^*(q_k, r_k)$的表达式组成的方程组。将它们代入公式(18)，就得出以下形式的相应条件：

$$\exp\Big(A_{00} + A_{10}\sum_{k=1}^{N} q_k + A_{01}\sum_{k=1}^{N} r_k + A_{11}\sum_{k=1}^{N} q_k r_k + \\ + A_{20}\sum_{k=1}^{N} q_k^2 + A_{02}\sum_{k=1}^{N} r_k^2\Big) = \text{最大值} \tag{20}$$

在这种情况下，公式(14)具有下面形式：

$$\int_a^b \int_c^d \exp\,(A_{00} + A_{10}q + A_{01}r + A_{11}qr + A_{20}q^2 + A_{02}r^2) \cdot dq \cdot dr = 1 \tag{21}$$

在这里，(a,b)和(c,d)——相应为q和r的可能取值区间。在q_k，r_k给定的情况下，满足条件(20)和(21)的系数A_{ij}的计算，是一个纯技术问题，它可根据已知的（变分法和数值积分）算法去解决。

当函数 $p^*(q,\ r)$ 为其它形式时，也可利用类似的方法解决这一问题。

当函数形式已知道，只要求估计其自由参数时，我们把寻求近似函数的算法称为一级近似算法。

我们常常无法确切地知道最佳函数 p^* 的形式，而必须根据事件 $q_k \cap r_k$ 从函数 $p_L^* = p_L^*(q,r)$ 的某个有序族 $\{p_L^*\}$ 中选出来。下面这样一些函数将包含在这样一个族内：其中每一函数的顺序编号 L 越大，它具有的自由参数的数目就越多；而且，对 p_L^* 中自由参数的任何值来说，可以在 p_{L+1}^* 中找到这样一些自由参取值，在这些自由参数值下，对任何自变量 q，r 来说，$p_L^*(q,r) = p_{L+1}^*(q,r)$。可以这样选择 p_{L+1}^* 的自由参数值：在 p_L^* 的（数目较小的）自由参数取任何可能的值时，在非零维的点集 $(q,\ r)$ 上 $p_L^*(q,r) \neq p_{L+1}^*(q,r)$。

例如，不同阶的指数抛物线族，或者是不同数目正态分布的组合族等，都可以是函数 p_L^* 的这种族。

要将每个函数 p_L^* 都作为可能的最佳函数来探讨，就必须对它采用一级近似算法。因此，下面将假设，函数 p_L^* 中的自由参数已经按照这种算法确定出来。接着利用上面所探讨过的基本近似原则来选取最佳函数。为此，就必须先找出在公式(17)中起量 u 的作用的特征。这一特征应当针对每个函数，根据所提出课题的条件，利用包含在训练资料中的信息，也就是包含在空间 V 中的点集 $\{(q_k, r_k)\}$ 位置上的信息计算出来。

任何点集 $\{(q,r)\}$ 的点与点 $(q_l,\ r_l)$ 的距离的数学期望，可以作为该点集相对于点 (q_l, r_l) 的位置的特征加以探讨。对观测到的分布 $p(q,r)$ 来说，根据定义，这就是

$$\rho_l = \int_V \mathscr{L}((q,r),\ (q_l,r_l)) \cdot p(q,r) \cdot dV \qquad (22)$$

在这里，$\mathscr{L}((q,r),(q_l,r_l))$——点 (q,r) 和点 (q_l,r_l) 间的距离，在这之中，如果 q 和 r 是一维的量，就可利用以下定义：

$$\mathscr{L}((q,r),(q_l,r_l)) = \sqrt{(q-q_l)^2 + (r-r_l)^2}。$$

如果分布$p(q,r)$是由点集$\{(q_k, r_k)\}$的点的位置确定的，那么，根据已知的原则，由公式（22）确定的量ρ_l，就可根据下式来估计。

$$\rho_l = 1/N \sum_{k=1}^{N} \mathscr{L}((q_k, r_k)\cdot(q_l, r_l)) \tag{23}$$

量ρ_l表征着由函数$p(q,r)$表述的对事件$q_k \cap r_k$观测到的分布（该分布叠加有随机干扰）。这个分布要用来建立对未知函数$p_\psi(q,r)$（它与$p(q,r)$不同，没有干扰分量）进行近似的函数$p_L^*(q,r)$。

与公式（22）相类比，我们可写出点(q_l, r_l)到其余各点距离的数学期望的表达式，这类表达式与分别由函数$p_L^*(q,r)$和$p_\psi(q,r)$表述的事件$q \cap r$的分布相对应：

$$\rho_{Ll}^* = \int_V \mathscr{L}((q,r),(q_l,r_l))\cdot p_L^*(q,r)\cdot dV \tag{24}$$

$$\rho_{\psi l} = \int_V \mathscr{L}((q,r),(q_l,r_l))\cdot p_\psi(q,r)\cdot dV \tag{25}$$

量$\Delta\rho_{\psi l} = \rho_{\psi l} - \rho_l$和$\Delta\rho_{Ll}^* = \rho_{Ll}^* - \rho_l$说明了以函数$p_\psi(q,r)$和$p_L^*(q,r)$描述的事件$q \cap r$的分布与反映事件$q_k \cap r_k$的总体的那种分布$p(q,r)$之间的差别。因而，$\Delta\rho_{\psi l}$和$\Delta\rho_{Ll}^*$可以用来得出近似准则。得出近似准则的可能方法之一如下。我们在空间V中给出一个由N个点(q_l, r_l)组成的点网，各点之间的距离与这一空间相应部分的值$p_L^*(q,r)$大致成反比❶。我们将点中相邻近的点编成点对$\{(q_l, r_l), (q_l, r_l)\}$。网中的同一个点可以参加不同的点对，构成一些点链。点对中点的顺序和进而对点对作探讨的顺序是任意定的。

---

❶ 引入这个建议，是为了以后大致均匀地计算分布在相当小的距离上的点(q_k, r_k)，在这些点上值$p_L^*(q,r)$却相当大。要指出的是，这个建议从原则上说（在计算精度相当高的情况下）并不重要，并且为了简化制图，可以在包含点集(q_k, r_k)的空间V中的最小凸形或平行六面体区域内，使用点(q_l, r_l)的均匀网。

现在，假设在公式（23）、（24）和（25）中，$l=i$和$l=j$，我们对每个点对中的每个点计算相应的$\Delta\rho^*_{Li}$和$\Delta\rho^*_{Lj}$值。设$n_c(p^*_L)$和$n_d(p^*_L)$分别为对一个点对中的两个点求出的$\Delta\rho^*_{Li}$和$\Delta\rho^*_{Lj}$符号相同及相异两种情况的数目。与此类同，与$\Delta\rho_{\varphi i}$和$\Delta\rho_{\varphi j}$的符号相对应的$n_c(p)$和$n_d(p)$的数目亦可确定出来。

可见，$n_c(p^*_L)+n_d(p^*_L)=n_c(p_\varphi)+n_d(p_\varphi)$——上述点对的总数，对给定的点集$\{(q_k, r_k)\}$，它是不变的。

我们给定符号：

$$\mathcal{Z}^*_L=\frac{n_c(p^*_L)-n_d(p^*_L)}{n_c(p^*_L)+n_d(p^*_L)},\quad \mathcal{Z}_\varphi=\frac{n_c(p_\varphi)-n_d(p_\varphi)}{n_c(p_\varphi)+n_d(p_\varphi)} \tag{26}$$

我们将证明，量$E(\mathcal{Z})$❶可以用作进行近似的标志u。为此需要证明，虽然函数p_φ是未知的，但值$E(\mathcal{Z}_\varphi)$显然是已知的，而值$E(\mathcal{Z}^*_L)$是随数目L变化的。

我们只探讨这样一种情况，即假定由于只出现非相关干扰，观测到的分布p与由函数p_φ所描述的分布有差别的情况，这时，对每个点对$\{(Q_i, R_i), (Q_j, R_j)\}$来说，值$\Delta\rho_{\varphi i}$和$\Delta\rho_{\varphi j}$的符号也是不相关的，因而，这些符号相同或不同的情况的比率具有同样的数学期望：

$$E\left\{\frac{n_c(p_\varphi)}{n_c(p_\varphi)+n_d(p_\varphi)}\right\}=E\left\{\frac{n_d(p_\varphi)}{n_c(p_\varphi)+n_d(p_\varphi)}\right\} \tag{27}$$

根据公式（26）中的第二个式子，这就是说

$$E(\mathcal{Z}_\varphi)=E\left\{\frac{n_c(p_\varphi)}{n_c(p_\varphi)+n_d(p_\varphi)}\right\}-E\left\{\frac{n_d(p_\varphi)}{n_c(p_\varphi)+n_d(p_\varphi)}\right\}$$
$$=0 \tag{28}$$

这样一来，就证明了$E(\mathcal{Z}_\varphi)$值是知道的。但仍然要证明，值$E(\mathcal{Z}^*_L)$是随数目L而变化的。

---

❶ 在这里，$E(\mathcal{Z})$——表示$E(\mathcal{Z}_\varphi)$和$E(\mathcal{Z}^*_L)$的统一符号。

每个函数$p_L^*(q,r)$的自由参数值可根据最大似然原则来找出。因而，如果增大数目L，那么，在所有的点(q,r)上，$p_L^*(q,r)$值就会尽量趋近于观测分布值$p(q,r)$。但根据公式(22)、(24)，函数$p_L^*(q,r)$和$p(q,r)$决定着p_{Li}和p_{Lj}^*值。因此，在增大数目L时，p_{Li}^*值就会接近ρ_l值（也就是说，对所有的指数l来说，$p_{Li}^* \to \rho_l$）。这就说明，出现p_{Li}^*和p_{Lj}^*两个值处于函数ρ_l一侧这一事件的几率降低，也就是说出现$\Delta\rho_{Li}^*$和$\Delta\rho_{Lj}^*$具相同符号这一事件的几率降低。反过来，根据n_c和n_d值的定义，这就意味着，这时$n_c(p_L^*)$减小，$n_d(p_L^*)$增加；而根据公式(26)，值$\mathscr{Z}_L^*$减小且可以变为负值。所确定的量的数学期望就等于这个值，也就是$E(\mathscr{Z}_L^*)$值也变小，并且可能成为负值。

同样地，假如函数p_L^*中的参数数目减少，那么它与函数p的的差别就增大，因而，根据公式(22)和(24)，函数ρ_{Li}^*与函数ρ_l在点(q_i,r_i)和(q_j,r_j)上的差别的数学期望也增大。这时，值ρ_{Li}^*和ρ_{Lj}处于函数ρ_l一侧这一事件的几率的数学期望也增加。因此，根据量$\Delta\rho_{Li}^*$和$\Delta\rho_{Lj}^*$，$n_c(p_L^*)$和$n_d(p_L^*)$的定义，$n_c(p_L^*)$的数学期望增大，$n_d(p_L^*)$的数学期望就减少，并且值$E(\mathscr{Z}_L^*)$增大。

这样一来，就证明了值$E(\mathscr{Z}_L^*)$是随数目L变化的，并且根据上述所谈情况，$E(\mathscr{Z})$可以作为标志u。

$E(\mathscr{Z})$是一维的，因此，对标志空间内值$E(\mathscr{Z}_L^*)$和$E(\mathscr{Z}_\varphi)$之间的距离来说，可以采用一个最简单的定义——这两个值之差的绝对值：

$$\mathscr{L}_u(E(\mathscr{Z}_L^*),\ E(\mathscr{Z}_\varphi)) = |E(\mathscr{Z}_L^*) - E(\mathscr{Z}_\varphi)| \qquad (29)$$

$\mathscr{Z}_L^*$已确定：它可根据所探讨函数p_L^*的训练资料来直接计算。因此，根据已确定量的数学期望的性质，$E(\mathscr{Z}_L^*) = \mathscr{Z}_L^*$。根据公式(28)，$E(\mathscr{Z}_\varphi)$值等于零。所以，用以确定近似准则的公式(29)便可采用这样的形式

$$\mathscr{L}_u(E(\mathscr{Z}_L^*), E(\mathscr{Z}_\varphi)) = |\mathscr{Z}_L^*| = \frac{|n_c(p_L^*) - n_d(p_L^*)|}{n_c(p_L^*) + n_d(p_L^*)} \qquad (30)$$

当$\mathscr{L}$

$$\mathscr{L}_L^*=\frac{n_c(p_L^*)-n_d(p_L^*)}{n_c(p_L^*)+n_d(p_L^*)} \tag{31}$$

最接近于零时，(30) 具有最小值。所以，$\mathscr{L}_L^*$也可以作为近似准则。这个值可以是正值，也可以是负值。该值与零值接近的程度，可通过大量的计算予以估计，这比估计量(30)接近最小值的程度要简单一些。由此可见，由给定的这类函数构成的最佳近似函数，就是那些在采用一级近似算法之后值$\mathscr{L}_L^*$最接近于零的函数。

应当指出，为了选定最佳近似函数，并不需要对所探讨类别的所有函数进行计算。假如对数目 L 作某种改变 $\mathscr{L}_L$ 便发生越过零的变化，那么，要选择最佳近似函数，只要在具有相近的 L 值且都发生这种变化的那些函数中进行就足够了。

上述从给定族中找出最佳近似函数的算法就称为二级近似算法。

可以利用平均接近度来代替点 (q_l, r_l) 与点 (q_k, r_k) 的平均距离，平均接近度可定义为距离倒数的平均值。这就是说，为了建立上述标志，可以不用公式 (23)、(24) 和 (25)，而相应地代之以下各式：

$$\rho_l=\frac{1}{N}\sum_{k=1}^{N}1/\mathscr{L}((q_k,r_k),\ (q_l,r_l)),$$

$$\rho_{Ll}^*=\int_V[1/\mathscr{L}((q,r),(q_l,r_l))]p_L^*(q,r)\cdot dV,$$

$$\rho_{\varphi l}=\int_V[1/\mathscr{L}((q,r),\ (q_l,r_l))]p_\varphi(q,r)\cdot dV$$

这种变换提出了一个值 $\Delta\rho_{Ll}^*$，该值的大小会随着比较靠近 (q_l, r_l) 的点 (q_k, r_k) 的分布位置改变而发生更急剧的变化，这就提高了上述准则的灵敏度，并可以对计算的不准确性多少有所补偿：随着$\mathscr{L}((q_k, r_k)、(q_l, r_l))$值下降到零，$\rho_l$ 值就无限地

增大。因此，如果给定某一 k 值会使 $\mathscr{L}((q_k, r_k), (q_l, r_l))=0$，那么，就可采用任何一个足够大的数（显著地大于 $\rho^*_{L,l}$）作为 ρ_l，因为，对上述算法来说，重要的只是要知道 $\rho^*_L-\rho_l$ 值的符号。可见，在利用准则 $\mathscr{L}^*_L$ 时，可以将一级算法变换一下：不是使公式（18）最大化，而是寻找能使每一个函数 p_L* 的量（31）最小化的参数。然后，如同上述的方法一样，二级算法就是要从函数 p^*_L 中选出一个其求得的极小化数值 $\mathscr{L}^*_L$ 接近于零的函数。这种办法特别适用于不满足条件（14）的函数，这在本人的一篇文章（杜鲍夫，1974）里做过较详细的讨论。

可能有这样的情况，有时，还必须解决根据训练资料从什么样的族中才能更好地选出近似函数的问题（在这里，我们不详细研究这个问题，对所探讨的情况来讲，要解决这一问题就象针对其它类型的函数所做的那样办就行了）（杜鲍夫，1970，1974）。这时，二级算法被用于每个相比较的函数族，结果从每个族中选出一个代表性函数。从这些标准函数中选出参数数目最少的函数。这就是三级近似算法。可能会出现有两个或多个函数同时满足所提出要求的情况。这时，为了缩小选择的范围，还需要有补充资料。

这里还要指出利用三级算法的一种特定的可能性。

与族 R_i 相对应的空间区域 V_Q，一般说来，可能以各种各样的方式展布。根据训练资料，应该选择能使在其中找到的函数 p^*_L 具最少数目的参数的那种展布。这时，就可对族 R_i 进行调整，使得所得出的解的精度对被处理资料而言是最大的。此外，这种调整表明类 R_i 之间的可能关系：它们对应的空间区域 V_Q 彼此越是接近，在某种意义上说，属于这些族的客体的可能差异就越小。

2．再来探讨这样的情况：在解决反演题时，对每个被研究的化学元素，都要按照对公式（15）和(16)所谈到的原则，找出用于近似函数 $C_\varphi(x, y, z)$ 的最佳函数 $C^*(x, y, z)$。

与前述的推论相类似，我们先假定函数 $C^*(x, y, z)$ 的形式已给出，需要的只是求出某种数目 L 的自由参数值。

观测浓度的集合$\{C(x_k, y_k, z_k)\}$，$k=\overline{1, N}$，在这里可以直接作为函数$C_e(x, y, z)$的标志来使用，而同一些点的函数$C^*(x, y, z)$的数值集合$\{C^*(x_k, y_k, z_k)\}$则可作为这个函数的标志来使用。这些标志间的距离，可以根据研究目的和研究条件用不同方法来确定。在许多情况下，可以认为，对值$C(x_k, y_k, z_k)$的干扰是由具对称统计分布的误差造成的。这时，最好取$C_k=C(x_k, y_k, z_k)$与$C=C^*(x_k, y_k, z_k)$的差的平均方差作为距离的量度，按照最小二乘法理论，也就是取量：

$$D^*=\frac{1}{N}\sum_{k=1}^{N}(C_k-C_k^*)^2 \tag{32}$$

如果不是C值本身的误差，而是它们的某个函数$g(C)$的值的误差按对称规律分布，那么就应进行相应的改变，这时公式就变成

$$D^*=\frac{1}{N}\sum_{k=1}^{N}(g(C_k)-g(C_k^*))^2 \tag{33}$$

例如，在地球化学研究时，常常采用快速光谱分析，其误差近似地按对数正态律分布。这时，$g(C)=\lg C$，在这种情况下，自由参数值可利用最小二乘法求出，这正是在根据公式（33）确定D^*值时能使该值极小化的方法。

在上述条件下，类似于前述一级近似算法，我们提出寻找近似函数$C^*(x, y, z)$的法则。

我们假定，函数$C^*(x, y, z)$应当从一定的有序函数族$\{C_L^*(x, y, z)\}$中选择。这个族的函数，其序号L越大，其自由参数的数目也就越多，而且对C_L^*中任何一个可能的参数值组合来说，都可在C_{L+1}^*中找到同样的一些自由参数值，当自由参数值取这种值时，对任何自变量(x, y, z)的值来讲，$C_L^*(x, y, z)=C_{L+1}^*(x, y, z)$。函数$C_{L+1}^*$的自由参数值也可用下述方法选择，即当函数$C_L^*$的（数目较少的）自由参数取任何值时，在非零维的点集(x, y, z)上都有$C_L^*(x, y, z)\neq C_{L+1}^*(x, y, z)$。

正如过去那样，我们先假定对每个函数都可使用一级近似算法。在这样的情况下，这些函数内的所有参数均可确定下来，并使最小可能值 $D^*=D_L^*$ 与每个函数相一致。这个值随着参数数目的增加而减少：根据最小二乘法的理论，如果 $L\to N$，那么 $D_L^*\to 0$。

根据监控取样资料，可以计算误差的估计值，因而也可计算出 D_φ 的估计值，而 D_φ 值是值 C_k 或 $g(C_k)$ 与相应真值的均方差。

在这种情况下，D_L^* 和 D_φ 值可以作为函数 C_L^* 和 C_φ 的标志来对待，而且根据近似法的基本原理，应当选择下述量

$$\mathscr{S}_4(D_L^*,D_\varphi)=|\Delta D_L^*|=|D_L^*-D_\varphi| \tag{34}$$

的值可达最小的函数作为最佳函数。

当量

$$\Delta D_L^*=D_L^*-D_\varphi \tag{35}$$

最接近零时，公式(34)便达到最小值。因而，公式(35)可用做近似准则。不难指出，与 $\mathscr{Z}(P_L^*)$ 类似，当数目 L 由很低的值变化到偏高的值时，量 ΔD_L^* 的值就会越过零。由此可见，借助 ΔD_L^* 准则找最佳函数 C_L^*，与借助 $\mathscr{Z}(P_L^*)$ 准则找最佳函数 P_L^* 是完全类同的。

与前面所讲的术语相对应，这种从给定族中找最佳函数 C_L^* 的法则，就是二级近似算法。

由于地球化学现象是多种多样的，因而我们要从中选出近似函数的函数族也是多种多样的。有时，人们不能在检验之前就知道函数族中的哪一个是最佳的，这时，就可利用前已指出的三级近似算法（杜鲍夫，1970，1974）。

3．上面所讲的用来寻找最佳近似函数 $p^*(Q,R)$ 和 $C^*(x,y,z)$ 的算法的区别仅在近似准则上，这种区别无关紧要，因为近似准则的选择仅仅取决于在类似情况下研究人员所拥有的资料情况。

有时，必须在缺乏误差资料的情况下解反演课题。这时，最好采用以最普遍的统计规律为基础的、与量(31)类似的准则，这种准则在作者过去的文章（杜鲍夫，1970，1974）中已详细地叙述过。

因此，我们所遇到的是最普遍的近似理论，在这里可以根据拟解课题的条件采用各种各样的准则。

上面我们探讨了一维准则，但一般情况下准则可能是多维的。由于本文篇幅限制，在此我们不能详细探讨这一问题的理论，仅仅指出，多维准则可以表示为一维准则的组合，因而可利用上述原理，一步步地从许多函数中选出所需要的函数。

实例 插图表示一个铜的残余晕的解释结果（是由 Л. С. 维里根据作者的算法计算的）。这种类型的晕很常见。早在这个术语被提出来之前，А. П. 索洛沃夫（1959）就开始对这类晕进行了研究。笔者在一篇文章（杜鲍夫，1964）中曾指出，在取样层位（靠近残坡积层表部）上，元素浓度在这些晕中的空间分布，消除了沿斜坡的位移之后，大致可由下述公式表述：

$$C_\varphi(x,y)=\frac{K}{2\pi\sigma^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}C_0(\xi,\eta)\cdot\exp\left(-\frac{(x-\xi)^2+(y-\eta)^2}{2\sigma^2}\right)\cdot d\xi\cdot d\eta \qquad (36)$$

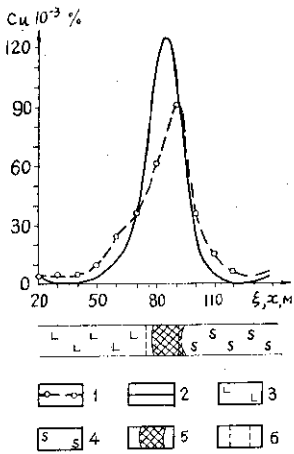
在这里，K ——一个系数，它取决于被研究化学元素在疏散岩石各粒级中的分布情况（这个系数应预先计算出来）；σ ——反映基岩物质内生改造作用的算子的参数（这个参数与晕形成时期内上述元素的质点所通过的平均距离成比例）；$C_0(\xi,\eta)$ ——称为算子的自变量（这是一个未知数，它反映元素在基岩上部层位中的空间分布；其参数是公式（36）的参数）。

按照前述，根据值 $C_k=C(x_k,y_k)$ 的集合来求这一课题的解，就表现为下述形式的近似函数：

$$C^*(x,y)=\frac{K}{2\pi\sigma^{*2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}C^*_{0L}(\xi,\eta)\cdot\exp\left(-\frac{(x-\xi)^2+(y-\eta)^2}{2\sigma^{*2}}\right)\cdot d\xi\cdot d\eta \qquad (37)$$

这个函数取决于参数 σ^* 和函数 $C^*_{0L}(\xi,\eta)$，在插图中所表示的情

况下，函数 $C^*_{0L}(\xi,\eta)$ 可从抛物线族（其自由参数的数目 L 是已知的）中选择出来。在这种情况下，对每个已确定的 σ^* 值，根据准则（34）或（35），就可以用上述方法求出数目 L 和能使所得函数 $C^*(x,y)$ 满足二级近似算法的相应的参数值。既然这些参数的值决定着函数 $C^*_{0L}(\xi,\eta)$，我们自然就可以得出所提出的反演问题的解。但是，在许多情况下 σ^* 值是未知的，这时，如果缺乏补充资料，其解就不能肯定：将不同的 σ^* 值代入公式（37），所得出的 L 是不同的，得出的函数 $C_{0L}(\xi,\eta)$ 的参数也是不同的，而这些参数则是完成使用准则（35）或（34）的二级算法的条件。



铜的残余晕的解释结果（哈萨克斯坦地球物理托拉斯的一个金属量测量地段）

1—疏松岩石取样得出的浓度值；2—基岩中铜含量曲线 C^*_L，它是根据解反演题得出的估计值作出的（$\sigma^*=12.5$ 米）；3—辉绿岩；4—细粒岩；5—据山地工程资料划出的矿体；6—由电子计算机圈出的矿体界线

掌握了被观测异常范围内某一数量 N_0 个点 (ξ'_i,η'_i) 上的实际值 $C_0(\xi'_i,\eta'_i)$，就足以用来作为补充资料。这些值可以在基岩的天然露头或人工露头取样时获得。如果进一步得出这些点上的值 $C^*_0(\xi',\eta')$ 与已知值 $C_0(\xi',\eta')$ 的最小平均偏差，就可对所提出的问题得出肯定的解。作为这种偏差的量度，也可以利用这些值的绝对平均偏差：

$$|\overline{\Delta C^*_0}|=\frac{1}{n}\sum_{i=1}^{N_0}|C^*_0(\xi'_i,\eta'_i)-C_0(\xi_i,\eta_i)|$$

因而，我们得出结论，在类似情况下必须利用两维准则（ΔD^*_L，$|\overline{\Delta C^*_{0L}}|$）。可见，知道一个点上的 $C_0(\xi',\eta')$ 值对此是不够的，因

为当$\sigma^* \neq \sigma$时，函数$C_0(\xi,\eta)$和用上述方法找出的函数$C_0^*(\xi,\eta)$是相交的❶。

随着σ^*改变，函数$C_0^*(\xi,\eta)$也发生变化，结果，交点就沿着函数$C_0(\xi,\eta)$滑动，并且这样一来，就可能与$C_0(\xi_i',\eta_i')$相吻合。所以，在$N_0=1$的情况下，即便使值σ^*与σ值差别很大，因而函数$C_0^*(\xi,\eta)$与$C_0(\xi,\eta)$的差别也很大，仍存在着满足上述的两维准则的可能性。

但是，在理论上，当值σ^*与σ存在很大差别时，函数$C_0^*(\xi,\eta)$和$C_0(\xi,\eta)$的交点同时落在两个或多个任意选出的具有已知值C_0的点上的概率为零。对不同的L来说，出现同样的$|\Delta D_L^*|$最小值的概率亦为零（即使对于除1以外的L值来说，这样的现象也不能排除）。所以在$N_0 \geqslant 2$时，满足了上述两维准则，无论在用函数$C_0^*(\xi,\eta)$估计函数$C_0(\xi,\eta)$方面，还是在用值σ^*估计值σ方面，几乎总能得出反演题的肯定的解。

在插图所举的实例中，计算是在有几个σ^*值的情况下利用准则(35)进行的；针对在这种情况下得到的C^*的函数计算了C_{0L}^*的最小值和相应的$|\overline{\Delta C_{0L}^*}|$值，其条件是假定实际的浓度最小值等于对这里所见到的岩石求出的背景值。然后，如有必要，应该准能使准则$|\overline{\Delta C_{0L}^*}|$接近于零的$\sigma^*$的值。

假如矿体有足够大的厚度（与σ值相比）和明显的地质界线，那么，这些界线可根据C_{0L}^*曲线的拐点（如插图所示情况）圈出。在所举的例子中，对于A_j值的计算来说，所得出的矿体厚度值及每个被研究元素的金属量（矿体界线之间的浓度积分）可以被包括于残余晕的标志q之中。而使用据一定的浓度C_{0L}^*确定经济界线的办法往往要更方便一些。

我们看到，观测值C的曲线和解反演题时所获得的值C_{0L}^*的曲线在插图上是相似的。但是，若不作上述计算，是不可能就C值的曲线形式作出上述推断的。此外，虽然二者相似，但C值曲

❶ 这是针对沿着单个剖面研究含量的情况而言，这已由П. И. 巴尔克(1974)所指出。

线不可能用来估计矿体厚度，与此同时，却可利用上述方法根据C^*_{0L}曲线（拐点）来估计矿体的厚度，并且与实际十分吻合。

在许多情况下，C和C^*_{0L}值要比插图上的差别大的多，这种情况过去我们已经描述过（杜鲍夫，1972）。

补 充 说 明

1．当我们根据$C^*_0(\xi,\eta)$值建立了曲线图、等浓度线平面图等之后，我们就获得了地球化学场的图示，这种图示与地球化学测量通常面临的反演题是极为相当的。插图就是这种以曲线图表示C^*_{0L}的一个实例。对地球化学场作面积性图示的实例我们在过去的文章里（杜鲍夫，1972）列举过。但是，对其必要的理论原则做过充分研究的只是残余地球化学晕和少数可以用公式(36)形式描述的其它地球化学晕（在杜鲍夫1974年的文章中指出的某些原生晕也属于这种情况）。所以，常常只限于对观测场作图示，而不是解反演题。这时，分级近似法有可能减少干扰的影响，从而有助于解释被观测异常。有时，人们为了减少计算不用三级算法，而是选用便于计算，只采用二级算法的函数族。由于在有色金属，稀有和分散元素的地球化学场中，噪音和有用信息都近似于对数正态统计分布，因此，利用对数抛物线和按滑动窗口对浓度对数作平均是合适的。详细的论证和操作步骤在以前的文章（杜鲍夫，1972б，1974）中曾描述过。但要指出的是，虽然这些方法在许多情况下效果好，并且具有相对简单、易懂和便于操作等优点，但仍有必要通过估计被研究客体的一些参数，即决定着它的经济价值及在其上进一步进行某种工作的效益的参数，力争更完备地解出反演课题。正如从本文一开始列举的一些考虑中得出结论那样，对这些参数进行了估计，就可以节省昂贵的野外工作经费。

2．上述近似原则不仅可以使用于本文直接谈到的那些目的，而且还可以应用于对某个过程可得出清楚或不清楚的逆反映的不

同情况，这时，结果不是以浓度的分布来表示，而是对另一些量的描述。

例如，在编制成矿图和预测图时，常常必须对与一定已知矿床有关的深断裂的位置作出估计，而且主要是根据这些矿床的位置来寻求解答。但是，这样的矿床并非直接产于拟寻的断裂中，而由于许多不完全清楚的原因离断裂有一定的距离。这些原因的影响可以用相互排斥力的算子的形式表示。所以，如果引入的是逆算子，即反映按一定规律相互吸引力的算子，就可明确说明上述断裂的位置。这种方法已由作者会同 С. И. 科斯特罗维茨基试验过，以便订正含金刚石的金伯利岩物质从深部侵入的构造带的位置（科斯特罗维茨基，杜鲍夫，1972）。

结　　语

1．在解释地球化学资料和对所研究的客体进行分类时，应该对这些客体的推定经济价值和在其上进一步进行各种可能工作所获效益的数学期望作出估计。

2．作这些估计所必要的数学步骤是建立在对统计规律以及空间地球化学场作等级近似的上述原则基础上的。用这种近似法，就能够在没有主观判断和只有最起码的先验信息的条件下，从可能的各函数中选择出所需要的函数。

列举的原则并非在地球化学涉及的所有情况下都是充分的。例如，这些原则就受到有无可能将一些近似函数确定地归为一类的限制。把研究范围扩大到也利用其它形式的函数是适宜的。

参 考 文 献

Балк П. И. К интерпретации вторичных геохимических ореолов. —В кн.: Ежегодник-1973. Новосибирск. "Наука", 1974, с. 254—258. (Отчет о работе Института геохимии СО АН СССР за 1973 г.).

Дубов Р. И. Концентрации химических элементов в ореолах рассеяния. —"Геол. и геофиз.", 1964, № 12, с. 44—55.

Дубов Р. И. Обобщенная регрессия и возможности ее применения для комплексной интерпретации данных. —В кн.: Ежегодник-1969 СибГЕОХИ. Иркутск, 1970, с. 289—296.

Дубов Р. И. К Теории классификации аномалий и решения обратных задач при геохимических поисках. —В кн.: Ежегодник-1971 СибГЕОХИ. Иркутск, 1972а, с. 305—309.

Дубов Р. И. Построение геохимических карт и разрезов с помощью электронно-вычислительных машин. —В кн.: Математические методы в геологии и геологическая информация. (Международный геологический конгресс, XXIV сессия, доклады советских геологов, проблема 16, симпозиумы, 104, 105). М., "Наука", 1972б, с. 80—88.

Дубов Р. И. Количественные исследования геохимических полей для поисков рудных месторождений. Новосибирск, "Наука", 1974. 278 с.

Костровицкий С. И., Дубов Р. И. Методика выявления линейных структур, контролирующих размещение кимберлитовых тел, с помощью ЭВМ. —В кн.: Гидрогеология и петрохимия Восточной Сибири, методика геологических исследований. Иркутск, Ин-т земной коры, 1972, с. 42—47.

Соловов А. П. Основы теории и практика металлометрических съемок. Алма-Ата, Изд-во АН КазССР, 1959. 266 с.

王家枢译

地球化学图的编制及地质解释实例

B. A. 奈格包埃尔

为了研究后贝加尔一些矿区的地球化学场，我们采用了不同的数学处理方法，方法的选择主要是依据测量的比例尺和研究的目的。

众所周知，任何浓度值都叠加有噪音——随机干扰（分析误差、天然的不均匀性等等）。因此，在地球化学测量中，为了区分出能据以查明矿体的有用信息——异常，并抑制观测场的噪音，使用数学方法处理地球化学数据是合适的。在对基岩进行取样时，由于浓度观测值的噪音分量特别大，最好对噪音进行抑制。而对疏松产物取样时，由于松散物质的天然混合提高了一定重量的样品的代表性，噪音要小一些。

我们利用P.И.杜鲍夫（1974）提出的等级近似法来区分浓度观测值场中的有用信息。杜鲍夫在关于这一问题的一篇文章（已选入本文集）中对这种方法的基本点作了论述，故这里不再赘述。在下面介绍的研究中，我们主要是利用第二级近似算法。

在所研究的场中，噪音的统计分布近似地服从对数正态律，因此，根据上述的近似理论，为了抑制噪音，事先将浓度换算成了浓度的对数，在做完必要的运算后，再查反对数换算回来。用了两类函数作为近似函数，即按滑动窗口求出的浓度对数的平均值和参数数目不等的对数抛物线，后者的参数根据观测浓度的对数用最小二乘法求出。

我们根据实际资料所做的研究，证实了杜鲍夫（1974）著作中提出的下述结论，即用最小窗口作多次平均的方法比一次用较大窗口的方法能更好地提取有用信息。其他研究者以前采用的是

后一种方法。多费点力并没有多大的关系，而应这样来选择窗口：即要使它最多只包括两个剖面上的两点。每当做第偶数次的平均时，窗口的中心就与采样点相重合。在做下一次平均时，对于位置处于边部的窗口来说，落入窗口中的点就不足四个（当窗口处于角上的位置时，为一个点；当窗口处于边部采样线上的中心位置时，为两个点）。所求得的平均值都放到窗口的中心点上去。第一次平均的结果便作为第二次平均的原始数据，依此办理，直至观测值（浓度对数）与平均值的绝对偏差最接近于观测值的平均绝对误差（它也是根据监控采样数据以浓度对数计算出来的）时为止。抑制噪音的这一方法很简单，也很便宜。

在观测网度较稀的情况下，为了尽可能准确地估计被研究场的极值的位置，采用了抛物线近似法。这种近似法在极值位于观测点之间，而且与它们的距离不等的情况下，就能将极值的位置判定出来；而求平均值法只有当极值在观测点上或者在它们之间的中点上时，才能指出极值的位置。抛物线近似法在于选择抛物线的阶次，以便使样品中所研究元素的浓度数平均绝对偏差与根据监控采样结果估计出来的浓度对数的平均绝对误差相差最小。如果地球化学观测的绝对误差是未知的，上述近似准则（浓度对数与近似函数之间的平均绝对偏差和这一误差值必须接近）就不适用。这时要用P.И.杜鲍夫（1974）提出的另一个准则：一对最接近的采样点上的浓度观测值应当同时高于相应点上的最佳近似函数值，或者同时低于相应的函数值，而这两种情况出现的次数应大致相等，就好象一对观测值中的一个较函数值高，而另一个较函数值低的情况一样。我们提醒一下，该准则应为第一种准则，而先前指出的准则应为第二种准则。下面，凡是未说明用何种准则的情况下，所用的都是第二种准则。

在后贝加尔那种条件下进行中、小比例尺（1:1万或者1:5万和更小比例尺）测量时，落入单个矿体晕圈中的取样点数不足以反映出这些晕的形态，而反映出晕圈的形态是对它们分别进行解释时必不可少的。因此，在这种测量过程中对被研究的场作统

计描述是需要的，这在某种程度上就相当于对详细测量时的异常组作特征描述。

这种做法与拟寻对象的概念应随比例尺改变而改变的观点是一致的(杜鲍夫,1974)：如果大比例尺普查时拟寻对象是矿体，则在小比例尺测量时寻找的就是矿床——常常是矿体的集合，乃至查明整个矿区和成矿省。单个矿体的晕在这种情况下甚至可以看作是噪音。当然，一些采样点可能偶然落在矿床范围以外的单独小矿体晕的极大值处，或落在矿床中心部分大量矿体晕之间的间隔中，这时如果不考虑这种偶然性（即不把它们作为噪音来抑制），便会导致对最有远景的详细普查地段的位置判断不正确。对由许多矿体组成的矿床的全部资料加以研究，并把那些按选定的测量比例尺应作为偶然信号处理的少数信号平滑掉以后，就可以将整个矿床造成的异常（对各个矿体异常来说它是综合性异常）圈出来。

浓度平均值或其函数（例如对数），离差、相关性和所谓相对远景率，都属于统计特征（杜鲍夫，1974）。需要提醒的是，根据上述著作中发展的理论原则，离差和相关一般不是直接从浓度来计算，而是从已确定的浓度的函数（使统计分布对称的）来计算。

方差图是利用滑动窗口作出的，它反映出地球化学过程的强度（杜鲍夫，1974）。在成矿作用过程的研究中这种方法得到了应用（塔乌桑，杜鲍夫、科兹洛夫，库兹明，1965）。滑动窗口的面积取决于要划分的有益信息的总规模。

配对相关图和局部相关图同样是利用滑动窗口作出来的。它们可用于确定异常性质和化学元素间的关系。窗口大小的选择和作方差图时的办法一样。所作的试验证明，不要根据各个样品中元素浓度的原始估计值，而要根据用求平均值法或者抛物线近似法作趋势面以后得到的数值来编制相关图，在理论上才是合理的。在杜鲍夫（1974）的著作中从理论上证明了这一论点，其要旨是：相关系数计算公式内的方差值，在抑制了噪音以后，在数值上就

会变得接近于被研究信息的天然方差值。

作图时对滑动窗口的每一个位置按公式计算相对远景率，这些公式是由杜鲍夫（1974）的下例公式演变而来的：

$$\Pi'_y=\int_{C\geq C_{н.б}} C.\ p(C)dC,$$

式中：C——样品中成矿元素的浓度；$p(C)$——C值的概率密度；$C_{н.б}$——从普查工业矿床的角度来说有意义的“最低有利”浓度值。在作次生晕测量时，$C_{н.б}$值要参照元素在松散岩石各粒级间的再分布系数来确定。该系数是А.П.索洛沃夫（1959）为了计算矿体露头的金属量而提出来的。这个系数可大于1，也可小于1，因为在采入样品的较小粒级的松散岩石中，各种元素的浓度比例关系与在基岩中形成的比例关系是不一样的。此外，在选择$C_{н.б}$值时还应考虑贫化系数，它反映着残余分散晕中由于物质扩散而使最高浓度降低的程度（杜鲍夫，1974）。

函数$p(C)$的参数既包括滑动窗口范围内的浓度平均值（或其函数），也包括反映地球化学过程强度的方差。该函数是对所有可能有的浓度分布的一种近似，并且根据它可以求得Π'_y值。Π'_y值是对浓度超过$C_{н.б}$并具经济意义的积聚体中的元素数量求出的估计值。由于取样点直接落入$C\geq C_{н.б}$的地点的概率很小，这样的浓度可能在样品中是发现不了的。

如果方差、相关性和相对远景率的数值偶然变化很大，并且这些值中的噪音分量很强，那么为了分出有用的信息就应该采用补充近似法，特别是求平均值法或者抛物线近似法，即类似于前面针对浓度场的近似提出的方法。

要提起注意的是，与根据较大的窗口并且只一次用它完成的对统计特征的计算不同，用于抑制所得场的噪音的平均法是多次使用最小窗口来完成的。

下面将列举根据后贝加尔一些具不同成矿特征的地区的岩石化学测量结果编制地球化学图的实例。由于技术上的原因，有时放弃采用更为精确的计算方法，但总的来说，所得结果说明了采

用地球化学数据的数学处理方法是可行的。这种实例对不曾进行过类似研究的专业人员在选择所需方法时将会有所补益。

后贝加尔铜矿区

研究区(图1)是铜矿省的一部分，原生沉积型铜矿化在早元古代岩石中广泛发育。铜是主要的矿产。这里已知有两个铜矿床（1号，2号）和一系列的铜矿点（3号，4号，5号）。

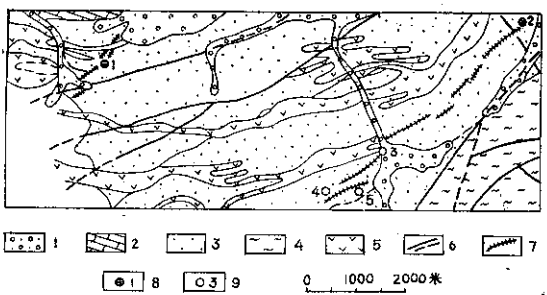


图1 后贝加尔某铜矿区地质略图

1—第四纪沉积物；2—钙质砂岩、白云岩、灰岩（下古生界）；3—石英-长石砂岩、粉砂岩（下元古界）；4—页岩、粉砂岩、砂岩、砾岩（下元古界）；5—辉质辉绿岩、辉长-正长-闪长岩（晚元古代）；6—查明的和推断的断裂；7—含铜砂岩和页岩；8—铜矿床；9—铜矿点

位于该区东北部的2号铜矿床为一中型矿床，矿区由细粒砂岩组成。含矿层断续地向北东方向延伸约6公里。在含铜砂岩剖面内，矿石矿物分布极不均匀。金属富集具工业意义的层状和透镜状矿体与实际上不含矿只含极少量浸染状铜矿物的砂岩成互层。位于该区西北部的1号矿段由细粒石英砂岩，少数情况下由细粒钙质砂岩组成。含铜砂岩产在亚组的上部，它与围岩的区别是存在铜矿化。根据一些细微的、有时勉强可辨的浸染状黄铁矿、黄铜矿和铜兰，含铜层可在走向上追索出1公里。矿层厚度

约40米。

主要根据太古代和早元古代岩石分布区的基岩取样结果，编制出铜浓度几何平均值（浓度对数算术平均值的反对数）图和铜浓度对数的方差图。

为了抑制与光谱分析误差和微小不均匀性有关的干扰，按图件比例尺以1×1公里的窗口对铜浓度对数实行最佳的三次平均，把该区经平均后的浓度图（图2*а*，）与地质图（见图1）进行对比，可以看出，已知的2号中型矿床被偏高的铜平均浓度（$18\cdot10^{-3}$%）等值线清晰地圈了出来。7.1—8.1·10^{-3}的等

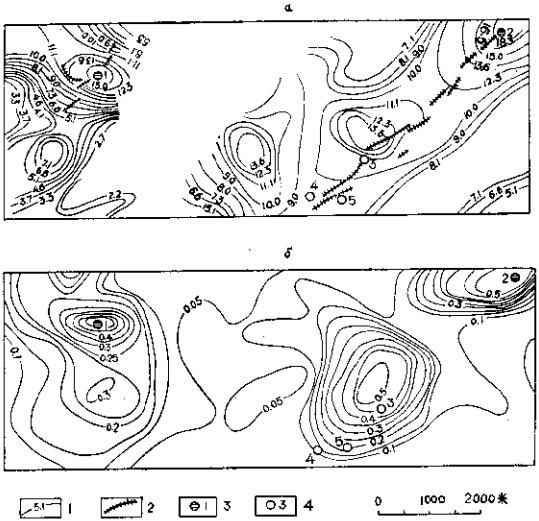


图2 后贝加尔某铜矿区根据基岩取样结果编绘出的铜浓度三次平均图（а）和根据铜浓度对数的三次平均结果编出的方差图（б）

1—铜浓度的等值线（单位为10^{-3}%）（а）和基岩中铜浓度对数方差的等值线（б）；2—含铜的砂岩和页岩；3—铜矿床；4—铜矿点

值线则将矿床周围的含矿岩组圈了出来。

含矿岩组范围内铜平均含量偏高区的形成与同生的原始沉积分散晕（有成矿迹象但并未成矿的岩带，据H.M.斯特拉霍夫）有关。4号矿点西北部的平均浓度最大值为 $13.6 \cdot 10^{-3}\%$，是由伴随有黄铁矿化的叠加成矿作用造成的。在该区西部的岩组中看到的晕的最大值为 $15 \cdot 10^{-3}\%$，它与已知的1号矿床相对应。

上述岩组除有铜富集外，钒的含量也比较高。用铜、钒和锌的偏高浓度可圈出辉长岩侵入体的露头。

与铜和钒不同，锌在沉积岩中的分布几乎是均匀的。在矿床以外的沉积岩中，只是在黄铁矿化强烈和裂隙度高的地段以及构造断裂带中，铜和钒的分散晕才会出现。

正如在三次平均的铜浓度图（见图2，*a*）上所看到的，平均浓度高的等值线的方向指示着控制铜矿化的地质构造的走向。例如通过北东向等值线就可以追索含铜层。

在根据铜浓度对数的三次平均结果作出的方差图（图2，*б*）上，方差最大值标出了1号和2号矿床。用于计算方差的滑动窗口的大小，按编图比例尺为 4×4 厘米。铜的偏高平均浓度与方差最大值的位置是一致的。

较高的方差值标出了铜矿点和含铜层，但它们的铜平均浓度比2号矿床稍低一些。只是在含矿层具有工业意义的地段，铜的平均浓度才会提高。从这种意义上来说，3号矿点的西北部方差值增高，大约是由沿溪谷的构造断裂带中出现了叠加矿化作用而造成的。

总之，在被研究的区域内，根据基岩地球化学取样结果能够确定铜的成矿聚积的位置。这时，可利用通过平均方法消除了干扰的浓度场的空间分布特点，以及反映出成矿作用强度的方差场。

涅尔钦-扎沃德地区

研究区位于东后贝加尔额尔古纳河地带。在该区的地质结构中有早、中古生代的沉积-变质岩，中生代的火山-沉积岩以及石炭纪和侏罗纪的侵入岩系（图3）。根据地质-地球物理资料，在该区可能存在有基米里岩浆旋回的侵入体，它在地表是根据局部露头填图填出来的，在深部是根据地球物理资料推断出来的。涅尔钦-扎沃德岩株（花岗岩、花岗斑岩）、戈尔内泽连图伊岩株（花岗闪长岩、花岗闪长斑岩）、伊凡诺夫岩株（闪长岩、石英闪长岩）均是基米里岩浆旋回的侵入体。

侵入体成分很可能决定着元素分散场的地球化学特征（华西里耶夫，格里戈里耶夫，1971）。

断裂发育是该区的一个特点，这些断裂在古老的基底上产生和发育，而后在中生代，在晚侏罗世和晚侏罗世后又重新活动。最后的一个构造期也决定着该区侵入体、岩墙和多金属矿床的空间分布。

研究区的研究程度较高。按500×100米网度对残积物进行的岩石化学取样资料被用来编制地球化学图。根据铅、锌和砷浓度作出了1:5万的原始浓度图。还以同一比例尺，根据上述元素的浓度对数编制出了下列地球化学图：平均浓度图、方差图、配对相关系数图和个别相关系数图。由于文章篇幅所限，这里仅就铅的图加以说明。当将铅的原始浓度图（图4）与地质图（图3）相对比时，发现松散岩石中的浓度异常与该区已知的多金属矿床的位置有些不一致。例如沃兹德维任矿床、列扎诺夫5号矿床、老沃斯克列先矿床、“三圣”矿床、老泽连图伊矿床、特罗伊茨矿场等就与铅晕的最大值稍有偏离。斯帕斯、格涅腊尔-萨文、叶卡捷里诺-布拉戈达特等矿床，则位于相邻异常之间的地带。

只有少数矿床（伊凡诺夫矿床和布拉戈达特矿床）的位置与

异常最大值是一致的。

上述大多数矿床的“位移”主要是由于天然不均匀性和光谱分析误差引起的干扰所造成的。

第一次平均时的窗口大小是：垂直于剖面（占据两条剖面）为2厘米，沿着剖面为0.5厘米。沿着剖面窗口的移动距为0.5厘米，垂直剖面窗口的移动距为1厘米。在第二、第三和第四次平均时，窗口大小相应为2厘米和1厘米（在剖面上包括前一次平均时得到数值的两个点），窗口移动距不变。在每次平均时，都作出等浓度平面图，以说明平均的程度。

在对铅浓度对数初次平均后的图上，晕的总体最大值圈出了列扎诺夫5号矿床、维尔赫涅-新沃斯克列先矿床、格涅腊耳-萨文矿床和沿列扎诺夫断裂的一系列矿床。在图上所看到的较高的铅平均浓度等值线呈北东向和近南北向延伸，与构造断裂的走向是重合的。

在铅的二次平均图上，偏高的浓度值与格涅腊耳-萨文矿床、沃兹德维任矿床和沿列扎诺夫断裂的一系列矿点相吻合。

在三次平均图（图5）上，最大的平均浓度值比二次平均图更明显地与已知矿床和矿点相一致。$6—15\cdot10^{-3}$%的等值线将涅尔钦-扎沃德和戈尔内泽连图伊两个矿田圈了出来。在新伊凡诺夫矿田上，铅的平均浓度值稍低，偏高的平均浓度值只见于伊凡诺夫矿床以北的一个地段上。

在分析该图的时候，研究区内铅的高浓度的分布主要与断裂，其次才是与岩石有确定的关系，这种情况是引人注意的。事实上，在铅浓度的三次平均图上，灰岩（下寒武统）和侏罗纪的花岗岩类是被$4—6.10^{-3}$%的等值线圈出来的。而砂岩-页岩岩系（下—中寒武统）的铅含量要更低一些。

在四次平均图上，被划分出来的与矿床相应的有用信息变得模糊了。鉴于这种情况，浓度的平均应该到第三次为止（这与第二级的近似算法是一致的）。

图6是根据涅尔钦-扎沃德矿区内残坡积物取样结果编绘的

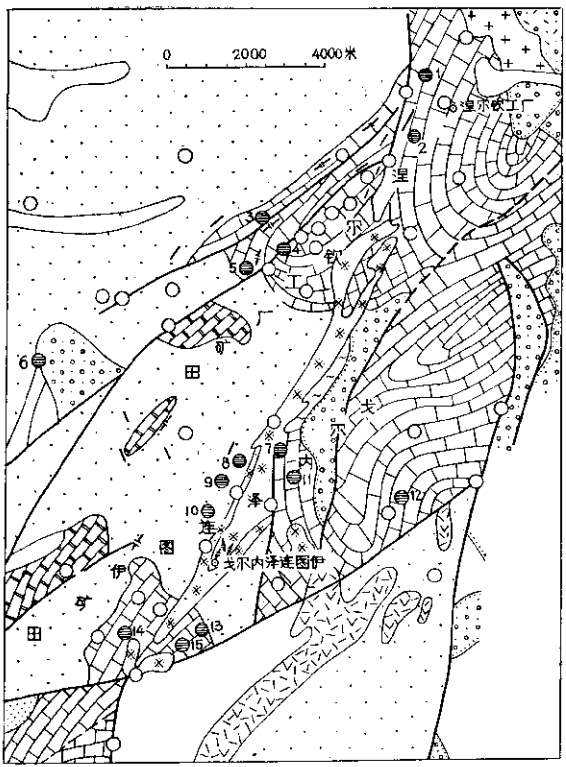

0
2000
4000米
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
a
б

图3 涅尔钦-扎沃德多金属矿床组的概略地质图

（根据A.A.洛克尔曼、Л.Н.列诺克、М.Л.斯捷丘克的资料编绘）

1—第四纪沉积；2—砂岩、砾岩、粉砂岩（下白垩统）；3—厚层泥岩、粉砂岩（上侏罗统—下白垩统）；4—砂岩、厚层泥岩（下—中侏罗统）；5—含泥质页岩夹层的白云岩（中寒武统—下志留统）；6—含灰岩和砂岩夹层的厚层泥岩（下—中泥盆统）；7—千枚状页岩、粉砂岩以及灰岩和砂岩夹层（下—中寒武统）；8—白云岩、含页岩夹层的灰岩（下寒武统）；9—a_1 花岗闪长斑岩；$б_1$ 石英斑岩（晚侏罗世）；10—花岗岩和花岗岩类（石炭纪）；11—斑岩类（早古生代）；12—a_2 石英斑岩，$б_2$ 辉绿斑岩（早侏罗世）；13—追索到的和推断的断裂；14—多金属矿床；15—多金属矿点。图中数字：1—彼尔沃-符托罗伊茨矿床；2—老沃斯克列先矿床；3—沃兹德维任矿床；4—维尔赫涅-新沃斯克列先矿床；5—列扎诺夫5号矿床；6—伊凡诺夫矿床；7—布拉科达特矿床；8—斯帕斯矿床；9—中心矿床；10—十月矿床；11—叶卡捷里诺-布拉戈达特矿床；12—马耳策夫-基耳京矿床；13—"三圣"矿床；14—格涅腊耳-萨文矿床；15—中新泽连图伊矿床

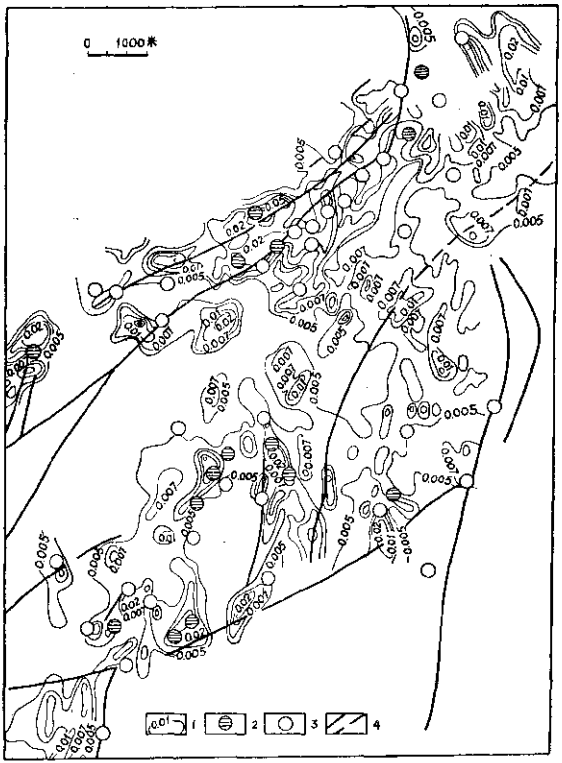


图 4 根据残坡积物岩石化学取样结果编绘的铅浓度图

涅尔钦-扎沃德多金属矿床组

1—铅浓度等值线（单位为%）；2—多金属矿床；3—多金属矿点；4—追索到的和推断的断裂

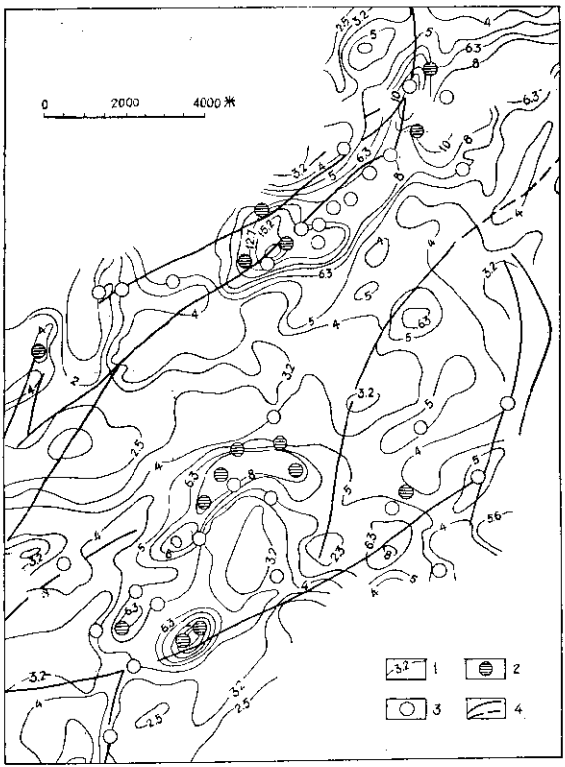


图5　经过三次平均平滑后的铅浓度图

配尔钦-扎沃德多金属矿床组

1—铅浓度等值线（单位为10^{-3}%）；2—多金属矿床；3—多金属矿点；4—构造断裂

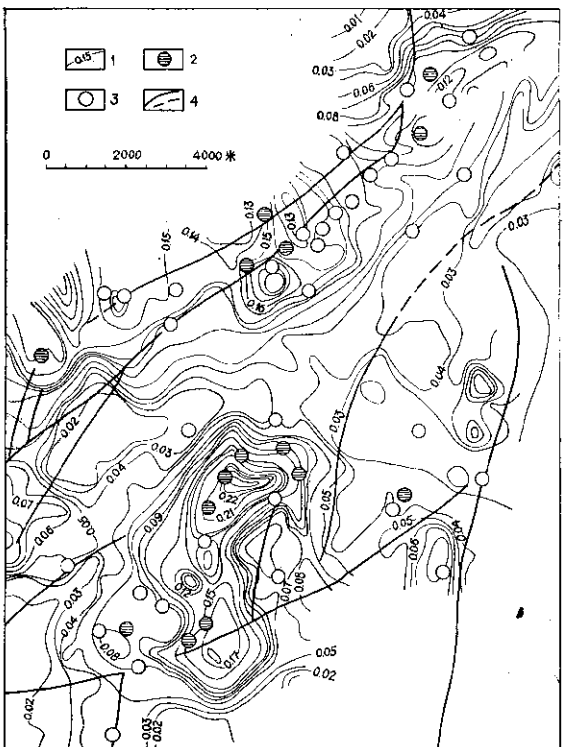


图 6 铅浓度对数的方差图

涅尔钦-扎沃德多金属矿床组

1—铅浓度对数的方差等值线；2—多金属矿床；3—多金属矿点；4—追索到的和推断的断裂

铅浓度对数方差图。

在这一图上划分出三个独立的、与矿田相应的地区，这些矿田已由其他研究者划出，并被在涅尔钦-扎沃德地区进行的详细地球化学测量（华西里耶夫，格里戈里耶夫，1974）所证实。正如在浓度平均图上看到的那样，方差异常正好位于矿化地段上，但在这里可以看到矿田被分成一些几乎是方形的块段。例如在戈尔内-泽连图伊矿田，块段的界线可由方差的高梯度显示出来，在每个块段范围内，方差变化不大。由密集的等值线标出的块段界线主要呈北东向和北西向。

所观察到的情况与多金属矿化在该区的分布和断块构造有关的假设是一致的。由于有松散盖层存在和成矿后的元素迁移作用的影响，只凭肉眼观察是很难证实这一假设的。

在许多情况下，根据煌斑岩的位移已用地质观测方法确定了北东向的古生代逆掩断层及与其伴生的北西向断裂发生过逆推-平移性的断裂错动。在矿化作用期内，一些岩石断块也发生过移位，不同断块遭受破坏的程度不同，因而沿裂隙运移的溶液使岩石发生蚀变的程度也是不同的。据地质图（图3）所见，这里分布着阿耳塔钦组砂页岩层，如果没有叠加的化学作用，这套岩层应该具有较低的铅浓度。因此，已观测到的偏高铅浓度就证明了有叠加的作用存在，而浓度对数的方差值则是叠加作用，特别是与构造破坏有关的叠加作用的反映。松散物质的混入，通常使方差值降低。因此可以设想，如果根据基岩取样结果作图，在图6上看到的情况将会更为明显。

在涅尔钦-扎沃德地区铅的相对远景率图（图7）上，当$C_{Pb}=0.03\%$时，所有的工业矿床都可以圈出来，只是格涅腊耳-萨文矿床和叶卡捷里诺-布拉戈达特矿床例外，前者氧化矿石延深达40米，后者埋藏很深。考虑到比例系数和贫化系数（在该区相应为1和25），残坡积晕中0.03%的铅含量，就相当于在矿体露头上的一米厚的晕中有0.7%的含量。根据工业上对多金属矿的要求，在矿床厚度为一米以上的情况下，0.7%的铅含量便属于工业品

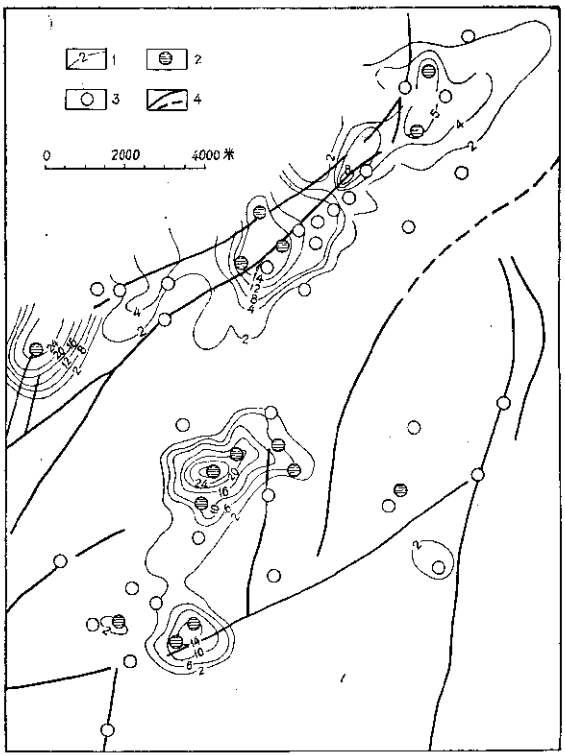


图 7 $C_{w\cdot6}\geq0.03\%$的铅的相对远景率图

涅尔扎-扎沃德多金属矿床组

1—相对远景率等值线（单位为 $10^{-3}\%$）；2—多金属矿床；3—多金属矿点；4—经证实的和推断的断裂

位。由此可见，该区松散岩石中0.03%的铅浓度就大致相当于基岩中的最低工业品位。

研究区各矿床范围内铅、锌和砷之间的数量关系反映在相关图上。相关图是用4×4平方厘米（依编图比例尺确定）的窗口对铅、锌和砷浓度进行三次平均以后编制的。从铅和锌的配对相关图（图8）上可以看出，在沃兹德维任矿床、列扎诺夫5号矿床、维尔赫涅-新沃斯克列先矿床、“三圣”矿床，残坡积物中铅和锌之间的相关关系是相当密切的。布拉戈达特矿床、叶卡捷里诺-布拉戈达特矿床以及它们的周围，处在正相关性很高的范围内。一个很有意义的现象是：在涅尔钦-扎沃德矿田范围内，在与各个断裂相对应的地段上，铅锌之间相关系数的估计值急剧变化：在沃兹德维任矿床—特罗伊茨矿点一线，与列扎诺夫断裂相邻的地区可见到正相关关系，而在沃兹涅先矿场—特罗伊茨矿点一线，与列扎诺夫断裂相邻地区则可见负相关关系。

在沃兹德维任矿区的“三圣”断裂、布拉达特断裂和扎列诺夫断裂的西支和南支地段，也可见到铅锌相关系数的高值。

根据所确定的相关关系，能够判断出研究区各个部分地球化学作用过程的不同特点，尽管不能把松散产物中的元素相关性与它们在基岩中的相关性等同起来（杜鲍夫，1974）。

不难看出，在戈尔内泽连图伊矿田和涅尔钦-扎沃德矿田，相关系数高值区主要与铅的方差偏高地段相吻合。这反映了一个自然规律性：对于一定类型矿床的形成来说，其必要和充分的条件是，除了某些一定的元素有相似的性状（这可以从它们的相关性上看出来）以外，还得有强大得足以改造岩石的地球化学作用（这从方差上看出来）。不言而喻，为此需要有足够数量的成矿元素，这从叠加有晕的基岩中成矿元素的平均浓度上反映了出来。

克利奇京矿区

克利奇京矿区位于普里额尔古纳河地区的西南部，包括十多

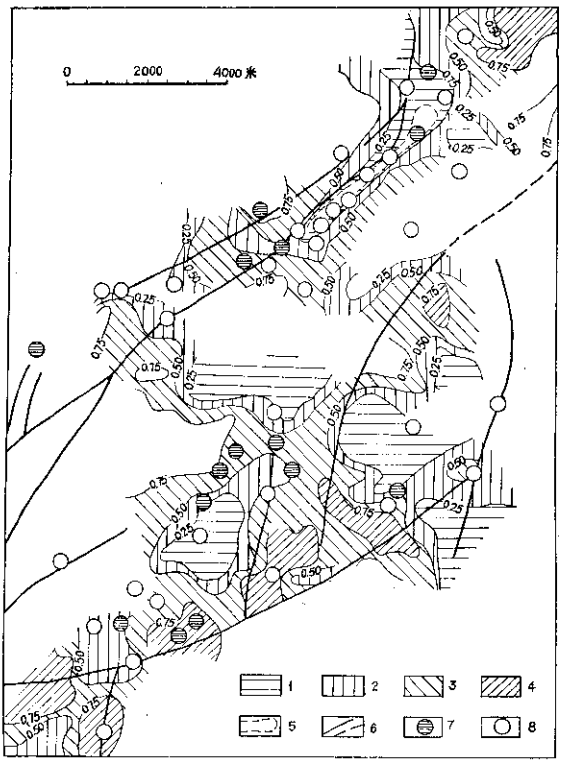


图 8　铅和锌浓度对数的配对相关图

根据浓度对数的三次平均结果作出。涅尔钦-扎沃德多金属矿床组

1，2，3，4—数值相应为0—0.25、0.25—0.5、0.5—0.75和0.75—1.0的正相关系数区；5—数值为－0.1至－0.3的负相关系数区；6—经证实的和推断的断裂；7—多金属矿床；8—多金属矿点

个金属矿床和矿点。该区主要由古生代和中生代的花岗岩类组成。晚元古代、早古生代的沉积变质岩以及侏罗纪和白垩纪的沉积碎屑岩和喷发岩居于次要地位（图 9，*а*）。

矿床赋存在早古生代的碳酸盐岩（灰岩、白云岩化灰岩、白云岩）和页岩交互层中。产在古生代花岗岩类中的几个矿床和矿点（卡缅矿床、法伊捷洛夫矿床等）属于例外情况。矿体主要由多金属矿石组成；石英-萤石矿化分布较少。稀有金属矿石则更

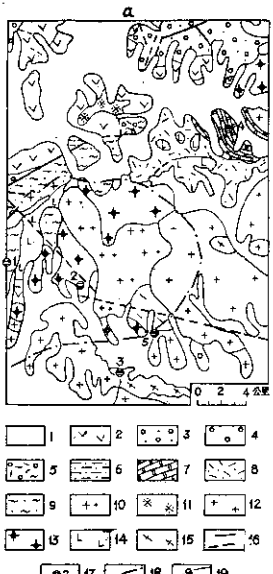
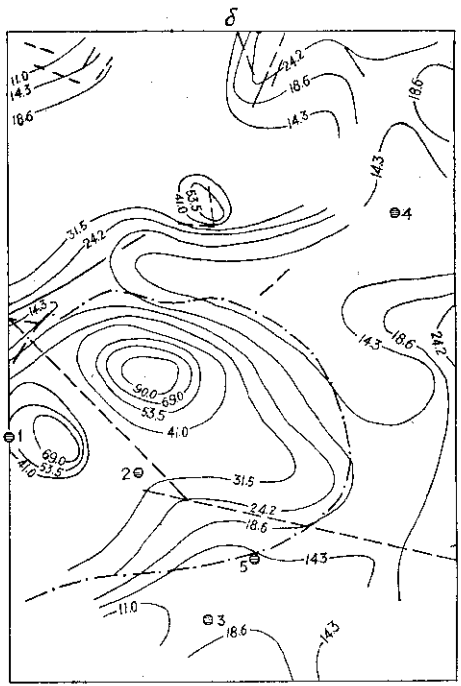
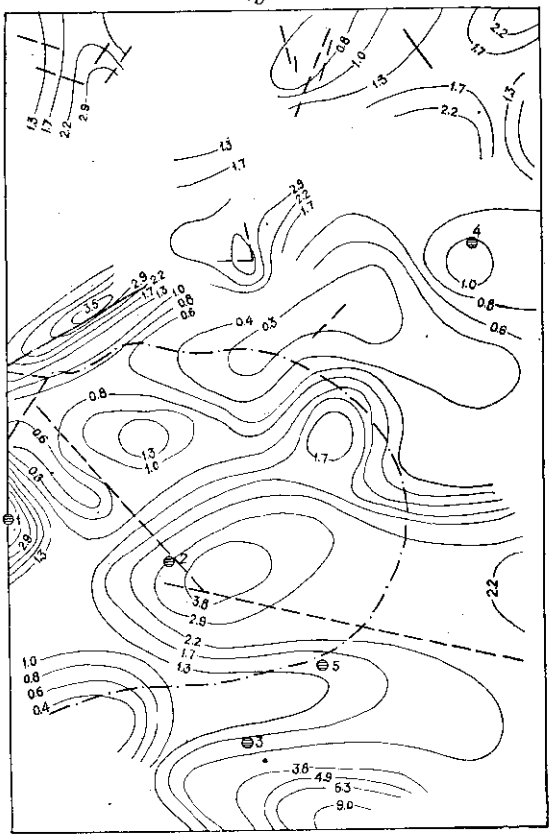


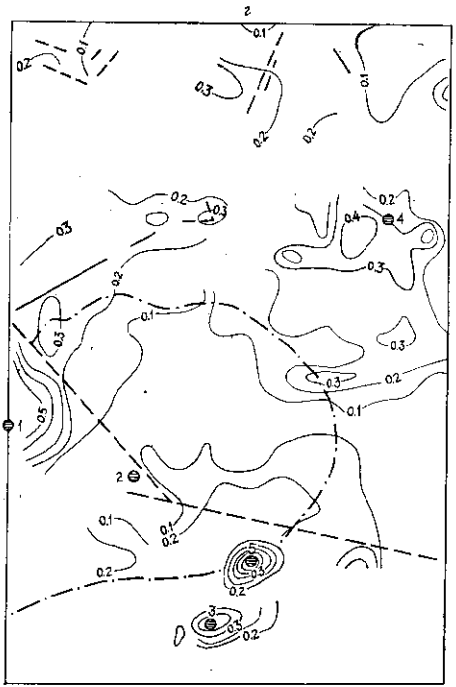
图 9

а—概略地质图（根据И. К. 阿布拉莫夫、В. М. 斯特罗耶夫、М. И. 斯捷丘克、Ф. Н. 阿列克谢耶夫、Д. Н. 阿列克谢耶夫的资料编绘）；*б*—经最佳的对数抛物线近似后得出的氟浓度场；*в*—铅浓度场；*г*—铅浓度对数的方差图。克利京矿区。

1—第四纪沉积；2—安山-玄武砂岩、非细岩、凝灰岩（上侏罗统）；3—砾岩、砂岩（中侏罗统）；4—砂岩、页岩（下侏罗统）；5—具砂岩夹层的砾岩（下罗侏统）；6—砂岩、页岩（下侏罗统）；7—白云岩、灰岩（下寒武统）；8—具灰岩夹层的砂岩、页岩（上元古界—下寒武统）；9—具碳酸盐岩夹层的页岩、砂岩、喷发岩（上太古界）；10—黑晶花岗岩（早侏罗世）；11—正长岩和闪长岩（早侏罗世）；12—黑云母花岗岩和白岗花岗岩（石炭纪）；13—花岗岩类岩石和白岗质花岗岩（石炭纪）；14—辉长-闪长岩、闪长岩（早古生代）；15—花岗片麻岩和片麻状花岗岩（早古生代）；16—查明的和推断的构造断裂；17—多金属矿床；18—根据地球物理资料圈出的黑晶花岗岩侵入体在地表的投影轮廓；19—氟（*б*）和铅（*в*）的浓度等值线（单位 10^{-3}%），铅浓度对数的方差等值线（*г*）；图上的数字表示矿床：1—沃兹涅先山矿床；2—卡缅矿床；3—梅耳尼科夫-霍尔基林矿床；4—丘平矿床；5—北矿井矿床


δ
11.0
14.3
18.6
24.2
18.6
18.6
14.3
14.3
53.5
41.0
31.5
24.2
4
14.3
14.3
18.6
24.2
90.0
69.0
53.5
41.0
1
69.0
53.5
41.0
2
31.5
24.2
18.6
5
14.3
11.0
3
18.6
14.3


B
1.3
1.7
2.2
2.9
0.8
1.0
1.5
1.7
2.2
1.3
1.3
1.7
2.9
2.2
1.7
4
1.0
0.8
0.6
1.5
2.9
2.2
1.7
1.3
1.0
0.8
0.6
0.4
0.5
0.8
0.6
0.8
1.3
1.0
1.7
1
2.8
1.5
2
3.8
2.9
2.2
1.7
1.3
5
1.0
0.8
0.6
0.4
3
3.8
4.9
6.3
9.0
2.2


2
0.1
0.2
0.1
0.3
0.2
0.2
0.1
0.2
0.3
0.3
0.2
0.4
4
0.3
0.3
0.2
0.1
0.3
0.3
0.2
0.1
0.5
1
2
0.1
0.1
0.2
5
0.3
0.2
0.2
3
0.3
0.2

为少见。矿床形成于晚侏罗世，受断裂带控制。

黑晶花岗岩与有锡、钨、铌和钽潜在含矿远景的库库别伊花岗岩类相似，因此可以推断，在克利奇京地区除了有铅-锌矿化以外，还可能存在着上述元素的矿化（塔乌桑等，1969）。

为了预测矿化的位置、界线和参数，应当将对黑晶花岗岩中元素统计分布的描述及与库库别伊花岗岩类的对比，改变为研究元素在整个矿区的空间分布。为了解决这一问题，利用了基岩地球化学取样结果，取样网度约相当于1:20万比例尺。

用最佳阶次的对数抛物线近似的地球化学浓度场，是根据氟、铌、铅、锌建立的。铅和锌浓度对数的方差图反映着物质在地质体中再分布作用的强度，是对地球化学场的一个补充。

在综合分析地球化学图和地质图时不难发现，在研究区内，在基尔-基林侵入体的黑晶花岗岩出露在地表的地方，与该花岗岩相应的上述元素的分散场具有比较高的浓度。根据地球化学资料解释的结果，基尔-基林黑晶花岗岩侵入体的面积是相当大的（塔乌桑等，1969）。经7阶对数抛物线近似后的氟浓度分散场（图9，*б*），极其明显地反映出黑晶花岗岩侵入体的规模和形态。0.0186%的氟浓度等值线的位置大致与侵入体在地表投影的北、东、南三条界线相吻合。在氟浓度偏高区内可划分出两个由0.09%和0.07%氟浓度等值线圈出的地段，它们分别与石炭纪的黑晶花岗岩和花岗岩类岩石相对应。这两个地段都有云英岩化带出现。氟浓度偏高，它的浓度对数的方差值也高（达0.4）。据此最大值可以确定，这里对寻找稀有金属是有意义的。由于在矿床周围发育着氟的矿染，借助氟来发现许多岩浆期后矿床（包括未出露地表的矿床）看来是可能的。因为氟是最活动的元素之一，它会进入大量能在很宽的物理化学条件下形成和保存下来的稳定矿物中。

在该侵入体范围以外，地球化学场的氟浓度递减顺序与下列岩石是相对应的：晚侏罗世正长闪长岩、石炭纪花岗岩侵入体的石榴石正长岩、侏罗纪火山沉积岩。最低浓度见于晚元古代—早

寒武世沉积岩中，例外的只是与黑晶花岗岩露头直接邻接的贝尔金组。在该区北部，氟浓度的近似场指出了一些推断的断裂。

和氟一样，研究区内铌的偏高浓度见于黑晶花岗岩侵入体范围内，并与侵入体近东西向的延伸方向一致。

用六阶对数抛物线近似出的铅浓度场（图9，б）也可圈出黑晶花岗岩侵入体，但不如氟浓度场那样清晰。在中生代花岗岩体中，可见到富集水平偏高的铅浓度（等值线最大值为0.0038%）。

除了上述矿床外，锌，特别是铅的高浓度场还见于将侏罗纪和晚元古代（贝尔金组）岩石分隔开的推断的构造断裂沿线。铅和锌浓度对数的高方差值区与该断裂带也是一致的。

可建议在与侵入体北缘相邻并由贝尔金组岩石（其中碳酸盐的比例很大）组成的地区普查多金属矿化。

铅和锌浓度场（沃兹涅先山矿床区）的特点是上述参数值偏高。这一异常与察干-佐洛图雅左坡的千枚岩组相对应，该千枚岩组是萨文5号矿床顶板千枚岩往东南的延伸。

利用浓度场近似法作出的浓度图，有助于查明元素浓度分布的主要规律。对浓度图起补充作用的浓度对数方差图，可以解释浓度场的内部结构。在浓度对数方差图上，可反映出与矿床和矿点相对应的的局部目标，但浓度场并不总能明显地把它们反映出来。例如，铅浓度对数的方差偏高值（图9，г）和锌的方差偏高值圈出了梅耳尼科夫-霍尔基林矿床和受从西向东延伸的巨大破碎带控制的北矿井矿床。产在石炭纪花岗岩类中的卡缅矿床在锌的方差图上得到了反映（数值为0.75—0.8），该矿床的矿石成分为闪锌矿和方铅矿。丘平矿床的特点是锌浓度对数的方差值偏高（0.75）。

应当指出，铅和锌的最高浓度主要分布在黑晶花岗岩侵入体之外或者接触部位上。在卡缅矿床附近两个断裂带交切的地方铅浓度出现最大值则是例外。在铅方差图上，可见到与侵入岩分布区相应的方差低值区；而主要由沉积岩组成的地段，其特点是方差具最大值和梯度很大。兼有铅和锌的高浓度及这两个元素的方

差值偏高两重特点，对预测多金属矿化是有利的。在根据地球物理资料确定的一个侵入体的北外接触带上，就看到了这种情况。

这就证实了一些作者（塔乌桑等，1969）的见解：在与黑晶花岗岩侵入体（深部填图圈出的）北部外接触带连接的含灰岩的沉积岩分布区寻找多金属是比较有远景的。

在研究乌柳姆矿田时，将直接根据松散岩石取样资料和根据金铋浓度场最佳抛物线近似结果编绘的金铋浓度对数相关图（图10）进行了对比。

正如所见，直接根据原始数值计算出的相关系数（图10，*а*）主要低于0.25的正值，这被认为是相关关系弱的表现。

用以对金铋浓度对数场进行近似的抛物线的最佳阶次是6和5。近似后得到的金铋浓度对数值，具有达0.5—0.75的相关系数。相关系数高值区明显处在所裂带和此处已知的自然电场异常上。显然，这样作出的图对研究地球化学现象可大有帮助。

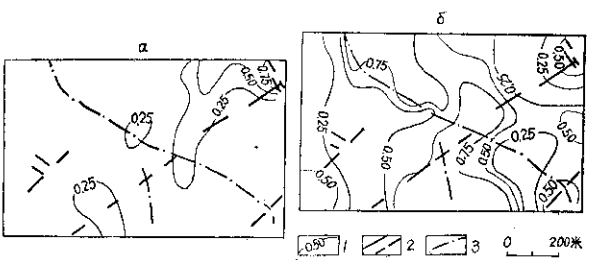


图 10　乌柳姆矿田中段金和铋的相关场

а—根据原始浓度编绘的；*б*—根据相应的最佳逼近抛物线值编绘的。1—配对相关系数值等值线；2—断裂带和高裂隙带；3—自然电场异常轴

普腊沃-库鲁耳提肯地段

（哈普切兰京矿结）

在进行地球化学调查（塔乌桑，贡多宾，祖布科夫，1974）

过程中，在哈普切兰京矿田以东查明了一系列局部锡异常。其中大部分与强烈硅化、绿泥石化和结核状硫化物有关。这些岩石的特点是锡、硼、锌、砷、铜、铅的浓度很高；而铋、银、铍、钨、锗，有时还有锑的浓度也偏高。上述这些元素的矿染也见于沿北东和北西向裂隙分布的石英脉和绿泥石化带中。斧石是锡-硼型矿化的特征矿物，它呈致密细粒集合体或者在围岩中呈浸染体产出。与斧石共生的有阳起石、水黑云母、绿泥石、白云母以及金属矿物。锡石倾向于呈细粒浸染状产在石英细脉之中。

后来，在普腊沃-库鲁耳提肯西部沿着中部断裂带也发现了锡-硼型矿化，矿化带长2公里多，如考虑钻孔见矿情况，则有3公里。

绝大多数斧石矿化出现在受北东向的中部断裂影响的地段，主要见于残坡积的碎石堆中。在个别地段可延伸50—200米。就已揭露出的原始产出情况看，矿化带的厚度是不稳定的，介于0.1—0.35米。锡-斧石矿化既见于矿区的不同部位，也见于距侵入体顶板的各不同高度上。在这方面，中部断裂带的资料是很说明问题的；在这里，斧石型矿化延伸3公里，并从黑云母的角页岩带中过渡到几乎未经蚀变的砂岩带中。

在图11上，表示出了根据普腊沃-库鲁耳提肯地段的基岩采样资料，用最佳对数抛物线（七阶）对锡浓度进行近似的结果。近似是利用第一种准则进行的。在这个平面图上标出了锡-硼的矿点及锡的含量。按100—150米的线距对残坡积物、残积物和基岩进行了取样。采样点点距为50米。

不难发现，沿中部断裂带分布的大多数锡浓度为0.02—0.07%的斧石-锡石矿化，都位于近似后的锡浓度异常的边缘。随着向异常最高值接近，锡含量提高到0.8%（在少数情况下达11%）。考虑到上述情况，以及在近似时已减少了与干扰有关的局部不均匀性的影响，建议在近似后的锡浓度最大值（$0.29\cdot10^{-1}$%）的范围内进行详细工作是合适的。也有可能这一最大值在深部是与两种类型的矿化（斧石-锡石矿化和哈普切兰京矿化）相对应

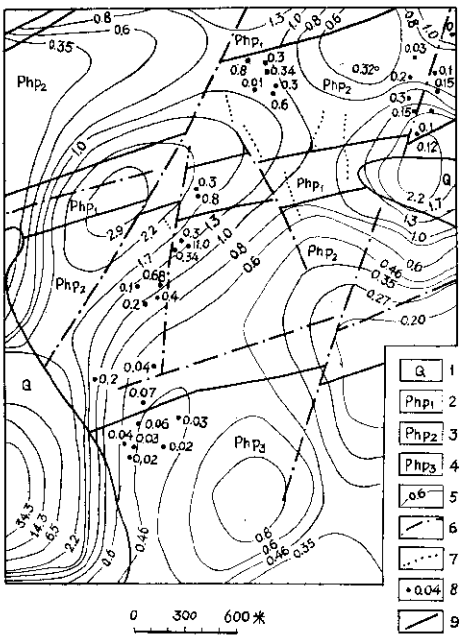


图 11 根据普腊沃-库鲁耳提肯地段基岩取样资料用最佳抛物线近似出的锡地球化学场

1—第四纪沉积；2、3、4—哈普切兰京组的第一、第二、第三亚组；5—锡浓度等值线（单位为10^{-3}%）；6—断裂；7—锡-多金属矿脉；8—锡-硼型矿点（数字表示锡浓度，%）；9—亚组的界线和第四纪沉积的范围

的。

在杜鲍夫、奈格包埃尔、巴尔克（1972）的著作中指出，在采样间隔比较大的情况下，为了找出地球化学异常最大值的位置，在一定条件下利用等级近似法是合理的。根据这一点，我们建议

在普拉沃-库鲁耳提肯河谷邻近地段，对根据取样数据内插结果得出的最大值用普查勘探工程进行验证。这一浓度为 $2.2\cdot10^{-3}$% 的最大值与东北部断裂带的位置是吻合的；根据地质资料，在研究区内这一断裂带对于找斧石-锡石型矿化是极为有利的。

在该地段范围内，除上述两个最大值外，还划分出第三个锡浓度为 $34.3\cdot10^{-3}$% 的最大值，它是通过浓度近似场的外推发现的。这一最大值位于提林河的第四纪沉积发育区内。在中部断裂的东南延伸部位打了一个钻孔，揭露出一个厚 2 米的斧石矿化带，锡浓度达1%。在上述最大值范围内发现工业意义的矿化，提高了在研究地段内预测锡矿的可靠性。

参 考 文 献

Васильев Е. М., Григорьев Н. Г. Использование металлометрической съемки для выявления рудной зональности (на примере Нерчинско-Заводского района, Вост. Забайкалье).—В кн.: Ореолы рассеяния месторождений Восточной Сибири. М., "Наука", 1971, с. 19—26.

Дубов Р. И. Количественные исследования геохимических полей для поисков рудных месторождений. Новосибирск, "Наука", 1974. 278 с.

Дубов Р.И., Найгебауэр В.А., Балк П.И. Применение ранговой аппроксимации для интерполяции геохимических данных при редкой сети опробования. —В кн.: Ежедгодник-1971 СибГЕОХИ. Иркутск, 1972, с. 314—318.

Соловов А. П. Основы теории и практики металлометрических съемок. Алма-Ата, Изд-во АН КазССР, 1959. 266 с.

Таусон Л. В., Гундобин Г. М., Зубков В. С. О новом

источнике оловянного оруденения в Восточном Забайкалье. —"Докл. АН СССР", 1974, т. 215, № 3, с. 704—707.

Таусон Л. В., Дубов Р. И., Козлов В. Д., Кузьмин М. И. Геохимическое значение дисперсии концентраций редких элементов в гранитоидах. —В кн.: Вопросы геохимии изверженных пород и рудных месторождений Восточной Сибири. М., "Наука", 1965, с. 12—22.

Таусон Л. В., Зорин Ю. А., Зорина Л. Д., Менакер Г. И. Саник Б. П. О некоторых закономерностях размещения месторождений Кличкинского рудного района (Вост. Забайкалье). — "Докл. АН СССР," 1969, т. 184, № 4, с. 929—932.

邱郁文译

利用内插法和趋势分析编制地球化学图和剖面图

A. A. 施曼斯基　Л. П. 普洛特尼科夫

地球化学场的天然起伏和取样误差会使地球化学图上出现复杂的图式，很难从这种图式中查明化学元素空间分布的总趋势。《金属矿床地球化学普查方法规范》（1965）提出一个划分地球化学场的原则：所有高于一定水平的化学元素含量都划为异常，而其余的则属于背景波动，不管这种波动是什么原因造成的。地球化学场的这种表示法会丢失有价值的信息，也不能对普查工作作出一个客观的评价。

在研究化学元素空间分布规律的时候，研究者要解决两个主要的课题：在区域内对观测值进行内插和把场（地球化学场、地球物理场等）分解为分量。当地球化学取样点在区域内呈散点状任意分布的时候，便产生第一种课题。这就需要用一个规则的等距离点网建立起场的原始数字模型，也就是需要解决在矩形网（也包括正方形网）的结点上内插一些数据基点❶的问题。

之所以产生分解场的课题，是由于有必要把局部异常地段和区域背景划分开来。有解决这一课题的两种主要方法。第一种是统计法，即计算落在某一范围内的所有观测值的平均值。该范围内的面积常被称为滑动窗口。第二种方法是趋势分析，就是选择某个能反映研究地段内整个场的主要规律的抛物面。❷

---

❶ 这里的数据基点，原文为исходные точки，指的是经内插后作为画等值线或作趋势分析的依据的点，它与原始数据点有联系而又有不同。——译校者注

❷ 这里用的术语与其他作者采用的术语不同。例如，上述两种方法可以看作是统计方法，因为它们都利用了数理统计学所研究的一些规律。趋势（тренд，源于英文trend）原则上可以用各种函数来表示，不仅是用早期研究中使用过、而且目前已使用颇广的抛物线来表示。——原编者注

统计内插和修匀

首先我们研究一下实现第一种方法（统计方法）的程序。这是一种通用程序，它既适用于内插，也适用于划分被研究场的区域分量，并且不与确定类型的函数相关联。此外，在该程序中，就内插误差最小的意义来说，实现了最佳性原则。在干丁(1963)的著作中，最佳内插原则是作为基础描述的。

方法的理论。要求根据周围n个点上的已知值$f_1, f_2, \cdots, f_n$，以最小的误差确定某个点上的场值f_0。我们将通过线性组合或者加权平均来确定f_0：

$$f_0=\sum_{i=1}^{n} f_i p_i+\delta, \sum_{i=1}^{n} p_i=1, \qquad (1)$$

式中：δ ——内插误差。

问题在于要给权p_i以能使得误差变为最小的值。最好是取误差的平方，其条件可写作：

$$\delta^2=\left[f_0-\sum_{i=1}^{n} f_i p_i\right]^2 \to \min \qquad (2)$$

求出δ^2关于p_i的导数，并使所得导数等于零：

$$\partial\delta^2/\partial p_i=-2f_0\cdot\sum_{i=1}^{n} f_i$$

$$+2\left(\sum_{i=1}^{n} f_i p_i\right)\cdot\sum_{i=1}^{n} f_i=0 \qquad (3)$$

或者，

$$\left(\sum_{i=1}^{n} f_i p_i\right)\cdot\sum_{i=1}^{n} f_i=f_0\cdot\sum_{i=1}^{n} f_i \qquad (4)$$

我们将括号中的符号i换成j，而在等式右边，我们将f_0移入总和号内，则得：

$$\sum_i \sum_j f_i f_j p_j = \sum_i f_0 f_i \qquad (5)$$

针对被作为f_0来研究的几个点都建立一个类似的方程，我们便得到一个用于计算权p_i的线性方程组：

$$\begin{cases} f_1 f_1 p_1 + f_1 f_2 p_2 + \cdots + f_1 f_n p_n = f_0 f_1 \\ f_2 f_1 p_1 + f_2 f_2 p_2 + \cdots + f_2 f_n p_n = f_0 f_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ f_n f_1 p_1 + f_n f_2 p_2 + \cdots + f_n f_n p_n = f_0 f_n \end{cases} \qquad (6)$$

在该方程组中，系数$f_i f_j$是相关矩的成分（这些系数在相加和除以n后，便得出相关矩）。

将$f_i f_j$和$f_0 f_i$用下式替换：

$$(f_i - \overline{f}) \cdot (f_j - \overline{f}) / \sigma^2 = r_{ij} \qquad (7)$$

$$(f_0 - \overline{f}) \cdot (f_i - \overline{f}) / \sigma^2 = r_{0i} \qquad (8)$$

式中：$\overline{f}$——平均水平；σ^2——在场的各点上f值的均方差；r_{ij}——各点的f值之间的规格化相关矩的成分，i，$j=1,2,\cdots,n$；r_{0i}——内插结点上的f值与第i个观测点上的f值之间的规格化相关矩的成分，$i=1,2,\cdots,n$。

此外，为了遵守权的规格化条件（1），我们将方程（6）的左边乘上比例因子K。结果得到一个新的线性方程组：

$$\begin{cases} K r_{11} p_1 + K r_{12} p_2 + \cdots + K r_{1n} P_n = r_{01} \\ K r_{21} p_1 + K r_{22} p_2 + \cdots + K r_{2n} p_n = r_{02} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ K r_{n1} p_1 + K r_{n2} p_2 + \cdots + K r_{nn} p_n = r_{0n} \end{cases} \qquad (8a)$$

如果将观测点$f_1, f_2, \cdots, f_n$的相互影响忽略不计，它们之间的相关矩就应当等于零。如果我们对这些矩的成分取下述值：

$$r_{ij} = \begin{cases} 1, & \text{当} i = j \\ 0, & \text{当} i \neq j \end{cases} \qquad (8б)$$

这时，平均来说误差就会很小。方程组（8a）中的非对角线系数应取零，这时它可变成：

$$\begin{cases} Kp_1 = r_{01} \\ Kp_2 = r_{02} \\ \cdots\cdots\cdots \\ Kp_n = r_{0n} \end{cases} \tag{9}$$

剩下的问题就是确定比例因子K的值，为此将方程式（9）的左边和右边分别相加起来：

$$K\sum_j p_j = \sum r_{0j}, \quad j=1,2,\cdots,n. \tag{10}$$

根据规格化条件，权的总和等于1，因此

$$K = \sum_j r_{0j} \tag{11}$$

因而，

$$p_i = \frac{r_{0i}}{\sum_{j=1}^{n} r_{0j}}, \quad i=1,2,\cdots,n. \tag{12}$$

这样，观测点的加权系数，便可通过这些点上的f值与内插结点上的f值的相关矩的规格化成分来确定。

假设，场相当均匀，并是各向同性的，也没有大的梯度或间断，这时就可以用一个自相关函数来描述它。正如经验所表明的，下列公式十分适用于这样的目的：

$$r_{0i} = (1 - r_{0i}/R)^3, \tag{13}$$

式中：r_{0i}——第i个观测点距内插结点的距离；R——内插半径，对内插结点产生影响的点的极限距离。必须指出的是，表达式（13）只有在$0 < r_{0i} < R$的条件下才有意义。

以表达式（1）、(12) 和 (13) 为基础，我们得到最佳内插公式

$$f_0 = \frac{1}{\sum_{i=1}^{i}\left(1 - \frac{r_{0i}}{R}\right)^3} \sum_{i=1}^{n} f_i \cdot \left(1 - \frac{r_{0i}}{R}\right)^3. \tag{14}$$

根据这一公式可以进行内插，并同时将区域分量划分出来。

为了在信息损失最少的情况下对矩形测网的结点作内插，必须这样来选择 R：在围绕大多数结点用半径画圆时，在各点周围起码会有一个点落入圆内。

如果在对矩形测网的结点作内插的同时还要划分出区域分量，R 值应当能与地质体的大小相比量，因为地质体的大小决定着研究者感兴趣的研究区构造特征，因而也决定着场的区域分量的特性及其所采用的自相关函数。在许多情况下，最佳 R 值接近于这些地质体平均规模的一半。一些作者（阿罗诺夫，戈尔丁，1971）也持有同样的见解。

电子计算机的算法和程序。预定要协同解决两个课题：

Ⅰ．进行统计内插并给出均匀矩形测网结点的数值；

Ⅱ．将观测场修匀并部分地抑制偶然误差和局部不均匀性。

在对算法作数学论证时，假设：

1．相邻点的场值是接近的和相互相关的。

2．场没有明显的线性不均匀性，场的自相关与方向无关，并且在研究区内没有大的变化。

3．相关函数可用根据下式确定的三阶抛物线分支来逼近：

$$K(r)=[1-(r/R_0)]^3, \qquad (15)$$

式中：r——观测点的距离；R_0——自相关半径，也就是测点之间的场值能保持相关性的临界距离（超过此距离相关性即消失，自相关函数就变得等于零）。当 $0\leqslant r\leqslant R_0$ 时，自相关函数值的范围为 $0\leqslant K(r)\leqslant 1$。

算法应按下列顺序完成：

1．在均匀测网中顺序计算的某个结点 M（这个点上的场值应根据最近的 n 个规则点上的场值来确定）周围以给定的半径 R 作圆。从 N 个观测点的集合中把位于该圆内的 n 个点划分出来。

2．根据公式

$$r_i=\sqrt{(x_i-x_M)^2+(y_i-y_M)^2} \qquad (16)$$

计算结点 M 至每一个划分出的点的距离，该式中：x_i，y_i——第 i

个观测点的座标；x_M，y_M——测网结点的座标。

3．根据公式

$$g_i=[1-r_i/R)]^3, \tag{17}$$

依观测点与结点M的距离计算各观测点的权。

4．针对被圈围限的地段根据n个观测点计算加权平均值：

$$Z=\frac{1}{\sum_{i=1}^{n} g_i}\sum_{i=1}^{n} f_i g_i, \tag{18}$$

式中：f_i——第i个点上场的观测值。给测网结点M以Z值。加权平均值Z依等式（17）中的R值不同而具有不同的含义。当$R=R_0$时，Z值有最大的可能就是与测网结点相吻合的圆心上的场值。在这种情况下，根据公式（17）对周围各点计算出来的加权系数，与相关函数（15）相对于中心的数值是相当的。如果$R>R_0$，抑制场的局部不均匀性的效果（修匀）就可以表现出来；这时，Z反映的仅是场的主要特征。在R值很大的情况下，局部异常对Z值的影响就被压制掉，只留下了场的系统分量（区域背景）。根据所得到的Z值的矩形矩阵（场的数字模型），进一步作出等值线图（作图是用“阿特拉斯”绘图仪完成的）。

趋势分析

趋势分析被广泛用来解决各种地质课题（阿罗诺夫，戈尔丁，1971；阿罗诺夫，戈尔丁，什尔戈诺娃，1971，等等）。趋势分析无论怎样都可以归结为计算其值与观测值最为接近的二维代数多项式（抛物面）的系数。多项式为双变量函数：

$$Z=E(x,y), \tag{19}$$

式中：x,y——研究区内的点的座标。对于N个观测点的集合来说，S^2值应当最小：

$$S^2=\sum_{i=1}^{N}(f_i-Z_i)^2=\min, \qquad (20)$$

也就是应使观测值与抛物面（趋势面）的差方和（偏差值平方之和）最小。在这种情况下，所说的是沿竖坐标轴的偏差。同时要假设x_i，y_i值不含偶然误差。这显然是一个十分重要的限制。如果能使所作的抛物面沿最短的路径（法线）最大限度地与观测点相接近，则这样作抛物面将是比较正确的。所提出的程序可实现这种可能性。

所采用的多项式的参数数目要选择得小于N。因此，在用这种方式的情况下，趋势面的建立就表现为对场的观测值进行平滑。这就减少了f_i值中随机变化的影响。观测点坐标值中的随机分量未被压抑。

为了在同时压抑f_i值中和观测点坐标x_i，y_i中的随机变化的情况下反映出趋势，我们不是使一维值z_i与f_i的差方最小，而是使三维点（x_i，y_i，f_i）与所作出的抛物面的差方最小。这种偏差是沿最短的路径，即从x_i，y_i，f_i点沿着法线至这个抛物面的距离来确定的❶。

建立趋势面的算法和程序 根据给定的观测点集合，可计算出下列抛物面方程式的系数（参数）a_{ij}：

$$Z(X,Y)=\sum_{j=1}^{t+1}\sum_{i=1}^{p}a_{ij}X^{(i-j+1)}Y^{(j-1)}, \qquad (21)$$

式中：p——多项式的阶次。参数的计算要按照原始观测点与近似出的趋势面的最短距离差方和最小的条件进行，即

❶ 在x_i，y_i，f_i的随机均方差（在数量上）大致相同的情况下，这个课题就是众所周知的合流分析的局部情况，在这些偏差差异很大的情况下，最好是改用更带普遍性和更为复杂的合流分析的方案。在最简单的情况下，要使x_i，y_i，f_i值的误差在数量上相等，只要通过把这些值的变量单位作相应的变化就可作到（如果误差与上述数值不相关，并且沿相应坐标轴不变的话）——原编者注。

$$D^2=\sum_{i=1}^{N}[(x_i-X_i)^2+(y_i-Y_i)^2+(f_i-Z_i)^2]$$
$$=\min, \qquad (22)$$

这里，如前所述，x_i，y_i——第i个观测点的坐标；f_i——该点上场的观测值；X_i，Y_i，Z_i——近似出的趋势面的相应点上正规分量的坐标和数值。X_i、Y_i、Z_i点是从第i个点（x_i，y_i，f_i）所作法线与趋势面的交点。

多项式的阶次p或者是给定的，或者是通过程序自动选择的。在这种情况下，要通过下述方式利用相邻点上场值的相关特性。多项式阶次的选择从一次多项式开始。对每一阶次，都计算出S_2值。相邻点上的相关性决定着S^2和D^2值的相近程度。开始，随着p值增大，S^2和D^2两个值都减小。但在多项式达到某个阶次$p+1$时，此两值之差由于S^2的剧增而大大加大。在其后发生这种变化的p值，被取为最佳值。

用本文作者所编出的程序，根据给定的观测点（达10000个）集合计算出多项式（21）的系数。接着根据表达式（21）计算出均匀测网各点上的Z值。将这些值输入“阿特拉斯”绘图仪，由它根据这些值画出趋势面。

用БЭСМ—4型计算机计算和绘制一个中型图幅（1000个观测点）所费的时间，很少超过一个半小时。

例　1

上述计算方法与被研究场的种类无关。f值和Z值既可理解为地球化学样品中的化学元素浓度、各种地球物理场值，也可理解为地球化学和地球物理数据的解释结果。因此在第一个例子中并没有直接使用地球化学的结果，而是使用地球物理的结果。后者可以十分充分地反映出上述方法。

1．我们将引用乌斯特-库特勘探区根据地震勘探资料确定反射面标高的处理结果。野外工作是由东部地球物理托拉斯以1:20万的比例尺进行的。观测点1067个，网度是不规则的（图1，a）。

平均标高为1919米，深度总方差等于1419（相应的均方差为37.7米）。工作区南北宽104公里，东西长120公里。

按照统计内插程序以三个 R 值（6、10、20公里）进行了处理。为了便于比较，在相应的图 1,*а*,*б*,*в* 的左上角标出了窗口的大小。

图1，*г*是按第二种程序处理后得到的趋势图。多项式的阶次值等于 4，它是经自动选择后得到的最佳值。趋势面的方程式为：

$$Z=-1790-64.7x-151y+8.03x^2+83.6xy+19.6y^2$$
$$-2.23x^3-0.70x^2y-23.7xy^2+5.44y^3+0.30x^4$$
$$-1.21x^3y+2.54x^2y^2+0.26xy^3-0.59y^4$$

趋势的性质或者“强度”一般可根据被研究地段 Z 值的方差

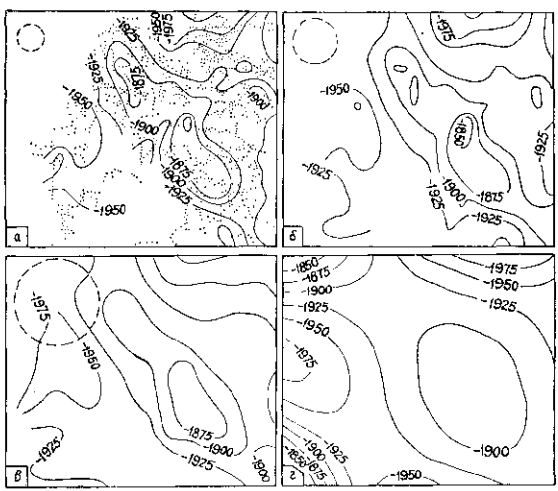


图 1　用统计内插法作出的反射面图

（*а*、*б*、*в*）和用趋势分析作的反射面图（*г*）

占f_i值总方差的比率来估计。在此例中，趋势方差为611。这仅占f_i值总方差1419的43%。其余的57%是由反射面起伏的局部不均匀性造成的与趋势面的偏差。

选择4次多项式的最佳性可由表1看出来。

与前4个p值相应的S^2和D^2值是接近的，并顺序变小，而且$D^2<S^2$。当阶次为5时，S^2值急剧增高，而D^2值还保持原来的减小趋势。因此，四次抛物面的S^2值最小，与观测总体吻合得最好。

表 1

p	S^2	D^2	p	S^2	D^2
1	1318	1316	4	813	813
2	1015	1012	5	217000	603
3	961	957			

表 2

元　素	趋势面方差（占总方差的%）	趋势面阶次
金	3.0	3
铅	20.5	3
锌	9.3	3
铜	25.3	4
砷	16.2	4

趋势面（图1，2）反映着该区构造最主要的方面——存在着一个由南东向北西延伸的大型背斜构造。同时，明显可见该构造在北西方向上有继续延伸和发展的趋向。在西部，标高稍有偏高。这可能是研究区之外又一个背斜的反映。

例　2

伊尔库茨克地质局曾在一个面积为0.9×1.5平方公里的地段

上进行金属量测量，以查明金及其伴生元素的次生分散晕。按50×20米的网度采了978个岩石地球化学样。用这些样品分析了金、铅、锌、铜和砷。样品分析方法是快速光谱分析（近似定量）。

根据这些数据，利用“阿特拉斯”绘图仪对每种元素编制和勾画了趋势面（图2）。

表2列出了趋势的“强度”值，即Z值的方差占被研究元素

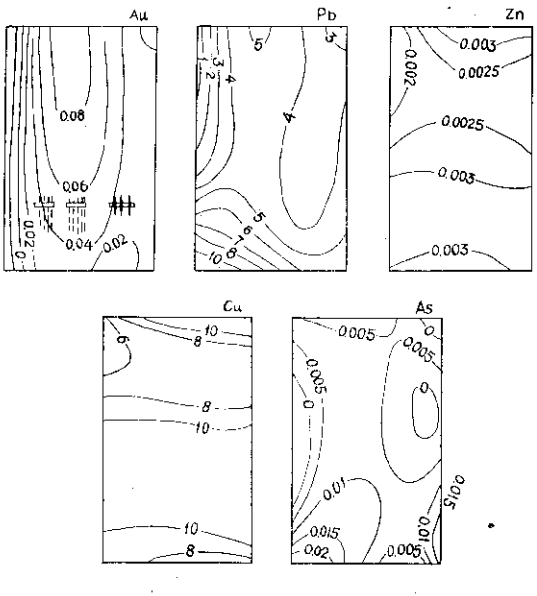


图 2　金、铅、锌、铜和砷浓度（假定单位）的趋势面
（比例尺1:1000）

1—变质强烈的砂岩、页岩；2—含金和硫化物的矿化带；3—石英脉

原始浓度值总方差的比率。这里，值相对较小表明偶然的无规律的浓度波动很大。金的趋势“强度”特别小，在我们看来，是由于“高”样品（浓度“高”达30克/吨）的影响所至。

表2还列出了趋势面的最佳阶次值（准确地说，是反映趋势面的多项式的阶次值）。这些值是自动选择的，但我们坚信它们在实际上是正确的：较低阶次的多项式提供的是过分平滑的图形，与所查明规律性的接近程度较低；当阶次较高时，场的无规律的偶然变化开始在趋势面的图形上起较大作用。特殊的是，从第三阶次开始，趋势的方差实际上几乎保持不变。

对结果进行分析可得出如下结论：

1．在金、铅、砷的浓度趋势图上，等值线的形态非常相似，即都呈南北向延伸，这与研究地段的砂岩层和页岩层的走向是吻合的。但金的浓度梯度的方向则与铅、砷不同。

2．铜和锌的浓度场非常相似，而与金、铅和砷的浓度场很不相同：铜和锌在趋势图上的等浓度线有颇大部分大体垂直于砂岩层和页岩层的走向，也垂直于金、铅和砷的趋势图上的等浓度线。这种现象的原因尚不清楚，仅仅可以推测，景观条件对铜和锌在研究地段第四纪沉积物中的分布影响颇大。

3．在该地段曾挖过一些探槽，揭露出了硫化物和金的矿化带（根据刻槽取样资料）。根据金的趋势面图，可以大胆建议在地段北部继续挖探槽。

上述资料证明，应用数学方法和电子计算机作地球化学图大大减少了手工劳动量，并为预测和确定进一步普查勘探方向提供了新的可能性。

参 考 文 献

Аронов В. И., Гордин В. М. О применении тренд-анализа в геологии. —“Труды ВНИГНИ”. вып. 103. М., 1971, с. 46—70.

Аронов В. И., Гордин В. М., Шаргинова А. И. К

вопросу о построении графиков и карт изолиний в геологии и геофизике с помощью ЭВМ.—"Труды ВНИГНИ", вып. 103. М., 1971, с. 71—124.

Гандин Л. С. Объективный анализ метеорологических полей. Л., ГИМИЗ, 1963. 287 с.

Дубов Р. И. Применение математических методов и счетных машин при геохимическом картировании. —В кн.: Эффективность геохимических методов поисков рудных месторождений и широкое внедрение их в практику геологической службы СССР (материалы к Всесоюзному семинару). М., ИМГРЭ 1967, с. 33—34.

Дубов Р. И. Количественные исследования геохимических полей—В кн.: Ежегодник-1968 СибГЕОХИ, Иркутск, 1969, с. 246—254.

Зайдель А. Р., Лохматов Г. И., Таусон Н. Н. Математичетские методы при построении карт в практике геологических работ. М., 1967. 73 с. (Труды ВНИИОЭНГ. Сер. Нефтегазовая геол. и геофиз.).

Инструкция по геохимическим методам поисков рудных месторождений. М., "Недра", 1965. 228 с.

Крамбейн У., Грейбилл Ф. Статистические модели в геологии. М., "Мир", 1969. 397 с.

Литвиненко О. К., Мелихов В. Р., Русьянов Ю. Т., Рукин М. Д. Применение электронных вычислительных машин в гравиразведке. М., ОНТИ ВИЭМС, 1969. 32 с. (БНТИ. Сер. Региональная, разведочная и промысловая геофиз., № 40).

Лохматов Г. И., Алаев Г. Т. О ппименении электронных вычислительных машин при тектоничес-

ком анадизе геологических структур. —В кн.: Математические методы геохимических исследований. М., "Наука", 1966, с. 99—118.

邱郁文译

论残留地球化学晕的解释

П. И. 巴尔克

我们知道，在凸型山坡上，矿体上方的疏松残-坡积物中发育着的残留分散晕，是相对容易研究的，因为在这种情况下，成矿元素赋存于在表生作用带中未遭受强烈溶解的原生矿物或次生矿物中。大量的稀有和有色金属矿床都有这样的残留分散晕，所以，研究这种分散晕有一定的实际意义。

用数学-物理模拟法研究残留分散晕始于 A. П. 索洛沃夫(1959)。特别是他确定了残留晕的金属量和产在晕下面的矿体露头的金属量之间大致的比例关系。后来，Р. И. 杜鲍夫(1964)又导出了残留晕和基岩中化学元素浓度相关性的一般性数学表达式。本文打算改进他所提出的对在近似一维的环境中（如在长条状矿体中部的上方）形成的残留晕进行解释的方法。

一种化学元素在残留晕中的异常浓度C(x)与尚未遭到破坏的基岩中该元素的异常浓度$C_0(\xi)$有关，这种关系服从下列积分方程（杜鲍夫，1964）：

$$C(x)=\frac{k}{\sqrt{2\pi}\cdot\sigma}\int_{-\infty}^{\infty}C_0(\xi)e^{-\frac{(\xi-x)^2}{2\sigma^2}}\cdot d\xi \qquad (1)$$

这里，σ——分散参数；k——比例系数，下面假定它为1；x——采样剖面的座标。晕的解释就是要根据松散产物中的元素浓度值$C(x)$估计出成矿元素的浓度（富集程度）$C_0(\xi)$，这就是所谓的解反演问题❶。在这种情况下，由于参数σ是不确定的，因而解反演题就变得十分复杂。由于参数σ的不确定性，不仅需要在解任何反

❶ 下面，我们将把“晕的解释”这一概念与其数学上的类似术语——“解反演题”和“解方程(1)”等同起来——原注

演题之前对解释的肯定性作专门研究，而且还使求得方程(1)的近似解（例如用人所共知的调整法去解）变得困难。方程(1)的解析解，只有在最简单的晕的解释条件下才能得到，这些条件是：异常$C(x)$须是由具有长方形断面的均匀矿柱，或由无限延伸的层状矿体，或由所研究元素含量不同的两种岩石的接触面引起的（杜鲍夫，1964）。但是，矿体的各种实际形态绝不止这几种。

本文想说明求解方程(1)的解析方法，这种解法可归结为对拟求函数$C_0(\xi)$进行足够适当的逼近，特别是还可用它找出对所研究的晕作肯定解释的某些充分条件。下面来探讨在数轴上处处有定义的线性独立的函数系统

$$\Psi=\{\psi_i(\xi)\}_{i=0}^{\infty} \qquad (2)$$

我们要在具有L个参数的函数ψ_i的线性组合中找出方程(1)的解：

$$C_0^*(\xi)=\sum_{i=0}^{L} a_i\psi_i(\xi) \qquad (3)$$

同时，系统Ψ要这样选择才合理，即应适当地选择其参数a_i的数目L和a_i的数值，以便使函数(3)尽量准确地逼近浓度$C_0(\xi)$的任一实际分布。例如，三角系统$\Psi_1=\{1,\{\sin i\xi,\ \cos i\xi\}_{i=1}^{\infty}\}$和代数单项式系统$\Psi_2=\{\xi^i\}_{i=0}^{\infty}$就符合这种要求，前者可以在长度$2\pi$的区间内平均地逼近任何连续函数，后者可以均匀地逼近连续函数。

我们分别用T和P表示三角多项式和代数多项式族，这些族分别由系统Ψ_1和Ψ_2中函数的所有可能的线性组合产生。我们还将由函数族T和P来选择近似分布(3)。利用表格积分法（格拉德什捷英，雷日克，1962），在一系列简单的变换之后，函数$C(x)$的积分表示式(1)以最终形式表达如下：

$$C(x)=\frac{1}{\sqrt{2\pi}\cdot\sigma}\int_{-\infty}^{\infty}\left[a_0+\sum_{i=1}^{L}(a_i\sin i\xi+a_{i+L}\cos i\xi)\right]\times$$

$$\times e^{-(\xi-x)^2/2\sigma^2}\,d\xi \equiv a_0 \sum_{i=1}^{L} (a_i \sin ix + a_{i+L}\cos ix)\cdot e^{-(i\sigma)^2/2} \tag{4}$$

$$C(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\int_{-\infty}^{\infty} \sum_{i=0}^{L} a_i \xi^i e^{-\frac{(\xi-x)^2}{2\sigma^2}}\,d\xi \equiv$$

$$\equiv \sum_{i=0}^{L} a_i \sum_{k=0}^{\lambda} (\delta^i + 2k)\,(2\lambda - 2k - 1)\,!!\times \sigma^{2(\lambda-k)}\, x^{\delta+2k} \tag{5}$$

这里：

$$\delta = \begin{cases} 0, & \text{如果 } i \text{—偶数} \\ 1, & \text{如果 } i \text{—非偶数，} \lambda = [i/2],\ (-1)!! = 1 \end{cases}$$

假设σ已经给定。这时，在族T和P中，方程(1)具有肯定的解，这是直接由恒等式(4)和(5)右边的三角函数和幂函数的线性独立性得出来的。但是，正如上面已经指出的，σ值实际上不知道。在这种情况下，在族T和P中，方程(1)是多解的。实际上，假设对于在族T中的某个函数$C(x)$和分散参数σ_0来说，方程(1)具有解：

$$C_0^*(\xi,\sigma_0) = a_0(\sigma_0) + \sum_{i=1}^{L} (a_i(\sigma_0)\sin i\xi + a_{i+L}(\sigma_0)\cos i\xi) \tag{6}$$

这时，在具有另外一些σ值的情况下，同一个函数$C(x)$就会相应地有一个无限的解集：

$$C_0^*(\xi,\sigma) = b_0(\sigma) + \sum_{i=1}^{L} (b_i(\sigma)\sin i\xi + b_{i+L}(\sigma)\cos i\xi) \tag{7}$$

其中函数b_i与a_i有关，关系如下：

$$b_0 = a_0,\ b_i e^{-\frac{(i\sigma)^2}{2}} = a_i e^{-\frac{(i\sigma_0)^2}{2}},\ b_{i+L}\ e^{-\frac{(i\sigma)^2}{2}} =$$

$$= a_{i+L}\ e^{-\frac{(i\sigma_0)^2}{2}},\quad i = \overline{1,\ L} \tag{8}$$

在族P中，有两种不同的分布：

$$C_0^*(\xi,\sigma_0)=\sum_{i=1}^{L}a_i(\sigma_0)\xi^i \tag{9}$$

和

$$C_0^*(\xi,\sigma)=\sum_{i=1}^{L_1}b_i(\sigma)\xi^i \tag{10}$$

而当$L=L_1$，$a_L=b_L$，$a_{L-1}=b_{L-1}$，

$$a_j=b_j+\sum_{k=1}^{\mu}\binom{2k+j}{j}(2k-1)!!(b_{2k+j}\sigma^{2k}-a_{2k+j}\sigma_0^{2k}) \tag{11}$$

$$\mu=(L-j)/2,\quad j=L-2,\ L-3,\cdots\cdots,1,0$$

的时候，并且只是在这时候，这两种分布才会使得 晕中 浓度$C(x)$有同样的分布。递推关系(11)决定着方程(1)的解为无限族。

为了消除解释的无限多值性(固定参数σ)，必须假设拟求函数$C_0(\xi)$在某个点ξ_0上的值$C^{(0)}$是已知的，这实际上总能办到，因为在研究地段内通常总会有天然的或人工的露头，能精确地知道所研究元素的含量。若在容矿岩石中适当地选定了ξ_0点，容矿岩石所特有的浓度值也可作为所要求的$C^{(0)}$值（杜鲍夫，1972）。

我们首先谈一下利用族T的函数来解释所研究的 晕的问题。给定近似三角多项式(6)的阶L和分散参数的任一值σ_0。根据分量函数

$$1,\ e^{-\frac{-(i\sigma_0)^2}{2}}\sin ix,\quad e^{-\frac{(i\sigma_0)^2}{2}}\cos ix,\quad i=\overline{1,\ L}$$

来分解函数$C(x)$(例如用最小二乘法)。根据公式(4)，该分解系数同时就是相应的系数$a_i(\sigma_0)$，$i=\overline{0,\ 2L}$，它肯定地决定着线性组合(6)中的近似多项式$C_0^*(\xi,\ \sigma_0)$。我们计算该近似多项式在点ξ_0上的值，并且假定$\xi_0\neq 0$，$\pm\pi/j$，$\pm\pi/2j$(这里$j=\overline{1,\ L}$)。一般情况下，这样的值与所给定的$C^{(0)}$值并不一致，所以所建立的多项式$C_0^*(\xi,\ \sigma_0)$不是拟求多项式。

我们仍然用$C_0^*(\xi,\ \sigma)$表示拟求多项式，用$b_i(\sigma)$表示它的系

数。

根据公式(8)，

$$C_0^*(\xi,\sigma)=a_0(\sigma_0)+\sum_{i=1}^{L}(a_i(\sigma_0)\sin i\xi+a_{i+L}(\sigma_0)\cos i\xi)\times \times e^{\frac{i^2(\sigma^2-\sigma_0^2)}{2}} \tag{12}$$

于是

$$C_0^*(\xi_0,\ \sigma)-C_0^*(\xi_0,\ \sigma_0)\equiv C^0-C_0^*(\xi_0,\ \sigma_0)=\sum_{i=1}^{L}(a(\sigma_0)\sin i\xi_0+a_{i+L}(\sigma_0)\cos i\xi_0)\cdot\left(e^{\frac{i^2(\sigma^2-\sigma_0^2)}{2}}-1\right) \tag{13}$$

后一个等式是关于未知的σ的方程式。在代入$e^{\sigma^2/2}=y$之后，个方程式就变成代数方程式

$$\sum_{i=1}^{L}A_i y^{i^2}-A_0=0 \tag{14}$$

这里

$$A_i=(a_i(\sigma_0)\sin i\xi_0+a_{i+L}(\sigma_0)\cos i\xi_0)\,e^{-\frac{(i\sigma_0)^2}{2}},\ i=\overline{1,\ L}$$

$$A_0=C^{(0)}-C_0^*(\xi_0,\sigma_0)+\sum_{i=1}^{L}(a_i(\sigma_0)\sin i\xi_0+a_{i+L}(\sigma_0)\cos i\xi_0)$$

我们确定它的根。这些根将不多于L^2个。按物理含意，$\sigma\geqslant 0$。考虑到这一点，我们从这些根中把大于1或等于1的根挑出来。假设它们是y_1，y_2，$y_3\cdots\cdots y_s$，$s\leqslant L^2$。现在我们就能求出拟求的分散参数：

$$\sigma=\sqrt{2\ln(y_j)},\quad j=\overline{1,\ S}$$

把实际上不现实和在物理学上不允许的σ和$C_0^*(\xi,\ \sigma)$值排除掉：例如我们已经知道，分散参数通常不会超过二、三十米，而在采样地段内被研究元素的浓度不应为负值，也不能超过100%。这就缩小了可允许的σ值的集合。在一般情况下，我们求得的公

式(12)的解的数目$n \leqslant L^2$，在这个公式中，右边的所有值都已确定了。在没有更多资料的情况下，这些解中的任何一个解都是最符合实情的，并具有同样的概率。

实际上，可接受的解的数目是不大的：根据近似准则，所取近似多项式$C_0^*(\xi,\sigma)$的参数数目不能比建立这个多项式所依据的浓度值的数目大。所以，由参数数目决定的L值是有限的。此外，根据解的平滑性要求，通常可以减少数目n。

浓度$C_0(\xi)$的拟求分布有可能不是三角多项式。由于这个原因，并且由于在用于解释的采样结果中存在着干扰，所以，在满足上述要求时，方程式(14)有可能不含有在ξ_0点上给出指定值$C^{(0)}$的解。这时，为了进行解释，可以选用在ξ_0点上给出与$C^{(0)}$值差别最小的浓度值的解。

在本专辑中P. И. 杜鲍夫关于等级近似的文章里，以及在杜鲍夫1974年的一本书里，都探讨了在更一般的情况下从各种可能的解中选择解的问题。

现在我们再来探讨利用多项式近似法解释分散晕的问题。和上述情况下一样，我们给出近似多项式(9)的某个幂L，以及分散参数的任一值σ_0。根据给定的函数$C(x)$我们求出决定它的分布$C_0^*(\xi,\ \sigma_0)$，并且为确定未知数σ建立方程式

$$\sum_{i=0}^{L} \xi_0^i (b_i(\sigma) - a_i(\sigma_0)) + C_0^*(\xi_0,\ \sigma_0) - C^{(0)} = 0 \qquad (15)$$

我们借助递推关系(11)，通过σ和已知值$a_i(\sigma_0)$及σ_0表示出(15)中未知的$b_i(\sigma)$。我们在这儿省略了中间的演算，只是写出方程(15)的变换结果：

$$\sum_{m=0}^{[L/2]} W_m \sigma^{2m} = 0 \qquad (16)$$

这里

$$W_m=\begin{cases}\sum\limits_{r=0}^{L}\xi_0^r(S_0^{(r)}-a_r(\sigma_0))+C_0^*(\xi_0,\ \sigma_0)-C^{(0)}, & m=0\\ \sum\limits_{r=0}^{L-2m}S_m^{(r)}\cdot\xi_0^r, & m\geqslant 1\end{cases}$$

$$S_n^{(r)}=\begin{cases}P_r,\ n=0\\ P_{r+2n}\sum\limits_{i=0}^{n}(-1)^iR_{n,i},\ 1\leqslant n\leqslant[\nu]-1\\ a_L\sum\limits_{i=0}^{\nu-1}(-1)^iK_i,\ (L-r)\text{—偶数}\\ a_{L-1}\sum\limits_{i=0}^{[\nu]-1}(-1)^ik_i,(L-r)\text{—非偶数}\end{cases}\Bigg\}n=[\nu]$$

$$K_i=\begin{cases}Q_{[\nu]},r,i=0\\ q_1(Q_{j1},r),i=1\\ q_i(T_{i,2}(T_{i,3}(\cdots(T_{i,i-1}(T_{i,i}(Q_{ji},r)))\cdots))),i>1\end{cases}$$

$$R_{n,i}=\begin{cases}Q_n,r,i=1\\ t_{n,2}(Q_{j_2},r),i=2\\ t_{n,i}(T_{i,3}(T_{i,4}(\cdots(T_{i,i-1}(T_{i,i}(Q_{j_i},r)))\cdots))),\ i>2\end{cases}$$

$T_{i,r},t_{n,i},q_i$——求和算子：

$$T_{i,r}(u)=\sum_{j_{s-1}}^{j_{s-1}}Q_{j_{s-1}-jr+1,r+2j_r+2i-2s}\cdot u$$

$$t_{n,i}(u)=\sum_{j_2=1}^{n-i+1}Q_{n-i-j_2+2,r+2j_2+2i-4}\cdot u$$

$$q_i(u)=\sum_{j_1=1}^{[\nu]-1}Q_{[\nu]-i-j_1+1,r+2j_1+2i-2}\cdot u$$

因此，$R_{n,i}$ 和 K_i 就是有限的嵌套和式串，其中，后一和式中每一和式的求和之上限值取决于前一和式求和的指标现行值

$$Q_{k,s}=\binom{2k+s}{s}(2k-1)!!,\ P_s=a_s+\sum_{k=1}^{[L-s/2]}$$

$$\times Q_{k,s}\cdot a_{2k+s}\cdot\sigma_0^{2k},\quad \nu=\frac{L-\gamma}{2}$$

我们确定方程(16)的正根。它们的个数不超过$[L/2]$。那么，$\sigma=\sqrt{y}$，这里，y是这些根中的任何一个根。排除掉在物理上不允许的σ值，并将留下的值代入递推关系式(11)，我们就能对它们中的每一个计算拟求的系数$b_i(\sigma)$。这样，在一般情况下，我们就得到在初始条件给定时方程(1)的$n\leqslant[L/2]$个形如(10)的解。

总之，知道了采样剖面某个点ξ_0上的拟求浓度值$C_0(\xi)$，在一般情况下，还不能避免解释的不肯定性。

但是，这里与上面指出的情况相类似，利用一定的限制来大大缩小方程(1)可能解的类，就可以从这些解的无限集合中选择出有限个解。

在本文之前的一些著作中（杜鲍夫，1972，1973），就曾使用拟求分布$C_0(\xi)$的抛物线近似法探讨过分散晕的解释问题。在那些著作里，积分方程(1)的解是用数值法导出的，因此选择所需要的σ值的工作，是通过在某些合理的范围内变换积分方程来完成的。接着，对每一σ值解反演题，而后，在具有已知值$C_0(\xi)=C^{(0)}$的点ξ_0上计算所得多项式$C_0^*(\xi,\ \sigma)$的值。一个σ^*值，只要满足$C_0^*(\xi,\ \sigma^*)\approx C^{(0)}$，就可以认为它是真实值，而与之相应的多项式$C_0^*(\xi,\sigma^*)$就可视为拟求的分布。因此，一般说只能查明可能解中的一个。利用上面所描述的原理，针对任一个$\sigma=\sigma_0$情况只需要解一次反演题，在这之后，所有可允许的σ值都可作为代数方程的根而确定出来，而要寻找与这些值相应的近似函数$C_0^*(\xi,\sigma)$的值，就不再需要解方程(1)了。

上述算法是通过适用于БЭСМ-4型电子计算机的算法程序实现的，并用许多实例作过试验，所得结果证明它是有效的。

最后要指出，采用双变量幂函数和三角函数系统Ψ_1和Ψ_2作为基，并且利用上述系统来研究方程(1)的解的肯定性的程序，通过用类似方法可以解释区域取样时观察到的分散晕。

参 考 文 献

Градштейн И. С., Рыжик И. М. Таблицы интегралов, сумм, рядов и произведений. М., Физматгиз, 1962. 1100 с.

Дубов Р. И. Концентрация химических элементов в ореолах рассеяния. —"Геол. и геофиз.", 1964, № 12, с. 44—45.

Дубов Р. И. Построение геохимических карт и разрезов с помощью электронно-вычислительных машин. —В кн.: Математические методы в геологии и геологическая информация. (Международный геологический конгресс, XXIV сессия, доклады советских геологов, проблема 16, симпозиумы 104, 105). М., "Наука", 1972, с. 80—88.

Дубов Р. И. Количественные исследования геохимических полей для поисков рудых месторождений. Новосибирск, "Наука", 1974. 278 с.

Соловов А. П. Основы теории и практики металлометрических съемок. Алма-Ата, Изд-во АН КазССР, 1959. 266 с.

Dubov R. I. A Statistical approach to the classification of geochemical anomalies. —In: Geochemical Exploration 1972. London, 1973, p. 275—284.

项仁杰译

利用最大似然法对地球化学异常进行分类的方法

H. M. 弗朗科夫斯卡娅

随着易发现矿床的数量减少，要查明矿床就要利用数量越来越多的受偶然因素影响的标志。在地球化学普查时需要对地球化学异常作评价。但是只根据一些成矿元素的浓度通常是不可能作到这一点的。将一些标型组合的化学元素用于地球化学异常的分类，就可以解决其中的某些问题（索洛沃夫、加拉宁，1972）。与此同时，要在多维标志空间中进行分类，也会出现一些特殊的困难。

在地球化学中广泛采用了判别分析。判别分析可归结为建立可以把某一观察客体 r 归入给定的集合（类）之一 R_1 或 R_2 中去的判别函数。判别函数的方程式应由似然比

$$D(x)=\frac{p(x|R_1)}{p(x|R_2)} \qquad (1)$$

求出，式中，$p(x|R_i)$——客体 $r\in R_i$ 的标志向量 X 的分布密度。最佳求解规则的要点是：计算出似然比，并把它与某一判别值 c 作比较。

在 $C=1$ 及先验概率 $P(R_1)$ 和 $P(R_2)$ 相等时，从几何意义说来，判别函数就是 q 维标志空间中的超曲面的方程式，导出该超曲面的方法是：要使超曲面的两侧各有一类客体的数目取极大值。

对由标志向量 x 描述的客体 r 进行分类，便归结为确定 x 与上述超曲面的相对位置。

有时候，如果考虑到量 $P(R_1)$ 和 $P(R_2)$ 会有差别，并且会有

把客体r错误地归入R_1类或R_2类的两种危险，那就要选择不等于1的C值。对于我们所研究的几种方法的对比来说，采用最简单的方案$C=1$就足够了。

对于标志集合X可用多维正态分布来描述的情况，判别分析研究得最为充分。这时，可利用线性判别函数和二次判别函数。在两个集合的协方差矩阵相等的条件下，建立线性判别函数。它表达为一个能最佳地把两个集合分割开来的超平面的方程式。两集合的交叉程度可由第一类和第二类误差来描述。

在协方差矩阵不相等的条件下，建立判别函数就有某些困难。在这种条件下，函数表达式复杂化了，所以客体分类只能用电子计算机进行。此外，估计分类的可能误差也成为十分复杂的问题。

在集合数目增到$N>2$时，就必须确定判别函数的数目，它等于C_N^2，即将N个集合两两组合的数目。众所周知，它在这时，以及在向量x中的标志数增加时，就产生了既要保持解的精确性又要缩减求解程序的问题，而要这样做，甚至在象正态分布这样简单分布的条件下，也会由于集合间界线的相互关系复杂而引起颇大困难。

根据判别函数的值把r归于R_i，就意味着被分类的客体r属于训练集合R_i中的一个。实际上还可能出现这样的情况：r并不属于这些集合中的那一类。在这种情况下，用判别分析就可能引出不正确的结论。

在杜鲍夫1971年的著作中提议，在地球化学异常分类时采用最大似然法，该法与判别分析相比有一系列优点。众所周知，最大似然法是把客体r归入量$p(x|R_i)$最大的R_i类。在直接依据概率密度值作客体分类时，不需要建立C_N^2个判别函数，只要把i值不同的概率密度进行N次比较就足够了。这时，虽然所利用的标志数增加，但并不降低其精度，并且与建立判别函数所进行的计算相比，计算是大大地简化了（特别当N很大时，更是如此）。许多作者得出结论，在协方差矩阵不相等时，为了对多维正态集

合作分类必须直接计算量 $p(x|R_i)$的比。他们把这种关系就看成是建立非线性判别函数（安杰尔先，1963；埃普什特因，1970），但实质上这就是似然原则，只是在形式上与制定判别分析前的似然原则相比稍为（和不适当地）复杂了一些：正如已经指出的，对于分类来说，量$p(x|R_i)$比它们的比值要更合适一些。

如果不仅根据概率密度 $p(x|R_i)$的比来对客体进行分类，而且也研究这些密度值本身及其分布，那么就可以获得某些补充资料。假设有两个类别 R_1 和 R_2，并给出了它们的多维概率分布密度 $\omega_1=p(x|R_1)$和 $\omega_2=p(x|R_2)$。我们把客体 $r\in R_1$ 和客体 $r\in R_2$ 的 ω_1 和 ω_2 值相应地投放到直角座标系的两个轴上去，就由原始标志的 q 维空间转换为概率密度的二维空间。这时候，训练集合中的每一个都可以用所建立的二维空间中点 x 的一定分布来表述。

把与新的拟分类客体的集合的向量 x 相应的点 (ω_1,ω_2) 标到这个空间中，就可看出它们的分布与训练集合的客体 r 所对应的点 (ω_1,ω_2) 是不同的（或相同的）。甚至在直接根据单个量 $p(x|R_i)$就可把拟分类客体的集合归入某个类别 R_i中去的时候，也能看出这种差别（例如，根据点分布区的延伸和其他形态细节）。由此可见，点 (ω_1,ω_2) 分布区的形态从总体上提供了有关客体集合的新信息。

二维向量的分布$W=\{\omega_1,\omega_2\}$可由下列联合分布函数给出：

$$F(W|R_i)=\int_{\omega_1}\int_{\omega_2} f(\omega_1,\omega_2|R_i)\cdot d\omega_1\omega_2 \qquad (2)$$

式中，$f(\omega_1,\omega_2|R_i)$——规定的概率密度，并且

$$\int_{\omega_1}\int_{\omega_2} f(\omega_1,\omega_2|R_i)d\omega_1 d\omega_2=1 \qquad (3)$$

显然，拟分类客体的量$F(W|R_i)$越大，它属于 R_i 类的概率就越大。所以，对每一类别R_i来说都可以选择一定的水平α_i，并且可以认为，如果任何一个点的 $F(W|R_i)$ 大于 α_i，那么这个点就属于这个类。相反情况下，该点就不属于训练类别R_i中的任何一

类。

如果在平面上取一条通过数值相等（$\omega_1=\omega_2$）的各点的直线作为类别间的界线，那么就可以估算出分类的可能误差：

$$P(2|1)=\int_{R_1} f(\omega_1,\omega_2|R_1)d\omega_1 d\omega_2$$

$$P(1|2)=\int_{R_2} f(\omega_1,\omega_2|R_2)d\omega_1 d\omega_2 \qquad (4)$$

式中，$P(2|1)$——将R_1中的客体归入 R_2的概率，$P(1|2)$——将R_2中的客体归入R_1的概率。

如果针对属于R_i类中某一类的向量X给出先验概率 $P(R_i)$，就可以根据贝叶斯（Байес）公式确定出后验概率：

$$P(R_i|x)=\frac{P(R_i)F(W|R_i)}{\sum_{j=1}^{2}P(R_j)F(W|R_j)} \qquad (5)$$

如果训练类别的数量$N>2$，原始q维标志空间就在概率密度的N维空间中反映出来。显然，对于$N=2$谈到的上述所有情况，在$N>2$时也还是正确的。

我们再研究对于 R_1 和 R_2两个训练类别其概率分布均可由多维正态律$N_p(\mu_1,\Sigma_1)$和$N_p(\mu_2,\Sigma_2)$来确定的情况，这里，μ_i，$i=1,2$，——第 i 个样本的平均值的向量；Σ_i，$i=1,2$，——第 i 个样本的协方差矩阵。多维正态分布的概率分布密度由下式确定：

$$p(x|R_i)=(2\pi)^{-q/2}|\Sigma_i|^{-1/2}\cdot \exp\left[-\frac{1}{2}(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i)\right] \qquad (6)$$

式中，$|\Sigma_i|$——矩阵Σ_i的行列式。

在$\ln p(x|R_i)$空间中进行分类是比较适宜的，如果取：

$$\omega_i=\ln\ p(x|R_i)=C-\frac{1}{2}(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i) \qquad (7)$$

式中，$C=-\frac{1}{2}(q\ \ln 2\pi+\ln|\Sigma_i|)$

我们再看一下ω_i值沿座标轴的分布。我们令：$V=\frac{1}{2}(x-\mu)'\cdot\Sigma^{-1}(x-\mu)$。

假设q维向量X是按$N_p(\mu,\Sigma)$律分布的。我们令$Z=x-\mu$，并假设向量Z是按$N_p(0,\Sigma)$律分布的。假设D为非退化矩阵，这就是$D\Sigma D'=I$，这里，I是单位矩阵。这时候，向量$Z^*=DZ$是按$N_p(0,1)$律分布的（安杰尔先，1963）。经过这些变换之后我们得：

$$V=\frac{1}{2}\cdot Z'\Sigma^{-1}Z=\frac{1}{2}\cdot Z^{*\prime}Z^*=\frac{1}{2}\sum_{i=1}^{q}Z_i^{*2}$$

即V服从于参数为$\frac{1}{2}q$的伽玛分布：

$$f(V,q)=\frac{1}{\Gamma\left(\frac{1}{2}q\right)}[V^{1/2(q-1)}e^{-V}] \qquad (8)$$

相应的，量$\omega=C-V$的分布密度可定义如下：

$$f(\omega,q)=\frac{1}{\Gamma\left(\frac{1}{2}q\right)}[(C-\omega)^{1/2(q-1)}e^{-(C-\omega)}] \qquad (9)$$

我们来研究量$V=\frac{1}{2}(x-\mu_1)'\times\Sigma_1^{-1}(x-\mu_1)$的分布，这里，向量$X$是按$N_p(\mu_2,\Sigma_2)$律分布的。令$Z=x-\mu_1$。这时，向量$Z$就按$N_p(v,\Sigma_2)$律分布，这里，$v=\mu_2-\mu_1$：

$$V=\frac{1}{2}Z'\Sigma_1^{-1}Z \qquad (10)$$

设非退化矩阵D是符合$D\Sigma_2D'=I$的条件的。我们定义：

$$Z^*=DZ,\quad \Sigma_1^*=D\Sigma_1D',\quad V^*=DV \qquad (11)$$

这里，Z^*具有$N_p(v^*,I)$分布（安杰尔先,1963），所以$V=\frac{1}{2}\cdot Z^{*\prime}\Sigma_1^{*-1}Z^*$。我们选择一个第一行为$q_{1i}=\frac{Z_i^*}{\sqrt{Z^{*\prime}Z^*}}$的$q\times q$阶正交矩阵。设$u=GZ^*$，$B=G\Sigma_1^*G$。这时，基于正交矩阵的定义，我们得（安杰尔先，1963）：

$$U_1=\sum_{i=1}^{q}q_{1i},\ Z_1^*=\sqrt{Z^{*\prime}Z^*},\ u_{j\neq 1}=0$$

$$V=\frac{1}{2}u'B^{-1}u=\frac{1}{2}(u_1\,0\cdots 0)\begin{pmatrix}b^{11}b^{12}\cdots b^{1q}\\ b^{21}b^{22}\cdots b^{2q}\\ \\ b^{q1}b^{q2}\cdots b^{qq}\end{pmatrix}\begin{pmatrix}u^1\\0\\\vdots\\0\end{pmatrix}=\frac{1}{2}u_1^2b^{11}$$

(12)

这里，$(b^{ij})=B^{-1}$。由此可得，$V=\frac{1}{2}u_1^2b^{11}=\frac{b^{11}}{2}Z^{*1}Z^*$。

前已指出，向量Z^*是按$N_p(\nu^*,\ I)$律分布的，所以量$Z^{*1}Z^*$具有非中心x^2分布（安杰尔先，1963），其非中心参数为$\tau^2=\nu^{*1}\nu^*=\nu^1\Sigma_2^{-1}\nu$。量$V^*$的分布密度由下式确定：

$$f(V^*)=\frac{1}{2^{\frac{1}{2}q}\sqrt{\pi}}e^{-\frac{1}{2}(\tau^2+V^*)}V^{*\left(-\frac{1}{2}q-1\right)}\sum_{\beta=0}^{\infty}\times$$

$$\times\frac{(\tau^2)^\beta(V^*)^\beta\Gamma\left(\beta+\frac{1}{2}\right)}{(2\beta)!\Gamma\left(\frac{1}{2}q+\beta\right)}\qquad(13)$$

令$b^*=\frac{1}{b^{11}}$时，我们得$V^*=2b^*V$。把这个表达式代入公式(13)，我们就得到量V的分布律的表达式：

$$f(V)=\frac{(b^*)^{\frac{1}{2}q}}{\sqrt{\pi}}e^{-\frac{1}{2}(\tau^2+2b^*V)}V^{\frac{1}{2}q-1}$$

$$\cdot\sum_{\beta=0}^{\infty}\frac{(\tau^2\cdot 2b^*\cdot V)^\beta\cdot\Gamma\left(\beta+\frac{1}{2}\right)}{(2\beta)!\Gamma\left(\frac{1}{2}q+\beta\right)}\qquad(14)$$

利用纯虚自变量的第一类贝塞尔函数，便可把这个公式变换成：

$$f(V)=b^{*}e^{-\frac{1}{2}(\tau^{2}+2b^{*}V)}\left(\frac{2b^{*}V}{\tau^{2}}\right)^{\frac{q}{4}-\frac{1}{2}}I_{\frac{q}{2}-1}(\sqrt{\tau^{2}2b^{*}V}) \tag{15}$$

式中，$I_{\frac{q}{2}-1}(\sqrt{\tau^{2}2b^{*}V})$——纯虚自变量的贝塞尔函数。如果把表达式$V=C-\omega$代入公式（15），我们就得到量$\omega=\ln p(x|R)$的分布律。显然，这一分布律与（15）式分布律的不同在于，分布密度的曲线向C值的方向移动。

总之，对于正态多维分布$N_p(\mu_1,\Sigma_1)$和$N_p(\mu_2,\Sigma_2)$所给定的R_1和R_2两类来说，在$x\in R_1$时，在点(ω_1,ω_2)的空间内ω_1的座标按公式（9）分布，ω_2的座标按公式（15）分布；在$x\in R_2$时，ω_1的座标按公式（15）分布，ω_2的座标按公式（9）分布。

可以认为，ω_1和ω_2的座标的分布是彼此无关的。所以，$f(\omega_1,\omega_2)$的二维分布可以定义为伽玛分布与非中心分布的乘积。由此可见，是可以用公式(2)、(4)和(5)来计算分类的可能误差的。

在直接利用量$p(x|R_i)$或$\omega_i=\ln p(x|R_i)$（而不是量$p(x|R_i)$的比值）进行分类时，有可能用直观的图解法来解问题。为此，与上述相对应，在$\omega_1\omega_2$的直角座标系中可反映出训练类别R_1和R_2，以及拟分类客体。这样就可以判断类别的相互位置、类别的交错域，并可估计分类的概率误差。图解法有可能用简单的作图来代替一部分费力的运算。

显然，如果任一拟分类客体落入某一训练类别的点的密集域内，那么在原始标志x的空间内它就属于这一类。如果它落在ω_1和ω_2的低值域内，那么它属于某一类的概率就十分小，这个客体多半不属于所给定的训练类。正如已经指出的，如果考虑到与拟分类样本相应的点大部分都被归入随便哪一类内，那么分类的错误就会减少。例如，对于图1上类别R_3的样本来说，可以看出，它属于类别R_1和R_2的概率很小，以至于不能很有把握地将这个样本归入训练资料的某一类别中去。然而，假定根据判别分析或只是根据量$p(x|R_1)$和$p(x|R_2)$的比值对这个样本进行分类，它

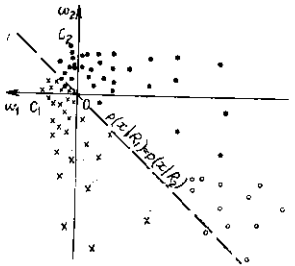


图 1　R_1和R_2的训练样本的图示以及在 $\omega_i=\ln p(x|R_i)$ 空间中对 R_3 进行判别的略图

（×——R_1；●——R_2；○——R_3）

就会被划入类别R_2。

这个方法曾通过许多例子作过检验。最初的例子是石油客体，后来的例子是金属矿客体。果然不出所料，就方法而论结果是类似的。

图 2 表明了在哈萨克斯坦根据甲烷、氮、铀、铜、铅、铬、镍、硼、锶的浓度平均值以及氧化还原电位 Eh 的平均值，利用判别分析查明油气区（构造）的情况。根据在热特拜含气构造（R_1）上打的18个钻孔以及在不含油气的沙拉拜-塞讷克构造（R_2）上打的同样数量的钻孔，取了上述各量作为训练资料。用来作判别的有乌什库云、南塔斯布拉特、塔斯布拉特、科库姆拜、科克舍等构造。

在图 2 的横座标上标出了线型判别函数的L'值，它是在用带有相同协方差矩阵的正态律对 $p(x|R_i)$ 的分布进行近似的条件下，按标准算法计算得的。沿着纵座标是针对训练资料的构造求得的 L'估计值的频率 p，以直方图的形式表示出来。横座标轴下面的点表示针对拟分类构造的不同钻孔求出的L'值。如果与该钻孔相应的点落在值$L'=0$的左边，那么就判定它为沙拉拜-塞讷克型构造，如果与该钻孔相应的点落在值 $L'=0$ 的右边，那么就判定它为热特拜型构造。

图 3 表示直接根据量 $p(x|R_i)$ 对钻孔进行分类的结果。$p(x|R_i)$ 的分布也用正态律来近似，但不受协方差矩阵相等的限制。

在图 3 表示的空间内，位于座标轴夹角平分线上的任一个点，都具有$p(x|R_1)=p(x|R_2)$，也就是说它们属于 R_1和R_2的概率是相同的。位于 $\ln p(x|R_2)$轴附近的点，$p(x|R_1)$值大于$p(x|R_2)$值，也就是说它们属于第一类，反之亦然。正如我们所见到

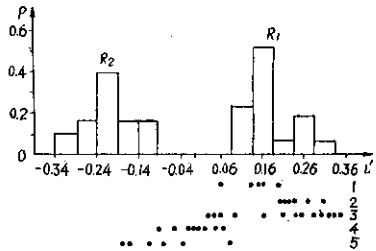


图 2　利用判别分析对构造分类

构造：1—乌什库云；2—南塔斯布拉特；3—塔斯布拉特；4—科库姆拜；5—科克舍

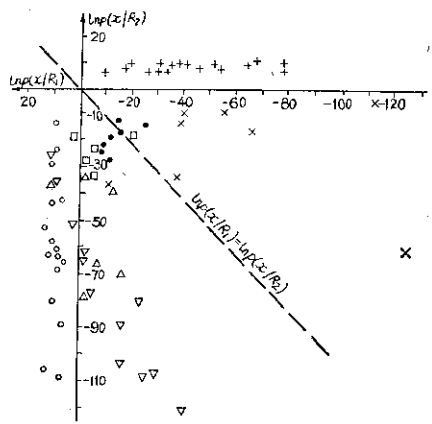


图 3　用最大似然法对南曼格什拉克的构造进行的分类。

构造的钻孔：□—乌什库云；△—南塔斯布拉特；▽—塔斯布拉特；●—科库姆拜；×—科克舍；○—热特拜；+—沙拉拜-塞讷克。

的，根据判别分析（见图2）属于 R_1 和 R_2 的概率大致相同的科库姆拜构造，现在可比较有把握地归为 R_1。科克舍构造被分散在 $\ln p(x|R_i)$ 值较低的空间内的点表示出来，因此可以推测，它所属的类在训练资料中是没有的。

现在，这些判别方法已推广使用，在使用中，人们往往把客体 r 归入其标志在某种意义上与它很相近的类别中去，而没有考虑不同类别的标志交错的概率和可能性。这里，结果取决于标志空间中距离的确定。这种“戈洛齐普”（Голотип）程序用上述训练资料作过检验（《地质学和数学》，1970）。塔斯布拉特构造上的13个钻孔中只有2个被判为属于油气构造，其余的钻孔都落在程序的拒绝分类区内。有关其他构造的资料也是未被分类的。

上述资料证明，对前述类型的实际问题来说，最大似然法是最好的方法。

这个方法也被用来查明哈萨克斯坦塔尔迪库尔干州科克赛铜—钼矿床上部中段、中部中段和下部中段的原生晕的差别。

细脉浸染型的科克赛矿床分布于科克赛向斜中，向斜核部产有加里东期花岗岩类，周围是志留纪的岩层。矿化产于科克赛花岗闪长岩侵入体中，该侵入体的热液蚀变岩广泛发育。人们还观察到，沿科克赛断裂，矿体在垂向上有大幅度地位移，位于下降盘中的矿体至今还没有为钻探所发现。

在某些勘探剖面上研究了内生晕。对Ⅱ号剖面作了最为完整的取样，根据它构成了训练资料。矿带的上部中段（60—190米）用69个样品来代表；中部中段（220—355米）——51个样品；下部中段（420—520米）——29个样品。

问题是在12个标志的空间中解决的，这12个标志是：Zn、Sr、Bi、Co、Pb、Ag、As、Sn、Ba、Hg、Mo和Cu的浓度。

最初利用判别分析取得了下列结果。上部中段与下部中段有十分明显的差异，分类的平均误差为23%，在区分上部中段和中部中段时，误差等于40%。区分中部中段和下部中段的误差为27%。

在运用最大似然法时分类误差减小了。图 4 表示的是对该矿床的矿带上部中段和下部中段的区分。$p(x|R_i)$ 分布也是用多维正态律来近似的。和上述这类插图一样，在图 4 所表示的空间内，位于座标轴角平分线上的任一个点，其$p(x|R_1)=p(x|R_2)$，也就是说属于R_1和R_2的概率是一样的。位于$\ln p(x|R_2)$轴附近的点，$p(x|R_1)>p(x|R_2)$，也就是说属于第一类。位于 $\ln p(x|R_1)$ 轴附近的点，情况正相反。这样一来，在区分矿带的上部中段和下部中段时，只有 3 个上部中段的点被错分为下部中段的（误差

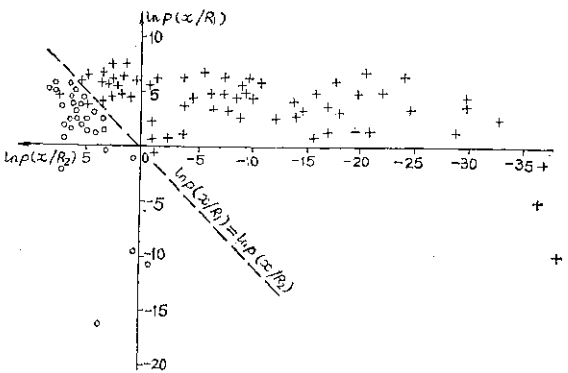


图 4　科克赛矿床（剖面Ⅱ）的上部中段和下部中段的 $\ln p(x|R_i)$ 值的分布

+—上部中段，○—下部中段

4%），下部中段的点没有一个被归到上部中段中去。

在区分上部中段和中部中段时，平均误差的估计值为15%。

沿剖面Ⅳ还提出了几个已被钻孔揭露的含矿截面来进行分类。这时，在大约比剖面Ⅱ中低200米的位置上取的样品，被归入了反映上部和中部中段的类别中。在这个基础上作出结论：剖面Ⅳ相对于剖面Ⅱ向下移动了200米的距离。地质学家分析了各个化学元素的含量和地球化学指数，也得出了这一结论（索洛沃夫、加拉宁，1972）。

用上述方法解决的又一个问题就是对米罗诺夫铋矿区的次生晕进行分类。取米罗诺夫矿床的组合次生晕和没有远景的19号异常作为训练资料。位于米罗诺夫矿床一侧并具有一定意义的1号异常被选来进行判别。

用78个样品来代表米罗诺夫矿床；用60个样品来代表19号异常；用43个样品来代表1号异常。分类是在10个标志的空间中进行的，这10个标志是：Mo、Be、W、Cu、Zn、Bi、Cr、Nb、Ag和Pb的浓度。

我们对用最大似然法、判别分析，以及根据“戈洛齐普”程序（《地质学和数学》，1970）所得的分类结果进行了比较。

判别分析的结果表示在图5上。沿横座标标出了线型判别函数的L'值，它是在用具有相同协方差矩阵的正态律对 $p(x|R_i)$ 的

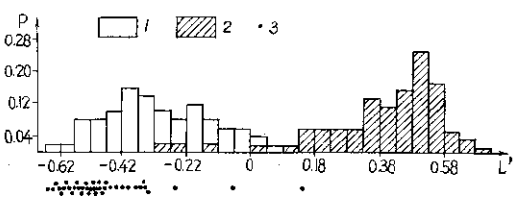


图 5　利用判别分析对次生晕进行分类

1—19号异常样品分布的直方图；2—米罗诺夫矿床样品分布的直方图；3—1号异常中的样品

分布进行近似的情况下，根据标准算法计算出来的。沿纵座标用直方图表示出针对训练资料求得的判别函数值L'的估计值的频率。横座标轴下面的点表示拟分类的1号异常的L'值。如果与该异常相应的点落在值$L'=0$的左边，那么它就被认为是属于米罗诺夫矿床型的；如果与该异常相应的点落在值$L'=0$的右边，那么它就被认为是属于19号异常型的。这种方式的分类误差估计值为7%。

图6表示了直接根据$\ln p(x|R_i)$值对1号异常进行分类的结果。代表米罗诺夫矿床的78个样品中有5个落在19号异常的区域内（误差为6%），同时，19号异常的点没有1个被归为米罗诺夫矿床的。

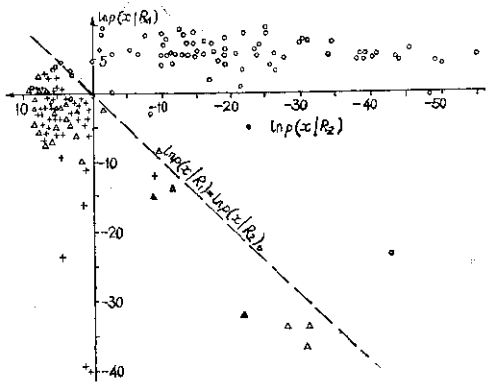


图 6　用最大似然法对次生晕进行分类

⊙—米罗诺夫矿床；＋—19号异常；△—1号异常

19号异常的大部分点分布在它自己的点的密集域内，只有个别例外，它们沿着等分线散布，并具有比较低的$\ln p(x|R_i)$值。所以，可以作出结论：在所给定的标志空间中，1号异常与19号

异常有颇大的交错域，但它同时又是与训练资料所代表的类别有所不同的总体。

根据“戈洛齐普”程序，所有具有 $\ln p(x|R_i)$ 低值的点，以及其他许多点都落在拒绝分类的区域中。

总之，上边引用的所有资料都证明，对所描述的各类实际问题来说，最大似然法是最好的方法。应当指出，用转换出量 $p(R_i|x)$ 的估计值和考虑不同解的好处改进此方法，将会进一步提高它的效果（杜鲍夫，1971）。

参 考 文 献

Андерсен Т. В. Введение в многомерный статистический анализ. М., Физматгиз, 1963. 500 с.

Бугаец А. Н. Статистические методы принятия решений при поисках и оценке пегматитов по геохимическим данным. М., “Недра”, 1970, 128 с.

Геология и математика. Новосибирск, “Наука”, 1970. 224 с.

Дубов Р. И. О применении принципа максимального правдоподобия для классификации геохимических аномалий. —В кн.: Ежегодник—1970 СибГЕОХИ. Иркутск, 1971, с. 362—366.

Соловов А. П., Гаранин А. В. Геохимические спектры и установление различий между сходными объектами. —В кн.: Литохимичские поиски рудных месторождений. Алма-Ата, “Наука”, 1972, с. 148—164.

Эпштейн Е. Н. Нелинейная решающая Функция. —В кн.: Геология и математика. Новосибирск, “Наука”, 1970, с. 42—64.

项仁杰译

矿体及矿体晕分带性经验模型的建立

H.A.基塔耶夫 B.H.叶夫道基莫娃 B.H.楚马金

问题的提出

在矿产普查勘探工作实践中，地球化学工作日益广泛的运用碰上了要对大量资料进行处理的问题。因此，很明显，必须研究以利用电子计算机为基础的数学处理方法，并且建立拥有一些用来解不同问题的标准程序的处理系统。在这个系统中特别应包括以地球化学场图件的形式自动输出结果的程序。

这些图件为地质工作者提供了重要的资料，因为，它们所表示出来的不同元素浓度的空间布局决定着地球化学分带、矿物分带，以及热液蚀变岩石和晕的分带和结构。大体上说，对地球化学场的定量描述就是建立一个以被研究化学元素的浓度及其他特征的相互关系为依据的含矿地质体的经验模型。

考虑到上述情况，我们研究出用特殊方式定量处理地质-地球化学信息的数学原理，并在电子计算机上予以实现。这种处理的最终成果是获得地球化学场的单元素图和多元素图（化学元素组合图）。该方法的实质如下：

就化学元素的浓度而言，地质体是不均匀的，研究它们就需要查明不均匀的程度、划分出相对均匀的地段，并且研究不均匀性产生的原因。因此，对所研究地质空间的定量描述自然可看作是这样一个任务，即在确定各元素组合的相似和相异程度的同时划分和圈出以化学元素的不同组合为其特征的场。特别是不同种类的热液蚀变岩石都有不同的元素组合。所以，划分这些组合与划分不同的热液蚀变带是一致的，也会反映出现象的成因本质。

由此可见，将所研究的地球化学场以组合（也就是最常见的共生化学元素的集合）的形式表示出来，是建立地球化学场经验模型的主要问题之一。

划分组合可以通过自动分类来进行，即把取样点的集合划分或若干个均匀的组（类），随后对它们进行分析。在m维特征空间内对点作分类时，与点在取样区内的座标（x,y）无关。

分类的问题具有解集❶。须由其中挑选出一些符合下列两项准则的解：1，属于同一类的点在特征空间中应该彼此相近；2，在划分出来的每一类中，都应该存在一种主要的元素比例关系（均一性标准），也就是说在每一类中都应该是一种元素组合。

把在化学元素分布场上所观测的经验资料用一个nxm阶的矩阵$\{Z_{ij}\}$表示出来是适宜的，这里，Z_{ij}为化学元素的百分含量（浓度）；n为矩阵的行数，它等于研究区内的取样点数；m是列数，等于被分析元素的数目。

我们将矩阵的每一行都看作是m维特征空间中的一个点-向量。

样品中各种化学元素的百分含量是不可相互对比的量。所以，在解决所提出来的问题时不是用绝对浓度，而是用它们与背景值的比——衬度系数（KK）。

衬度系数是无量纲的量，因此，对于具有不同绝对含量水平的各种化学元素来说，该系数是可以彼此对比的。同时，它们也是岩石的重要地球化学特征。

划分元素集合主要是根据元素的衬度系数值，同时要考虑它们的方差和不同元素衬度系数之间的相关性。

如果存在着一种主要的元素比例关系，类的主要元素已经确定，并且已给出了它们的数字特征，如平均浓度、方差、衬度系数等，那么这个类（组合）就可认为是已被划分出的类。只要所划分出的类还没有满足解的最终准则，就还可进一步划分。这个

❶ 分类的结果不是唯一的，由所有可能的结果组成的集合称为解集——审者注

最终准则就是下列要求：同一类中的各元素的剩余方差不应超过某个事先给定的（足够小的）值。引用这个准则就可以根据经验资料（在此必须这样做）建立谱系结构的分类系统。

衡量所划分组合的准则是：1，m维特征空间中属于同一组合或相邻组合的点，在取样地质空间中的分布也应该是彼此邻近的。2，在圈划不同组合的场时，它们在地质空间中的位置应按照其地质-构造特点来圈定。

在必要时，在编制多元素图（组合图）之前应先编单元素图（更准确地说是单元素场的图）。这时，为了取得取样点之间的点上的场值，便进行内插。

按这种思路提出问题，已研究出并运用了一种能自动构建矿田经验模型的方法（叶弗多基莫娃，基塔耶夫，1973；基塔耶夫叶弗多基莫娃、丘马金、西克连，1974）。这种构建分为几个阶段：分析经验资料，圈划地球化学场（元素组合场）和描述所划分的组合。

经验资料的分析

在下面叙述的算法中，可根据样品中元素衬度系数的相近性自动地将经验资料划分成事先不知其数目的K个类。如果在一个类中已确定的主要元素比例关系说明了事先确定的该类中元素总方差的百分数，那么这个类就被认为是可划分出来的均匀的类。要给定可成功地进行分类的方差限，势必要考虑到诸因素的总和，这些因素是：矿田地质构成的特点、物质分析的精度、异常中浓度的方差，等等。

主分量法可用于分类的目的，该法在文献中也称作“卡鲁纳-洛艾夫分解”（разложение Карунена-Лоэва），或在随机函数理论中称作特征正交分量分解（拉奥，1968；瓦塔纳贝，1959）。在这种情况下，不是直接利用给定的观测值的集合，而是这些集合的最佳近似参数。为了避免成因实质不同的各类的数据掺混，

主分量法只针对每一个类分别予以使用。

B. И. 叶尔金娜和И. М. 扎戈鲁衣科1968年提出来的用超球面作分类的算法是最为适用的，下面就是该文作者提出来的计算格式。

我们来研究特征空间中的n个点组成的集合A：$A=\{A_1, A_2, \cdots\cdots, A_i, \cdots\cdots, A_n\}$。每个点$A_i$都由$m$个座标的征集$(Z_{i1}, \cdots\cdots, Z_{ij}\cdots\cdots, Z_{im})$所表征，也就是说，每个点$A_i$都是$m$维特征空间中的一个向量，这里，$i$——向量(样品)号；$j$——所研究的每个化学元素的编号。所研究的点集必须划分成若干个(K个)类。

我们求点A_i的整个集合的重心$C_0(Yo_1, Yo_2, \cdots\cdots, Yo_j, \cdots\cdots, Yo_m)$。它的座标是

$$Yo_j=\frac{1}{n}\sum_{i=1}^{n} Z_{ij}, \quad j=\overline{1,m}$$ ❶

我们计算由点C_0到所有A_i点($i=\overline{1,n}$)的距离。这些距离中的最大者被取作为以点C_0为中心的超球面的半径R_0：

$$R_0=\max_i R(A_0, A_i)=\max_i\left\{\sqrt{\sum_{j=1}^{m}(Y_{0j}-Z_{ij})^2}\right\}$$ ❷

所有的点$A_i, i=\overline{1,n}$都落在这个超球面中。

我们给出某个正值ΔR，并计算新的半径$R_1=R_0-\Delta R$。我们再用半径R_1在点$A_1(Z_{11}, Z_{12}, \cdots\cdots Z_{1j}, \cdots\cdots Z_{1m})$的周围作一个超球面$S_{R_1}(A_1)$。取最靠$C_0$点的点作为$A_1$点是合适的，因为在$C_0$点附近的区域是背景域。

---

❶ $j=\overline{1,m}$相当于$j=1,2,\cdots\cdots m$——审者注

❷ 原文根号内的$\sum_{j=1}^{m}$误写为$\sum_{i=1}^{m}$，据文意改正。原文类似错误有多处，不一一注明——审者注

我们将会划分出落在超球面 $S_{R_1}(A_i)$ 内的点集$B^{(0)}$，$B^{(0)}=A_i:R(A_1,A_i)\leqslant R_1$。我们再计算集$B^{(0)}$的重心$C_1=(Y_{11},Y_{12},\cdots\cdots,Y_{1j},\cdots\cdots Y_{1m})$，这里：

$$Y_{1j}=\frac{1}{n_1}\sum_{i=1}^{n_1}Z_{ij},\quad j=\overline{1,m}$$

式中：n_1——集$B^{(0)}$的点数。

我们用同样的半径R_1由C_1点作超球面$S_{R1}(C_1)$，并且划出集$B^{(1)}$：$B^{(1)}=\{A_i:R(C_1,A_i)\leqslant R_1\}$。

以此类推，我们再求出集$B^{(1)}$的重心$C_2(Y_{21},Y_{22},\cdots\cdots,Y_{2j},\cdots\cdots,Y_{2m})$，并分出集$B^{(2)}$。

把超球面$S_{R_1}(C_i)$的重心C_i移到重心C_{i+1}的这种移位程序反复进行，一直到重心C_i和C_{i+1}的座标重合为止。落在这个"终止"的超球面内的点的集合，我们以Q^1类标记之。

用类似的方式，在选择点 A_2(除点A_1之外下一个最靠近C_0的点）作为起始点后，我们便可得到Q^2类，如此等等。我们把Q^α作为表示这些类中任何一类的一般符号，这里$\alpha=\overline{1,K}$。我们把这样一个类中的任意一个向量-点标记为$A_p^\alpha(Z_{p1}^\alpha,Z_{p2}^\alpha,\cdots\cdots,Z_{pj}^\alpha,\cdots\cdots,Z_{pm}^\alpha)$，这里$p=\overline{1,n_\alpha}$，$n_\alpha<n$，$n_\alpha$——落在$Q^\alpha$类中的集合$A$的点数。

利用主分量法可以找到这样的一些随机系数$T_1^1\cdots\cdots,T_m^1$，以及非随机向量

$$U_1(U_{11},U_{12},\cdots\cdots,U_{1j},\cdots\cdots,U_{1m})$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$U_i(U_{i1},U_{i2},\cdots\cdots,U_{ij},\cdots\cdots,U_{im}$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$U_h(U_{h1},U_{h2},\cdots\cdots U_{hj},\cdots\cdots,U_{hm})$$

它们应满足下面条件：

$$\sum_{i=1}^{n_a}\sum_{j=1}^{m}\left(Z_{ij}^{a}-\sum_{l=1}^{h}T_i^l u_{lj}\right)^2=\min,\ (h\leqslant m)$$

向量u_1，……，u_h是下列协方差矩阵

$$\{\sigma_{j,j+k}^{a}\},\ k=\overline{1,m}\ ,j=\overline{1,m};\ \ \sigma_{j,\ j+k}^{a}=\sum_{i=1}^{n_a}\frac{Z_{ij}^{a}\cdot Z_{i,j+k}^{a}}{n_a}$$

的特征向量。

协方差矩阵$\{\sigma_{ij}^{a}\}$的特征值，经过相对于它们的和：$\rho_1^{(a)}$，……，$\rho_l^{(a)}$……$\rho_h^{(a)}$规格化之后，就可作为每个向量分量的显著性的量度，或向量分量的概率量度。

我们把满足下列条件看作是分类成功的准则：对每个类来说都应该有一个主要的元素比例关系，或有一个这样的特征向量u_l，其条件是对该向量来说：

$$C_2\leqslant\rho_l^{(a)}\leqslant C_1 \qquad (1)$$

这里，C_1，C_2——事先给定的接近于1的数。

在满足条件（1）时，第j个座标（因素）的概率量度$\rho_j^{(k)}$按下列公式计算

$$\rho_j^{(k)}=\frac{u_{lj}}{\sum_{i=1}^{m}u_{li}},\quad j=\overline{1,m}$$

如果

$$\rho_{\gamma_L}^{(k)}>C_3 \qquad (2)$$

其中C_3也是事先给定的数，$0<C_3\leqslant 1$，那么我们就可以说类是按座标γ_L来划分的，并且这个类就具有这个座标的名称。如果对所有这些座标（因素）γ_1，γ_2，……γ_n来说条件（2）都得以满足，那么就可根据座标来划分类。

由此可见，要认为一个类是均匀的，其主要条件就是存在有一个在所有剩下的特征值❶中最大的特征值。或者，同样地，我

❶ 指所有满足(2)的特征值——审者注

们把满足下列要求看作是分类成功的准则：对每一类来说都应该有一个主要的比例关系，或者存在着一个满足于（1）式的特征向量u_l。

如果在所划分出的类中，条件（1）没有得到满足，那么就要改变用来划分这个类的半径R_l，并且，分类程序还要反复进行，直至条件（1）得到满足为止。

对剩下的点进行类的划分和描述，其程序与上类同。

我们来确定分类终结的条件。正如上面所指出的，所考虑的点的经验集合$\{Z_{il}\}$是在m维特征空间中研究的。但是，特征（元素）的重要性是不一样的。衬度系数高或其变化大的元素是最有意义的。在所提出的程序中高衬度元素起主要作用。为了查明低衬度元素的作用，问题应该再解一次。

在所得的每一类$Q^\alpha(\alpha=\overline{1,K})$中，在不用这一类的主座标（因素）$\gamma_1,\gamma_2\cdots\cdots,\gamma_g$的情况下，用上面叙述过的同样方法进行再分类，也即划分Q^α类的亚类。需要进行这种划分的条件是，除了座标$\gamma_1,\gamma_2,\cdots\cdots\gamma_g$之外，还有一些与其对应的元素的平均值大大超过背景含量的座标。

建立经验模型的实例

按照上面描述的方法编绘地球化学场，就可以从反映运移化学元素的相互关系和由矿田形成过程造成的元素在地质空间中的分布规律的角度对地球化学客体进行研究。因此，本文的作者之一曾提议把早已熟知的“化学元素组合”的概念引入晕的论题中（基塔耶夫、楚马金，1973；基塔耶夫、叶夫道基莫娃、楚马金、西克连，1974）。该方法也可以在定量基础上查明晕的成因类型，这些成因类型既可以是由矿田形成的多阶段性造成的，也可以是由于存在不同的矿化类型而产生的（塔乌桑，1974；基塔耶夫、叶夫道基莫娃、楚马金、西克连，1974）。所有这一切都为十分详细而明显地查明矿田和矿体的分带性提供了可能性。此

外，根据组合的定量和定性参数可以区分矿致异常与非矿异常。

在处理西伯利亚的金、锡、钨、多金属、稀有金属和伟晶岩矿床的原生晕、次生晕和分散流的资料时运用了这个方法。下面就举几个建立经验模型的实例。

作者不打算在本篇详细描述矿田的成因和地质构造。在所举的第一个例子中将比较详细地描述一下所研究方法的实质和细节。

伊林金矿床

早在1876年就已知道的伊林矿床，其成因最初是由B.A.奥勃鲁契夫（1918）确定的，他认为它是火山口相火山岩中的低温热液矿床。这个观点直到目前还有人支持（弗格尔曼，1964）。与此同时，还有其他一些不同的观点，如认为矿体所在的角砾岩是与火山机体无关的构造角砾岩（库兹明、舒宾，1966）。根据我们踏勘工作取得的初步资料来看，矿床是不规则的似网脉状体，富含金，分布在硫化物矿化的破碎带中。这些网脉体是碎裂花岗岩类及其带有浸染状硫化物矿化（主要是黄铁矿，其次是毒砂）的角砾岩。在剖面上部见到一些小的石英-电气石细脉。金呈游离状态存在于长石和石英颗粒中，或者分散在硫化物里。

图1和图2是金矿山露天采场一侧的展开图，沿着近地表的和地表以下50米、80米的三个截面采了样。样品作了快速的（近似定量的）光谱和伦琴光谱硅酸盐分析。

借助于电子计算机按衬度系数的谱系划分了五种元素：金、砷、银、硼和铋，这五种元素是这里观察到的地球化学晕所特有的。衬度系数相对比较小（1—3）的是锰和锂。所研究的其余元素（W、Mo、Cu、Zn、Ni、Co、Li、Sn）只见有弱的异常浓度。图1是得到的多元素图的一个例子。图1上所表示出来的组合的参数列在附表中。

我们专门阐述一下“组合中衬度系数百分比”这一栏。实际

上就是它决定了组合的划分和命名：在这一栏中显示出了所分析元素衬度系数的“权”，并且要根据它来确定谱系。元素的“权”，就是以百分比表示出来的所研究元素的衬度系数在谱系里所有元素衬度系数总和中占的比率。

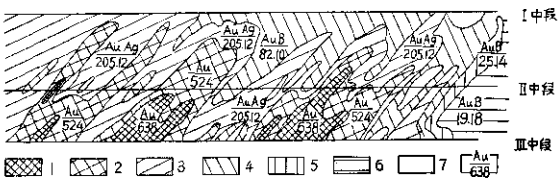


图 1　多元素地球化学场。伊林矿床，金矿山的露天采场（剖面）

1—7—附表里所列组合的顺序号；8—组合中的主要元素及其衬度系数

衬度系数大于 1 的所有元素决定着它们参与带入作用，而衬度系数小于 1（含量低于背景值）的那些元素在矿田形成时以带出作用为特征。

根据谱系中衬度系数的百分比（“权”），可相对地划分出：主要元素（KK❶ >20）、次要元素（KK为5—20）和混入元素（KK <5）。组合名称由衬度系数最大的二到三个元素来决定，而且命名时先标出衬度系数相对较小的元素，后标出相对较大的元素。

在伊林矿床的剖面中（见图 1），金矿柱可由一系列元素确定下来，这些元素是Au、As、Pb、Ag、W、Bi、Mo、B、Mn、Li，它们的衬度系数分别为628、28、11、8、7、5、4、4、3、2（四舍五入成整数）。这里，按衬度系数来说起主要作用的是金，它的“权”等于99%。

因此，元素的这种结合构成了一个组合，命名为“金”组合。与金共存的有其他一些衬度系数较高的元素，这证明它们富集的地质和物理化学条件有共同性。因此，上述这些元素可认为

❶ KK为衬度系数（Коэффициенты контрастности）的缩写，下同。——译者注

伊林矿床成矿元素组合的参数

组合号	样品数	组合中元素的平均含量，单位10^{-3}%									
		Au	Ag	As	W	Mo	Pb	Li	Bi	Mn	B
1	11	0.35	0.003	30	0.3	0.2	2	1	0.4	40	10
2	14	0.26	0.006	18	0.14	0.2	1	1	0.7	30	50
3	7	0.11	0.02	6	0.05	0.3	2	3	0.3	34	10
4	10	0.04	0.003	3	0.05	0.3	0.7	2	0.3	26	10
5	5	0.013	0.003	3	0.05	0.3	1	2	0.3	8	18
6	28	0.005	0.003	1	0.05	0.1	0.5	4	0.08	19	15
7	9	0.04	0.001	5	1	0.4	1	7	0.7	9	267

组合号	样品数	KK 的 平 均 值									
		Au	Ag	As	W	Mo	Pb	Li	Bi	Mn	B
1	11	628	8	23	7	4	11	2.3	5	3	4
2	14	524	12	35	3	4	13	2.3	14	3	25
3	7	205	35	11	1	7	16	4	7	4	5
4	10	82	6	7	1	6	7	8	7	2.8	10
5	5	25	7	3	1	6	1	2.7	6	0.8	14
6	28	19	6	3	1	3	5	14	2	2	18
7	9	81	2	10	2	9	12	7	14	1	134

组合号	样品数	组合中KK的百分比						组　合
		Au	Ag	As	Bi	B	Pb	
1	11	99	—	—	—	—	—	Au
2	14	81	2	6	2	3	2	Au
3	7	66	11	4	2	2	5	Au,Ag
4	10	59	4	5	4	7	5	Au,B
5	5	35	8	9	8	14	1.4	Au,B,As
6	28	21	9	5	3	15	1	Au,B
7	9	29	0.7	3	5	45	4	B,Au,Bi

是伴生元素，它们的偏高浓度也是金矿化的普查标志。

2号组合所代表的带围绕着前一个带产出。在这个组合中金的衬度系数（524）比1号组合中的（628）要稍低一些，而“权”

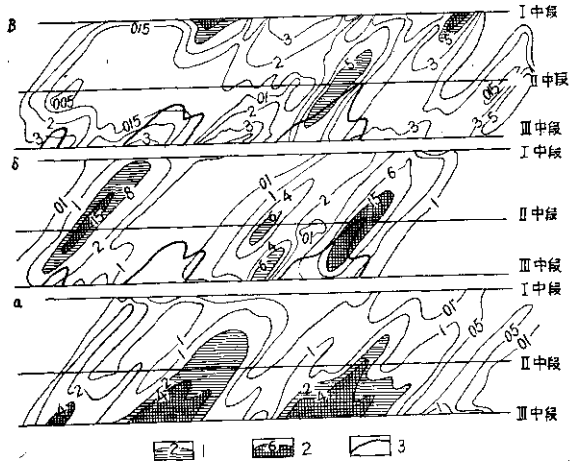


图 2 金（а）、硼（б）、锰（в）的原生晕

含量以假定单位表示，其单位分别为10^{-3}%和10^{-0}%。伊林矿床，金矿山的露天采场

1、2—浓度偏高地段及其浓度值；3—在硼和锰的浓度图上标出的金浓度为4个和4个以上假定单位的轮廓线

也由99%减到81%。在金的“权”减小的同时，砷、硼、银和铋的衬度系数却增大，它们的“权”相应为6、3、2和2。由于金和其他元素在“权”上仍有明显的差别，所以该组合的名称也和第一个组合一样，称为“金”组合。

3号组合围绕着前一个带，其特点是金的衬度系数更低，而银的衬度系数相对较高，这为把它命名为“银-金”组合提供了依据。其他组合如此类推。第7号组合无论就空间位置，还是就衬度系数来说都占有特殊的地位。这里，金、砷、铅、钼、铋，尤其是硼的衬度系数都大大地高于上述各带。在空间上，这个组合（铋-金-硼）与含有石英-电气石细脉的角砾岩带是重合的。

在图1上可明显看出，随着往上离最大矿化地段的距离加大，一个组合依次被另一个组合所取代。我们发现，砷的最大衬度系数分布在金矿地段附近，但是按衬度百分比（按“权”）来看，它占据的位置要更高一些。在表里所列的其他参数中，对于确定分带序列来说，衬度系数百分比是个主要参数。在所举的例子中，得出下列序列：Au—(Ag,As)—(Li,B,Mn)。

为了查明矿体及其晕中各种元素分布的细节，常常编制地球化学场的单元素图，虽然在许多情况下单元素图是必要的，但它们并不能取代多元素图。

金、硼和锰的单元素图可用图2作为一个例子。正如我们所见，金的最大浓度产于三个透镜状的地段，而这三个地段趋向于产在中部中段，并特别趋向于产在下部中段，这证明矿化往深部继续延伸。硼的偏高含量在金的最大浓度地段之上呈“火炬”状分布，有时候也由这个地段往四面延伸一定距离。锰晕多半在离金矿柱较远处形成偏高浓度（同时衬度系数较低），但是在某些情况下，在富含金的地段附近也见到该元素浓度增高。看来，这是由于在所研究矿田的不同部位存在有不同的含锰矿物所造成的。

柳巴文—哈佛京金矿结

矿结分布在达乌尔中生代拱形隆起的南部。柳巴文深断裂是主要的控矿构造，它控制着一系列酸性岩的板状岩株和岩墙，也控制着金矿化。在这里，为了预测深部和侧翼的矿化，以及为了评价该矿结总的含矿远景，进行了综合性的科研—生产地球化学调查。问题是通过面积性的分散流和原生晕研究来解决的。

在大约500平方公里的面积上以常规方式进行了分散流的研究：沿所有的河流和溪流按100米间隔对淤泥-砾石粒级的河流沉积物进行取样。

分散流的所有特征元素按衬度系数划分成两类。第一类元素

(Au、Ag、As) 是主要的，其衬度系数达 100—300。第二类元素 (Bi、W、Mo、Sn、Pb、Cu、Zn) 仅超过背景值三倍，少数为 5—10 倍。Au、Ag、As、Bi 这样的元素组合表征着与金矿脉和矿床直接有关的分散流；而W、Mo、Sn、Pb，有时还有 Ag 和 As 这样的元素组合表征着与整个矿田有关的分散流。每一类金矿化都特有其具有一定特征的分散流，应该强调一下一个非常重要的特征元素—Bi 的意义。由于铋只是在富的金矿石中才大量存在，所以，它在分散流中的出现证明在近处就存在着原生的金矿床。

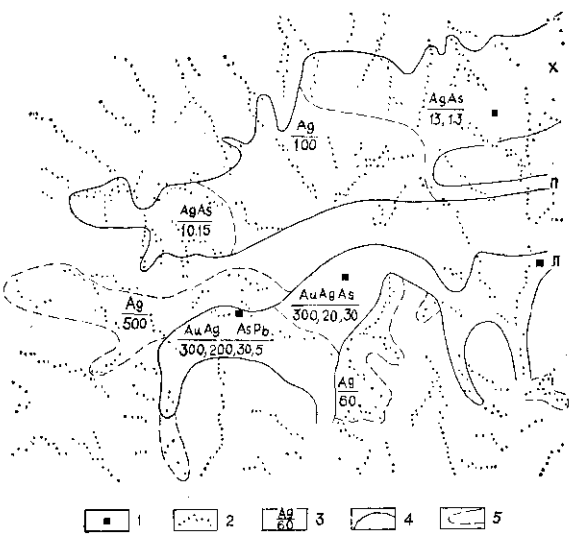


图 3　根据分散流取样资料作的多元素地球化学场图

(据B. A. 科列斯尼科夫和H. A. 基塔耶夫的资料。柳巴文—哈佛京矿结(白图)。

1—矿床；2—取样点；3—组合中的主要元素及其衬度系数；4—矿带；X—哈佛京矿带；II—中间矿带；Л—柳巴文矿带；5—不同的化学元素组合所占据的范围

图 3 表示出了柳巴文—哈佛京矿结的多元素地球化学场。

该矿结中的柳巴文矿带，正如图 3 上所见到的，主要由砷—金组合反映出来，而且随着往西银的衬度系数增高，并出现铅。这两个元素正好是矿田西翼的矿石所特有的成份。

哈佛京矿带由银（中部）和砷（侧翼）的较高的衬度系数反映出来。

为了确定柳巴文矿带深部层位的远景，对尼古拉耶夫、耶夫格拉福夫和大费多罗夫等矿床的原生晕进行了研究。这些矿床是金—石英建造的典型矿床，其围岩是砂页岩，较少情况下为侵入体。总的来说，矿田具有断块构造，断块由西向东单斜状跌落。由于这个原因，矿田西部被侵蚀切割得更深一些。位于矿田中部的耶夫格拉福夫矿床的原生晕，以及该矿床的远景，在许多著作（塔乌桑，1974；基塔耶夫、楚马金，1973；基塔耶夫、叶夫道基莫娃、楚马金、西克连，1974）中早就指出来了。这里只是简要地探讨一下尼古拉耶夫和大费多罗夫矿床，前者位于矿田的东部，后者位于矿田的西部。

尼古拉耶夫矿床的原生晕（图4，*a*）沿着垂直矿体的剖面在四个中段上作了研究。晕的结构显示出它与矿化的特点有明显的关系。正如在图4，*a*中所见到的，由矿柱中部往四面八方发生地球化学组合的交替：由金的组合到汞—金组合、金—汞组合，最后为汞的组合。

在大费多罗夫矿床，所研究的剖面（图4，*б*）垂直穿过两条矿脉。在地表附近矿脉倾斜平缓，往下倾角增大。金的矿化强度由地表往下部中段逐渐减弱。元素组合在原生晕中的分布与尼古拉耶夫矿床一样也是有规律的。随着向四面远离矿体，可以看到先是砷-银-金组合，而后是汞-金-银组合，最后变为银-金-汞组合。在垂直方向上沿着矿体的倾斜也发生类似的组合更替。与尼古拉耶夫矿床不同的是，在大费多罗夫矿床的原生晕中，实际上所有元素的衬度系数都相当高。应该强调的是，由矿柱往各个方向，包括在它们的下部（沿着倾斜），汞的作用都是增大的（就组合中

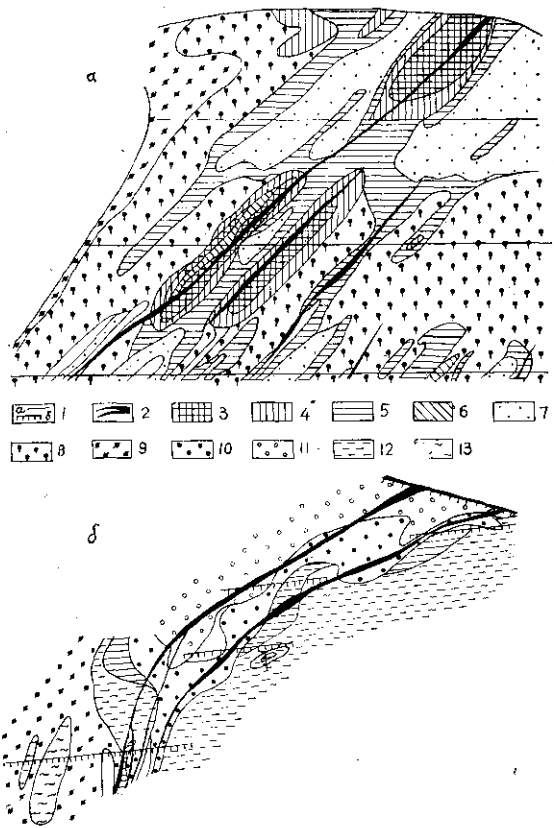


图 4　多元素地球化学场

а—尼古拉耶夫矿床（剖面）；*б*—大费多罗夫矿床（剖面）
1—取样的中段；2—矿体。元素组合：3—金组合；4—砷-金组合；5—砷-汞-金组合；6—汞-金-砷组合；7—钼-金-汞组合；8—砷-金-汞组合；9—汞组合；10—汞-金-锑组合；11—砷-银-汞组合；12—银-金-汞组合；13—硼-汞组合

各元素衬度系数的百分比而言)。这两种情况证明，所研究的矿体（和矿柱）位于厚大的热液蚀变岩石带中，看来蚀变带由已被矿山坑道揭露的矿体往下延伸得还相当深。并且不排除这里有可能存在着新的矿柱。

科帕耶夫铁矿床

矿床产于穿切近于水平产出的沉积岩层的凝灰爆破筒中。根据Л. Г.斯特拉霍夫和B.E.波查罗夫（1972）的资料，经常呈柱状的陡倾矿体在水平断面上一般具有五个带（由中心向外）：裂隙脉型矿石、交代矿石、无矿或弱矿化的砂卡岩、中-低温交代岩、弱热液蚀变岩石。但只是在极少数的情况下，才能在一个矿床中同时划分出这五个带。

矿体中分布最广的矿物是：磁铁矿、赤铁矿、细晶磷灰石、黑镁铁锰矿、鲁道尼特矿（Рудонит）；在硫化物中出现有：磁黄铁矿、黄铁矿、黄铜矿、闪锌矿、方铅矿、辉钴矿；脉石矿物是：斜长石、石英、绿帘石、角闪石、绿泥石、方解石，少数情况下有萤石、菱锶矿和重晶石。Л.Г.斯特拉霍夫和B.E.波查罗夫根据方解石的爆裂测温分析查明了线性变化的温度垂直分带，其梯度为每加深100米温度增高4°到11°。H.B.巴甫洛夫在1961年、Л.Г.斯特拉霍夫和B.K.瓦西里耶夫在1974年查明，镁磁铁矿的类质同象成份随着矿石形成的深度而改变：在覆盖的玄武岩中出现镁铁矿；在0.4公里深处形成的中温矿石中，镁磁铁矿里的镁分子占80%，在1.6—1.8公里深处形成的矿石中，镁磁铁矿里的镁分子则降到5%以下。

在查明矿石的原生分带时，除了矿物成分外，地球化学分带也是很有意义的，但它的资料却很少。根据Б.З.李特文的取样资料用电子计算机建立经验模型时所得的结果，大概在某种程度上能弥补这个缺点（图5）。

首先看出，可划分出两类地球化学组合（图5中粗线限定的

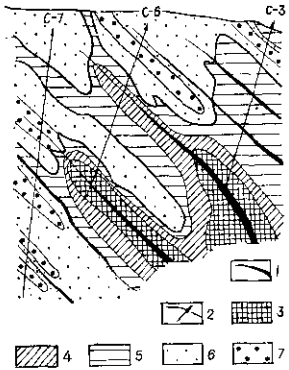


图 5　多元素地球化学场。
科帕耶夫矿床

（剖面，据Б.З.李特文的资料）
1—交代矿体；2—钻孔，元素组合；3—铅-铜-铁组合；4—铬-铜-铁组合；5—铜-铬 铁组合；6—铬-铁-汞组合；7—铬-钡-锶组合

部分）。在谱系中以铁和铜的衬度系数居优势的第一类组合（No1—4）将矿体圈了出来；以汞、锶和钡的衬度系数居优势的第二类组合（No5—10）主要分布在矿体上部之上。下面我们稍为详细地谈一谈这两类组合的结构。

第一个组合（铜—铁组合）与谱系中各元素的下列顺序相对应：铁、铅、锶、汞、铬、钴、镍和钡。头一个元素铁和最后一个元素钡的衬度系数相应为121和4。已确定出来的其余元素——As、Ag、B、Be、Sn、Zn、Mo、Mn、V、Ge、Ti的衬度系数较低，它们不起多大作用。第二个组合包围着第一个组合，分布于无矿带之中，其特点是衬度系数的谱系具有同样的结构，但所有元素的衬度系数都较低。在第三个组合中，除了铬的衬度系数增大外，几乎所有元素的衬度系数都在减少，在谱系中铬的衬度系数占第三位；此外，还见到钴的衬度系数稍有增大的趋势。在第四个组合中，铁、铜和钴的衬度系数进一步减小，铬和汞的衬度系数则增大。

第二类计有六个组合（锶-钡-汞）。它们分布在剖面的上部，这里，碳酸盐化带强烈发育。锶、汞，其次还有钡在衬度系数的谱系中占前几位。铜-锶组合好象是上两类组合之间的过渡组合：在该组合中，衬度系数谱系中的前二个位置分别由第一类组合的元素铜和第二类组合的元素锶所占据。由图5可以看出，这个组合在空间中也占据着过渡位置。

总之，这两类组合的空间位置证明，随着远离矿体的中心部位，成矿元素逐渐发生量变，因此就反映出下列的地球化学分带：铁-铜-铅-铬-钴-锶-汞。

所查明的沉淀物的地球化学分带证明这类矿床是热液成因的，看来，它也反映了上述矿物最大数量所在地带的分布特点。在矿体和近矿空间，主要是含铁和铜的矿物，而在边部则主要是钴、锶、钡和汞的载体矿物。

上述实例说明多维分析法适用于处理在不同类型金属矿床上进行的各种地球化学工作所得的地球化学资料。根据已提出来的并在生产中得到部分运用的多维分析法用电子计算机建立的地球化学场表明，该法用来处理地球化学资料是有效的。

参 考 文 献

Ватанабе С. Разложение Карунена-Лоэва и факторный анализ. —В кн.: Автоматический анализ сложных изображений. М., "Мир", 1959, с. 254—275.

Евдокимова В. Н., Китаев Н. А. Обработка данных геохимических съемок на основе анализа рудных полей. —В кн.: Геохимические поиски месторождений в Сибири и на Дальнем Востоке по вторичным ореолам рассеяния. Иркутск, 1973, с. 212—225.

Распознование образов в социальных исследованиях. Новосибирск, "Наука"1968, 195 с.

Китаев Н. А., Евдокимова В. Н., Чумакин В. Н., Сыхлен С. Е. Классификация геохимических свойств геологических объектов в связи с изучением их зональности и генезиса. —"Геол. и геофиз.", 1974, № 6, с. 135—139.

Китаев Н. А., Чумакин В. Н. Ореолы ассоциаций химических элементов Любавинского рудного поля.

—В кн.: Ежегодник-1972 СибГЕОХИ. Иркутск, 1973, с. 143—147.

Кузьмин А. М., Шубин Г. В. Илинское месторождение(Восточное Забайкалье).Томск, Изд-во Томск. ун-та, 1966. 90 с.

Обручев В. А. Илинское золоторудное месторождение в Восточном Забайкалье. —"Материалы по геологи России", 1918, вып. 26, с. 195—287

Рао С. Р. Линейные статистические методы и их применение. М., "Наука", 1968, с. 530—533.

Страхов Л. Г., Бочаров В. Е. Температурная зональность железорудных месторождений Ангаро-Илимского типа. — "Геология рудных месторождений", 1972, № 2, с. 110—113.

Таусон Л. В. Закономерность образования эндогенных ореолов рассеяния в полях рудных месторождений. —"Геол. и геофиз. ", 1974, № 5, с. 113—123.

*Фогельман Н.А.*Эксплозивно-инъекционные золотоносные брекчии Илимского месторождения в Забайкалье. —"Бюл. МОИП. Отд. геол.", 1964, т. 39, вып. 2, с. 90—99.

项仁杰译

主分量分析法在解释地球化学资料中的应用

A. A. 施曼斯基　Л. П. 普洛特尼科夫

在普查和勘探矿床时，广泛地应用地球化学和地球物理方法，使得每个观测点的地质信息容量迅速扩大。通常可以把每个观测点所有数据的总体作为阿几里德多维标志空间中的一个点来对待，因而可使用多元统计分析法进行数据的处理。

随着标志空间维数的增加，研究由这些标志描述的对象也就愈加困难。于是，产生了在不损失重要信息的情况下，如何以少数变量代替许多观测标志的问题。在这方面，主分量分析法值得注意。其主要优点如下：

1．它是唯一能够保证解的客观性的通用计算方法。

2．该方法能保证得到最合理的观测标志组合，而且每个标志都具有与其信息量和与其它特征的配合程度相应的贡献。

3．所得到的标志组合乃是一些数目较少且相互独立的新合成变量。这些变量的总方差与原始数据的方差相等，也就是说，它们实际上载有离于大量原始数据中的全部有用信息。因此，可分别对每个新变量的信息量进行估计，并可确定它们与各独立自然（地质）因素的对应关系。

为转换出新变量，通常要利用协方差（相关）矩阵的内部结构特征。

按照主分量分析法，要针对第 j 个主分量 γ_j 求出一些不相关的、其方差已极大化的规格化线性组合：

$$\gamma_j=\sum_{i=1}^{k}W_{ji}x_i \qquad j,i=1,2,\cdots,k \qquad (1)$$

式中，W_{ji}表示第j个分量在第i个观测变量x上的因子权。k表示协方差（相关）矩阵被分解出的正交分量的数目。

从几何上看，求主分量的过程就是要实行向新的正交坐标系的变换：要使与第一个坐标轴相应的线性形式的方差构成总方差的最大部分，与第二个坐标轴（与第一个正交）相应的方差尽可能多地包含总方差的剩余部分，第三个坐标轴亦按类似要求确定。

考虑到变换的正交性，根据下式❶的关系就可将新坐标恢复成旧坐标：

$$x_i=\sum_{j=1}^{k}W_{ji}\gamma_j \qquad (2)$$

此式为主分量分析法的基本关系式，其中，k个主分量囊括了原始变量的全部方差。

在有现代计算技术的情况下，要实现主分量分析法是比较简单的，只采用两个计算步骤即可：

1．建立协方差矩阵S。为此，$m\times p$维的观测值矩阵X（其中m为p个标志的观测次数）左乘以它的转置矩阵X'：

$$S=X'X \qquad (3)$$

2．用旋转法（雅可比法）求协方差矩阵的特征值和特征向量，结果得：

$$S=W'\Lambda W$$

式中W为$p\times p$维矩阵，它的列为矩阵S的特征向量：

$$W=\begin{pmatrix}W_{11} & W_{12}\cdots & W_{1p}\\ W_{21} & W_{22}\cdots & W_{2p}\\ \cdots\cdots & \cdots\cdots & \cdots\\ W_{p1} & W_{p2}\cdots & W_{pp}\end{pmatrix} \qquad (5)$$

矩阵中的W_{ji}表示第j个分量在第i个观测变量中的因子权；Λ表

---

❶ 原文误为$x_i=\sum_{j,i=1}^{k}W_{ji}\gamma_j$——审者注。

示一个$p\times p$维的矩阵，其主对角线是矩阵S的特征值：

$$\Lambda=\begin{pmatrix}\lambda_1^2 & 0 & \cdots & 0\\ 0 & \lambda_2^2 & 0 & 0\\ \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & 0 & \lambda_p^2\end{pmatrix} \qquad (6)$$

$\lambda_1^2>\lambda_2^2>\cdots>\lambda_p^2$。

$\lambda_1^2+\lambda_2^2+\cdots+\lambda_p^2=S_p^2$是所有观测变量的总方差。

第一个主分量是组合

$$\gamma_1=W_{11}x_1+W_{21}x_2+\cdots+W_{p1}x_p \qquad (7)$$

其方差为λ_1^2。第二个主分量相应地是组合

$$\gamma_2=W_{12}x_1+W_{22}x_2+\cdots+W_{p2}x_p \qquad (8)$$

其方差等于λ_2^2。其余主分量依此类推。

由于正交性质，主分量之间是不相关的。因此人们把用主分量代替观测变量称为将观测变量向正交分量的分解。

如果前面k个主分量的方差之和构成了总方差的主要部分，则k个（$<p$）主分量的数据就可代表所有被研究的元素。

在将矩阵W的列按相应的特征值规格化（第j列的每个元素平方后再乘以λ_j^2）后，就得出方差分解矩阵：

$$F=\begin{pmatrix}f_{11} & f_{12} & \cdots & f_{1p}\\ f_{21} & f_{22} & \cdots & f_{2p}\\ \cdots & \cdots & \cdots & \cdots\\ f_{p1} & f_{p2} & \cdots & f_{pp}\end{pmatrix} \qquad (9)$$

如果认为λ_j^2是由第j个因子作用于所有标志的值而造成的方差，则f_{ij}应看作第j个因子对第i个元素值造成的方差。在解释因子时，所探讨的是前面k个列的矩阵元素的绝对值和符号。

下面我们举一个在地球化学资料解释中应用主分量分析法的实例。

在东西伯利亚的一个稀有金属伟晶岩矿床中查明，许多元素的原生分散晕都是内部结构复杂，界线不清晰。基岩中锡、铍和稀有碱金属浓度的分布特征如图1所示。要确定各个元素原生晕

的界线是极其困难的。这样便产生了一个问题：为了查明异常，应如何表示地球化学取样的数据呢？

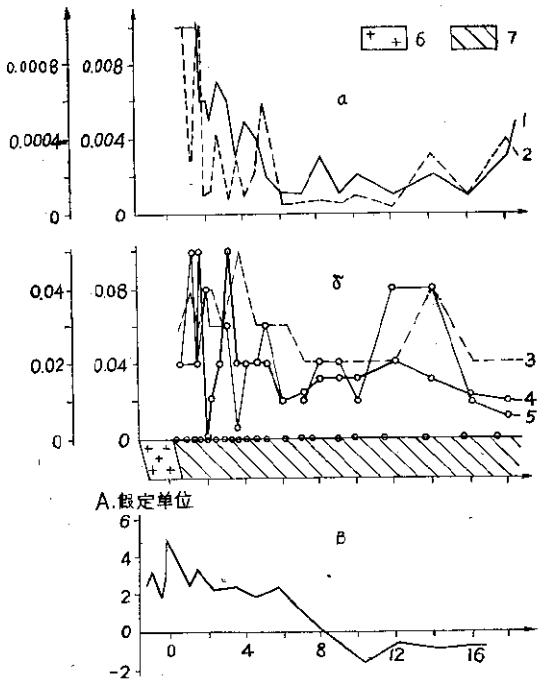


图 1　在伟晶岩的原生分散晕中铍、锡、稀有碱金属和第一主分量的分布特征

а—铍和锡的分布特征（左纵坐标轴表示铍含量，%；右纵坐标轴表示锡含量，%）；*б*—稀有碱金属的分布特征（左纵坐标表示铷、铯含量，%；右纵坐标轴表示锂含量，%）；*в*—第一主分量的分布特征。1—锡；2—铍；3—铷；4—锂；5—铯；6—伟晶岩；7—黑云片岩（围岩）；圆圈表示采样位置（沿水平坐标轴，以米为单位）

表 1　相关分析和分量分析的结果

标　志	相关系数					因子权
	锂	铷	铯	锡	铍	
离开伟晶岩的距离	−0.70	−0.80	−0.67	−0.70	−0.48	−0.81
元素含量						
锂	1.00	0.70	0.60	0.55	0.65	0.71
铷		1.00	0.73	0.61	0.45	0.79
铯			1.00	0.41	0.41	0.62
锡				1.00	0.49	0.59
铍					1.00	0.48
因子方差，%						67.00

注：相关系数和因子权不等于零，当其绝对值 $r \geqslant 0.32$ 时，置信度为 0.95

在表 1 中列出了针对第 1 主分量所做的相关分析和分量分析的结果。

通过相关关系的研究，可推测距离矿体的远近对围岩样品中上述所有化学元素的含量有重要影响。通过分量分析查明了五个主分量，其中只有第一分量是重要的，其方差占总方差的67%。第一主分量与离伟晶岩的远近成负相关，而与所研究元素的浓度成正相关。这个分量反映原生晕中稀有金属矿化指示元素分布的总特征，因而可以用它成功地划分原生分散晕（见图 1）。

让我们再来研究一个同时应用上分量分析和趋势分析方法的实例，这种做法在解释地球化学测量结果时是富有成效的。

在某一地区，进行了次生分散晕地球化学测量（据伊尔库次克地质局资料），沿剖面系统地采了 208 个样品，分析了十二个元素。

所研究地段是由片麻岩区和一个大花岗岩侵入体组成的（图 2,*a*）。建立单个元素趋势面的尝试没有成功，因为由趋势面反映出的系统分量的比率太小（图2,*б*）。所作的分量分析的结果列于表 2 。

第一主分量占总方差的40%。所有系数（因子权）都是正的，

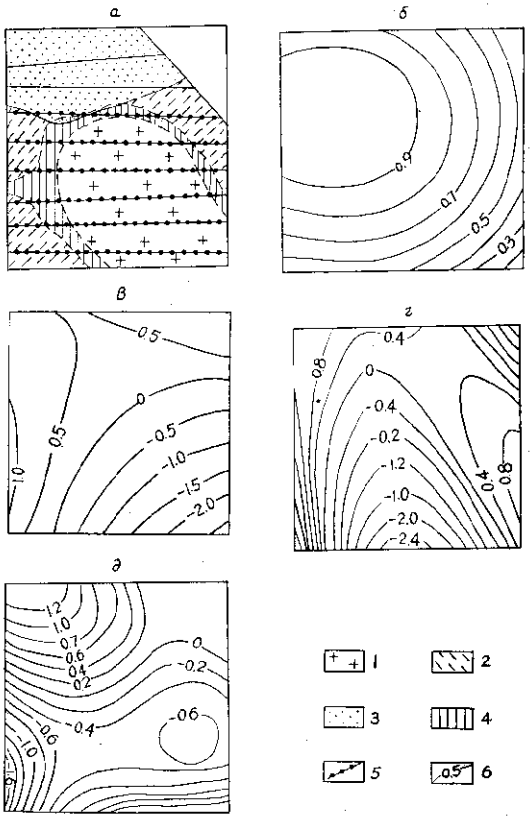


图 2　钼和前三个主分量的趋势图

а—研究地段的地质略图；*б*—钼浓度的趋势分析结果（单位为 10^{-3}%）；*в*、*г*和*д*—分别为第一、第二和第三主分量的趋势分析结果(假定单位)。1—花岗岩；2—退混合岩化黑云母片麻岩；3—混合岩化黑云母片麻岩，夹栉榴石黑云母片麻岩层；4—混合岩化黑云母片麻岩和角闪石-石英岩石；5—地球化学采样位置；6—钼浓度等值线

表 2　分量分析结果

元　素	因　子　权		
	第　一	第　二	第　三
钛	0.351	−0.207	−0.031
钒	0.319	−0.188	−0.431
铬	0.232	0.461	−0.182
锰	0.314	0.016	0.413
钴	0.356	0.107	−0.226
镍	0.363	0.148	−0.255
铜	0.256	0.428	−0,124
锌	0.258	0.200	0.342
钼	0.228	−0.215	0.346
锡	0.263	−0.365	0.158
钡	0.314	−0.090	0.387
铅	0.116	−0.514	−0.355
因子方差,%	40.0	13.0	9.5

并且与零值相差甚大。第一主分量的地球化学含义可确定为疏松沉积物被上述十二种元素混染的程度。根据这个分量绘制的趋势面，反映了所有元素的区域分布情况。例如，花岗岩侵入体中元素含量比片麻岩发育带中低很多（图 2，*a*）。

第二主分量占总方差的13%。它反映两个元素组合的性状。第一个元素组合包括铬、铜、锌、镍、钴，第二个元素组合包括铅、锡、钼、钛、钒、钡。根据这个主分量绘制的趋势面表明，在片麻岩区第一组元素的浓度偏高，而在花岗岩中第二组元素的浓度偏高（图 2，*b*）。

第三个主分量只占总方差的9.5%，它是两个元素组合，它们在地球化学测区内的分布明显地受东西向断裂构造控制。因而主分量的零值线实际上与断裂构造相吻合（图 2，*d*）。

由上可见，前三个主分量占了总方差的62.5%，故可代表由十二个元素构成的整个总体。

第一主分量反映该地段的一般规律，即金属元素在片麻岩中

的含量比花岗岩中的高。第二和第三主分量可以看作是对一般规律的详细表述，考虑到该地段的地质构造特点，可更精确地反映出一般规律。

计算主分量的算法是在БЭСМ-4型电子计算机上依程序实现的。该程序能压缩信息、划分系统分量、过滤观测变量和进行方差分析。

参考文献

Андерсен Т. Введение в многомерный статистический анализ. М., Физматгиз, 1963. 500 с.

Белонин М. Д., Татаринов И. В., Калинин О. М., Шиманский В. К., Бескровная О. В., Гранский В. В., Похитонова Г. Е. Факторный анализ в нефтяной геологии. (обзор). М., изд. ВИЭМС, 1971. 56с.

Бондаренко В. Н., Верховская Л. А., Сорокина Е. П. Применение многомерных статистических методов для изучения структуры геохимических полей. —В кн.; Математические методы при геохимических исследованиях. (Тезисы к семинару "Методы математичесеой обработки результатов геохимических поисков"). Свердловск, Уральское территориальное геологическое управление, 1973, с. 23—25.

Дуденко Л. Н. Использование теории игр и факторного анализа при прогнозе рудных месторождений. —В кн.; Применение ЭВМ в геологии и в горном деле. Вып. 1. Л., 1971, с. 43—54.

Загоскин В. А., Шиманский А. А. Строение первичных ореолов рассеяния редкометальных пегматитов натро-литиевого типа. —В кн.; Геохимия пегматитов Восточной Сибири. М., "Наука", 1971, с.

195—206.

Крамбейн У., Грейбилл Ф. Статистические модели в геологии. М., "Мир", 1969. 397 с.

Доули Д., Максвелл А. Факторный анализ как статистический метод. М., "Мир", 1967. 144 с.

Миллер Р., Кан Дж. Статистический анализ в геологических науках. М., "Мир", 1965. 482 с.

Налимов В. В. Теория эксперимента.М., "Наука", 1971. 208 с.

Плотников Л. П. Корректировка многомерных наблюдений. —В кн.: Тезисы докладов конференции ИПИ. Иркутск, 1971, с. 49—50.

Шиманский А. А. Многомерные модели. —В кн.: Каждан А. Б. Методологические основы разведки полезных ископаемых. М., "Недра", 1974, с. 137—143.

陈昭华　程金柱译

地球化学和地球物理资料的综合解释

Г. С. 瓦赫罗麦耶夫　А. Ю. 达维坚科

应用地质、地球物理和地球化学综合方法进行找矿，需要研究综合解释所得结果的专门原理。至少可以指出两个解决这个问题的途径。传统的微分方法就是分别逐步分析每一种方法（甚至化学元素）的数据，而后将结果归纳起来。积分方法则截然不同，它以信息解释理论原理为依据（哈尔芬，1958）。

地球物理场和化学元素分散场定量评价方法的数目迅猛增加，电子计算机的引进更是推波助澜，遗憾的是，这并没有使普查工作的效果相应地提高。在我们看来，其原因在于，大多数已知的以数学物理的经典概念为基础的定量解释方法，是用来计算激发体的形状、产状、异常物理性质的参数的，而这些参数与拟寻矿化的联系，或者已经证实，或者是推断而未证实的。有鉴于此，将异常划分为远景（矿致）异常和无远景（非矿）异常，是一个尚未解决的最重要而又最复杂的问题。

地球化学和地球物理综合测量结果的解释过程，是一个复杂的、独立的方法问题，它要解决如下基本问题：

（1）在测区内划分出异常；

（2）将划分出的异常区分（分类）为远景异常和无远景异常（那些同拟寻客体的相关概率足够高的异常，我们就认为是远景异常）；

（3）远景异常的定量解释，主要任务是估计矿产的预测储量。

这几个解释阶段的内容较详细地列于下表：

地球化学和地球物理资料解释的总体方案

第 I 阶 段

1. 原始资料：A_1场的剖面图，剖面平面图，等值线平面图

2. 场的统计参数的计算：背景值A_0，异常下限A_n，方差σ_B，相关半径r等。

3. 划分异常：（1）超过A_n值的异常；（2）不超过噪声水平的弱异常。

第 II 阶 段

4. 用信息解释论方法将异常区分（分类）成远景和无远景异常。

5. 定性解释由拟寻客体引起的远景异常。

第 III 阶 段

6. 选择（测量）一些剖面，以便定量计算所解释的异常。

7. 计算激发体的形状和大小。

8. 计算激发体的深度和产状。

9. 估计远景异常范围内矿产的预测储量。

10. 编制解释成果图并评价解释的可靠性。

不难看出，拟寻客体形状和产状的定量计算，在整个解释过程中还不占很多的位置。此外，对于每一种普查方法来说，这些定量计算也是各有特色的。因此，下面我们研究一下前两个，也是最重要的两个解释阶段。

解释的第I阶段——划分异常

划分幅度明显地超过噪声水平的异常，通常是根据关于观测场结构的概率统计概念进行的。在这种情况下，异常划分是从能确保以给定的概率查明有用信号的某一异常下限A_n开始的。为了计算A_n，必须确定正常场内场值的数学分布律，并给定保证查明异常的概率。例如，在正态和对数正态分布律的情况下，广泛采用三倍标准差准则，能以99.86%的概率划分出异常。

为了划分弱异常，也就是幅度相当于或小于平均噪声水平的异常，需要应用专门的方法，首先是方差分析方法，它能计算测

量误差 $\Delta_{测}$ 对总方差 $\Delta_{总}$ 的影响部分和观测场的固有截口 Δ_{0}。如果在这种情况下证明 $\Delta_{测}$ 大大超过固有的方差，那么为了提高地球化学方法解决问题的能力，我们可以提高分析的精度和重现性，而在地球物理测量中，就可以采用更完善的仪器或专门的方法，以保证提高观测精度。

在 $\Delta_{0}>\Delta_{测}$ 的条件下，得用其他一些方法，这些方法大体可分为两大类：（1）地质简化法；（2）基于异常频率滤波的观测场变换法。

用地质简化法划分异常的依据是物理场的叠加原理，按照这个原理，各种地质因素引起的效应是可以求代数和的。降低那些与矿化无关的地质因素（例如疏松沉积的厚度，多年冻土等，其资料可用其他方法获得）的影响，就能使由成矿客体造成的分量突出出来。目前已有在金属矿地球物理方面应用地质简化法的实例（雅罗什，波利亚科夫，1963；卡尼别尔，苏契科夫，1968），在地球化学方法中，原则上也可以应用这种方法（杜鲍夫，索洛沃夫，1974）。

采用逐个排除所有造成异常的已知因素的办法，需要详细研究干扰客体，因而花费很大，正因为如此，各种变换地球化学和地球物理场的方法便得到广泛应用。

最流行的变换方法是用各种方法进行平滑、趋势面分析、位场的往上、下半空间的解析延拓、计算高阶导数。

根据关于观测场本质的谱带合成概念，可以认为观测场是三分量之和：与区域场相应的低频分量、与局部异常相应的中频分量和随机干扰造成的高频分量。平滑方法和把场延拓到上半空间的方法，会消除谱带的高次谐波，从而使有用信号在区域场背景上突出出来，或者突出区域分量本身。在很高的高度上进行实际测量（航空地球物理），或者使用采组合样（在地球物理中是组合测量）的方法，即在测桩附近的几个点上取样（读数），然后计算所有样品（读数）的平均浓度，也能收到类似效果。后一种方法可以同电子工程中常用的有用信号叠加法相比。

观测场换算成各种导数并延拓到下半空间，使观测场的高频分量增强，这样就能划分出局部异常，特别是在噪声水平低或用平滑方法使噪声水平预先减弱的情况下。

近一个时期，为了划分局部异常，越来越多地利用趋势分析。在多数作者看来，趋势图的概念与“区域背景”一词相当。如果把观测场$A(x,y)$看作是系统分量$f(x,y)$(亦即趋势面)与随机分量$\varepsilon(x,y)$的和，则趋势分析实际上可根据点的二次近似公式完成。区域分量可用m次的多项式$P_m(x,y)$来代替，它的系数用最小二乘法根据如下条件求出：

$$\Sigma[A(x,y)-P_m(x,y)]^2=\text{极小值} \qquad (1)$$

与物理场和化学元素分散场的其他变换方法相比，趋势分析技术的优点是，在按不均匀观测网得出原始数据的情况下也能进行计算。

在基于对所研究的场作随机分析而划分和追索弱异常的方法中，还应指出逆概率方法（尼基京，塔尔霍夫，1970），还有以多元统计分析方法为基础的修自系统（самообучающаяся система）（杰穆拉，卢基纳，尼基京，塔尔霍夫，1973）。

尽管有许多观测场的变换方案，它们统统是过滤有用信号的方法。这样一来，对变换观测场最终结果有决定性影响的，不是变换的方式，而是变换参数的选择，这些参数是取平均值的矢径、换算高度、近似多项式的阶次、导数的阶，等等。

不管原始场的变换方法如何，具有最高峰值的场的风暴值通常对划分区域分量的精度都有很大影响。在大多数方法中，这些风暴值使由数据转换成的曲线图上产生“冗繁”效应，结果使区域分量和局部异常（从观测曲线中减去区域场得出）发生畸变。

防止这种不良现象的主要办法，是深入研究一些求观测场平均值[1]的方法，用这些变换方法要计算被变换场的权值，即被变

---

[1] среднение, усреднение，一般译作求平均值或取均值，但它们的含意并不特指求算数平均值，而是泛指用某一层中的统计数值（如算术平均值、几何平均值、中位数、调平均值等）来代表一组或一套数据的运算，位宜译作取中值，以区别于人们习惯的求算术平均值的概念。但本书及本文仍循成规译作求平均值或取均值，请读者注意此术语在各处的不同含意——校者注

换场的数值在各区间内的实际出现频率。利用权函数选择最佳趋势面的方法（帕拉马尔丘克，1972），近似算法的应用（杜鲍夫，1974），以及C.B. 维特罗夫（1972）提出的中位数法和埃吉沃特原理的利用，都可以作为实例。在中位数方法中，在每一区间值的递降序列内选择居中值作为平均估计值，这就剔除了位于序列边缘的最突出的风暴值。在区域（低频）分量具有明显梯度的情况下，要对各取均值区间内（包括滑动窗口中）的第 i 个点上的区域分量进行近似，用下列线性关系效果较好：

$$Y_i = Ax_i + B \qquad (2)$$

式中x为流动坐标。A和B是常数。这时我们就有方程组

$$\Delta_i = Y_i - (Ax_i + B) \qquad (3)$$

根据偏差模数和的极小化条件

$$\sum_{i=1}^{n} |Y_i - (Ax_i + B)| = \text{极小值} \qquad (4)$$

选择出区域分量的参数A和B，我们就可以解方程组（3）。这就是埃吉沃特原理，它实质上是广义的中位数法，因为中位数取均值法也使偏差模数和达到极小。

以偏差模数和极小化为基础的方法的缺点是，在滑动窗口宽度不大的情况下，有几个值可能同时满足这个条件。例如，对于由 1，2，6，10这几个数组成的区域来说，不仅中位数 $M_e=4$，而且2，3，5，6这几个数都满足偏差模数和（等于13）极小这个条件。为了排除这种不确定性，可以提出另一种在滑动窗口中求平均值的方法。

假定在宽度为三个测桩的取均值滑动窗口的不同情况下，出现了图 1 所示的特征 A 的数值的几种组合。按逻辑假设，各个取均值区间内的数值是这样分布的，即具较大概率的事件会在较大程度上出现几个相近的值，而概率较小的事件则不是这样。特别是对正态分布和对数正态分布而言，这种假设是正确的。对于取均值窗口的头一种情况（图1,a）来说，自然可以推测A_1，A_2，

A_3诸点具有大体相同的权，而在第二种情况（图 1,б）下，A_1和A_2的权应当大于A_3的权，因为A_1和A_2相近。第三种情况（图1,в）介于上述二种情况之间，但很可能点 A_3的权要比相互接近的点A_1和A_2小。

由此可见，每一个值A_i的权，都可以作为数轴上各点之间距离的某一函数来定义，也可以作为这些距离（增量）之间的比例的某一函数来定义。在不同的求均值窗口的情况下，滑动值A_H的计算可按下列公式进行：

$$A_H=A_0+\sum_{i=1}^{n}(A_i-A_0)\cdot h_i \qquad (5)$$

式中，A_0是某个A_i值，h_i是差（A_i-A_0）的权，n是滑动窗口内点的数目。

求权 h_i的函数是

$$h_i=-kq_i\log q_i \qquad (6)$$

式中

$$q_i=\frac{|A_i-A_0|}{\sum_{i=1}^{n}|A_i-A_0|}, \qquad (7)$$

$k=1/\log(n-1)$，是为了把h_i变换成与对数底无关的相对值而引进的系数。不难证明：

$$0\leqslant -k\sum_{i=1}^{n}q_i\cdot\log q_i\leqslant 1$$

式中 n 是滑动窗口内点的数目。

根据式（7）可以得出$\sum_{i=1}^{n}q_i=1$。

之所以选择函数（6）作为权，是因为当q_i接近于 0 或 1 时，它最小；而当$0.2<q_i<0.6$及h_i的函数取最大值时，它的变化又比较平稳（雅格洛姆A. M.，雅格洛姆И. М.，1973）。因此，A_0点必须这样来选择，即须让那些最有代表性的（占多数的）A_i值比

那些对给定样本来说不太典型的值有较大的权。在下列条件下，从A_1，A_2，A_3，…，An构成的总体中选择A_0，是可以达到这一要求的。该条件是：

$$H(A_0)=\sum_{i=1}^{n} h_i(A_0)=\text{极大值} \qquad (8)$$

$H(A_0)$是一个函数，用来描述在n个测点组成的滑动窗口中特征A的值相对于A_0点来说分布的均匀性。这个函数是通过与以下函数类比得出的：

$$H=-\sum_{i=1}^{n} p(A_i)\cdot\log p(A_i) \qquad (9)$$

函数H是申农为了求具有几个结局的试验A的熵(不确定性度量)而提出来的，其中$p(A_i)$为结局A_i的概率。对于具有等概率事件的试验来说，函数H达到极大值。

在这种情况下，极大值$H(A_0)$意味着，与其他任何一个$A_i\neq A_0$的点相比，在A_0点上的特征A的值与滑动窗口所有其他的特征值是最“等距的”。由此可得出结论，进入求均值窗口的风暴值越大，它对A_H的最终求均值的结果影响就越小，因为当A_0等于这个异常风暴值时，函数$H(A_0)$将取最大值，而且相对于A_0来说，A_i的所有其他的值都将是相近的。

在图1的a，$б$，$в$三种情况下，按准则（8）确定的A_0点分别是A_2，A_3，和A_2。图上也标出了根据这些点计算出的平均值A_H。

用上述的“熵法”计算滑动值的算法，由以下概括即可明瞭：

1．选择落入求均值窗口的值A_1，A_2，A_3，…，A_n。

2．根据这一总体求使式（8）达到极大值的$A_i=A_0$的点。

3．计算差(A_i-A_0)，求模数和$\sum_{i=1}^{n}|A_i-A_0|$，分别按式（7）和（6）计算q_i和h_i。

4．按式（5）计算平均值A_H，并将其挪到求均值窗口的中心。

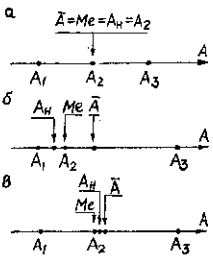


图 1 算术平均值 $\bar{A}$、中位数Me和熵平均值 A_H 之间的关系

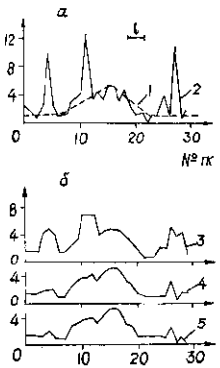


图 2 三种方法滑动的结果（实例）
1—给定的异常；2—总曲线（干扰+给定的异常）；3—$\bar{A}$；4—Me；5—A_H；l—滑动窗口宽度

5．把窗口移动一个测桩，再从第一项开始全部重做一遍。

这种方法用手算极其繁难，故运算是在 Минск-22型电子计算机上完成的。

图 2 表示的是一个典型实例和用不同方法根据三个测点求均值的结果，这三种方法是：求算术平均值法，求中位数法和求“熵”平均值法。原始曲线表示一个异常，其上叠加着平均值为零的正态分布的干扰和三个局部风暴高值（图 2，а）。取均值的结果表明，这些具有峰值特征的局部异常，使A曲线畸变的程度相当大，而使Me和A_H曲线畸变的程度较小。给定异常的取均值曲线的适宜程度，可根据相对于理论曲线的均方差的大小来确定。采用算术平均值$\bar{A}$、中位数Me和我们提出的A_H等方法来求均值，均方差分别为2.0，0.9和0.7。由此可见，“熵”平均法得出的结果与真值最接近。

最后我们要强调的是，尽管广泛地引进了划分物理场的数学

方法，但划分区域背景的直观图解法到目前为止仍不失其价值。而且如果一个研究人员有足够的经验的话，他不用作复杂的数学运算，有时亦可确保成果出得又快又好，因为一个专家可直觉地利用观测场中迄今尚难给予数学模拟的细微差别。

解释的第二阶段——划分远景异常和无远景异常

对根据综合地球化学和地球物理调查的结果划分出的异常的远景作比较评价，是依据信息解释理论用各种方法完成的。其中最流行的是模型逻辑识别方法，它有这样一些众所周知的程序，诸如"Кора-3"，"Голотил"，"Образ-3"，"Тау"等等（鲍罗夫科，1972）。此外，用拜耶斯假说的定理，相关分析和回归分析方法，以及统计判定理论进行综合解释的处理方法（尼基京，1970），都日趋完善。

在应用地球化学和勘探地球物理中，解疑难课题的途经之一，就是把所有普查方法的测量结果都结合起来，建立综合指数。伊尔库次克工学院为此提出了一种建立综合指数函数（ФКЛ）的方法，其依据就是对被观测物理场作随机分析。广义的综合指数函数是将各种方法（测量结果均以数字表示）的信号混成一体的一种结果。这个函数可用来：（1）压缩信息；（2）划分幅度与干扰的平均水平相当的弱异常；（3）将划分出的异常区分为远景（矿致）异常和无远景（非矿）异常。

为了表示原生分散晕多元素测量结果，已提出了建立累加晕和累乘晕的方法（奥甫钦尼科夫，格里戈梁，1970）和计算成矿（成晕）能的方法（萨弗罗诺夫，拉普，麦谢里亚科夫，1970）。累加晕是把几个元素的浓度根据背景规格化后相加得出的，加一的元素须是在对该类型矿床确定的分带序列中占有相近位置的相组元素。累乘晕是通过把样品中的元素浓度连乘而得出的，它相当于浓度对数的累加晕。

地球化学资料用累加的、累乘的原生晕或成晕能表示出来，增加了原生晕的有效规模和衬度（由于通过相加压抑了相对于背景的偶然误差），使原生晕的带状结构更加明显，以便于确定侵蚀截面水平。

在两种方法中，组合在一起的是根据背景浓度规格化后得出的无量纲的量（或它们的函数），但这些量的方差不同，其大小主要取决于给定元素的背景绝对值。

由方差的已知性质可知，随机量$y=Kx$的方差等于

$$\overline{A}(y)=K^2\overline{A}(x) \tag{10}$$

式中K是常数因子。因此，用背景值去除浓度，甚至在方差相同而背景值不同的情况下，也会使无量纲的量有不同的方差。这样的话，把这些规格化的量加起来时，我们所采用的就会是精度不同的测量值。在不同元素的背景浓度之间不存在相关关系的情况下，累加晕背景的方差将等于

$$\overline{A}=\sum_{j=1}^{n}\overline{A}_j \tag{11}$$

式中，$\overline{A}_j$是第j个元素规格化浓度的方差，按下列公式计算：

$$\overline{A}_j=\frac{\overline{A}(C_j)}{C_{\phi j}^2} \tag{12}$$

式中，$\overline{A}(C_j)$是第j个元素浓度的方差，而$C_{\phi j}$是其背景值。

地球化学和地球物理的原始数据用综合地球物理指数θ的函数来表示是比较合理的（瓦赫罗麦耶夫，1970）。这种方法要求事先把每一种方法的数据都按下式进行变换：

$$\gamma_{ij}=(A_{ij}-\overline{A}_j)/S_j \tag{13}$$

式中，γ_{ij}是按标准S_j规格化的与背景$\overline{A}_j$的偏差；A_{ij}是j方法在剖面的i点上测得的特征值。这个公式在背景值为正态分布律时是对的，而在对数正态分布律的情况下，应当利用公式

$$\gamma_{ij}=(\lg A_{ij}-\lg\overline{A}_j)/\lg\varepsilon_j \tag{13'}$$

式中，ε_j是无量纲的标准因子。

按公式(13)和(13')变换后的正常场的特征值，其分布服从于

具有零平均值和单位标准差的高斯律。Г. А. 沃斯特罗克努托夫（1972）提出了一种将分布律不同于正态分布或对数正态分布的值规格化的方法。在这种情况下，量γ和被变换特征之间的关系不能用线性函数（13）和（13′）来近似，而要用一个较高次的多项式来近似。由于利用无量纲值γ，就能够借助严格的统计准则估计地球化学和地球物理综合异常的显著性。

根据具体地质环境、物理性质和规定的任务，在选择某一种ФКЛ时，要使标准矿体上的综合异常衬度最大。例如，有这样一个最简单但相当有效的函数：

$$\theta_i = \sum_{i=1}^{p} \gamma_{ii} - \sum_{k=1}^{q} \gamma_{ik} \qquad (14)$$

式中，θ_i是剖面中i点上的ФКЛ值；p和q是分别依据正异常和负异常来划分矿体的方法的数目。

Н. Н. 索切瓦诺夫（1972）提出的函数θ具有同样的含义，不过他是把按背景而不是按标准差规格化的值加起来，这就降低了这种方法的效果。我们举一个简单例子来说明这一点：需要建立铅、锌、铜的组合累加晕和组合ФКЛ晕，这几个元素的背景浓度分别是10，20，50（假定单位），而所有元素的标准差都相等，等于1。它们在晕中的最大浓度分别是15，30，75。在根据式(12)按背景规格化之后，得出$S_{Pb}=0.1$；$S_{Zn}=0.05$，$S_{Cu}=0.02$。

由式（11）可知，累加晕的标准差等于

$$S_{\text{加}} = \sqrt{S_{Pb}^2 + S_{Zn}^2 + S_{Cu}^2} \approx 0.11$$

规格化后的最大异常含量的和和背景含量的和分别是

$$C_{\text{加}}^{\max} = 1.5 + 1.5 + 1.5 = 4.5$$

$$C_{\text{背}}^{\phi} = 1 + 1 + 1 = 3$$

累加晕异常的衬度是

$$\gamma_{\text{加}} = (C_{\max} - C_{\phi})/S_{\text{加}} = (4.5 - 3)/0.11 = 14$$

按公式（14）计算出的θ函数的同一些参数是

$$\theta_{\max} = \gamma_{Pb} + \gamma_{Zn} + \gamma_{Cu} = 5 + 10 + 15 = 40$$

$S_{Pb}=S_{Zn}=S_{Cu}=1$，　　$S_\theta = \sqrt{3} \approx 1.73$　　$\theta_{\text{фон}}=0$

组合异常的衬度是

$$\gamma_\theta=\theta_{\max}-\theta_{\text{背景}})/S_\theta\approx 23$$

这样，根据函数θ计算出的组合晕的衬度，比累加晕的衬度大。只有在每个元素背景的变差系数相同的情况下，用两种方法才能得出相同的结果，但实际上这种情况显然相当罕见。

利用函数（14）可以按照通用的三倍标准差准则估计异常下限θ_{an}。实际上，不相关随机量的和或差的方差可按公式（11）计算。因为γ_i值已经根据标准差S_i规格化了，所以我们有

$$Д(\gamma_1)=Д(\gamma_2)=\cdots=Д(\gamma_j)=\cdots=Д(\gamma_m)=1$$

式中，$m=p+q$。这时根据（11）得出$Д(\theta)=mД(\gamma_i)=m$或

$$S_\theta=\sqrt{m} \tag{15}$$

式中，$Д(\theta)$和S_θ分别是正常场中θ的函数的方差和标准差。

使用三倍标准差准则的异常下限按下式计算：

$$\theta_{an}=3S_\theta=3\sqrt{m} \tag{16}$$

因此，模数超过值θ_{an}的任何一个值θ_i，都可以以0.9986的概率认为是非随机的。组合异常的幅度与m的数目是成正比的，这时，由式（13）和（15）可知，有用信号强度与干扰强度的比值将与$\sqrt{m}$成比例地增加。对于根据剖面上n个观测点或n个剖面划分出的异常来说，θ_{an}相应地减少：

$$\theta_{an}=3\sqrt{\frac{m}{n}} \tag{17}$$

式中$n\leqslant 9$。

图3表示录自一本书（雅尼舍夫斯基等人，1963）的一个处理原生分散晕资料的实例。在有两个相邻矿体的地段，铀、铅和钼的晕同时出现；在建立函数θ（图3，6）之后，就可以很有把握地根据数值$\theta>\theta_{an}$把含矿带划分出来，而单个矿体的晕之间也有明显不同。正如前面已经说到的，公式（15）和（16）适用于不相关背景值相加的情况。更常见的情况是，要在背景值之间存在线性相关的情况下估计S_θ，Г. А. 沃斯特罗克努托夫（1972）已对此作了研究。

用手算γ_{ij}，实际上和按背景值进行规格化一样繁难。最好以下述方式进行计算：

1．根据第 j 个特征的正常场的样本，求出背景值$\overline{A_j}(\overline{\lg A_j})$和标准差$S(\lg\varepsilon)$。

2．由于γ_{ij}和$A_{ij}(\lg A_{ij})$之间的相关关系是线性的〔见（13）和（13′）〕，便可作成一个线性函数曲线：

$$y=kx+B \tag{18}$$

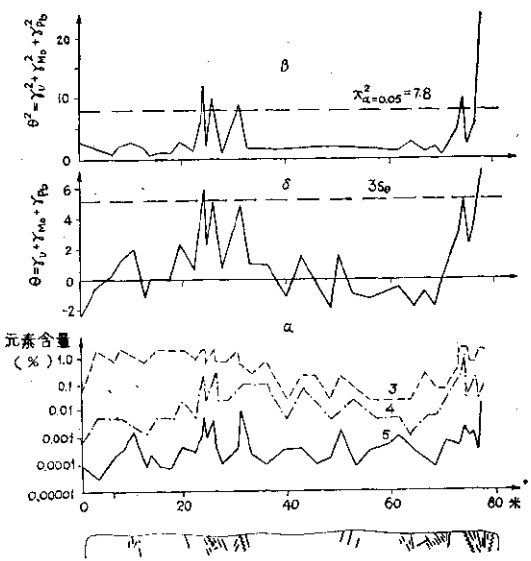


图 3　根据U、M_o和Pb单岩石地球化学测量数据（a）和用两种方案（$б$，$в$）变换出的ΦКЛ划分相邻矿带

1—构造裂隙；2—含矿裂隙；3—铅；4—铀；5—铀

式中，在背景值作高斯律分布的情况下，$k=1/S_j$，$B=-\bar{A}_j/S_j$，或者在对数正态律分布时，$k=1/\lg\varepsilon_j$，$B=\overline{\lg A_j}/\lg\varepsilon_j$。在处理对数正态分布的数值时，横坐标轴用对数尺度。

根据规定的任务不同，综合指数的函数可有各种形式（瓦赫罗麦耶夫，舍斯塔科夫，1972），而在多数情况下用的是函数

$$\theta_i^2=\sum_{j=1}^{m}\gamma_{ij}^2-\sum_{k=1}^{p}\gamma^2{}_{ik} \qquad (19)$$

式中，m和p是被综合的方法的数目，两个代号分别表示其异常是控矿的有利因素和不利因素的方法。利用公式(19)求出的ΦКЛ，可以用来在自由度为（$m-p+1$）的情况下根据χ^2准则估计结果的显著性。图3表示一个按公式（19）处理数据的实例。

通过对原始数据作预先变换，可提高ΦКЛ解决问题的能力。例如，计算滑动算术平均值，就可降低观测场的高频分量(干扰)，从而使背景方差减少$\frac{n-1}{n}$（n是滑动窗口的测点数目）。由式(13)可知，这种情况下异常的衬度会增加到$\sqrt{n}$倍。因为普查网的布置都希望以不多的观测点划分出异常，所以为以后的变换而作滑动时，务必使滑动窗口的宽度取最小值（沿剖面不超过3—5个测点）。在这种情况下，信号不会有大的畸变。

图4表示一个划分综合异常的实例。干扰是具有零平均值的正态分布。衬度γ_{ij}是根据三点滑动平均后的原始数据计算的。通过比较未平滑值（图4，*б*）和平滑值（图4，*в*）的衬度累计结果可以看出，在进行平滑的情况下，用几个θ值超过三倍标准差的点就可很有把握地把组合异常划分出来。

不是“经典的”异常，而是在场的分布上具特殊性的地段，往往是一种找矿标志。例如，在围岩和拟寻客体之上场的背景值相同的情况下，可根据场的偏高或偏低的可分割性及其他特征，把拟寻客体划分出来。为了把这些数据用于ΦКЛ，可将其变换成方差曲线；标准差和其余的参数也要变换。

图5表示在普查稀有金属伟晶岩脉集中地段时处理综合方法

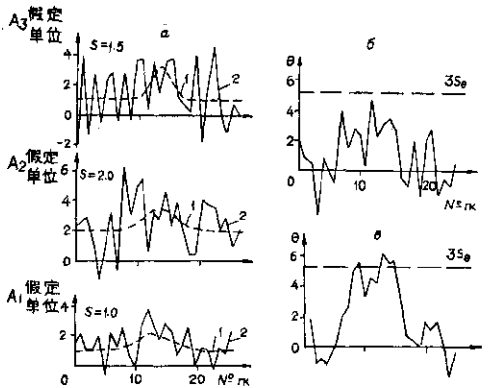


图 4　用三种方法(A_1, A_2, A_3)利用综合指数的典型实例。S是标准差(其单位是假定单位)

а—模型的原始曲线；*б*—根据原始数据计算的结果；*в*—根据经过预先平滑的数据计算的结果；1,2—见图 2

(磁测，伽玛能谱测量)数据的一个实例。图中，这些地段可根据方差偏高的磁场划分出来，这种现象是由控制稀有金属矿化的正角闪岩的交代蚀变造成的。坑道中的测量表明，伟晶岩脉可根据铀含量偏高（2—3 个背景值）和钍含量偏低(比背景值约低$\frac{1}{3}-\frac{1}{2}$)划分出来。因此，在伟晶岩脉集中的地段之上，应当观测到铀含量偏高和钍含量偏低的趋势。为了把这些数据组合成ΦKJI的形式，可把磁场的垂直分量 ΔZ 用宽度为 3 个测点的滑动窗口法换算成标准差。按下列公式计算：

$$S_i^z=\sqrt{\frac{\sum\limits_{k=1}^{n}(\Delta Z_k-\Delta\bar{Z}_i)^2}{n-1}} \qquad (20)$$

式中，S_i^z是标准差，$\Delta\bar{Z}_i$是以剖面上的 i 点为中心的滑动窗口内

场的算术平均值；$n=3$，是落入滑动窗口的观测点的数目。然后分别根据S_i^2以及铀和钍的浓度计算出衬度，把它们按下式相加：

$$\theta_i=\gamma_S+\gamma_U+\gamma_{Th} \qquad (21)$$

式中，γ_S、γ_U和γ_{Th}分别是磁场标准差，铀和钍浓度的衬度，与θ_i一样，它们都是剖面上的i点的数值。

由图5看出，伟晶岩脉集中分布的地段，已根据综合异常可靠地划分出来了；除此以外，200—400米这一区段的异常，也是有稀有金属矿化远景的。

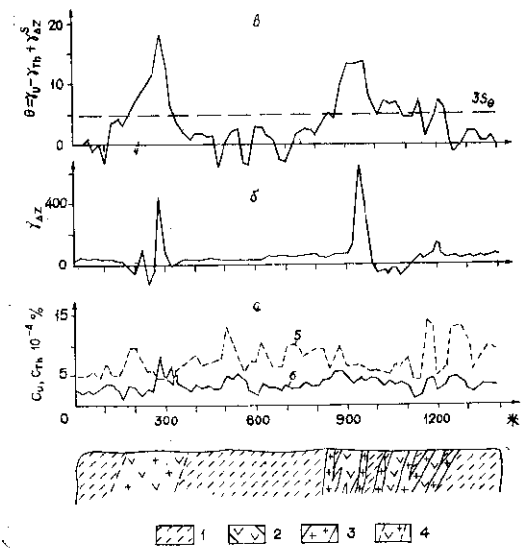


图5 稀有金属伟晶岩脉发育地段的伽玛能谱测量（а）、磁测（б）结果和ФКЛθ（в）

1—石英黑云片岩；2—正角闪岩；3—伟晶岩脉；4—推断的正角闪岩中的伟晶岩脉富集带；5—Th；6—U

在根据个别方法所得异常的形态差别很大的情况下把数据变换成曲线的问题，在有的著作(瓦赫罗麦耶夫，舍斯塔科夫，1972)中已经作了探讨。

为了进行 ФКЛ 计算，已经编成了在 Минск-22 型电子计算机上用的程序。由于计算不太复杂，万不得已时也可手算。

利用综合指数函数对结果进行综合解释的方法，在东西伯利亚、外贝加尔和蒙古人民共和国铁、有色金属（钴、铜）矿床、多金属和稀有金属矿床、萤石以及压电石英的地球物理和地球化学普查中，已经通过了技术鉴定。

参 考 文 献

Бороеко Н. Н. Применимость и эффективность различных схем распознавания в типичных геологических задачах. —В кн.: Методы разведочной геофизики. Вып. 15. Л., "Недра", 1972, с. 5—14.

Вахромеев Г. С. Общие принципы комплексирования геофизических методов при поисках рудных месторождений. Иркутск, 1970. 120 с.

Вахромеев Г. С., Шестаков Ю. Г. Обобщение результатов геофизических и геохимических съемок с помощью функции комплексного показателя. — "Разведка и охрана недр", 1972, № 5, с. 38—42.

Ветров С. В. О возможности применения метода медианы и принципа Эджворта для обработки геофизических данных. —В кн.: Геофизические исследования при решении геологических задач в Восточной Сибири. Вып. 6. Иркутск, Вост.-Сиб. изд-во, 1972, с. 85—90.

Вострокнутов Г. А. Некоторые статистические методы выделения геохимических аномалий. —В

кн.: Геохимические поиски рудных месторождений. М., "Недра", 1972, с. 207—218.

Демура Г. В., Лукина О. П., Никитин А. А., Тархов А. Г. Выделение геохимических аномалий с помощью самообучающихся фильтров. —"Изв. вузов. Сер. Геология и разведка", 1973, № 9, с. 103—109.

Дубов Р. И. Количественные исследования геохимических полей для поисков рудных месторождений. Новосибирск, "Наука", 1974. 278 с.

Дубов Р. И., Соломов А. П. Линейка и круговая палетка для расчета вторичных ореолов рассеяния. —"Разведка и охрана недр", 1974, № 4, с. 23—27.

Канибер Ю. П., Сучков А. П. О возможности учета гравитационного влияния погребенного рельефа с использованием данных электроразведки. —В кн.: Глубинные поиски рудных месторождений. Вып. 2. М., 1968, с. 90—101.

Никитин А. А. Применение теории статистических решений для классификации геологических объектов по данным комплекса геофизических методов. —"Изв. вузов. Сер. Геология и разведка", 1970, № 10, с. 135—144.

Никитин А. А., Тархов А. Г. Статистические приемы извлечения информации при обработке геофизических данных. —"Геофиз. сб. АН УССР", 1970, вып. 35, с. 78—90.

Овчинников А. Н., Григорян С. В. Закономерности состава и строения первичных геохимических ореолов сульфидных месторождений. —В кн.: Научные основы геохимических поисков глубокозалегающих

рудных месторождений. Иркутск, 1970, с. 3—36.

Паламарчук В. К. Опыт разделения аномалий методом тренда.—"Геол.и геофиз.", 1972, №2(146), с. 86—92.

Сафронов Н. И., Лапп М. А., Мещеряков С. С. Научные основы поисков глубокозалегающих месторождений. —В кн.: Научные основы поисков глубокозалегающих рудных месторождений. Иркутск, 1970, с. 37—78.

Сочеванов Н. Н. Новые методы расчетов и изображения результатов при изучении многокомпонентных геохимических ореолов.—В кн.: Геохимические поиски рудных месторождений. М., "Недра", 1972, с. 218—222.

Халфин Л. А. Информационная теория интерпретации геофизических исследований. —"Докл. АН СССР", 1958, т. 122, № 6, с. 1007—1012.

Яглом А. М., Яглом И. М. Вероятность и информация. М.-"Наука", 1973. 511 с.

Янишевский Е. М., Григорян С. В., Баранов Э. Н., Вертепов Г. И., Каблуков А. Д. Эндогенные ореолы рассеяния некоторых гидротермальных месторождений. М., Госгеолтехиздат, 1963. 123 с.

Ярош А. Я., Поляков А. Б. Поиски и разведка медноколчеданых месторождений на Урале гравитационным методом. М., Госгеолтехиздат, 1963. 131 с.

纪忠元译

论选择、论证和建立数据数学处理方法的理论问题

P. И. 杜鲍夫

前　言

我们认为，各种各样的观测结果，包括测量结果，只要必须根据它们来得出有关被研究客体 r 的某种解，都可以理解为待处理数据。数据处理就是对数据进行各种便于求解的运算。

本文的目的是：研究数据处理的方法，解决为提高方法的应用效果而与这些方法的选择、论证和建立有关的问题。

数据的数学处理现在十分普遍，但由于方法选择不当，它的使用往往并未奏效。同样的数据一下子用几种方法来处理，不见得就有好处，一来花费大，二来往往得出矛盾的结果。根据方法作者的推荐来选定方法常常是困难的，因为许多作者对所提出的方法指出了很广的应用范围，且互相有重叠，同时还没有严格的论证。某些研究者甚至认为存在着一些原则上无须论证的方法。这些所谓启发式的方法现在被广泛地使用着，有时会得到良好效果（《地质学和数学》，1970），但在这种状况下难以事先就预见到可成功地使用这些方法的范围。

近些年来，方法选择理论的缺乏已开始明显地遏制了地质学中数学研究的发展。在建立数据处理的自动化系统时，这种感觉尤其突出。因此，出现了一些对不同方法作对比分析的文献。其中，杜鲍夫（1973a,б, 1974a）和帮达连科等（1974）的著作和某些其它著作即属此类。在这些著作中，都表示了这样的意见：

总的说来，将一定类别的方法用于地质学、地球化学或者科学的其它大的分支，是有优越性的。方法选择理论发展的下一个阶段，显然应该是制定一套规则，其中每一条规则都可缩小具体条件下对所需方法作选择的不肯定性的范围。本文将引入一些方法选择理论的基础所必需的概念；并探讨在一系列有重要实践意义的情况下可大大减少最佳方法选择的不肯定性的规则。

我们所找到的原理，不仅对于选择现有的数据处理方法是可行的，而且对于论证和建立新方法也是可用的。

我们将用一些实例来说明理论上的结论，每个实例可能对于分析具体数据处理方法的应用可能性，都具有独立的意义。❶

一般原理和定义

我们首先指出，所谓数据处理方法，就是从观测结果中抽取有用信息的算法，而这种算法又应是用于某种非单一事件的集合的。如果某一个算法只可用于解决一个具体的课题，那它就不是一种方法，而只是这个课题的解法。

我们引入下列主要符号。

某个被研究的第 n 个方法，我们将以 F_n 标记之。这个符号也将用于在使用该方法求解时所需借助的算子（算法）。

$Q=\{q_i\}$——为研究某个客体 r 而收集到的待处理数据。q_i 的数值可以是地球化学场的观测结果（岩石、矿体和矿体晕中化学元素的浓度值），以及地球物理场、岩石学特征、经济条件和其它特征的观测结果。数据的总体 Q 可以视为客体 r 的标志（在一般情况下，它是与 q_i 值的数目相应的多元标志），我们有时将使用这个名称。为了表示出标志 Q 属于一定的客体 r，我们引入

❶ 在此，实例的解析都是从最一般的原理出发而以达到得出算法为准。为了理解所介绍的原理，广大地质工作者希望引用一些处理具体数据直到用它得出数字结果的实例。但是，这样的实例要占用相当大的篇幅，因此，拟在晚些时候于专著中发表——原注。

一个符号$Q=Q(r)$。反过来，符号$r(Q)$将表示具有标志Q的客体r。

我们用符号$H=\{h_j\}$表示客体r的对我们有重要意义的特征的总体。我们将认为，所有的特征或者它们的一部分是直接未知的（要不然根据数据Q进行研究似乎就没有意义了）。例如，在地球化学测量的地区内未揭露矿体的规模，客体r所属类别的名称等等，都是这样的特征。特征h_j中的某一些是能够直接观测到的。这时，这些被观测到的特征也作为元素q_i❶加入Q之中。

$r^{(0)}$——已被研究过的客体；$Q^{(0)}=\{q_i^{(0)}\}$——已研究客体的标志〔$Q^{(0)}=Q(r^{(0)})$〕。数据$q_i^{(0)}$——与对q_i所作观测类同的观测结果（与客体$r^{(0)}$有关的同样名称的场所作测量的记录）。$H^{(0)}=\{h_j^{(0)}\}$——客体$r^{(0)}$的与特征h_j类同的特征，不过它是已知的。

关于客体$r^{(0)}$的资料$Q^{(0)}$和$H^{(0)}$，属于习惯上称为训练资料的一部分。这些资料可以用分别属于各个具体客体的具体数字的形式存储起来，或者用从理论上加以概括后的相关性的形式存储起来。在后一种情况下，这些资料与已认定存在这种相关性的客体的无穷集合的具体资料是等效的。例如，这种相关性可通过残留晕中浓度的分布与矿体参数之间关系的公式表示出来（杜鲍夫，1974б）。下面我们使用的符号$r^{(0)}$、$Q^{(0)}$、$H^{(0)}$，将不涉及它们是与各个具体客体有关的，或是与经理论概括的相关性有关的。

训练资料和待处理数据Q结合起来，便构成用于求得解A_k^*的原始资料。

方法的差别

我们将从原始资料和所获结果A_k^*的形式方面探讨各方法的差别。在这个探讨中，不涉及各方法在运算、计算的繁简和准确

---

❶ 这里所说的元素并非化学元素，而是数学上的元素，下文亦有类似情况——译注。

性方面的差别，这些差别在现有的数学文献中已有探讨。例如，在选择解线性方程组、内插格式等问题的方法时，就会碰到这类差别。

我们将对地球化学和其它地质数据处理中目前采用的几类主要方法，引证一些它们之间在所讨论范畴内的差别的例子。同时，我们将澄清某些术语的涵义，这对于它们的使用可能是有益的。

1．解反演问题

根据观测与客体有关的场来得出客体的特征，在数学上称之为解反演问题。这与正演问题是相反的。解正演问题时，研究者负有的任务是对客体的特征作描述，并弄清应该怎样用观测结果把这些客体反映出来。这就意味着，在处理地质数据时要解的所有问题，都是广义的反演问题。

许多地质工作者，特别是地球物理工作者，都狭义地理解解反演问题，他们认为，反演问题的解是借助与在正演问题中表示场的物理模型的一定算子相反的某个算子F_k^*来得到的。正演问题中的物理模型概括了关于观测场与造成观测场的客体之间联系特征的资料，在解反演问题时，它起着训练资料的作用。关于具体客体$r^{(0)}$的资料，在这里并不直接需要（但要使用的所谓理论模型，可以事先利用这些资料建立起来，特别是以由实验确定的物理定律的形式建立起来）。数据Q应当包括场的观测值和观测点的座标。处理结果表示为对客体r及其状况（矿体的厚度、延伸长度、产出深度、矿石数量、测定点的座标等等）的数字的和函数的表述。

2．合流分析和回归分析

在这里，被用来作为训练资料的是函数F_k^*，它反映着观测数据$q\in Q$和研究者感兴趣的量A之间关系的估计量。通常，这个函数的形式，或者甚至某些参数的值，要预先根据客体$r^{(0)}$上的观测值$Q^{(0)}$计算出来，而对该客体来说，量A的数值$A^{(0)}$同样应是已知的或者已观测（测量）过的。数值集合$Q^{(0)}$和$A^{(0)}$在一般情况下都含有噪音分量（包括测量误差），依函数的建立方式不

同，噪音对函数F_k*的影响可以是不一样的。选择能减小噪音影响的函数建立方式，就称为对噪音分量（或简称噪音）的压抑。

如果函数F_k*是在不仅压抑了$q^{(0)}$值中的噪音，而且也压抑了$A^{(0)}$值中的噪音的情况下建立的，那么使用这个函数去处理数据Q就称为合流（Конфлюэнтный）分析。如果噪音分量的压抑只限于$A^{(0)}$值，那么建立和利用所得到的函数关系就称之为回归分析。在后一种情况下，F_k*便叫做回归函数。

我们把"回归函数"这一名称也推广使用于在合流分析中运用的函数F_k*。如果在建立函数F_k*时噪音的压抑可以在不同数目的被使用量内进行，"回归分析"这一术语就可以更一般化地予以使用。如果在建立函数F_k*时对被使用的所有量中的噪音都进行了压抑，那么就可将合流分析视为回归分析的一种特殊情况。

建立函数F_k*就是对某种通常未知的相关性F进行近似（在个别情况下，当相关性F已确切知道且不过于复杂时，它就可直接作为函数F_k*来使用）。近似可以依据具体客体$r^{(0)}$的数据集合$Q^{(0)}$和根据理论假设（即根据事先的经验总结）来进行。因此，当在函数中使用的是前已述及的有物理学依据的算子时，解上述狭义的反演问题就可视为回归分析的一种特殊情况。回归分析中的待处理数据Q和所得到的结果具有第一点中谈到的那种形式。

我们转而谈谈所谓的趋势分析，它常常被看成是一种独立的方法，从实质上说，它也是建立回归函数。如果把数据$q_i^{(0)} \in Q^{(0)}$换成空间座标，而把观测场在此座标各点上的数值视为数值$A^{(0)}$，上述建立函数的各种要求对于建立趋势分析的回归函数来说，就都是正确的了。与回归分析不同，在趋势分析中以得到回归函数为最终结果，可是回归分析却是要利用事先得到的回归函数。

3．观测场的统计特征

属于此类特征的有下列各参数的估计值：各种平均值、方差、相关关系、熵，等等。在这里，训练资料须具有表达这些估计值的十分确定的算子形式。具体客体$r^{(0)}$的特征常常并不被直接使用。待处理数据Q可以用各种观测值的集合的形式分别给出，用

时要标明座标，座标精度只须能将一种数据集合与另一种区分开来也就行了。结果以被统计特征的数值形式给出。

4．分类

所谓分类一般指的是将一个客体归属到有限总体的某个类中去。但严格说来，类别的种数原则上是不限的。因而广义地说，随便一种什么样的解法，包括回归，都可以理解为分类，因为任何一种数值都可以看作一定类别的标志。例如，计算矿床中矿石数量的估计值，就可以看作按储量对矿床作分类。

分类可以借助已被充分或不充分肯定了的判定规则（算子）F_k^*来进行。

在第一种情况下，习惯上把这种方法称为无训练分类，尽管实际上算子F_k^*的建立是有理论依据的，而且这种依据是以经验（即早先总结出的训练资料）为基础的。在此应指出的是，甚至就像欧几里德几何学这样一种基础数学理论，也是建立在从经验得出的公理的基础上的。归纳法也是经验的潜意识的总结。

现在，某些（启发式的）方法所以被推荐使用，其依据就在于它们在一定条件下通过试验得出了良好结果。因而这些试验的结果就可用来选择算子F_k^*，这就是说它们起着训练资料的作用。

在第二种情况下，如果算子F_k^*尚未充分肯定，为了充分肯定它就需要使用在解题时选出的具体客体$r^{(0)}$的训练资料。这种条件下的方法就称为有训练分类。不同方法所使用的数据Q可以是不一样的：数据的数目不等，项目各异；带有观测点座标和不带观测点座标；等等。结果也可能是不同的，而且，按照信息论，结果的熵（不肯定性）总是不低于原始数据的熵。

目标课题的求解

当所得到的结果并不是目标本身，而只是在求解更远课题（目标课题）的道路上迈出了一步时，在数据处理方法的选择上

就会遇到困难。

实际上，如果某种方法利用所收集到的数据 Q 完全地解决了一个目标课题，那么，它就分明比任何一种在所获信息方面与之不同的其它方法更优越一些。但问题在于，现有的任何一种方法往往只能给出关于解这种课题的不完全的信息。

在现有的文献中，目标课题常常并未明确表述出来，但总是含于其中。例如，在一篇论文中（基塔耶夫、叶夫道基莫娃、楚马金、西克连，1974）叙述了一种划分地质剖面中具有一定化学元素组合的带的方法。从文章所举的例子中可明显地看出，作者并未把这种划分作为目标本身，而是作为获取可能与这些带有关的矿体产出位置的信息的手段。

在处理小比例尺地质测量数据时，目标课题可以是解决进一步详细工作地区和工作形式的问题；在矿床勘探中，可以是解决矿床的经济评价问题；如此等等。

现在问题已经清楚，为了对各方法就其在处理数据 Q 方面的优度进行对比，必须提出（表述）这些方法共有的目标课题。下面，我们将把这种课题的解记为 A。

一个目标课题可以是求解另一个目标课题的过渡阶段。例如，在完成根据小比例尺地球化学测量数据 Q 选定的详细工作之后，可能就需要选择揭露异常的山地工程，此后，可能还需要选择用以勘探被发现矿体的勘探工程，等等。下面，只要不附加说明，我们所说的目标课题将都指的是“最小的”目标课题，即在那些可供对比的处理数据 Q 的各种方法所共有的课题中，属于初级的和最简单的目标课题。

在具备必要的训练资料的情况下，某种可从数据 Q 中提取较多关于解 A 的信息的方法，我们将称之为较有信息的方法。显然，最有信息的方法也就是所有参加对比的方法中最优的方法，只要该方法所需的资料齐备就行。如果训练资料不充分，最优的方法也就是在这种资料条件下所能使用的诸方法中最有信息的方法。

对各种信息的估计，如所周知，甚至就其本身而言也是一个

复杂的问题，因此需要专门分析。为了得到我们所需要的、能对各方法的信息量作对比的足够肯定的规则，我们来分析一下求得目标课题的明显最佳解的过程。

目标课题的构成就意味着在该课题的解A与某个因素集合$H=\{h_i\}$间建立起联系，即建立表达式

$$A=\Phi(H), \qquad (1)$$

式中，$\Phi(H)$——目标函数，它用一个数字或名称表示出想要对客体r认识到的问题。例如，Φ可以是表示矿体中有用矿物工业储量的数字，而H可以是关于该矿体的线性规模和矿体中有用矿物富集程度的信息的集合。在个别情况下，例如当H是客体r所属类别的名称时，Φ与H可以是一致的。

根据表达式（1）可以针对任一个客体r计算出A来，只要对该客体而言，所有的值$h_i\in H$都是已知的。

我们首先假设已具备了训练资料，在每个客体$r^{(0)}$的训练资料中存在有资料$H^{(0)}=H$。这时，按照公式（1）可以对每个客体$r^{(0)}$计算出数值$A^{(0)}$：$A^{(0)}=\Phi(H^{(0)})$。训练资料中的每一个这样的数值都与它自己的数据集合$Q^{(0)}$相对应。我们假设，在这个集合中明显或不明显地包含有关于集合$H^{(0)}$的信息，并已找到了$A^{(0)}$的这种数值与$Q^{(0)}$之间以某个算子F体现出来的确切而肯定的对应关系：$A^{(0)}=F(G,Q^{(0)})$，这里，G——被称为自由度的参数g的集合（g应该根据数据$Q^{(0)}$来确定）。

这时，根据客体r与客体$r^{(0)}$类似，便可以认为：算子F同样可把客体r的未知数值A和数据Q联系起来：

$$A=F(G,Q)。 \qquad (2)$$

于是，现在就可以依据数据Q计算出被研究客体r的数值A，这也就是课题的解。这时，根据数据Q可对客体r做出$A=\Phi(H)$的结论，这个结论是从该客体的特征$h_i\in H$得出的，尽管这些特征我们还不能直接认识到。

我们来探讨在求解目标课题时，在训练资料中缺少关于特征$h_i^{(0)}$中的某些特征的具体数值资料，以及在$q_i^{(0)}$和q的数值中存

在有噪音（测量误差和其它随机干扰）的情况。我们同样假定，在为精确地求解课题所必需的数据q_i中，某些数据在集合Q中可能是没有的。

设：H^*是特征h_i的集合，其资料将被用于求解既定课题。根据条件，$H^* \subseteq H$。目标课题的最佳解是可以求得的，其条件是，在前述方程组中，将函数Φ用它的最佳估计值Φ^*（据所利用的数值$h_i^{(0)}$求出），而前述算子F须用它的最佳近似算子F^*代替之。我们用A^*表示求得的目标课题的解。A^*就是解A的最佳估计值。

按照本文的议题，所谓的最佳估计值，在此就是一些量、函数和算子的估算值，而且必须是能对目标课题的解A给出最大信息的估计值（假定在计算时没有信息量的损失）。因此，在这里就要谈谈对于信息量的优度问题。

为了得到函数Φ的最佳（按信息量说）估计值Φ^*，必须对函数Φ作一变换，使之与未知特征h_i无关，同时又与其余特征$h_i \in H$保持既定的相关性。在这种条件下，变换应当这样进行，即要使$M(\mathscr{L}(\Phi^*,\Phi))$的数值最小。而$M(\mathscr{L}(\Phi^*,\Phi))$就是在与目标课题相应的函数空间中函数$\Phi^*$和$\Phi$的距离的数字期望。

上述将目标函数进行变换使之与未知特征h_i不发生关系的转化是可以实现的，其条件是把因客体$r^{(0)}$而异的h_i值去掉，而代之以经过专门选择的、对所有客体都一样的、可满足所谓优度条件的$\bar{h}_i$值。

$\bar{h}_i$值的估计值的计算，应当使用所有现存的关于相应特征h_i可能数值的分布的资料。

在我们看来，在一种极端的情况下，即当缺乏对于某个量在某一可能的数值区间内的分布情况的资料时，似乎就可以先验地认为该量在这个区间内是均匀分布的。如果在获得确切的资料之前就需计算估计值，也应该认为该量的分布是均匀的（如所周知，均匀分布的熵是最大的，若使用非均匀分布的资料，熵就减

小)。由此可见，$\overline{h_j}$值❶原则上在任何时候都是可以估计出来的。

$\overline{h_j}$值可以视为相应的可能数值h_j的某种平均值。因此，为了简便起见，可通过用$\overline{h_j}$值代替未知特征的方法来消除后者的变异性对函数Φ的影响。这种方法就称为求相应特征（参数）的平均值法。

$\overline{h_j}$值的选择和对距离ρ的概念下定义，可以成为一项独立研讨的科目。但是，在实践中遇到的许多情况下，可以使用下述显而易见无须论证的原则。

在许多课题中，目标函数Φ只是表达一种对比关系，即在一些客体r中，对于什么样的一个客体来说Φ'值将是较大的，Φ'就是各特征h_j的一定的对比关系。如果某个特征h_j在函数$\Phi'(H)$中作为被加数或被减数出现，那么要使函数Φ与这个变量的变化无关，就可以通过用0值作为$\overline{h_j}$值代入的办法来实现。如果变量h_j在函数$\Phi'(H)$中只作为乘数或除数出现，那么要使函数Φ与该变量的变化无关，就可以通过用任何一个不使函数Φ'的符号改变非零值作为$\overline{h_j}$值代入其中的办法来实现。在这中间，1是一个的可以采用的适当数值。

例如，我们假设：$h_1(r)$——某个客体r中的矿石储量（重量）；$h_2(r)$——这个客体的产出深度；$h_3(r)$——单位重量矿石的出售价格；这时要对两个客体r_1和r_2作对比，建立起了一个指出最有价值客体的目标函数，其形式如下：

$$\Phi(H)=\begin{cases} r_1, \text{如果} \Phi'(H(r_1))\geq\Phi'(H(r_2)) \\ r_2, \text{如果} \Phi'(H(r_1))<\Phi'(H(r_2)) \end{cases} \qquad (3)$$

式中，$\Phi'(H(r))=h_1(r)\cdot h_2(r)/h_3(r)$；$\Phi(H)=r_1$和$\Phi(H)=r_2$意味着客体$r_1$或$r_2$相应地为最有价值的客体。

很明显，如果在客体$r=r_1$和$r=r_2$时$h_3(r)$值都是未知的，就应该用量$\Phi'^*(H(r))=h_1(r)\cdot h_2(r)/\overline{h_3}(r)$来代替量$\Phi'(H(r))$，式中，$\overline{h_3}(r)$—由$h_3(r)$的可能数值求得的平均值。这个数

❶ 原文为$\overline{h}$值，似为$\overline{h_j}$之误——译注。

值对于两个客体r_1和r_2来说是一样的。根据公式(3)，在用数值1代替$\overline{h}_3(r)$的值时，即在使$\Phi^*(H(r))=h_1(r)\cdot h_2(r)$时，我们也会得到同样的结果（选择客体$r_1$或$r_2$）。既然这种代换（按所获结果来说）与求平均值法是等效的，我们就可把这种代换用“求平均值”的术语称呼之，在此这样做是不会引起误会的。

在被求平均值的参数的数目等于零时，函数Φ就可被视为函数Φ^*的实际的局部表达式。下面我们将使用这种办法，只要这样做合适，我们就用符号Φ^*来表示所有上述情况下的目标函数。

在用上述方式得到函数Φ的最佳估计值Φ^*之后，我们就按上面指出的办法针对每一个客体$r^{(0)}$计算$A^{*(0)}$的估计值：

$$A^{*(0)}=\Phi^*(H^{*(0)}) \qquad (4)$$

与上面引用的符号类似，在此，$H^{*(0)}$——在特征$h_j\in H^*$的训练资料中拥有的数值$h_j^{(0)}$的集合$H^{(0)}$，也就是说$H^{*(0)}=H^{(0)}\cap H$。

现在，我们来根据$A^{*(0)}$的这些数值和集合$Q^{(0)}$来求包含有参数G^*的集合的算子F^*，该算子应是在压抑噪音的同时能最佳地表达出对应关系

$$A^{*(0)}=F^*(G^*,\ Q^{(0)}) \qquad (5)$$

的算子。

应指出的是，算子F^*可以是解狭义反演问题时使用的函数、回归函数，或将客体划分为有限数目的类别时的判定规则。这个算子是算子F的估计值，它与算子F的区别在于对集合H中的参数求了平均值，因为算子F^*是通过数值$A^{*(0)}$确定的，而$A^{*(0)}$则是根据含有集合H^*的公式（4）计算出来的。因而，如果有了可明确给出目标课题的解A的算子F，并且知道求参数平均值的运算对解A的影响，就可根据算子F立即建立起算子F^*，而无须再根据公式（4）和（5）进行计算。

换句话说，如果算子F的一般形式是已知的，但某些具体的量（加入该算子的参数、函数和泛函数）是未知的，就可以通过对这些未知量求平均值的办法将算子F直接转换成算子F^*。我

们注意到，根据观测结果对任何函数φ所作的精度最佳的近似，可视为求未知量平均值的同义术语。实际上，在这种近似中被查明和被利用的是所研究量的分量，即其信息含于观测结果中的那部分分量，而且运算的进行是以使数值$M(\mathscr{L}(\varphi,\varphi^*))$达到最小为目的的。在表达式$M(\mathscr{L}(\varphi,\varphi^*))$中，$\varphi^*$——任意一函数$\varphi$的近似值，而符号$M$和$\mathscr{L}$的意义与上面在引入求目标函数的参数平均值的概念时所指出的意义相同。因此，下面我们将使用“求参数平均值”这一术语，在这个概念里就包含有上述那种近似。

在提出目标课题的同时直接确定出算子F的情况，在实践中是颇常遇到的。这从下边选择方法的例子中将反映出来。因此我们规定某些似乎无须证明、几乎一看便知的规则，它们在许多情况下对上述那些转换都是适用的。

1．如果关于某个函数的资料是不充分的，因此对于该函数近似方法的选择也是不确定的，这时，就可在能对现有资料作令人满意的近似的诸方案中选择一种最简单（自由参数数目最小）的作为最佳的函数近似方案。这个规则相当于第三秩算法(алгоритм третьего ранга)，后者在杜鲍夫的一篇文章（1974б）中及本文集的另一篇关于近似法的论文中都比较详细地论述过。由此规则可引出进一步探讨所必需的下述规则。

2．我们假设在算子F中包括有可约量（在量纲上可作对比而在所取数值上可能不同的参数、函数、泛函数）的比较。这时，如果没有说明它们的差别，在获得确切资料之前，就应该在运算时认为这些量的数值是一样的。例如，在下述方案中就需要使用这一规则：如果在函数F中含有比值$P(R_1)/P(R_2)$，那么在没有关于事件$r\in R_1$和$r\in R_2$的概率的资料时，在建立函数F^*时就应该取$R(R_1)/P(R_2)=1$。

3．如果在算子F中含有某个函数$p(Q/r)$的反对称的量，那么在缺少关于该量的资料时，我们就认为反对称取正值和负值的概率相等。这时，反对称的可能数值区间两端的绝对值应当相等，而反对称的平均值就应当取作零。由此又引出后边将要使用

的下述规则，

设：仅仅知道$Q_1^{(0)}$——客体$(r_1^{(0)}\in R_1$的标志，$Q_2^{(0)}$——客体$r_2^{(0)}\in R_2$的标志，Q——被研究（拟分类）客体r的标志，而且在标志空间中已确定了距离$\mathscr{L}$❶。这时，在取得精确的资料之前，不等式$\mathscr{L}(Q,Q_1^{(0)})<\mathscr{L}(Q,Q_2^{(0)})$就意味着不等式$P(R_1|Q)>P(R_2|Q)$（相反的不等式亦有类似的关系）。

实际上，由于缺少关于$p(Q|R_1)$和$p(Q|R_2)$的分布律的反对称数值的资料，根据规则3就应该把它们先验地视为对称的定律。而根据规则1，这些定律应该近似地看作为单模态的定律，其最大值在$Q_1^{(0)}$和$Q_2^{(0)}$点上。此外，根据规则2可取$P(R_1)=P(R_2)$。这时，正如从几何上不难看出的那样，随着点Q接近于点$Q_1^{(0)}$，或者接近于点$Q_2^{(0)}$，$p(Q|R_1)$和$p(Q|R_2)$的数值是单调而同等程度增大的。因此，如果点Q离点$Q_1^{(0)}$比离点$Q_2^{(0)}$要更近一些，则$p(Q|R_1)>p(Q|R_2)$，反之亦然。要证明的就是这一点。

这样一来，利用这种或那种构成算子F^*的办法（通过Φ^*和$A^{*(0)}$的数值，或者通过对参数求平均值和对算子F中的函数进行近似），我们便得到一个求解所提出目标课题的明显最佳方法。与表达式（2）相类比，在算子F^*已经求得的条件下，这个解可写成如下形式：

$$A^*=F^*(G^*,\ Q) \qquad (6)$$

根据建立表达式（6）的条件，客体r的解的估计值A^*，反映着有关该客体的特征h_j的信息，尽管我们并不直接知道$h_j(r)$的数值。该信息包含于数据Q之中，可在将被研究客体r与训练资料所描述客体$r^{(0)}$作类比的基础上把它抽取出来。这就意味着，在表达式的任何构成中，A^*的估计精度都取决于客体r与客体$r^{(0)}$的相似程度，或者说取决于训练资料在多大程度上与拟处理资料相

❶ 标志空间距离$\mathscr{L}$的选择问题已超出了本文探讨的范围。我们只提一下，它与拟解的目标课题和关于不同变量q_l的变异特点的概念有依从关系——原注。

一致。因此，将来最好专门研究一下这种一致性的估计问题，本文对此不加探讨。

在算子F^*的表达式集合中和与之相应的估计值A^*的表达式集合中，包括有算子F及与之相应的解A，它们是在$\Phi^*=\Phi$和没有噪音（计算误差和自然干扰）的条件下实现的个别表达式。

方法的选择

选择应在K个方法的有限集合中进行。该集合保存在人们的记忆中，也以专门选配好的文件等形式储存在计算机的存储器中。

对于处理数据Q的每一种方法（第k个方法），我们都可以与等式（6）类比写出：$A_k^*=F_k^*(G_k^*, Q)$。在这里，我们现在把F_k^*理解为这样一个算子，它不仅包含目前普遍使用的表述方法的通用运算，而且也包含着求解目标课题所必需的补充运算。例如，我们假设，某个分类方法被用来求解应当在客体r上进行何种工作的问题。这时，算子F_k^*既包含着对客体所属类别名称的说明，也包含着对被推荐工作种类的解答，但在现时的描述中，后者只是以不明显的形式暗示出来。与算子F_k^*相对应，现在A_k^*表示的是用第k个方法（有时进而将它称为方法F_k^*）求得的解A的估计值。与上类似，我们用$H_k^{*(0)}$、Φ_k^*等符号来表示用方法F_k^*研究的特征$h_i^{(0)}$的集合、目标函数的相应估计值，等等。

任何两种方法F_{k1}^*和F_{k2}^*，如果将计算误差忽略不计，它们可以对任何一个客体r取$A_{k1}^*=A_{k2}^*$，我们就称它们为求解既定课题的等效（按信息量说）方法。这时，$H_{k1}^*=H_{k2}^*$，$\Phi_{k1}^*=\Phi_{k2}^*$，但这两个方法在运算上，在上面提到的与目标课题有关的补充运算的制约条件上，以及诸如此类的方面，是有区别的。

很明显，在存储器中只应当保存最低数目的方法，这些方法须是求解可能有的目标课题所必需的，同时须符合一定的经济要

求（某些方法在精度方面不如其它方法，但保存在存储器中也是有益的，因为在某些条件下使用它们生产效率较高，而成本较低，等等）。在一定阶段上，为了简便起见，我们将假设方法选择应在两两不等效的方法中进行。

目标课题一经提出之后，就可以对求解此课题所要采用的方法进行划分。这些方法须满足三个条件。

第一个条件是理之当然的：Q和$Q^{(0)}$应该能转换成可被方法F_k^*利用的形式（例如，它们应含有样品中化学元素的浓度值，并标明采样点的座标，如果它们是算子F_k^*的运算对象的话）。

第二个条件是显而易见的：应当满足$H_k^{(0)} \subseteq H^{(0)}$的要求，因为在相反条件下，训练资料对于方法$F_k^*$的使用来说就是不充分的了。

第三个条件：函数Φ_k^*或算子F_k^*应当是可以通过求参数平均值的办法从目标函数Φ或算子F求出来的，因为这些参数资料是不能被方法F_k^*直接使用的。在相反的情况下，用方法F_k^*求出的不是既定课题的解，而是另外某个目标课题的解。如果没有两两等效的方法，以及在上述关于函数Φ_k^*和算子F^*的条件下，剩下的可供选择的方法在集合H_k^*上就会是不一样的。

要从一个有K个方法的被研究（有限）集合中选出一种方法，可以通过将这些方法两两对比来选定。因此，下面为了简明起见，我们把问题限定在只对两种方法F_{k1}^*和F_{k2}^*的对比上。

设：$\mathscr{E}_A(H_k^*)$——包含于集合H_k^*中的决定解A的集合H的信息。这时，按照上面确定的根据信息量来选择方法的原则，如果$\mathscr{E}_A(H_{k1}^*) > \mathscr{E}_A(H_{k2}^*)$，对于求解既定的目标课题来说，方法$H_{k1}^*$就比方法$F_{k2}^*$要好一些。实际上所探讨的两个方法就条件来说都是适用的，不等式意味的是方法F_{k1}^*比方法F_{k2}^*有更多的信息。

探讨对比$\mathscr{E}_A(H_k^*)$值的一般原则不是本文的任务，但我们请读者注意$\mathscr{E}_A(H_k^*)$值的下列明显的性质，它们在颇广的情况下均可利用。

如果$H_{k2}^* \supset H_{k1}^*$，在其它条件相同时，显然$\mathscr{E}(H_{k2}^*) > \mathscr{E}(H_{k1}^*)$。

实际上，为了估计出A值，在这时方法$F^*_{k_2}$使用的资料与方法$F^*_{k_1}$是一样的，不过它还使用了方法$F^*_{k_1}$不使用的资料$H^*_{k_2}\backslash H^*_{k_1}$的集合。

如果N_{k_1}和N_{k_2}——分别为方法$F^*_{k_1}$和方法$F^*_{k_2}$所使用的特征$h_j{}^{(0)}$的数目，在其它条件相同时，如果$N_{k_2}>N_{k_1}$，则$\mathscr{E}(H^*_{k_2})>\mathscr{E}(H^*_{k_1})$。实际上，如果在用方法$F^*_{k_1}$和$F^*_{k_2}$求解时反映出来的客体$r$的所有特征$h_j$先验地是等价的（就其带有的关于解$A$的信息而论），那么，那个能把较多数决定解$A$的特征的变化情况反映出来的方法，就是较好的方法。

下述性质常常也是有用的：对计算客体$r^{(0)}$的函数$\Phi^*_k(H^{*(0)}_k)$时所使用的因子h^0_j估计得越准确，在其它条件相同时，这些因子的熵（不确定性）就越小，相应的$\mathscr{E}(H^*_k)$值就越大。这个性质表达出了熵与观测信息和观测误差之间人所共知的关系：观测误差越大，熵也越大，所获结果的准确性和信息量就越小。

令$\sigma(h_j)$为特征h_j的统计权，它取决于该特征对A值影响的大小和对此特征估计精度的高低。在变量h_j是一些互不相关的因子时，就可以认为是$\mathscr{E}_A(H^*_k)$与量$\sum_{j=1}^{N_k}\sigma(h_j)$是成比例的，这里$h_j\in H^*_k$。这个规则不仅在使用特征$h_j$的直接估计值时，而且在使用它的间接估计值时，都可以采用。这里所说的间接估计值，其中包括能以某种精度反映出客体中矿石数量的地球化学场的统计参数（这类客体应是与该地球化学场有关的）。

所谓最佳方法，就是在所有可能的已知或未知方法中相对最好的方法。按照定义，它应该具有最大可能的$\mathscr{E}_A(H^*_k)$值，即就是说，这种方法应该使用对变量A的估计有意义的、以$h_j{}^{(0)}$值的形式含于训练资料中的特征h_j的全部集合。

这种方法应该与明显的最佳方法F^*是等效的，关于后者在前面一节中已经谈及。因而，应当有$H^{*(0)}_k=H^{*(0)}=H^{(0)}\cap H$。

最后面的条件意味着，随着包含于训练资料中和参加于H中的信息减少，信息量越来越少的方法就会变成最佳方法，反之亦

然。

实　例

我们假设，要对残留晕的地球化学采样数据进行处理（数据为成矿元素浓度的集合及与之相应的采样点的座标），而且需要从下列四种方法中作一选择。

F_1^*——求解上面指出过的狭义反演问题的方法，其结果以下列各量的数字估计值的形式输出：h_1—被疏松产物覆盖的矿体露头上的成矿元素浓度的平均值；h_2——这个露头的平均视厚度；h_3——露头的延伸长度；（即就是说，$s=h_1h_2h_3$——矿体露头的金属量。）

F_2^*——回归法，这种方法可以根据在已知客体上作的观测，估计出任何一个特征h_i与任意标志q_i的集合的相关性。

F_3^*——以在已知客体上所作的类似观测为基础，把数据集合（按照它们可能属于工业客体和非工业客体）划分成R_1^*和R_2^*两类的方法。

F_4^*——把数据集合划分成在每个类别内部具相对同类的数据、即自身相对近似的两类的方法。在此提示一下，这种方法是所谓“无训练”方法的一个典型代表。

我们假设，一个最小的目标课题是对疏松产物之下的可能矿体的露头作评价，以便进而选择出最有工业开发价值的矿体。这个课题可通过下式来定义：

$$\Phi(H)=\begin{cases}s, & \text{如果 } (s\geqslant s_m)\wedge(h_1\geqslant c_m),\\ 0, & \text{如果 } (s<s_m)\vee(h_1<c_m),\end{cases} \qquad (7)$$

式中，$\Phi(H)=s=h_1h_2h_3$就意味着承认客体是有工业意义的，其意义与$h_1h_2h_3$的乘积是成比例的（从统计上说，该乘积又与矿石储量是大致成比例的）；$\Phi(H)=0$——就意味着将客体鉴定为不具工业意义的；s_m和c_m——当客体被认作为推测的工业矿体时，它们分别为矿体露头金属量的最低值和成矿元素浓度的最低值。

现在我们假定，训练资料包括针对被研究元素的残留晕而求解前已论及的反演问题的算法，以及这种晕的金属量与矿体露头金属量的对应表格。

这时，F_1是最优的方法。事实上，这种方法宜于用来在已有的训练资料的条件下求解既定的目标课题，而且是用被研究客体所有特征h_i的信息来求解A的。在此，算子F_1^*表现为根据针对残留晕求解反演问题的公式来求得这些特征的估计值h_i^*，进而按下式算出解：

$$A=\begin{cases}s^*，\text{如果 } (s^*\geqslant s_m)\wedge(h_1^*\geqslant c_m)，\\ 0，\text{如果 } (s^*<s_m)\vee(h_1^*<c_m)，\end{cases}$$

式中，$s^*=h_1^*h_2^*h_3^*$。

方法F_2^*的信息量较小，因为在给定的训练资料的条件下，它的结果只给出$s=h_1h_2h_3$的估计值，而不是h_1、h_2、h_3的个别的估计值。换言之，方法F_2得出的是含有关于矿体露头金属量信息的资料，但并不给出在这个露头上达到没有达到有用组分工业富集的信息。

方法F_3的信息量要更小一些，它给出的不是乘积$s=h_1h_2h_3$的数字估计值，而仅仅是指出不等式$s^*\geqslant s_m$成立或不成立。然而在求解给定的目标课题时，重要的是须从具备$s^*>s_m$的诸矿体中把具有尽可能大的s值的矿体划分出来。

最后，方法F_4^*所具有的信息量是最小的。实际上，它和方法F_3^*一样，是把一些客体r的集合分成两类。然而在方法F_3^*的判定规则中是直接使用关于各客体特征的信息，这种信息从目标课题的观点看来是重要的信息（针对$r^{(0)}\in R_1$或$r^{(0)}\in R_2$的客体，不等式$s^{(0)}\geqslant s_m$相应地成立或不成立）。在方法F_4^*中，分类并非是在直接考虑这种信息的情况下完成的。在这里，对客体集合所作的划分，依据的是它们的特征的相对近似程度，这种划分取决于地质因素和所采用的特征空间的量度，它可以与根据经济要求（这些要求与目标课题和现有训练资料是一致的）所做的划分是不一致的。

由此可见，在所研究的条件下，方法F_1^*是最优的。

当上述训练资料减少时，规则就要有所变化。如果在训练资料中缺少了针对被观测元素的残留晕解反演问题的算法，那么方法F_1^*就变成不能采用的了，而方法F_2^*便成为最优的方法。此外，如果金属量 s 的值在训练资料中不是以数字给出的，而是仅仅指出它满足于不等式 $s \geqslant s_m$或者 $s < s_m$，那么方法 F_2^* 也就变成不能采用的了，方法F_3^*就成了最优方法。最后，如果完全没有理论资料，也完全没有在所研究的那种客体上作过观测，方法F_4^*便成为最优方法。采用这种方法之所以能得到好处与下述情况有关：在其余条件相同的情况下，在根据某种标志（例如成矿元素的浓度）的同类性原则对客体集合进行划分以后，还可进一步根据另一种标志（例如工业价值）对客体再作划分，而且后一种划分的平均误差比起乱猜瞎碰还要小一些。这种情况可从地质上作如下说明。

在许多情况下，矿体具有相对分明的界线，并在成矿元素的浓度上与围岩有区别。在这种情况下，根据成矿元素浓度的相对同类性对采自不同地质体中的地球化学样品进行划分，就可做到基本上把采自无矿岩石的样品归入一类，而把采自矿体的样品归入另一类。在后一类中，也可能会有采自无工业意义的贫矿体的样品，但划分出此种类别毕竟会使工业矿化的寻找变得容易一些。

同时，也存在着这种分类法导致重大错误的情况。许多矿体没有截然的地质界线。这类矿体的相当一部分在成矿元素的浓度上与近矿围岩是有差别的，但这种差别不如这部分矿体与绝无仅有的特富矿体或个别块段之间的差别大。在这种情况下，我们所探讨的这种对被采样客体作分类的方法，就会把许多工业矿体与无工业意义的岩石归入一类。自然，在根据相对的同类性原则和根据直接与既定目标课题有关的重要标志划分类别时，类别之间的界线也可能出现另一种情况。

由此可见，只是在训练资料中含有的信息对于采用上述各方

法是不充分的时候，方法 F_1 才是最优方法。在这种情况下，人们假定由方法 F_1 确定的类别界线与给定目标课题的最佳界线是尽可能接近的（哪怕是对客体的某一部分来说也好），而且这种假定起着训练资料的作用。在上面已讨论过的条件下，要不使用这种方法就只好胡猜瞎碰，会酿成更大的谬误。

补充意见

许多采用了数学方法而未能奏效的情况，是因为数据处理方法选择得不正确。例如大家所知道的，“无训练”的方法在无论有没有关于特征 $h_j^{(0)}$ 的数据的情况下都被采用着：颇为流行的“直观推断”法不要求事先对现有训练资料的完备性及其使用的正确性作检验。同时，某些研究者认为，不对特征 h_j 作统计，不顾及这些特征的数值，得出的结果也是颇为精确的。现在已很清楚，这种办法是不对的。地质工作者如若能更多地（哪怕是不自觉地）利用关于因子 h_j 的影响的信息，尽管由直觉想像的估量是粗略的，也可能得到比较接近于真实的结果。

应该注意这样一个问题，对已有的训练资料考虑不周，就意味着会产生本来可以避免的要对（对该课题非常重要的）参数求平均值的问题。这种求平均值的做法，不仅能使拟计算特征 h_j 的数目减少，而且也会使对这些特征的变化性作近似时的精度降低。

当对某课题甚为重要的一些特征 h_j 属于一个连续的集合时，人们却采用了划分出几个离散类别的分类方法，这是选择数据处理方法时颇为常见的一种错误。

类似的错误在上面提到的把地球化学和地球物理异常的集合划分成推测的“矿致异常”和“非矿异常”时也会出现，这时，目标课题就是要根据储量值对异常作分类。实际上，储量值是一个连续的值，即可以作为数值的无限集合来看待的值。而上述分类法相当于把这样一个集合只表达成两个数值，分别以推定的

“矿致异常”和“非矿异常”两个词来代表之，这就意味着在一些很宽的数值区间内作了平均。因此，据上所述，当具备必要的信息时，在这种情况下利用储量数值具连续性的特点，采用（广义的）回归分析会更优越一些。而为了利用这种特点却得不到信息的情况十分少见。这是显而易见的，因为只要具备观测值Q，哪怕它们是在两个具备所必需的已知特征的客体$r^{(0)}$上观测到的，也可完成最简单的线性回归分析。当然，当客体$r^{(0)}$的数目更多时，还可以采用较复杂的最佳回归分析。

如果客体类别在h_j的数值空间中实际上具有分明的界线，当要求研究者对这种客体作划分时，就可看到另一种情况。例如，当从褐铁矿铁帽中划分出黄铁矿矿床的铁帽时，就会产生这样的课题（沃斯特罗克努托夫，1973）。这时的课题不再是按照储量数值作划分，而是按照特点不同、化学元素浓度组合迥异的铁帽类型来作划分。在这种情况下，甚至用最简单的算法也会得到良好的分类结果。

我们要指出，对于采自不同类别地质产物的样品来说，被确定用于计算的不同元素浓度值的组合可能是很不相同的。而每种单个元素的浓度可能就没有这么明显的差别。单个元素的浓度在每个类别中都有变化，在不同的类别中会有重迭。在这种情况下，如果用上面论及的数学分类法效果就很好，而使用直观作图法，就只能局限于将浓度作直接对比，可能会含有许多重大错误。

方法的论证

按照普遍采用的直观概念，如果有证据表明，某个方法在某些确定条件下其优越性不亚于任何已知的其它方法，那么这个方法就是有依据的。

正如在前言中已提及的，许多现在正被采用的数据处理方法，由于对其必需的原则研究得不深入而没有加以论证。现在，

作这种论证的可能性具备了：根据本文前一节阐述的原则，可以针对不同的目标课题和训练资料就各方法的优度作对比。如果在对比时确定，当目标课题和训练资料构成某种组合时，所探讨的方法不比任何其它在其之前采用的方法差，那么，证明该论点的正确性并指出该方法的使用条件，也就是对方法的论证。

显然，若有证据表明，在一定训练资料的条件下，所探讨的方法对解一定的目标课题是最佳的方法，那就更好了。实际上，这样的证据是指出一种条件，在这些条件下，被论证的方法不仅不亚于现有的方法，而且也不亚于将来可能研究出来的方法。其实，这样的论证在某种意义上说是比较简单的。只要确定在给定的目标课题和一定的训练资料下，所探讨的方法 F_k^* 与按照“目标课题的求解”一节中叙述的规则建立起来的明显最佳算法 F^* 是等效的，也就足够了。

这就意味着，给出某个目标函数 Φ 和训练资料 $H^{(0)}$ 之后，就需要在该函数中对那些在所探讨方法的相应函数 Φ_k^* 中没有的参数 h_j 求平均值。如果求平均值之后得出的函数 Φ^* 与函数 Φ_k^* 是等效的，那么，只要训练资料中存在有和仅存在有符合等式 $H^{(0)} \cap H = H^{*(0)}$ 的资料，方法 F_k^* 对于解给定的目标课题来说就是最佳的。当在 $H^{(0)}$ 中没有含于 $H_k^{*(0)}$ 中的资料时（这时方法 F_k^* 不适用），或者当存在包含于 H 中的特征 h_j 的资料但方法 F_k^* 并不使用它的时候，上述等式都不能满足。在后一种情况下，方法 F_k^* 不是最佳的，因为它没有把现有的全部重要信息都用上。例如函数 Φ^* 和 Φ_k^* 的上述等效性未能实现，无论训练资料如何，方法 F_k^* 对于求解给定的目标课题就都不是最佳的，因为如上节所述，这时它给不出解 A 的估计值。

这里提醒一点，不进行函数 Φ 向函数 Φ^* 的转换，而是把算子 F 转换成算子 F^*，也是可以的。

如果用这种办法对一个所有可能目标课题的集合进行研究，那么就可把方法 F_k^* 最佳性的范围充分确定下来。这时，如果所有目标课题之间的规律性联系已被揭示，就不必冗繁地把集合中

的所有课题都研究一遍。例如，下面这种做法就是可行的，如前所述，一些不同的目标课题可以看成是一个总课题的局部情况。今后如有可能，我们将对这个问题进行适当的专门研究。但是，如果要对方法只进行最必要的论证，仅要求对一个目标课题指出它的最佳性条件，那么以上边列举的规则为依据通常也就可以进行了。下面我们将举例说明此点。

我们再指出一个在许多情况下可简化方法论证的显而易见的规则。

我们假设，已经有证据说某个方法$F^*_{k_1}$是处理属于某一定集合的数据的最佳方法。这时，另外的任何一种方法$F^*_{k_2}$只是在处理该集合的某个子集的数据时才是最佳方法，因为仅仅在此子集合内方法$F^*_{k_2}$才与方法$F^*_{k_1}$是等效的。这个规则的应用问题，将在实例2和3中说明。

实例 1：

我们先来探讨一种几乎人人熟知的情况——论证根据最大似然性原则作分类的方法（为简便起见，下面把这种方法称为最大似然法），对此方法进行论证的途径与已有的论证是相近的。

应当指出，用最大似然法作分类，就意味着已经针对已知类别R_1和R_2确定了连续函数$p(Q|R_1)$和$p(Q|R_2)$，并在训练资料中给出了这种函数。$p(Q|R_1)$和$p(Q|R_2)$就是在客体$r\in R_1$和$r\in R_2$中出现Q值的概率密度。令R^*_1为分类后被认为是属于类别R_1的客体r的集合。实际上，由于分类的错误，属于类别R_2的客体r也会以某种比率落到该集合中。因此，类别R^*_1与类别R_1并不相合，只与它近似。与此类似，我们用符号R^*_2来表示与类别R_2近似的类别。在描述分类方法时，通常并不对类别R和R^*引入不同的符号，但在我们讨论的课题中，采用不同符号使之更加严谨是恰当的。这时，用最大似然法进行分类，就是根据$p(Q|R_1)/p(Q|R_2)>\mathscr{K}$或者$p(Q|R_1)/p(Q|R_2)<\mathscr{K}$，相应地得出$r\in R^*_1$或$r\in R^*_2$的结论。这里，$\mathscr{K}$是先验地给定的某个数值（风险函数）。选择$\mathscr{K}$值的算法不属于最大似然法的内容。$p(Q|R_1)$

$/p(Q|R_2)=\mathscr{K}$，就意味着可以不带偏见地把客体 r (Q)归入 R_1^* 或R_2^*中（这是一种很常见的情况，讨论就到此为止）。

以上所述可表示如下

$$F_k^*=\begin{cases} r\in R_1^*, & \text{如果} p(Q|R_1)/p(Q|R_2)>\mathscr{K} \\ r\in R_2^*, & \text{如果} p(Q|R_1)/p(Q|R_2)<\mathscr{K} \end{cases} \tag{8}$$

这就是最大似然法的算法（算子）。

我们假定，为了寻找最大似然法的最佳性条件，选择了下述目标课题：求解客体 $r=r(Q)$ 应当属于哪一个类别（R_1^*抑或R_2^*）。这一课题可用下列习见的目标函数来表达：

$$\Phi=\begin{cases} r\in R_1^*, & \text{如果 } r\in R_1, \\ r\in R_2^*, & \text{如果 } r\in R_2。 \end{cases} \tag{9}$$

在此，特征集合H就是可以决定任一已知客体 $r^{(0)}$ 属于类别R_1或类别R_2的任一些标志的集合。

现在，为了得到算子F必须寻找在（9）式中指出的Φ值与$Q=Q(r)$❶值之间的最佳对应关系。目标课题的提出即意味着：如果错误的数目最小，即当$r\in R_2$时将 r 错误地归入R_1^*的数目最小，或者当$r\in R_1$时将 r 错误地归入R_2^*的数目最小，该目标课题就是以最好的方式完成的。按照概率论，这就意味着必须取

$$F=\begin{cases} r\in R_1^*, & \text{如果} P(R_1|Q)/P(R_2|Q)>1 \\ r\in R_2^*, & \text{如果} P(R_1|Q)/P(R_2|Q)<1 \end{cases} \tag{10}$$

式中，P——概率符号，$P(R_i|Q)$——当客体 r 中存在有标志Q时，事件 $r\in R_i$❷的概率。结论$F=r\in R_1^*$或者$F=r\in R_2^*$即就是解A，它是用客体 r 属于类别R_1的概率比属于类别 R_2 的概率要大（或者相反的关系）的形式表示出来的。常见的情况$P(R_1|Q)$

---

❶ 在这种情况和其后将论及的情况下，既便不指出函数Φ，算子F也是显而易见的。但我们指出它来是为了举例说明所叙述的这种方法，以便在后面所选用的另一些更复杂的情况下也可以说它是正确的——原注。

❷ 原文为R_i，似有误，应为事件$r\in R_i$的概率——译注。

$/P(R_2|Q)=1$意味着客体 r 属于类别R_1和R_2 的 概率相 等，就不再予以探讨了。

根据人所共知的贝叶斯公式：

$$P(R_1|Q)=P(R_1)\cdot p(Q|R_1)/p(Q),$$
$$P(R_2|Q)=P(R_2)\cdot p(Q|R_2)/p(Q), \qquad (11)$$

式中，$P(R_1)$，$P(R_2)$——相应为事件 $r\in R_1$ 和 $r\in R_2$ 的无条件概率；$p(Q)$——获得数据Q的无条件概率密度。由此推得，

$$P(R_1|Q)/P(R_2|Q)=P(R_1)\cdot p(Q|R_1)/[P(R_2)\cdot p(Q|R_2)] \qquad (12)$$

把这个表达式代入算子(10)，我们便得到下式：

$$F=\begin{cases} r\in R_1^*, & \text{如果} p(Q|R_1)/p(Q|R_2)>\mathscr{K} \\ r\in R_2^*, & \text{如果} p(Q|R_1)/p(Q|R_2)<\mathscr{K} \end{cases} \qquad (13)$$

式中，

$$\mathscr{K}=P(R_2)/P(R_1)。 \qquad (14)$$

如上所述，为了阐明方法F_k^*的最佳性条件，只 要说 明 它在什么时候与算子 F或者它的估计值F^*是等效 的 就足 够了。我们来研究几个方案。

方案 1　如果对表达式（8）和表达式（13）取一样的由公式（14）表达的$\mathscr{K}$值，那么这两个表达 式就是一致 的。同时，为了使用表达式（8），必须具备$P(R_1)$、$P(R_2)$ 的数值和 函数 $p(Q|R_1)$、$p(Q|R_2)$（或者可用以代替它们的估计值）。

这样一来，在所提出的目标课题的条件下，在训练资料中存在有$P(R_1)$、$P(R_2)$ 的数值和函数$p(Q|R_1)$、$p(Q|R_2)$（或者它们的估计值，这些估计值可当作被估计数值本身来用），以及 $\mathscr{K}$ 值是按照等式（14）得出的时候，最大似然法就是最佳的。

方案 2　现在假设，在上述资料中的同一个目标课题中，在训练资料中只具备函数 $p(Q|R_1)$、$p(Q|R_2)$或者它们的估计值，而不具备关于$P(R_1)$、$P(R_2)$的数值的资 料。这时，按照求参数

的统计平均值的规则，如果取$P(R_2)/P(R_1)=1$的话，算子F的最佳估计值F^*是可以得到的。根据等式（14），这就意味着必须取

$$\mathscr{K}=1。 \tag{15}$$

因而，在给定的目标课题和$\mathscr{K}=1$的条件下，如果训练资料中只含有函数$p(Q|R_1)$、$p(Q|R_2)$或它们的估计值，最大似然法就是最佳的。在具备关于$P(R_1)$、$P(R_2)$的数值的资料时，最大似然法的这个方案仅仅在明确知道$P(R_1)/P(R_2)=1$时才是最佳的。

方案3　现在我们来探讨对比较复杂然而在实践中更为常见的目标课题寻找最佳性条件的问题。假设要选择能在客体r上得到最大好处的某种工作形式$J=J_1$或$J=J_2$。例如，待处理资料可能是小比例尺地球化学测量数据，工作J_1可能意味着须进行详细的采样或者补充的地球物理调查，而工作J_2可能意味着终止该地段的研究，放弃揭露已发现的异常而转入另一个地区。

我们也可假定，对于已知客体r来说，工作J_1只是在$r\in R_1$的情况下是有利的，而工作J_2只是在$r\in R_2$的情况下是有利的（例如在R_1和R_2分别为有工业意义和无工业意义的类别时就可能有这种情况）。

应当指出的是，在这种情况下，现在一般是先进行客体的分类，只能在此之后再决定在客体上应该进行什么样的工作。关于这种做法的缺点和转变为按工作形式进行分类的合理性问题，在收入本文集中的笔者的另一篇论文中业已论及。这篇论文谈的是秩近似和效益估计问题。下面，在回答所提出的主要问题的同时，也顺便特意证明一下上述传统的两阶段求解法只是在一定条件下才可能是最佳的。

令$a_{J_1}(R_1)$和$a_{J_1}(R_2)$相应为在已知客体$r\in R_1$和$r\in R_2$上进行工作J_1所得好处的数学期望；$a_{J_2}(R_1)$和$a_{J_2}(R_2)$分别为针对工作J_2表示的同类数值。在一般情况下，这些数值可以是正值或负值（如果相应的工作使付出的代价大于收益，这些值便为负值）。

工作J_1只在$r\in R_1$的条件下是有利的，工作J_2只是在$r\in R_2$的条件下是有利的，这就意味着：$a_{J_1}(R_1)>a_{J_2}(R_1)$，$a_{J_2}(R_2)>a_{J_1}(R_2)$，$a_{J_1}(R_1)>a_{J_1}(R_2)$，$a_{J_2}(R_2)>a_{J_2}(R_1)$。

根据定义，进行工作J_1和J_2带来好处的数学期望分别等于：

$$\bar{a}_{J_1}(r)=a_{J1}(R_1)\cdot P(R_1|Q)+a_{J1}(R_2)\cdot P(R_2|Q) \tag{16}$$

和

$$\bar{a}_{J_2}(r)=a_{J_2}(R_1)\cdot P(R_1|Q)+a_{J_2}(R_2)\cdot P(R_2|Q) \tag{17}$$

令$\max(\bar{a}_{J_1}(r),\bar{a}_{J_2}(r))$表示$\bar{a}_{J_1}(r)$和$\bar{a}_{J_2}(r)$数值中最大的数值。这时，与前面详细研究过的情况类比，不难看出，提出的新目标课题相应有算子

$$F=\begin{cases}J_1, & \text{如果}\bar{a}_{J_1}(r)=\max(\bar{a}_{J_1}(r),\ \bar{a}_{J_2}(r))\\ J_2, & \text{如果}\bar{a}_{J_2}(r)=\max(\bar{a}_{J_1}(r),\ \bar{a}_{J_2}(r))\end{cases} \tag{18}$$

式中，$F=J_1$和$F=J_2$分别为以选择工作J_1或J_2的形式表示出的解A。当$\bar{a}_{J_1}(r)=\bar{a}_{J_2}(r)$时，无论进行哪种工作（$J_1$或$J_2$）都一样（这里的附加条件下面就不再说明了）。

令$\Delta\bar{a}_{1,2}(r)=\bar{a}_{J_1}(r)-\bar{a}_{J_2}(r)$。这时，表达式(18)可改写成

$$F=\begin{cases}J_1, & \text{如果}\ \Delta\bar{a}_{1,2}>0\\ J_2, & \text{如果}\ \Delta\bar{a}_{1,2}<0\end{cases} \tag{19}$$

根据表达式（16）和（17），$\Delta\bar{a}_{1,2}=0$的情况意味着

$$\begin{aligned}a_{J_1}(R_1)\cdot P(R_1|Q)+a_{J_1}(R_2)\cdot P(R_2|Q)&=a_{J_2}(R_1)\cdot P(R_1|Q)\\&\quad+a_{J_2}(R_2)\cdot P(R_2|Q),\end{aligned}$$

或

$$\begin{aligned}&[a_{J_1}(R_1)-a_{J_2}(R_1)]\cdot P(R_1|Q)\\&=[a_{J_2}(R_2)-a_{J_1}(R_2)]\cdot P(R_2|Q)\end{aligned} \tag{20}$$

利用表达式（12），从等式(20)可推得

$$\frac{p(Q|R_1)}{p(Q|R_2)}=\frac{[a_{J_2}(R_2)-a_{J_1}(R_2)]\cdot P(R_2)}{[a_{J_1}(R_1)-a_{J_2}(R_1)]\cdot p(R_1)}$$

当$\Delta\bar{a}_{1,2}>0$时，经类比得

$$\frac{p(Q|R_1)}{p(Q|R_2)}>\frac{[a_{J_2}(R_2)-a_{J_1}(R_2)]\cdot P(R_2)}{[a_{J_1}(R_1)-a_{J_2}(R_1)]\cdot P(R_1)}$$

而当$\Delta\bar{a}_{1,2}<0$时，

$$\frac{p(Q|R_1)}{p(Q|R_2)}<\frac{[a_{J_2}(R_2)-a_{J_1}(R_2)]\cdot P(R_2)}{[a_{J_1}(R_1)-a_{J_2}(R_1)]\cdot P(R_1)}$$

这就意味着，由等式（19）定义的算子F也可以写成：

$$F=\begin{cases}J_1, & \text{如果}p(Q|R_1)/p(Q|R_2)>\mathscr{K}\\ J_2, & \text{如果}p(Q|R_1)/p(Q|R_2)<\mathscr{K},\end{cases}\qquad(21)$$

这里，

$$\mathscr{K}=\frac{[a_{J_2}(R_2)-a_{J_1}(R_2)]\cdot P(R_2)}{[a_{J_1}(R_1)-a_{J_2}(R_1)]\cdot P(R_1)}。\qquad(22)$$

既然有上面约定的条件，结果$r\in R_1^*$和$r\in R_2^*$相应地指选择工作J_1和J_2。因此，如果这个表达式中的$\mathscr{K}$值满足于等式（22），表达式（8）与算子（21）就是等效的。由此可以得出结论，在训练资料中具备$P(R_1)$、$P(R_2)$、$a_{J_1}(R_1)$、$a_{J_2}(R_2)$的数值和函数$p(Q|R_1)$、$p(Q|R_2)$，或它们的估计值时，如果用以分类的$\mathscr{K}$值满足于等式（22），在解给定的目标课题时最大似然法就是最佳的方法。

前面已提到，现在的普遍做法是，只在得出关于客体r属于何种类别的结论之后，才能参照经济依据对在该客体上进行的工作J作出判定。在考虑将由分类而选定的工作的经济特征a_J之前就用最大似然法进行分类，就意味着须使用满足等式（14）或（15）（依有无关于$P(R_1)$和$P(R_2)$值的资料而定）的$\mathscr{K}$值。若用这种办法来确定$\mathscr{K}$值，附加有当$r\in R_1^*$时选择工作J_1和当$r\in R_2^*$时选择工作J_2的结论的算法（8），显然与须按公式（22）确定$\mathscr{K}$的明显最佳算法（21）是不等效的。

因此，当具备关于$a_J(R)$值的资料时，上述颇为流行的分两个阶段作判定的方法不是最佳的。

实际上，根据表达式（21）和（22），量$a_{J_1}(R_1)$、$a_{J_1}(R_2)$、$a_{J_2}(R_1)$、$a_{J_2}(R_2)$与$P(R_1)$、$P(R_2)$可能有下述关系：进行工作J_1不仅在$P(R_1|Q)>P(R_2|Q)$时是适宜的，而且有时候当$P(R_1|Q)<P(R_2|Q)$时也是适宜的（若量$a_{J_1}(R_1)$足够大）。相反，进行工作J_2可能不仅在$P(R_2|Q)>P(R_1|Q)$时是适宜的，而且有时候

当$P(R_2|Q)<P(R_1|Q)$时也是适宜的（若量$a_{J_2}(R_2)$足够大）。按上面谈的办法，即先进行分类，然后再以分类结果为基础并参照经济依据来选定工作形式，在所讨论的情况下，如果$P(R_1|Q)>P(R_2|Q)$，总是应该选择工作J_1，如果$P(R_2|Q)>P(R_1|Q)$，总是应该选择工作J_2。

要指出的是，这些规则表达了有经验地质工作者的“健康的想法”，并使之更为准确了。这可以从一个简单的例子看出来。假设：Q——地球化学异常，r——与该异常有关的客体，R_1——工业矿体的类别，R_2——非工业客体的类别，J_1——为揭露客体r而进行的山地工作，J_2——放弃对异常作进一步研究。

这时，分阶段作判定的第一种办法意味着，在异常与工业矿体有关的概率大于与非工业客体有关的概率的情况下，而且仅仅在这种情况下，地质工作者才应当在异常上进行山地工作。第二种办法则意味着，甚至在异常与工业矿体有关的概率小于与非工业客体有关的概率时，只要已有在类似情况下可以见矿的实例证明投资是正确的，地质工作者也应当安排山地工作。

显然，在类似的情况下，在揭露之前先不把客体r称之为“矿体”或“非矿客体”（“工业客体”或“非工业客体”），而是说它值得不值得进行一定的工作，才是比较恰当的。不然的话，即使$a_{J_2}(R_1)$的值足够大，在被划归“矿体”（“工业客体”）类别的客体中，也可能大部分是非工业客体。把一定的工作形式与客体r相比拟，也是一种对这些客体按工作形式分类的方法。

在（22）式的条件下，运用算子（8）或（21）求解目标课题的算法，也可称之为广义的最大似然法。

我们注意到，如若分类不发生错误，且目标课题是给定的，广义的和一般的（分两阶段作判定的）最大似然法所得结果就表现不出什么差异。实际上，如果当$p(Q|R_2)>0$时必定有$p(Q|R_1)=0$，那么，无论是什么样的有限的$\mathscr{N}$值，每个客体$r\in R_1$都将被归入类别R_1^*，并都要建议对它进行工作J_1，无论用算法（8）还是算法（21）都一样。与此类似，如果当$p(Q|R_1)>0$

时必定有$p(Q|R_2)=0$，那么用两种算法都会得出进行工作J_2的建议。

方案4　当在训练资料中缺乏关于$P(R_1)$、$P(R_2)$的数值的资料，而具备上述其它资料时，与前述情况（认为$P(R_2)/P(R_1)=1$）类比，可从（22）式得出广义的最大似然法的最佳性条件：

$$\mathscr{K}=\frac{a_{J_2}(R_2)-a_{J_1}(R_2)}{a_{J_1}(R_1)-a_{J_2}(R_1)} \tag{23}$$

方案5　从公式（22）与（14）和公式（23）与（16）的差别中应该看出，在考虑量$a_J(R)$的情况下，目标课题的解的准确度会大为提高。然而，在训练资料中往往缺少有关这些数值的具体资料。例如，现在就没有关于用探槽揭露疏松产物之下的矿体露头或用钻孔发现某个深度上的矿体会给我矿带来何种益处的资料。将来，最好从被勘查客体与所得到的评价（例如在它们工业开采之后）相符的概率出发，对这类资料进行估计。这就需要进行专门的地质-经济学研究。但是，对量a_J的相互关系作大致的估计，现在一般已经可以做了。尽管这种估计的准确度低，但比放弃作估计，即等于认为它们都是一样的（根据所采用的求平均值的规则）要好得多。实质上，在现时流行的最大似然法的方案中，正是通过这种大致的估计来求$\mathscr{K}$值这个风险函数的，但在这里估计是根据对于所谓第一种错误和第二错误的不同风险的直观概念来进行的（这两种错误用我们的符号分别表示为：当$r\in R_2$时结论为$r\in R_1^*$，当$r\in R_1$时结论为$r\in R_2^*$）。现在，这两个概念针对已有资料的不同情况分别通过表达式（22）和（23）以最佳的方式具体化了。

我们来探讨一种局部的情况，即给出拟解课题训练资料的人常常认为，错误地将客体r归入可能有的类别中的某一类中就会导致某种损失（既便是精神上的和威信上的损失），以致于抵消了在对类别作正确“推测”时所获的成果。我们假设，在训练资料中既没有关于量$P(R_1)$、$P(R_2)$的资料，也没有关于a_J的具体

数值的资料，即就是说要在训练资料中只有上述设想和关于函数 $p(Q|R_1)$、$p(Q|R_2)$的资料的情况下来求解课题。

在这种情况下，我们将通过把量a_j的上述估计值代入表达式（23）来得到最大似然法的最佳性条件：根据上述设想，应当取

$$a_{J_1}(R_1)=a_{J_2}(R_2)=-a_{J_1}(R_2)=-a_{J_2}(R_1)$$

把这些表达式代入等式(23)，我们得到$\mathscr{K}=1$。应注意的是，这个等式对$\mathscr{K}$值的要求与上面探讨过的方案2中的最大似然法的最佳性条件是一致的。

由此可见，表示为（22）、（23）式的条件及与之类似的条件不是僵死不变的：它们并不限制广义最大似然法的适用性，因为它们很容易根据现有资料推导出来，并给出适于计算的选择$\mathscr{K}$值的规则。

实例2：

在连续的标志空间的条件下，现在广为采用的分类方法之一是判别分析。该方法可表达为，以训练资料为依据，标志Q的空间S可按照类别R_i的数目分割成几个子区。在两个类别R_1和R_2的条件下，这样的子区就是S_1和S_2，如果$Q(r)\in S_1$或$Q(r)\in S_2$，客体r便相应地被归入类别R_1^*或R_2^*，同时，分类是根据r的点相对于空间区界线的位置来进行的。界线是在对客体r作分类前根据训练资料划定的，以关于函数$p(Q|R_1)$和$p(Q|R_2)$的资料为依据，划界时要使在界线上的Q的点满足等式

$$p(Q|R_1)/p(Q|R_2)=\mathscr{K} \qquad (24)$$

式中，$\mathscr{K}$就是最大似然法中的那个量。在S_1或S_2区内，分别包含着比值$p(Q|R_1)/p(Q|R_2)$大于或小于$\mathscr{K}$值的标志Q。换言之，与最大似然法不同，在这里分类不是根据比值$p(Q|R_1)/p(Q|R_2)$直接进行的，而是根据标志Q落入已专门划分和限定的标志空间区来进行的。判别分析的特点是，这种分类要根据所谓判别函数的值来进行。对于界线上的点来说，该函数与表达式（24）是等效的；而对于分别处于界线两侧的点来说，该函数则

相应地取较大或较小的值。

从定义应当看出，在加于最大似然法的限定之外又增加了一些限定（要求量$p(Q|R_i)$具确定性，而这又要求标志空间具连续性），在这里，也就是意味着有可能进一步划分出数目有限的相互关联的子区S_1和S_2。如若没有添加的限定，要预先计算出这些子区间的全部界线实际上是不可能的。在上面说到的条件下，就方法的构成来说，判别分析与最大似然法是等效的，若违反这些条件，就不能采用判别分析。

由此可见，判别分析比最大似然法的优越之点可能仅仅是技术上的。只是在把数据Q代入判别函数求值比直接计算比值(24)要来得简单一些时，这种优越性才表现出来。此外，一次计算出的判别函数应当经常地使用，以补偿在根据训练资料构成判别函数和保存它时所花费的时间和经费。

为了简化上面论及的计算，用得较广的是线性判别分析。按此方法，S_1和S_2区的界线是用线性判别函数来近似的。在一般情况下，这个界线可能实际上不是直线。这时，甚至在遵循上述其它条件时线性判别分析的结果也和最大似然法的结果有差别。既然在这种条件下最大似然法是最佳的，相形之下，判别分析法就不是最佳的。

总之，在判别分析为最佳方法的所有情况下，最大似然法都是最佳的；而当最大似然法为最佳方法时，仅仅在一部分情况下判别分析才是最佳的。因而，判别分析要么得出与最大似然法相同的结果，要么当两个结果有差别时就应当认为它是优度（就信息量而言）较低的方法。在收入本文集的H. M. 弗朗科夫斯卡娅的文章中，以实际例子对最大似然法和判别分析法作了对比。从这些实例可以看出，上述差别实际上是很大的。

在现代技术条件下，要计算出最大似然法所必需给定的函数$p(Q|R_1)$和$p(Q|R_2)$的值，通常并不很困难。而根据这些函数在空间S中划定判别分析所必需的界线，既便它十分简单，也需要补做专门的研究和建立程序。这就缩小了判别分析的适用范围。

实例 3：

现在我们来寻求一种典型的直观推断法的最佳性条件。我们称它为Γ法，它是一种与《地质学与数学》(1970)一书中所述的《Голотип》法类似的方法。

令$Q_1^{(01)}$、$Q_2^{(01)}$、$Q_3^{(01)}$、……、$Q_{n_1}^{(01)}$——分别为对已知客体$r_1^{(01)}$、$r_2^{(01)}$、$r_3^{(01)}$、……、$r_{n_1^{(1)}}^{(01)}\in R_1$得到的数据$Q^{(0)}$的集合；$Q_1^{(02)}$、$Q_2^{(02)}$、$Q_3^{(02)}$、……、$Q_{n_2}^{(02)}$——分别为对已知客体$r_1^{(02)}$、$r_2^{(02)}$、$r_3^{(02)}$、……、$r_{n_2^{(2)}}^{(02)}\in R_2$得到的数据$Q^{(0)}$的集合。假定标志（$Q$的点）空间$S$是可量度的，$\mathscr{L}(Q_i, Q_j)$是空间中点$Q_i$和$Q_j$的距离。$\overline{\mathscr{L}}(Q_i^{(01)}, Q^{(01)})$——点$Q_i^{(01)}$与同一集合中其余点$Q_j^{(01)}$之间的平均距离；$\overline{\mathscr{L}}(Q_i^{(02)}, Q^{(02)})$——点$Q_i^{(02)}$与同一集合中其余点$Q_j^{(02)}$之间的平均距离。现在我们提出两个数$\delta_1$和$\delta_2$，用它们表示数$n_1$和$n_2$的某一部分。

把$\overline{\mathscr{L}}(Q_i^{(01)}, Q^{(01)})$的值排成一个递增序列，并用$\overline{\mathscr{L}}\delta_1$表示该序列各项的前几个$\delta_1 n_1$中最大的一个。用同样办法，我们可以找出$\overline{\mathscr{L}}\delta_2$值——$\overline{\mathscr{L}}(Q_j^{(02)}, Q^{(02)})$值的类似序列各项的前几个$\delta_2 n_2$中最大的一个。

用S_1'表示满足不等式

$$\overline{\mathscr{L}}(Q, Q^{(01)}) < \overline{\mathscr{L}}\delta_1 \tag{25}$$

的点Q的区域，而用S_2'表示满足不等式

$$\overline{\mathscr{L}}(Q, Q^{(02)}) < \overline{\mathscr{L}}\delta_2 \tag{26}$$

的点Q的区域。

这时，如果与客体r对应的标志$Q(r)$属于区域S_1'或S_1'，就可以认为该客体相应地属于类别R_1^*或R_2^*。

δ_1和δ_2值的选定要使不等式（25）和（26）不可能同时满足。如果这两个不等式连一个也未被满足，那就要拒绝进行分类。该方法只有在点的集合仅满足不等式（25）或（26）之一，并且该集合不是零集时才可使用。

首先应指出的是，在适用于Γ法的所有情况下，根据前面实例 2 中所述，判别分析都是求解课题的最佳方法。实际上，既然

在其中施用Γ法的空间S是可量度的，那么它也就是一个连续空间。区S'_1和S'_2是被作为实例2中所述的有限数目（两个）的相关联子区来看待的（它们本身正是这样的子区）。训练资料中包含着关于客体$r^{(0)}\in R_1$和$r^{(0)}\in R_2$的标志$Q^{(0)}$，因而，可据以求出函数$p(Q|R_1)$和$p(Q|R_2)$的最佳估计值。这样一来，判别分析的所有最佳性条件在此都具备了。

反过来，从所得的结论应该看出，当Γ法与判别分析等效的时候，也仅仅在这时，Γ法才是最佳方法。

在此基础上我们指出，在精度不超过用计算（当n_1和n_2数目有限时）建立点区的误差的条件下，下述原则是正确的：Γ法在而且只是在下述情况下，即点Q的分布在$\delta_1=\delta_2=1$和任何（大到适当程度的）n_1和n_2的条件下允许建立起有限个不交迭点区S'_1和S'_2的情况下，才是最佳的。

开始时我们已假设上述条件是具备的。这时所有的点$Q(r^{(0)}\in R_1)$都落在S'_1中，所有的点$Q(r^{(0)}\in R_2)$都落入S'_2中。判别分析要得出一个判别函数，并由一个超曲面反映出来，该曲面把空间S分成S_1和S_2两个区域，就作图来说，它应从S'_1和S'_2两区之间通过。因此，S'_1和S'_2中的每一个都是而且仅仅是区域S_1或S_2中的一个子区。为了明确起见，我们引入这样的符号，令S_1包括着S'_1，S_2包括着S'_2。这时，Q的点落入由Γ法所确定的S'_1或S'_2区的某一个中时，也必定落入由判别分析法所确定的区域S_1和S_2的某一个中，而且仅仅只落入一个区域中，反之亦然。因而，在这种情况下，Γ法和判别分析法的结果是相互等同的，即就是说，这时Γ法与最佳的判别分析法是等效的，因此它也是最佳的方法。

最后还要指出，当上述条件被破坏时，Γ法就不与判别分析等效了。

实际上，如果δ_1和（或）δ_2的值不等于1，$Q(r^{(0)})$的一部分点就会落在所划分的S'_1和（或）S'_2区之外。根据$r^{(0)}$和r之间的类似性，这就意味着对$Q(r)$而言落到上述两区之外的概率也不等于零。在这种情况下，Γ法拒绝作分类，可是判别分析则可进行

分类。因而，既然上述条件已这样被破坏了，Γ法和判别分析就不是等效的。

由此可见，要使Γ法与判别分析法等效，S'_1区和S'_2区合起来必须包含$Q(r^{(0)})$和$Q(r)$的所有的点。

S'_1和S'_2区的有界性是由包括在Γ法的算法中的不等式（25）和（26）决定的，因此如果不是拒绝采用该方法，是不可能使这种有界性破坏的。上述空间区的不交迭性也包括在Γ法的算法之中，因此如果不是拒绝采用该法，也不可能使这种不交迭性破坏。至此，关于Γ法最佳性条件的上述原则就充分地证明了。

上述论证说明，Γ法只是在作十分严格的限定时才不亚于判别分析法，而标志Q的上述分布（要求$Q(r\in R_1)$和$Q(r\in R_2)$的点域是有界的和连续的）是很罕见的。这样，据上所述，Γ法任何时候也不是比判别分析更好的方法。在研究者必须取δ_1和（或）δ_2值不等于1时，他分明是在Γ法的非最佳条件下使用它的，可是在这种条件下，判别分析法（还有最大似然法）却是最佳方法。

数据处理算法的建立

在“目标课题的求解”一节中，已建立了数据处理最佳算法的理论。然后又阐述了可以怎样利用这个理论来论证已建立的方法。这种理论也可以在很宽的条件范围内直接用于数据处理，而不用管已建立的方法是不是针对这些数据的。

实际上，在得到目标课题和原始数据之后，总可以在该课题相应的函数Φ中对训练资料里没有其资料的参数求平均值。在此以后，为了得到作为备用计算程序的算子F^*，仍然要对包括在该算子中的函数进行近似。已研究制定出了各种各样条件下的近似规则。在一些现在常用的数据处理方法的算法中，也使用着这些规则。例如，按照笔者在本文集和专著（19746）中就此问题阐明的算法，就可对一个广的函数集合进行划一的近似。

在某些情况下，用于进行近似的现有算法不是最佳的（就精确性而言），但如上所述，可以根据数学中现有的规则选出其中最好的算法（准最佳算法）。因此，所研究的划一的数据处理新方法，就这个词的完整意义上说，并不总是最佳的，但与采用一般方法相比，它总是明显较好的方法。随着近似理论的发展，在可以使用该方法的全部范围内，它应该得出越来越近似于最佳（就这个词的完整意义上说）方法的结果来。

显然，在已知方法中没有所需要的方法，而当新办法可以得出比现有办法更好的处理数据的算法时，采用新办法的合理性是不会引起异议的。同时应该强调指出，有可能用这种方式划一地构成与已知最佳方法等效的算法，在需要用这种算法代替那些须在存储器中长期储存的方法集合的形势下，是一件重要的事情。

已有的方法积累得越多，在存储器中储存和检索它们越困难，而又要能更快地重新得到它们，上述办法就变得越发有益了。随着时间的推移和科学的发展，所探讨的这种方法必然要被应用和发展。如果考虑到下述因素，这一点就变得很明显了。这个因素是，当存储器内局部算法的数目无限制增多时，求解问题所必需的运算数目也会无限制地增多，但在这些算法中有没有所需要的算法呢？其中哪一个是欲解课题条件下的最佳算法呢？如果它是存在的，为了从存储器中检索并提取出这种算法，所必需进行的运算数目也要增加。但是，建立一个可用于计算的算法，则要求其运算的数目是有限的。因此，由于处理数据的局部方法数目不断增加，总有一天与其从存储器中提取和使用原有的算法（哪怕是一部分算法）不如重新建立显然较好的算法要更有利一些。

在一般的办法中，由于存储器可供使用的容量是有限的（尽管它随着时间的推移会扩大），也会带来一些人所共知的困难。

上述新办法减少了一些这样的困难，虽然它还没有完全排除需在存储器中保存的算法数目仍会增多的问题。例如，为了以解上述狭义反演问题的形式进行较好的近似，在存储器中必须有一

套反映关于不同现象本质的理论概念的函数和算子。目前还没有构成这样一些函数和算子的通用算法。因此，为了扩大被解问题的集合，现在还必须扩大储存在存储器中的上述函数和算子的集合。但这个集合所需的存储器位置大大少于该集合中各函数和算子所能构成的组合的数目，而在先前的办法中，这些组合必须作为各种方法的组成部分储存在存储器中。从上述集合中选择所需的算子或函数，以及由它们构成多种多样的组合（表现为不同的类别），在许多情况下可以自动地完成，例如可按照笔者叙述过的三阶秩近似算法自动地完成（杜鲍夫，1974б）。

那么，使用新办法缩减算法数目的效果究竟如何呢？可以根据下述理由大致说明如下。

令E为局部方法F_k的数目，而其中每一个方法，当训练资料中具备关于特征$h_j^{(0)}$的集合$H_k^{(0)}$的资料时，对于求解与该方法相应的某目标课题的解A的估计值A_k^*来说，都应该是最佳的；n_k—集合$H_k^{(0)}$中的元素数目；N—在课题的解A中所要使用的特征集合H中的元素数目，如果训练资料含有关于所研究客体$r^{(0)}$的所有资料的话。

一套完整的可在给定课题和某种训练资料的条件下处理数据的方法，应当包括具备所有可能的各种集合$H_k^{(0)}\subseteq H$的方法F_k^*。

在这些方法中，每个元素$h_j^{(0)}$含有N个集合。每两个元素含有C_N^2个（从N个元素中每次取出两个进行组合的种数）集合。总之，每i个元素含有C_N^i个集合，$i=\overline{1,N}$。这样，这种集合的总数以及方法F_k^*的数目是

$$m=\sum_{i=1}^{N}C_N^i \qquad (27)$$

按这种新办法，不需要把这么多数目的方法都储存起来，只需要储存一种算法F（它是在原先意义上处理数据的诸方法F_k^*中满足$H_k^{(0)}=H$的那种方法）和求参数平均值的运算。这个运算对于不同目标课题来说是共同的。因此可以认为，进行这个运算不会对求解一个课题平均所需的存储器容量有很大影响。

应提起注意的是，一些目标课题可以视为另一些更广义的目标课题的局部情况。这时，这两种目标课题的相互关系与算子F_k^*和算子F是一样的，即就是说，与它们等效的数学表达式可以通过较广义课题中的参数统计平均法来得到。所有这些都使我们认为，采用新办法有希望大大地节省存储器，当然，以后最好对此问题进行更精确的研究。

通过上述不难看出，已经存在着某种与所提办法类似的办法，尽管它极其简单。必要的结果不是存在存储器里，而是每当需要的时候根据广泛使用着的函数表通过现代计算技术求出来。大家知道，在使用高速自动计算机的条件下，与其让已计算好的数值以表格的形式占满计算机的存储器，再从中寻找所需的结果，不如保存多种函数的计算算法，再根据需要对任意自变量的数值求出所需的数值要更好一些。

现在假定走下一步：不往存储器里输入和储存为数众多的局部算法，而是从更一般的原理出发根据需要建立这些算法或与它等效的算子。在存储器中储存的应当主要是算法的变换原理和规则。只是在需要经常使用某些局部算法的时候，才应当把这些局部算法存储起来（就和储存某些函数的单个常用值一样）。

新办法除了上述优点之外，还应指出：用求参数平均值的办法由算子F得出的最佳算法F^*，就其结构来说具有便于求解给定目标课题的形式。不考虑给定的目标课题而建立的等效算子F_k^*，可能具有在求解该课题时不太有利于计算的形式。下边，我们通过几个例子从这方面对用此办法得出的广义最大似然法的变种特别讨论一下。

每一个按照特定需要建立的局部算法都不是为了要多次使用它们。在这种情况下，根据上面给出的关于方法的定义，这些算法已经不起方法的作用了。从这一角度来看，可作为方法的应是一些较广义的算法，它们含有目标课题的数学表达式，并含有从中得出数据处理的局部算法的规则。就其使用的可能性来说，这种广义的算法与一些已知的但还没有按现代表达方式建立起来的

方法是等效的。随着时间的推移，作进一步的总结可能将是恰当的。

最后我们要指出，对未知参数求平均值不仅可以在函数 Φ 或算子 F 中进行，而且可以在已经由它们推出的表达式 Φ^* 和 F^* 中，以及在现有方法的算法中进行。在这几种情况下建立起的方法（或者供一次使用的算法）可按信息量递减的顺序排出一个序列。

实例 1：

我们假定，数据 Q 的处理是要解这样一个目标课题：选择一种宜在客体 $r(Q)$ 上进行的工作 J_1 或 J_2。我们要在训练资料已明确了 $r(Q)\in R_1\cup R_2$ 并且包括下列补充资料的条件下，用上述方法找出处理数据的最佳算法。

1. 数值 $a_{J_1}(R_1)$、$a_{J_1}(R_2)$、$a_{J_2}(R_1)$、$a_{J_2}(R_2)$、$P(R_1)$、$P(R_2)$ 和函数 $p(Q|R_1)$、$p(Q|R_2)$ 均为已知。

在这种情况下，按目标课题说来，其最佳解应由（18）式表示的算子 F 来确定，该式所需的所有原始数据在上述训练资料中都已具备。如果进行下列代换，原始数据齐全的情况就一目了然了。

根据公式（11）、（16）、（17），可以导出下列符号：

$$\begin{aligned}\bar{a}'_{J_1}(r)&=\bar{a}_{J_1}(r)\cdot p(Q)=a_{J_1}(R_1)\cdot P(R_1)\cdot p(Q|R_1)\\&\quad+a_{J_1}(R_2)\cdot P(R_2)\cdot p(Q|R_2),\\\bar{a}'_{J_2}(r)&=\bar{a}_{J_2}(r)\cdot p(Q)=a_{J_2}(R)\cdot P(R_1)\cdot p(Q|R_1)\\&\quad+a_{J_2}(R_2)\cdot P(R_2)\cdot p(Q|R_2)。\end{aligned}\tag{28}$$

由此应得，若 $\bar{a}'_{J_1}(r)>\bar{a}'_{J_2}(r)$，则 $\bar{a}_{J_1}(r)>\bar{a}_{J_2}(r)$，反之亦然。因此，表达式（18）可改写为：

$$F=\begin{cases}J_1, & \text{如果}\bar{a}'_{J_1}(r)=\max\left(\bar{a}'_{J_1}(r),\bar{a}'_{J_2}(r)\right)\\J_2, & \text{如果}\bar{a}'_{J_2}(r)=\max\left(\bar{a}'_{J_1}(r),\bar{a}'_{J_2}(r)\right)\end{cases}\tag{29}$$

这里引入的量 $\bar{a}'_{J_1}(r)$ 和 $\bar{a}'_{J_2}(r)$ 可根据数据 Q 和训练资料直接计算出来。在这种情况下直接表示出目标课题最佳解（就信息量而言）的公式（29），在我们上边提出的办法中也应当使用。

上一节曾经指出，在这种情况下，使用算子（18）和（22）式为条件的最大似然法仍然是最佳解法，不过要补充一些专门的运算-判定条件：若$r\in R_1^*$，则$J=J_1$；若$r\in R_2^*$，则$J=J_2$。正如所看到的，这种解法虽然与前述解法是等效的，但在所用的表达式方面却有区别。下面在一个实例中将要探讨前面提出的新办法的一个很普遍的优点，并把它更清楚地显示出来。

2．假若在上一条所列数据中$P(R_1)$和$P(R_2)$是未知的。关于量$a_{J_1}(R_1)$、$a_{J_1}(R_2)$、$a_{J_2}(R_1)$、$a_{J_2}(R_2)$只知道下列信息：对类别R_1最好进行工作J_1，对类别R_2最好进行工作J_2。选错了工作种类就会带来损失（物质的或精神的），其大小与正确求解所得好处在绝对值上是一样的。

在这种情况下，按照求未知参数平均值的法则，公式（28）中的量$P(R_1)$、$P(R_2)$应当用某个同一的正值（以α表示）代替之。同时，参数$a_{J_1}(R_1)$、$a_{J_2}(R_2)$也应该用另一个同一的正值（以β表示）代替之。而参数$a_{J_1}(R_2)$、$a_{J_2}(R_1)$则用绝对值与其相等的负值（以$-\beta$表示）代替之。

这时，根据等式（28），若$\alpha\beta p(Q|R_1)-\alpha\beta p(Q|R_2)>-\alpha\beta p(Q|R_1)+\alpha\beta p(Q|R_2)$，则$\bar{a}'_{J_1}(r)>\bar{a}'_{J_2}(r)$。把前一个不等式以$\alpha\beta$约简，我们便得到与之等效的表达式$p(Q|R_1)-p(Q|R_2)>-p(Q|R_1)+p(Q|R_2)$，或者在合并共同项后得：$p(Q|R_1)>p(Q|R_2)$。与此类比，若$p(Q|R_1)<p(Q|R_2)$，则$\bar{a}'_{J_1}(r)<\bar{a}'_{J_2}(r)$。因而在这种情况下表达式（29）可转化为：

$$F^*=\begin{cases}J_1, & \text{如果}\bar{a}''_{J_1}(r)=\max(\bar{a}''_{J_1}(r),\bar{a}''_{J_2}(r))\\ J_2, & \text{如果}\bar{a}''_{J_2}(r)=\max(\bar{a}''_{J_1}(r),\bar{a}''_{J_2}(r))\end{cases}\qquad(30)$$

式中

$$\bar{a}''_{J_1}(r)=p(Q|R_1)\ ,\ \bar{a}''_{J_2}(r)=p(Q|R_2)\qquad(31)$$

不难看出，所得出的算法与以$\mathscr{K}=1$为条件并附加了运算-判定条件（若$r\in R_1^*$，则$J=J_1$；若$r\in R_2^*$，则$J=J_2$）的最大似然法是等效的。

3．现在我们假定训练资料只是由下列资料构成：已知的仅仅

是分别代表类别R_1和R_2的两客体$r_1^{(0)}$和$r_2^{(0)}$的标志$Q_1^{(0)}$和$Q_2^{(0)}$。在标志空间内，这两个类别分别由明显相关的点Q组成的两个点域反映出来。标志空间的量度$\mathscr{L}$前已给出。此外，关于量$a_{J_1}(R_1)$、$a_{J_1}(R_2)$、$a_{J_2}(R_1)$和$a_{J_2}(R_2)$也采用与第二条中一样的假定。

在这种情况下，根据求未知参数统计平均值的法则，为了得到明显的最佳算法F^*，作为第二条中所进行代换的补充，应当取：若$\mathscr{L}(Q, Q_1^{(0)}) < \mathscr{L}(Q, Q_2^{(0)})$，则$p(Q|R_1) > p(Q|R_2)$，反之亦然。

这时，我们即可从表达式（30）得到处理数据Q的最佳算法，其形式如下：

$$F^* = \begin{cases} J_1, & \text{如果} \mathscr{L}(Q, Q_1^{(0)}) = \min(\mathscr{L}(Q, Q_1^{(0)}), \\ & \mathscr{L}(Q, Q_2^{(0)})) \\ J_2, & \text{如果} \mathscr{L}(Q, Q_2^{(0)}) = \min(\mathscr{L}(Q, Q_1^{(0)}), \\ & \mathscr{L}(Q, Q_2^{(0)})) \end{cases}$$

实例 2：

我们将举例说明是可以根据一个较广义课题的现成算法来求解某个目标课题的。这要求助于对广义课题中那些非主要（对给定的局部课题而言）参数作求平均值的运算。

设拟解课题是：必须把一个取得了数据Q的客体r归入类别R_1^*或R_2^*（一如前述，显然须假定$r \in R_1 \cup R_2$）。实例 1 中提出的课题在此就作为较广义的目标课题。

我们先来研究在训练资料中包含有广义课题的所有重要资料的情况。这时，如前所述，它的解可由公式（29）和（28）来表示。不过在此不是研究工作J_1和J_2，而是要得出拟解课题所要求的关于客体r应归入类别$R_1^*(r \in R_1^*)$还是类别$R_2^*(r \in R_2^*)$的结论。分类的必要性，即把一些客体分别归入类别R_1^*和R_2^*的必要性，意味着人们想做到当实际上$r \in R_1$和$r \in R_2$时能够相应地得出$r \in R_1^*$和$r \in R_2^*$的结论。满足了这一要求，就意味着得到了某种收益$a_{J_1}(R_1)$和$a_{J_2}(R_2)$；要是搞错了，就意味着带来失误$a_{J_1}(R_2)$和$a_{J_2}(R_1)$。在被求解的课题中，并不在乎这些量的绝对值的差别大小，只要得出收益为正值失误为负值就很清楚了。如果在训

练资料中有资料说明某种工作在客体$r\in R_1$和$r\in R_2$上进行都会得到收益（两种收益是不同的），那么被限定的目标课题就可把这种资料略去。于是，按照求参数平均值的法则，为了进行计算就需要取：

$$a_{J_1}(R_1)=a_{J_2}(R_2)=-a_{J_1}(R_2)=-a_{J_2}(R_1)$$

而这时，与上面已作过的代换相类比，我们可以从表达式（28）得出：如果$P(R_1)\cdot p(Q|R_1)-P(R_2)\cdot p(Q|R_2)>-P(R_1)\cdot P(Q|R_1)+P(R_2)\cdot p(Q|R_2)$，或者说$P(R_1)\cdot p(Q|R_1)>P(R_2)\cdot p(Q|R_2)$也一样，那么$\vec{a}'_{J_1}(r)>\vec{a}'_{J_2}(r)$。同样，如果$P(R_1)\cdot p(Q|R_1)<P(R_2)\cdot p(Q|R_2)$，那么$\vec{a}'_{J_1}(r)<\vec{a}'_{J_2}(r)$。

令符号$\vec{a}''_{k_1}=P(R_1)\cdot p(Q|R_1)$，$\vec{a}''_{k_2}=P(R_2)\cdot p(Q|R_2)$。这时，利用这些关系式和上面对工作$J_1$和$J_2$所作过的代换，就可以从公式（29）得出求解被探讨局部课题的算法：

$$F^*=\begin{cases}r\in R_1^*, & \text{如果}\vec{a}''_{k_1}=\max(\vec{a}''_{k_1},\vec{a}''_{k_2})\\ r\in R_2^*, & \text{如果}\vec{a}''_{k_2}=\max(\vec{a}''_{k_1},\vec{a}''_{k_2})\end{cases}\qquad(32)$$

不难看出算法（32）与$\mathcal{K}=P(R_2)/P(R_1)$时的最大似然法是等效的，这与前面引入的最大似然法最佳性条件是相符的。

当训练资料中没有任何可供算法（32）使用的资料时，另外的求解上述目标课题的最佳算法，可以像在实例1中已做过的一样，根据训练资料用求参数平均值的办法得出来。

实例3：

现在我们来说明用新办法可得出比用已知方法显然要简便的算法的情况。

设目标课题是要选择在客体r上进行的工作J_i。训练资料包括$P(R_i)$、$a_{J_i}(R_i)$的数值集合和函数$p(Q|R_i)$，这里，$i=\overline{1,n}$，n—客体r所可能归属的类别的数目，它与对这些客体最“有利”的工作的数目（与客体类别相当）相等。根据上一节所指出的，在这种情况下可以采用最大似然法作为最佳方法（考虑到当类别数目有限时，分类可简化为分别顺序地在每两个类别之间进行）。这时，可先针对任何两个类别R_{i_1}和R_{i_2}计算比值$P(Q|R_{i_1})/$

$p(Q|R_{i_2})$，然后评定该比值是否超过了针对这一对类别给定的与公式（22）类似的量$\mathscr{K}=\mathscr{K}_{i_1,i_2}$的值。根据此结果选定了类别$R_{i_1}$或$R_{i_2}$后，再计算该类别与另一类别的类似比值，将所得数值与新的$\mathscr{K}$值作对比，如此等等。最后，用这种方法从最后一次对比中得到某个类别R_i之后，便得出结论：$r\in R_i^*$，因而也就是$J=J_i$。

按照推荐的这种办法，这个结果可用下述方法得到。

根据目标课题的题意，是要在客体r上找到这样一种工作，使用它所得收益的数学期望最大，亦即$\bar{a}_{J_i}(r)$值为最大，如上所述，这与要求量

$$\bar{a}'_{J_i}(r)=\sum_{i=1}^{n}a_{J_i}(R_i)\cdot P'(R_i|Q) \tag{33}$$

达到极大是等效的。式中，$P'(R_i|Q)=P(R_i)\cdot p(Q|R_i)$。因此，最佳算法具下述形式：

$$F^*=(J:\bar{a}'_{J_i}(r)=\max_j\{\bar{a}'_{J_i}(r)\})$$

换言之，对于每一个j来说，所要计算的仅是量$\bar{a}_{J_i}(r)$，然后再选出使$\bar{a}_{J_i}(r)$达到最大的j值，显然，与上述最大似然法的一长串计算相比，这种计算要简单得多。当n很大时，这种简化可能是很重要的，特别在不采用快速自动技术而又必须迅速完成计算的情况下更是如此。

极为明显的是，与最大似然法不同，所述办法可以在不必事先分析哪一种工作对某个类别“有利”的情况下使用，因为它不含有需依据客体对某个确定类别R_i^*的归属性而选择工作的运算。工作种类的数目在这里可以不等于在最大似然法中使用的类别R_i的数目。由似可见，在求解给定的目标课题时，所提出的办法是最大似然法的方案之一，而这种方案与一般的最大似然法不同，它不受后者须恪守的一些重要限制的约束。前述的不需要按类别R_i^*对客体集合作划分的广义的最大似然法的公式（21），与一般的最大似然法相比要好一些，但它同样需要计算比值$p(Q|R_{i_1})/$

$p(Q|R_i)$和$\mathscr{K}_{i1i2}$值。因此，与从我们举的例子中得出的算法相比，它仍然是不太经济的。

最后，与已采用的最大似然法的公式不同，这种新方案可以在类别空间为连续空间或区段-连续空间的情况下，即类别R的数目是无限的情况下使用。这时，按照一般的最大似然法要把所有的类别都两两搭配成对，并将与之相应的概率密度比值加以对比，自然是不可能的。在这种情况下，如果采用上述方案，只需要找到连续函数或区段-连续函数

$$\overline{a'_j}=\int_V a_j(R)\cdot p(R|Q)dV$$

达到极大值的点R就行了（式中，dV—类别空间V中包含着R点的基元空间）。

如所周知，连续函数或区段-连续函数极大值的位置，可以用任何事先给定的精度通过有限数目的运算估计出来。

H. M. 弗朗科夫斯卡娅以实例验证了在类别R_i的数目n有限的情况下，上述这种最大似然法方案的某些特点。她的这篇论文已收编于本文集中。

例实 4：

我们来探讨一下综合普查工作的领导者现在面临的求解目标课题的理论方面的问题。

我们设想，在某个区域（面积）Π内进行了地球化学和地球物理调查，结果，在测网的每个点上都得到数值$q_i\in Q$——地球化学样品中某些化学元素的浓度和各种地球物理场的数值。

一般情况下，数据Q中不仅包含有直接在观测点上得到的结果，也包含有在观测点周围得到的结果；而且为了在观测之后得出结论，既要利用被观测场的数值，也要利用被观测场值的空间变化（即异常的强度和形态）参数。

人们所需要的是，以这些数据为基础对一些点$\rho_i\in\Pi$的远景作出评价（以便编制测区的远景图，选择作为进一步工作地段中心的最有远景的点，等等）。

我们用下述方式来定义点的远景的概念。

设：V——在其范围内可能存在矿化的一个地质空间；ρ_r——该空间的一个点；$b(\rho,\rho_r)$——产在ρ点附近的矿石每单位体积可提供赢利的数学期望（假使把这些矿石都开采出来并把它们运送到ρ_r点上）。这时，空间$dV\ni\rho$的体积增大就会使点ρ_r的远景增大，其增值与$b(\rho,\rho_r)$及dV是成正比例的。因而，点ρ的远景取决于是否在整个空间V中都有矿石存在，它等于：

$$\Phi(\rho_r)=\int_V b(\rho,\rho_r)dV \qquad (34)$$

现在我们来探讨当训练资料中除了有集合$Q^{(0)}$之外还有下列资料时如何求解给定的目标课题的问题。

设：对一个面$\Pi^{(0)}$上的有限数目的点$\rho_r^{(0)}$来说，我们已经具备了数据集合$Q^{(0)}$和使用公式（34）时必须有的所有数据。换句话说，在先前已做过良好研究的不同地段上，在足够大的距离内和可以达到的整个深度上，成矿元素的空间分布情况会是已知的，在这一空间内，可能存在一些$b>0$的点（即含有工业矿化的点），并具备所有必要的经济评价的资料。

在这种情况下，根据训练资料中的数据我们可进行下列计算。

我们在任一经研究地段的面$\Pi^{(0)}$上选出某个点$\rho_r^{(0)}$。由于空间$V^{(0)}$的各点都有相应的训练资料来说明它，我们就可针对每一个点$\rho^{(0)}$计算出$b(\rho^{(0)},\rho_r^{(0)})$的值。再根据公式（34）计算出$\Phi(\rho_r^{(0)})$的值。对于不同研究地段的不同的点$\rho_r^{(0)}$都要按此程序完成计算。在实际计算中当然可以使用近似法。

$b(\rho,\rho_r)$的值属于一个连续的集合。由于量Φ是由这些值根据公式（34）确定出来的，因此Φ的值也属于一个连续的集合。所以就可用一个（广义的）回归方程把求出的$\Phi(\rho_r^{(0)})$值与$Q^{(0)}(\rho_r^{(0)})$值间的对应关系表达示来。例如，利用前已述及的秩近似法，就可以根据事先编制的$A^{(0)*}=\Phi(\rho_r^{(0)})$值和与之对应的$q_i^{(0)}$值的表格把这种方程建立起来。这时，我们便得到算子$F^*$，它

在这里表达着回归函数：

$$A^{(0)*}=F(G^*, Q^{(0)}) \qquad (35)$$

如在“目标课题的求解”一节中谈过的那样，由此我们便得到处理数据Q的表达式：

$$A^*=F^*(G^*, Q) \qquad (36)$$

式中，$A^*=A^*(\rho_r)$——拥有数据$Q=Q(\rho_r)$的点ρ_r的远景估计值。

计算函数$b(\rho, \rho_r^{(0)})$需要有大量的资料：有关所有成矿元素浓度的空间分布资料和经济标准的资料。应予提示的是，经济标准是因矿石积聚量的大小而异的。因此点ρ上的$b(\rho, \rho_r)$值不仅取决于该点的成矿元素浓度，也取决于该点周围的成矿元素浓度。在需要处理数据Q的研究者手中可能会没有这样的资料（其原因之一是对客体研究得不充分，尽管在训练资料中包含有关于这些客体的资料）。在这类情况下，就要用相应的求未知资料平均值的办法建立函数$b(\rho, \rho_r)$。

我们假设，在一些其资料已包括于训练资料中的客体上，除了数据$Q^{(0)}$以外，仅仅知道成矿元素浓度的空间分布$C_i=C_i(\rho)$（这里，$i=\overline{1, N}$，N则为成矿元素的数目），以及每种元素单位重量的出售价格g_i。此外，研究者假设在点ρ上从开采矿石得到的赢利与$\mathrm{h}(\rho)+k\cdot L(\rho, \rho_r)$成反比（这里：$\mathrm{h}(\rho)$——点$\rho$的深度，$L(\rho, \rho_r)$——点$\rho_r$与点$\rho$在通过点$\rho_r$的水平面上的投影间的距离，$k$——某个常量系数）。

在这种情况下，就不必使用量$b(\rho, \rho_r)$，而应该用量

$$b^*(\rho, \rho_r)=1/[\mathrm{h}(\rho)+K\cdot L(\rho, \rho_r)]\sum_{i=1}^{n} g_i\cdot C_i(\rho)$$

来代替它。

其它计算依然如前。

现在，我们假设训练资料还要少一些：除了数据Q外，仅知道已研究矿体重心的位置$\rho_j(j=\overline{1, n})$和矿体的矿石储量$B_j$，关于赢利与量$\mathrm{h}(\rho)$和$L(\rho, \rho_r)$的关系的假设仍如前述。这时，不用量$b(\rho, \rho_r)$而应代之以量

$b^*(\rho_j, \rho_r)=B_j/[\mathrm{h}(\rho_j)+k\cdot L(\rho_j, \rho_r)]$，不要用公式（34）而应代之以公式

$$\Phi(\rho_r)=\sum_{j=1}^{n} b^*(\rho_j, \rho_r)$$

不难发现，在这里对每个矿体范围内的矿石的特征参数都采用了求平均值的方法。

实例 5：

此例是为了告诫人们，如果稀里糊涂地求参数平均值就会犯错误。

假设提出了一个评价点ρ_r的远景的目标课题。但与前述情况不同，远景是由下列目标函数来确定：

$$\Phi(\rho_r)=\int_H \frac{C(\rho_r, \mathrm{h})}{\mathrm{h}}\cdot d\mathrm{h} \qquad (37)$$

式中，$C(\rho_r, \mathrm{h})$——点ρ_r之下的深度 h 上某成矿元素的浓度；H——可能有矿化的全部深度范围。

在寻找具近水平层状矿体的矿床时，特别可能提出这样的目标课题。这时，凭直觉就可知道，在某个点之下的垂直矿量（垂直方向的矿石量）越大，矿石埋藏越浅，这个点的远景就越大。

我们假设，在训练资料中除了数据Q之外只有$C(\rho_r^{(0)}, \mathrm{h})$的数值，而且相应的 h 值也已给出。

在这种情况下，应该根据公式（37）对每一个点$\rho_r^{(0)}$计算量$A^{(0)}(\rho_r^{(0)})=\Phi(\rho_r^{(0)})$，再把这些数值与数据$Q^{(0)}(\rho_r)$作对比，找出能最佳地表达关系式$A^{*(0)}(\rho_r^{(0)})=F^*(G^*, Q^{(0)})$的算子$F^*$。此后，亦如前所述，数据$Q$的处理就表现为计算表达式

$$A^*(\rho_r)=F^*(G^*, Q) \qquad (38)$$

但下面这种解法也是可取的。根据数据$C(\rho_r^{(0)}, \mathrm{h})$针对每个点$\rho_r^{(0)}$求出在整个深度范围内的垂直矿量

$$B(\rho_r^{(0)})=\int_H C(\rho_r^{(0)}, \mathrm{h})d\mathrm{h}$$

和在该点之下矿化（具有C值的矿石柱）重心的深度$h_c(\rho_r^{(0)})$。然后求算子F_1^*和F_2^*，它们分别表示每个点$\rho_r^{(0)}$之下的垂直矿量值和矿化重心埋藏深度对于$Q^{(0)}$的回归：

$$B^*(\rho_r^{(0)})=F_1^*(G_1^*, Q^{(0)}),\quad h_c^*(\rho_r^{(0)})=F_2^*(G_2^*, Q^{(0)})。$$

借助这两个算子，也可以针对每个点ρ_r分别对矿化的垂直矿量和矿化重心产出深度进行估计：

$$B^*(\rho_r)=F_1^*(G_1^*, Q) \tag{39}$$

$$h_c^*(\rho_r)=F_2^*(G_2^*, Q) \tag{40}$$

有了这两个数值，就可以针对每一个点ρ_r把量A的估计值计算出来，它就是这两个量的比值：

$$A^*(\rho_r)=B^*(\rho_r)/h_c^*(\rho_r) \tag{41}$$

这样的解比上边由公式（38）表示的解可能还要更直观一些。然而解（41）比解（38）的信息量要小一些。

实际上，在计算量$B(\rho_r^{(0)})$对$Q^{(0)}$的回归时，并不考虑这个量与h的关系，这就相当于对参数h求了平均值。在计算量$h_c(\rho_r^{(0)})$对于$Q^{(0)}$的回归时，也未考虑这个量与B的关系，这就相当于对参数B求了平均值。因此，在结果表达式（41）中就包含了对参数求平均值，不过其形式不明显罢了。应当指出，矿石量与其产出深度之间的依从关系可以是很明显的。如所周知，既有主要在相对的低温低压下于近地表条件下形成的矿床，也有其形成需要较高的温度和压力的矿床。此外，矿石量与其产出深度之间的依从关系还取决于从矿床形成到研究之时止侵蚀截面所达到的深度，以及一些其它因素。

因此，用公式（39）和（40）来计算的量$B^*(\rho_r)$和$h_c^*(\rho_r)$只能作为辅助量使用，例如在用它们来解释何以会出现远景估计值A^*的极大值时，就只是个辅助量。比如说，A^*极大值的出现既可解释为是由观测面以下的大型深部矿床造成的，也可解释为是由离观测面不远的小矿化引起的，两种解释可能都有意义，如此等等。量A的估计值本身应当根据公式（38）计算，据等式（37）看，公式（38）反映着远景与矿床的垂直矿量和产出深度

的依从关系，且没有上述求参数平均值的问题。

结　论

1．在处理任何具体数据之前，可以而且应该形成一个对所有在检验前被认为是可取的数据处理方法来说共同的目标课题。

2．训练资料是以具体客体的标志的形式还是以早先得出的理论结论的形式提出来，就决定了在数据处理结果中对未直接观察过的客体作怎样的描述。

3．根据目标课题的表达方式和具有的训练资料，可以划分出处理具体数据的最优方法和最佳（按信息量说）方法。

4．已得出的那些规则可以用来对处理数据的各种方法进行论证，其中包括目前习惯地认为毋须论证的（可直观推断的）方法。

5．处理各种具体数据的明显最佳算法（按信息量说）可以按照划一的方法求得，即依据训练资料对未知参数作求平均值的专门运算，通过建立给定目标课题的数学表达式来求得。要实现这种可能，就要建立一些新的广义方法，每一个广义方法都与一个现有的和尚未建立的处理数据的最佳（从当时的观点来看）方法的集合是等效的。将来在存储器中最好储存这类广义方法，而不储存数目很多的常见的局部方法（十分常用的方法除外）。

为了解决本文论及的问题，最好在考虑经济效果的前提下去发展选择和建立数据处理算法的理论研究。在某些情况下，若有一种方法其信息量虽不是最大的，但与最佳方法相差不多，且具有工作效率高和相当经济的优点，就可以使用它。这时，最好计算一下方法的总精确度（它取决于可参考到的各种因素和计算误差）。

可以认为，对于在从较一般的解中得出局部解时的函数近似规则和求未知参数平均值的规则，还需要进行补充研究。本文只探讨了比较简单然而也很常见的情况。

最后我们指出，希望对训练资料的选择问题进行研究：根据前面阐述的原理，最优方法的选择与训练资料是有关系的。

本文已详细探讨过的那些原理，很可能还需要再加研究，使之更为精确和完满。

参 考 文 献

Бондаренко В. Н., Верховская Л. А., Григорян С. В., Позднякоа Э. Н., Сорокина Е. И. Проблема алгоритмизации литогеохимических поисков рудных месторождений. —"Сов. геол.", 1974, № 11, с. 79—89.

Вострокнутов Г. А. Интерпретация геохимических данных на основе интегральных показателей. —В кн.: Математические методы при геохимических исследованиях. (Тезисы к семинару "Методы математической обработки результатов геохимических поисков"). Свердловск, Уральское территориальное геологическое управление, 1973, с. 37—19.

Геология и математика. Новосибирск, «Наука». 1970. 224с.

Дубов Р. И. К теории поиска оптимальных решений. —В кн.: Математические методы в геологии и геостатистика. М., «Наука», 1973 а, с. 14—28

Дубов Р. И. О методах классификации и интерпретации геохимических аномалий. —В кн.: Математические методы при геохимических исследованиях (Тезисы к семинару «Методы математической обработки результатов геохимических поисков»). Свердровск, Уральское территориальное геологическое управление, 1973б, с. 16—17

Дубов Р. И. Принцип выбор и обоснования ме-

тодов обработки наблюдений. —В кн.: Тезисы семинара "Применение математических методов и ЭМВ в геологии", Алма-Ата, КазИМС, 1974а, с. 53—55.

Дубов Р. И. Количественные исследования геохимических полей для поисков рудных месторождений. Новосибирск, "Наука", 1974б. 278с.

Китаев Н. А., Евдокимова В. Н., Чумакин В. Н., Сынлен С. Е. Классификация геохимических свойств геологических объектов в связи с изучением их зональности и генезиса. —"Геол. и геофиз.", 1974, № 6, с. 135—139.

Родионов Д.А. К вопросу о статистической теории однородности геологических совокупностей. —"Геология", 1965, № 4, с. 466—473.

Родионов Д. А. Статистические методы разграничения геологических объектов по комплексу признаков. М., "Недра", 1968. 158 с.

Таусон Л. В., Дубов Р. И., Козлов В. Д., Кузьмин М. И. Геохимическое значения дисперсии концентраций редких элементов в гранитоидах. —В кн.: Вопросы геохимии изверженных горных пород и рудных месторождений Восточной Сибири. М., "Наука", 1965, с. 12—22.

吴传璧译