

应用概率图解确定地球化学数据异常下限的方法

前言

概率格纸可用来简明直观地表示各种类型的地球化学数据,是推断地球化学分析结果的一种辅助手段。已发表过不少文章,但对异常下限的确定都未作详细论述。在化探工作中,异常下限是指能够将不同原因、基本性质不同的高低数值区分开的一个特定的值,将较高的异常从较低的背景值辨认出来。异常值往往与含矿岩石有关,根据异常值可能圈出有经济价值的矿床。所以确定异常下限对指导勘探工作有很大意义。

确定地球化学数据异常下限的方法有好几种。其中有一种是将异常下限确定在数据的平均值加两倍其标准离差的值上。这在某些情况下是适用的,但不顾事实如何,将超过两倍标准离差范围的值一律定为异常,并没有太大的依据。另外这种方法没有考虑到异常与背景总体在某些情况下有较宽的重迭范围,同时因为它们是两个总体,从整个数据组导出的平均值与标准离差,实际上缺少统计上的真实性,只不过是一些数字。此缺点已被许多依靠主观视觉办法研究数据直方图来确定异常下限的工作者所公认。

另一种方法是将累积概率曲线最大曲率的点上的数值定为异常下限。此法需要用一直线来逼近概率曲线的一段,而将这条“直线”线段与概率曲线交点的纵座标值定为异常下限。这只不过是一种近似方法,可能漏掉很多异常。

需要有一种确定异常下限的好方法,能尽最大可能将异常辨认出来,而同时使包含

在异常数据中的背景值的数目减到最少。累积概率曲线提供了一种达到这一目的的有效图示方法。

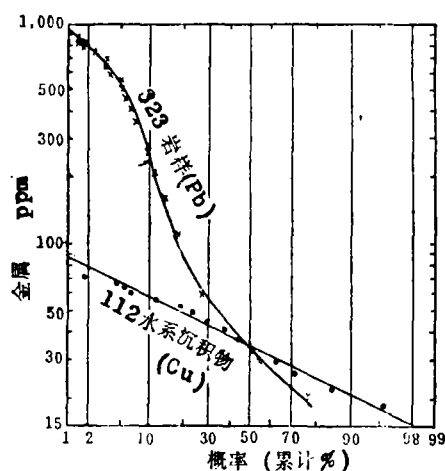


图 1

概率格纸

算术概率格纸,其纵座标代表元素含量,用算术标尺表示;其横座标代表概率分布函数(或累积频率百分数),这样绘出的正态(高斯)累积分布曲线是一直线。对数正态概率格纸与前者的差别在于它的纵座标也是对数标尺。单峰对数正态分布的算术数值的分组,与构制累积直方图所要求的方式完全一样,绘在对数概率纸上是一条直线。由两个对数正态总体组成的双峰对数正态分布图是一条曲线。单峰的和双峰的对数正态分布的例子如图1所示。在本文所举例子中,两种分布均用含量值的(从高含量的一端开始)累积频率绘制的。

概率格纸的优点是:1.可以检验数据组

的密度分布形式；2.能快速估算正态分布与对数正态分布总体的参数；3.在同一张图纸上可以表示几个数据组，比多元直方图更为清楚；4.可用视觉方法比较几个数据组的曲线，很快地看出其相同或不相同之处。另外还可以将多峰(多众数)分布分成各个总体，参看下面的例子便可明瞭了。其不足之处是：1.数值可能不呈正态分布，或是不呈对数正态分布；2.尽管这种方法适于处理少量数据，但绘制一张概率图解一般至少要有上百个数据；3.在概率图解上如果数据分散太开，就得出置信的分析结果。虽然存在着这些缺点，但在概率格纸上能有把握地分析大部分地球化学数据。

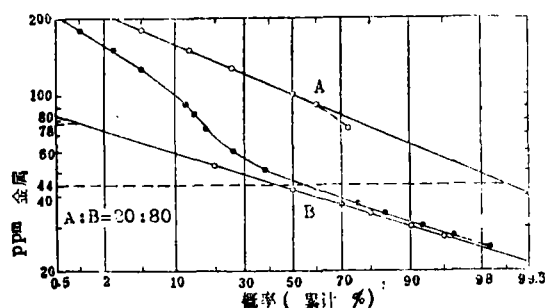


图 2

多峰(多众数)分布的划分方法

这是指从含有两个或两个以上的总体组合组成的多峰分布中提取各个总体的方法。先研究双峰分布的情况，假设各总体具有正态的或对数正态的密度分布，将其绘在适当的概率格纸上，根据曲线的拐点或是曲率方向的改变来估算它们的总体的比例，例如在图2中，累计频率为20的拐点(以箭头标示的)表明高含量的总体A有20%，低含量的总体B有80%。曲线形状表明有两个重叠的总体。斜率相当平缓的中间线段，表示两总体有相当一部分重叠。

纵座标为180ppm的点，即标在曲线上位置最高的那一点，其累计频率为1%。用它算出的总体A的累积频率为5(=1/20×100)，

因为在数据组这一端不受总体B分布的影响，因此在含量为180ppm累积频率为5%处确定了总体A上的第一个点。按同样方式可用纵座标为150ppm点的累积频率算出总体A累积频率为13(=2.6/20×100)，从而得到总体A上的第二个点。重复这种过程直到获得足够的点可绘成一条直线来确定总体A时为止，或者是直到得出的点开始偏离直线，表示有大量的总体B存在时为止。得到这样一条直线后即可对总体A进行参数的估算。

用同样的方法可求得总体B，但这时概率标尺应按其余值读数。例如，累计频率90%应读成累积频率10%(即100-90=10)。计算出的总体A与总体B都以圆圈示于图2。

可用纵座标分别为20%与80%的A、B总体的组合模型来检验这两个总体模型的确实性。在此假想例子中，因为是在理想情况下构制的，所以没有将检验结果所得的点表示在图上。但在以后的例子中，检验结果均以△号表示。检验方法是利用关系式

$P_M = f_A P_A + f_B P_B$ 计算不同纵座标上已划分的总体的理想组合模型的 P_M ， P_M 是所要计算的“组合”总体的概率； P_A 与 P_B 是从概率格纸纵座标上分别读出的总体A和总体B的累积概率； f_A 是分出来的总体A的比例， f_B 是分出来的总体B的比例(=1- f_A)。实际上，因为精确地确定拐点有困难，需要校核几次才能使理想的“组合”总体与实际的数据有比较好的拟合。在大多数情况下，多峰分布的划分方法是很简单的。在其它情况下，在处置实际数据时，需对方法作少许修改。在下面的例子中能清楚地看到这点。

划分包含总体三个以上的多峰曲线是比较复杂的，但作法一样。一般是从概率曲线末端所代表的总体开始划分，其次再划分位于曲线中部的总体。

在此假想例子中，可以估算已划分的各个总体的参数。在累积频率50%处可读出每个总体的几何平均值，在累积频率为84%与

16%处可以确定包括含量值的68%的范围。包含2倍标准离差的这一范围围绕着几何平均值是不对称的。这里所采用的表示方法是将几何平均值用括号括起来,括号中后面的值是包括含量值68%的范围。已划分的总体A和B的参数是100ppm(144.71)与42ppm(55.33)。

异常下限的确定

图2的假想例子中说明了一种普遍常见的情况,高、低含量总体有一个有效的重迭范围。如果不存在有较大的数值重迭,则曲线中部陡度不大的一段将变得近乎直立,就可很快地在其中点上确定出一个异常下限。但在一般情况下选择异常下限比较复杂。

设想在图2中在经过划分的总体A与总体B的累积频率为99%与1%的点上分别任意确定了两个异常下限(记着,A与B是按 $A/B=20/80$ 比例划分的)。这两个累积频率在纵坐标上含量值为44ppm与78ppm处将全部数据分为三组。组合总体全部数据其中有16%的含量值超过了78ppm的异常下限的上限。在假定的100个样品数中,其中高含量的一组将含有大约15个总体A的样品和1个总体B的样品。低于44ppm的一组含有组合总体全部数据的46%,由1%总体A(在此情况下最多只有1个样品)与57%的总体B(大约有46个样品)组成。两个异常下限之间的中间组,含组合总体全部数据的38%,由总体B42%与总体A23%组成。这在假设的样品数中,相当于4个(或7个)A样品与34个(或33个)B样品(表1)。

此法虽武断,但能有效地把数据分成三组,其中有两组含有大部分高含量的总体A,第三组几乎专一地代表低含量的总体B。我们现在暂时假设A和B分别代表异常总体和背景总体。超过异常下限上限的较高的含量组要优先予以检查验证,因为实际上全部含量值都是异常。中间这一组是次优的一组,因

表1

	组合总体		总体A		总体B	
	频率(%)	样品数	频率(%)	样品数	频率(%)	样品数
第一组	16	16	76	15.2	1	0.8
第二组	38	38	23	4.6	42	33.6
第三组	46	46	1	0.2	57	45.6
	100	100	100	20	100	80

•样品数=100,其中总体A占20,总体B占80

为该组实际上也是异常,但也含有背景样品,因此需要对每个异常样品增加人力来检查它们并将它们从同一背景范围里面挑选出来。

关于用百分位数来确定异常下限是无可非议的。在此情况下,A和B总体分别为99%和1%的累积频率所对应的含量即为组合总体的异常下限的上、下限。用适当方法划分出来的总体的98%和2%的累积频率也同样能够很好地确定异常下限。无论怎样确定,确定分组中每个总体所含数目是可能的。根据作者的经验,上面提到的两组数字证明最有用,但可以选择不同的值,这取决于数据的性质,以及所有异常值保留在较高的两组内所要求的概率。

注意在此假想例子中,异常下限确定为平均值加两倍标准离差,将会把大部分异常值与背景值放在一起。将此法作一些变动,

高于 $2\frac{1}{2}\%$ 的值作为异常,将得到同样的效果。如果用三个直线段来逼近概率曲线,其交点将大约在103ppm与55ppm给出异常下限。一般采用上限值作异常下限的方法,将导致有50%的异常值被舍弃。即使选择较低的值也会造成5%的异常值被丢掉。

剑特洛湖地区(中不列颠哥伦比亚)土壤中的锌

图3为取自剑特洛湖附近已知Mo-Cu矿区内B层土壤中173个土壤Zn的分析结

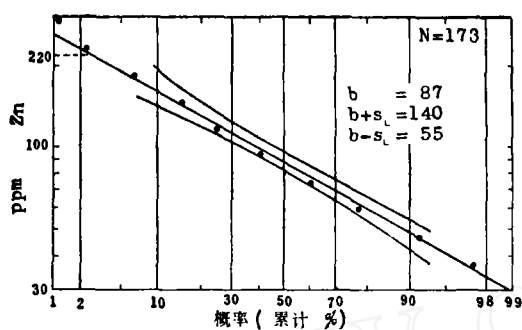


图 3

果,绘制的概率曲线。下伏岩石为闪长岩,在构造上和含矿方面都很均匀,节理发育。节理含矿,主要有石英与黄铁矿,但在某些地方辉钼矿很多,还有少量黄铜矿。除有星散的圆形露头外,该地区为不厚的覆盖层所掩盖。

如果忽略不考虑两端有少许张开(一般是由于采样误差所造成),其概率曲线是直线。因此,利用通过所绘出的点的那条直线,可以求得分布的估计参数。用图解方法确定其总体的95%置信界限、有一种有用的方法能在概率图解上分辨有意义的曲率,该方法假定只有单一的总体存在,构绘出其95%的置信带,对概率图解有意义的曲率是在置信带外边的点上。剑特洛湖地区的数据所绘的点没有位于用95%置信界限所圈定的带之外的,说明它只有单一的总体存在。

在此情况下数值的范围与概率曲线的形状表明数据代表的是单一的背景总体。因此,合理的方法是假定,除非有其它证明,只有几个最高的数值是异常。在为确定第二个总体异常样品太少的情况下,这是一种方便而保险的方法。为处置这样的数据使方法标准化,比较方便的方法是把平均值加两倍标准离差的纵座标值确定为任意的异常下限。在此方法中假定数值大约超过 $\frac{1}{2}$ %标准离差的值才是异常,否则应有其它的证明。此方法只在根据对概率曲线的观察看出是一个总体时才适用。在此例中,概率网格图上上部5

个Zn含量值是较高的,但都远离已知矿区。

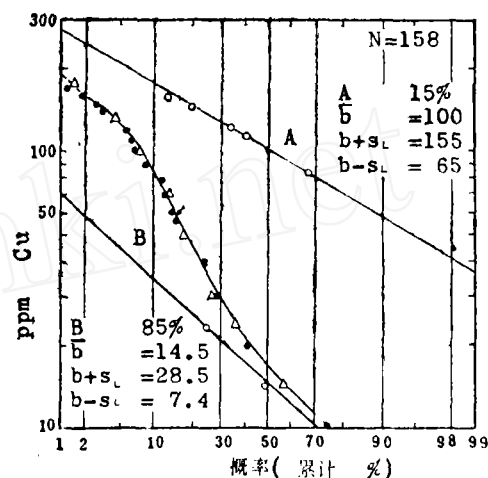


图 4

兰森山区(育空地方)水系沉积物中的Cu

采集了158个水系沉积物样品,将铜的分析结果作出的概率图解示于图4。通过数据点的圆滑曲线有着双峰的密度分布形式,拐点在累积频率15%处。利用上述方法将曲线划分获得的总体A和总体B,及估算的参数列于表Ⅰ。用纵座标铜含量值不同的两个划分出来的总体,按15%A与85%B的比例组合,对划分方法进行了检验。检验的点绘在图中以△号表示,可以看出与实际曲线重合。在此例中,某些高含量值与产在斑岩侵入体内的已知Cu—Mo矿床有关,推断解释认为有两个总体(A为异常,B为背景)似乎是合理的。

根据曲线上总体B与总体A的累积频率分别为1%与99%所对应的含量值,可以立即确定出两个任意的异常下限。其异常下限值分别为70ppm与37ppm。因此数据被分成3组。高含量组以异常为主,低含量组以背景值为主,中间组既有异常也有背景。158个样品中,大约23个是异常,135个是背景。异常值有80%(或者说有18个)超过70PPm异常下限,有5个低于70ppm异常下限,就实际效果讲是落在中间组范围内。在135个背景样品中,有91.5%的(即124个)的样品低于异

常下限的下限,其余11个超过异常下限的下限落在中间组中。

因此,异常样品位于两个ppm区间,都要给以优先的检验。实质上高于70ppm的所有样品都是异常,都要优先检查验证。落在中间组的16个样品,其中的有5个是异常,应放在第二位。

理论上落在两个异常下限之间的样品,不能划入总体A,也不能划入总体B。因为落在此范围内的样品三个中只有1个是异常,所以大约需要用检验Cu高于70ppm的样品所需要的工作量的3倍来检验每个异常,因此,这是优先检查验证这两组的原因。实际上能够很确切地辨认出落在中间组的某些异常样品。例如,其中有一些可以期望在离开将予以最优先检查验证的异常样品的下游出现。在采样的水系平面图上如果样品按颜色编码来分组,这种地理位置的关系显得特别明显。在许多场合,实际上所有落在中间组内的样品可用同样方法很确切地将异常样品从中辨认出来、当处理已确定有两个异常下限的土壤或全岩分析结果时,可用类比法。那些落在中间组的样品与地理位置已知异常归在一组的,一般也可以认为是异常。这样第二位优先检查异常的工作量可以削减到最少,而且在许多情况下,可以完全免去。

兰森山区划分水系沉积物中Cu的估算参数 表2

总 体	比 例 (%)	样 品 数	Cu的含量(ppm)		
			$b+SL$	$b-SL$	
A(异常)	15	24	101	155	63
B(背景)	85	134	14.7	28.5	7.4
A+B	100	158			

霍普地区(南不列颠哥伦比亚) 土壤中的镍

图5是在该区一个已知矿带上采得的166个土壤样品Ni的分析结果的对数概率图解。矿带产在区域变质细粒碎屑沉积岩里面的超基性岩中。通过实测数据绘出的圆滑曲

线,根据累积频率为5.5%与25%处的拐点看出,曲线至少有三个总体。A和C总体是前述方法作出的。总体B是后来利用下面的关系式估算的:

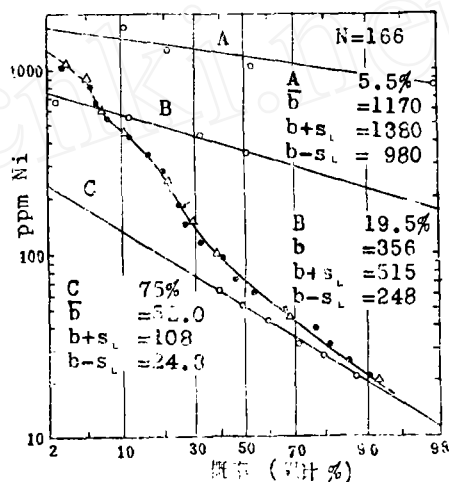


图 5

$$P_M = f_A P_A + f_B P_B + f_C P_C$$

式中: $f_A=0.055$, $f_B=0.195$, $f_C=0.75$, P_M, P_A, P_C 可从图中纵座标上读出。 P_B 是唯一的未知数,可对不同纵座标进行估算,将结果绘在图上,用直线通过所计算得的点,对总体B的参数作出估计。将三个划分的总体(A、B、C)按上面提出的占全部样品数的比例(5.5/19.5/75)组成理想的组合总体来检验划分方法。将检验值以 Δ 号表示于图5中,几乎与原始数据点的平滑曲线重合。

总体A的点不完全在估算的直线上,显然,总体A的划分不够好。原因是组合总体中代表总体A的数据太少,所以只能根据很少的点(在此例中只有4个)来估计它的参数。总体B与总体C的划分较好,主要原因是按19.5%和75%的比例组成的理想组合总体与实际的数据曲线很符合。三个总体的估计参数列入表3中。根据划分的总体,可以选定780Ni作为唯一的异常下限来有效地区分总体A和总体B,总体B与总体C稍有重叠,必须确定两个异常下限。这两个异常下限任意地确定在总体C累积频率为2%(即

236ppmNi), 和总体B的累积频率为98% (即170ppmNi) 处。

霍普地区土壤中镍划分总体估计的参数 表3

总体	比例 (%)	样品数	Ni含量ppm		
			b	b+SL	b-SL
A(异常)	5.5	9	1170	1380	980
B(背景, 超基性岩)	19.5	32	359	515	243
C(背景, 变质沉积岩)	75	125	52	108	24.5
A+B+C	100	166			

此三个异常下限值将数据组分成4组, 其中有三组各有一个总体的样品, 第四组含有两个总体的样品(表4)。现在可利用这些异常下限值作为网格平面图上的等量线, 或是用颜色或符号在平面图上进行数码编号, 帮助推断解释每个总体的显著性。在此情况下, 总体A与Ni矿有关, 是异常总体。总体B与超基性岩有关, 总体C出现在下面为变质沉积岩的地区。

霍普地区所估计的土壤中镍的异常下限 表4

异常下限	组中主要成分
780	大部分为总体A
236	大部分为总体B
170	由总体B和总体C组成
	大部分为总体C

异常下限值是任意确定的。例如, 将总体B和总体C的两个异常下限分别确定在其累积频率为1%与99%处, 或是在2.5%与97.5%处, 等等, 也都是可以的。在进行确定时, 要使确定出的两个总体有不大的重迭范围, 并使相邻范围实际上只含一个总体的样品, 而其它总体的样品数非常少, 可以忽略不计。

斯迈哲斯地区(不列颠哥伦比亚) 土壤中的铜

795个土样中铜的分析结果的概率曲线如图6所示。因为数据组中数值很多, 曲线

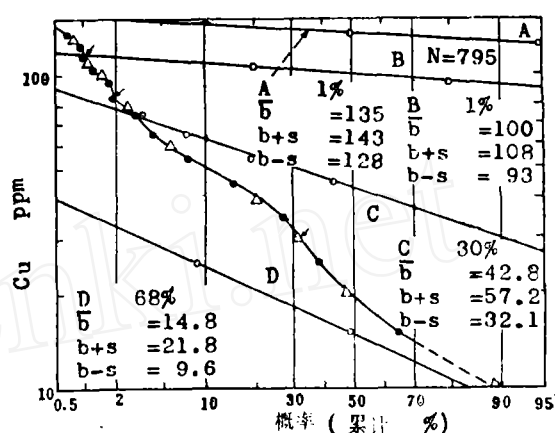


图 6

呈弯曲特征可能是真实的。这种类型的数据具有在比较短的时间获得大量资料的普查测量所获得的那种特征。采样地区下面的岩石主要是切过厚层的、单一火山系的酸性到中性的侵入岩。

拐点大致在累积频率为1%, 2%与32%处, 表明至少有4个总体存在。逐段划分曲线可以对这些总体作出估算。在此情况下, 由总体C开始最方便, 因其大部分数据点都可采用。一待总体C确定后, 利用C与原始数据曲线可估计总体D。因总体A、B在全部数据中所占比例小, 用于划分的点数也小, 所以只能近似地确定。根据有限数据估算的总体A总体B也示于图6。

对划分的、以比例1/1/30/68组合的A、B、C、D总体, 计算得的一些检验点以△号表示。这些点几乎与实测数据的圆滑曲线完全符合, 表明这种划分的总体代表一种似乎有理的数据模型。划分的总体其参数列于表5中, 将数据与采样区的地质图比较, 看出总体C、D分别代表火山岩与深成岩上面土壤中Cu的背景样品。而总体A、B则分别代表下伏火山岩和深成岩地区的异常总体。

在用确定的异常下限来区分异常样品和背景样品时, 不需要考虑总体A和总体B、概率曲线的转折部分是总体B与总体C重迭的范围。

我们知道约有 2 % 的样品 (即 16 个) 是异常, 其中, 像总体 C 的 1 个样品一样, 有 11 个样品 Cu 超过 100 ppm。于是, 超过 100 ppm Cu 的 12 个样品中有一个不是异常, 可以确定 100 作为任意确定的异常下限的上限。

斯迈斯地区, 土壤中铜, 划分出的总体的参数 表 5

总体	比例	样品数	Cu 的含量 ppm		
			b	b + sL	b - sL
A	1	8	135	145	128
B	1	8	100	108	93
C	30	239	42.8	57.2	32.1
D	68	540	14.8	21.8	9.6
A + B + C + D	100	795			

实质上所有异常总体的样品, Cu 都超过 85 ppm。所以 Cu 为 85~100 ppm 的这一范围含有余下的 5 个异常样品。此范围还含有 1.0 % 的背景总体 C (大约有两个样品)。于是确定出两个异常下限, 从实际效果来看, 辨认出全部异常样品, 含的背景样品极少。

这一例子说明方法中的几个要点:

1. 为检验所作出的解释的确实性, 研究复杂的分布, 最好用一完善的划分方法来进行。

2. 甚至在各个总体不能分别精确地被划分时, 一般还是可以确定出异常下限值, 而且具有足够的精度。

3. 根据很多数据给出概率曲线, 具有其真实的拐点, 可用作推断解释的基础。

4. 另一方法是根据下伏的火山岩或深成岩的存在, 把数据分成两个亚组。在这里没有用这种方法, 只是因为在进行更详细的分析时, 不花费另外的人力, 就可得出合适的异常下限。

5. 尽管底下的总体 D, 只是根据两个点划分出来的也还是合适的。

上述的几个例子说明概率图解的主要优点在于它能够有效地分组。其目的不只是要求出异常与背景总体之间的异常下限, 而且更普遍地是要求出各个总体之间的异常下限, 有助于对数据的意义作出总的

全面的推断解释。

水系沉积物的 pH 的测定

pH 的测定一般是水系沉积物研究工作中不可缺少的部分。测量了南不列颠哥伦比亚某地区水系沉积物的 pH 值; 所获得的概率曲线如图 7 所示。曲线绘在算术概率纸上

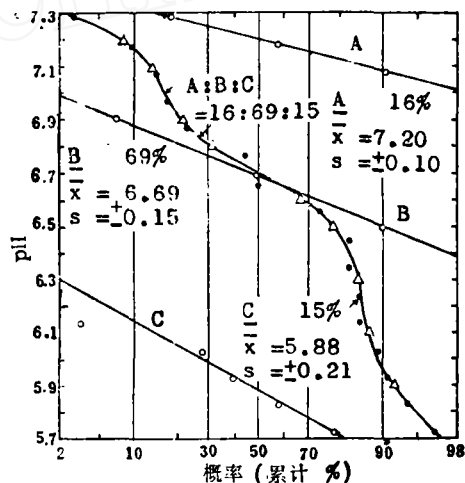


图 7

(由于 pH 值本身的性质, 对原来的数据进行了对数转换)。通过数据点的光滑曲线具有三峰分布的形式。拐点位于累计频率 16 % 与 85 % 处。用上述方法划分出总体 A、B、C。对三个总体按 16/69/15 的比例组成的理想组合进行了检验。检验结果与实际数据曲线非常吻合。

异常下限随意地确定在总体 A 和 B 累积频率 99 % 处, 与总体 B 和总体 C 累积频率 1 % 处, 得到的结果列入表 6 中。

估计的 pH 值的异常下限 表 6

pH	频率 (%)	
7.00	总体 A 为主	15
6.93	总体 A + 总体 B	4.5
6.37	总体 B 为主	64.5
6.36		
6.35	总体 C 为主	16

因此在进一步分析和推断解释以前, 根

据pH测量结果将全部数据分成了四组。这种分组在推断解释微量元素数据时,由于pH值对金属的分散有影响,所以具有很大的意义。

对分析精度的要求

在估计异常下限的方法中,要求分析的结果十分精确。在测定的数据中含有取样误差和分析误差。有些数值虽超过异常下限,但实际上是属于异常下限以下的,也有与这种情况相反的。通常这种混淆只影响数据很小的一部分,但如果分析精度愈差时,这种影响就显得愈大。为此,把异常限定在一个数值范围内,其中心值是唯一的异常下限,是用完全已知的数值求得的。上述范围是依据数据的精度所确定的置信范围。确定这样一个异常下限范围,其一般精度常常是合适的。但精度确随所估计的变量的绝对大小的变化而变化。当有适当的数据可资利用时,对这点应予考虑。在等量线图这种异常下限范围圈出来的是一个狭窄的带。

这种方法可能会使异常的样品数“增多”,检验这些“增多”的样品需要的时间与费用也增多,但研究这些增多的样品相对于已知异常样品的地理位置,能使所花的时间与费用减到最少。

讨 论

所介绍的确定异常下限的方法适用于处理大量地球化学数据。如有质量合适、数量足够的数据,就可以划分总体,还可以用于多峰分布。可将数据的数值分组,这对推断解释有帮助。此法比其它常用的方法可能更有用。尤其是在背景与异常总体有很多重迭的情况下,用此法圈定异常效果颇好。

此法不限于确定背景与异常总体之间的异常下限。它实际上是一种最通用的方法,可把具有适当密度分布的数据分成许多组。这种数据的概率图解分析既简便又快,很适

于野外应用。如果根据某种物理的或化学的准则,先把数据分组,一般都可以得到简单的解释,但如果数据中存在有三、四个总体,用这种方法也可以作出较确切的解释。

结 论

1.地球化学分析结果一般均呈对数正态密度分布,用对数概率格纸能充分地将其分布表示出来。

2.只要数据组含有足够的数据(最少要有100多个),即可划分总体的累积概率曲线,估算各个总体构成的多峰总体分布的参数。

3.划分出来的总体可用来确定任意的(但是是有意义的)异常下限,将数据分成有基本意义的组群。

4.在异常与背景总体之间没有有效重迭的特殊情况下,可确定一个异常下限。在一般有两个重迭的异常和背景总体的简单情况下,要确定两个异常下限,将数据分成三组,高含量组主要为异常样品,低含量组为背景样品,中间含量组合有异常与背景样品。

5.由两个以上的总体组成的地球化学数据的多峰分布,一般可用对待双峰分布相同的办法,求得有用的异常下限。但在某些情况下,方法可以简化,根据某些基本特征将数据分成组群,以便得到较简单的概率曲线,使划分的总体与推断解释有较大的置信程度。

6.所论述的确定异常下限的方法,不限于划分异常与背景,对任何类型的数据组,只要各总体近于对数正态密度分布,这种方法都适用。幸好,这些准则,对大量地球化学数据都可以满足。

摘译自《Journal of Geochemical Exploration》, 1974, Vol. 3, No. 2, P. 129—149

作者: A.J. 辛克莱 吴奇石译

初绍华校