

关于若干化探数据处理方法的讨论

纪宏金¹ 林瑞庆² 周永昶¹

(1. 吉林大学, 长春 130026; 2. 栖霞地矿局, 栖霞 265300)

[摘要] 化探数据处理中的系统误差校正和异常识别方法很多, 移动标准化法、分区标准化法和衬度系数法是其中的几种方法, 主要对这几种方法做了简单的介绍并对其效果进行了对比, 最后着重讨论了C型变换法的优点。

[关键词] 地球化学找矿 系统误差校正 异常识别 标准化变换 C变换

[中图分类号] P632 [文献标识码] A [文章编号] 0495-5331(2001)04-0056-04

化探数据处理中的系统误差校正和异常识别是两个复杂而有意义的问题。本刊1999年第2期发表的题为“化探背景与异常识别的问题与对策”^[1](简称“对策”)一文, 文章在总结了有关方法进展之后, 重点评述了衬度系数法、分区标准化和移动标准化等类方法中存在的“问题”, 并分别给予了全部或部分否定。作为解决“问题”的对策, 作者推荐了C型变换法, 并论证了其“优点”。

虽然文中所述大多属于一元方法, 较为简单, 但在基本概念理解上与笔者存在分歧, 故写此拙文, 不当之处, 望读者批评指正。

1 关于被否定方法

1.1 移动标准化法

移动标准化方法^[2]是针对区域化探数据特点而设计的一种异常识别方法。它明确定义:

“设某移动窗口内单元背景数据子集服从正态分布, 其子样均值为 \bar{X} , 方差为 S^2 , 则该窗口中心处数值(一般为网格化数据) X 有以下标准化形式:

$$T = (X - \bar{X}) / S \quad (1)$$

由上式及其变形

$$X = \bar{X} + TS \quad (2)$$

易见, 移动标准化变量 T …直观表达原始数据 X 属于异常的概率意义和正负性质, 例如 $T=0$ 为背景值, $T=3$ 为均值加三倍标准差, …当该窗口中心不断移动并用(1)式作相应计算, 便可估计出图幅内各点处的变量 T , 这就是移动标准化分析。”并反复强调上式中的分布参数应为“背景分布参数”, 指出了“效果好坏取决于计算方法的稳健性”, “要移动窗口大小选择适当”, “合理消除异点影响”, “满足或近似

满足稳健性要求”等。

“对策”也叙述了一种移动标准化方法, 并当作靶子进行批评, 不妨称其为“靶移动标准化”。尽管文中对其叙述简略, 但结合插图可知, 该错误的表现形式是, 标准化后的背景总体与异常总体都变成均值为0方差为1的同一分布, 这不仅是该文所批评的缩小了它们的差异, 而是完全抹杀了它们的差异, 必然导致异常识别的错误; 其错误的技术关键是, 计算分布参数时异点未被剔除, 或窗口选得过小而其中全部为异常数据, 使得(1)式中的均值与均方差不一定是背景分布参数, 甚至完全是异常分布参数。

与“靶移动标准化”相反, 移动标准化方法强调了避免上述错误的概念, 即要求保证上式中的分布参数为“背景分布参数”。移动标准化仅使得正态假设下的背景总体变成标准正态分布, 而异常总体仍呈长尾分布于其后。容易理解, 若疏忽这些基本概念, 把移动窗口选得过小或不作稳健处理, 得不到背景分布参数, 移动标准化确实能蜕变为“靶移动标准化”。例如, 当窗口大小与异常地质体尺寸大小相当时就是如此。

遗憾的是, 在评述移动标准化时, “对策”不但出现了这种疏忽, 而且还恰恰加以“移动窗口大小必定要与研究区内各地质体的平均尺寸相当”的前提, 将移动标准化转换为“靶移动标准化”。

1.2 分区标准化法

分类标准化、分区标准化和分幅标准化^{[3][4]}是为消除不同岩性、不同介质的元素背景差异, 或不同图幅间系统分析误差而设计的校正方法。

现以分幅标准化法^[4]为例讨论。该方法以某图幅元素含量的背景均值和均方差为参数, 用(1)式的

[收稿日期] 2000-01-01; [修订日期] 2000-04-21; [责任编辑] 余大良。

形式对该图幅内原始数据作标准化,使不同图幅背景部分的标准化均值都为0,方差都为1,以此校正不同图幅的背景差异。

“对策”批评的分区标准化法也可称为“靶分区标准化”。它在以(1)式的形式对某图幅内原始数据作标准化时,使用的参数是元素含量在全图幅的均值和均方差。这自然就出现了各图幅标准化均值为0,方差为1的不合理现象,被“对策”批评为“对化探工作者来说同样不能接受”,因为“每个图幅元素含量的平均值都为0,方差均为1”。

这一批评有两个特点,第一是被批评的错误正是来源于“对策”本身。因为它将分区标准化中的“背景均值和均方差”曲解为“全图幅的均值和均方差”。第二是把“标准化数值”与“元素含量”这两个不同概念混为一谈。

1.3 衬度系数法

衬度系数法也是一种背景差异或系统误差校正方法。“对策”认为,衬度系数法的缺点是“把各个图幅各个元素的平均含量都统一到数值1”,质疑说“任意分割的两个不同地区(图幅)的元素平均含量都相等吗?”

这一批评也有两个显著特点。第一是其形式和性质与对分区标准化的批评完全一样,不再重复。第二是恰好颠倒了衬度系数法的优点和缺点。由下文讨论可知,衬度系数法的优点正是被否定了的各图幅均值为1;其缺点正在于被作为优点继承于C型转换的各图幅均方差不等。

2 有关方法效果对比

对化探拼图中的差异校正问题,“对策”推荐了C型转换法,认为它具备了诸多方法的优点,但是其效果究竟如何,就需要与其它方法作出对比。

为明确答案,不妨给出一个化探数据系统误差校正、解决相邻图幅拼接问题的假想例子,对待校正数据作出如下假设:

①某元素在同一图幅的含量有:任意邻域的均值等于全图幅均值,任意邻域的方差等于全图幅方差;

②某元素在不同图幅的含量有:均值不等,均方差不等,这些差异全部由分析误差引起。显然,校正后不同图幅均值应相等、均方差也应相等。现用不同方法对上述差异作出校正,对比其效果优劣。

2.1 标准化方法

由于问题的简单性,分幅标准化、分区标准化甚至移动标准化方法都能达到同样校正效果。这里不妨用分幅标准化法。设 X_{ij} 为某元素在图幅 i 样品 j 中的含量, X_i 和 S_i 分别为该元素含量在图幅 i 的背景均值与均方差(本例中也是全图幅均值与均方差,下同),可得该元素在图幅 i 样品 j 的分区标准化数值为

$$T_{ij} = (X_{ij} - X_i) / S_i$$

$$i = 1, 2, \dots, p; j = 1, 2, \dots, N_i \quad (3)$$

其中, p 为图幅个数, N_i 为图幅 i 中样品数。易证,该元素在图幅 i 的标准化背景均值 T_i 与均方差 U_i 分别为

$$T_i = \frac{1}{N_i} \sum_{j=1}^{N_i} T_{ij} = 0$$

$$i = 1, 2, \dots, p \quad (4)$$

$$U_i = \sqrt{\frac{1}{N_i} \sum_{j=1}^{N_i} (T_{ij} - T_i)^2} = 1$$

$$i = 1, 2, \dots, p \quad (5)$$

可见,它们都与图幅 i 无关,即不同图幅上背景均值一致,变化范围相同。因此,该方法校正了上例中假想的系统误差。

2.2 衬度系数法

设 W_{ij} 为某元素在图幅 i 样品 j 的衬度值,其它记号同上,则有

$$W_{ij} = X_{ij} / X_i$$

$$i = 1, 2, \dots, p; j = 1, 2, \dots, N_i \quad (6)$$

可证元素在图幅 i 的衬度背景均值 W_i 与均方差 V_i 分别为

$$W_i = \frac{1}{N_i} \sum_{j=1}^{N_i} W_{ij} = 1$$

$$i = 1, 2, \dots, p \quad (7)$$

$$V_i = \sqrt{\frac{1}{N_i} \sum_{j=1}^{N_i} (W_{ij} - W_i)^2} = S_i / X_i$$

$$i = 1, 2, \dots, p \quad (8)$$

可见,元素的衬度背景均值与图幅 i 无关,但均方差一般与图幅 i 有关,即在不同图幅上的背景均值一致,但变化范围一般不同。要取得好的效果,必须保证以下条件成立:

$$V_1 = V_2 = \dots = V_p \quad (9)$$

但这是难以满足的。故该方法一般可部分校正上例中假想的系统误差。

2.3 C 型转换法

根据“对策”的定义,某元素在图幅 i 样品 j 中的 C 型转换值可表示为

$$C_{ij} = (X_{ij}/X_i)(S_j/\bar{X}_j) \\ i = 1, 2, \dots, p; j = 1, 2, \dots, N_i \quad (10)$$

其中, \bar{X}_j 和 S_j 分别为图幅 i 样品 j 附近某邻域内元素的含量均值和均方差, X_i 为图幅 i 的元素含量均值,其它记号同上。在本例假设下,上式可简化为

$$C_{ij} = (X_{ij}/X_i)(S_i/X_i) \\ i = 1, 2, \dots, p; j = 1, 2, \dots, N_i \quad (11)$$

不难证明,元素在图幅 i 的 C 背景均值 C_i 与均方差 Q_i 分别为

$$C_i = \frac{1}{N_i} \sum_{j=1}^{N_i} C_{ij} = S_i/X_i = V_i \\ i = 1, 2, \dots, p \quad (12)$$

$$Q_i = \sqrt{\frac{1}{N_i} \sum_{j=1}^{N_i} (C_{ij} - C_i)^2} = (S_i/X_i)^2 = V_i^2 \\ i = 1, 2, \dots, p \quad (13)$$

显然,它们一般都与图幅 i 有关,即在不同图幅上的背景均值一般不等,其变化范围也一般不同。要取得好的效果,也是保证条件(9)成立。由于该条件难以满足,故该方法一般不能校正上例中假想的系统误差。

由此还可发现 C 型转换法与衬度系数法之间的半依赖关系:若条件(9)不成立,衬度系数法可校正背景均值, C 型转换法则完全无校正能力;若条件(9)成立,二者可同时作出完全校正。出现这种关系的原因,正是 C 型转换放弃了衬度系数法的优点,而继承了其缺点。

3 关于 C 型变换法

由以上论证可知,在所述方法中, C 型转换效果一般不如其它方法。但“对策”认为:“校正的目的不是让某个特征参数等于某个固定值,而是要使所有的数据具有统一的地质和找矿意义”、“也就是说,在化探专家脑海中存在一种无形的标准,可以用来识别有意义的异常。总结这种标准并把它转换为数学方法对校正系统误差和消除背景差异无疑是十分有益的”,“期望有识之士加强该方面研究”。根据介绍,由这种“无形的标准”理论总结出来的 C 型转换最少有两个优点。现在我们简要讨论这两个优点。

3.1 第一优点

据“对策”语意, C 型转换的第一个优点是,“确

认 C 型转换可把异常和背景总体的差异扩大”,“为进一步划分背景与异常提供了良好条件”。

但值得疑虑的是, C 型转换在力求扩大背景与异常差异的同时,也可能扩大了图幅间的背景差异或系统误差。由(12)、(13)式知, C 值在图幅 i 的均方差 Q_i 与均值 C_i 之间有如下关系:

$$Q_i = C_i^2 = (S_i/X_i)^2 \\ i = 1, 2, \dots, p \quad (14)$$

由此产生两方面的问题:一方面,各图幅 C 均值间差异难以预料,即可能很小,也可能很大;另一方面,各图幅 C 均方差间差异对 C 均值的差异非常敏感,只要某图幅的 C 均值稍有增大,其均方差就会呈平方形式扩大。若把 C 均值比作海平面, C 均方差比作波浪高度,那么,在 C 均值较小的图幅将是微波荡漾,而 C 均值较大的图幅则呈现巨浪滔天。

如此图幅拼接,很难说有“统一的地质和找矿意义”,也不会“对校正系统误差和消除背景差异无疑是十分有益的”从而“为进一步划分背景与异常提供了良好条件”。

“对策”在强调“在能够正确划分背景与异常之前,往往需要采用一定的手段把背景的起伏变化扣除,把背景场转化成平面的”的同时,又反对移动标准化等方法中把背景场转化成平面场的作法;前面强调了系统误差校正的重要意义,后面又说:“如果只看到其中一个地球化学图,虽然该图幅与相邻图幅之间系统误差未经校正,有经验的化探专家仍可以通过分析得出合理的关于该图幅内矿产情况的结论”,此言本意并无大碍,但在讨论多图幅系统误差校正时,其言外之意令人费解,而 C 型转换正是在这种认识基础上提出的。

3.2 第二优点

“对策”说,“ C 型转换的另一个优点是:在二阶平稳条件下, C 型转换不影响元素之间的相互关系,因此不影响多元统计分析”,并用 C 型转换前后相关系数的不变性为依据,给出了一个证明。但是,在论证这一优点时,存在三个明显的缺点:

第一,“不影响元素之间相互关系”的含义不确切。元素间的相互关系,可以是在单图幅内的相互关系,也可以是在多图幅内的相互关系。在讨论相邻图幅误差校正效果时,区别这两种情形是必要的。但“对策”未加区分。

第二,“不影响元素之间的相互关系”的证明不完整。“对策”在证明这一结论时,只用了一个图幅,因上述原因,该证明显然不具完整性。

第三,将“不影响元素之间的相互关系”作为误差校正方法的优点不合理。系统误差可使两个给定元素由单图幅内的不相关变为多图幅内的负相关或正相关,也可由单图幅内的正相关变为多图幅内的负相关或不相关等。简言之,系统误差对元素间原本存在的相关关系有破坏作用。因此,一种好的误差校正方法应能对相关系数起校正作用,而不是“不影响元素之间的相互关系”。

从几何上看,在正态假设下,相关元素(变量)所成二维空间中的样品点呈单一椭圆分布。当存在系统误差时,不同图幅的样品点可形成多个椭圆,其中心、大小和方向分别由元素在各图幅的均值、方差和协方差或相关系数确定。一种好的误差校正方法,应该能把这些不同的椭圆校正到单一椭圆,即把各图幅的均值、方差和协方差或相关系数都校正到同一数值。本文未讨论协方差,但仅由上节给出的各方法对不同图幅均值、方差的校正效果不难看出,在校正相关系数或协方差时,同样是分区标准化的效果较好,衬度系数法次之,而C型转换较差。

“对策”还列举了重大项目的实例来说明C型转换“效果优于其它方法”,但由于对“其它方法”的曲解和C型转换理论中存在的上述问题,其可信程度有待进一步验证。

4 结论与讨论

1)移动标准化和分区标准化法被否定的原因是,“对策”疏忽了其基本概念、曲解了方法原意,将“背景均值和均方差”疏忽为全域“均值和均方差”,把“要移动窗口大小选择适当”曲解为“移动窗口大小必定要与研究区内各地质体的平均尺寸相当”。

2)“对策”颠倒了衬度系数法的优点和缺点;C型转换法继承了和扩大了其缺点。

3)C型转换不是校正而是很可能扩大图幅间的背景差异或系统误差。各图幅C背景均值间差异难以预料,C背景均方差依赖均值的微弱增加而大幅度扩大。

4)“对策”关于C型转换“不影响元素之间相互关系”的含义不确切,其证明不完整,将“不影响元素之间的相互关系”作为误差校正方法的优点也不合理。好的方法应能同时校正背景均值、方差和相关系数。

5)在上述各种误差校正方法中,理论上分区(幅)标准化方法效果较优,衬度系数法效果次之,C型转换效果较差。

本文强调了被批评方法的优点,但它们并非十全十美。例如,对背景分布参数只能作出近似估计,正态假设也往往只能近似满足,这在文献[2]中已作出暗示,也是统计分析的一般问题。此外,一元方法存在很大局限性。我们确信,化探数据处理将在多元方法方面取得突破。

本文也强调了“对策”的缺点,但其中并非一无是处。例如有关划分背景与异常的7条关键中就有六条值得重视。其大胆批评精神也深为笔者钦佩,因为严谨的学风和必要的学术讨论是避免错误、推动学术发展的动力之一。

[参考文献]

- [1] 陈明,李金春.化探背景与异常识别的问题与对策[J].地质与勘探,1999(2):25-29.
- [2] 纪宏金,仲崇学.区域化探中若干新图件的计算与制作[J].物探化探计算技术,1991(2):96-104.
- [3] 纪宏金,连长云,杜庆丰.地球化学数据的标准化与衬度变换[J].物探化探计算技术,1993(1):19-25.
- [4] 陈永清,纪宏金.标准化区域地球化学图的编制方法及应用效果[J].长春地质学院学报,1995(2):216-221.

A DISCUSSION ABOUT SOME DATA PROCESSING METHODS IN GEOCHEMICAL EXPLORATION

Ji Hong-jin, LIN Rui-qing, ZHOU Yong-chang

Abstract: There are many methods to revise system error and identify anomaly in geochemical data processing e.g. moving standardization, partition standardization and contrasting coefficient. Those methods have been briefly introduced and compared in application effect. Finally, Merit of C-type transformation method is emphasized discussed.

Key words: geochemical exploration, correction of systematic error, recognition of anomaly, standardisation-transformation, C-transformation

[第一作者简介]

纪宏金(1950年-),男,1976年毕业于长春地质学院物探系,现任吉林大学地球探测与信息技术学院应用地球化学系副教授。主要从事勘查地球化学的教学与科研工作。

通讯地址:长春市地质宫 吉林大学地球探测与信息技术学院地球化学系 邮政编码:130026