

文章编号: 0254- 5357(2004) 01- 0052- 04

用稳健方差分析评估地球化学调查 土壤中有机质的采样质量和分析质量

江泓

(福建省地质测试研究中心, 福建 福州 350002)

摘要: 用稳健方差分析评估地球化学调查土壤中有机质的采样和分析质量, 表明稳健方差统计比传统方差统计更能反映地球化学场的真实分布, 提出重视深部土壤采样设计方案。

关键词: 稳健方差分析; 地球化学调查; 土壤有机质; 质量评估

中图分类号: O213.1

文献标识码: A

地球化学调查通过重复采样和重复测试结果的方差分析或三重套叠方差分析来评估采样质量和实验室的分析测试质量^[1]。

传统方差分析受少部分的异常值影响很严重^[2~4]。它过分地强调少部分异常值, 使估计的标准偏差偏大, 从而导致不合理的结果。如果采用统计办法将异常值舍弃, 虽然提高了统计的精密性, 却低估了标准偏差, 产生错误的测试结果的重复性和再现性。现代质量管理理论认为在控制条件下产生的测试结果都有其存在的合理性, 如果因原因不明就用数学统计方法将异常值舍弃是没有道理的^[5]。Thompson、Rammsey 和董德明等人应用英国分析方法委员会 (Analytical Methods Committee AMC) 推荐的稳健方差分析方法 (Robust ANOVA)^[4,6]来评估采样和分析质量, 提出“基于稳健方差理论的全程质量控制模式 (Sampling and Analytical Quality Control Scheme, SAX)”^[5,7~10]。他们通过模拟试验证明这种方法比传统的方差统计结果更可靠, 并应用于环境土壤调查, 取得不少成果^[3,7,11~17]。何锡文等也对它作了详细的介绍和研究^[3,17]。稳健方差分析法对异常值不是舍弃它, 而是在统计中根据它偏离中心值的程度, 确定它的权, 从而减小它对结果的影响。全程质量控制是地球化学调查的重要手段^[18]。

一批测试数据所反映的某一组分在地球化学

场的空间分布的离散性 (σ_{total}) 由三个部分组成^[5]。首先是该组分空间分布的真实离散性 (σ_g), 它是该地球化学场本身所固有的特性。在特定的条件下它是固定不变的, 也是我们所要求的。第二是采样误差带来的离散性 (σ_s), 即同一采样点, 不同采样位置 and 采样操作所产生的误差。第三是分析误差产生的离散性 (σ_a)。理论证明三者互相独立, 所以它们的方差具有加和性:

$$\sigma_{\text{total}}^2 = \sigma_g^2 + \sigma_s^2 + \sigma_a^2 \quad (1)$$

通过方差分析可以解析出样本的标准偏差 s_{total} 、 s_g 、 s_s 和 s_a (符号下标意义同 σ) 作为 σ_{total} 、 σ_g 、 σ_s 和 σ_a 的估计值。

从满足地球化学调查质量控制的目的出发, s_g 、 s_s 和 s_a 三者之间应有一定的比例关系, 这样符合所谓的 FFP (fitness for purpose) 原则^[19]。它既可防止浪费大量的人力和物力, 盲目地追求尽量小的 s_s 和 s_a , 又能防止对 s_s 和 s_a 要求太宽, 使分析数据不能正确反映地球化学场真实面貌。董德明等人根据实践经验^[10]建议, 控制的临界值为:

$$1\% s_{\text{total}}^2 < s_{\text{meas}}^2 < 20\% s_{\text{total}}^2 \text{ 和}$$

$$1\% s_{\text{meas}}^2 < s_a^2 < 20\% s_{\text{meas}}^2 \quad (2)$$

式中, $s_{\text{meas}}^2 = s_s^2 + s_a^2$ 称为测试方差。因此, $s_{\text{total}}^2 = s_g^2 + s_{\text{meas}}^2$ 。如果测试方差与总方差的比在 1% ~ 20%, 表明采样过程和分析过程处于受控状态。若

收稿日期: 2003-05-27; 修订日期: 2003-07-25; 作者简介: 江泓 (1943-), 男, 福建福州人, 高级工程师, 岩矿测试和质量管理工作。

c 中国地质调查局. 覆盖区多目标地球化学调查暂行规定 (DD2002)。

它们之比大于 20%, 则组分的空间分布的离散性将被测试离散性所掩盖, 将得不到真实的空间分布图。同样, 若分析方差与测试方差之比大于 20%, 也即占总方差的 4% 以上, 则分析的误差占全程质量控制的主导地位, 也会使组分的空间分布图失真。当然, 要求 s_{meas}^2 小于 s_{total}^2 的 1% 或 s_a^2 小于 s_{meas}^2 的 1% 也是毫无必要的。

本文试图用稳健方差分析法来评估某地球化学调查报告中有有机质的采样质量和分析质量。

1 数据来源和处理

该地球化学调查项目共送 2 775 个样品到实验室, 其中浅层土样 2 229 个, 深层土样 546 个。根据有关规范^[1]的要求, 浅层土样包含 44 组重复采样和重复分析样品共 88 对, 深层土样包含 10 组重复采样和重复分析样品共 20 对样。它们以密码方式随机分布在样品中。在严格的质量监控条件下测试后, 实验室提交了 2 775 个样品有机质的测试数据。从送样单位的采样记录中抽出 54 组的重复采样, 108 对重复测试, 216 个样品的样号和分析数据做稳健方差分析统计。

统计是在电子表格上运行。

2 结果与讨论

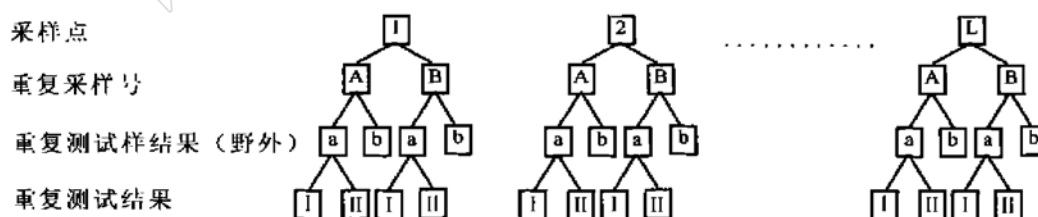


图 1 采样示意图

Fig. 1 Schematic diagram for sampling

表 2 浅层土壤样品稳健方差统计结果

Table 2 The results of robust ANOVA for top layer soils

项 目	$\bar{w}$ (org. m)	s_a^2	s_s^2	s_{meas}^2	s_g^2	s_{total}^2
分析结果 / 10^{-2}	1.350	0.028 3 ²	0.097 9 ²	0.101 9 ²	0.509 5 ²	0.519 6 ²
s_a^2 占 s_{meas}^2 的比例 (%)		7.727	92.273	100	—	—
s_a^2 占 s_{total}^2 的比例 (%)		0.297	3.555	3.852	96.148	100

表 2 可见, 浅层样品的采样和分析的精密密度比较高, 它们的方差相对于有机质在本测区内空间分

2.1 数据处理结果

浅层样品测试结果用稳健方差统计的测区有机质的平均值和其它参数见下表 1。

表 1 稳健方差统计结果

Table 1 The Results of robust ANOVA

项 目	$\bar{w}$ (org. m)	s_a^2	s_s^2	s_{meas}^2	s_g^2	s_{total}^2
分析结果 / 10^{-2}	1.362	0.062 9 ²	0.069 7 ²	0.093 9 ²	0.519 0 ²	0.527 4 ²
占 s_{meas}^2 的 比例 (%)		44.829	55.171	100	—	—
占 s_{total}^2 的 比例 (%)		1.403	1.766	3.169	96.830	100

表 1 第 3 行显示分析的方差占测试方差的比例达 44.8%, 这与实验室内部的平行测试质量监控的结果不相符。经调查, 原来重复分析样品是在野外将样品缩分成两份, 再送实验室分别加工测试的。由于缩分时样品颗粒比较粗, 加上野外人员的混匀缩分技术等原因, 分样的代表性差。分样的误差叠加到分析过程误差中, 使分析误差比例变大。

为此, 按照采样示意图(图 1)将重复测试样品的 a 号样全部取出, 重新编成密码, 交分析人员重新测试, 测试结果为结果 II。原来 a 号样测试的结果为结果 I。将此重复测试结果, 重新作稳健方差分析统计, 统计结果见下表 2 和表 3。

布的不均匀性来说都很小 (仅占 3.852%)。分析质量和采样质量均能满足地球化学调查的要求。

表 3 深层土壤样品稳健方差统计结果

Table 3 The results of robust ANOVA for deep layer soils

项 目	$\bar{w}$ (org. m)	s_a^2	s_s^2	s_{meas}^2	s_g^2	s_{total}^2
分析结果 / 10^{-2}	0.482	0.025 1 ²	0.127 8 ²	0.130 2 ²	0.216 6 ²	0.252 8 ²
s_a^2 占 s_{meas}^2 的比例 (%)		3.712	96.288	100	—	—
s_a^2 占 s_{total}^2 的比例 (%)		0.986	25.563	26.549	73.451	100

但表 3 的第 4 行显示深层样品的测试方差占总方差的比重为 26.55%, 大于 20% 限定值。从第 3 行可见, 测试误差主要由采样误差产生(占测试方差的 96.29%, 总方差的 25.56%)。在土壤深层, 人类活动影响很小, 生物遗体腐烂后分布不均匀, 造成局部的有机质分布不均匀, 因而采样误差

变大。这个误差可能模糊了深层土壤有机质的空间分布。但它可以通过设计更合理的采样方案, 比如, 增加重复采样次数^[2,3]或加密采样点来解决。

2.2 稳健方差分析结果与传统方差分析结果比较

上述所得的数据也用传统的方差分析处理, 将两种统计方法所得的结果列于表 4。

表 4 稳健方差分析与传统方差分析比较

Table 4 The comparison of the results from robust with these from classical ANOVA

$w_B/10^{-2}$

样品 sample	统计方法 method	$\bar{w}(\text{org. m})$	s_a^2	s_s^2	s_{meas}^2	s_g^2	s_{total}^2
浅层 surface	稳健 robust	1.35	0.028 3	0.097 9	0.101 9	0.509 5	0.519 6
	传统 classic	1.36	0.027 6	0.122 4	0.125 4	0.507 1	0.516 2
深层 deep layer	稳健 robust	0.48	0.025 1	0.127 8	0.130 2	0.216 6	0.252 8
	传统 classic	0.82	0.035 9	0.270 7	0.273 1	0.879 3	0.902 5

统计结果表明, 对于浅层土壤而言, 采样和分析质量都比较好, 两种方法统计的数据比较接近。深层样品统计数据差异比较大, 平均值相差了近一倍。

深层 10 个采样点测试结果列于表 5。

表 5 深层 10 个采样点测试结果

Table 5 The analytical results from 10 sampling spots of deep layer soils

采样点 spot	重复采样 sampling	$\bar{w}(\text{org. m})/10^{-2}$	
		结果 I analysis I	结果 II analysis II
1	A	0.26	0.26
	B	0.33	0.34
2	A	0.63	0.51
	B	0.56	0.51
3	A	0.43	0.44
	B	0.38	0.38
4	A	0.22	0.18
	B	0.29	0.28
5	A	0.34	0.34
	B	0.21	0.25
6	A	0.39	0.37
	B	0.74	0.73
7	A	1.74	1.75
	B	2.70	2.70
8	A	2.41	2.49
	B	3.02	3.16
9	A	0.48	0.44
	B	0.55	0.50
10	A	0.38	0.36
	B	0.34	0.34

表 5 可见, 平行测试的结果 I 和结果 II 之间偏差很小, 但同一点重复采样点 A 和 B 两结果差异比较大。其中第 7 和 8 采样点有机质含量远高于平均值, 且重复采样测试结果偏差也较大。传统方差统计在处理这 10 个点的数据时, 不加区别地等

同对待, 稳健方差统计则根据偏差大小采用权的处理, 克服了这两个异常点过分的影响, 所以整个测区的有机质平均含量不是 0.82%, 而是 0.48%。其它统计数字也是这样。它说明当测试对象在地球化学场中分布不均匀时, 采用稳健方差统计数据比较合理。

2.3 异常值的判别

方差统计理论^[3]认为, 当

$$|X_{iA} - X_{iB}| / \sqrt{MS_s} > 2 \quad (3)$$

时重复采样可能有异常值存在。式中 MS 为各参数的平均平方值, MS_s 可从 $S_s^2 = (MS_s - MS_a)/2$, $S_a^2 = MS_a$, 两式中求得。 MS_k 为各参数的平均值, 下标 k 为 $a, s, \text{meas}, g, \text{total}$ 。 X_{iA} 、 X_{iB} 分别表示第 i 采样点两次重复采样分析结果。将稳健方差和传统方差统计的 MS_s 代入式(3)计算的异常临界值:

$$|X_{iA} - X_{iB}| < 2 \sqrt{MS_s} \quad (4)$$

分别为 0.365% 和 0.769%。表 5 中第 7、8 两点的重复采样 A 和 B 结果偏差分别为 0.95% 和 0.64%, 都大于稳健方差的异常临界值, 而传统方差统计只能判别第 7 点重复采样有异常。通过异常的识别, 可以对 7、8 两点样品进行复验检查, 同时进行现场调查寻找异常存在的原因。

同样, 当

$$|X_i - \bar{X}| > 2 \sqrt{(L-1)MS_g/(4L)} \quad (5)$$

时, 该点可能存在有机物含量异常。本次测定中 $L = 10$, 稳健方差和传统方差统计的异常临界值为 0.22% 和 0.85%。稳健方差统计认为第 4 采样点为含量异常低点, 7 点和 8 点为高异常点。而传统方差统计只识别出 7 和 8 两点为高异常点。传统

方差统计在计算测区有机质含量平均值时,一种是全部数值等同地参加平均。本例中有机质平均含量为0.82%,明显高估了平均值;一种按规范^c用剔除异常值后的算术平均值,本例中如果剔除7、8两点的数值,求得的平均值为0.40%,可能会造成低估的效果。如果按规范^①将大于原始数据平均值加减两倍标准偏差的值为异常值(本例中为 $0.82\% + 1.80\% = 2.62\%$),剔除第8点数值,平均值为0.60%,也有偏高的趋势。但这些剔除仅仅是统计上的意义,它否认了客观存在,因为这两点确实存在着高含量的有机质;若是按规范^c取原始数据的中位值为平均值,虽然0.45%与稳健方差统计值相近,但取中位值则完全忽略了异常值的作用。说明稳健方差分析在控制采样质量和识别地球化学异常上比传统方差分析优越。

3 结语

用稳健方差分析对某地球化学调查项目的有机质含量进行全程质量评估,证明对于浅层土壤样品,采样质量和分析质量都满足质量要求,而深层土壤样品的采样误差比较大,可能会掩盖了有机质空间分布的真实面貌。通过这个实例比较稳健方差分析与传统方差分析,说明稳健方差分析在全程质量控制上比传统方差分析优越。

上述分析也说明,在大面积的地球化学调查开始前,先小规模地进行采样试验,再用稳健方差统计来分析,然后制定可靠的采样方案,对于提高地球化学调查的质量是很有意义的。

4 参考文献

- [1] 中华人民共和国地质矿产行业标准 DZ/T 0167-1995, 区域地球化学勘查规范(比例尺 1:200000) [S].
- [2] Thompson M, Magurie M. Estimating and Using Sampling Precision in Surveys of Trace Constituents of Soils [J]. *Analyst*. 1993, 118: 1107-1110.
- [3] 何锡文, 郭薇, 胡军. 稳健方差法探讨土壤取样[J]. *分析化学*. 1997, 25(3): 253-257.
- [4] Analytical Method Committee. Robust Statistics-How Not to Reject Outliers Part I. Basic Concepts [J]. *Analyst*. 1989, 114(12): 1963-1967.
- [5] Rammsey M H, Argyraki T, Thompson M. Estimation of Sampling Bias Between Different Sampling Protocols on Contaminated Lands [J]. *Analyst*. 1995, 120: 1353-1356.
- [6] Analytical Method Committee. Robust Statistics-How Not to Reject Outliers Part II. Inter Laboratory Trials [J]. *Analyst*. 1989, 114(12): 1969-1702.
- [7] Rammsey M H. Sampling and Analytical Quality Control (SAX) for Improved Error Estimation in the Measurement of Pb in the Environment Using Robust Analysis of Variance [J]. *Appl Geochem*. 1993, suppl(2): 149-153.
- [8] Thompson M, Rammsey M H. Quality Concepts and Practices Applied to Sampling-An Exploratory Study [J]. *Analyst*. 1995, 120: 261-270.
- [9] 董德明, 房春生, 王菊, 等. 土壤样品采集和化学元素分析全过程质量控制比较研究[J]. *分析化学*. 2002, 30(10): 1158-1162.
- [10] 董德明, 房春生, 朱先磊. 环境样品的采集和分析质量保证研究的现状与进展[J]. *吉林大学科学学报*. 2000, 4(3): 85-90.
- [11] Squire S, Rammsey M H, Gardner M J. Sampling Proficiency Test for the Estimation of Uncertainty in the Spatial Delineation of Contamination [J]. *Analyst*. 2000, 125(5): 2026-2031.
- [12] Squire S, Rammsey M H, Gardner M J. Collaborative Trials in Sampling for the Spatial Delineation of Contamination and the Estimation of Uncertainty [J]. *Analyst*. 2000, 125(1): 139-145.
- [13] Deming D, Xiaolei Z, Zhonglei X, et al. Assessment of Sampling and Analytical Quality Control in Determination of Al in Soils Using Chemical Extraction [J]. *Microchem*. 1999, 62(3): 386-393.
- [14] Deming D, Rammsey M H, Thornton I. Sampling and Analytical Quality Control of the Determination of Aluminum in Soybean leaves [J]. *Analyst*. 1997, 122(5): 421-424.
- [15] 董德明, 杨彬, 刘淼. 化学连续浸提法对土壤中 Zn 和 Mn 化学形态的研究[J]. *吉林大学自然科学学报*. 1998, 1: 62-66.
- [16] 谢志雷, 董德明, 杜尧园, 等. 茶叶铝含量与茶园土壤 pH 值的关系[J]. *吉林大学自然科学学报*. 1998, (2): 89-92.
- [17] 高志, 何锡文, 李一峻. 测试取样理论的进展[J]. *分析化学*. 2000, 28(4): 497-506.
- [18] 董德明, 朱先磊, 房春生, 等. 土壤样品分析全过程质量控制方法及应用计算程序[J]. *吉林大学自然科学学报*. 2000, (1): 83-86.
- [19] Thompson M, Fearn T. What Exactly is Fitness for Purpose in Analytical Measurement [J]. *Analyst*. 1996, 121: 275-278.

(英文下转第61页)

① 中国地质调查局. 覆盖区多目标地球化学调查暂行规定(DD2002).

(英文上接第 55 页)

Estimation on Sampling and Analytical Quality in Geochemical Survey for Organic Materials in Soils with Robust ANOVA

JIANG Hong

(The Geoanalytical Laboratory of Fujian, Fuzhou 350002, China)

Abstract: Estimation of sampling and analytical quality in geochemistry survey for organic materials in soils with robust ANOVA is reported in this paper. The results show that the estimation results from robust ANOVA much truly reflect the distribution of geochemical field than that from classical ANOVA. And the results also show that much attention should be paid to the design of deep soil sampling.

Key words: robust ANOVA; geochemistry survey; organic material in soil; quality estimation