

文章编号: 0254-5357(2013)01-0096-05

正态分布在区域地球化学调查样品分析质量评价中的应用探讨

赵志飞, 闫 晖, 姚 岚, 方金东, 唐兴敏

(湖北省地质实验研究所, 湖北 武汉 430034)

摘要: 区域地球化学样品长期以来一直是采用间接方法对实际样品元素分析质量进行评价, 有效地提高了区域地球化学样品的整体分析质量, 但是当实际样品和外部质量控制样品含量范围差别较大时, 间接评价的效果会减弱。本文从区域地球化学样品大部分元素符合标准(对数)正态分布的规律出发, 运用 SPSS、Excel 等软件对区域地球化学调查样品元素含量正态分布情况进行研究, 通过计算元素含量实际正态分布图和标准正态分布图的重合度直接评价元素的分析质量。运用整套方法对实验室分析的江西某地多目标区域地球化学样品的分析数据进行了质量评价, 正态分布检验表明微量元素需要通过取对数才能进行正态分布分析, 该地区钽、钨、铈等元素不适合用本方法进行质量评价; 钴、汞、镓等 18 个元素的重合度都在 0.9 以上; 氧化钙和氧化钠重合度小于 0.9, 样品实际结果分析表明需降低分析方法检出限, 提高低含量氧化钙和氧化钠的分析质量。本方法对区域地球化学调查样品的分析数据质量的直接评价作了有益的尝试, 可以作为现有区域地球化学调查样品质量评价办法的参考和补充。

关键词: 区域地球化学调查样品; 质量评价; 正态分布; 对数正态分布

中图分类号: O213

文献标识码: A

我国开展区域地球化学调查工作以来, 化探样品的实验室分析质量得到了高度的重视, 首先在全国范围内研制了水系沉积物、土壤和岩石系列标准物质, 制定了实验室内部质量监控方案, 经过不断的补充和完善逐渐形成了地质矿产行业规范^[1-2]; 21 世纪初期开展的多目标区域地球化学调查融合了基础地质、农业地质和环境地质, 分析元素数量有所增加, 精度、准确度等质量要求进一步提高; 叶家瑜等^[3]提出的实验室外部和实验室内部质量控制相结合的方案在多目标区域地球化学调查中得到了实施, 使区域地球化学样品的分析质量控制方案进一步得到完善。全国地质实验室在样品分析时对质量的控制都相当重视, 除了按照相应规范进行了内部和外部质量监控外, 还各自采取了一些其他措施来提高区域地球化学样品分析质量^[4-7]。管理办法的完善使得区域地球化学样品分析技术水平和质量较以前有了较大的提高, 高质量的分析数据取得了较好的元素成图效果, 从而使我国区域地球化学调查工作在矿产勘查、基础地质、生态农业环境等方面取得了突出的成果。

正态分布是统计学领域最重要的连续概率分布模式。地质统计学研究中也时常运用正态分布模型

对实验现象进行研究分析^[8-13], 区域地球化学研究的主要对象为土壤和水系沉积物; 在同一地质体区域内, 将土壤和水系沉积物中元素的含量作为研究变量, 取样数量越多, 元素含量的概率分布形式越趋近于正态分布(对数正态分布)。以往研究也表明^[14]: 单一地球化学作用产物中元素多呈对数正态分布, 多次地球化学作用综合产物中元素分布形式趋于正态分布; 常量元素大多数服从正态分布, 微量元素大多服从对数正态分布。

现行样品质量控制方案都是以加入的平行样、内部控制样、外部控制样的质量水平, 从而对实际样品的分析质量进行间接评价。本文利用元素正态分布的特征, 对江西某地多目标区域地球化学样品的实验室分析数据进行含量概率作图, 从实际图形和标准图形的拟合度对数据分析质量的直接评价作一些探讨。

1 评价方法

1.1 样品分析数据的选择

本文以本实验室分析的江西某地多目标区域地球化学调查样品为研究对象。多目标区域地球化学

收稿日期: 2012-03-27; 接受日期: 2012-08-17

作者简介: 赵志飞, 工程师, 地球化学样品分析研究。E-mail: zzfeel@126.com。

调查规范规定,实验室实际分析样品每 50 件中插入 2 件重复样、4 件外部标准控制样、4 件精密度控制样品。由于正态分布检验的是该地区实际样品的概率分布,参加数理统计的样品应该剔除上述重复样和质量控制样后的样品。实际作图也反映出人为地插入样品改变了实际样品的分布规律。图 1 为剔除监控样后的 SiO_2 实际样品含量频率直方图,符合正态分布的特点,图 2 为 SiO_2 全部样品含量频率直方图。从图 2 可以看出:在插入的 4 个内部监控样含量(62.6%、73.35%、74.72%、58.61%)的附近频率分布得到了增加,改变了分布趋势。本文后续所有统计样本都剔除了监控样和重复样。

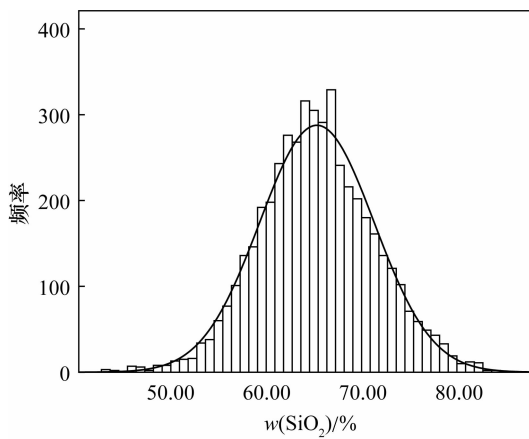


图 1 二氧化硅实际样品的直方图

Fig. 1 Histogram of SiO_2 in real samples

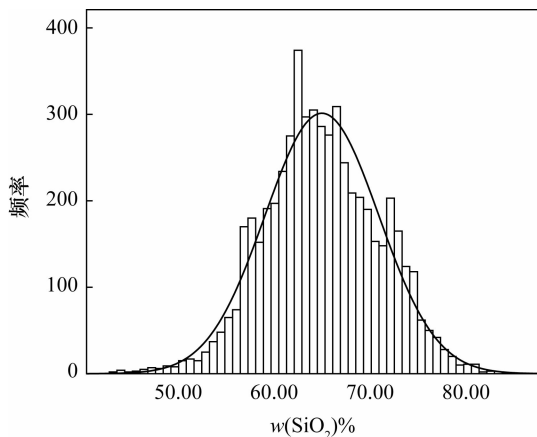


图 2 二氧化硅全部样品的直方图

Fig. 2 Histogram of SiO_2 in all samples

1.2 元素正态分布检验方法

地质样品中元素正态分布和对数分布的成因多有探讨。一般认为,正态分布是大量相互独立又相对微小的随机变量共同作用的结果,反映的是一个

渐变过程和平稳过程;对数正态分布是某些因素起了突出作用但是未起到左右全局的结果,典型矿区个别地质因素主导了成矿元素的形成和分布,成矿元素的分布往往呈现偏态分布,没有标准建模形式,因此不经过转换无法运用标准正态模型进行评价。区域地球化学调查虽然是以采集汇水盆地下游的水系沉积物和第四纪覆盖区的土壤为分析介质,地质作用上大多数反映的是长时间的渐变过程,但是也不乏突变地质作用对一些元素的影响,因此对原始数据进行正态检验,考察评价地区的所有元素是否符合正态分布是后续工作的前提^[15-16]。

最常用的数理统计检验软件 SPSS 中对正态分布进行检验有计算法和图示法。计算法包括峰度系数检验、偏度系数检验、Kolmogorov - Smirnov 检验 (D 检验)和 Shapiro - Wilk 检验 (W 检验);图示法包括 $Q-Q$ 图、 $P-P$ 图、直方图等^[17-19]。SPSS 中 D 检验和 W 检验对样品数量有很强的使用范围 ($n \leq 1000$),本文样本数量较大 ($n = 4762$),采用图示法,将进行检验的数据导入 SPSS 软件中,运用分析 - 描述性统计 - 探索的技术路线,检验样本的正态分布特性,分析数据近似呈一条直线说明被检验的数据服从正态分布。图 3 和图 4 分别是钴元素的 $Q-Q$ 图和 $P-P$ 图,期望累计概率和观测值近似为直线,可以认为考察地区的钴元素符合对数正态分布。

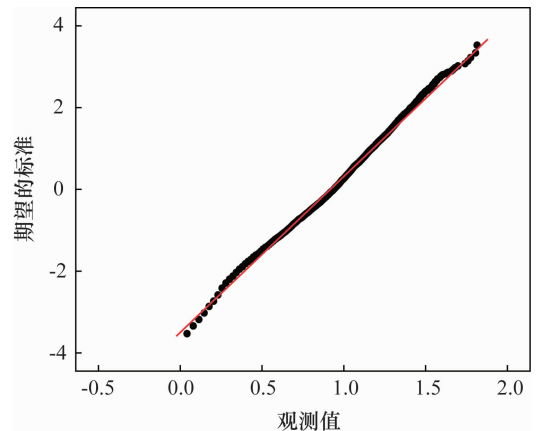


图 3 钴元素的 $Q-Q$ 正态分布图

Fig. 3 $Q-Q$ plots of Co

1.3 正态分布图的制作与比较

将每个元素的全部分析数据按照标准或者对数转换后,计算转换后样本的平均值 μ 、标准偏差 σ ,两者确定后根据正态分布公式 $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$,样本的标准正态分布图就唯一确定;将样本数据在平均

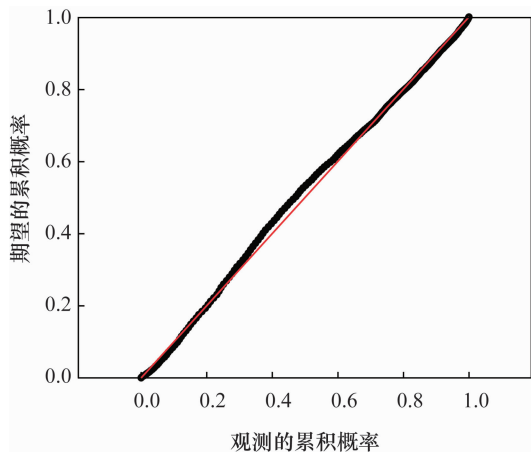


图 4 钴元素的 P-P 正态分布图

Fig. 4 P-P plots of Co

值 μ 上下按照一定组距分成若干组,运用 Excel 中 Countif 函数统计样本在每组间隔之内的数据数量,根据总量计算统计点的概率密度,制作实际正态分布图;运用 Normdist 函数计算统计点的标准概率密度值,制作样本的标准正态分布图^[20]。将钴元素含量进行对数转换,统计计算样本平均值 $\mu = 0.917$,标准偏差 $\sigma = 0.262$ 。按照上述方法作图,见图 5。

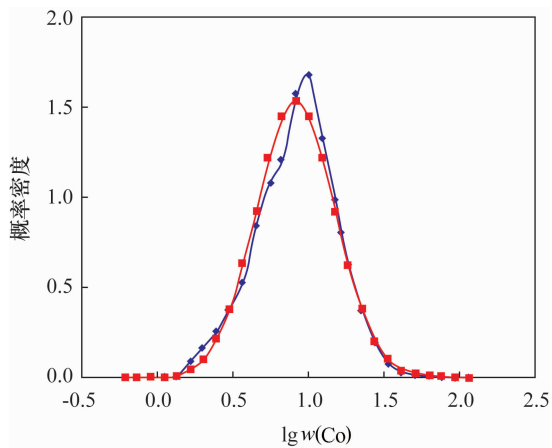


图 5 钴元素的正态分布图

Fig. 5 Normal distribution of Co

红色曲线为标准正态分布趋势,蓝线曲线为真实的概率分布趋势。

2 结果与讨论

2.1 元素分析数据的正态分布检验

一般来讲微量元素的对数呈正态分布,常量元素服从标准正态分布;对于特殊矿区微量元素的对数也需要经过转换才符合正态分布。在江西多目标区域地球化学样品元素正态分布检验中,一些土壤的常量组分氧化镁、氧化钙、氧化钠等因在这一地区含量普遍偏低(平均含量 $< 0.5\%$),镁、钙、钠三个

元素需经过对数转换后才呈正态分布。

图 6 为常量组分氧化镁没有经过对数转换的 Q-Q 检验图,图 7 是氧化镁经过对数转换后的 Q-Q 检验图。从图 6 可以看出常量元素应该按照测量样本的平均含量来判定,不能仅仅按照平时的主微量元素分类笼统地加以区分;对微量元素应该首先将对数转换后的数据进行 Q-Q 检验,对于本地区钛、钨、铬、铷、钍等线性较差的元素,不能单纯地用本文办法进行质量评价。

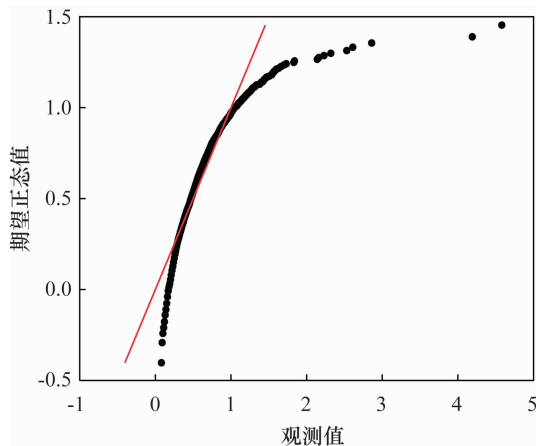


图 6 氧化镁的 Q-Q 正态分布图

Fig. 6 Q-Q plots of MgO

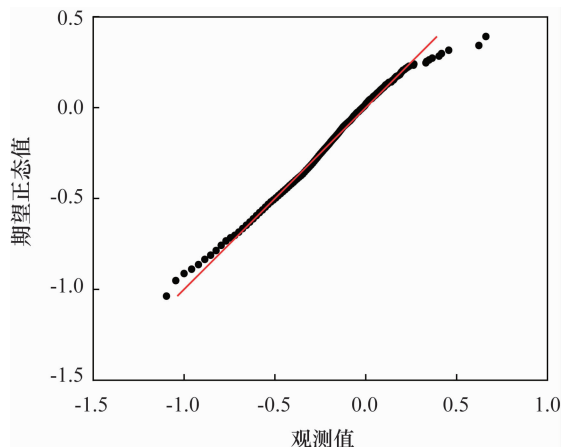


图 7 lgw(MgO) 的 Q-Q 正态分布图

Fig. 7 Q-Q plots of lgw(MgO)

2.2 正态分布图组数和组距

正态分布的组数 n 和组距 h 的确定通常有两种方法:一种方法是建立在个人经验基础之上的主观判断来进行;另一种方法是根据 Sturges 经验公式计算分组,组数 $K = 1 + 3.322 \lg n$ 。两种方法都存在缺陷,第一种方法很难做到分组的唯一性,第二种方法

只考虑了数据的个数,当总体单位数过少时,计算的组数可能偏多,而当总体单位数很多时又可能偏少。

标准正态分布中样本(均值 ± 3 倍标准偏差)之内的面积占全部面积的 99.7%,即 $P(\mu - 3\sigma < X < \mu + 3\sigma) = 99.7%$;图形中峰值左右各 3 个点决定了正态分布的大概形状。本批样品数量 $n = 4762$,根据 Sturges 公式计算得到组数 $K = 13.2$,组间距 h 在 $\sigma/3 \sim \sigma/2$ 。考虑到重合度计算时多个数据点能够更详细地反映真实情况,本样本组间距 h 选择为 $\sigma/3$,共 18 组数据绘制图形,对分析质量可以有一个比较真实的评价。

2.3 元素分析数据的质量评价

根据绘制的实际样品正态分布图和标准正态分布图,可以对分析数据的整体质量和分析细节进行评价。运用 Excel 函数计算元素实际正态分布图和标准正态分布图的重合度(完全重合作 1)可以对元素的整体分析质量进行一个评判,根据图中每点实际和标准图形概率密度的差异,返回元素含量范围可以评价各个含量段元素的测量质量。

表 1 列出了正态分布检验较好的 20 个元素实际正态分布图和标准正态分布图的重合度,可以看出氧化钙、氧化钠重合度均 < 0.9 ,分析数据结果表明江西该地区氧化钙、氧化钠元素含量较低,全谱直读等离子体发射光谱测定氧化钙、氧化钠的检出限为 0.05%,分析方法 3 倍检出限以内的含量范围分析结果误差较大,本地区全部样品中氧化钙和氧化钠含量小于 0.15% 的样品分别有 3425 件和 2372 件,样品分析结果含量不在仪器优势分析含量范围内,需要降低分析方法检出限,提高氧化钙、氧化钠含量在 0.15% 以下的样品分析质量。

表 1 正态分布图重合度

Table 1 Overlap ratio of normal distribution

元素	重合度	元素	重合度
Co	0.942	Ga	0.952
Hg	0.934	Sc	0.951
Li	0.920	Cl	0.905
Mn	0.973	Br	0.941
Pb	0.930	SiO ₂	0.968
Sr	0.902	Al ₂ O ₃	0.946
Zn	0.951	TFe ₂ O ₃	0.963
Ge	0.938	MgO	0.935
Ce	0.950	CaO	0.843
Se	0.902	Na ₂ O	0.856

图 8 为氧化钠经对数转换后的正态分布图,从氧化钠的实际(蓝色)和标准(红色)正态分布图可以看出,同一横坐标两点差别较大时说明此区间含量范围样品分析结果不太理想。 $\lg w(\text{Na}_2\text{O})$ 为 $-1.12 \sim$

-0.78 之间的四个点相差较大,即氧化钠含量为 0.07% ~ 0.16% 的一些检测结果不理想,此部分含量在 3 倍仪器检出限之内,样品分析结果准确度不高。

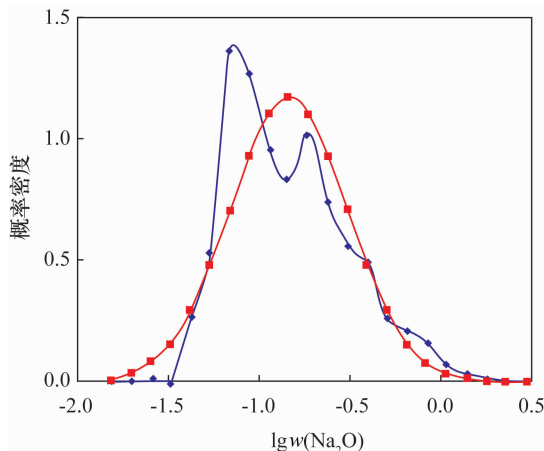


图 8 氧化钠的正态分布图

Fig. 8 Normal distribution of Na₂O

3 结语

由于地质作用的长期性和多因素性,区域地球化学调查样品中大多数元素的含量符合标准(对数)正态分布。本文运用 SPSS 统计软件中 P-P 和 Q-Q 检验方法对主元素和对数转换后的微量元素进行了正态分布检验,通过正态分布检验的 20 项元素的实际正态分布图和标准正态分布图进行了重合度计算,结果表明钴、铅等 18 项元素分析质量较好,实际样品中氧化钙、氧化钠含量较低,需要通过降低分析方法检出限来提高元素分析质量。

本方法不受样品元素含量范围的影响,对不同含量的样品作出有效的评价,可以作为实验室内部质量控制的参考和区域地球化学样品质量管理办法、实际地球化学成图效果的一个补充,对分析测试质量进行更为全面的评价和分析。但由于地质作用的突变性,本办法对于不太符合正态分布的元素不能作出有效的评价,需要探讨元素正态转换方法,转换后通过正态分布检验可以用本法进行评价;同时本办法无法分辨和评价样品分析结果的系统误差,因为样品分析结果的实验室内系统误差只能改变样品含量平均值(即正态分布图峰值横坐标),使图形 X 轴平移,而对概率图形没有任何改变。

4 参考文献

- [1] 中华人民共和国地质矿产部. DZ/T 0167—1995, 区域地球化学勘查规范(比例尺 1: 200000) [S].
- [2] 中国地质调查局. DD 2005-01, 多目标区域地球化学调查规范(1: 250000) [S].

- [3] 叶家瑜,姚岚. 区域地球化学样品分析质量控制方法探讨[J]. 岩矿测试,2004,23(2): 137-142.
- [4] 杨锦发. 多目标生态地球化学土壤样品高精度测试与质量控制[J]. 岩矿测试,1997,18(3): 36-39.
- [5] 王连和,李曼. 区域地球化学样品分析质量管理计算机控制[J]. 岩矿测试,2008,27(3): 143-147.
- [6] 陈立军,汪涛. 1:20万区域化探样品的分析测试质量控制[J]. 岩矿测试,2004,23(2): 219-222.
- [7] 王焯. 应用6 σ 质量管理方法尝试评价多目标样品氯等项目检测质量水平[J]. 岩矿测试,2011,30(3): 281-284.
- [8] 吕连宏,张征,迟志森,李道峰,尚晓颖. 地质统计学在环境科学领域的应用进展[J]. 地球科学与环境学报,2006,28(1): 101-105.
- [9] 李黎,王永刚. 地质统计学应用综述[J]. 勘探地球物理进展,2006,29(3): 163-169.
- [10] 王学松,秦勇. 利用对数正态分布图解析徐州城市土壤中重金属元素来源和确定地球化学背景值[J]. 地球化学,2007,36(1): 98-102.
- [11] Huang Y X, Dagne G. A bayesian approach to joint mixed-effects models with a skew—Normal distribution and measurement errors in covariates [J]. *Biometrics*, 2011,67(1): 260-269.
- [12] Zikovskiy L, Chah B. The lognormal distribution of radon concentration in ground water[J]. *Ground Water*,1990,28(5): 673-676.
- [13] 李松,饶竹,黄毅,贾静. 正态分布检验在地下水检测质量控制中的应用[J]. 岩矿测试,2010,29(5): 585-589.
- [14] 罗先熔,文美兰,欧阳菲,唐甲光. 勘查地球化学[M]. 北京:冶金出版社,2008: 5-6.
- [15] 李晓晖,袁峰,白晓宇,张明明,贾蔡,周涛发. 典型矿区非正态分布土壤元素数据的正态变换方法对比研究[J]. 地理与地理信息科学,2010,26(6): 102-105.
- [16] 韩春明. 对数正态分布密度函数有关参数的计算及其成因讨论[J]. 新疆工业学院学报,2008,27(3): 222-226.
- [17] 朱红兵,何丽娟. 在SPSS10.0中进行数据资料正态性检验的方法[J]. 首都体育学院学报,2004,16(3): 123-125.
- [18] 韩永志. 统计学在理化检验中的应用 第六讲 正态分布及其检验 [J]. 理化检验: 化学分册,2000,36(2): 98-102.
- [19] 韩永志. 统计学在理化检验中的应用 第六讲 正态分布及其检验(续) [J]. 理化检验: 化学分册,2000,36(3): 138-141.
- [20] 刘清珺,陈婷,陈舜琮,刘清. 正态分布积分近似计算公式及其在实验结果判定中的应用[J]. 现代测量与实验室管理,2009(3): 21-23.

Application of Normal Distribution to Estimate the Quality of Regional Geochemical Survey Samples

ZHAO Zhi-fei, YAN Hui, YAO Lan, FANG Jin-dong, TANG Xing-min
(Hubei Geological Research Laboratory, Wuhan 430034, China)

Abstract: The quality of regional geochemical survey samples is currently estimated indirectly by internal and external laboratory quality control. The method described here effectively improves the quality of geochemical samples. However, the improvement of the method is limited when the contents of the actual samples and external quality control samples have large differences. Since most elements are in accord with a normal distribution in a regional geochemical survey sample, a comparison was made between the real and standard normal distribution of the elements, which passed the normal distribution test and gave an estimate of the quality by using a comparison of the real and standard graph's superposition using SPSS and Excel software. This method was directly applied to estimate the quality of analysis data for geochemical survey samples in Jiangxi province. The result of the normal distribution test indicated that the analysis data of the trace elements need a logarithmic conversion before the normal distribution can be studied. The quality evaluation method in this paper was not suitable for Ti, W and Cr etc in this area. The overlap ratios of 18 elements such as Co, Hg and Ga were more than 0.9. The overlap ratios of CaO and Na₂O were less than 0.9. The results of the actual samples indicated that the detection limits need to be reduced in order to improve the analysis qualities of CaO and Na₂O. It was a valuable attempt to directly evaluate the quality of analysis data for regional geochemical survey samples by using the method reported in this paper, which was a reference and supplement for the current methods of quality evaluation.

Key words: regional geochemical survey sample; quality evaluation; normal distribution; logarithmic normal distribution