

用聚类分析法研究土类和母质 对背景值的影响

潘佑民 吴定安 黄璋*

戎捷 杨国治

(湖南环保所)

(中国科学院南京土壤研究所)

讨论影响土壤中化学元素地球化学背景值因素的文章虽时有报道^[1-3],但它们大多是从定性的角度进行讨论的。本文拟将数理统计中的多元分析技术——聚类方法——用于研究湖南省境内土类和母质对土壤中Cu、Ni、Co、As四个元素含量的影响程度。以不同土类不同母质的土壤样品为观察单元,以元素含量为观察指标的聚类结果表明,土壤中元素背景值主要受成土母质母岩的影响,而受土壤类型的影响则是次要的,不明显的。

一、影响土壤背景值的主要因素

地质(母岩)、气候、地形、生物和时间等因素不仅影响土壤的形成过程,而且还影响和制约着土壤中化学元素的自然含量,但各因素的影响程度和结果并不是等同的。母质母岩是土壤形成的物质基础,是土壤中化学元素的源泉;气候因素决定着土壤的发育方向和过程。这两个因素,在生态学中被认为是状态因子,其作用是第一位的。地形、生物和时间的影响,被认为是转换因子,它们的影响受控并响应于气候条件和地质环境,因而其作用是第二位的。因此,本文运用聚类分析方法研究土壤中化学元素自然含量的影响因素时,只局限于土类和母质母岩这两个因素。

二、样品的来源与预处理

供研究的48个样品系由“湘江谷地土壤背景值协作组”提供的,它们采自湖南省内的不同土类、不同母质和不同的地形部位。从土类上来说,包括红壤、黄壤、山地黄棕壤、山地草甸土和水稻土。从成土母质母岩来说,有花岗岩、石灰岩、玄武岩等岩石风化物及第四纪红色粘土。根据“六·五”国家重点科技攻关对土壤背景值研究的要求,采样点尽可能选在未受人类活动影响的地方。从海拔高度来说,从1600米(山地草甸土)到100米(第四纪红色粘土)内的不同高度都设置了采样点。因此,样品具有相当的代表性。土样均采自土壤表层,在室内铺开自然风干后,挑去植物残体和砾石,用玛瑙研钵磨细过筛备用。Cu、Ni和Co元素的分析,采用王水-高氯酸-氢氟酸消化,用原子吸收光谱测定;As元素的测定是用浓盐酸浸泡土样,经处理后用原子荧光光度计测定。

三、聚类分析方法的具体应用

简单地说,聚类分析方法是依“大异而分群”,据“小异而聚类”的原理而建立的。在本研

* 黄璋同志现在广东省环监站工作。

表 1

样品序号、元素含量 (ppm) 和“距离”

序号	样品编号	土类	母质母岩	Cu	Ni	Co	As	距离值 D(a ₁ , a _j)
a ₁	CRG ₁ -47	红壤	花岗岩	14	11	4.0	5.28	0 (I)*
a ₂	CRG ₂ -48			25	22	7.0	14.73	18 (I)
a ₃	CRG ₃ -49			15	19	6.0	6.43	9 (I)
a ₄	CRG ₇ -415			20	16	8.0	12.95	11 (I)
a ₅	CRG ₈ -416			20	10	7.0	8.67	7 (I)
a ₆	CRG ₉ -417			40	24	10.0	7.53	29 (I)
a ₇	CRF ₂ -34	红壤	第四纪红色粘土	33	32	6.0	32.05	39 (I)
a ₈	CRF ₃ -35			32	26	11.0	19.31	29 (I)
a ₉	CRF ₄ -36			30	36	19.0	15.14	34 (I)
a ₁₀	CRF ₅ -37			24	32	11.0	16.43	27 (I)
a ₁₁	CRF ₆ -38			25	27	11.0	17.56	25 (I)
a ₁₂	CRF ₇ -39			33	38	15.0	18.74	38 (I)
a ₁₃	XRL ₂ -100	红壤	石灰岩	40	57	21	34.94	60 (II)
a ₁₄	XRL ₃ -101			25	70	17	35.00	68 (II)
a ₁₅	XRL ₄ -102			35	42	20	21.45	44 (II)
a ₁₆	XRL ₅ -103			37	49	29	55.19	71 (II)
a ₁₇	XRL ₆ -104			50	115	23	40.0	115 (IV)
a ₁₈	XRL ₁₀ -108			37	66	27	24.0	67 (II)
a ₁₉	XRE ₁ -127	红壤	玄武岩	60	131	40	5.28	134 (IV)
a ₂₀	XRE ₂ -128			67	111	66	3.39	129 (IV)
a ₂₁	XRE ₃ -129			50	107	35	11.09	106 (IV)
a ₂₂	XRE ₆ -132			72	270	70	96.68	288 (IV)
a ₂₃	XRE ₇ -432			39	55	30	12.82	120 (IV)
a ₂₄	XRE ₈ -430			59	123	12	12.41	121 (IV)
a ₂₅	CPG ₁ -20	水稻土	花岗岩	15	8.0	6.0	7.60	11 (I)
a ₂₆	CPG ₂ -21			20	12	7.0	8.40	6.3 (I)
a ₂₇	CPG ₃ -22			25	12	7.0	10.73	7.1 (I)
a ₂₈	CPG ₅ -24			17	21	10	2.57	12 (I)
a ₂₉	CPG ₆ -25			17	16	13	6.47	9.0 (I)
a ₃₀	CPG ₇ -26			25	21	13	5.47	12 (I)
a ₃₁	ZYG ₁ -243	黄壤	花岗岩	22	25	9.0	19.97	12 (I)
a ₃₂	ZYG ₂ -134			20	26	11	5.57	13 (I)
a ₃₃	XYG ₃ -135			32	14	13	22.6	17 (I)
a ₃₄	XYG ₄ -136			20	23	4.0	20.39	11 (I)
a ₃₅	XYG ₅ -137			3.0	11	1.0	11.09	19 (I)
a ₃₆	XYG ₆ -138			33	31	16	14.18	19 (I)
a ₃₇	XMG ₁ -160	山地黄棕壤	花岗岩	23	26	12	11.58	12 (I)
a ₃₈	XMG ₂ -161			20	24	27	14.27	20 (I)
a ₃₉	XMG ₄ -163			8.3	10	5.0	13.89	14 (I)
a ₄₀	XMG ₅ -165			17	17	7.0	7.91	6 (I)
a ₄₁	XMG ₇ -166			30	37	12	12.40	22 (I)

续上表

序号	样品编号	土类	母质母岩	Cu	Ni	Co	As	距离值 D(a _i , a _j)
a ₄₂	ZGG ₁ -241			16	29	12	16.93	15(I)
a ₄₃	ZGG ₂ -242	山	花	17	27	5.0	25.73	19(I)
a ₄₄	XGG ₁ -167	地		30	24	12	16.63	15(I)
a ₄₅	XGG ₂ -168	草	岩	29	31	9.0	18.16	18(I)
a ₄₆	XGG ₄ -170	甸		27	24	16	14.86	14(I)
a ₄₇	XGG ₅ -171	上	岗	25	28	10	11.87	15(I)
a ₄₈	XGG ₆ -172			20	23	9.0	14.32	11(I)

* 最后一项为分类结果

究中的具体作法是：把样品作为观察单元，而样品中化学元素的含量作为观察指标。如果以 a_i 表示第 i 个观察单元，以 C_{ik} 表示第 i 个观察单元 k 个观察指标，则两两观察单元之间的距离为：

$$D(a_i, a_j) = \sqrt{\sum_{k=1}^m (C_{ik} - C_{jk})^2}$$

式中：i, j = 1, 2, ……………n (本文 n = 48) k = 1, 2, ……………m (本文 m = 4)

当 i = j 时，为同一个样品，距离 D(a_i, a_j) = 0；当 i ≠ j 时，为不同样品，距离 D(a_i, a_j) ≠ 0。D(a_i, a_j) 值的大小，直观地表述了两个样品之间差别的大小。选用不同的 D(a_i, a_j) 值为区间值，即可以将被观察单元分成若干类。

为简便起见，本文采用“原点距离聚类方法”，即以某一观察单元(样品)作为原始点，与其它样品进行比较。本文选用 a₁ 号(编号 CRG₁-47)样品作为距离的原点，该样品具有各元素含量(观察指标)的加和值 ($\sum_{k=1}^4 C_{1k}$) 最小的特点，具有放大差别的作用。而观察指标的选择，则

应注意以下三点：

1. 各观察指标的值基本上应在同一数量级，以避免个别数量级大的指标对距离 D(a_i, a_j) 值起决定作用；
2. 各观察指标的变化方向应基本上一致。逆向变化的指标应尽可能还选。因为逆向变化的指标使距离 D(a_i, a_j) 值变小，不利于样品的分类。
3. 各观察单元之间的观察指标(元素含量)值应有明显可见的差别，即有较大的波动，才可以使距离 D 值较大，便于分类。

根据以上考虑，我们选定 Cu、Ni、Co 和 As 四个元素在样品中的含量为观察指标，对 48 个样品进行聚类。

样品序号及其元素含量(ppm)和“距离”以及各样品的土类和母质母岩的称谓列于表 1。

四、结果与讨论

将其它 47 个样品与“原点”样品 a₁(红壤、花岗岩)的距离 D(a_i, a_j) 值进行聚类，结果列于表 2。

从表 1 和表 2 我们可以看出：

1. 落在第(I)类的 28 个样品的土壤，分别属于水稻土、红壤、黄壤、山地黄棕壤和山地草甸土等 5 个土类，但其母岩全部是花岗岩的；第(II)类的 8 个样品，只有两个不属于第四

表 2

47 个样品与“原点”样品 a₁ 聚类结果

分 类	类 间 距 离	各 类 所 含 样 品 序 号	土 类	母 质 母 岩
(I)	$D \leq 20$	a ₁ , a ₂ , a ₃ , a ₄ , a ₅ , a ₂₅ , a ₂₆ , a ₂₇ , a ₂₈ , a ₂₉ , a ₃₀ , a ₃₁ , a ₃₂ , a ₃₃ , a ₃₄ , a ₃₅ , a ₃₆ , a ₃₇ , a ₃₈ , a ₃₉ , a ₄₀ , a ₄₁ , a ₄₂ , a ₄₃ , a ₄₄ , a ₄₅ , a ₄₆ , a ₄₇ , a ₄₈	红壤、水稻土、 黄壤、山地黄 棕壤、山地草 甸土	花岗岩
(II)	$D > 20$ ≤ 40	a ₆ , a ₇ , a ₈ , a ₉ , a ₁₀ , a ₁₁ , a ₁₂ , a ₄₁	红壤、山地黄 棕壤	第四纪红色粘 土、花岗岩
(III)	$D > 40$, $D \leq 80$	a ₁₃ , a ₁₄ , a ₁₅ , a ₁₆ , a ₁₈	红壤	石灰岩
(IV)	$D > 80$	a ₁₇ , a ₁₉ , a ₂₀ , a ₂₁ , a ₂₂ , a ₂₃ , a ₂₄	红壤	玄武岩、石灰岩

纪红色粘土发育的样品，占25%；第(III)类的5个样品，其母岩均为石灰岩；第(IV)类的7个样品，有6个是玄武岩为母岩的，占85.7%。

2. 以花岗岩为母岩的30个样品，尽管分属五个不同的土类，但有28个聚于(I)类，占93.33%；6个第四纪红色粘土为母质的样品，全部落在第(II)类，为100%；6个石灰岩为母岩的样品有5个在第(III)类，占83.33%；玄武岩为母岩的6个样品，100%地聚类于第(IV)类。这说明按母质母岩不同对样品进行分类，符合率不低于80%。

3. 24个红壤样品，分散于(I)、(II)、(III)、(IV)四个类，大致上是平均分配，每类5—7个样品。聚于(I)的样品，有水稻土、红壤、黄壤、山地黄棕壤、山地草甸土等，但全为花岗岩母岩。

可见，以距离聚类方法对48个不同土类、不同母质母岩和不同海拔高度的样品进行聚类，其分类结果与样品所属的母质母岩相符合，符合率不低于80%。样品的距离分类与按土类的分类完全不符合。这表明，成土母质母岩对土壤中金属元素含量的影响，远远大于土类的影响。因此，调查研究土壤背景值时，对于样点的分布和统计单元的划分，首先应考虑土壤母质母岩的影响。

参 考 文 献

- [1] 杨国治，天津地区土壤环境中若干元素的群分析。环境科学学报，第3卷，3期，1983。
- [2] 杨学义，南京地区土壤背景值与母质的关系。环境中若干元素的自然背景值及研究方法，科学出版社，1982。
- [3] 中国科学土壤背景值协作组，广东省区域土壤中某些元素的自然背景值。环境中若干元素的自然背景值及研究方法，科学出版社，1980。
- [4] 黄正南等，医用多因素分析，湖南科学技术出版社，1980。