

横断山区森林土壤的数值分类*

肖笃宁 盛世俊

(中国科学院林业土壤研究所)

摘 要

在土壤分类中综合运用多种数值分类方法可以取得较好的分类效果。本文应用了主组元分析、聚类分析,多组判别分析和最小支撑树等方法,将横断山区有代表性的30个森林土壤剖面分为六组,各组土壤间不仅在理化性质上有鲜明特征,而且与最重要的成土条件和森林植被类型基本吻合。在西南高山林区,树种与下木对土壤形成有重要作用,主要成土过程为酸性淋溶与腐殖化过程。通过选取的六个理化性质指标,主组元分析结果反映出了上述主导成土过程,而多组判别分析则进一步给予每组土壤以明确的分类域。这两种方法的结合可作为土壤数值分类的基本方法。

近年来我国土壤学者应用主组元分析、聚类分析和判别分析,获得可喜成果^[1-4]。实践证明,以上几种数学方法对于土壤分类既是有效的,又是不完全的,各有其局限性。我们认为有效的途径是综合运用现有的多元分析方法,用多种方法的计算结果互相补充,从而取得较满意的分类效果。本着这种认识,我们试图以一个大地区的一种土壤系列为例,来进行土壤数值分类的研究,重点在于方法应用效果的比较分析,而非该区若干有争议的土壤发生问题的讨论。

横断山区位于川西、滇北,土壤分布复杂。主要森林类型有暗针叶林,云南松林,常绿阔叶林和高原块状针叶林等,相应发育着不同的森林土壤,大致上云南松林区以红壤为主,常绿阔叶林区以黄壤为主,暗针叶林区则为各种属于棕壤系列的土壤^[1]。

本文利用的土壤资料主要来自于我们六十年代在该区工作时的成果,部分引用了其他作者的成果^[5,6]。

一、数据的准备与处理

本文选用的30个森林土壤剖面的地理情况见表1。因为所研究的分类客体是棕壤系列的土壤,酸性淋溶、灰化、粘化和腐殖质化是主要的成土过程,因此我们选取了六个理化性质作为指标(表2): x_1 为 A₁ 层(亚表层) pH 值(H₂O 提); x_2 为 A₁ 层盐饱和度(%); x_3 为粘粒比值 B 层/A 层; x_4 为硅铁铝率比值 A₁ 层/B 层; x_5 为 A₁ 层腐殖质(%); x_6 为 A₁ 层 C/N 比值。

* 本项研究的计算工作是在美国明尼苏达州立大学土壤系利用大型电子计算机完成。R. H. Rust 教授和 Jim Crum 博士给予热情支持和协助,特此致谢。

1) 中国科学院西部地区南水北调综合考察队: 川西、滇北地区的森林土壤。

表 1 横断山区森林土壤剖面的地理情况

Table 1 Geographic conditions of forest soil profiles in the region of Hengduan Mountains

编号 Sample No.	地点 Location	海拔 (m) Elevation	森林类型 Vegetation type	土壤野外定名 Soil name in field	母质 Parent material	剖面野外号 No. of profile
1	雅江	4120	杜鹃-冷杉	灰棕壤		61-南-425
2	丽江	3710	大杜鹃-冷杉	腐殖质淀积灰化土		173
3	南坪	3400	杜鹃-冷杉	泥炭潜育化灰棕壤	花岗岩	60-南-355
4	云岭西坡	3490	竹类-冷杉	泥炭潜育化灰棕壤		172
5	米易	3200	箭竹-冷杉	弱潜育化灰棕壤		60-南-11
6	维西	3110	箭竹-冷杉	泥炭潜育化灰棕壤	片岩	60-南-311
7	盐源	3360	箭竹-云杉	灰棕壤	变质岩	60-南-208
8	稻城	4120	杜鹃-云杉	灰棕壤	花岗岩	61-南-445
9	中甸	3510	箭竹-云杉	棕壤	板岩	60-南-335
10	沙鲁里山南部	3400	大竹子-云杉	棕壤		E ₃
11	云岭西坡	3120	灌木云杉	棕壤		E ₁
12	沙鲁里山南部	3240	沟旁云杉	棕壤		E ₃
13	松潘	2800	巴氏云杉	残余碳酸盐棕壤	石灰岩	61-南-94
14	沙鲁里山南部	3300	高山松	棕壤	砂岩	C ₁
15	道孚	3820	高山栎	棕壤		61-南-298
16	沙鲁里山南部	3640	高山栎	灰棕壤	页岩	D ₁
17	贡山	2660	常绿青杠	腐殖质棕壤	片麻岩	60-南-315
18	宁蒗	3100	楠竹-云南松	棕壤	砂页岩	60-南-269
19	昭觉	3220	高山栎	棕壤	钙质页岩	60-南-72
20	冕宁	2820	杜鹃-云南松	红棕壤	片岩	60-南-230
21	壤塘	3450	草类-冷杉	生草棕壤		61-南-240
22	甘孜	3840	草类-巴氏云杉	生草棕壤	云母片岩	61-南-264
23	红原	3580	草类-冷杉	生草棕壤		61-南-60
24	马尔康	3700	藓类-冷杉	灰棕壤		57-标-17
25	碧江	2550	常绿青杠	腐殖质棕壤	花岗岩	60-南-372
26	维西	3720	杜鹃-冷杉	腐殖质淀积灰化土	砂岩	174
27	马尔康	3750	杜鹃-冷杉	腐殖质淀积灰化土	片岩	No. 2001
28	黑水	3630	杜鹃-冷杉	棕色灰化土	片岩	61-南-56
29	马尔康	3780	杜鹃-冷杉	棕色灰化土	片岩	61-南-212
30	松潘	3460	冷杉林	棕壤	石灰岩	61-南-80

各土壤个体在选取变量(指标)上的数据见表 2。

我们重视亚表层所提供的土壤信息,是因为它最能反映棕壤化过程和灰化过程的本质差别。同时我们也选取了表土层腐殖质积累和分解的指标,以反映某些林型下土壤泥炭化和表层潜育化过程的特征。

为了消除原始数据量纲不一在计算中可能产生的影响,需将全部原始数据进行标准化处理。本文中根据不同统计方法的要求分别采用了数据中心化,离差标准化和标准差标准化等几种方法。

二、主组元分析

主组元分析(PCA)是所有近代排序方法中应用最广泛的,其突出优点是可以在关系错综复杂的多变量中找出影响它们的共同因素和特殊因素,从而用若干个数目较少的

表 2 横断山区森林土壤理化性质

Table 2 Physical and chemical properties of forest soils in the region of Hengduan Mountains

样品号 Sample No.	pH (E 层)	盐基饱和度 (E 层) B. S.	粘粒比 B/E Clay	$\frac{\text{SiO}_2}{\text{R}_2\text{O}_3}$ E/B	有机质 (A 层) O. M. (%)	C/N (A 层)
1	3.5	25	0.42	1.15	53.4	22.3
2	3.4	3	0.44	4.42	48.5	15.6
3	3.7	20	0.86	1.55	17.4	44.0
4	4.1	8	0.81	1.93	17.0	14.1
5	4.2	4	0.93	0.69	17.8	13.1
6	4.4	17	1.02	1.27	39.1	12.0
7	5.1	10	1.29	0.99	7.2	9.4
8	5.0	17	1.75	0.92	10.7	11.9
9	5.6	9	0.95	0.66	5.8	9.1
10	4.8	21	1.23	1.17	14.4	16.7
11	4.6	7	2.25	1.07	24.5	16.0
12	5.1	40	1.44	1.13	19.7	14.1
13	6.5	99	2.15	1.00	20.3	13.5
14	5.8	50	3.06	1.57	9.2	13.8
15	4.7	20	2.54	0.66	19.2	12.7
16	4.6	7	2.43	1.17	5.9	15.1
17	5.1	47	1.80	0.88	13.9	6.7
18	5.8	42	1.42	1.25	6.2	10.2
19	6.0	81	2.07	1.27	13.2	11.2
20	5.5	83	1.20	0.90	1.9	3.3
21	5.3	72	1.05	0.85	29.1	14.2
22	6.2	89	3.16	0.90	20.5	11.9
23	6.3	91	3.84	0.82	20.1	6.7
24	4.4	33	2.15	1.13	17.6	13.2
25	5.3	23	1.22	0.90	11.7	10.6
26	4.0	7	1.22	4.05	20.0	12.5
27	4.6	4	0.80	1.24	7.3	12.0
28	3.9	30	1.53	1.64	6.0	11.1
29	4.0	31	1.64	1.20	29.7	52.0
30	5.2	78	1.20	1.27	35.2	11.2
平均值	4.9	35.5	1.60	1.32	18.79	14.70
标准差	0.81	30.54	0.81	0.84	12.46	9.70

独立新变量来表达所观测的原始数据,取得良好的降维效果^[1,7,8]。

我们采用的是按费歇准则编制的电子计算机程序^[14],所得计算结果见表 2, 3。从表 3 可见,第一主组元占总信息的 79.3%,第二主组元占 14.5%,二者合计约占 94%。可见用第一和第二主组元作为横坐标和纵坐标所画出的二维散点图(图 1),足以概括原始资料所提供的绝大部分信息,从图中可将 30 个土壤剖面分为六组:第 I 组包括剖面 1, 2, 3, 6, 29;第 II 组包括剖面 4, 5, 7, 8, 9, 11, 16, 26, 27;第 III 组包括剖面 10, 15, 24, 25, 12, 28;第 IV 组包括剖面 14, 17, 18;第 V 组包括剖面 13, 19, 21, 22, 23, 30;第 VI 组包括剖面 20。

表 3 第 1, 2, 3 个主组元的方差贡献率及属性对主组元的权系数

Table 3 The contribution and weight coefficients of the first, second, and third principal components

主组元 P.C.	特征值 λ_T Eigenvalue	贡献率 Contribution %	累计贡献率 Sum contribution %	权系数 Weight coefficient						
				x_1	x_2	x_3	x_4	x_5	x_6	绝对值 累计
1	27290	79.3	79.3	0.020	0.997	0.013	-0.008	-0.009	-0.073	1.116
2	4996	14.5	93.8	-0.026	0.039	-0.015	0.019	0.910	0.411	1.420
3	2134	6.1	99.9	0.005	-0.062	0.009	-0.018	-0.412	0.909	1.415

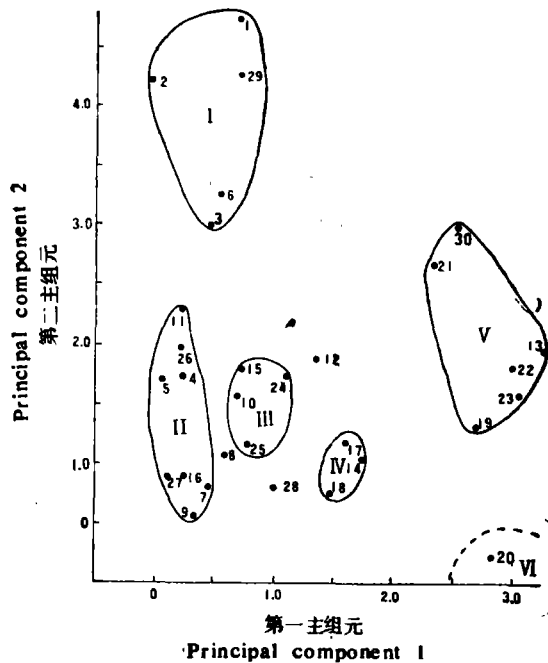


图 1 主组元分析的二维散点图

Fig. 1 Classification of soil samples by principal component analysis

从表 3 可知,亚表层的盐基饱和度(x_2)在第一主组元的构成中占最主要地位,权系数的绝对值最高,所占份额为 89%;而表土层腐殖质含量(x_5)在第二主组元的构成中所占份额为 64%;至于第三主组元则以表土层 C/N 比值(x_6)所占最高。由此可以认为,第一主组元值的增大,反映了本区森林土壤的主导成土过程——酸性淋溶过程的减弱。从第 II 组到第 V 组土壤,即由盐基强烈不饱和的强酸性土壤,很有规律地递变为盐基饱和的中性土壤。第二主组元的构成中,表土层的腐殖质含量和 C/N 比合占份额达 93%,可见它所反映的成土过程正是土壤腐殖质积累过程。

这样,我们所划分出的六个土壤组合,也就有了相应的分类学意义。第 I 组土壤特征是强酸性和还原条件下的强有机质聚积,剖面发育常见有表层泥炭化和表面潜育化特征,泥炭粗腐殖质层厚度为 8—15 厘米,因此我们把这组土壤命名为泥炭潜育化的冷棕壤。第

II 组土壤有着明显的灰化过程,从分析结果看,代换性盐基、铁、铝均有移动,部分剖面并有铁质和腐殖质淀积,暂称为棕色灰化土。第 III 组土壤基本特征类似 II 组,但酸性淋溶作用较弱,称为灰棕壤。第 IV 组盐基饱和度达 50%,腐殖质聚积强度中等,已属于棕壤范畴。第 V 组土壤代换性盐基近于饱和,表土层腐殖质强烈聚集并且矿化程度较高,显示出森林土壤的生草化特征,故暂名为生草棕壤。第 VI 组土壤本次研究中虽只有一个剖面,但其生物地理环境和土壤性质均有很大不同,故仍以野外命名红棕壤名之。

三、最小支撑树法 (MST)

最小支撑树法是图论中的一种常用方法。求取最小支撑树的方法不止一种^[10]。图 2 是在主组元分析结果上画出的最小支撑树。如分类合理,则在最小支撑树中,同一类的点的边长较短,而类间的边长较长。故只要在最小支撑树中找到几个长边,并将其抹去,使最小支撑树分成互不联接的若干部分,这几部分的点就组成了所划分的类。为了找长边,则需要事先设定两个参数,即步数 K 及长边与其两端 K 步以内边长平均值之比 R 。以图 2 为例,若我们定 $K = 1, R_0 = 1.5$, 则长边可按如下方法进行。例如,边 3—11 的边长为 0.83, 邻边平均长为 $(0.26 + 0.37) / 2$, 则 $R = 2.63 > R_0$, 边 3—11 即为所求长边之一。

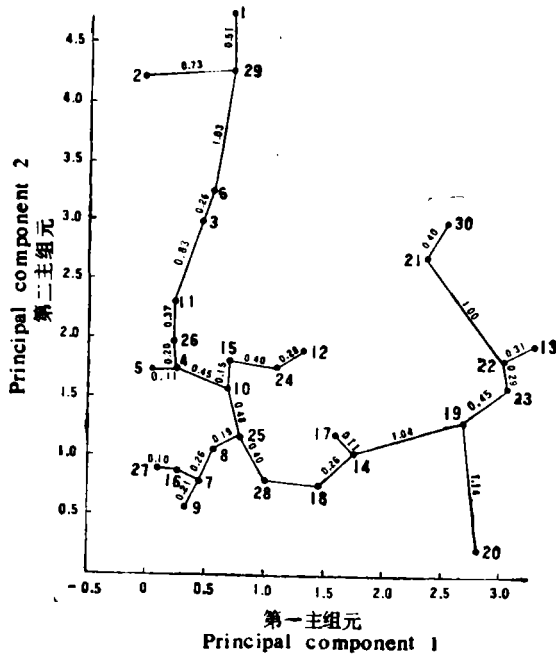


图 2 最小支撑树结构图

Fig. 2 Groupings of 30 forest soils by first principal components in a minimum spanning tree (MST)

若是抹去上述几条长边,被分割的最小支撑树所形成的分类组合类似前述。但第 I 组可分为剖面 1, 2, 29 和 3, 6 两个亚组;第 II 组可分为 11, 26, 4, 5 和 27, 16, 7, 9 两个

亚组。

四、聚类分析

聚类分析是依据样品之间的相似性而把它们逐步归组成群的一种方法，计算结果通常用聚类枝状图表示^[2,8,11]。反映样品间的距离通常有两种方法：

1. 普通聚类分析：把土壤样品看成是多维空间中的点，用点间距离来表示土壤样品间的亲疏关系，距离越小反映关系愈密切而归为一类，反之则划为另一类。通常用作这种计算的统计量有两种：即绝对值距离和欧氏距离。

采用离差标准化法，我们比较了上述两种距离的聚类枝状图（图 3），其中绝对值距离的分类效果优于欧氏距离，这和其他研究者的结论是一致的^[2]。

分类阈值的选取是一个比较困难的问题。如从图 3 绝对值距离 83.5 处画一垂直于横轴的直线，可将 30 个剖面分为七组，其中第 1 和第 2 组合并后即相当于主组元分析所得的 I 组，第 3, 4, 5 组分别相当于 II, III, IV 组，第 6 和第 7 组合并后相当于第 V 组。但仔细审视可发现剖面 1, 8 被列入第 III 组，剖面 20 被列入第 V 组。而在 MST 图和 PCA 散点图上剖面 1 均归为 I 组，剖面 20 以单独归组为好。关于剖面 8 的归组问题以后另行讨论。

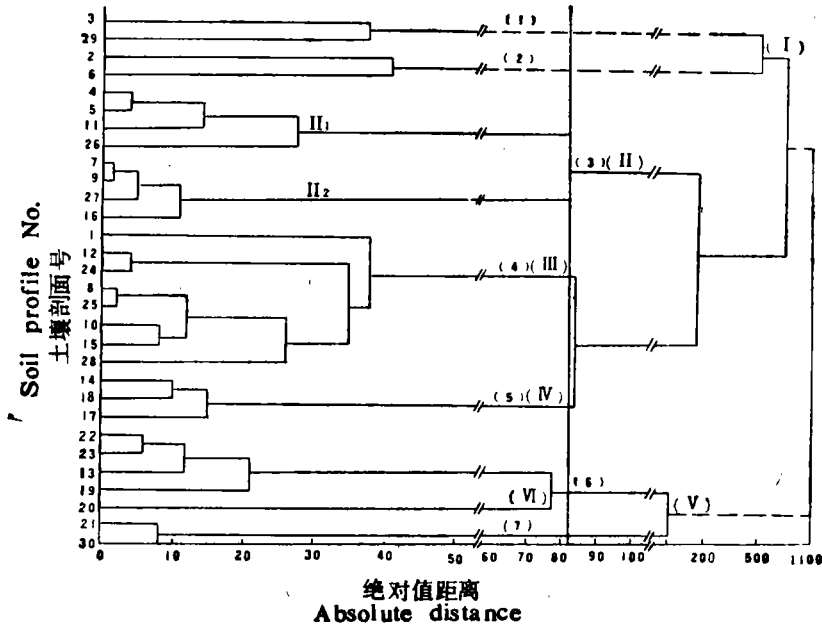


图 3 绝对值距离指标的聚类枝状图
Fig. 3 Cluster distribution of 30 forest soils

2. 相似性指标法：按照 Hole 等的方法^[12]，土壤样品间的相似系数 $I = 100 \times 2W / (A + B)$ ，其计算步骤是，首先转换原始数据，将同一变量值中最高者赋值为 100，最低

者赋值为 0, 然后按比例换算出其余数值。每一个变量的数据照此变换后, 再将样品的各变量值累加而得 A 或 B 。表 4 是从表 2 按上法变换而来。 W 为任意两个样品中每一变量值分别比较, 选取最小值累加之和。若要比样品 1 和样品 2 的相似性, 从表 4 中读取 $A = 178$, $B = 216$, $W = 0 + 0 + 0 + 13 + 90 + 25 = 128$ 。

$$I = 100 \times 2 \times 128 \div (178 + 216) = 65$$

本次计算所选 30 个土壤剖面两两之间的相似性统计量共有 435 个数值。其中剖面 22 和 23 的相似性指标最高, 达 94; 而剖面 2 和 20 的相似性指标最低, 只有 4。按照前

表 4 土壤剖面相似性指标计算表
Table 4 Calculation of similarity index

样本	x_1	x_2	x_3	x_4	x_5	x_6	总计 Sum
1	3	23	0	13	100	39	178
2	0	0	1	100	90	25	216
3	42	18	13	24	30	84	211
4	23	5	11	34	29	22	124
5	26	1	15	1	31	20	94
6	32	15	18	16	72	17	170
7	55	7	25	9	12	12	114
8	52	15	39	7	17	18	148
9	71	6	15	0	8	12	112
10	45	19	24	14	24	27	153
11	39	4	53	11	44	26	177
12	55	39	30	15	35	22	196
13	100	100	51	9	36	21	317
14	77	49	77	24	14	21	262
15	42	18	62	0	34	19	175
16	39	4	59	14	8	24	148
17	55	46	40	6	23	7	177
18	77	41	29	16	8	14	205
19	84	81	48	16	22	16	267
20	68	83	23	6	0	0	180
21	61	72	18	5	53	22	231
22	90	90	80	6	36	18	320
23	94	92	100	4	35	7	332
24	32	31	50	15	30	22	180
25	61	20	23	6	19	15	144
26	19	4	23	90	35	19	190
27	39	1	11	15	10	18	94
28	16	28	32	26	8	16	126
29	19	29	36	14	54	100	252
30	58	78	23	16	65	16	256

述分组方案, 组内样品的相似性最高, 也就是说分类基本上是合理的。但也有例外, 如剖面 8 和 II 组各剖面的平均相似系数为 69.9, 而和 III 组的平均相似系数为 80.3, 显应划入 III 组。另外两个分类位置不好确定的剖面, No. 12 和 III 组的平均相似系数为 82.3, 和

IV 组为 81.3; No. 28 和 III 组的平均相似系数为 67.4, 和 IV 组为 66.3。这说明它们处于 III 组和 IV 组间的过渡地位, 而仍以划入 IV 组为宜。

此外, 我们还可以通过相似性指标来比较组内样品聚积的紧密程度。各组剖面间的平均相似系数如下:

I	II	III	IV	V
58.6	66.3	77.5	77.3	82.0

显见第 I 和 II 组内样品的聚类较松散, 即组内各剖面间的性质差异较大, 这与图 2 和图 3 所反映出的特征是一致的。按照 Sokal 等的方法^[5], 以表 4 的数据为基础, 可进一步得到又一聚类枝状图(图 4), 其形状和主组元分组结果吻合。

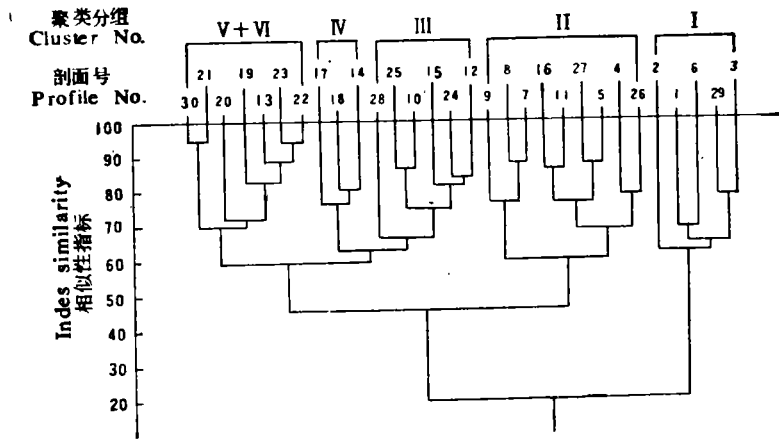


图 4 按照相似性指标的聚类枝状图

Fig. 4 Dendrogram of 30 forest soils according to similarity indexes

五、判别分析

判别分析通过对样品的多变量数据(分类信息)求出其判别函数, 来计算被分类客体的判别值以确定其归属。用之于土壤分类, 它可检验和调整原有分类方案, 也可决定新土

表 5 正规化判别函数的构成

Table 5 Eigenvalues, percent variance, cumulative percent variance and canonical correlation from discriminant analysis

判别函数 Function	特征值 Eigenvalue	所包括的变量(%) Percent variance	累计(%) Cumulative percent	正规化相关系数 Canonical correlation
1	25.2128	82.75	82.75	0.98
2	4.7258	15.51	98.26	0.91
3	0.4219	1.38	99.64	0.55
4	0.0905	0.30	99.94	0.29
5	0.0184	0.06	100.00	0.13

壤样品的归属。在主元分析或聚类分析的分类基础上,判别分析可给出准确的边界值,这是它的突出优点^[6,9,16,17]。本节使用的是多组判别方法,它不同于只有一个判别函数的两组判别,而具有 $P-1$ 个(P 为变量个数)判别函数。从表 5 可知,判别函数 F_1 的权系数为 82.7%, F_2 为 15.5%,二者合计已超过 98%,故我们用这两个函数作坐标轴而绘制的散点图,足以反映原有分类信息。在表 6 中只列出了 F_1 和 F_2 的函数值,其余从 F_3 到 F_6 的函数值所占比重很小,故从略。从表 7 可知,函数 F_1 在 1% 置信度以下,多元平均值之间在统计上有显著差异,判别有效。

表 6 正规化判别函数的系数

Table 6 Standardized canonical discriminant function (F) coefficients for the 6 index variables selected in this study

函数 F	x_1	x_2	x_3	x_4	x_5	x_6
1	0.0734	1.1668	-0.3099	0.2302	0.0594	-0.3423
2	0.4917	-0.0271	-0.1364	0.2129	1.1738	0.9323

表 7 判别函数的检验

Table 7 Significance after function

判别函数检验 After function	λ	χ^2 Chi-squared	自由度 D. F.	置信度 Significance
0	0.0042	125.77	30	0.0000
1	0.1106	50.64	20	0.0002
2	0.6333	10.51	12	0.5715
3	0.9004	2.41	6	0.8781
4	0.9817	0.42	2	0.8104

一般说来 K 组间的判别界限的个数为 $C_k^2 = k(k-1)/2$ 。六组间的判别界线共有 15 条。但图 5 的六组中,有判别意义的组间界线实际只有 9 条。下述公式,就是这 9 条基本判别界线的线性函数式^[13]

$$W_{ij} = F_1(\bar{f}_{1i} - \bar{f}_{1j}) + F_2(\bar{f}_{2i} - \bar{f}_{2j}) - \frac{1}{2}(\bar{f}_{1i} - \bar{f}_{1j}) \cdot (\bar{f}_{1i} + \bar{f}_{1j}) - \frac{1}{2}(\bar{f}_{2i} - \bar{f}_{2j})(\bar{f}_{2i} + \bar{f}_{2j})$$

\bar{f} 代表各组形心(变量均值点)的判别函数值(表 8), i 和 j 代表不同的分类组合。本次研

表 8 各组形心的正规化判别函数值

Table 8 Canonical discriminant functions (F) evaluated at group means (group centroids)

函数 F	I	II	III	IV	V	VI
\bar{F}_1	-2.694	-4.078	-1.290	1.641	7.532	7.792
\bar{F}_2	3.863	-1.169	-0.989	-1.351	0.767	-3.407

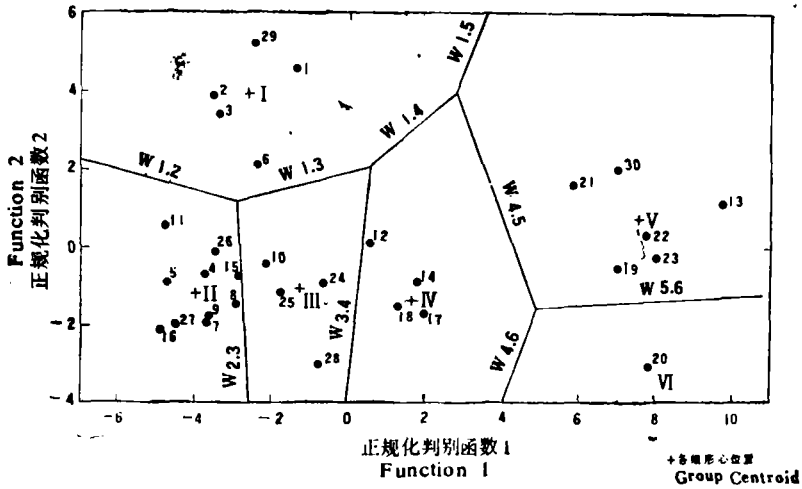


图 5 多组判别分析得出的分类区域图

Fig. 5 Classification regions for 6 groups of 30 forest soils based on discrimination

究得出的计算结果如下:

$$\begin{aligned}
 W_{1,1} &= 1.38U_1 + 5.03U_2 - 2.09 \\
 W_{1,2} &= -1.40U_1 + 4.85U_2 - 9.76 \\
 W_{1,3} &= -4.33U_1 + 5.21U_2 - 8.83 \\
 W_{1,4} &= -10.22U_1 + 3.10U_2 + 17.55 \\
 W_{2,3} &= -2.78U_1 - 0.18U_2 - 7.67 \\
 W_{3,4} &= -2.93U_1 + 0.36U_2 + 0.93 \\
 W_{4,5} &= -5.89U_1 - 2.11U_2 + 26.40 \\
 W_{4,6} &= -6.15U_1 + 2.05U_2 + 33.90 \\
 W_{5,6} &= -0.26U_1 + 4.17U_2 + 7.49
 \end{aligned}$$

从这些函数中我们可以建立起六个分类区域的分类规则:

- | | |
|--|-------|
| $W_{1,1} > 0, W_{1,2} > 0, W_{1,3} > 0, W_{1,4} > 0$ | I 组 |
| $W_{1,2} < 0, W_{2,3} < 0$ | II 组 |
| $W_{1,3} < 0, W_{2,3} < 0, W_{3,4} > 0$ | III 组 |
| $W_{1,4} < 0, W_{3,4} < 0, W_{4,5} > 0, W_{4,6} > 0$ | IV 组 |
| $W_{1,1} < 0, W_{4,5} < 0, W_{5,6} > 0$ | V 组 |
| $W_{5,6} < 0, W_{4,6} < 0$ | VI 组 |

计算机根据以上规则可自动判定点的归属, 给出分类结果。计算机给予原有分类方案正确性的评价是 93%。按判别分析结果, 剖面 12 应由 III 组划入 IV 组 (可能性为 77%); 剖面 15 划入 II 组的可能性为 51%, 划入 III 组的可能性为 49%。考虑到其森林类型是高山栎林, 故仍以划在 III 组为好。

六、讨 论

1. 在土壤分类中综合运用多种数值分类方法,可以互相检验和补充,取得较好的分类效果。本文将横断山区有代表性的 30 个森林土壤剖面分为六组,各组土样间不仅在理化性质上有鲜明特征,而且与最重要的成土条件和森林类型基本吻合,显示了土壤发育与森林类型间的密切联系。第 I 组冷棕壤全部发育于藓类-杜鹃-冷杉林下,下木杜鹃是这种冷湿生态环境的重要指标。第 II 组棕色灰化土主要发育于箭竹-云、冷杉林下,下木箭竹的灰分组成可能对土壤灰化过程有所影响。第 III 组灰棕壤主要发育于灌木-云、冷杉林下,下木较杂,反映出土壤条件比较多样。第 IV 组棕壤分布于海拔较低的山地垂直带上,多发育于高山松和云南松之下,滇西并有部分常绿阔叶林的分布。第 V 组生草棕壤全部分布于本区北部的高原块状林区,发育在草类-巴氏云杉(或鳞皮冷杉)之下,气候较干燥,多系碳酸盐母质。第 VI 组红棕壤则属于云南松林分布区。

2. 在本文应用的多种方法中,主组元分析是基础,因为主组元的物理意义可以依靠专业知识来鉴别,这就便于我们确定分组时可相应考虑土壤发生学的规律,在数值分类方法与传统的土壤分类方法间架起一座桥梁。同聚类分析相比,这是它的优点。但与此同时,这样得出的初始分类也容易带上主观烙印,某些样品的分类地位带有不确定性,这就应该参照聚类分析和 MST 方法,特别是多组判别分析的结果,通过判别函数给予每组土样以明确的分类域,从而调正原有分类中的不合理处,并可准确判断未知土壤的归属。所以我们认为,主组元分析和多组判别分析相结合可作为土壤数值分类的基本方法。

3. 本文计算的 30 个森林土壤剖面中,有三个明显处于过渡位置,不好确定其归属。表 9 列出了这三个剖面在不同数值分类方法中被列入的位置。从中可以看出,在所使用的这些方法中,判别分析用来鉴别剖面归属最为有力。它可以把剖面 8, 12 和 28 清楚地分

表 9 不同分类方法时三个土壤剖面的分组位置

Table 9 Differing group location of 3 soil resulting from the numerical methods used in this study

土壤剖面 Soil	主组元分析 PCA		最小支撑树 MST	聚类分析 Cluster analysis		相似性分析 Similarity analysis	判别分析 Discriminant analysis
	中心化 Centraliza- tion	标准化 Standardi- zation		绝对值距离 Absolute dist	欧式距离 Euclidean dist		
No. 8	III	III 或 II	III	III	II	III	II
No. 28	III	III	III	III	II + III	III	III
No. 12	IV	IV	III	III	III + IV	III	IV

开,而其它方法却做不到,尽管这些土壤有着大抵相同的森林植被类型(暗针叶林),但它们分布于不同的地理区域,海拔高度不同,林型有差,土壤性质呈现差别,盐基由不饱和到饱和,因此把它们分开是合理的。

4. 在数值分类中土壤信息的选择和处理十分重要。主组元分析选择那些变异系数

(CV) 大的变量, 在我们应用六个指标所作的分析中, 最有用的是亚表层盐基饱和度, 其次是表土层的有机质含量。粘粒比值和硅铁铝率比值是反映土壤粘化和灰化过程的重要指标, 但在本区的棕壤系列土壤中, 差别不如其它指标明显, 因而所起作用不如在非数值土壤分类中重要。若对这两个指标进行适当的数学处理, 或选用其它相似指标如粘粒的淋失累积率和土壤风化系数作为指标, 可能分类效果会更好一些。至于土壤剖面的层次结构与形态特征, 以及成土条件指标的选取, 则需要另文讨论。

参 考 文 献

- [1] 刘多森, 1979: 主组元分析在分辨土壤类型及风化—成土过程上的应用。土壤学报, 第 16 卷 2 期, 172—183 页。
- [2] 刘多森等, 1980: 聚类分析在太湖地区水稻土物质分类上的应用。土壤学报, 第 17 卷 4 期, 374—381 页。
- [3] 卜兆宏、万洪富, 1982: 土壤分类分级中的综合数值分析法的初步研究。土壤学报, 第 19 卷 3 期, 283—295 页。
- [4] 熊国炎, 1980: 两组线性判别分析在土壤分类中的应用。土壤, 第 5 期。
- [5] 刘寿坡, 1960: 横断山脉的高山灰化土。土壤学报, 第 8 卷 2 期, 122—128 页。
- [6] 张万儒, 1962: 青藏高原东南部边缘地区的森林土壤。土壤学报, 第 10 卷 2 期, 107—144 页。
- [7] Arkley, R. J., 1968: Statistical methods in soil classification. Trans. 9th Intean. Congr. Soil Science, 4: 187—192.
- [8] Bidwell, O. W., 1963: Numerical taxonomy and soil classification. J. Soil Sci., 97: 58—62.
- [9] Cipra, J. E., 1970: Numerical taxonomy of soils from nine orders by cluster and centroid-component analysis. Soil Sci. Soc. Amer. Proc., 34: 281—287.
- [10] Gower, J. C., 1969: Minimum spanning trees and single linkage cluster analysis. Appl. Statistica, 18: 54—64.
- [11] Grigal, D. F., 1969: Numerical classification of some forested minnesota soil. Soil Sci. Soc. Amer. Proc., 32: 433—438.
- [12] Hole, F. D., 1960: An enperiment in ordination of some soil profiles. Soil Sci. Soc. Amer. Proc., 24: 309—312.
- [13] Morrison, D. F., 1975: Multivariate Statistic Methode. McGraw-Hill.
- [14] Nie, N. H., 1975: Statistical Package for the Social Science. McGraw-Hill.
- [15] Sokal, R. R. and Mickener, C. D., 1958: A statistical method for evaluating systemaric relationships. Univ. of Kansas Bull, 38: 1409—1438.
- [16] Webster, R., 1974: Multiple discriminant analysis in soil survey. J. Soil Sci., 25: 120—134
- [17] Webster, R., 1977: Quantative and Numerical Methods in Soil Classification and Survey. Clarendon Press.

NUMERICAL CLASSIFICATION OF FOREST SOILS IN THE REGION OF HONGDUAN MOUNTAINS

Xiao Duning, Sheng Shijun

(Institute of Forestry and Soil Science, Academia Sinica)

Summary

Principal component analysis, cluster analysis, discriminant analysis and a graph-theoretical method were used for the classification of forest soils in region of the Hongduan Mountains. 30 forest-soil profiles were divided into six groups based on their chemical and physical properties. The results obtained from the numerical methods well conformed to the most important genetic conditions of the soils, i.e. the forest vegetation types. In the high-mountain forest region of Southwest China, the major soil forming processes are eluviation (acid leaching) and humification. Results of principal component analysis of 6 physical and chemical indexes of the soils indicated that the soils grouping could reflect the major processes mentioned above, while discriminant analysis of the soils gave the distinct delimitation of each soil group. The two methods in combination may be the basic method for soil numerical classification.