ACTA PEDOLOGICA SINICA

# 土壤分类分级中的综合数值 分析法的初步研究\*

ト 兆 宏 万 洪 富 (中国科学院南京土壌研究所)

近十年来,国内外采用某些数学方法与土壤学相结合,取得了一定的进展<sup>[3,8,9]</sup>。 然而,在切实解决土壤分类问题方面,特别是土壤性质的分类界限指标方面,尚有许多工作要做。

数学中有许多方法,各有其适用情况。但就分类精度而言,以主成份分析法为优。这在土壤分类中是如此,在遥感信息分类中也是如此<sup>[6]</sup>。虽然它可以获得比较严密的土壤主成份方程,但是要完满地解决土壤分类问题,单靠它是不可能的。例如,对于绘制散点图上类间界线及其数学方法的问题;对于土壤类间理化性质数据的分类界限指标等问题,主成份分析就无法解决。因此,尚须寻找适于土壤分类的数学方法。

近来,实验研究了使两种数学方法综合为一体的方法,可称之为"综合数值分析法",已试用于碱化土壤分级中。在应用这种综合数值分析法对土壤波谱反射特性数据进行土壤分类的试验中,显示出它用于解决土壤分类问题的可能性。 现将其基本原理、应用程序、试验结果和精度讨论等问题,作一介绍,以求在广泛试用中更臻完善。

# 一、基本原理

综合数值分析法,主要包括有主成份分析和判别分析。 在寻求土壤分类分级界限指标中,它是以相关分析求出供主成份分析的相关系数方阵,进而求出土壤的主成份方程和散点图,以判别分析求出类间的综合系数,并在散点图上确切地绘制出类间的界线,从而寻找到土壤的各项性质分类界限指标。

相关分析是综合数值分析法的基础。相关系数的计算公式为:

$$r_{il} = \frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_i) (x_{il} - \bar{x}_l)}{\sqrt{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2 \sum_{i=1}^{n} (x_{il} - \bar{x}_l)^2}}$$
(1)

式中 $r_{il}$ 为i,l两项土壤理化性质(或称两变量)的相关系数;i为土壤剖面编号; $x_{ij}$ , $x_{il}$ 

<sup>\*</sup> 此项工作曾得到本所地理室、盐土室和地化室各位领导的支持与帮助;本文蒙席承藩教授、王遵亲副教授等审阅指正,特此致谢。

为某一土壤的两项性质的理化分析数据;  $\bar{x}_i$ 、 $\bar{x}_i$  为其所有参与统计土壤的单项性质的均值。当每个土壤剖面的理化数据都有五项或八项时,便可组成五阶或八阶方阵。

主成份分析是综合数值分析的主体。主成份方程的表达式为:

$$Y_i = \sum_{j=1}^{v_i} W_{ji} \cdot \frac{x_{ij} - \overline{x}_j}{S_i},$$

经变换为:

$$Y_{i} = \sum_{i=1}^{\nu} \frac{W_{ii}}{S_{i}} \cdot x_{ij} - \sum_{i=1}^{\nu} \frac{W_{ii}}{S_{i}} \cdot \vec{x}_{i}$$
 (2)

式中  $W_{i}$  为第 i 个特征值的第 i 个特征向量,或称之为第 i 个变量在第 i 个特征值主成份方程的权系数;  $x_{ij}$  和  $\bar{x}_i$  与 (1) 式中的同义;  $S_i$  为参与计算的各土壤第 i 项性质的标准差; v 为土壤理化性质项目数(或称变量数)。

为了解算(2)式中的权系数  $W_{ii}$ ,首先要解算出相关方阵的特征值  $\lambda_i$ 。解算  $\lambda_i$  和  $W_{ii}$  的方法很多,如克雷洛夫法、迭代法、反幂迭代法等 (i,i) 。 这些解法一般较烦琐难掌握。 雅可比法虽然可以同时获得全部特征值和权系数值,但只适用于大型电算机解算。在综合数值分析法试验中选用的是可以手算的主成份分析简便解算法 (i) 。

依主成份值  $Y_1$ ,  $Y_2$  制作的散点图,虽然可以将土壤初步分成几类,但不能解决前面提到的土壤分类问题。判别分析正是因此被提出来综合使用的。

综合数值分析法中的判别分析,采用的是两组线性判别分析<sup>[1]</sup>。它通常是在室内外按经验或按聚类分析初步分类后,使用的较严密的数学方法。虽然综合数值分析法中的判别分析,也需解算判别方程  $(R_{sp})$ 、类组均值的判别值  $(R_s$  和  $R_p)$ 、类间判别点的判别值  $(R_{sp})$ ,也可解决介于两类组间待定点的归类问题,但它与常规的两组判别分析有显著不同。首先,综合数值分析法中所用数据,是各土壤的主成份值  $Y_1$ ,  $Y_2$  (累计贡献率应达70%),因只有两个变量(但却包含原有多项性质的信息),故解算简便,F 检验显著。常规判别分析用的是土壤的许多项性质数据,不仅费工,且往往因各类组的剖面点少或点数不等,F 检验不显著,而无法作判别分析。其次,综合数值分析法中判别分析的主要目的不在于判别,而在于求出综合系数  $(K_{sp})$ ,把判别分析结果与主成份分析结果联结起来。而常规判别分析中是找不到  $K_{sp}$  值的。

综合系数  $(K_{\mathfrak{sp}})$ ,实质上是类间判别点至两邻类均值点的距离比,故也可称之为"类间距比"。根据两组线性判别分析的基本原理,我们导出  $K_{\mathfrak{sp}}$ ,的计算公式为:

$$K_{gp} = \frac{R_g - R_{gp_o}}{R_{gp_o} - R_p} \tag{3}$$

式中g,p为相邻两类组的编号,规定p > g。类组的编号,特规定以Y,值的类均值大小为序,从正到负编类组号为I、II、····K。

按照上述编类组的规定和综合系数的实质含义,我们导出类间判别点的主成份值公式:

$$Y_{1gp_o} = \overline{Y}_{1p} + (\overline{Y}_{1g} - \overline{Y}_{1p})/(K_{gp} + 1) 
Y_{2gp_o} = \overline{Y}_{2p} + (\overline{Y}_{2g} - \overline{Y}_{2p})/(K_{gp} + 1)$$
(4)

<sup>1)</sup> 卜兆宏: 主成份分析简便解算法及其应用问题(待刊稿)。

式中  $\overline{Y}_{1p}$ ,  $\overline{Y}_{1g}$  和  $\overline{Y}_{2p}$ ,  $\overline{Y}_{2g}$  分别为 g, p 两邻类组  $Y_1$  和  $Y_2$  值的均值。 依  $Y_{1gPo}$ ,  $Y_{2gPo}$  值在散点图上标出的点,就是两邻类组的判别点,它必位于两类均值点的联线上。 为了绘出两邻类组的判别界线,尚须按下式计算判别界线在  $Y_1$ ,  $Y_2$  轴上的数值:

$$Y_{100} - R_{gP_0}/f_1, Y_{200} - 0 
Y_{100} - 0, Y_{200} - R_{gP_0}/f_2$$
(5)

式中  $f_1$ ,  $f_2$  分别为判别方程  $R_g$ , 中的两个系数值。联结  $Y_1$  和  $Y_2$  轴上两点所作的联线,必通过已标注出的判别点。这条线,便是两邻类间的判别界线。对于同一土类分级而言,级间分级界线一般是一条直线 (见图 2)。对于若干土类而言,分类界线则是由多条判别界线相交组成的多角折线 (见图 1)。这就是判别界线的数学绘制方法。

在不同土类间的判别界线,往往相邻两类间有数条穿过,然而其中只有一条的可靠性最大。 为了找出这一条最可靠的判别界线,作为分类界线,尚须依次遵循如下取舍三原则.

- 1. 只留取属于该两类的类间判别界线,而会去其它判别界线;
- 2. 只留取类间距较短的两邻类的判别界线,而舍去类间距较长的判别界线;
- 3. 只留取同类名的判别界线相交的点,作为分类界线的转折点,而舍去非同类各线交点。

计算出各项土壤性质分类界限指标  $(X_{igPo})$ ,是综合数值分析法的一项重要成果,以供土壤分类中应用。计算  $X_{igPo}$  的公式,我们仿求判别点坐标公式导出为:

$$X_{igp_o} = \bar{X}_{ip} + (\bar{X}_{ig} - \bar{X}_{ip})/(K_{gp} + 1)$$
 (6)

式中  $\bar{X}_{is}$ 、 $\bar{X}_{ip}$  分别为相邻类组的第 i 项性质数据的均值。

应用  $X_{isr_o}$  于土壤分类时,待定土壤出现在  $g(\mathbf{g}_p)$  类的概率  $P_g(\mathbf{g}_p)$  按下式计算:

$$P_{g(\mathbf{g}(p))} = \sum_{1}^{N} T_{j1} \cdot W_{1} + \sum_{1}^{N} T_{j2} \cdot W_{2}$$
 (7)

式中  $T_{i1}$  和  $T_{i2}$  为指标权,分别由  $W_{i1}$  和  $W_{i2}$  算得; $W_1$  和  $W_2$  分别为  $Y_1$  和  $Y_2$  值在判别方程中的权值,称判别权;N 为待定土壤的性质  $(x_{ij})$  大于(或小于)分类界限指标  $(x_{igPo})$  的项目数。当  $x_{ij} = x_{jgPo}$  时,则在  $P_g$  和  $P_P$  中都应计算在内。 根据计算结果,以  $P_g$  , $P_P$  的大小来决定该土壤归为 S 类或 P 类。

综上所述,综合数值分析法的出发点和归宿点,均为反映土壤类型差异和成土过程差 异的各项性质数据。 其所取得的各项土壤性质的分类界限指标,有可能为生产和研究中的大量土壤分类工作,提供较为确切的依据。因此,综合数值分析法有广为应用的可能。

# 二、应用程序

基于上述基本原理,综合数值分析法应用于土壤分类中的程序可分为: (一)数据准备; (二)求出相关方阵; (三)主成份分析; (四)判别分析; (五)求出各项土壤性质的分类界限指标; (六)验证与应用等阶段。

#### (一) 數据准备阶段

主要是将具有土壤分类意义的野外观测资料和室内理化分析资料进行取值,并整理成表。在土壤遥感信息分类试验中,我们选用了 0.4—0.5, 0.5—0.6, 0.6—0.7, 0.7—0.8, 0.8—1.1 微米五个波段范围各种土壤及有关地物的波谱反射特性数据,其中后四个波段与美国陆地卫星的工作波段一致<sup>[7]</sup>。由于土壤性质 (j) 数据仅有五项,而土壤剖面点数和类型数均较多,故在介绍该法应用程序时,以此作为实例按步介绍。 在数据准备阶段,整理成表的格式如表 1。

性质号 j No. of properties  $|0.4-0.5|0.5 \rightarrow 0.6|0.6-0.7|0.7-0.8|0.8-1.1\mu$ Υ. Y2 . 3 5 1 2 4 剖面号 1 No. of soil profile 10.80 16.05 21.41 23.72 25.68 -0.2901 - 0.39841 (淤泥土) 14.05 20.98 27.15 30.45 34.61 1.2075 - 0.55402 (油沙土) 3 (黄壤) 13.42 24.72 35.89 41.55 41.52 2.5229 0.0002 10.75 21.19 32.30 36.03 41.88 0.4119 6 (红黄壤) 1.6354 : : : : : : 2.6971 39 (绿黄梧桐叶) 6.99 25.93 43.11 61.27 61.82 4.0083 40 (黄棕壤) 4.48 5.80 7.70 9.75 13.03 -3.26310.1162 均值 $\bar{X}$ , 10.2798 16.3280 21.9305 26.6105 31.2105 Mean 标准差 S, 4.2023 6.6350 9.4152 11.6775 13.2648 Standard deviation

表 1 土壤的各项性质数据\*

#### (二) 建立相关方阵阶段

根据表 1 数据,按(1)式(或用计算器直接)计算出相关方阵中的元素,即相关系数  $r_{12}$ ,  $r_{13}$ ,  $r_{14}$ ,  $r_{15}$ ,  $r_{23}$ ,  $r_{24}$ ,  $r_{25}$ ,  $r_{34}$ ,  $r_{35}$ ,  $r_{45}$  和表 1 中的各项性质均值  $\overline{X}_j$ 、标准差  $S_j$  值。 因  $r_{11} = r_{22} = \cdots = r_{jj} = 1$ ,可不必计算。由  $r_{jj}$  和  $r_{jj}$  组成的相关方阵 R,列在表 2。 相

表 2 相关(系数)方阵(只写出下三角元素)

j(或l)	1	2	3	4	5
1(或1)	0.4—0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-1.1µ
1 0.4-0.5	1				
2 0.5-0.6	0.8799	1			
3 0.6-0.7	0.7346	0.9653	1		
4 0.7-0.8	0.4816	0.7857	0.8410	1	
5 $0.8-1.1\mu$	0.3929	0.6912	0.7464	0.9815	1

Table 2 Correlation matrix

Table 1 Data of properties in profile of soils

<sup>\*</sup> 本表数据实质上是不同土壤及其植物表面与绝对白体表面反射太阳光的波谱段能力的比值,即波谱段反射率, 又称波谱反射特性。

关方阵实际上是对称阵,故只需写出上(或下)三角元素。

#### (三) 主成份分析阶段

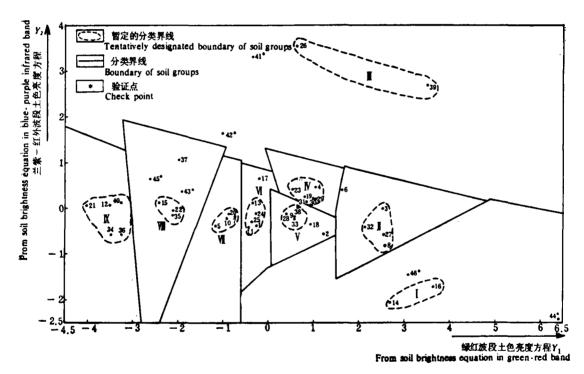
表 2 经主成份分析简便解算法获得的主成份方程为:

 $Y_1 = 0.0915x_{i1} + 0.0725x_{i2} + 0.0509x_{i3} + 0.0392x_{i4} + 0.0322x_{i5} - 5.2862$ 

 $Y_2 = -0.6610x_{i1} - 0.0415x_{i2} - 0.0092x_{i3} + 0.0367x_{i4} + 0.0410x_{i5} + 0.2405$ 

依  $W_{ii}$  (或  $T_{ii}$ ),  $Y_i$  可命名为"绿红波段土色亮度"方程; $Y_2$ 为"兰紫红外波段土色亮度"方程。 其所含五项性质数据的土壤信息(或称贡献率)分别达 80.5% 和 16.5%,累计达 96.5%。

将表 1 中各项性质数据代入主成份方程,计算出各土壤剖面的  $Y_1$  和  $Y_2$  值,并填在表 1 中。依  $Y_1$  和  $Y_2$  值制作的散点图,如图 1 所示。图中分类界线是在判别分析后绘制的。



I. 碱土 II. 小粉土(塝坪田) III. 植被覆盖的土 IV. 黄土(岗丘地) V. 黄白土(小冲田) VI. 黄马肝土(塝冲田) VII. 马肝土、青泥土(大冲、河湖畔) VIII. 湿淤泥土、草甸黑土 IX. 黄棕壤(山丘上)

I .- IX. -- Different soil groups

图 1 绘有分类界线的散点图

Fig. 1 The scattering points delimited in higher category of soil classification

#### (四)判别分析阶段

1. 判别分析: 首先,根据散点图上明显聚合一处的土壤点,暂用闭合虚线粗划为一类,而对于离散较大的点列为待定点。然后按类组整理主成份分析成果为表 3。 根据表

# 表 3 主成份分析成果整理表

Table 3 Results of principal composition analysis adjusted

g, ng 类组号及巳 归类点数 No. of group and number of points classified	剖面号 No. of soil profile	$\begin{array}{c} Y_1 - Y_2 \\ (r_{Y_1Y_2}) \end{array}$	g, ng 类组号及巴 归类点数 No. of group and number of points classified	剖面号 No. of soil profile	$Y_1$ $Y_2$ $(r_{Y_1Y_2})$
	14 16	2.6914 -2.0589 3.6446 -1.6990		26 39	0.6209 3.5195 4.0083 2:6971
1 n <sub>1</sub> = 2 (灰白色表土类)	类均值 $\overline{Y}_{t1}$ Mean of group 标准差 $S_{Yt}$ Standard deviation	3.1680 -1.8790 0.6740 0.2545 (1)	III n <sub>III</sub> = 2 (植被覆盖的土类)	类均值 $\overline{Y}_{till}$ Mean of group 标准差 $S_{Yt}$ Standard deviation	
A CONTRACT OF THE PROPERTY OF	28, 33 38	0.5633 -0.1852 : : 0.6409 -0.0359		10, 7	-1.1798 -0.3937 : : -0.8839 -0.1166
V n <sub>V</sub> = 4 (黄白土类)	类均值 $\overline{y}_{tV}$ Mean of group 标准差 $S_{Yt}$ Standard deviation		VII n <sub>VII</sub> = 4 (青泥土类)	类均值 $\overline{Y}_{tVII}$ Mean of group 标准差 $S_{Yt}$ Standard deviation	-0.9310 -0.2321 0.1828 0.1178 (0.8434)

3 中的类组均值、标准差、相关系数  $(r_{Y_1,Y_2})$ ,便可很快解算出两邻类组的判别方程  $(R_{gp})$ 、判别点的判别值  $(R_{gp_0})$ 、判别权  $(W_1, W_2)$ ,并按 (3)、(4)、(5) 式解算出综合系数  $K_{gp}$ 值,以及各判别点的坐标值。判别分析的数据成果列于表 4 中。

2. 类间判别界线的绘制和取舍:按表 4 中三个判别点的坐标值,在散点图上标剌出

#### 表 4 判别分析成果表

Table 4 Result of discriminatory analysis

分类界线号 No. of boundary	Discri	权(%) min <b>ato</b> ry, eight	判別值 Discriminatory value	综合系数 Synthetical coefficient	nthetical Verification of visibility		三个判别 Coorginate iscriminat		!
of groups	$W_1$	$W_{1}$	Res.	K <sub>t</sub> ,	$F_{N_1,N_2}$	Yiero	Y2870	Y <sub>1 %h</sub>	Y <sub>286</sub>
IV 分类界	37.96	62.04	-674.5981	2.0000	$F_{2,3} = 729.34**$	1.4136	-0.7228	3.2059	-1.2929
II V	100.46	-0.46	-98.62 <b>08</b>	1.0000	$F_{2,} = 105.40**$	1.4920	-0.2948	1.4835	51.2050
II IV	91.55	8.45	-52.8628	1.5000	$F_{2,7} = 63.77**$	-1.5003	-0.0040	1.5011	-7.5722
III VIII	26,69	73.31	924.3430	1.5000	$F_{z,z} = 1408.10**$	-0.4160	1.2273	4.4537	1.1225
IV V	24.58	75.42	-15.5372	0.6667	$F_{2,7} = 21.88**$	0.7358	0.1161	0.9909	0.4510
v vi	100.97	-0.97	-8.1562	1.2500	$F_{2,6} = 97.63**$	0.0495	-0.1898	0.0452	2.0014
VII VIII	93.80	6.20	51.0042	0.7500	$F_{2,4} = 34.04**$	-1.49 <b>0</b> 5	-0.1441	-1.4366	3.8395
VIII IX	94.29	5.71	60.7672	1.6667	$F_{2,j} = 20.67**$	-3.0295	-0.1260	-3.0437	_27 <b>00</b> 38

<sup>\*</sup> 表中所列  $F_{N_1,N_2}$  数值为计算值,凡大于显著性水平  $\alpha = 0.05$  查表值者为显著;凡大于以  $\alpha = 0.01$  查表值者为非常显著;\*\*计算了十九条界线,有一条不显著,十七条非常显著。

表 5 土壤性质分类界限指标

Table 5... Delimitation index of soil properties

反算判別值	R's Discriminatory value of anticomputation		ŧ	-674.48		-98.75			-15.53		•	-924.30		51.00		
反算类间判别点坐标值	Y' <sub>ifPo</sub> , Y' <sub>ifPo</sub> Coorginate of anticomputation of discriminatory point			1.415, -0.722		1.494, -0.294			0.733, 0.117			-0.415, 1.227		-1.490, -0.144		
0.1819@ 0.2957@	5	ä	38,165	35.316	33.892	37.531	41.170	37.442	36.010	33.892	67.745	37.648	17.583	21.847	25.045	•••
0.2094@	4 6 7 0		35.405	30.690	28.332	33.004	37.675	32,110	30.599	28.332	55.785	31.168	14.757	17.857	20.182	
0.2296@ 0.0075©			33.425	27.298	24.235	29.016	33.798	27.026	25.910	24.235	22.065	18.318	12,487	15.283	17.380	
0.2313Q	2 3		29.290	21,920	18.235	21.618	25.002	18.737	18.536	18.235	17.865	12.694	9.247	11.810	13.732	
0.1478© 0.4369®	1 2 0 -4 0		21.995	15.106	11.662	13.482	15.302	10.395	10.902	11.662	6.145	6.390	6.553	8.294	9.600	
各项性质的指标权与权序 T <sub>H</sub> Weight of index of properties and its order T <sub>H</sub>	3.p 类间性质界限指标 Xists。 Delimitation index of properties of	8, p groups	I类 \ \X_\tag{X}	IV 类界 - XiIV。	, V 类 \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \	IIV 类界 Xhiv。	II 类 、 Z <sub>fII</sub> *	IV 类 \(\bar{X}_{\pi \sum_{\pi \sim_{\pi \sum_{\pi \sum_{\pi \sum_{\pi \sum_{\pi \sum_{\pi \sim_{\pi \sum_{\pi \sum_{\pi \sim_{\pi \sim_	IV V类界 Xiiv v。	V 类 \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \	111 类 \[ \bar{\bar{\bar{\bar{\bar{\bar{\bar{	III VIII类界 Xjiii viii。	VIII 类 X̄ <sub>IVIII</sub> *	VII VIII 类界Xjvii viii。	VII 类 \(\bar{X}_{\invIII*}\)	
项性质的指 ight of in and		W <sub>2</sub>		62.04		-0.46			75.42			73.31		6.20		
<b>₩</b>	判别权(%) Discriminatory weight	74		37.96		100.46	,		24.58			56.69		93.80		

\* 为8或户类的性质类均值。

点位。一般至少可标刺出两个点。联结这些点所作的直线就是类间判别界线。 同时,应以三个点是否在同一直线上,或以类均值点联线检查类间判别点是否在其上,作为检验前面各阶段计算正确与否的措施。然后,按判别界线的取舍三原则绘出分类界线。例如,图 1 中 I 类与 VI、VII、VIII、IX 类虽都作过判别分析计算,但按取舍三原则只留取了 I VI 界线,而舍去了其余界线。又如,III VIII 分类界线只留取很短一部分。

由于判别界线均按严格的数学公式解算,所以分类界线一经绘出,其余尚未归类组的点(如 2 和 6 等),也就解决了它们的最后归类。一般无须用判别方程来判定待定土壤的归属。例如位于 II、IV 界线旁的 6 点,以其  $Y_1Y_2$  值代人判别方程得  $R_{II IV} = -54.7170$ ,与  $R_{II IV} = -52.8628$  比应判归为 II 类。这与用图 1 上的分类界线来判定归类的结果是一致的。

### (五) 求出分类界限指标阶段

- 1. 计算每类各项性质的类均值  $(\bar{X}_{ig, k, ip})$ : 在计算类均值时,新判定的点一般不参与计算。 如果要把同一类界内的所有点都参与类均值计算,则须重新判别分析和算出新的  $K_{gp}$ , 并重新计算和绘制判别分类界线。这样可以使类均值更接近于实际。
- 2. 按 (6) 式计算各项性质的分类界限指标 ( $X_{igPo}$ ): 利用土壤波谱反射特性数据 进行分类的界限指标值,共有 19 条分类界线的  $X_{igPo}$  值。表 5 中只详列了 V、VII 两类与其它邻类的  $X_{igPo}$  和  $\overline{X}_{pRP}$  值。

至此,综合数值分析法得到三项主要成果:一是土壤的主成份方程;二是绘有分类界 线的散点图;三是各项性质的分类界限指标。

## (六) 验证与应用阶段

这是检验综合数值分析成果是否符合实际的过程,也是其在生产和研究中发挥实际效用的过程。所谓验证,就是将已知类型的土壤来验证成果的可靠性。所谓应用,就是用 其成果将未知的待土壤较确切地归类。

对于绘有分类界线的散点图,在验证和应用时的程序为: 1. 将待验证或待定土壤的性质数据代人主成份方程,计算出 Y<sub>1</sub> 和 Y<sub>2</sub>值; 2. 依 Y<sub>1</sub> 和 Y<sub>2</sub>值在散点图上标出点位。然后看该待定验土壤点位所处类型是否与已知的类型名称同。如果相同,则表明成果是可靠的。对于待定土壤,则散点图上所标注的类型,就是该待定土壤应归的土壤类型。

对于土壤性质分类界限指标成果,在验证和应用时有两种情况: 一是待验或待定土壤的各项性质数据均在分类界限指标  $X_{igPo}$  一边的情况; 二是有的性质数据在  $X_{igPo}$  这一边,有的又在那一边的情况。前一种情况,则可根据  $X_{igPo}$  立即作出判断,验证其可靠性或待定土壤应归的类。后一种情况就需用 (7) 式计算待验或待定土壤出现在 8 类和 P 类的概率,以概率大小定其归类。 此外,也可采用将其性质数据代入主成份方程算出  $Y_1$  和  $Y_2$  值,再用相应判别方程进行归类。不过,这样就没有用  $X_{igPo}$  值那么简便。

为了验证综合数值分析成果的可靠性,我们选择了一些已知其类型的波谱反射特性数据  $\rho_{\Delta 1}$ ,如表 6 所示。依其性质数据代人主成份方程算出  $Y_1$  和  $Y_2$  值,并标在图 1 上,表明散点图上的分类界线是可靠的。如  $41^*$ , $42^*$  点落在范围最大的 III 类,表明它确实

表 6 待验证的土壤波谱反射特性数据与验证归类结果

Table 6 Data of spectrum reflectivity of soils and their grouping

的 No. of	土壤名称	测试仪器与方法 Instrument		被 Reflec	放段反射率 pai Reflectivity of wave band	ροι wave		主成份值 Value of principa composition	份值 principal sition	概率(%) Probability	(%) bility	应归类号 No. of classified
profile	NAME OF SOIL	and method	1	2	3	4	5	Υı	Υ,	$P_{\boldsymbol{g}}$	Pp	group
41*	长有小麦(盖度 90%) 的土 302	302 仅野外测试	4.1	8.0	5.2	48.4		-0.3498	3,3085	56.5 -0.3498 3.3085 P <sub>III</sub> 81.58	PvIII 18.42 III	***
42*	长有小麦(盖度 80%)的土	302 仪野外测试	5.7	9.2	11.0	33.0	38.5	-1.0044	1.6505		PvIII 18.42 III	*** E
43*	红鹤林士	USV-2G 仪室内测试	5.93	9.20	14.10	18.10	22.40	-1.9281	0.3789	18.10   22.40   -1.9281   0.3789   PIII 35.97	PvIII 64.03	VIII ₩
44*	龟裂性土	USV-2G 仪室内测试	28.80	40.23		47.90   50.50	51.89	6.3542	-2.4191	6.3542 -2.4191 P1 100.0	$P_{\mathbf{V}} = 0$	<b>™</b>
45*	红颜林士	302 仪室外模拟测试	4.02	6.74		10.69 14.98	21.30	21.30  2.6125   0.6532   Pv11 0	0.6532	Pv11 0	$P_{\rm VIII}$ 100.0	VIII ∰
46*	龟裂性土	302 仪室外模拟测试	20.21	26.67	35.60	35.60 36.48	38.70	3.0763	-1.4473	38.70 3.0763-1.4473 P <sub>1</sub> 100.0	$P_{\mathbf{v}} = 0$	**

表 7 碱化土壤的分级界限指标

Table 7 Delimitation index of alkaline soils

0.1141@ 0.0095@	全盐(%) Na+ Total salt (Ca+++Mg++)				因 7. 值较小,分级作用不大,故略去。		
0.0563® 0.1577®	全盐(%) Total salt				1较小,分级1		
0.1347@ 0.0699@ 0.0563@ 0.7026@ 0.1577@	代換量*				B T₁ (		
0.1347© 0.7026©	Hď		9.559	9.288	0.030	8.803	8.913
0.1528@ 0.0137@	代換性钠* Exchangeable	SOCIUM	2.524	1.447	1.057	0.397	0.710
0.1549®	总碱度* Total alkalinity	(503 + BCO3)	0.464	908.0	0.235	0.182	0.196
0.1554©	碱化度 (%)	ESF	40.315	21.625	12.918	4.123	8.765
0.1620D 0 ®	現余碳酸钠* Residual sodium carbonate (CO*+HCO-) -	$(C_a^{++} + Mg^{++})$	0.395	0.249	0.167	0.062	0.099
T <sub>11</sub>	及其界限指标 odivisions and	ration index	Хи п,	X)11 111,	X;111 IV.	XIIV Vo	X,111 V.
t权和权序 of propertie rder	分级界号及其界版 No. of subdivision	its delimitation	II II	111 111	VI III	IV V	111 V**
各项性质的指标权和权序 Weight of index of properties and its order	(%) natory ght	14,	5.80	96.6	-0.97	41.71	5.59
各項 Weight	判别权(%) Discriminatory weight	<u>,</u>	94.2	90.04	100.97	58.28	94.41

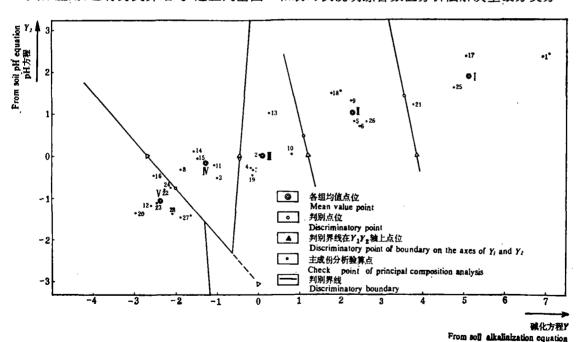
\* 单位为; (med/100g soil); \*\* 个别情况下 III 与V的分级界限指标。

属于覆盖有植被的土壤类。又如灰白色的龟裂性土 44\*, 46\* 皆落在 I 类, 与表层灰白的 碱十、水泥地同属一类,也是符合实际的。

如果仅用分类界限指标成果(见表5),同样也可验证和应用。例如将41\*,42\*的性 质数据对照  $X_{\text{JIII VIII.}}$  值,有第 2,3 两项在 VIII 类一边,有第 1,4,5 三项在 III 类一边, 按(7)式算得在 VIII 类的概率均为 18.42%, 而在 III 类的为 81.58%, 故长有小麦(盖 度大于 80%) 的土壤应归 III 类。43\*, 45\* 若以 X<sub>iIII vIII</sub>。判, 它们的 P<sub>III</sub>、P<sub>VIII</sub> 均分别 为 35.97% 和 64.03%, 应归 VIII 类。由于新疆红柳林土的土色与长江边湿润的淤泥土 的土色是一样的,故以波谱反射特性数据分类就将它们归为同一类了。通过验证表明,综 合数值分析法求出的土壤波谱性质数据的分类界限指标,对于表土的土色分类是可靠的。

# 三、在碱化土壤分级中的应用

综合数值分析法在碱化土壤分级中,选用了黄准海地区 28 个土壤剖面、八项性质数 据。 按前述原理和程序,获得碱化土壤的主成份方程、综合系数  $K_{so}$  值、指标权和判别 权,以及绘有分级界线的散点图(见图 2)和碱化土壤分级界限指标值(见表 7)等。有关其 具体应用,已有另文介绍<sup>1</sup>。这里列出图 2 和表7,仅说明综合数值分析法解决土壤分类分



III. 碱化土

l. 磁土 Solonetz II. 强碱化土 Strongly solonetzic soil Solonetzic soil IV. 弱碱化土 Weakly solonetzic soil V. 非磁化士 Non-solonetzic soil

图 2 绘有分级界线的散点图

Fig. 2 The scattering points delimited in lower category of soil classification

<sup>1)</sup> 万洪富: 黄淮海平原碱化土壤分级的初步研究。

级问题的能力。

图 2 显示出,在一般情况下可将黄淮海平原土壤的碱化分五级,即碱土(I)、强碱化土(II)、碱化土(III)、弱碱化土(IV)和非碱化土(V); 个别情况下因 IV V 与 III IV 两界线相交,可分为四级。

表 7 除列出分级界限指标外,还有指标权、权序和判别权,它们阐明了各项性质在碱化分级中的作用大小。作用最大的是残余碳酸钠,其次是碱化度……,pH 仅在非碱化与弱碱化分级中起较大作用,可达 37.15% 的概率。所获得的碱化土壤分级界限指标,证实了国内由经验总结出的有关论述的科学性<sup>[2,5]</sup>。例如,他们认为,碱化度达 20—30% 为中度碱化土,大于 50 为碱土;而本文表 7 是 21.63—40.32% 为强碱化土,大于 40.32% 为碱土。 这说明综合数值分析法是使经验分类数值建立在更加科学基础上的有效手段,同时也显示出它在解决土壤分类分级问题方面的能力。

# 四、精度讨论

一种数学方法的精度涉及两方面: 一是应用精度;另一是方法本身的精度。 综合数值分析法,同样也是如此,应从这两方面来讨论。

综合数值分析法应用于土壤分类所得结论的精度,除由方法本身的精度决定外,尚受参与计算的数据精度的影响。数据精度,包括土壤剖面选择的代表性,理化分析的可靠性,野外观察特征数字取值的合理性和具有较大分类意义的性质数据漏失程度。因此,为了确保该法获得的分类结果可靠,首先应依据土壤学原理检查和保证数据的精度,选用具有分类意义的性质作变量。同时,应使剖面数量至少应大于变量数的三倍,最好是四倍以上。就碱化土壤分级而言,开始漏失了残余碳酸钠,只做了七变量统计,虽然分出的级组与八变量一致,但求出的分级界限指标就不如八变量的可靠。

综合数值分析法的本身精度,包括有方法原理是否严密,解算方法是否可靠。这是综合数值分析法能否应用于土壤分类的关键。

虽然综合数值分析法袭用了相关分析中的相关系数计算,采用了主成份分析简便解算法,改革了判别分析,使之更简便并综合为一体,但是仍可证明其原理的严密性。我们曾采用常规判别分析作过八项原始数据的判别分析,所费时间比综合数值分析法中的多数十倍,其结果如表 8。表中的  $K_{gp}$  值与括号内综合数值分析法的  $K_{gp}$  值,几乎完全相等,最大误差为 0.0002,小于限差值  $\delta$  (=  $0.0001 \times \sqrt{9}$ )。 这说明综合数值分析法中的判别分析原理是严密的。由于其变量是取自主成份分析结果,所以两种  $K_{gp}$  值的相等也间接证明了主成份分析简便解算法的原理是严密的。此外,从表 5 中"反算类间判别点坐标  $Y_{1gpo}$ ,  $Y_{2gpo}$  值"和"反算判别值  $R_{gpo}$ ",与表 4 中按原理公式算得的相应值几乎相等的情况,不仅说明分类界限指标是散点图上分类界线的数据描述,而且也证明了综合数值分析法的严密性。

经改革后的判别分析,由于计算较简单,故不存在解算方法的可靠性问题。解算方法的可靠性,主要存在于所采用的主成份分析简便解算法。这种简便解算法的精度已有专文介绍了,表明它与大型电算机的计算结果相同,精度是可靠的。

	表 8	用八项性质	数据常规判别	分析的结	i果*	
Table 8	result of	conventional	discriminatory	analysis	of eight	variables

判别方程	Dis	判 别 值 criminatory v	alue	综合系数 Synthetic coefficient	F <sub>N1</sub> , N,检验 Verification	
Discriminatory equation R <sub>sp</sub>	R <sub>g</sub>	Repo	R,	K <sub>tp</sub>	of visibility	
$R_{111} = -7.3512x_1 + 1072.4x_2  -2.6x_3 + 2.2x_4$	-121.3721	-114.0147	-108.1288	1.2500	由于变量数为	
$\begin{array}{l} -870.4x, +1106x_6 \\ -6.3x, +28.6x_4 \end{array}$	(-18.630)	(-12.970)	(-8.442)	(1.2500)	8 (即 N, = 8), 两级组点数	
T) (ab.)	115.5838	89.2958	67.3887	1.2000		
$R_{\rm II\ III}=($ <b>略</b> $)$	(-42.900)	(-20.485)	(-1.805)	(1.2002)	+ n, - 8 - 1 有可能为负数)	
D. (alta)	-1075.0240	-1039.4617	-986.1181	0.6667	故无法做F <sub>N,N</sub>	
$R_{\text{III IV}}=(\mathbf{B})$	(-1.186)	(5.159)	(14.677)	(0.6666)	检验或很不显著。	

<sup>\*</sup> 括号内数据为综合数值分析法的结果。

因此可以认为,综合数值分析法的原理是严密的,解算方法是可靠的。

# 五、结 语

综合数值分析法是一种使主成份分析和判别分析综合成一体的数学方法。在土壤波谱数据分类试验中,它不仅能将参与计算的土壤分成不同类型,而且能确切地求出各项土壤性质数据(波段反射率  $\rho_{\Delta\lambda}$ )的分类界限指标。因此,它在土壤分类分级的生产和科研中有着广泛的应用前景。

在寻求土壤碱化分级界限指标的应用中,它显示了简便可靠和不受条件限制的优点。 所获得的分级界限指标、主成份方程和绘有分级界线的散点图等结果,可供黄准海平原进行土壤碱化分级时参考和验证。

综合数值分析法的基本思想,不仅可用于土壤分类,也为土壤波谱反射特性数据处理 提供了一种方法。它在遥感图象的识别分类中也有应用的可能。

#### **卷 考 文 献**

- [1] 中国科学院地质研究所,1978:数学地质引论。地质出版社。
- [2] 田兆顺、董汉章, 1965: 华北平原瓦碱的特性和形成。土壤学报, 13 卷 1 期, 24-37 页。
- [3] 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,16卷2期,172-183页。
- [4] 武汉测绘学院,1961:最小二乘法。中国工业出版社。
- [5] 俞仁培、蔡阿兴, 1978: 盐渍土分类及改良分区专题讨论概况。土壤, 4期, 132-135页。
- [6] Collins W. F., Chang S. H., Kuo J. T., 1980: Detection of hidden mineral deposits by airborne spectral analysis of forest canopies. Alag Colubia University, New York.
- [7] Taraik J. V., 1978: Characteristics of the landsat multispectral data system. U. S. G. S. Openfile Report pp 78—187, Sioux Falls, South Dakota.
- [8] Webster R., 1977: Quantitative and numerical methods in soil classification and survey. Clarendon Press, Oxford.
- [9] Рожков, В. А., 1975: Метод главных компонент и его применение в почвоведении. Почвоведение, № 10, 141—151.

# PRELIMINARY STUDY ON THE APPLICATION OF ANALYTIC METHOD OF SYNTHETIC VALUE IN SOIL CLASSIFICATION

Pu Zhao-hong and Wan Hong-fu
(Institute of Soil Science, Academia Sinica, Nanjing)

#### Summary

The analytic method of synthetic numerical value is a mathematical method in which the principal composition analysis and the discriminatory analysis are integrated into a single combination. In experiment of classification of soil spectra, with this method, not only the soil samples can be divided into different types by calculation, but also the delimitation index of various soil properties (reflectivity of wave band  $\rho_{\triangle\lambda}$ ) can be calculated. Therefore, it may be regarded as the method with a bright prospect in the practice and research of soil classification.

In the application to the subdivision of alkaline soil, this method exhibits its simplicity and reliability to a larger extent. The results obtained in the experiment of subdivision of alkaline soil by this method may be useful for the classifying the alkaline soils of Huang-Huai-Hai plain into lower categories. The basic concept of analytic method of synthetic numerical value, in addition to its application in soil classification, may also provide a way for processing the data involved in spectrum reflective characteristics of soils and may be possibly used in the soil classification and identification of remote-sensing image as well.