

卫星图象土壤类型自动识别与制图的研究

I. 计算机分类及其结果的光谱学和地理学分析*

曾志远

(中国科学院南京土壤研究所)

一、研究区简况与工作步骤

研究区位于干旱的新疆南部,塔里木盆地北缘。大致以阿瓦提县城为中心,西近柯坪县城,东近阿拉尔,北近阿克苏,东南接大沙漠。东西长165公里,南北宽75公里。总面积为12,375平方公里。

本区地势呈簸箕形:西、北、南三面高,中东部低。西部和西北部是天山及其山前洪积扇,柯坪河在其西南角向东流出山口而散失;中北部是位于阿克苏河三角洲上的耕垦活动频繁的大面积(片状)绿洲,阿克苏河由西北流向东南;东北部是洪积平原,为北面几条小河的散流消失之地;南部是广大的冲积平原,喀什噶尔河、叶尔羌河由西南流向东北,形成几条走廊式(线状)绿洲;冲积平原东端是塔克拉玛干大沙漠的北缘,和田河由南流向北。上述诸大河在本区中东部汇合成塔里木河,向东流去(图1)。

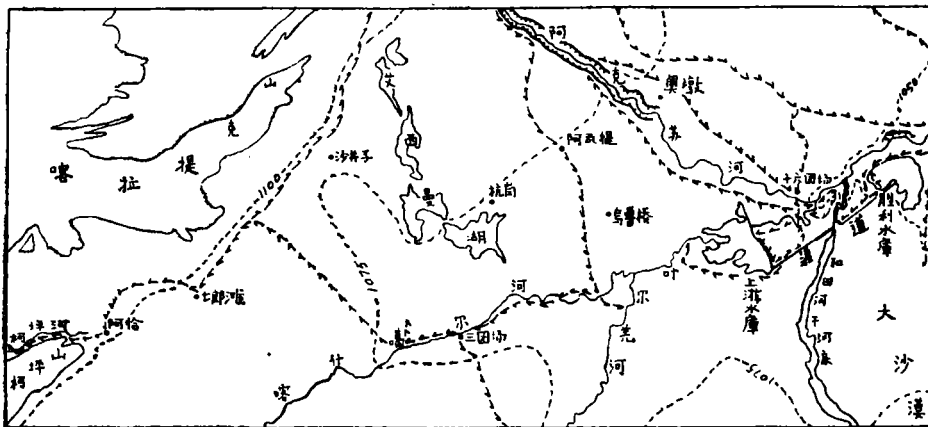


图1 研究区示意图

Fig. 1 Sketch map of the studied region

○ 居民点 → 作者参加考察的路线 ---1075--- 主要等高线 ~~~~~ 河边陡坎

本文作者参加了本区的土壤考察,并对本地区的卫星图象作过初步研究^[2]。

* 计算机处理由北京大学数学系信息实验室张绪定、孙立二同志帮助完成;雷文进同志在本研究中给予帮助。在此一并致谢。

本研究使用的是阿瓦提象幅的图象。系美国 Landsat-1 1972 年 11 月 15 日拍摄。比例尺为 1:100 万。

具体工作步骤是: 先将 4 个波段的透明片(主要是正片)在 JOYCE LOEBLE 微密度计上扫描。每隔 0.5 毫米取一条扫描线, 线上每隔 0.5 毫米取一个采样点。采样点面积取为 $200\mu \times 250\mu$, 目的是保证每一数据是较多象元(约 11 个)的平均。这种小密度采样, 使得能够一次作出较大面积地区的土壤图, 同时又能满足中、小比例尺土壤图的精度要求¹⁾。扫描得到的数据存入计算机, 用非监督和监督分类的方法进行各种实验, 得出适合研究区的参数和具体策略, 然后作出正式的非监督和监督分类图——自动识别图。再参考野外调查及文献资料, 对这种图进行分析、解释, 并将它们变成土壤图。

二、自动识别分类¹⁾

(一) 分类前的处理

1. 4 个波段扫描数据的行和列, 根据明显地物点进行配准。然后去掉不重合的行和列。

2. MSS-4 图象上有几条宽约 0.5 毫米的贯穿全图的横纹。其异常数据用线性插值法校正。

3. MSS-5 图象底片上有一小片乳剂脱落。异常数据除采用线性插值法求得脱落数据外, 还用文献 [2] 介绍的方法, 测定正片和负片(负片上无乳剂脱落)灰标等级与密度关系曲线, 并假定同一地物在正、负片上测得的灰标等级相等, 从而由负片数据算得对应的正片数据。最后取两种校正数据的平均, 作为校正数。这样, 既保证了原有数据的密度水平, 又能反映原有数据可能存在的非线性过渡。

4. 可能是由于扫描过程中仪器背景光发生突变或其它原因, 致使第 4 波段图象扫描数据发生不正常系统跳跃: 第 23 行以后的数据, 系统地增加一个密度等级。为了校正这种异常数据, 我们设计了两种方法: 一种是求出跳跃行前后若干行(如第 20 至第 25 行)各行的平均密度值及其两两邻行之差, 然后以 23 行与 22 行的差减去其它几个邻行差值的平均数, 以此作为异常数。另一种方法是, 如果某些列的数据从第 22 行到 23 行在 5、6、7 三个波段中均无变化或变化极小, 那么将这些列在第 4 波段中的 23 行与 22 行数据的差值取平均, 就是异常数。两种方法算得的异常数很相近。将第 23 行及以后各行的所有数据减去某一异常数, 即得正确数。

5. 照片数字化时, 4 个波段图象的第一条扫描线均扫在底片的影象与片基的交界处。得到的一行异常数据全部删除。

上述校正都得到了很好的效果。

(二) 非监督分类

我们用迭代法进行非监督分类。通过试验确定了下述策略: (1) 找到比较适合分类

1) 参看曾志远、张绪定、孙立《小密度采样卫星扫描数据的微型机处理及其在土壤分类制图中的应用》, 载《遥感技术研究与应用汇编》(国家科学技术委员会成果办公室编辑, 将由科技文献出版社出版)。

区的类别数,并使它自始至终保持不变;(2)保证有足够多的迭代次数,不使它收敛太快。为此,我们通过试验将类别数定为17。基本上用101图象处理系统的方法定初始聚类中心,得到16类,又人为附加一类。然后将最大象元数(SIZEMAX)和拒绝门坎(Rejection threshold)定大,将停止门坎(Stopthreshold)、最少象元数(SIZEMIN)和最小距离(Tooclose)定小。所用具体参数如下:

Initial class number = 17
 Largest number = 30
 SIZEMAX = 0.40 SIZEMIN = 0.001
 Rejection threshold = 10,000
 Stopthreshold = 1.00
 Tooclose = 2 Feature = 4

用上述参数和作法,得到了较好的非监督分类结果。迭代次数为69。

表1 监督分类各类的含义、样本数、先验概率和输出符号

Table 1 The implication, sample capacity, prior probability and output symbols of the categories in the supervised classification

计算机打印序号 Computer printing order numbers	代表的主要土壤(或景观)类别 Main representative soil (or landscape) categories	样本数 Sample capacity	估计的先验概率 Evaluated prior probability	规定的输出符号 Given output symbols
1	水(无陆地出露,无植被覆盖的开放水面)	165	0.010	#
2	草湖(间有陆地出露或间有植被覆盖的水面)	113	0.015	\$
3	沼泽	92	0.015	/
4	沼泽盐土;盐化沼泽化草甸土	156	0.030	!
5	水分较多的绿洲潮土(绿洲潮土 I)	162	0.060	&
6	地面潮湿、植物较密的典型盐土	115	0.045	[
7	植物较密、地面潮湿的红柳林盐土;红柳林土;胡杨林土	122	0.055	:
8	水分较少的绿洲潮土(绿洲潮土 II)	190	0.060	%
9	轻度荒漠化或疏林下的胡杨林土	188	0.050	-
10	绿洲黄土;荒漠化红柳林盐土;强度荒漠化胡杨林土	160	0.070	□
11	有沙土包的荒漠化胡杨林土;植物稀疏的草甸盐土	225	0.050	*
12	荒漠化胡杨林盐土(或间有荒漠化草甸土)	209	0.035	+
13	典型的半固定风砂土	458	0.080	-
14	砂性龟裂状残余盐土(间有龟裂性土)	169	0.040	(
15	流动风砂土或流动性较大的半固定风砂土;植物很稀疏、地面有盐霜的典型盐土	367	0.060	>
16	有密集砂土包的龟裂状残余盐土;有稀疏植被、地面呈灰色的(典型的)砂砾质棕漠土	283	0.080	~
17	龟裂状残余盐土(间有龟裂性土);龟裂状棕漠土	268	0.070	:
18	流动性很大的半固定风砂土或流动风砂土	383	0.070	x
19	流动砂丘;地面基本裸露的复盐霜典型盐土	123	0.035	↑
20	地面基本裸露的复盐霜典型盐土;流动砂丘	253	0.070	.

(三) 监督分类

我们用贝叶斯最大似然法^[3]进行监督分类。先根据实地调查,并参考现有的土壤图^[1]和非监督分类结果,选出 7880 个训练样本,进行监督分类预试验。预试验的土壤类别定得多而单纯,共 34 类。打印出分类结果图、训练样本混淆矩阵和 136(34 × 4) 个训练样本直方图。

由预试验结果来看,有些土壤分不开,个别土壤类直方图有双峰。我们将 136 个直方图两两对比,对原定的 34 类进行调整,最后缩减到 20 类。这 20 类都是至少有一个波段的直方图与其它类不重叠或很少重叠的,因而是可以分开的。但这样调整后的类别,多数都不是原来 34 类简单去留后的单纯类,而是两个甚至两个以上的类的集合了。原来的样本中有一些已为其它类的样本所代表,就删去了。

正式监督分类时,实际输入训练样本 4201 个,平均每类 210 个。各类训练样本所代表的主要土壤类别、训练样本数以及估计的先验概率²⁾和规定的输出符号如表 1。

由此得到的监督分类结果是比较理想的。它和非监督分类结果十分一致,但更加详细、准确。

三、光谱学分析及其意义

自动识别的非监督分类图,实际上不是土壤图而是谱类图^[3]。监督分类图虽然在本义上应当就是土壤图,但由于区域广大,对每种土壤类型都选取足够的训练样本并精确定位常常是办不到的^[3];加上光谱本身的限制,也不可能使每种谱类都正好对应于一种土壤,所以监督分类图也仍然不是土壤学家可以接受的土壤图。我们宁可把它也看成是谱类图。

在这两种谱类图上,除了一种谱类正好对应一种土壤单元的情况外,都不同程度地存在着这样的问题:(1)同一种谱类符号可能包括两种或两种以上的土壤单元;(2)两种或两种以上的谱类符号的组合,可能仅代表一种土壤单元;(3)几种同样的谱类符号以不同的比例组合,又可能代表不同的土壤单元。

我们认为,要把谱类图变成土壤图,必须参照实地考察和文献资料,进行谱类组合规律和地理分异规律的研究,从而解决谱类的组合和分裂,并进而解决勾绘土壤界线的问题。

下面先谈谱类的分析。

(一) 谱类的内在序列

各谱类的计算机打印序号,并不反映它们所代表的客观土壤类别的内在序列关系。必

1) 参看雷文进、刘立诚等编《塔里木盆地北部地区土壤图》

2) 先验概率是根据卫片目视判读、非监督分类结果和对这个地区各类土壤面积关系的粗略了解估计的。先将要划分的土壤类别的概率粗估计为 7%、4.5%、2% 三个等级,然后在每一等级内部进一步估计各类出现概率的差异。结果共分 11 个等级。

需建立合适的谱类排列顺序。我们考虑了三种排序方法。一是按第 5 波段密度值由大到小排列。它在一定范围内是各谱类的植物覆盖由密到疏的反映。二是按 4 个波段密度值的总和由大到小排列。它是各类的总辐射水平由低到高的反映。三是考虑类间距离。从密度值最大的类开始,到密度值最小的类结束,用一条距离最短的线把所有的类连成一个系列。对于非监督分类结果的各谱类,此法排成的序列如表 2。

表 2 非监督分类 17 个谱类按总距离最短排成的序列

Table 2 The least-total-distance sequence of the 17 spectral categories in the unsupervised classification

序列号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
打印序号	17	16	14	8	12	15	6	7	13	11	4	10	5	3	2	9	1
类符号	#	\$	/		&	[%	=	@	+	*	(~	-	>	:	.
类间距离	173.1	113.4	58.0	60.9	31.7	26.9	27.5	22.7	24.4	20.4	19.8	22.2	17.0	16.2	24.4	31.6	

上述三种排列方法得到的序列大同小异。非监督分类和监督分类的排列结果也基本一致。我们把它们综合为表 3。它是依监督分类各类的总密度顺序排列的。按此顺序编定类序号。

(二) 谱类的群分析

由非监督分类或监督分类的各类类间距离矩阵¹⁾, 可以将各谱类合理地分为若干群, 并构成一个完整的系统聚类, 从而更清楚地了解各谱类之间的内在关系。

我们以类间距离为依据, 用一般认为较好的联接方法——非加权配对算术平均法 (UPGMA) 和加权配对算术平均法 (WPGMA)²⁾ 进行联接。以监督分类为例, 两种联接法都经过 7 级联接达到最后的聚类。两种联接法在最低两级联接和最高两级联接方面完全一致, 仅在中间的三级联接方面有些差异, 主要是过渡类型的处理有所不同。图 2 是 WPGMA 法得到的监督分类各类的分类系统。它较理想地反映了实际地物的分异。

(三) 谱类的波段辐射曲线

按照文献[2]介绍的方法, 测定每个波段照片各灰标等级 G 的密度 D , 绘制 $G-D$ 关系曲线; 据此曲线将各谱类在各个波段中的平均密度值 D^0 变换为灰标等级值 G ; 用作者提出的下列公式, 计算各谱类在 4 个波段中的辐射值 R (毫瓦/厘米²·立体角):

$$R_1 = 2.657 - 0.1771G_1$$

$$R_2 = 2.143 - 0.1429G_2$$

$$R_3 = 1.886 - 0.1257G_3$$

$$R_4 = 4.929 - 0.3286G_4$$

1)、2) 同第 185 页 1)

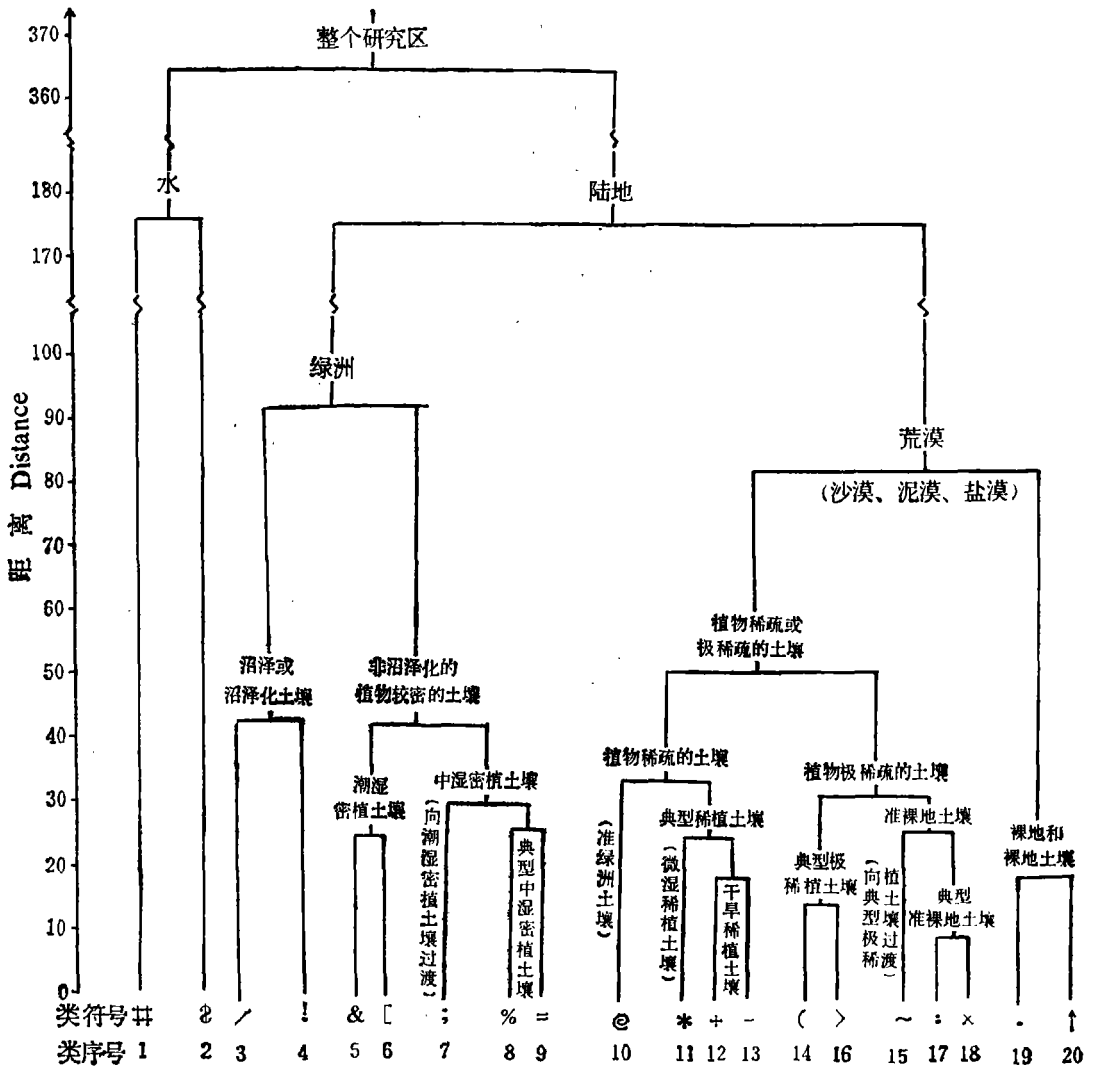


图2 用 WPGMA 法联结的监督分类各类的树系图

Fig. 2 The dendrogram(cluster analysis) of the spectral categories in the supervised classification, using WPGMA connection method

式中R和G的下标4、5、6、7为波段号。得到的辐射值不是真实值,但它提供了一个相对量,据此可以绘出各谱类的平均波段辐射曲线,从而可以对照实地资料了解各谱类的相对辐射关系和土壤学含义。

以监督分类为例,20个谱类的平均波段辐射曲线如图3。因为其中每条曲线都是几百到几千个土壤或景观个体的辐射状况的平均,所以它们可以作为南疆的这些土壤或景观的代表性波段辐射曲线。

根据这些曲线在第5波段中的辐射值 R_5 ,可以把它们分为4大类(纲): (I) 超低辐

表 3 非监督分类和监督分类各谱类的内在序列及序号、符号、点数、密度值和距离对照表
Table 3 The internal sequences of the spectral categories in the unsupervised and the supervised classification and their sequence numbers, symbols, point amount, density values and distance

		173	113	58	61	32	27	23	28	28	27	27	20	20	20	27	29	16	24	32		
非监督分类		距离	498	459	438	436	425	421	413	410	407	401	400	400	391	393	393	386	376			
		5 波段密度	2089	1916	1803	1784	1723	1691	1664	1666	1643	1636	1626	1617	1605	1597	1593	1569	1538			
		总密度	414	715	705	1304	3980	4374	3348	3278	3276	2705	2115	5717	1430	5439	2955	4654	2552			
		类符号	#	\$	/	1	&	[%	=	⊙	*	+	(-	~	>	:	.			
		类序号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17			
监督分类		类序号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
		类符号	#	\$	/	1	&	[;	%	=	⊙	*	+	(~	>	:	x	.	↑	
		点数	394	478	577	1619	2223	1698	1918	6027	2832	3128	1044	2755	6455	1033	4320	2838	2209	2258	936	
		总密度	2088	1912	1805	1762	1725	1702	1693	1666	1664	1649	1638	1629	1618	1606	1598	1576	1568	1549	1529	
		5 波段密度	498	454	439	430	428	417	422	414	411	407	399	403	399	390	393	386	386	380	374	
距离		176	107	43	37	43	37	25	33	27	26	27	39	27	19	26	34	19	31	10	19	20

表 4 监督分类各类土壤景观(各谱类)的斜率比 K 及其分组(型)
Table 4 The slope ratios K and their grouping (types) of the soil or landscape categories (the spectral categories) in the supervised classification

类序号	类符号	K	根据 K 值划分的型																				
			下凹型								平直型								微凸型				特凸型
2	\$	0.47	3	8	1	5	7	4	9	16	6	13	12	19	11	15	20	14	10	18	17		
			/	%	#	0.57	0.57	0.58	1.00	1.00	1.00	1.07	1.08	1.15	1.61	1.69	1.85	2.08	2.16	2.20	2.20	9.09	
			1577	6027	394	2223	1918	1619	2832	3279	1698	6455	2755	2258	1044	4320	936	1033	3128	2209	2838	2838	

射纲($R_s < 0.35$):主要是水和沼泽。有趣的是,它们的 R_7 均低于 R_4 , 为其它纲所未见。
 (II) 低辐射纲 ($0.35 < R_s < 0.51$): 主要是地面覆盖着较好植被的土壤。是沙漠中的绿洲。
 (III) 中辐射纲 ($0.51 < R_s < 0.66$): 部分地面为植被覆盖、部分地面裸露的土壤。裸露部分为砂、盐或粘土状物质。
 (IV) 高辐射纲 ($R_s > 0.66$): 地面基本裸露的土壤或景观: 砂、盐或粘土状物质(图 3)。

根据其它波段的辐射值,基本上也能分出和上面相同的 4 大纲,只是以 R_s 划分得最好。分辨土壤和景观大类的能力顺序是 $R_5 > R_7 > R_4 > R_6$ 。

某些谱类波段辐射曲线所代表的地面实况见图版 I 照片 1—6。

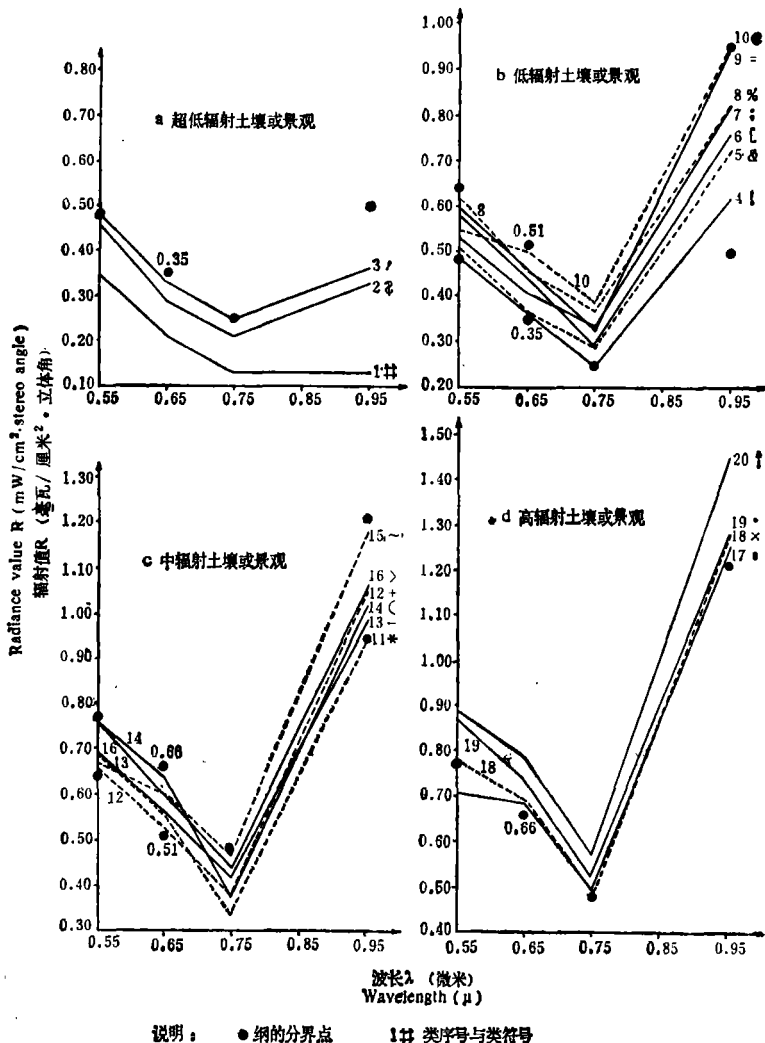


图 3 监督分类各类土壤或景观(20个谱类)的波段辐射曲线

Fig. 3 The band-radiation curves of the soil or landscape categories (20 spectral categories) in the supervised classification

1) 与文献[2]中不同,这里的 R 值因只涉及一幅图象,故未经相邻两幅图象重叠测量以进行灰标等级校正。

(四) 谱类曲线的斜率比系数

上述波段辐射曲线中,除了水的曲线(符号“#”)大致呈幂函数曲线 $y = ax^b$ 的形式外,其它曲线都大致呈向上开口的二次曲线 $y = ax^2 + bx + c$ 的形式。如果我们仅观察从 4 波段到 6 波段这一段由两个线段组成的折线,那么显然有些呈平直型,有些呈下凹型,有些呈上凸型。我们定义一个斜率比系数:

$$K = \frac{R_5 - R_6}{R_4 - R_5}$$

来定量地描述这个现象。算得的监督分类 20 类的 K 值见表 4。

根据 K 值,可把曲线分为 4 组(型): (I) 下凹型 ($K = 0.47 - 0.58$)。主要是植物茂密的耕地土壤和天然森林或灌木林下的土壤,还包括水和沼泽。它们的 R_5 相对较低。(II) 平直型 ($K = 1.00 - 1.15$)。地面部分为植被覆盖、部分裸露的各种土壤。它们从 R_4 到 R_5 到 R_6 几乎呈直线降低。是一种过渡类型。(III) 微凸型 ($K = 1.61 - 2.20$)。主要是流沙、盐霜、粘土等覆盖地面而植物很稀疏的各种土壤。它们的 R_5 相对较高,但 R_4 也不低。(IV) 特凸型 ($K = 9.09$)。只有龟裂性土和典型的龟裂状残余盐土(二者分不开)才属此类。文献[2]中的龟裂性土和所谓的荒漠化草甸土(粘土状平坦地面)与此类似。它们的 R_5 相对较高而 R_4 较低。

我们可以说,下凹型是植被和水面辐射的特征,微凸型是砂、盐辐射的特征,特凸型是平坦粘质地面(特别是带红色时)的辐射特征,平直型是过渡状态地面的辐射特征。不同 K 值是这些地面构成要素以不同比例配合的结果。

每类地物(土壤或景观)的 K 值也是根据几百到几千个个体统计出来的,因而也可以作为这些地物的代表性 K 值。

(五) 光谱学分析的意义

光谱学分析在土壤类型识别中起着重要作用。例如在非监督分类结果中有谱类“\$”,这是分类前未料到的,根据谱类辐射曲线,它处在典型水体的曲线“#”和沼泽曲线“/”之间,由此可知它是典型水体与沼泽间的一种过渡类型。联系实地情况,我们将它定为间有植被或陆地出现的水体(草湖)。又如监督分类中的谱类“C”,根据波段辐射曲线、斜率比和群分析,可知它处于中辐射纲中,并接近于中辐射纲的上界和高辐射纲的下界,曲线属于上凸型,土壤接近于龟裂状残余盐土而又带有一定程度的砂性。联系实地情况,我们将它定为砂性龟裂状残余盐土。

光谱学分析在土壤类型细分和谱类符号组合意义的确定上也起着重要作用。例如典型盐土可细分为 5—6 种小类型,而各小类的分界线基本上对应于系统聚类中的第 5 级划分。又如“%”和“—”两个谱类按总密度或总辐射水平而言十分相近,但二者辐射的波段分配不同,且谱类“%”按斜率比属下凹型,说明它以植被覆盖为主;谱类“—”属于平直型,说明地面裸露面积较大。在片状绿洲中,“%”主要是绿洲潮土 II,“—”主要是典型盐土 18b。故当“%”和“—”形成双符号组合时,若以“%”为主,则定为夹有典型盐土的绿洲潮土 III,若以“—”为主,则仍定为典型盐土 18b。

光谱学分析在谱类分组和土壤区域划分中也起着重要作用。例如我们由群分析结果、波段辐射曲线以及谱类内在序列的排序,可以确定绿洲和荒漠的地域界线以及片状绿洲和线状绿洲的地域界线。这对于后面要谈的地理控制区的建立是非常重要的。

总之,熟悉一个地区各种土壤和景观所属的谱类,认识这些谱类的内在成因联系和它们的波段辐射曲线及其参数的特征,掌握这些谱类的组合、分组规律,对于该地区地物的自动识别分类和其它遥感工作,是很有帮助的。

四、地理控制系统及其意义

(一) 地理控制系统

怎样解决同一谱类符号代表不同土壤单元的问题呢?我们在分析两种自动识别图时都发现,同谱类的不同土壤或景观,往往分处于互不相同的特定地理环境中。例如,谱类“#”和“\$”可以代表水和草湖。又可代表山地阴影;但前二者出现在绿洲中和荒漠平原上,后者则出现在山地中。谱类“;”可以代表胡杨林土,又可以代表红柳林盐土,但前者一般出现在荒漠平原上的走廊式绿洲中,后者多半出现在人类耕垦灌溉活动激烈的片状绿洲中。又如谱类“!”在片状绿洲中可以代表淡草甸土,又可代表沼泽盐土和盐化沼泽化草甸土,但前者在河流两岸,后者不在河流两岸。如此等等。

由此可以设想,如果我们能划出若干地理界线,确定若干特定的地理区域(如山地、荒漠平原、片状绿洲、线状绿洲等)或地理环境(如河流、非河岸等),我们就能赋予相同的谱类符号以不同的土壤学意义。

从整个分类区考虑,我们先划分出四个大的地理-土壤区域(我们叫它控制区):(I)片状绿洲区;(II)冲积平原区;(III)洪积平原区;(IV)山地及山前洪积扇区。

在控制区之下,或划分控制亚区;或考虑其它控制因素和环境:(1)是否在河流或大干渠两岸;(2)是否与某种地形部位相联系;(3)是否与某种景观相联系;(4)是否与某种土壤共轭;(5)大面积连续分布或小面积零星分布;等等。

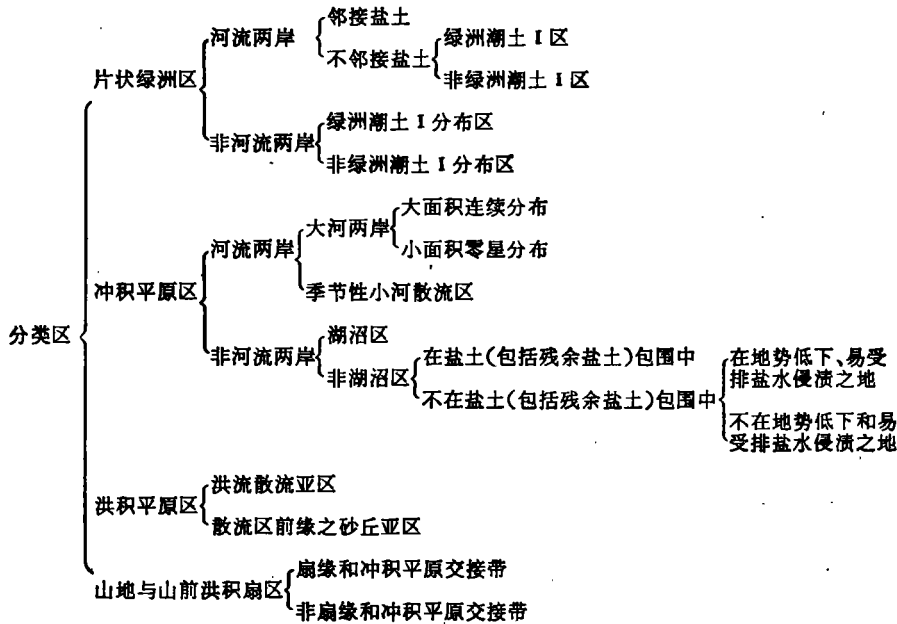
考虑到上述种种因素之后,我们就可设计出一个将同一谱类分为不同土壤或景观的地理控制系统(表5)。

(二) 地理控制系统的理论根据和意义

地理控制系统的理论根据是土壤和景观是一个统一整体的学说。二者密切不可分割。这一点反映在遥感图象上,就是土壤信息往往混迹于其它地理要素的信息之中,而以景观信息或地理综合体的信息表现出来。不同的土壤有可能在各种地理要素的配合下表现出相同的谱特征。但是,因为一定的土壤只能出现在一定的地理环境中,而不出现在与它极不相干的另一地理环境中,我们就不能仅根据谱特征相同就把处在极不相同的地理环境中的不同土壤类型划为同一土壤类型。要划分它们,单凭遥感信息是作不到的,必须使用非遥感的地理学或土壤学参考信息。这就有必要先划分地理环境或建立地理控制系统。这种系统建立得正确与否,依赖于对地形、母质、植被、水文以及人类经济活动等成土因素的分异规律研究的深度,是一个地理-土壤学问题。有了合适的地理控制系统,我们就能

有效地对遥感数值分类结果进行专题解释和勾绘专题界线。

表 5 地理控制系统表
Table 5 A geographical control system



进行了上述谱类分析和地理分析之后，我们对于计算机划分出来的类别与实地土壤类型的关系就比较清楚了。我们可以整理出非监督分类和监督分类的谱类符号与它们所代表的土壤类型关系表，据此可在计算机输出的数字分类图象上勾绘土壤界线，作出正式的土壤图。

从我们的结果来看，用本文所述的方法编出来的土壤图，比常规方法编出的土壤图要详细、精确得多。故这种方法是可行的，是能够在中、小比例尺土壤清查和制图中加以采用的。

因限于篇幅，计算机输出的非监督分类图和监督分类图连同勾绘出来的土壤界线，据此作出的两张 1:40 万土壤图和它们与现有土壤图的比较，以及以卫星图象计算机自动识别分类为基础编制中、小比例尺土壤图的可行性估计，将另文发表¹⁾。

参 考 文 献

[1] 中国科学院地质研究所编著，1977：数学地质引论。地质出版社。
 [2] 曾志远，1981：卫星图象密度-辐射量算与土壤资源探测。自然资源，第 4 期，52--66 页。
 [3] Swain, P. H., Davic, S. M., 1978: Remote Sensing: The Quantitative Approach, McGraw-Hill, Inc.,

1) 曾志远：卫星图象土壤类型自动识别与制图的研究 II. 自动识别结果的成图及其与常规土壤图的比较，将在本刊另期发表。

AUTOMATIC RECOGNITION AND MAPPING OF SOIL TYPES USING LANDSAT IMAGES

I. COMPUTER CLASSIFICATION AND THE SPECTROSCOPICAL AND GEOGRAPHICAL ANALYSES OF THE RESULTS

Zeng Zhiyuan

(Institute of Soil Science, Academia Sinica, Nanjing)

Summary

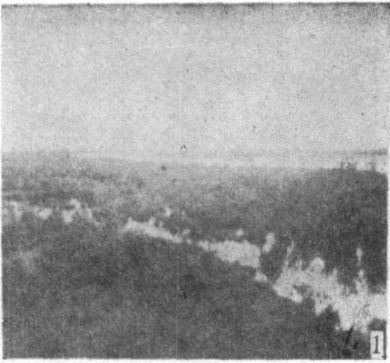
Unsupervised and supervised classification of soils of an arid area in Xingjiang was made using landsat data which were derived from scanning of the images on a transparent negative hard-copy and processed by micro-computer. The map compiled from these data is on a scale of 1:400,000 and covers an area of 12,375 km².

Before classification, the scanned data of the four channels were matched by columns and rows according to some special topographic features on the image. All anomalous data were rectified by the methods designed by the author.

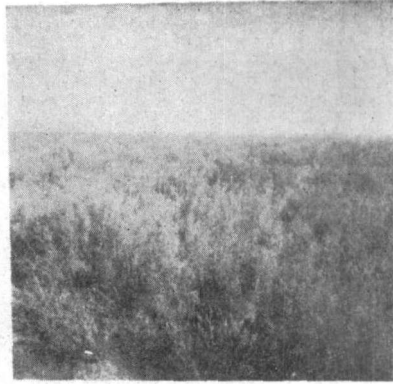
For the unsupervised classification, the key tactics adopted by the author was firstly to find out and fixed the suitable number for the division of soil groups in this region, and then to select the proper parameters in order that the iteration could be continued freely until satisfied results were obtained.

For the supervised classification, training samples were selected and soil groups were stipulated as many as possible. Then, according to the classified map, the discrimination ratio of the training samples, the histograms of the training data of various soil groups, etc. obtained from the first classification, the number of groups was revised and the training samples were modified by deletion, split or merge for the formal classification.

The data obtained from the methods mentioned above were further analysed spectroscopically (the internal sequence, the dendrogram, the band radiation curves and their slope ratios of the spectral categories) and geographically (a geographical control system), and on the basis of these analytical data the soil map was compiled.



1. 谱类“/”：沼泽(左边中部和右半部)
谱类“1”：盐化沼泽化草甸土 13_a(左下角)



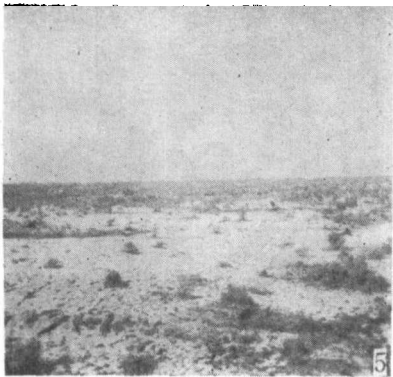
2. 谱类“;”：植物较密的红柳林盐土 17_b



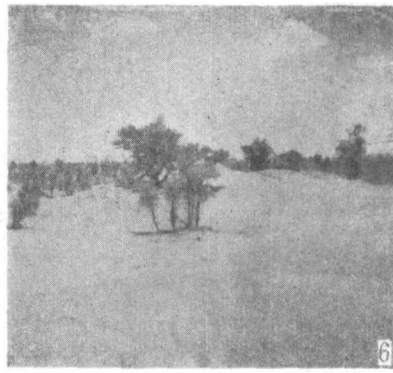
3. 谱类“=”：轻度荒漠化胡杨林土 12_a(上部)
谱类“#”：开放的水面 W₁(左边中下部)



4. 谱类“+”：荒漠化胡杨林盐土 19₂



5. 谱类“×”：植物极稀疏的复盐霜典型盐土 18_a



6. 谱类“↑”：流动砂丘地 30_a(近景)
谱类“@”：强度荒漠化胡杨林土 12_c(远景)

注：1、2、4、5 摄于本研究区；3、6 摄于本区之外，但能代表本区的谱类

照片 1 某些谱类的地面实况代表

Photo. 1 The representative ground conditions of some spectral categories