

主组元分析在分辨土壤类型及 风化—成土过程上的应用*

(以水稻土为例)

刘多森

(中国科学院南京土壤研究所)

数学向土壤学的渗透,是四个现代化的需要。随着仪器分析自动化程度的提高,我们在土壤学领域中取得的数据必将大量积累。为了充分利用大量的土壤学数据提供的信息,以深刻认识成土过程,准确区分土壤的发生学类型,就有必要运用有关多元分析的数学方法。多元分析的计算工作量,随元数的增加而成几何级数增加。电子计算机的逐步普及,则为充分运用多元分析解决土壤学问题提供了实际可能。

主组元分析(Анализ главных компонент)是多元分析的方法之一。数学研究者对这一方法有详尽论述¹⁾(Rao, 1973)。主组元分析在科学技术各个领域的应用相当广泛。七十年代以来,国际上较多地用主组元分析解决土壤分类问题²⁾(Cuanalo et al., 1970; Рожков, 1975)。

土壤剖面可由反映各发生层的物理性质、化学性质、生物学性质的多种指标(元)表示。由于风化—成土过程的影响,在不同指标之间往往存在着程度不等的相关性,其中若干对指标的相关性甚至达到显著性水准。土壤指标之间的相关性,使它们提供的土壤学信息大部分发生重叠。这就有可能用数学方法找出新的、个数比原指标少得多的综合指标,并使这些综合指标彼此无相关性,而又反映了原来较多的指标所提供信息的主要部分。这些综合指标就是原指标的主组元。第 t 主组元值 y_t 是土壤的各个指标值 x_i 的线性组合:

$$y_t = \sum_{i=1}^m \zeta_{it} x_i$$

其中 i 为土壤指标的编号, m 为土壤指标的总个数, ζ_{it} 为第 i 指标对于第 t 主组元的权系数。

主组元分析也称为主成分分析或主分量分析。为了避免与土壤的“成分”(化学成分、物理成分……)及计算过程中各种向量的“分量”相混淆,所以我们没有采用后二种名词。

* 本文第一部分经南京大学数学系唐述钊、夏定中老师审阅,谨此致谢。本文曾在1978年5月中国土壤学会理事暨土壤分类学术交流会上提出,这次刊载前进行了修改。

1) 方开泰,1974: 主成分分析。中国科学院数学研究所编,多元分析资料汇编,(1),5—40页。

2) Kyuma, K. 等: 日本水稻土的土壤物质分类。陈志诚译自 Soil Sci. Plant Nutr., 22(2): 111—124, 1976。

一、土壤学中主组元分析的一般方法

(一) 土壤指标值(原始数据)的整理

假设所研究的土壤剖面的总个数为 n , 每一土壤剖面经野外观测及实验室分析均取得 m 个具有土壤发生学意义的指标, 则第 i 个土壤剖面的 m 个指标值可表示成一个 m 维随机行向量 x_i :

$$x_i = (x_{i1}, x_{i2}, \dots, x_{im})$$

该向量的维数(指标数) m 等于土壤各发生层测定项目(土壤的各种性质及各发生层厚度均可作为测定项目)数之和。例如, 各土壤剖面均取三个发生层 A 、 B 、 C , 而 A 层、 B 层、 C 层均分别测定 10 个、8 个、4 个项目, 则 $m = 22$ 。

n 个土壤剖面共有 $n \times m$ 个指标值。它们可表示成一个 $n \times m$ 矩阵 X :

$$X = (x_i) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix} = (x_{ij})$$

$$(i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m)$$

其中 i 为土壤剖面的编号; j 为土壤指标的编号。

(二) 求土壤指标值的标准化值 x_{ij}^* , 矩阵 X^* 及土壤指标的相关阵 R

土壤的不同指标往往有不同的量纲。例如腐殖质含量、容重、发生层厚度的量纲彼此殊异。即使同一指标, 也可人为地采用不同量纲。例如, 发生层厚度可用米、厘米、毫米表示而分别得到迥然不同的数字。为了消除因量纲不同而带来的不合理影响, 则应使指标值, 即 $n \times m$ 矩阵 X 的元素标准化。标准化有不同方法, 我们采用标准差标准化方法, 即对 x_{ij} 作如下变换:

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{s_j}$$

$$(i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m)$$

其中 s_j 为阵 X 第 j 列元素的标准差, 即土壤的第 j 指标的标准差:

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

则得 $n \times m$ 矩阵 X^* :

$$X^* = (x_{ij}^*)$$

阵 X 的列向量经标准化后, 其 m 阶协方差阵 V 的元素 v_{il} 变换为:

$$v_{il} = \frac{\sum_{i=1}^n (x_{ij}^* - \bar{x}_j^*)(x_{il}^* - \bar{x}_l^*)}{n-1}$$

$$= \frac{\sum_{i=1}^n \left(\frac{x_{ij} - \bar{x}_j}{s_j} - 0 \right) \left(\frac{x_{il} - \bar{x}_l}{s_l} - 0 \right)}{n-1}$$

$$= r_{jl}$$

式中 r_{jl} 是土壤的第 j 指标与第 l 指标之间的相关系数; s_j 和 s_l 分别是土壤的第 j 指标和第 l 指标的标准差。因此, 阵 X 的列向量标准化后的 m 阶协方差阵 V 与原指标的 m 阶相关阵 R 相同。这样, 就可从阵 R 出发求主组元。我们先求出土壤指标的 m 阶相关阵 R :

$$R = (r_{jl}) = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{m1} & r_{m2} & \cdots & r_{mm} \end{pmatrix}$$

$$r_{11} = r_{22} = \cdots = r_{mm} = 1$$

$$r_{jl} = \frac{\sum_{i=1}^n x_{ij}^* x_{il}^*}{n-1}$$

$$(j, l = 1, 2, \dots, m)$$

显然, 阵 X^* 的转置阵 $X^{*'}$ 与阵 X^* 之积同阵 R 有如下关系:

$$X^{*'} X^* = (n-1)R$$

(三) 求土壤指标相关阵 R 的特征值 λ_t , 特征向量 ξ_t , 正规化特征向量 ζ_t 及正交阵 Z

在有关数学书中, 对于方阵的特征值和特征向量的各种解法已有详尽论说, 这里不再赘述。但为了有助于解决土壤学的实际问题, 我们将提及若干基本概念。

m 阶相关阵 R 的特征方程

$$|\lambda E - R| = 0$$

的 m 个根 λ_t , ($t = 1, 2, \dots, m$), 即相关阵 R 的特征值。 E 为 m 阶单位阵。

m 阶相关阵的特征值 λ_t , 按其大小排列为:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$$

式中等号表示可能出现重特征根。但在土壤学的实际问题中, 相关阵的特征方程出现重根的情况是十分罕见的。同时, m 阶相关阵的各特征值有如下重要关系:

$$\lambda_1 + \lambda_2 + \cdots + \lambda_m = m$$

相关阵为非负定的实对称阵, 故 m 阶相关阵的特征值 λ_t 必为 m 个(包括重根)非负实数; 而相应的特征向量 ξ_t , 不管是否出现重特征根, 必可取为 m 个实的非零正交向量 $\xi_1, \xi_2, \dots, \xi_m$ 。

特征向量并不是一个完全确定的向量。令 a 为一常数, 则 $a\xi_t$ 仍为属于 λ_t 的特征向量。将各特征向量 ξ_t 正规化, 得相应的正规化特征向量 ζ_t 。视为列向量, 则有:

$$\zeta_t = \frac{\xi_t}{\sqrt{\xi_t' \cdot \xi_t}}$$

$$\zeta_t' \zeta_t = 1$$

$$(t = 1, 2, \dots, m)$$

将各正规化特征向量 ζ_t 作为列向量, 则可在 m 维线性空间构成 m 阶正交阵 Z :

$$Z = (\zeta_{jt}) = (\zeta_{j1}, \zeta_{j2}, \dots, \zeta_{jm})$$

$$(j, t = 1, 2, \dots, m)$$

$$Z'Z = ZZ' = E$$

(四) 求第 i 个土壤剖面的第 t 主组元 y_{it} 及主组元矩阵 Y

$$y_{it} = \sum_{j=1}^m x_{ij}^* \cdot \zeta_{jt}$$

$$= (x_{i1}^*, x_{i2}^*, \dots, x_{im}^*) \begin{pmatrix} \zeta_{1t} \\ \zeta_{2t} \\ \vdots \\ \zeta_{mt} \end{pmatrix}$$

$$(i = 1, 2, \dots, n; t = 1, 2, \dots, m)$$

以主组元 y_{it} 为元素, 构成 $n \times m$ 主组元矩阵 Y :

$$Y = X^*Z = (y_{it})$$

$$= \begin{pmatrix} y_{11} & y_{12} & \dots & y_{1m} \\ y_{21} & y_{22} & \dots & y_{2m} \\ \dots & \dots & \dots & \dots \\ y_{n1} & y_{n2} & \dots & y_{nm} \end{pmatrix}$$

$$(i = 1, 2, \dots, n; t = 1, 2, \dots, m)$$

阵 Y 每一列的 n 个元素的算术平均值 \bar{y}_t 为零。

相关阵经其特征向量阵可化为对角阵, 且此对角阵的主对角线元素为相关阵的特征值。因此, 主组元的协方差阵 V 等于对角阵 Λ ——

$$\begin{aligned} V &= \frac{1}{n-1} Y'Y \\ &= \frac{1}{n-1} (X^*Z)'(X^*Z) \\ &= \frac{1}{n-1} Z'X^{**}X^*Z \\ &= \frac{1}{n-1} Z'(n-1)RZ \\ &= \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \dots & \\ 0 & & & \lambda_m \end{pmatrix} \\ &= \Lambda \end{aligned}$$

式中 Λ 为 m 阶对角阵, 其主对角线上的元素依次为 m 阶相关阵 R 的特征值 $\lambda_t, t = 1, 2, \dots, m$, 而其余元素均为零。

(五) 土壤剖面的分类

有代表性的土壤剖面的分类, 是作为统计学母体的全部土壤的分类估计。

第 t 特征值 λ_t 是第 t 主组元 y_t 的方差的估计量。因此, 第 t 主组元 y_t 对于总方差的贡献率为:

$$\frac{\lambda_i}{\sum_{i=1}^m \lambda_i} = \frac{\lambda_i}{m}$$

而主组元 y_1, y_2, \dots, y_k 对于总方差的累计贡献率为:

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^m \lambda_i} = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_k}{m}$$

如果说“三七开”是人们日常区分主流与支流的数量界限的话,那末我们可以人为地取前几个主组元,并使之对于总方差的累计贡献率达到 $\geq 70\%$ 。假若前二个主组元 y_1, y_2 达到了这一要求,即前二个主组元能把全部土壤指标提供的信息的70%以上反映出来,则以每个土壤剖面的第1和第2主组元 y_{i1}, y_{i2} ($i = 1, 2, \dots, n$),在 y_1, y_2 为横、纵坐标的二维平面上点图时,同一土壤类型的不同剖面的点相当充分地聚集于平面的同一围线之内,不同土壤类型的剖面的点相当充分地聚集于不同围线内而彼此分开。每一围线的重心,可作为该围线土壤类型的典型剖面。

由于第1和第2主组元只是反映了主要信息,而不是全部信息,所以主组元分析用于土壤分类并不是总有成效的。特别当选取的土壤指标不当,各土壤指标缺乏共同的制约因素,各特征值差别不大时,土壤剖面的类型在第1和第2主组元坐标图上的分辨将不明显。

(六) 土壤指标的分类与主要风化—成土过程的指标组合

如果前二个主组元对于总方差的贡献率在70%以上,则以第1和第2正规化特征向量 ζ_1, ζ_2 (或特征向量 ξ_1, ξ_2) 的各对应分量 ζ_{j1}, ζ_{j2} ($j = 1, 2, \dots, m$),在 ζ_1, ζ_2 (或 ξ_1, ξ_2) 为横、纵坐标的二维平面上点图时,性质近似的指标的点相当充分地聚集于平面的同一围线之内,性质不同的指标的点相当充分地聚集于不同围线内而彼此分开。这样,我们就能把土壤的 m 个指标按其性质分类。

各个主组元之间彼此无相关关系,它们各自反映了彼此不同的、独立的风化—成土过程。前几个主组元,则反映在所取指标提供的全部信息内的主要风化—成土过程。主组元反映的风化—成土过程,是被研究的所有土壤剖面共同的过程。正因为各土壤剖面在这几个共同的主要风化—成土过程上存在差异,才能把它们按主组元区分为不同的土壤类型。

主组元的具体物理意义,要结合土壤学知识加以理解。各指标对于第 k 主组元 y_k 的权系数,即第 k 正规化特征向量 ζ_k 对应于第 j 指标的分量 ζ_{jk} ($j = 1, 2, \dots, m$),是第 j 指标对第 k 主组元的负荷的相对大小和作用方向的反映。选出绝对值最大的几个 ζ_{jk} ,则与它们相应的指标的组合,将反映同第 k 主组元相应的风化—成土过程的主要特征;而各个 ζ_{jk} 中的正分量与负分量,则表示相应的指标在第 k 主组元代表的风化—成土过程中的作用是相逆的。

(七) 指标的选择与主组元分析在土壤学中的其他应用

主组元分析作为一种数学方法,可以应用于科学技术的各个领域;而在土壤学领域中

的应用,也不限于土壤分类和主要风化—成土过程的确定。

在运用主组元分析分辨土壤的发生学类型及主要风化—成土过程时,被选择的土壤指标必须是具有土壤发生学意义的指标。

主组元反映了在所取指标范围内提供的全部信息的主要信息。在解决不同的土壤学问题时,土壤指标的选择原则应当有所不同。如果只选择土壤的 SiO_2 、 Fe_2O_3 、 Al_2O_3 含量作为指标,其结果将可能分辨土壤母质类型及相应的化学风化过程。为了解决污染和环境保护问题,我们选择某污染地区各有害元素在土壤中的含量作为指标,就可能分辨土壤的污染类型及相应的污染过程。为了研究生物—土壤—水—大气生态系统,我们选择化学元素在不同空间的含量作为指标,就可能分辨生态系统的类型及相应的生态过程。如果用土壤遥感获得的光学数据为指标,也可以分辨土壤类型;但如何定量地理解光学数据的土壤学意义,是一项复杂的工作。

二、主组元分析在分辨水稻土上的应用¹⁾

选取广东、广西、云南、江西、湖南、湖北、安徽、江苏水稻土 82 个,分析耕层如下项目: 1. 砂粒(> 0.05 毫米)含量(%), 2. 粗粉粒(0.01—0.05 毫米)含量(%), 3. 粘粒(< 0.001 毫米)含量(%), 4. 有机质含量(%), 5. 全氮含量(%), 6. 结构系数, 7. 浸水容重(克/厘米³)。结构系数用 $(1 - \frac{b}{a})100\%$ 表示,其中 a 为机械分析得出的 < 0.001 毫米粒级的含量, b 为微团聚体分析得出的 < 0.001 毫米粒级的含量。选取耕层的上述分析项目作为土壤指标,且按上述顺序对土壤指标进行编号。

各指标在 82 个水稻土中的算术平均值(\bar{x}_j)、标准差(s_j)及变差系数(CV),列于表 1。从表 1 看出,砂粒含量、有机质含量的分布比较离散,浸水容重、结构系数的分布比较集聚,而其余指标值分布的离散程度则居于前二组之间。

表 1 82 个水稻土各指标的均值(\bar{x}_j)、标准差 s_j 及变差系数(CV)

指 标	1. 砂粒(%)	2. 粗粉粒(%)	3. 粘粒(%)	4. 有机质(%)	5. 全氮(%)	6. 结构系数(%)	7. 浸水容重(克/厘米 ³)
\bar{x}_j	15.4	30.2	21.8	2.93	0.157	69.1	0.60
s_j	19.88	12.40	10.27	2.889	0.1152	19.41	0.106
CV	1.29	0.411	0.471	0.986	0.734	0.281	0.18

(一) 水稻土指标的相关阵及其启示

水稻土指标的相关阵用表格形式表示,即表 2。从表 2 看出:

1. 水稻土耕层砂粒含量同粘粒含量、粗粉粒含量成负相关,同浸水容重成正相关,且其相关性均非常显著;砂粒含量同全氮含量、有机质含量成负相关,其相关性是显著的;砂粒含量同结构系数的关系不明显。

1) 所用水稻土 7 项指标的原始数据均由龚子同同志提供。

表2 水稻土指标相关阵(自由度=80) $r_{0.05} = 0.217$ $r_{0.001} = 0.357$

相关系数 指标	1.砂粒含量	2.粗粉粒含量	3.粘粒含量	4.有机质含量	5.全氮含量	6.结构系数	7.浸水容重
1	1	-0.500**	-0.589**	-0.239*	-0.279*	0.074	0.515**
2		1	-0.257*	-0.077	-0.078	0.312*	0.172
3			1	0.368**	0.412**	-0.144	-0.580**
4				1	0.965**	0.123	-0.549**
5					1	0.120	-0.601**
6						1	0.405**
7							1

* 在概率 0.05 的显著性水准上,可以接受相应指标间相关的估计;

** 在概率 0.001 的显著性水准上,可以接受相应指标间相关的估计。

2. 粗粉粒含量同粘粒含量成负相关,同结构系数成正相关,其相关性都是显著的;粗粉粒含量同有机质含量、全氮含量、浸水容重的关系均不明显。

3. 粘粒含量同全氮含量、有机质含量成正相关,同浸水容重成负相关,且其相关性均非常显著;粘粒含量同结构系数的关系不明显。

4. 有机质含量同全氮含量成正相关,同浸水容重成负相关,且其相关性均非常显著;有机质含量同结构系数的关系不明显。

5. 全氮含量同浸水容重成负相关,且其相关性非常显著;全氮含量同结构系数的关系不明显。

6. 结构系数同浸水容重成正相关,且其相关性非常显著。

(二) 水稻土风化—成土过程的分辨

水稻土指标相关阵的前三个特征值及主组元的贡献率列于表 3。相应于前三个特征值的正规化特征向量,用表格形式表示,即表 4 的数据部分。

水稻土在所取 7 项指标提供的全部信息范围内的各个风化—成土过程,可由各个主组元的权系数加以分辨;其主要过程,则由前三个主组元的权系数进行分辨。

表 3 水稻土前三个主组元的贡献率

主 组 元	特征值 λ_i	贡献率 (%)	累计贡献率 (%)
1	3.084	44.06	44.06
2	1.563	22.33	66.39
3	1.354	19.34	85.73

第 1 主组元的物理意义,是水稻土含氮有机质形成过程的量度。全氮含量、有机质含量、粘粒含量对第 1 主组元具有较强的正向负荷,而浸水容重、砂粒含量则具有较强的逆向负荷。粘粒可与有机质形成有机-无机复合胶体,粘粒含量高时有利于有机质的保持和积累,因而粘粒含量对第 1 主组元反映的含氮有机质形成过程具有较强的正向负荷。相反,砂粒含量及浸水容重高时,则不利于含氮有机质形成过程的发展,即其对第 1 主组元显示较强的逆向负荷。

第 2 主组元是水稻土在含氮有机质形成过程强度大体相同条件下粗粉粒团聚化过程

的量度。粗粉粒含量、结构系数对第 2 主组元具有较强的正向负荷,而砂粒含量则具有较强的逆向负荷。粗粉粒含量同结构系数之间的相关系数 r_{26} 为 0.312,砂粒含量同结构系数之间的相关系数 r_{16} 为 0.074 (表 2),且有:

$$r_{26} = 0.312 > 0.283 = r_{0.01}$$

$$r_{16} = 0.074 < 0.0756 = r_{0.5}$$

其中 $r_{0.01}$ 和 $r_{0.5}$ 分别为在概率 0.01 和 0.5 时相关系数的临界值(自由度 = 80)。这说明,在水稻土粘粒参与形成水稳性团聚体的过程中,粗粉粒团聚化作用要比砂粒团聚化作用

表 4 水稻土前三个主组元的权系数及主要指标组合

指标		主组元		
		1	2	3
1. 砂粒含量		-0.348	-0.365	0.532
2. 粗粉粒含量		-0.0579	0.709	-0.283
3. 粘粒含量		0.425	-0.118	-0.235
4. 有机质含量		0.464	0.107	0.445
5. 全氮含量		0.483	0.106	0.402
6. 结构系数		-0.0900	0.548	0.448
7. 浸水容重		-0.488	0.179	0.180
主要指标组合	正向指标	全氮, 有机质, 粘粒	粗粉粒, 结构系数	砂粒, 结构系数, 有机质, 全氮
	逆向指标	浸水容重, 砂粒	砂粒	

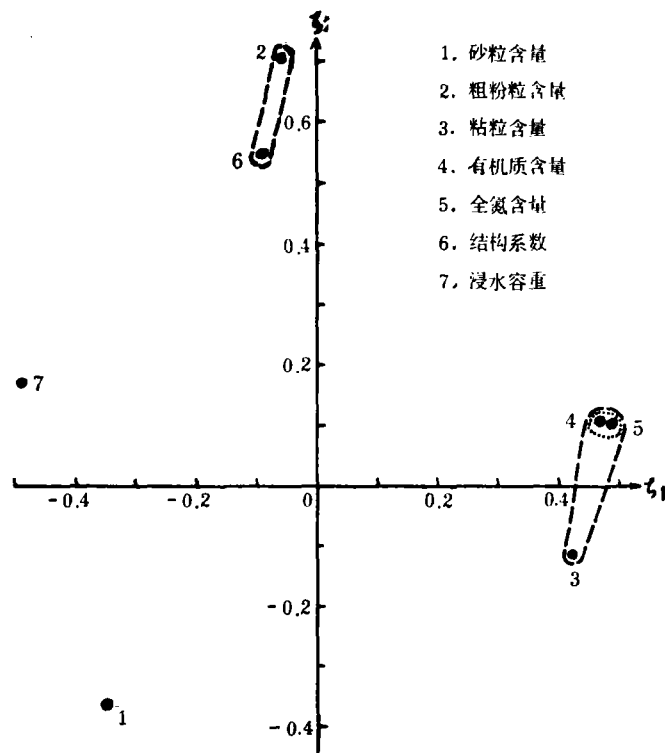


图 1 水稻土 7 项指标按第 1 和第 2 正规化特征向量(ζ_1, ζ_2)分类

更为强烈。

第3主组元是水稻土在含氮有机质形成过程、粗粉粒团聚化过程强度大体相同条件下的砂粒(可以推论,主要是细砂粒)团聚化过程的量度。砂粒含量、结构系数、有机质含量、全氮含量对第3主组元具有较强的正向负荷。有机质和全氮含量对第3主组元的负荷显著大于对第2主组元的负荷,说明砂粒团聚化比粗粉粒团聚化需要更多的含氮有机质。

(三) 水稻土各项指标的类型分辨

取表4第1、2列的7对数据,即 ζ_1 和 ζ_2 的对应分量,在 ζ_1 和 ζ_2 为横、纵坐标的二维平面上点图(图1)时,有机质含量和全氮含量的点非常接近,故可将其划为性质极为近似的一类指标。有机质含量、全氮含量又可同粘粒含量进一步划为一类;这三个指标划为同一类指标的物理意义,是它们在水稻土含氮有机质形成过程中均具有显著的正向作用。粗粉粒含量、结构系数也可划为一类指标;其物理意义是二者在水稻土粗粉粒团聚化过程中均具有显著的正向作用。砂粒含量、浸水容重则分别视为不同类型的指标。

(四) 水稻土的类型分辨

第1和第3主组元贡献率之和为 63.40%, 略小于前二个主组元的累计贡献率 66.39%。但因第3主组元反映的砂粒团聚化过程比第2主组元反映的粗粉粒团聚化过程更难进行,因而更具有特征性,砂粒含量分布比粗粉粒含量分布更加离散(表1),即第3主组元比第2主组元的土壤学意义更大,故用第1和第3主组元分辨水稻土类型将比用前二个主组元更为优越。当然,用前三个主组元综合分辨水稻土类型是最理想的,但三维空间不如二维平面直观。

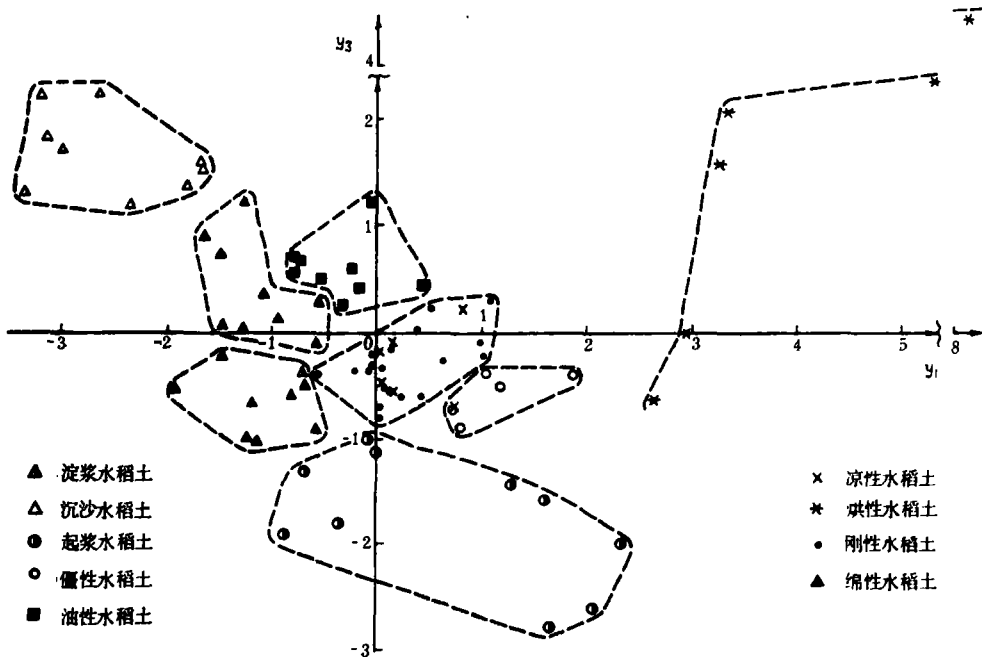


图2 82个水稻土按第1和第3主组元(y_1, y_3)分类

用每个水稻土的第 1 和第 3 主组元 y_{1i} 和 y_{3i} 在 y_1, y_3 为横、纵坐标的二维平面上点图(图 2)时,除 6 个凉性水稻土外,其余 76 个水稻土则按其类型分别聚集于 8 个围线之内而彼此分开。这 8 个类型是: 淀浆、沉沙、起浆、僵性、油性、烘性、刚性、绵性水稻土¹⁾(龚子同, 1979)。由于未选取说明水分状况的指标,故对凉性水稻土不能加以分辨。6 个凉性水稻土在图 2 中分别落入僵性、刚性水稻土围线之内,说明这些凉性水稻土在施用排水措施之后将演化为僵性、刚性水稻土类型。

图 2 启示,淀浆、沉沙水稻土经过绵性水稻土可演化为油性水稻土,起浆、僵性水稻土经过刚性水稻土也可演化为油性水稻土。主组元分析说明,低肥力的各种水稻土可以定向培育成高肥力的油性水稻土。烘性水稻土,主要指雷文进等(1964)记述的“烘土”,其特点是: 全年淹水,浮泥层深厚,有机质含量很高,胀缩性能很大。烘性水稻土在排水之后可经过凉性水稻土演化为油性水稻土。

图 2 说明,高肥力的油性水稻土,是在第 1 主组元 y_1 处于闭区间 $[-0.77, 0.44]$ 条件下,第 3 主组元最高的一类水稻土。换言之,油性水稻土是在含氮有机质形成过程强度中等的条件下,砂粒团聚化过程强度最高的一类水稻土。由此可见,高肥力的土壤,并不是含氮有机质形成过程强度最高的土壤,亦即不是全氮和有机质含量最高的土壤。高度的土壤肥力,是土壤有关指标适当配合的综合反映。油性水稻土则表现为土壤颗粒各粒级含量适当,全氮、有机质含量适当,浸水容重与结构系数适当,而且上述指标之间的配合适当。

(五) 未知类型的水稻土的类型分辨

根据表 1 及表 4 数据,可建立主组元方程。

第 1 主组元(y_1) 方程为:

$$y_1 = -0.0175x_1 - 0.00467x_2 + 0.0414x_3 + 0.161x_4 \\ + 4.19x_5 - 0.00464x_6 - 4.60x_7 + 1.462$$

第 3 主组元(y_3) 方程为:

$$y_3 = 0.0268x_1 - 0.0228x_2 - 0.0229x_3 + 0.154x_4 \\ + 3.49x_5 + 0.0231x_6 + 1.70x_7 - 2.837$$

其中 x_1, x_2, \dots, x_7 顺序表示表 1 中的 7 项指标,各指标的量纲与表 1 相同。主组元计算结果取至小数点后二位。

表 5 未知类型的二个水稻土的类型分辨

土号	地点	砂粒(%)	粗粉粒(%)	粘粒(%)	有机质(%)	全氮(%)	结构系数(%)	浸水容重(克/厘米 ³)	y_1	y_3	类型估计
33862	云南省曲靖县	4.5	15.5	39.3	2.89	0.171	39.5	0.50	1.64	-1.17	→起浆水稻土
33935	广西省陆川县	36.4	17.8	20.4	3.21	0.130	60.5	0.80	-1.31	0.97	绵性水稻土

表 5 列出了二个未知类型的水稻土耕层 7 项指标值。为了分辨其类型,将二个水稻土的这 7 项数据分别代入第 1 和第 3 主组元方程,即得 y_1, y_3 , 并由此在图 2 上分别得到二个水稻土的点。33862 号水稻土的点在僵性与起浆水稻土围线之间,但较接近起浆水稻

1) 将水稻土区分为上述类型的设想是龚子同同志 1978 年提出的;本文仅从数学角度证明了这一设想的科学性。

土围线,故 33862 号水稻土应为僵性与起浆水稻土的过渡类型,但较接近起浆水稻土。33935 号水稻土的点落入绵性水稻土围线之内,故应为绵性水稻土。

三、小 结

阐述了土壤学中主组元分析的数学原理。从土壤剖面各指标的协方差阵的概念出发,经过若干计算程序,最后归结为主组元的协方差阵是一对角阵,且其主对角线元素依次为土壤指标相关阵的特征值。

阐述了主组元分析用于分辨土壤指标的类型和土壤剖面的类型的一般方法。

各个主组元之间无相关关系,它们各自反映了彼此不同的、独立的风化—成土过程。前几个主组元,则反映在所取指标提供的全部信息内的主要风化—成土过程。各个主组元的权系数,是分辨各个风化—成土过程的依据。

在所取 7 项指标提供的信息范围内,水稻土的主要成土过程依次是: 1. 含氮有机质形成过程; 2. 粗粉粒团聚化过程; 3. 砂粒团聚化过程。这 7 项指标是耕层的砂粒、粗粉粒、粘粒、有机质、全氮含量(%), 结构系数(%), 浸水容重(克/厘米³)。

按第 1 和第 2 主组元的权系数,可将水稻土上述指标根据其性质分为四个类型: 1. 有机质、全氮、粘粒含量; 2. 粗粉粒含量, 结构系数; 3. 砂粒含量; 4. 浸水容重。

淀浆、沉沙、起浆、僵性、油性、烘性、刚性、绵性等八个类型 76 个水稻土,按第 1 和第 3 主组元获得了相当好的分辨。仅 6 个凉性水稻土因为没有水分状况的指标而未能分辨。

建立了第 1 和第 3 主组元方程。如将未知类型的水稻土的 7 项指标值分别代入方程,求其第 1 和第 3 主组元,就能对其类型作出估计。

参 考 文 献

- 龚子同, 1979: 水稻土耕的性及其在土壤分类上的意义。土壤学报, 第 16 卷 2 期, 85—93 页。
 雷文进、朱洪官, 1964: 江苏里下河地区的土壤和改良利用途径。土壤专刊, 36 号, 140 页。
 Cuanalo, de la C. H. E., Webster, R., 1970: A comparative study of numerical classification and ordination of soil profiles in a locality near Oxford. *J. Soil Sci.*, 21(2):340—352.
 Rao, C. R., 1973: *Linear Statistical Inference and Its Applications* (second edition). 590—593, John Wiley.
 Рожков, В. А., 1975: Метод главных компонент и его применение в почвоведении. *Почвоведение*, № 10, 141—151.

THE APPLICATION OF PRINCIPAL COMPONENT ANALYSIS ON THE IDENTIFICATION OF SOIL TYPES, WEATHERING AND SOIL FORMING PROCESS

Liu Tuo-sheng

(Nanjing Institute of Soil Science, Academia Sinica)

Summary

This article deals with the application of the mathematical method of principal component analysis in soil research.

No correlation is found among these principal components reflected the different and unique processes of their weathering and soil forming process respectively.

Using paddy soil as an example, the application of the principal component analysis on the identification of its soil forming process, the type of index and soil type was discussed.