

# 土壤数值分类模型的建立及应用\*

## ——RSCM 的建立

侯彦林

(中国科学院自然资源综合考察委员会, 100101)

### 摘 要

本文以吉林省农安县盐渍草甸草原区域土壤为例, 以多种多元统计分析方法相结合作为骨架, 在土体构型函数表达式等一系列新概念的基础上, 建立了区域土壤数值分类模型 (RSCM) 它有效地解决了土体构型的数值化, 从而建立了任何剖面相互比较的定量基础。

**关键词** 土壤数值分类, 理想剖面, 深度函数参数, 剖面综合值

近十几年来, 土壤数值分类成为土壤分类研究中的一个非常活跃的研究领域, 人们试图通过土壤数值分类的研究进一步寻求土壤定量分类的标准, 以推动土壤分类量化进程<sup>[1-2]</sup>。

土壤数值分类实质是对土壤类型的划分或对土壤某些属性的分级, 由于土壤是一个三维连续性的、具有一定土体构型和过渡性很强的历史自然体, 加之土壤剖面特性千差万别, 每个剖面中土层出现的深度、厚度及个数不同, 每个层次内土壤属性各异, 这就使得土壤整个剖面之间很难直接采用统计的方法进行比较<sup>[3]</sup>。RSCM (Regional Soil Classification Model) 正是为了解决这些问题而建立的。

## 一、区域土壤数值分类模型(RSCM)的建立

RSCM 是针对某一区域土壤定量分类, 在计算机支持下, 利用现代多元统计分析方法而建立的系统模型。

### (一) 土壤数值分类的理论基础

土壤数值分类应当坚持土壤发生学派土壤分类的基本原则和观点<sup>[4]</sup>。土壤分类的实质是如何科学地划分单个土体, 即区分单个土体在  $z$  轴方向(垂直方向)上的属性差异。前人对土体构型的理论和实践意义曾经进行过阐明<sup>[5,6]</sup>, 但还未进行数值化试尝。作者从单个土体和土体构型的概念出发, 将单个土体沿  $z$  轴方向上的特性 (以下简称土体构型, 即指单个土体在分类控制层断内的土体特征) 表示成如下函数关系:

$$P = \sum_{k=1}^m f(D_k, I_k, V_k) \quad (1)$$

\* 本研究得到肖笃宁研究员、刘兆荣教授的帮助; 龚子同研究员审阅了本文, 在此一并致谢。

式中,  $D_k$ 、 $I_k$ 、 $V_k$  分别代表土体构型从地表开始向下第  $k$  个发生层次出现的深度上限、厚度、属性 (包括定量和定性属性);  $m$  为土体构型在分类控制层段内的所有层次数。公式 (1) 的含义是土体构型主要取决于发生层的深度、厚度、属性及层次的配置方式, 它是土体构型数值化和进行土壤数值分类的理论基础。

**(二) RSCM 的结构与功能**

RSCM 基本结构见图 1。RSCM 体现两方面内容: 数学方法主要选取了主成份分

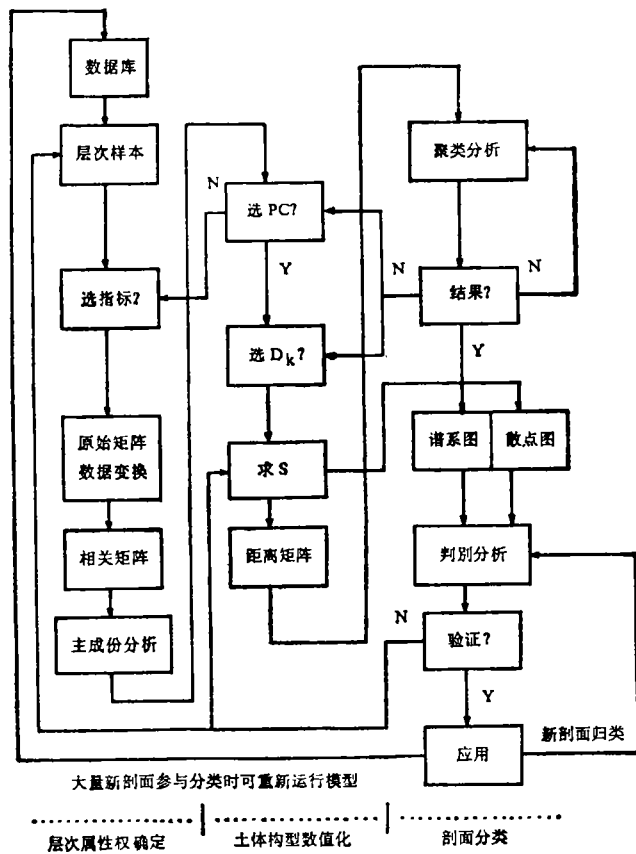


图 1 RSCM 框图  
Fig. 1 The scheme of RSCM

析、聚类分析 (系统聚类和动态聚类)、判别分析, 并结合使用; 专业问题的数学处理是 RSCM 的核心内容, 首先对土壤层次样本进行主成份分析, 在理想剖面的基础上, 利用层次属性主成份值和深度参数函数实现土体构型的数值化, 以数值化后的剖面综合值作为指示通过聚类分析 (聚类谱系图是分类结果的一种形式) 或直接绘制散点图来表示分类结果, 最后用判别分析确定各类型间的界线和指标。

下面就图 1 来讨论 RSCM 的主要功能, 有关计算公式和方法评价见本文所列参考文献。

1. 分类样本选择原则 土壤样本应包函区域内的所有土壤类型, 尤其要有过渡类

型的剖面;严格按照发生层次采集分析样本。土壤剖面总数应大于所分组数的4倍以上<sup>[6]</sup>,各组内样本数目要求比较均匀。

2. 土壤层次样本的属性分析 RSCM 首先是对土壤层次样本属性进行分析。指标选择主要由研究目的和研究对象的特点所决定,一般应尽量选取那些对区分对象有意义的指标,即具有土壤发生学意义的指标,并且这些指标要具有相对的稳定性,同时也较容易获得。

相关分析是综合数值分类的基础,它有助于分析各变量间的相互关系,帮助选择原始指标。

主成分分析能够在保证不损失大部分信息的前提下,大幅度地压缩变量维数的目的。对层次样本进行主成份分析后,获得层次综合计量属性,达到对层次属性压缩的目的(指标权确定的过程)。

3. 土体构型数值化 迄今为止,土壤数值分类中没有很好地处理和解决土壤剖面的层次问题,即没有一个很好的通用办法来实现土壤剖面层次之间的比较问题。前人研究中都是在主成份分析之前,采用属性按层次的加权平均值或某一层次的特征值作为分类的原始指标<sup>[6,7]</sup>,它忽视了土体构型的因素,不是真正的土壤剖面的分类。为了将土壤剖面各层次的综合计量属性值换算成一个数值,作者提出深度参数函数的概念,其核心内容是土壤各层次的属性在土壤分类上的作用是不同的,用  $D_k$  表示深度参数函数。如草甸碱土区域碱土的碱化层次距地面越近、厚度越大、盐渍越强,土体构型越差,对植物的生长和土壤生产力的发挥影响越大,相应的在分类上的作用也不同。

现在引入理论剖面的概念,所谓理想剖面是将原来按照发生层次采集的土壤剖面从地表开始每隔 1 cm 划分为一个假想层次,每一层次综合计量属性为原来剖面相应位置的综合计量属性。以 1 cm 作为一个假想层次是和野外观测的发生层深度最小值相适应的,有了理想剖面和  $D_k$ ,任何剖面之间均可以进行数量上的比较,从而实现了土体构型的数值化,即:

$$S_{i,t} = \sum_{k=1}^m D_k \times PC_{i,k,t} \quad (2)$$

式中,  $S_{i,t}$  为剖面  $i$  的第  $t$  个主成份综合值(剖面综合值);  $PC_{i,k,t}$ 、 $D_k$  分别为第  $i$  个土壤剖面第  $k$  个理论层段(均为 1 cm 厚)第  $t$  个主成份值(可为任意实数)和深度函数参数;  $m$  为用以分类的控制层断内的理论层次数,在数值上它等于控制层段内的厘米深度。 $PC_{i,k,t}$  相当于(1)式中的  $V_k$ ,所不同的是  $V_k$  为所有原始指标属性的集合,而  $PC_{i,k,t}$  是代表土壤某方面属性的综合值(即主成份值),这时厚度因隐含于  $PC_{i,k,t}$  之中,即厚度均为 1 cm。(2)式的含义是土壤剖面某方面属性综合值等于土壤理论层次的该方面综合计量属性与其深度权重乘积之和。

4. 土壤剖面分类及土壤类型界限划分 以剖面综合值( $S$ )作为指标,计算两两剖面间的距离,形成距离矩阵供聚类分析使用。聚类分析是解决未知土壤样本分类的有效方法。

本文在聚类分析获得满意结果的基础上进一步应用判别分析,目的是寻找各类型间的分类界限和进行样本的归类。对土壤这样复杂的分类系统而言,多组判别分析方法效

果较好<sup>7)</sup>。

模型运行至此,得到了四项主要成果:土壤层次样本属性的主成份方程;剖面综合值;聚类谱系图和绘有分类界线的散点图;土壤数值(综合值)分类界线指标,即类间判别点的综合值。

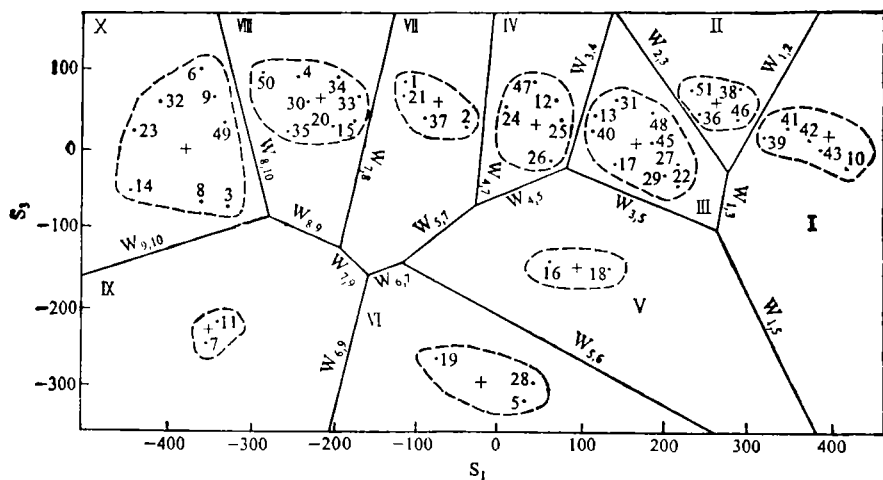
5. 精度讨论 RSCM 的应用精度取决于数学方法本身的精度和应用精度。前者已经证实是可靠的;后者主要取决于数据精度,即土壤剖面的代表性、野外数据取舍的合理性和具有较大分类意义的性质数据漏失程度<sup>5,6)</sup>及深度函数参数的确定等有密切的关系。

### 二、RSCM 应用举例

以吉林省农安县盐渍草甸草原区域土壤土类划分作为简例,来说明模型运行过程和功能。

以 51 个剖面,131 个土壤层次样品作为样本,50 cm 作为控制层段,以 12 个指标作为原始变量,进行主成份分析,得到第一、三主成份主要反映的是土壤盐碱特性和残余碳酸钠含量(两主成份的累积贡献率为 69.57%);而第二主成份主要反映有机质、交换量和粘粒含量,即有机无机矿物质的多少。因此,选取第一和三主成份作为盐渍土土类划分的指标是合适的。D<sub>k</sub> 根据区域特点确定。

以 S<sub>1</sub> 和 S<sub>3</sub> 作为指标绘制散点图(图 2),图中虚线为聚类分析结果,实线为多组判别分析结果,限于篇幅,略去对分类结果和分类结果检验的讨论。



I 混合盐土; II 苏打盐土; III 强碱土; IV 中碱土; V 潜育碱; VI 盐渍潜育土; VII 弱碱土; VIII 中盐渍草甸土; IX 重盐渍草甸土; X (轻盐渍)草甸土; +各组中心坐标

图 2 盐渍土分类结果

Fig. 2 Results of numerical classification of salt-affected soils

## 参 考 文 献

1. 龚子同、雷文进、熊国炎, 1986: 土壤分类的国际趋势。土壤学进展, 第 20 卷 1 期, 1—13 页。
2. 李天杰、王华东、许嘉琳, 1979: 数学分类在土壤中的应用。土壤分类及土壤地理论文集。241—247 页, 浙江人民出版社。
3. Rozanov. B. G., 1986: 国外主要土壤分类的评价。国外农学-土壤肥料, 第 3 期, 6—11 页。
4. 周明秋, 1987: 试论我国土壤基层分类。土壤通报, 第 18 卷 3 期, 102—104 页。
5. 卜兆宏、万洪福, 1982: 土壤分类分级中的综合数值分类的初步研究。土壤学报, 第 19 卷 3 期, 283—295 页。
6. 杨国荣、孟庆秋、王海岩, 1986: 松嫩平原苏打盐渍土数值分类的初步研究。土壤学报, 第 23 卷 4 期, 291—297 页。
7. 肖笃宁、盛世俊, 1987: 横断山区森林土壤的数值分类。土壤学报, 第 24 卷 2 期, 180—192 页。
8. Little, I. P. and D. R. Ross., 1985: Thelevenshtein metric, a new for soil classification tested by data from a sand-podol chronosequence and evaluated by discriminant Function analysis. Aust. J. Soil Res. No:23, 115—130.

## ESTABLISHMENT AND APPLICATIONS OF NUMERICAL SOIL CLASSIFICATION MODEL —ESTABLISHMENT OF RSCM

Hou Yanlin

(Commission for Integrated Survey of Natural Resources, Academia Sinica, 100101)

### Summary

By a combination of many multi variate statistical analysis methods as a skeleton, taking the saline meadow prairie soils in Nongan, Jilin Province as an example, a regional soil classification model (RSCM) was established in the article on the basis of new concepts like pedon structure functional formula. RSCM effectively works out the numerical values of pedon structure and sets up a quantitative criterion for the comparison of profiles.

**Key words** Soil numerical classification, Ideal soil profile, Depth parameter, Integrated value of profile