

基于模糊集理论提取土壤—地形定量关系 及制图应用*

赵量^{1,2} 赵玉国^{1†} 李德成¹ 孙孝林¹ 张甘霖¹

(1 土壤与农业可持续发展国家重点实验室(中国科学院南京土壤研究所), 南京 210008)

(2 南京农业大学资源与环境科学学院, 南京 210095)

摘要 通过对研究区地形因子的模糊聚类,提取了地形因子组合与特定土壤属性的定量隶属度关系,然后对隶属度高值区土壤进行目的性采样为隶属度函数赋值,制作研究区土层厚度连续分布图。通过野外实地验证,将观测值与图中预测值比较,结果显示该方法制图精度在82%左右,具有一定的可靠性。进一步考察认为该模型在地形部位较低,地势较为平坦,土壤发育较好,土层较厚,成土环境相对稳定的地区预测效果更好,适用性更强。该方法能提高土壤制图效率,降低制图成本,提高制图精度,对土壤微域变异的表现更为详细,图面信息负载量更高。应用该方法制作大比例尺土壤详图不失为土壤调查与制图领域一种可行的新方法。

关键词 模糊集理论;土层厚度;土壤—景观模型;土壤制图
中图分类号 S159.9 **文献标识码** A

精准农业、环境管理以及生态过程模拟等的发展产生了对高精度土壤信息的大量需求。然而传统的土壤制图是基于多边形的图形表达和手工完成的过程。基于多边形的土壤图中,由于比例尺的限制,面积较小的土壤个体在图上无法体现,而且很小的上图斑块在制图综合时也常被综合掉,从而产生土壤图在空间上的简化;又由于土壤被图斑界线分割,变异被认为只发生在边界上,不能表达多数情况下土壤性状的渐变和连续,从而产生属性上的简化。因此,如何客观地描述土壤空间连续分布规律,并提供高精度大信息量土壤图成为一个新的课题。

“3S”技术的应用,为土壤制图提供了新的技术与数据支持。以土壤—景观为代表的模型外推方法,是目前较为主流的、适用于区域范围的土壤定量预测制图方法。该方法基于经典的土壤发生学理论,即Jenny方程: $S = f(Cl, O, R, P, T, \dots)$,式中S表示土壤,C_l, O, R, P, T被称为影响土壤成土的环境因子,分别表示气候、生物、地形、母质和时间。该方程预示了这样一个假设:相同的土壤发生景观对应于相同的土壤类型;一旦获知土壤与景观单元之

间的对应关系,就可以通过景观单元的特征去推断土壤类型^[1]。在土壤—景观模型的建立中,气候、植被、母质等信息可以通过分析遥感数据获取,利用同位素地球化学的方法可以了解土壤的年龄。然而地形直接决定了水分运动过程以及坡面物质的迁移和再分配,这在土壤发生学上意义明确,因此,地形因子是应用最普遍、最有效的定量预测因子。已发表的研究中,采用地形作为景观变量之一的占80%^[2],尤其在较小范围内,气候、植被、母质差异不大,地形因素是土壤发生发育和类别差异的决定性因素。据McSweeney的研究,在小流域尺度内,高程、坡度、平面曲率、剖面曲率及复合地形指数这五个地形因子可以反映土壤的发生和发育^[3]。

很多研究探索了多种建立土壤—景观模型的方法,其中模糊聚类方法是一种非监督分类方法,它根据分类对象在多属性空间的距离进行自动分类,得到各聚类类别与各聚类中心的一组相似度向量,称之为隶属度(Fuzzy membership)^[4]。该方法用连续划分的模糊隶属度代替了传统模型中非此即彼的二值假设,可以更准确地表达连续性地理现象在空间的渐变和

* 国家自然科学基金项目(40401024)和江苏省自然科学基金项目(BK2006526)资助

† 通讯作者, Tel: 025-86881299, E-mail: ygzhaol@issas.ac.cn

作者简介: 赵量(1981~), 男, 河南项城人, 硕士研究生, 从事资源与环境信息系统研究。E-mail: lzhaol@issas.ac.cn

收稿日期: 2007-01-16; 收到修改稿日期: 2007-03-27

过渡,符合土壤系统连续性变化的本质。Bruin 和 Stein 研究发现模糊聚类得到的隶属度和表土黏粒含量之间存在较好的相关性^[5];Lark 的研究则表明,对地形因子进行计算得到的模糊隶属度与土壤属性之间的相关性强于地形因子与土壤属性之间的相关性,这说明模糊聚类方法可以表示土壤属性与地形之间复杂的非线性关系^[6];Zhu 等应用模糊推理原理建立了 SoLIM 模型,预测土壤类型、A 层深度、导水率,在地形起伏较大的地区应用得到较好结果^[7,8]。

在土壤的诸多属性中,土层厚度是一个重要的基本特性。土层厚度指可以生长植物的土壤厚度,一般指 A+B 层。该层土壤已经脱离母质性状,其厚度能够反映土壤发育的程度,同时该土层是维系植物生长的基础,其厚度和性状是土壤质量的重要表征指标。

本研究使用模糊 c 均值聚类(FCM)方法对决定性地形因子进行聚类分析,得到土层厚度与地形之间的对应关系,在此基础上进行野外目的性采样,用获取的土壤属性为隶属度函数赋值,最后用线性加权的方法得到研究区土层厚度连续分布图。

1 研究区概况

研究区位于安徽省宣城市金坝乡,地理位置在北

纬 30°51'10" ~ 30°52'30",东经 118°37'37" ~ 118°40'04" 之间,面积为 9.3km²。该区是中亚热带与北亚热带的过渡地区,属于亚热带湿润季风气候,年均温度 16℃,年均降水量 1300mm。地貌类型属低丘岗地,海拔高度为 32 ~ 80m。成土物质有白垩纪砂岩、页岩、泥砾岩风化物 and 黄土沉积物。植被类型为常绿针叶林和落叶阔叶林,目前原生植被较少,多为自然次生植被和人工植被。土壤类型有富铁土、淋溶土、雏形土和水耕人为土^[9]。

2 研究方法

2.1 地形因子计算

在 Arc/info 中数字化宣城市周王茶场 1:1 万地形图(图幅编号:H-50-46-(27)),生成 10m 分辨率 DEM。借助数字地形分析技术对 DEM 进行计算得到聚类分析所采用的五个地形因子:高程、坡度、平面曲率、剖面曲率及复合地形指数。高程决定地势起伏、势能、气候、植被类型等;坡度反映径流流速、植被、土壤水分;两个曲率描述的是分水/汇水流域、土壤水分含量、径流加速度、侵蚀/分解速率等;复合地形指数则表现土壤湿度、流量累积、水位深度、土壤水分蒸发、土壤中的有机物、含沙量分布等^[10]。关于这五个地形因子的定义及算法见表 1。

表 1 地形因子定义及计算方法^[10]

Table 1 Definition and calculation of terrain factors

地形因子 Terrain factors	定义 Definition	算法 Calculation
高程 Elevation (H), m	地面点沿铅垂线到大地水准面的距离 The height of a given point on the land surface above sea-level	
坡度 Slope (θ), °	地面点法线与铅垂线之夹角 An angle between a tangent plane and a horizontal one at a given point on the land surface	$= \arctan \sqrt{p^2 + q^2}^{(1)}$
平面曲率 Planform curvature (C _t), (100m) ⁻¹	过地面上任一点的水平面沿水平方向切地形表面所得的曲线在该点的曲率值 A curvature of a normal section of the land surface. This section is orthogonal to the section of profile curvature at a given point on the land surface	$C_t = - \frac{q^2 r - 2 p q s + p^2 t}{(p^2 + q^2)(1 + p^2 + q^2)^{1/2}}^{(1)}$
剖面曲率 Profile curvature (C _p), (100m) ⁻¹	对地面坡度的沿最大坡降方向地面高程变化率的度量 A curvature of a normal section of the land surface by a plane, including gravity acceleration vector at a given point	$C_p = - \frac{p^2 r + 2 p q s + q^2 t}{(p^2 + q^2) \sqrt{(1 + p^2 + q^2)^3}}^{(1)}$
单位汇水面积 Specific catchment area (CA), m ² m ⁻¹	地表上任意点或等高线之上游汇水区的平面投影面积与等高线段长度之比 A ratio of an area of an exclusive figure formed on the one hand by a contour intercept with a given point on the land surface and, on the other by flow lines coming from the upslope to the ends of this contour intercept, to the length of this intercept	根据秦承志等提出的多流向算法 MFD-fg 计算 ^[11] Calculated by Qin's MFD-fg method
复合地形指数 Compound topographic index (CTI)	$CTI = \ln(CA/\tan \theta)$	如定义,见单位汇水面积和坡度的计算 Referring to the definition

1) 将 DEM 数据模拟为一个连续的曲面 $H(x, y)$, x 和 y 为地面点的平面坐标值, $H(x, y)$ 为地面点高程值,则上表各式中符号的算式为:
 $p = \frac{\partial H}{\partial x}, q = \frac{\partial H}{\partial y}, r = \frac{\partial^2 H}{\partial x^2}, s = \frac{\partial^2 H}{\partial x \partial y}, t = \frac{\partial^2 H}{\partial y^2}$ Making the datas of DEM to a continuous surface $H(x, y)$, x and y are the values of plane rectangular coordinates system, $H(x, y)$ is the value of height, then the meanings of letters in the table are: $p = \frac{\partial H}{\partial x}, q = \frac{\partial H}{\partial y}, r = \frac{\partial^2 H}{\partial x^2}, s = \frac{\partial^2 H}{\partial x \partial y}, t = \frac{\partial^2 H}{\partial y^2}$

2.2 对地形因子进行模糊聚类提取土壤—地形关系

本研究采用模糊 c 均值 (FCM) 方法进行聚类分析。FCM 的基本原理是利用统计方法计算多属性空间中的距离使得每个数据点距离自己所属类别中心点最近,并且以此给每个数据点分配隶属度。用隶属度函数定义的聚类损失函数表示为:

$$J_m(U, v) = \sum_{i=1}^n \sum_{k=1}^c u_{ik}^m d_{ik}^2 \quad (1)$$

$$d_{ik}^2 = \|y_i - v_k\|_A^2 = (y_i - v_k)^T A (y_i - v_k) \quad (2)$$

式中, y 为数据集, c 为聚类数, m 为模糊指数, n 为数据集 y 中的数据点数, U 为 y 的一个模糊 c 划分, v 是聚类中心的向量, A 为距离权重矩阵, d_{ik} 为数据点 y_i 到第 k 类聚类中心 v_k 的加权距离, u_{ik} 为第 i 个数据点属于第 k 类的隶属度函数。

模糊 c 均值算法要求样本对各个聚类的隶属度之和为 1, 即:

$$\sum_{k=1}^c u_{ik} = 1 \quad (3)$$

在条件(3)下用迭代法求解式(4)、式(5)得样本隶属度和各类别聚类中心,并满足准则函数式(1)最小化目标:

$$u_{ik} = \frac{d_{ik}^{-2/(m-1)}}{\sum_{k=1}^c d_{ik}^{-2/(m-1)}} \quad (4)$$

$$v_k = \frac{\sum_{i=1}^n (u_{ik})^m y_i}{\sum_{i=1}^n (u_{ik})^m} \quad (5)$$

当算法收敛时,就得到了各类别的聚类中心 v_k 和各个数据点对于各聚类类别的隶属度值 u_{ik} ,从而完成了各地形因子的模糊聚类^[12]。

在 FCM 算法中,模糊指数 m 为一个可控制聚类结果和类别交叠程度的常数。当 $m = 1$, 聚类结

果趋向于硬划分,失去模糊划分能力;当 $m \rightarrow \infty$ 时,FCM 得到的隶属度均为 $1/c$, 样本隶属于各类的程度相等,使得分类结果太模糊而且得到 c 个一样的聚类原型,达不到聚类的目的。一般地, m 有效值在 1 和 30 之间变化,实验研究多取值在 1 至 3 之间。很多研究采用 $m = 2$ ^[13]。Bezdek 在研究中发现, $m = 2$ 的 FCM 算法具有明确的物理意义^[14]。Pal 等从聚类有效性的实验研究得到 m 的最佳选取区间为 $[1.5, 2.5]$, 在不做特殊要求时,可选用区间的中值 $m = 2$ ^[15]。

为了确定最优分类数 c , 这里引入分割系数 (Partition coefficient, F) 和分类熵 (Classification entropy, H) 的概念^[13,16]:

$$F_c(u) = \frac{\sum_{i=1}^n \sum_{k=1}^c (u_{ik})^2}{n} \quad (6)$$

$$H_c(u) = - \frac{\sum_{i=1}^n \sum_{k=1}^c (u_{ik} \log_a(u_{ik}))}{n} \quad (7)$$

分割系数取值在 $1/c$ 至 1 之间,熵取值在 0 至 $\log_a(c)$ 之间。 F 测量不同类别之间的重叠度,与总体模糊集内不同类别之间平均重叠度成反比, H 为模糊分类 U 模糊度的测量。随着聚类数的增加, H 增大, F 减小,当达到某一类数时,熵变化较前后两类变化高,熵值的增大量减小;同时,分割系数变化较前后两类变化低,分割系数的减少量减小,之后熵变化和分割系数变化趋于稳定,认为该聚类数即为最优分割。根据以上理论,经过反复试验,本研究取 $m = 2.0$ 时, $c = 10$ 为最佳聚类数,即:模糊指数 m 取 2,由地形因子聚类将该地区土壤属性划分为 10 类。至此,得到了特定的地形因子组合与特定土壤属性组合之间的空间对应关系。聚类过程如图 1 所示。

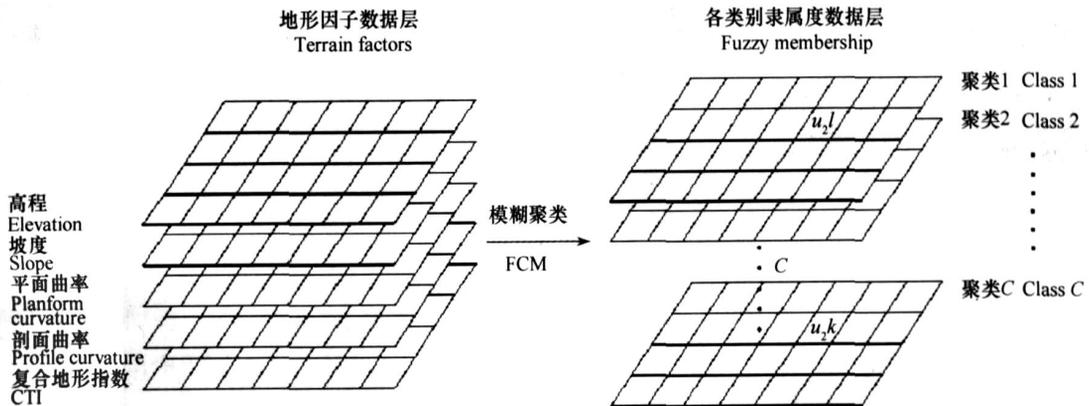


图 1 聚类过程示意

Fig. 1 Clustering procedure

3 结果与讨论

3.1 模糊聚类结果

对上述五个地形因子进行模糊聚类,得到 10 个聚类类型的隶属度分布图。在各图中隶属度高值区域的

斑块中间部位,选择隶属度大于 0.9 的栅格点,认为该处土壤的属性即为该种聚类类型土壤的典型属性。为了减少误差,每类选择三个点,使用 GPS 在野外找到该点并采样。采样时尽量确保三处土壤的植被、母质、坡向、微域地形部位相似。取这三处土壤土层厚度的平均值作为该聚类类型土壤的土层厚度(见表 2)。

表 2 各聚类类型土层厚度典型值

Table 2 Typical thicknesses of soil horizons of each class in clustering

类别 Category	土层厚度 Horizon depth (cm)								
Class1	67	Class3	118	Class5	66	Class7	79	Class9	110
Class2	40	Class4	77	Class6	73	Class8	110	Class10	110

土壤属性图的编制是通过点 i 的土壤环境与各聚类中心 k 典型环境条件的相似度 u_{ik} 的隶属度线性加权的方法来进行,这主要是基于当地土壤的成土环境与给定土壤类型的成土环境相似,则当地土壤的性状就与给定土壤类型性状相似的假设^[8]:

$$V_i = \frac{\sum_{k=1}^c u_{ik} \cdot V^k}{\sum_{k=1}^c u_{ik}} \quad (8)$$

式中, V_i 表示某点 i 的土壤属性值, V^k 为各聚类中心 k 的土壤属性值, c 为这一地区给定的土壤类型的总数。

将各个聚类中心的土层厚度值(见表 2)代入上式,即得到该研究区土层厚度连续分布图(图 2)。

的准确程度,评价该方法的可靠性。根据统计学上的中心极限定理,从一个总体抽取样本,当抽样数 $n > 30$, t 分布才稳定,检验才较为有效,本研究在野外采集 34 个点作为验证点。

将观测集与预测集作相关分析,计算得 Pearson 相关系数为 0.668 **,表明预测值与观测值之间是高度相关的。对两组数据进行方差分析(F 检验), $F = 0.34 < F_{0.05} = 3.99$ 故在显著性水平 0.05 下,认为预测集与观测集之间的方差差异不具备统计显著性;对两组数据进行成对双样本均值分析(t 检验), $t = 0.92 < t_{0.05/2} = 2.03$,故在置信度为 0.05 的情况下,认为预测集与观测集的平均值之间的差异不具备统计显著性。

在此基础上,进一步用以下三个指标来评价模型的准确程度,分别是均方根误差(Root mean square error)、相对误差(Relative error)和吻合度(Agreement coefficient)^[10]。均方根误差(RMSE)是对模型精度的度量,值小说明模型的精度越高;相对误差(RE)能够客观地反映预测结果的准确度,通常用百分比表示,其值越小,说明预测值与观测值越接近;吻合度(AC)反映预测值与观测值的吻合程度,其值在 0 至 1 内变化,越接近 1,表明预测值与观测值之间吻合程度越好,反之则差。吻合度的计算公式如下^[8]:

$$AC = 1 - \frac{N \cdot RMSE^2}{PE}$$

$$PE = \sum_{j=1}^n (|P_j - \bar{O}| + |O_j - \bar{O}|)^2 \quad (9)$$

式中, N 、 P 、 O 分别为观察数、预测值、观测值。

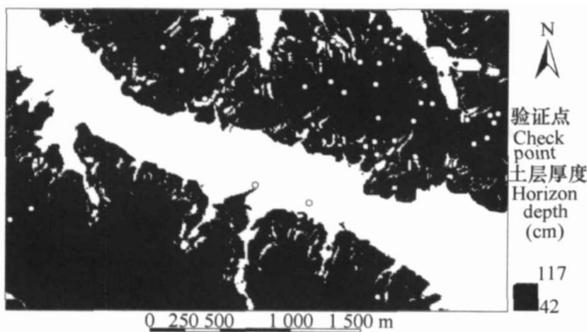


图 2 研究区土层厚度连续分布

Fig. 2 Continuous soil horizon thickness distribution map of the study site

3.2 模型验证与精度评价

为了验证该制图方法的精度,在周王茶场研究区内实地采集验证剖面(称之为观测集),采样在各个不同的地形部位上随机进行,同时考虑到微地貌差异。对预测集与观测集进行比较来验证预测模型

精细,有的粗放,在开垦时间上有的几十年有的只有几年,所以其土壤厚度变异较大,相对误差较高。

总体而言,该模型在地形部位较低、土壤发育较好、土层较厚、成土环境相对稳定的地区预测效果较好,适用性较强。

4 结 论

本研究通过对地形因子的模糊聚类得到了地形因子组合与土壤属性组合的定量隶属度关系,然后在隶属度高值区进行目的性采样为隶属度函数赋值,制作研究区土层厚度连续分布图。通过野外随机采样,将观测值与图中预测值比较,进行精度检验,模型的精度在 82% 左右,从土壤制图的精度及需求考虑,具有一定的可靠性。对相对误差作进一步的剖析,发现预测模型在地形部位较低、土壤发育较好、土层较厚、成土环境相对稳定的地区预测效果较好,适用性较强。

本研究在对地形因子聚类得到隶属度图的基础上,只进行了很少的目的性采样——10 个聚类类型,每类采集 3 个点,共 30 个土壤剖面就得到了土壤属性与地形因子之间的对应关系,满足了制图要求,提高了采样效率,极大地减少了工作量,从而降低了制图成本。

预测得到的土层厚度分布图厚度变异是连续的、渐变的,没有人为划分的图斑界线,更符合土壤空间连续变化的实际情况;由于空间上的细小变化在计算机成图时可以通过小像元来体现,所以对土壤微域变异的表现也更为详细,图面信息负载量更高,从而制图精度更高。该制图方法与传统制图方法相比,大部分工作都由计算机完成,制图周期短,更新速度快。

总之,研究表明模糊聚类方法是提取土壤属性和地形因子关系的一种有效途径,应用该方法制作大比例尺土壤详图是土壤调查与制图领域一种可行的新方法。

致 谢 中国科学院地理科学与资源研究所朱阿兴教授等为本研究提供了大量帮助,中国科学院南京土壤研究所杜国华研究员在野外调查工作中给予重要的指导和帮助,在此一并表示感谢。

参 考 文 献

[1] Hudson B D. The soil survey as a paradigm based science. *Soil Sci-*

ence Society of America Journal, 1992, 56: 836 ~ 841

- [2] McBratney A B, Mendon a Santos M L, Minasny B. On digital soil mapping. *Geoderma*, 2003, 117:3 ~ 52
- [3] McSweeney K, Slater B K, Hammer R D, *et al.* Towards a new framework for modeling the soil-landscape continuum. In *Factors of Soil Formation: A Fiftieth Anniversary Publication*, R. Amundson (ed.), Madison, WI: Soil Science Society of America, 1994. 127 ~ 154
- [4] 朱阿兴,李宝林,杨琳,等. 基于 GIS、模糊逻辑和专家知识的土壤制图及其在中国应用前景. *土壤学报*, 2005, 42(5): 844 ~ 851. Zhu A X, Li B L, Yang L, *et al.* Predictive soil mapping based on a GIS expert knowledge, and fuzzy logic framework and its application prospects in China (In Chinese). *Acta Pedologica Sinica*, 2005, 42(5):844 ~ 851
- [5] Bruin S, Stein A. Soil-landscape modeling using fuzzy c -means clustering of attribute data derived from a Digital Elevation Model (DEM). *Geoderma*, 1998, 83:17 ~ 33
- [6] Lark R M. Soil-landform relationships at within-field scales: an investigation using continuous classification. *Geoderma*, 1999, 92: 141 ~ 165
- [7] Zhu A X, Band L E, Dutton B, *et al.* Automated soil inference under fuzzy logic. *Ecological Modelling*, 1996, 90: 123 ~ 145
- [8] Zhu A X, Band L, Vertessy R, *et al.* Derivation of soil properties using a Soil-Land Inference Model SoLIM. *Soil Science Society of America Journal*, 1997, 61: 523 ~ 533
- [9] 顾也萍,钱进,吕成文,等. 安徽宣城样区土系的划分. *土壤*, 2001, 33(1):7 ~ 12. Gu Y P, Qian J, Lu C W, *et al.* Soil series in Xuancheng studied area, Anhui Province (In Chinese). *Soils*, 2001, 33(1): 7 ~ 12
- [10] 周启鸣,刘学军. 数字地形分析. 北京:科学出版社,2006. Zhou Q M, Liu X J. *Digital Terrain Analysis* (In Chinese). Beijing: Science Press, 2006
- [11] 秦承志,朱阿兴,李宝林,等. 基于栅格 DEM 的多流向算法述评. *地学前缘*, 2006, 13(3):91 ~ 98. Qin C Z, Zhu A X, Li B L, *et al.* Review of multiple flow direction algorithms based on gridded digital elevation models(In Chinese). *Earth Science Frontiers*, 2006, 13(3): 91 ~ 98
- [12] Tan M Z, Xu F M, Chen J, *et al.* Spatial prediction of heavy metal pollution for soils in Peri-urban Beijing, China based on Fuzzy Set theory. *Pedosphere*, 2006, 16(5):545 ~ 554
- [13] 高新波. 模糊聚类分析及其应用. 西安:西安电子科技大学出版社, 2004. Gao X B. *Fuzzy Cluster Analysis and Its Applications* (In Chinese). Xi'an:Xi'an Electronic Science and Technology University Press, 2004
- [14] Bezdek J C. A physical interpretation of fuzzy ISODATA. *IEEE Trans. SMC*, 1976, 6(3): 387 ~ 390
- [15] Pal N R, Bezdek J C. On clustering for the fuzzy c -means model. *IEEE Trans. FS*, 1995, 3(3): 370 ~ 379
- [16] Bezdek J C, Ehrlich R, Full W. FCM: The fuzzy c -means clustering algorithm. *Computers and Geosciences*, 1984, 10 (2/3): 191 ~ 203

DIGITAL SOIL MAPPING BY EXTRACTING QUANTITATIVE RELATIONSHIPS BETWEEN SOIL PROPERTIES AND TERRAIN FACTORS BASED ON FUZZY SET THEORY

Zhao Liang^{1,2} Zhao Yuguo^{1†} Li Decheng¹ Sun Xiaolin¹ Zhang Ganlin¹

(1 *State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China*)

(2 *College of Resources and Environmental Science, Nanjing Agriculture University, Nanjing 210095, China*)

Abstract A study was conducted to extract relationships between terrain factors and soil properties by clustering terrain factors based on the fuzzy set theory. Similarity of each pixel to the typical ones was obtained from the classification results, with degrees of similarity referred to as fuzzy membership. Then, a continuous soil horizon thickness distribution map was drawn with the aid of ARC/INFO software based on samples collected at the area with high value of fuzzy membership and population of the similarity model using the linear and additive weighting function. Comparison of the predicted value set from the map with the independently collected field sample set showed the derived soil map achieved 82 % of accuracy. Further investigation showed that the model worked well in the area of low altitude, well-developed soil, thick soil horizon, stable surrounding for soil formation. The new approach was found to be higher in efficiency and accuracy of soil mapping, lower in mapping cost, and more capable of demonstrating microregional variance and accommodating higher information load for mapping. Hence, it is a feasible new method worth extrapolating in soil survey and mapping.

Key words Fuzzy set theory; Soil horizon depth; Soil-Landscape model; Soil mapping