

DOI: 10.11766/trxb201503260023

# 基于传统土壤图的土壤—环境关系获取及推理制图研究\*

黄魏<sup>1</sup> 罗云<sup>1</sup> 汪善勤<sup>1</sup> 陈家赢<sup>1</sup> 韩宗伟<sup>1</sup> 祁大成<sup>2</sup>

(1 华中农业大学资源与环境学院, 武汉 430070)

(2 湖北省红安县农业局土壤肥料工作站, 湖北黄冈 738400)

**摘要** 在数字土壤制图研究中, 从历史资料中提取准确的、详细的土壤—环境关系对于土壤图的更新和修正十分重要。从传统土壤图中提取土壤类型并从地形数据中提取环境参数, 采用空间数据挖掘方法建立土壤—环境关系, 并进行推理制图和精度验证。以湖北省黄冈市红安县华家河镇淝水河流域为例, 首先选取成土母质和基于地形数据提取的高程、坡度、坡向等7个环境因子; 然后利用频率分布原理得到包含土壤类型与环境因子信息的典型样本数据1410个; 采用See5.0决策树方法进行空间数据挖掘, 建立土壤—环境关系; 将其导入SoLIM中进行推理制图; 最后利用270个实地采样点验证所得土壤图的精度。土壤图的精度提高了约11%, 证明了本研究方法对土壤类型和空间分布推理的可靠性。

**关键词** 环境因子; 频率分布; 数据挖掘; 土壤—环境推理模型 (SoLIM)

**中国分类号** P934 **文献标识码** A

数字土壤制图以Jenny的土壤发生学理论<sup>[1]</sup>和Huggett的土壤—景观模型理论<sup>[2]</sup>为基础, 其基本思想为环境要素与土壤的空间差异存在协同关系, 一旦获取这种关系, 就可以依据环境要素的空间分布来推测土壤的空间变化, 并以此为基础进行土壤推理制图<sup>[3]</sup>。土壤—环境关系知识的有效性决定着土壤制图的精度, 因此, 准确获取土壤与环境关系成为精细土壤制图的关键。

现阶段土壤—环境关系知识的获取主要源于两种方式: 由土壤普查专家提供<sup>[4-5]</sup>或通过数据挖掘方法从已有数据中提取<sup>[6]</sup>。前者获得的知识是土壤普查专家通过大量野外调查形成的经验性知识, 需要长期积累才能形成; 后者采用不同数据挖掘算法, 以传统土壤图、案例或大量野外样点作为数据来源获取土壤—环境关系知识<sup>[7]</sup>, 这种方法弥补

了某些地区土壤专家缺乏的局面。我国在第二次土壤普查中形成了完整的土壤志和土壤图(全国土壤普查办公室, 1993—1996), 如何充分利用这些历史积累的宝贵资料, 以获取精细土壤制图所需要的知识成为目前研究的重点和难点<sup>[3]</sup>。

传统土壤图是土壤普查专家进行野外调查的产物。在调查制图的过程中, 土壤专家通过大量的野外调查同时结合地形图、地质图等对航片进行解译, 识别具有特定环境特征的景观单元, 并以此为基础勾绘土壤多边形<sup>[8]</sup>。因此其中蕴含了土壤专家对土壤与环境间关系的经验性知识, 土壤类型多边形与多个在一定属性范围内变化的环境因子组合存在一定的对应关系<sup>[6]</sup>。根据土壤普查专家绘制土壤图的过程, 从多边形土壤图中提取知识的基本思想即为土壤图绘制过程的逆过程<sup>[6, 9]</sup>。在这一

\* 国家自然科学基金项目(41171174)、国家863项目(2013AA102401-3)、中央高校基本科研业务费专项资金项目资助(2010QC035)资助 Supported by the National Natural Science Foundation of China (No.41171174), the National High Technology Research and Development Program (No.2013AA102401-3), and the Central Universities Fundamental Research Funding (No.2010QC035)

作者简介: 黄魏(1975—), 男, 汉族, 博士, 副教授, 研究方向为精细数字土壤制图、资源环境信息工程。E-mail: ccan@mail.hzau.edu.cn

收稿日期: 2015-03-26; 收到修改稿日期: 2015-06-24

过程中，与土壤专家识别景观单元的过程相似，对环境因子（如高程、坡度等）进行综合可以得到具有特定环境特征的环境因子组合<sup>[10]</sup>。因此，本研究以传统土壤图作为知识获取源，探讨从传统土壤图中获取土壤—环境知识用于更新土壤图的方法，并验证该方法的有效性。

## 1 材料与方 法

### 1.1 研究区及数据

红安县位于湖北省东北部，鄂豫两省交界处，地跨东经114° 23′~114° 49′，北纬

30° 56′~31° 35′之间，气候较为温润，夏季较为炎热。成土母质主要为片麻岩、石灰岩、近代河流冲积物、泥质岩、基性岩和石英片岩。土地利用类型以耕地、林地为主，其他为水域、道路及建筑用地，其中林地主要是以次生林为主，耕地以水田为主。按土壤发生分类县内土壤可分为4个土类，7个亚类，26个土属类型。研究区域华家河镇淝水河流域位于红安县西北部，地势北高南低，属于半山区半丘陵地区，面积约为27 km<sup>2</sup>。本研究所使用的传统土壤图（1：50000）来源于第二次全国土壤普查，以发生分类下的土属类型作为制图单位（图1）。

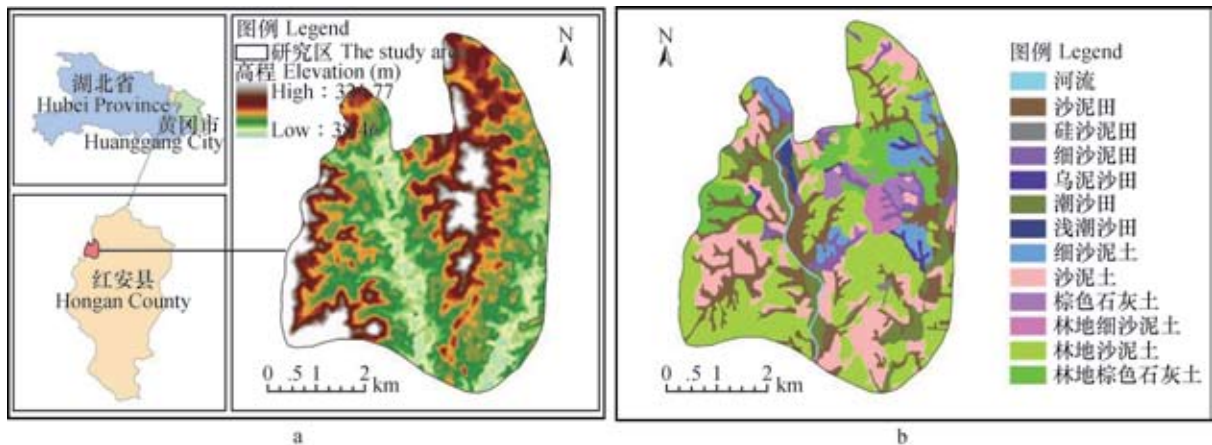


图1 研究区示意图、DEM图 (a) 和传统土壤图 (b)

Fig.1 Location map and DEM of the study area (a), conventional soil map of the study area (b)

### 1.2 环境因子数据选择

为了更好地获取隐含在土壤图中的土壤—环境关系，选择能够有效描述当地土壤发生环境的相关因子十分重要<sup>[11]</sup>。影响土壤形成的主要因子有母质、气候、生物、地形、时间等，而在较小的空间范围内，气候、时间因子可以忽略，地形是最能反映土壤空间差异的环境因子，也是最为重要的因子。它不仅影响土壤内部的水分运动和物质迁移，而且影响土壤母质的形成<sup>[12]</sup>。地形属性根据能否直接由数字高程模型（Digital Elevation Model, DEM）获得分为基本地形属性（包括坡度、坡向、曲率等）和派生地形属性（如地形湿度指数）<sup>[13]</sup>。

本研究选择母质和地形作为土壤制图的环境要素。由于没有详细的成土母质图层，利用传统土壤图中各土壤类型对应的母岩信息生成替代的母质层，根据母岩类型的不同将研究区划分为6种母质单元，并生成母质图。研究区的地形特征可通过6

个地形因子来描述，即高程、坡度、坡向、沿平面曲率、沿剖面曲率、地形湿度指数。在ArcGIS9.3平台下，利用来源于湖北省测绘局的10 m等高线数据，采用TopoGrid方法生成DEM。由DEM获取6个地形变量，其中地形湿度指数采用多流向算法得到的上游汇流面积及最大下坡坡度计算而成，因为多流向算法在坡面等地形上的漫散流动模拟非常有效，精度高，更符合水流流向的实际状况<sup>[14-15]</sup>。

### 1.3 数据预处理

在传统土壤制图过程中由于受制图比例尺的限制，小斑块土壤类型在图中不显示，但却作为土壤包含物存在于其他土壤类型的多边形内<sup>[6]</sup>。因此需要对数据进行预处理来减少地图中包含的噪声、错误值等。而与土壤类型有关的环境因子在土壤图中有着高频率的参数域分布，那这一土壤类型中的错误则存在于低频部分<sup>[16]</sup>。根据研究的预处理结果，每种土壤类型斑块必须有15个采样点才能保

证数据描述的有效性。计算面积和像元间距后,选取 $100 \times 100$ 行列的均匀网格生成原始样点6571个。数据预处理过程采用频率分布原理,利用环境直方图拟合与土壤类型相关环境条件的频率分布,通过对图上采样点的处理获得能体现土壤—环境关系知识的代表性样本集。其中每种土壤类型的样本集选取均遵循以下两个步骤:一是选取每个环境因子落入峰值区的数据,即提炼代表性高的典型样本集;二是对每种土壤的7个环境因子的典型样本集进行交叉操作,获取至少落入两个直方图峰值区的数据,即该土壤的样本集。最后将所有土壤类型的样本集综合起来即为数据预处理后的最终样本集1410个。

#### 1.4 数据挖掘

数据挖掘算法有多种,如神经网络<sup>[17]</sup>、决策树<sup>[18]</sup>、判别分析<sup>[19]</sup>等,不同的分类方法和技术可以通过比较其预测速度、准确率等指标来进行评估。而决策树算法与其他分类算法相比,它的输入变量多样化,既可以是连续的,也可以是离散的,甚至可以是定名变量,此外其分类结果为易于解译的分类判别准则文件且样本处理效率高<sup>[20]</sup>。See5.0是美国地质调查局(USGS)在进行土地覆盖分类项目过程中开发的支持决策树分类的软件模块<sup>[21]</sup>,该算法与其他分类算法的不同之处在于强调基于if-then规则的分类器,更容易理解,每个规则可以单独检验和验证,更适宜于土壤推理过程的研究。本研究采用See5.0决策树软件提取土壤—环境知识,从训练样本中找出类的特征并自动建立决策树及决策规则,并且利用该软件对分类树进行修剪可以避免数据的过度拟合问题,以减少所要制定的规则。

#### 1.5 推理制图

土壤—环境推理模型(Soil-Land Inference Model, SoLIM)是利用GIS、模糊逻辑和专家知识建立的应用于土壤制图的模型,它可以有效地推理土壤的空间分布,提高土壤制图的速度和精度<sup>[8]</sup>。SoLIM模型根据已经建立的土壤—环境知识对区域内每个像元依次进行计算,得到所有像元的相似度向量,即该像元土壤与某种土壤类型的相似度<sup>[5]</sup>,并以模糊隶属度图<sup>[22]</sup>表示。由于参与推理的7个环境因子有比率量(高程、坡度等)和名称量(成土母质),因此采取的隶属度函数的形式也不同<sup>[3]</sup>。比率量一般采用连续型隶属度函

数,如钟型、S型和Z型,本研究中坡向数据采用 $0 \sim 360^\circ$ 之间的连续数值表示,因此采用连续型隶属度函数;名称量则采用阶跃型隶属度函数表示。将每一个像元相似度向量中最大隶属度值所代表的土壤类型作为该像元的土壤类型,即对模糊隶属度图进行硬化,得到土壤类型分布图<sup>[23]</sup>。

#### 1.6 精度验证

利用野外实地采取验证集对传统土壤图和更新后的土壤图进行精度评估。验证点数量的设计遵循以下原则:在每种土壤类型分布区内所确定的样点数量与该土壤类型的面积成正比,并且每种土壤类型内至少布设2个采样点。本研究采用均匀采样(126个)、主观采样(116个)和横截面采样(28个)三种采样方式获取验证点<sup>[24]</sup>,分布情况如图2所示。均匀采样和主观采样用来检验所获土壤信息的整体分布情况;横截面采样用于检验推理图能否有效展现土壤信息的空间分布变化,本研究设计横穿山坡和沟谷的4条线路,使28个采样点分布在环境类型发生变化的区域。由于验证点中有6个样点落入乌泥沙田和硅沙泥田中,而它们的空间分布范围由传统土壤图直接获得,因此,去除这6个样点对于更新后的土壤图的精度评价更有意义。通过对剩余的264个样点确定土壤—环境之间关系的准确性,验证推理所得土壤图的精度。

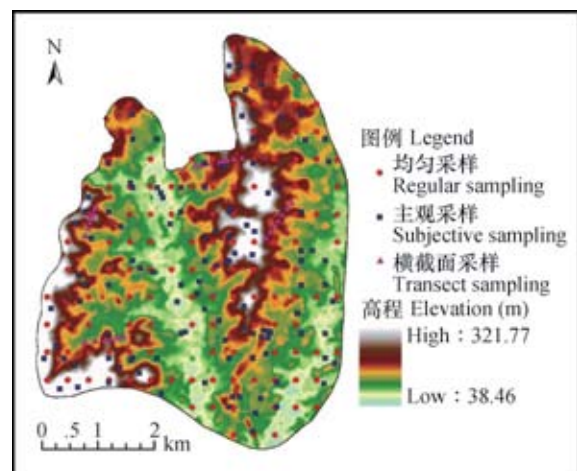


图2 270个验证点分布图

Fig.2 Distribution of 270 validation points

利用SoLIM生成的混淆矩阵得到Kappa系数、生产精度和用户精度等指标来分析验证更新后的土壤图的精度。其中Kappa系数用于衡量更新后土壤图与实际土壤类型之间的一致性程度<sup>[25]</sup>,生产精

度代表的是土壤类型图所能正确分类的水平，用户精度代表了土壤类型图所表达的土壤类型分类的可靠程度<sup>[11]</sup>。

## 2 结 果

### 2.1 获取土壤—环境关系知识

利用成土母质图对研究区进行分区，对于仅包含一种土壤类型的母质单元（如基性岩或石英片岩）不进行数据挖掘，直接将传统土壤图中该土壤

类型的多边形作为这种土壤的空间分布范围。在其他母质单元中，将预处理后的最终样本集导入 See5.0 软件，采用数据挖掘算法对 7 个环境因子进行分析并生成决策树，得到土壤—环境知识来预测研究区的土壤类型。决策树的分类结果表明，在本研究内沿平面曲率和沿剖面曲率因子对土壤类型的影响很小，在知识整理过程中去除这两个因子使研究结果更有效。结合红安县土壤志和普查资料对获取的原始土壤—环境知识进行整理，得到各土壤类型发生环境条件的描述表（表 1）。

表 1 各土壤类型发生环境条件的描述表

Table 1 Descriptions of environmental conditions for genesis of various soil types

土壤类型 Soil type	成土母质 Parent material	高程 Elevation ( m )	坡度 Slope ( ° )	坡向 Aspect ( ° )	湿度指数 Wetness
细沙泥土 Fine sand soil	泥质岩 Argillaceous rock	63.34 ~ 79.34 79.34 ~ 84.38			
林地细沙泥土 Forest sand soil	泥质岩 Argillaceous rock	> 109.61			
细沙泥田 Fine sand field	泥质岩 Argillaceous rock	< =63.348 79.34 ~ 84.38 84.38 ~ 109.61		< =162.92	
沙泥土 Sand soil	片麻岩 Gneiss	< =121.95	< =19.54	< =213.42	< =3.63
林地沙泥土 Forest sand soil	片麻岩 Gneiss	> 121.95 < =121.95	> 19.54		
沙泥田 Sand filed	片麻岩 Gneiss	< =121.95	< =19.54		> 3.63
棕色石灰土 Brown rendzina	石灰岩 Limestone	< =119.14		< =100.65	
林地棕色石灰土 Forest brown rendzina	石灰岩 Limestone	< =119.14 > 119.14		> 100.65	
浅潮沙田 Shallow moist sand field	近代河流冲积物 Modern river alluvial material	< =50.99 > 81.93		> 217.48	
潮沙田 Moist sand field	近代河流冲积物 Modern river alluvial material	50.99 ~ 81.93 > 81.93		< =217.48	

### 2.2 土壤制图结果

将土壤—环境关系知识和环境数据输入 SoLIM 模型，得到每一种土壤类型的隶属度分布图<sup>[26]</sup>。将隶属度分布图硬化得到每一个栅格单元内的土壤

类型，再与传统土壤图的乌泥沙田和硅砂泥田的空间分布图集成，得到栅格大小为 10 m 的更新后的土壤图（图 3）。与传统土壤图（图 1b）相比，更新后的土壤图的图斑数量明显增加，空间分布更加详



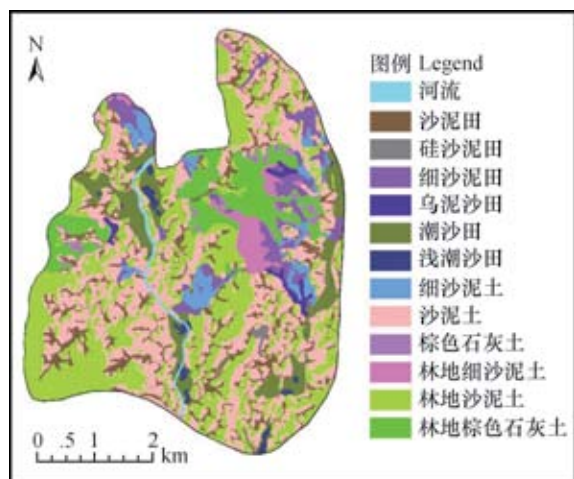


图3 更新后的土壤图

Fig.3 Updated soil map

细, 展现了更多细节信息。

### 2.3 制图精度验证

将验证点的真实土壤类型与传统土壤图和更新后的土壤图中对应位置的土壤类型进行对比(表2), 结果表明更新后的土壤图的总体精度可达75%, 高于传统土壤图的64%, 且三种采样方式下的验证精度均高于传统土壤图。Kappa系数为

70%, 一致性程度为显著, 能较好反映研究区的土壤分布情况。这说明更新后的土壤图不仅可以展现土壤信息在空间上的整体分布, 更能反映地形变异与土壤类型空间分布的协同趋势。

传统土壤图与更新后的土壤图的各土壤类型的生产精度和用户精度(表3)表明, 除林地沙泥土等个别土壤类型外, 研究区内大部分土壤类型的生产精度和用户精度均有一定程度的提高。其中林地细沙泥土、林地棕色石灰土的生产精度和用户精度均有很大提高, 这是由于它们分布在陡峭山体, 高程相对较高, 与地形变化呈现协同一致趋势; 而沙泥土、棕色石灰土等精度也有一定提高, 其分布于坡上, 坡度适中, 湿度较小, 由于研究基本单元精度的提高(10m分辨率), 更易展现空间细节; 由于分布于河谷平原以及沟谷地带, 地形最为平坦, 因此潮沙田、浅潮沙田等土壤大部分精度有所提高, 原由是原始成图过程中被合并忽略的图斑被还原。而林地沙泥土和沙泥田的生产精度有所降低, 说明当前的分类水平不够, 主要是因为这两类土壤的发生环境条件与其他土壤相似, 就现有的环境因子条件来说区别并不明显。因此更新后的土壤图主要是对研究区的整体土壤分布状况进行一个尽可能

表2 不同采样方式下验证点的精度比较

Table 2 Comparison in accuracy of validation points between different sampling strategies

采样方式 Sampling strategies	土壤图 Soil map	正确样点个数 Correct number of sampling points	总样点个数 Total number of sampling points	精度 Accuracy (%)
均匀采样 Regular sampling	传统土壤图 Conventional soil map	78	125	62
	更新后土壤图 Updated soil map	91	125	73
主观采样 Subjective sampling	传统土壤图 Conventional soil map	70	111	63
	更新后土壤图 Updated soil map	84	111	76
横截面采样 Transect sampling	传统土壤图 Conventional soil map	19	28	68
	更新后土壤图 Updated soil map	20	28	71
全部采样点 All sampling points	传统土壤图 Conventional soil map	168	264	64
	更新后土壤图 Updated soil map	198	264	75

精细的预测，推理的结果相较于原始土壤图精度有一定的提高。

综上所述，经264个野外验证点评价结果表明，更新后的土壤图的总体分类精度达到75%，高出原始土壤图约11%，图斑数量更多，细节更明显。因此这种利用原始土壤图获取土壤与环境关系

的数字土壤制图方法是可行的。此外，与传统土壤制图方法相比，对土壤类型的判断和边界的确定更加准确，能有效更新土壤图，并且本研究大部分工作利用计算机完成，在降低了人力、物力消耗的同时还缩短了制图周期，使得土壤调查更具时效。

表3 更新后的土壤图与传统土壤图的各土壤类型生产精度和用户精度比较

Table 3 Comparison between the conventional and updated soil maps in accuracy of production of soil types and of users (%)

土壤类型 Soil type	生产精度 Producer' s accuracy		用户精度 User' s accuracy	
	传统土壤图 Conventional soil map	更新后的土壤图 Updated soil map	传统土壤图 Conventional soil map	更新后的土壤图 Updated soil map
	细沙泥土 Fine sand soil	53	59	53
林地细沙泥土 Forest sand soil	40	67	86	100
林地棕色石灰土 Forest brown rendzina	69	96	95	96
沙泥土 Sand soil	35	82	51	60
林地沙泥土 Forest sand soil	78	76	71	84
棕色石灰土 Brown rendzina	50	50	11	50
浅潮沙田 Shallow moist sand field	60	60	75	75
细沙泥田 Fine sand field	50	80	28	40
沙泥田 Sand filed	69	47	55	68
潮沙田 Moist sand field	95	95	90	90

### 3 结 论

利用模糊推理得到的土壤图无论是在空间详细度还是在属性精确度上均有了较大的提高。与传统土壤调查方法相比，本研究采用了更详细的环境数据，包括改进后的成土母质图及由10 m分辨率DEM生成的地形因子数据，使得更新后的土壤图结果更

加合理。研究采用SoLIM模型来量化土壤与环境之间的关系，利用相似度来预测土壤类型。大量案例已证明SoLIM较传统手工、主观的土壤制图方法更精确。然而本研究主要考虑的环境因子是成土母质和地形数据，对于个别土壤的判定并不明显。而对于县域级土壤来说，土地利用方式和植被覆盖等因子对土壤类型的影响也很大，因此如何合理有效

地将这些因子加入推理模型中将下一步研究的重点。此外,本研究中数据预处理过程的有效性已得到验证,但如何在这一过程中提高典型样点的获取精度以更好地提取土壤—环境关系知识也是下一步需要完善的地方。

## 参考文献

- [ 1 ] Jenny H. Factors of soil formation. *Soil Science*, 1941, 52 ( 5 ) : 415
- [ 2 ] Huggett R J. Soil landscape systems: A model of soil genesis. *Geoderma*, 1975, 13: 1—22
- [ 3 ] 朱阿兴, 李宝林, 裴韬, 等. 精细数字土壤普查模型与方法. 北京: 科学出版社, 2008  
Zhu A X, Li B L, Pei T, et al. Fine digital soil census model and method ( In Chinese ). Beijing: Science Press, 2008
- [ 4 ] Shi X, Zhu A X, Burt J E, et al. A case-based reasoning approach to fuzzy soil mapping. *Soil Science Society of America Journal*, 2004, 68: 885—894
- [ 5 ] 刘京, 朱阿兴, 张淑杰, 等. 基于样点个体代表性的大尺度土壤属性制图方法. *土壤学报*, 2013, 50 ( 1 ) : 12—20  
Liu J, Zhu A X, Zhang S J, et al. Large-scaled soil attribute mapping method based on individual representativeness of sample sites ( In Chinese ). *Acta Pedologica Sinica*, 2013, 50 ( 1 ) : 12—20
- [ 6 ] Qi F, Zhu A X. Knowledge discovery from soil maps using inductive learning. *International Journal of Geographical Information*, 2003, 17 ( 8 ) : 771—795
- [ 7 ] 孙福军, 雷秋良, 刘颖, 等. 数字土壤制图技术研究进展与展望. *土壤通报*, 2011, 42 ( 6 ) : 1502—1507  
Sun F J, Lei Q L, Liu Y, et al. The progress and prospect of digital soil mapping research ( In Chinese ). *Chinese Journal of Soil Science*, 2011, 42 ( 6 ) : 1502—1507
- [ 8 ] 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
- [ 9 ] Schmidt K, Behrens T, Friedrich K, et al. A method to generate soilscape from soil maps. *Journal of Plant Nutrition and Soil Science*, 2010, 173 ( 2 ) : 163—172
- [ 10 ] 杨琳, 朱阿兴, 李宝林, 等. 应用模糊c-均值聚类获取土壤制图所需土壤—环境关系知识的方法研究. *土壤学报*, 2007, 44 ( 5 ) : 784—791  
Yang L, Zhu A X, Li B L, et al. Extraction of knowledge about soil-environment relationship for soil mapping using fuzzy c-means ( FCM ) clustering ( In Chinese ). *Acta Pedologica Sinica*, 2007, 44 ( 5 ) : 784—791
- [ 11 ] 杨琳, Sherif F, Sheldon H, 等. 基于土壤—环境关系的更新传统土壤图研究. *土壤学报*, 2010, 47 ( 6 ) : 1039—1049  
Yang L, Sherif F, Sheldon H, et al. Updating conventional soil maps using knowledge on soil environment relationships extracted from the maps ( In Chinese ). *Acta Pedologica Sinica*, 2010, 47 ( 6 ) : 1039—1049
- [ 12 ] 张淑杰, 朱阿兴, 刘京, 等. 基于样点的数字土壤属性制图方法及样点设计综述. *土壤*, 2012, 44 ( 6 ) : 917—923  
Zhang S J, Zhu A X, Liu J, et al. Sample-based digital soil mapping methods and related sampling schemes ( In Chinese ). *Soils*, 2012, 44 ( 6 ) : 917—923
- [ 13 ] 韩宗伟, 黄魏, 张春弟, 等. 基于土壤养分—景观关系的土壤采样布局合理性研究. *华中农业大学学报*, 2014, 33 ( 1 ) : 56—61  
Han Z W, Huang W, Zhang C D, et al. Rationality of sampling strategies based on soil-environment relationships ( In Chinese ). *Journal of Huazhong Agricultural University*, 2014, 33 ( 1 ) : 56—61
- [ 14 ] Qin C Z, Zhu A X, Pei T, et al. An approach to computing topographic wetness index based on maximum downslope gradient. *Precision Agriculture*, 2011, 12: 32—43
- [ 15 ] 张维, 杨昕, 汤国安, 等. 基于DEM的平缓地区水系提取和流域分割的流向算法分析. *测绘科学*, 2012, 37 ( 2 ) : 94—96  
Zhang W, Yang X, Tang G A, et al. DEM-based flow direction algorithms study of stream extraction and watershed delineation in the low relief areas ( In Chinese ). *Science of Surveying and Mapping*, 2012, 37 ( 2 ) : 94—96
- [ 16 ] Qi F. Knowledge discovery from Area-Class resource maps data preprocessing for noise reduction. *Transactions in GIS*, 2004, 8 ( 3 ) : 297—308
- [ 17 ] Li Q, Cheng J C, Hu Y M. Spatial interpolation of soil nutrients based on BP neural network. *Agricultural Science & Technology*, 2014, 15 ( 3 ) : 506—511
- [ 18 ] Carvalho D R, Freitas A A. A hybrid decision tree/genetic algorithm method for data mining. *Information Sciences*, 2004, 163 ( 1 ) : 13—35

- [ 19 ] 邱琳, 李安波, 赵玉国. 基于Fisher判别分析的数字土壤制图研究. 土壤通报, 2012, 43 ( 6 ) : 1281—1286  
Qiu L, Li A B, Zhao Y G. Digital soil mapping based on Fisher discriminant analysis ( In Chinese ) . Chinese Journal of Soil Science, 2012, 43 ( 6 ) : 1281—1286
- [ 20 ] 周斌, 王繁, 王人潮. 运用分类树进行土壤类型自动制图的研究. 水土保持学报, 2004, 18 ( 2 ) : 140—143, 147  
Zhou B, Wang F, Wang R C. Automated soil mapping by using classification tree algorithm ( In Chinese ) . Journal of Soil and Water Conservation, 2004, 18 ( 2 ) : 140—143, 147
- [ 21 ] Quinlan R. See5: An informal tutorial. URL: <http://www.rulequest.com>, 2001
- [ 22 ] Zhu A X, Yang L, Li B L, et al. Construction of membership functions for predictive soil mapping under fuzzy logic. Geoderma, 2010, 155: 166—174
- [ 23 ] Zhu A X, Hudson B, Burt J. Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Science Society of America Journal, 2001, 65 ( 5 ) : 1463—1472
- [ 24 ] 杨琳, 朱阿兴, 秦承志, 等. 一种基于样点代表性等级的土壤采样设计方法. 土壤学报, 2011, 48 ( 5 ) : 938—946  
Yang L, Zhu A X, Qin C Z, et al. A soil sampling method based on representativeness grade of sampling points ( In Chinese ) . Acta Pedologica Sinica, 2011, 48 ( 5 ) : 938—946
- [ 25 ] 田苗, 王鹏新, 严泰来, 等. Kappa系数的修正及在干旱预测精度及一致性评价中的应用. 农业工程学报, 2012, 28 ( 24 ) : 1—7  
Tian M, Wang P X, Yan T L, et al. Adjustment of Kappa coefficient and its application in precision and agreement evaluation of drought forecasting models ( In Chinese ) . Transactions of the Chinese Society of Agricultural Engineering, 2012, 28 ( 24 ) : 1—7
- [ 26 ] 杨琳, 朱阿兴, 秦承志, 等. 运用模糊隶属度进行土壤属性制图的研究——以黑龙江鹤山农场研究区为例. 土壤学报, 2009, 46 ( 1 ) : 9—15  
Yang L, Zhu A X, Qin C Z, et al. Soil property mapping using fuzzy membership—A case study of a study area in Heshan farm of Heilongjiang Province ( In Chinese ) . Acta Pedologica Sinica, 2009, 46 ( 1 ) : 9—15

## Knowledge of Soil-landscape Model Obtain from A Soil Map and Mapping

HUANG Wei<sup>1</sup> LUO Yun<sup>1</sup> WANG Shanqin<sup>1</sup> CHEN Jiaying<sup>1</sup> HAN Zongwei<sup>1</sup> QI Dacheng<sup>2</sup>

( 1 College of Resource and Environment, Huazhong Agricultural University, Wuhan 430070, China )

( 2 Soil and Fertilizer Station of Agricultural Bureau of Hong'an County, Huang Gang, Hubei 438400, China )

**Abstract** Conventional soil maps are what soil survey experts turn out after field soil survey and interpretation of corresponding aerial photos, and often used as major data sources of information about spatial distribution of soils, which is essential to watershed management and eco-hydrology. With the development of geographic information technique, traditional soil survey methods are already far from efficient to meet the requirements of soil information services. As they used to be based on the experts' empirical model of thinking, their products are often hard to express, exchange and store; the qualitative characteristics they described of a soil entity are often inconsistent with the characteristics of its spatial distribution, which tends to lead to low accuracy of the survey; and they are very costly and also limited to certain regions, which makes it hard to have information updated. Therefore, how to make full use of the existing historical resources and data is very important to retrieving efficiently soil maps higher in accuracy from the available information in Digital Soil Mapping ( DSM ). In this study, from the conventional soil maps and terrain data extracted were data of soil type and environment factors, based on which, a soil environment relationship model was established using the spatial data mining method, and finally, reliability and accuracy of the mapping was validated by field sampling. The Nieshui river basin in Huajiahe Town, Hongan County, Huanggang City, Hubei Province was selected for case study. The conventional soil maps of the study area plotted during the Second National Soil Survey were used to demonstrate processes of the research. The



proposed method consists of five major steps. 1) Select seven environmental factors that were closely related to the process of pedogenesis and establish a geographic information system (GIS) database, which should contain a modified soil parent material map and data of terrain factors (elevation, slope, aspect, plan curvature, profile curvature and topographic wetness index) extracted from 10 m resolution Digital Elevation Model (DEM). 2) Extract 1410 typical sample data of soil types and environment factors by following the principle of frequency distribution, so as to reduce noises and abnormal data that would often occur in traditional soil mapping, because traditional soil mapping used to be done manually and contain some hard-to-reflect knowledge (or noise) of the experts' about proper relationship models. It is, therefore, essential to have the data properly pretreated. 3) Retrieve detailed expertise implied in the soil map product, using the spatial data mining techniques. Compared with the other algorithms, the decision tree algorithm is the most suitable one for extracting and expressing knowledge of the soil-environment model. So, the See5.0 decision tree algorithm is selected to perform spatial data mining and hence, obtain knowledge of soil and environment relationships. 4) Predict soil spatial distribution through inferring and mapping in Soil-Land Inference Model (SoLIM) based on the soil-environment knowledge and environment data obtained. SoLIM uses similarity degree as measurement parameter and fuzzy logic as basis to calculate similarity between soils. Within a given pixel, a number of corresponding soils have a variety of similarity degrees, which can be represented in fuzzy membership degree. Finally, the soil type represented by the highest fuzzy membership degree among the similarity vectors of a pixel is defined as the soil type of the pixel. A soil type distribution map can be obtained by hardening the fuzzy membership degree map. A large number of case studies have demonstrated that SoLIM is a more accurate than the traditional manual subjective method in soil mapping. 5) Verify accuracy of the proposed method through sampling at 270 field validation points using three sampling strategies: regular sampling, subjective sampling and transect sampling. Results show that the soil map obtained through fuzzy inference provides more detailed information about soil spatial distribution than its corresponding conventional soil map and is about 11% higher in accuracy and significantly higher in number of patches. It is therefore concluded that the proposed method which retrieves soil-environment relationships from a traditional soil map is more accurate than the conventional mapping method in judging and delineating and more convenient for use to update soil maps.

**Key words** Environment factor; Frequency distribution; Data mining; Soil-Land Inference Model (SoLIM)

(责任编辑: 檀满枝)