

DOI: 10.11766/trxb201706220234

基于不确定性模型的土壤—环境关系知识 获取方法的研究*

黄 魏 许 伟 汪善勤 袁雅萍 望陈运

(华中农业大学资源与环境学院, 武汉 430070)

摘 要 土壤与环境关系知识的获取是精细数字土壤制图的关键, 如何快速准确地获取该知识成为现阶段研究的重点。以湖北省黄冈市红安县华家河镇为例, 利用土壤—环境推理模型 (Soil-Land Inference Model, SoLIM) 得到土壤类型的夸大和忽略不确定性分布图, 依据不确定性分布图在可信度高的位置重新采集样点, 对样点进行数据挖掘, 获取环境因子组合, 建立其与土壤类型的对应关系。结合原始规则, 生成新的土壤—环境关系知识, 并将其用于土壤推理制图, 获得新的土壤类型分布, 利用 253 个野外独立样点进行精度验证。结果表明: 推理土壤图显示了更加详细的空间分布信息, 经野外验证点验证, 总体精度为 86.9%, 高于原土壤图精度约 13%。因此, 利用不确定性模型重新获取土壤—环境关系知识的方法是可行且有效的, 该方法不仅增加了土壤图的空间详细度, 而且提高了土壤图的精确度。

关键词 数字土壤制图; 决策树; 不确定性; 知识融合

中图分类号 S159 **文献标识码** A

数字土壤制图以土壤发生学理论和土壤—环境模型为基础, 其核心是土壤空间差异与环境要素之间存在着协同关系^[1], 依据这种关系可以对土壤空间变化进行推测与制图。目前, 已有大量学者对知识的获取方法展开了研究, 主要包括三大类: 依赖土壤普查专家知识的土壤专家方法^[2]; 基于决策树、神经网络等算法的空间数据挖掘方法^[3-4]; 利用模糊数学从传统土壤图中获取知识的方法^[5-6], 但是这些方法均存在一定的局限性^[7-8]。

土壤作为一个自然连续体, 它的连续性不仅表现在属性的渐变上, 同时也表现于地理空间分布上^[9-10]。因此, 土壤实体表现出对不同土壤类型的多重相似性, 在知识获取过程中的不确定性会影响土壤推理制图的准确性^[11]。朱阿兴^[12]提出在数字土壤制图过程中, 获取土壤与环境关系知识时会产生两种不确定性: 忽略不确定性和夸大不确定性, 用于描述土壤类别划分的可信程度。

本研究利用不确定性模型实现知识的重新获取, 将基于模糊推理生成的土壤图与不确定性图进行融合, 在可信度高的区域重新采集样点, 挖掘不确定性图所蕴含的土壤分布规律, 探索最优土壤—环境关系知识组合。研究目的是利用原始推理图、不确定性模型等资源, 有效地获取最佳土壤—环境关系知识组合, 以提高土壤推理制图精度。

1 材料与方法

1.1 研究区概况及数据来源

研究区 (图 1) 位于湖北省黄冈市红安县境内 (114°26' ~ 114°33'N, 31°23' ~ 31°30'E), 该区地处湖北省东北部大别山南麓, 鄂豫两省交界处。属于亚热带季风气候, 年平均气温为 15.7℃, 较为温润, 夏季较为炎热。全区均为半山半丘陵地区, 地势北高南低。成土母质主要包括片麻岩、石灰

* 国家自然科学基金项目 (41171174) 和国家重点研发计划项目 (2017YFD0202001) 资助 Supported by the Natural Science Foundation of China (No.41171174), National key research and development project (No.2017YFD0202001)

作者简介: 黄 魏, 博士, 副教授, 主要从事精细数字土壤制图、资源环境信息工程研究。E-mail: ccan@mail.hzau.edu.cn

收稿日期: 2017-06-22; 修回日期: 2017-08-29; 优先数字出版日期 (www.cnki.net): 2017-10-12

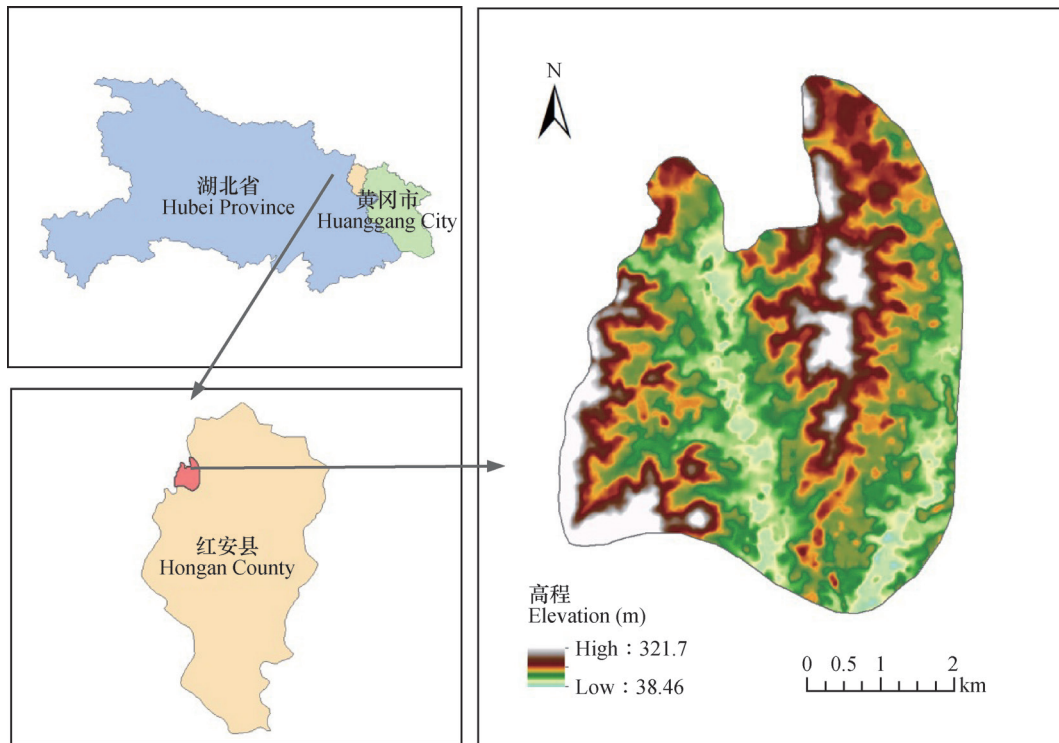


图1 研究区示意图

Fig. 1 Location map of the study area

岩、近代河流冲积物、泥质岩、基性岩和石英片岩。土壤可分为四个土类，七个亚类，可细分为26种土属。研究中用到的数据包括：从10 m分辨率的等高线（来源于湖北省测绘局）提取的各类环境因子，包括高程、坡度、坡向、曲率、地形湿度指数等；利用决策树模型从传统土壤图获取的土壤—环境关系组合，以及通过其关系组合推理生成的各类型土壤隶属度图（具体研究数据和方法可参考文献[9]）。

1.2 不确定性模型

在土壤—环境推理模型（Soil-Land Inference Model, SoLIM）中采用相似度模型来表达土壤在空间上的变化^[1]，其基本理论是某点土壤与多种土壤类型之间存在着不同程度的相似性，以隶属度来表示某点土壤与土壤类型 k 的相似度。将相似度向量中最大的隶属度值所代表的土壤类型作为该点的土壤类型，即土壤图硬化过程^[13-14]。在硬化的过程中，每个像元被赋予了唯一的土壤类型，但是该像元也存在着与其他土壤类型之间的相似性，因此，产生了忽略不确定性和夸大不确定性。

在相似度模型的基础上，对分类误差进行全面定义，从而可以精确详细地描写分类结果中不确定

性在空间上的变化^[15]，即忽略不确定性和夸大不确定性^[16-17]。忽略不确定性跟某点土壤与所有土壤类型的相似性有关，像元点的土壤与各类型土壤的隶属程度越接近，即相似度向量中的隶属度越平均（分散），忽略不确定性就越大。忽略不确定性可通过熵来计算^[17]，具体公式如下：

$$U_{ij} = \frac{1}{\ln N} \sum_{k=1}^N (S_{ij}^k \ln S_{ij}^k) \quad (1)$$

式中， U_{ij} 为忽略不确定性，范围从0至1， S_{ij}^k 为点 (i, j) 对土壤类型 k 的隶属度， N 为研究区内的土壤类型数。

夸大不确定性即某点被确定为某土壤类型的夸大程度，与像元所属类型的隶属度呈反比。该点在其所属类型的隶属度越高，夸大程度就越低。计算公式^[17]可简单表达为

$$F_{ij} = 1 - S_{ij}^g \quad (2)$$

式中， F_{ij} 为夸大不确定性， S_{ij}^g 为点 (i, j) 对土壤类型 g 的隶属度。

1.3 土壤推理制图

土壤—环境推理模型（Soil-Land Inference Model,）是一种成功应用于土壤制图的自动推理

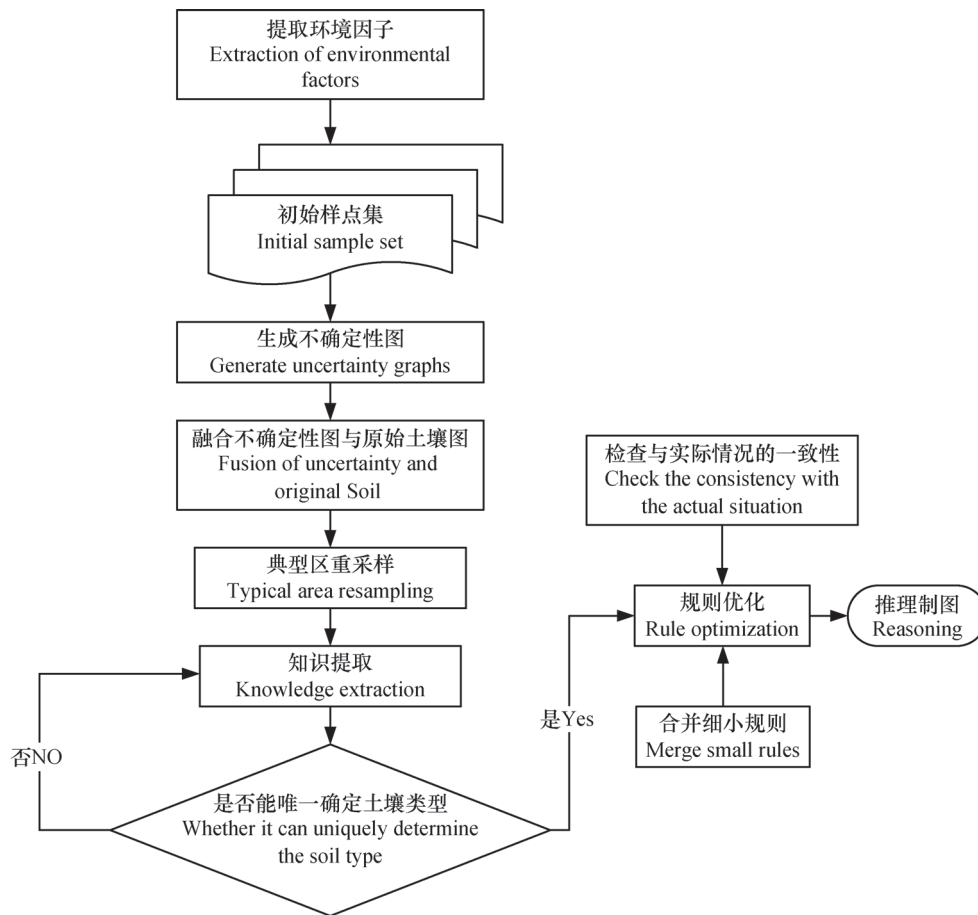


图2 方法流程图

Fig. 2 Flow chart of the method

模型，它将土壤—环境关系知识与环境数据结合，得到每类土壤对应的隶属度分布图^[18]。通过硬化，即将每一像元所有土壤类型隶属度中的最大值所代表的土壤类型作为该像元的土壤类型^[13]，可得到研究区土壤类型栅格图。利用SoLIM可得到土壤类型的夸大和忽略不确定性分布图。

1.4 知识提取的基本思路

依据不确定性分布图在可信度高的位置重新采集样点，对样点进行数据挖掘，获取环境因子组合，建立其与土壤类型的对应关系，结合原始规则方法的总体流程如图2所示。

2 结果与讨论

2.1 不确定性融合图

基于SoLIM5.0平台，将利用决策树模型生成的各类型土壤隶属度图进行硬化。相似度向量硬化时会产生不确定性图，夸大不确定性（左）和忽略不

确定性（右）的分布如图3所示，反映了土壤分类时各个像元的不确定性程度。图中颜色越亮的区域表示该像元推理的不确定性越高；反之，颜色越暗表示不确定性越低。在夸大不确定性分布图中，除去河流及周边支流部位，其余区域夸大不确定性值均很低，表明这个区域中土壤推理结果与实际类型基本符合。

将硬化后的土壤类型栅格图与不确定性图进行融合，便于直观地了解土壤分类的准确度。图4中左右分别为土壤类型栅格图与夸大不确定性和忽略不确定性的融合图，图中颜色越白的区域不确定性越高，颜色越纯，表示该颜色代表的土壤类型的可信度越高。例如图4中，A区域的土壤属于林地棕色石灰土的可信度高，而B区域附近的土壤属于林地沙泥土的可信度相对较低。进一步研究表明，不确定性值较大的区域多分布在两种土壤类型的过渡区，而不确定性值小的区域基本存在于土壤斑块的内部，这主要是因为过渡区的土壤与周边各类土壤

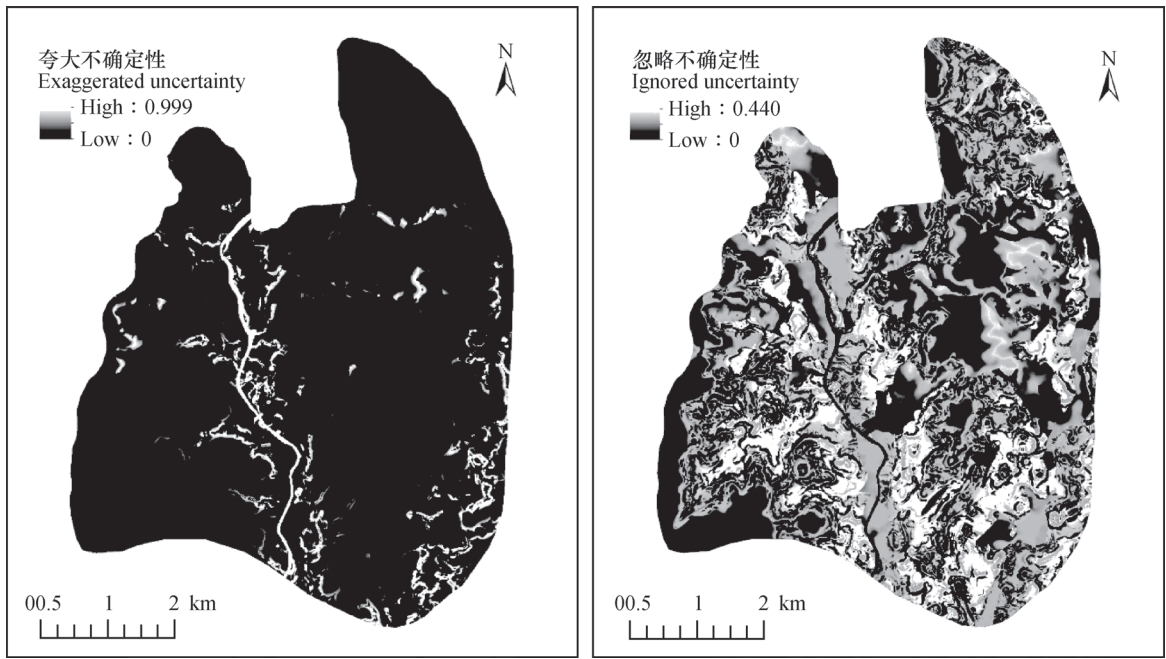


图3 不确定性分布图

Fig. 3 Distribution of uncertainty

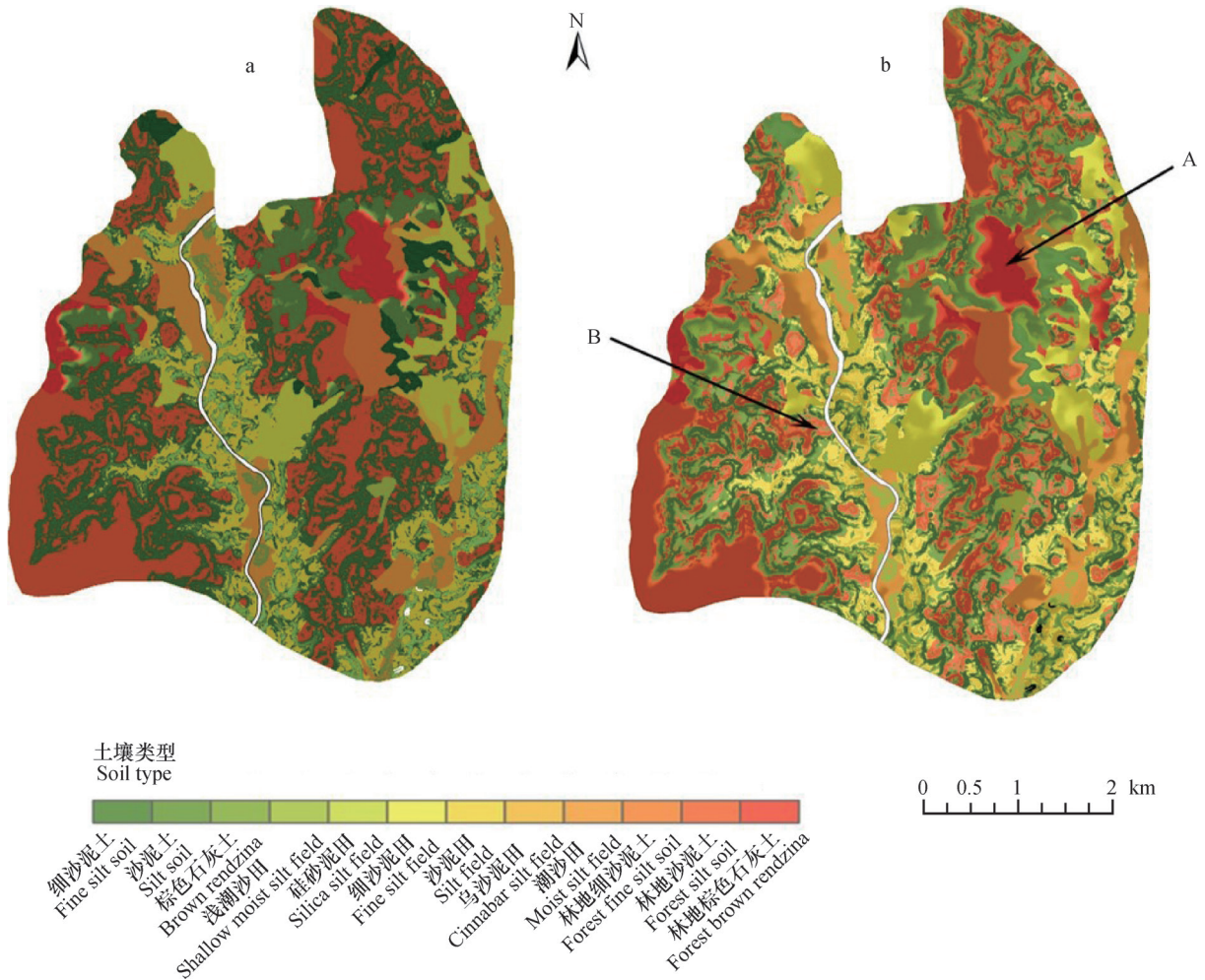


图4 土壤类型与不确定性融合图

(a为土壤类型与夸大不确定性的融合图, b为土壤类型与忽略不确定性的融合图)

Fig.4 Fusion graph of soil type and uncertainty

(A is a fusion graph of soil type and exaggerated uncertainty, and b is a fusion graph of soil type and ignored uncertainty)

均有一定的相似度,因此将该区域的土壤划分到任何一类土壤均会产生较大的忽略不确定性。由于夸大不确定性值较低,且分布范围较小,本文将基于忽略不确定进行研究。

2.2 获取可靠的土壤—环境关系知识

本研究目的是降低成图过程中人为因素导致的知识误差,以挖掘可靠的土壤—环境关系知识。SoLIM模型可以生成土壤类型分布图和不确定性分布图,并可对两类图进行融合。融合图中颜色越白的地区不确定性越高;而颜色越纯,则该颜色所代表的土壤类型的可信度就越高^[1]。根据土壤类型与忽略不确定性的融合图,选取不确定性较低(可信度较高)的部位作为代表性的区域进行重采样。采样遵循独立性原则,采样点必须覆盖不同海拔、不同坡向上的坡顶、坡中、坡底等典型景观位置,具体采样点分布如图5,合计样点数980个。

提取各样点对应的环境因子值,包括母质类型、高程、曲率、坡度、坡向、地形湿度指数。以母质为依据,将研究区内的土壤样点分为6大类,分析确定每大类下的环境因子重要性,按照重要性

排序依次对土壤样点进行分类,直到确定出唯一的土壤类型。根据研究区对应的原始土壤图和土壤志等已有资料,参考前期的决策树规则,对新提取知识进行优化。优化的具体流程如下:一方面依据原来的规则,合并细小的规则;另一方面对规则进行整理,在整理时要依据研究区现有资料和实地考察情况,了解规则与所对应的实际地表描述是否统一。如林地沙泥土,在母质为片麻岩的前提下,高程和曲率为主要的影响环境因子,而这类土壤基本存在于高程属于80~121 m且曲率小于-0.0013或高程大于121 m的区域,由此获得新的规则。

2.3 优化前后规则的对比

与原始规则相比,优化后的规则在环境因子的选取与组合上有了较大变化,特别是林地沙泥土,沙泥田,浅潮沙田等几类土壤。以沙泥土,林地沙泥土,沙泥田三类分布面积较大的土壤类型为例,对其推理规则进行对比(表1)。原始规则中利用母质、高程、坡度、坡向和地形湿度指数为推理因子,优化后的规则则是利用母质、高程、曲率和地形湿度指数为推理因子。发现这三类土壤可通过这

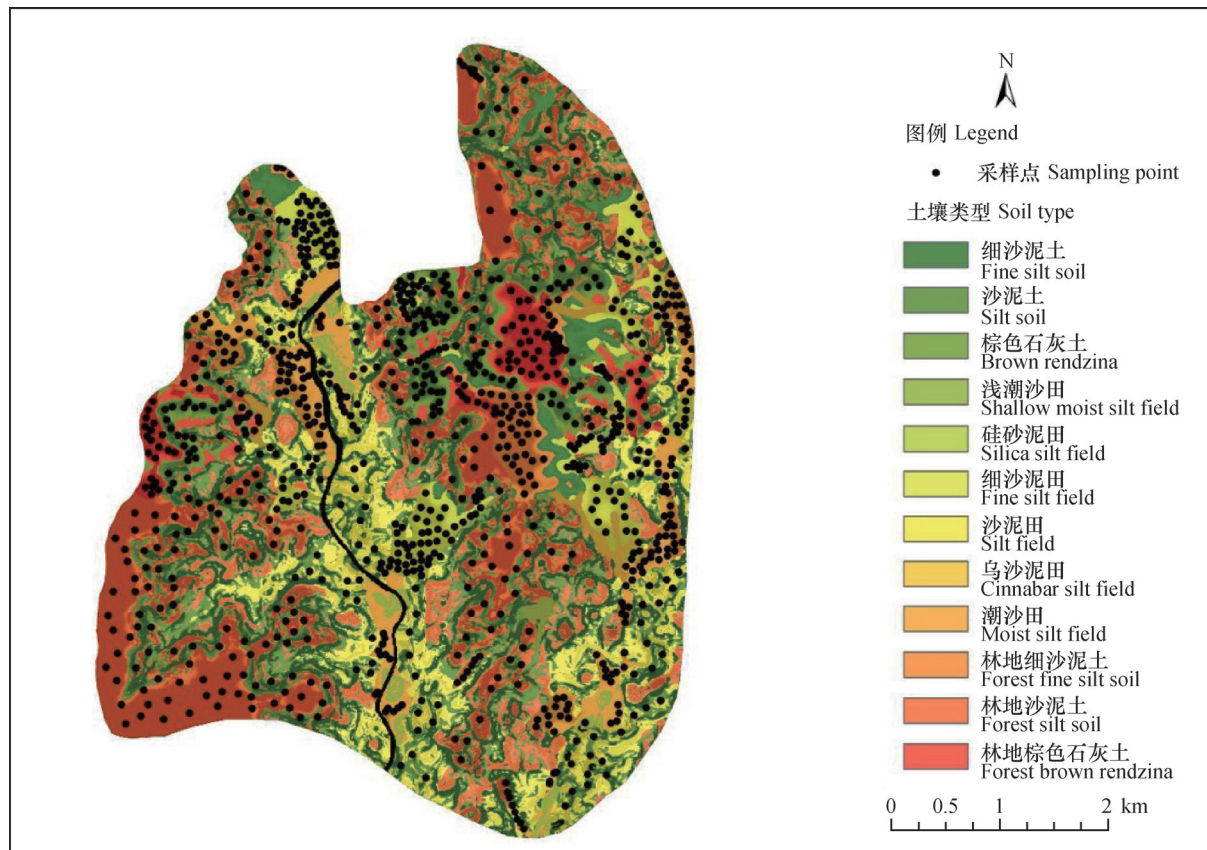


图5 样点分布图

Fig. 5 Distribution of sampling point

表1 局部规则对比

Table 1 Comparison of local rules

优化后规则 Updated rules	原始规则 Original rules
母质=片麻岩	母质=片麻岩
: ...高程 \leq 80.53	: ...高程 \leq 121.95
: : ...湿度指数 \leq 3.85	: : ...坡度 \leq 19.54
: : : ...曲率 \leq 0.042 9: 沙泥田	: : : ...坡向 \leq 213.42
: : : ...曲率 $>$ 0.042 9: 沙泥土	: : : : ...湿度指数 \leq 3.63 : 沙泥土
: : ...湿度指数 $>$ 3.85: 沙泥田	: : : : ...湿度指数 $>$ 3.63 : 沙泥田
: ...高程 $>$ 80.53 and 高程 \leq 121.83	: : : ...坡向 $>$ 213.42
: : ...曲率 \leq -0.001 3: 林地沙泥土	: : : : ...湿度指数 \leq 3.63 : 林地沙泥土
: : ...曲率 $>$ -0.001 3: 沙泥土	: : ...坡度 $>$ 19.54 : 林地沙泥土
: ...高程 $>$ 80.53 : 林地沙泥土	: ...高程 $>$ 121.95 : 林地沙泥土

四个环境要素不同条件下的组合加以明显区分，其均由片麻岩发育而来，沙泥土基本分布在坡脊部位，表面地形呈凸向变化，湿度较小，林地沙泥土分布在坡顶或坡上岗顶以下，局部表面地形呈凹向变化，沙泥田则主要分布在坡肩及以下部位，地形变化较为平缓，湿度较大。结合实际地形地貌，坡脊位置由于风蚀，雨蚀等影响，大量土壤颗粒和水分在重力作用下往地势较低的坡肩及平缓地带流动并堆积，形成较湿润的土壤环境。浅潮沙田由近代河流冲积物发育而来，高程较低，主要分布在坡脚部位，局部表面地形呈凹向变化，用曲率替换了原始规则中的坡向，说明优化后的土壤环境关系更能反映土壤—环境特征。

2.4 推理制图及精度验证

采用优化后的规则对土壤类型的空间分布进行推测，结果如图6所示，其中左边为原始规则推理图，右边为规则优化后的推理图。对比规则优化前后的土壤类型图，可以发现整体上呈现出一致性趋势，潮沙田、林地棕色石灰土、林地沙泥土以及沙泥土分布范围基本一致，潮沙田由近代河流冲积物发育而来，主要分布在河流两岸以及西部地区；林地棕色石灰土分布于北部丘陵地区，林地沙泥土在整个研究区域内均有分布，并与沙泥土相互夹杂，这是由于两种土壤类型的成土环境相似，因此这两种土壤往往相伴而生。局部细节中，相比于原始规则推理的土壤图，沙泥田与沙泥土在优化后的土壤图中图斑明显增多，空间细节更加丰富，且地处低山丘陵区河流冲积平原，精细程度与河流的分布规

律一致，由此可见土壤类型的空间分布与景观特征相吻合，这也进一步验证了土壤类型空间分布的准确性。

利用野外实地采集的独立验证点对原始推理图和优化后的推理图进行精度评估，用于精度检验的野外实地样点共253个，分别采用横截面采样（28个）、主观采样（109个）和均匀采样（116个）这三种采样方式，通过土壤专家实地辨别的方法获取各样点的土壤类型。对比优化前后推理图，优化后的推理图正确分类的样点个数为219个，总体精度为86.9%，高于原始规则的推理图（73%）；而Kappa系数为0.842，高于0.8，表示一致性程度为显著，能较好地反映研究区的土壤分布情况。

利用SoLIM软件建立混淆矩阵，以生产精度和用户精度为指标验证优化后的土壤图精度。优化后推理图与原始推理图的生产精度和用户精度统计结果如表2所示，除细沙泥土外，研究区内其他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，细沙泥田、林地沙泥土、沙泥田、浅潮沙田这四类土壤的生产精度和用户精度均有明显提高，这是由于林地沙泥土分布在陡峭山体或山坡位置，海拔相对较高，与地形变化呈现协同一致趋势；细沙泥田、沙泥田、浅潮沙田这三类土壤主要分布于河谷平原以及沟谷地带，地形最为平坦，海拔较低，湿度较小，随着研究区基本单元精度的提高（10m分辨率），更易展现空间细节。沙泥土、林地棕色石灰土、潮沙田等土壤精度也有所提升，这主要归因于原始成图过程中被合并忽略的图斑被还

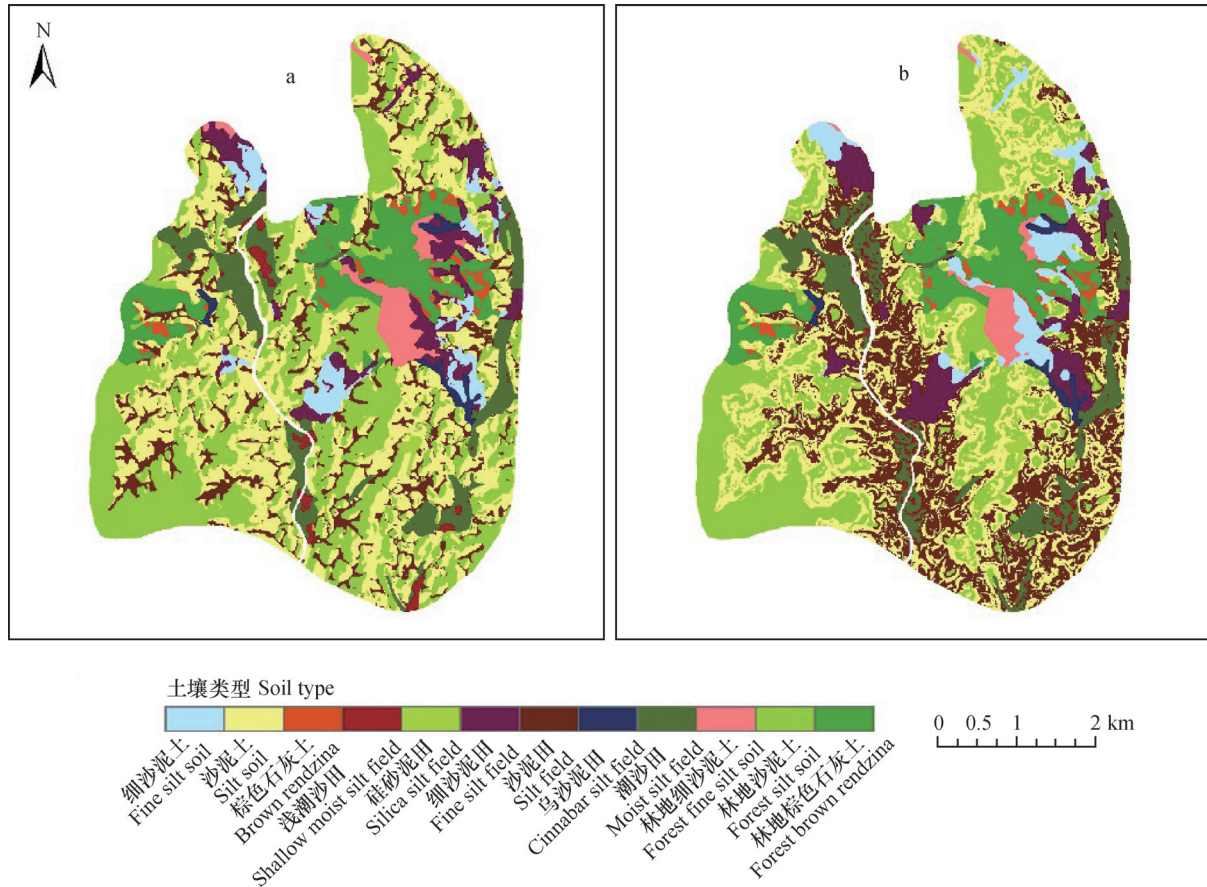


图6 土壤类型栅格图 (a为原始规则推理图, b为规则优化后的推理图)

Fig. 6 Soil type grid maps (a is the map inferred by original rules, and b is the map inferred by optimized rules)

表2 优化后推理图与原始推理图的各土壤类型生产精度和用户精度比较

Table 2 Comparison between the original and updated soil maps in accuracy of soil types for production and for users (%)

土壤类型 Soil type	生产精度 Accuracy for production		用户精度 Accuracy for users	
	原始推理图 Original soil map	优化后推理图 Updated soil map	原始推理图 Original soil map	优化后推理图 Updated soil map
	细沙泥土 Fine silt soil	59	67	83
林地细沙泥土 Forest fine silt soil	67	79	100	100
细沙泥田 Fine silt field	80	90	40	69
沙泥土 Silt soil	82	86	60	78
林地沙泥土 Forest siltsoil	76	86	84	92
沙泥田 Silt field	47	84	68	84
棕色石灰土 Brown rendzina	50	50	50	100
林地棕色石灰土 Forest brown rendzina	96	100	96	96
浅潮沙田 Shallow moist siltfield	60	80	75	100
潮沙田 Moist silt field	95	100	90	95

原。优化后的推理图总分类精度高于原始推理图约13%，预测精度有明显提高。

本研究中提出的土壤—环境知识获取方法虽然取得了较好效果，但仍存在一些问题需要深入探讨。在构建GIS数据库时，采用的环境因子只包括了成土母质及高程、坡度等6个因子，缺乏对地貌数据（径流系数）、植被等更多环境因子的筛选。而能够反映土壤与环境之间关系的环境因子有很多，对于小区域尺度下，土地利用方式也会对土壤类型的判别与推理产生影响。因此，科学筛选环境因子，并应用到推理制图中，使得提取的土壤—环境关系知识更加准确，这将作为后期的研究重点之一。在知识获取过程中，样点采集的数量相比于用决策树进行挖掘数据时所需要的样点明显减少了很多，但是否还可以再减少样点数而达到同样的效果，降低研究过程中的工作量。此外，用于精度验证的样点均通过野外实地调查所得，这些样点受到研究者可达性的限制，因此样点的数量和分布可能会对精度产生一定影响，是否可以利用高精度高分辨率的遥感影像图来验证推理土壤图的精度，使结论更具客观性、科学性。同时，今后将进一步通过实例验证该方法在其他研究区的适用性。

3 结 论

本研究提出了一种基于不确定模型的土壤—环境关系知识获取方法，将不确定性图与土壤类型图进行融合，然后在可信度高的地方进行重采样，进一步挖掘可靠的土壤—环境关系知识，采用SoLIM进行制图并利用野外独立验证点进行评价。结果表明：优化后的土壤图展示了更加详尽的空间分布信息，一些在原始成图过程中被合并忽略的图斑被还原，同时本文所提方法的推理制图结果更符合研究区土壤分布的特征；此外，应用该方法推理制图，其精度高于已有土壤图精度约13%。因此，本研究提出的基于不确定性模型的土壤—环境关系知识获取方法，可为土壤普查和制图等相关研究提供理论依据和方法。

参 考 文 献

- [1] 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: 21—26
Zhu A X, Li B L, Pei T, et al. Model and method of
- fine digital soil survey (In Chinese). Beijing: Science Press, 2008: 21—26
- [2] Zhu A X. A personal construct-based knowledge acquisition process for natural resource mapping. *International Journal of Geographical Information Science*, 1999, 13 (2): 119—141
- [3] Moran C J, Bui E N. Spatial data mining for enhanced soil map modeling. *International Journal of Geographical Information Science*, 2002, 16: 533—549
- [4] Chang P C, Fan C Y, Dzan W Y. A CBR-based fuzzy decision tree approach for database classification. *Expert Systems with Applications*, 2010, 37 (1): 214—225
- [5] 杨琳, 朱阿兴, 秦承志, 等. 基于典型点的目的性采样设计方法及其在土壤制图中的应用. *地理科学进展*, 2010, 29 (3): 279—286
Yang L, Zhu A X, Qin C Z, et al. A purposive sampling design method based on typical points and its application in soil mapping (In Chinese). *Progress in Geography*, 2010, 29 (3): 279—286
- [6] 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
- [7] 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 (3): 885—894
- [8] 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
- [9] 黄魏, 罗云, 汪善勤, 等. 基于传统土壤图的土壤—环境关系提取与预测制图研究. *土壤学报*, 2016, 53 (1): 72—75
Huang W, Luo Y, Wang S Q, et al. Acquisition of soil-environment relationships based on conventional soil map and predictive mapping (In Chinese). *Acta Pedologica Sinica*, 2016, 53 (1): 72—75
- [10] 黄魏, 韩宗伟, 罗云, 等. 基于地形单元的土壤有机质空间变异研究. *农业机械学报*, 2015, 46 (4): 161—167

- Huang W, Han Z W, Luo Y, et al. Spatial distribution of soil organic matter based on topographic unit (In Chinese). Transactions of the Chinese Society for Agricultural Machinery, 2015, 46 (4): 161—167
- [11] 卢岩君, 秦承志, 邱维理, 等. 基于少量典型样点土壤属性空间分布推测模型中的土壤属性参数敏感性分析. 地理科学, 2011, 31 (12): 1549—1554
- Lu Y J, Qin C Z, Qiu W L, et al. Sensitivity analysis of soil property parameter in typical-sample-based prediction model of digital soil mapping (In Chinese). Scientia Geographica Sinica, 2011, 31 (12): 1549—1554
- [12] Zhu A X. Measuring uncertainty in class assignment for natural resource maps under fuzzy logic. Photogrammetric Engineering and Remote Sensing, 1997, 63 (10): 1195—1202
- [13] 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
- [14] Zhu A X. A similarity model for representing soil spatial information. Geoderma, 1997, 77 (2): 217—242
- [15] Burt J E, Harrower M. Depicting fuzzy soil class uncertainty using perception-based color models. Proceedings of the Eleventh International Fuzzy Systems Association World Congress, 2005: 2
- [16] Fisher P, Comber A, Wadsworth R. Approaches to uncertainty in spatial data. Fundamentals of Spatial Data Quality, 2006 (2): 43—59
- [17] Qi F, Zhu A X. Comparing three methods for modeling the uncertainty in knowledge discovery from area-class soil maps. Computers & Geosciences, 2011, 37 (9): 1425—1436
- [18] Zhu A X, Hudson B, Burt J, et al. Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Science Society of America Journal, 2001, 65 (5): 1463—1472

Extraction of Knowledge about Soil-environment Relationship Based on an Uncertainty Model

HUANG Wei XU Wei WANG Shanqin YUAN Yaping WANG Chenyun

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

Abstract 【Objective】 Digital soil mapping is based on Jenny's classic theory, of which the core is that soil is the yield of the interactions among numerous soil forming factors. To establish relationships between environmental variables and soil attributes, a soil-landscape model is built up and used to predict soil types or attributes. Although this model has been extensively used in digital soil mapping, investigations are still on the way on what the relationships between environmental variables varying with the region and soil attributes are. Therefore, how to extract rapidly and accurately knowledge of soil environment has become the key to the current researches in this field. Knowledge acquisition based on sampling points is often affected by the number of sampling points, errors in sampling processes and representativeness of sampling points. Traditionally, soil mapping is based on manual soil surveys and tends to have errors in the following two fields, i.e. enclosure and displacements of boundaries. The knowledge acquired from soil maps cannot be used to predict local soil conditions, and especially lacks details specific to soil grade and issues the users are interested in. To solve these problems and attain soil environment information high in accuracy, a method based on knowledge intercomplementation and fusion is set forth. 【Method】 This paper uses decision tree coupled with a uncertainty model to extract soil environment information. Through limiting threshold values, the decision tree model can be used to predict soil type rapidly and efficiently. With the soil lowering in type level, the prediction lowers steadily in accuracy. Therefore, ignored uncertainty and exaggerated uncertainty can be used to judge truly and efficiently accuracy of the inferred map to a certain extend and

hence to realize accurate quantitative evaluation of the inferred map. So coupling of the two models can not only save money and time, but also raise efficiency and realize scientific re-extraction and fusion of soil information. The NieshuiRiverBasin in Huajiahe Town, Hongan County, Huanggang City of Hubei Province was selected as a case for study. The proposed method proceeded in three steps. 1) By means of the standard See 5 algorithm, decision trees were constructed and used to extract soil-environment information and hence spatial distribution for soil mapping; 2) With the aid of the SoLIM software, a spatial distribution map of exaggerated and ignored uncertainties was plotted. The two kinds of uncertainties appeared in the processes of classifying a geographic entity, i.e. ignored uncertainty, which is attributed to the similarity of the studied soil to all the soil types, and exaggerated uncertainty, which is associated with the deviation of the studied soil from the prototype specified in the processes of soil hardening. A similarity model can be used to estimate the two uncertainties; and 3) Soil samples were collected again based on the spatial distribution map of the uncertainties. The higher the values of exaggerated and ignored uncertainties, the higher the probability of a soil being mis-classified. So resampling should be performed in location low in uncertainty value. The soil environment information acquired from the resampled soil sample set combined and updated or optimized the original knowledge. In the end, soil mapping by inferring was performed with the aid of the SoLIM software and based on the eventually obtained soil-environment knowledge, and validated with the validation set of 253 sampling sites in the field for accuracy. **【Result】** Results show that the soil map plotted through inferring contains more specific spatial distribution information and reaches up to 86.9% in accuracy as validated with the field validation sites. Obviously it is 13% higher than the soil maps so far available. Moreover, its Kappa coefficient is 0.842, higher than 0.8, indicating its high degree of consistency, and conformity with the attributes and spatial distribution of the soils in the study area. **【Conclusion】** It is, therefore, concluded that the proposed method which acquire the knowledge of soil-environment relationship using the uncertainty models feasible and effective. This method not only increases spatial detailedness of the soil map, but also improves its accuracy.

Key words Digital soil mapping; Decision tree; Uncertainty; Knowledge fusion

(责任编辑：檀满枝)