

基于 HSI 模型的土壤氮素空间预测 不确定性的可视化表达*

檀满枝^{1, 2} 陈杰^{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 中国科学院研究生院, 北京 100049)

摘要 任何模型都无法完全准确地再现真实世界, 因此各种空间预测始终面临着不确定性的困扰。从 20 世纪 90 年代开始, 土壤空间预测研究中输出结果的不确定性分析与评价日益受到重视。由于不确定性信息的可视化表达有助于更加直观地揭示与展示真实世界, 因此是一种增强不确定分析与决策支持之间联系的有效途径。本文首先对不确定性可视化表达的主要技术与方法进行了简要论述, 进而以北京市城市边缘带样区基于地统计学的土壤表层全氮含量的空间预测作为案例研究, 以相对误差表示预测结果的不确定性, 应用色调-饱和度-亮度(HSI)色彩模型实现土壤属性预测结果和不确定信息的可视化同步表达。在 HSI 模型中, 色调值被用来表示预测值, 白度表示不确定性。在案例研究的全氮预测输出结果中, 颜色发白的区域表示预测结果的不确定性较高, 需要做进一步补充采样。实际应用过程中发现, 二维图例对不确定性的可视化表达效果有一定的限制作用, 仍需在二维图例完善或创建新型图例方面开展深入的研究工作。

关键词 土壤空间预测; 不确定性; 可视化表达; HSI 模型

中图分类号 TP302; S158 文献标识码 A

20 世纪 80 年代以来, 数学模型开始广泛应用于土壤和其他景观属性的空间预测研究。土壤作为地球表面一个极为复杂的非线性动力学系统(Nonlinearity Dynamics System, NDS), 任何一种模型都无法完全精确地描述和反映真实世界中土壤的空间可变性, 因此空间预测中的不确定性(uncertainty)永远存在。空间不确定性具有两个方面的内涵: 属性不确定性与空间位置不确定性, 产生于土壤属性数据测定、统计学计算以及空间预测模型建立等各个环节。更为重要地, 外在(原始数据)与内在(模型)不确定性通过空间预测模型传播, 导致最终输出结果以及对输出结果进行误差评估同样蕴涵不确定性^[1]。从 20 世纪 90 年代开始, 对土壤空间预测结果的不确定性评价日益受到重视, 其评价结果越来越广泛地被纳入诸如污染土壤空间界定、土壤适宜性分区等决策支持层面^[2]。判断土壤空间预测制图质量的逻辑标准是信息含量, 它由土壤属性信息含量、不确定性分析、信息识别潜力等多个内容组成。其中不确

定性的可视化表达是决定土壤图件的最关键因素之一。随着 GIS 技术的不断发展, 不确定性空间表达正在成为土壤制图一个极为重要的新内涵。总体而言, 不确定性的可视化表达划分为两种不同方式: 一是独立可视化, 不确定性空间表达作为土壤空间预测制图独立的附属图件(adjacent map); 二是同步可视化, 实现不确定性分析与土壤空间预测结果在同一幅图件上表达。

直到 20 世纪 90 年代中期, 土壤空间预测研究中的不确定性分析的最主要手段是非线性克里格方法, 它是通过评价一个特定非采样点的目标属性大于阈值的概率并设置概率置信区间而实现的, 不确定分析结果通常无法与土壤空间预测结果实现同步可视化表达^[3-6]。同步可视化表达通常由以下两种途径实现: 在土壤空间预测图层上创建表达不确定信息的新图层, 或创建蕴涵不确定信息的土壤空间预测图层。前者主要以透明度(transparency)、晕渲(haze)、雾状(fog)、边界虚化(blurring)、纹理(texture)

* 国家自然科学基金项目(40701070, 40571065)和中国科学院南京土壤研究所创新前沿项目(ISSASIP0716)资助

† 通讯作者, E-mail: jchen@issas.ac.cn

作者简介: 檀满枝(1978~), 女, 安徽望江人, 助理研究员, 博士研究生。主要从事土壤资源演变、土壤空间预测及土壤调查制图研究。已发表论文 10 余篇

收稿日期: 2006-11-24; 收到修改稿日期: 2007-06-12

和图案(patterns)等为主要手段,通过改变下覆土壤空间预测图层信息的可视效果实现不确定性表达^[7-10]。而后者则通过改进空间预测制图技术,将不确定性信息融合于同一可视图层中。2003年 Hengl^[11]在进行土壤表层厚度空间预测研究中,首次应用 HSI 色彩模型实现了土壤表层厚度空间预测结果与不确定分析在同一图层上的表达,国内这一领域的研究尚未见报道。本文以北京市边缘带样区表层土壤全氮空间预测研究为案例,以相对误差表示预测结果的不确定性,应用 HSI 色彩模型实现土壤全氮预测结果和不确定信息的可视化同步表达。

1 材料与方法

1.1 样品采集与分析

研究区位于北京城近郊 2 600 km²城市边缘带

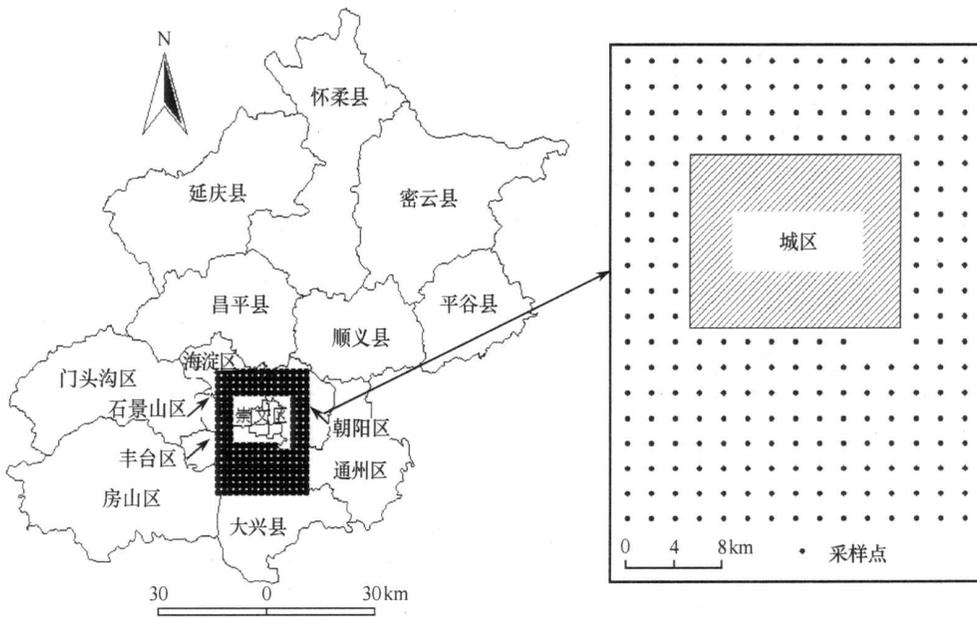


图 1 研究区样点分布图

Fig 1 Distribution of soil sampling sites in study area

1.2.2 HSI 模型 RGB 色彩立方体中与连接黑色(0, 0, 0)和白色(255, 255, 255)对角线正交的截面,即为 HSI 色彩模型,它是一视觉心理引力(psychologically appealing)色彩模型。沿着对角线有无数个这样的截面,每一个对应一个亮度(Intensity),或者称为明度(图 2a)。任何一个亮度值都有一个轮子形状的调色板,称为“HSI 色盘(colour wheel)”,它的亮度相同,色调(Hue)和饱和度(Saturation)不同。

内(图 1),土壤样品采集工作完成于 2001 年 4 月~ 5 月。以 2 km× 2 km 为网格单元系统布点采样,采集 0~ 20 cm 土壤表层样品,共完成 220 个样点的采集,基本覆盖研究区域。采样点用 GPS 精确定位,并进行周围景观描述记录。土壤样品在室内自然风干,分别过 10 目、60 目、100 目筛。全氮采用半微量开氏法测定。

1.2 研究方法

1.2.1 地统计学方法 运用 GS+ for windows 软件计算出研究区全氮的半方差函数值,然后在 ArcGIS 软件中采用普通克里格方法(Ordinary kriging),对研究区未采样区域的土壤全氮含量进行最优、无偏估值,得到土壤全氮空间预测结果分布图和预测标准误差图。

色调代表对颜色类型的视觉感知,通过围绕对角轴线的程度进行计算。饱和度代表色调颜色的程度,通过距离对角轴线的半径距离进行计算(图 2a)。亮度代表对明度的视觉感知。白度(whiteness),即白色的数量,在 RGB 颜色立方体中表示为任意点到白色顶点的最短距离,随饱和度值和亮度值呈线性变化。Jiang⁽¹⁾证实白度是表示不确定性合理的色彩变量,因此 HSI 模型是对不确定性进行可视化的理想模型。

(1)Jiang B. Fuzzy Overlay Analysis and Visualization in Geographic Information Systems. PhD Thesis. Utrecht: University of Utrecht, 1996

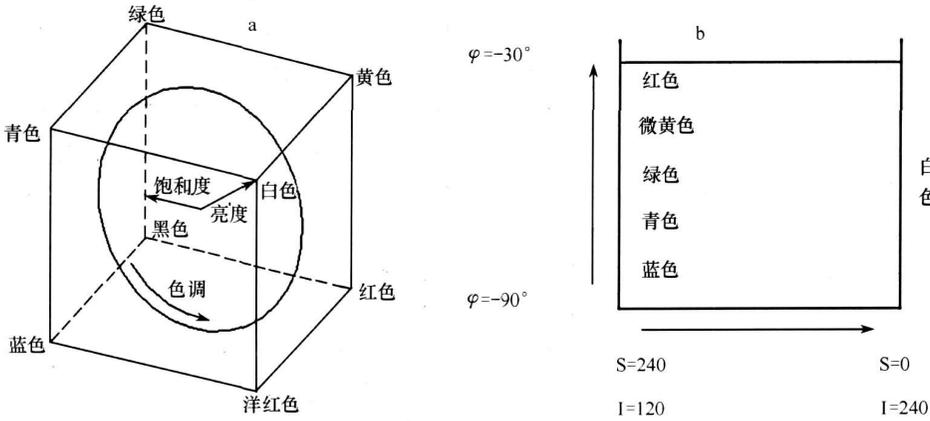


图 2 RGB 颜色立方体中 HSI 色彩模型(a) 和连续变量预测不确定性可视化的二维图例(b)

Fig. 2 HSI colour model in the RGB colour cube (a) and the two dimensional legend used for visualization of uncertainty in continuous variable predictions (b)

1.2.3 不确定性表达方法 用预测标准误差(prediction standard error)的标准偏差(简称相对预测误差)表示不确定性:

$$\sigma_{E,r}(S_0) = \frac{\sigma_E(S_0)}{S_z} \times 100\% \quad (1)$$

式中, $\sigma_E(S_0)$ 表示预测标准误差, S_z 表示观察样本的标准偏差。满意的预测结果是模型能够解释 85% 以上的总体变异。经验认为^[11-13] $\sigma_{E,r}(S_0)$ 值接近 40% 时意味着相当满意的预测精度。相反, 如果值达到 80% 以上, 在验证样点中模型解释不到 50% 的变异, 预测结果是不理想的。

在进行制图之前, 首先对预测结果和相对预测误差进行极差标准化处理:

$$z_r = \frac{\hat{z} - z_1}{z_2 - z_1}; \quad u_r = \frac{\sigma_{E,r} - u_1}{u_2 - u_1} \quad (2)$$

式中, z_r 为标准化预测值, u_r 为相对预测误差, 范围在 [0, 1] 之间。 \hat{z} 为预测结果图, $\sigma_{E,r}$ 为相对预测误差图。 z_1 和 z_2 分别为预测值的最小值和最大值, u_1 和 u_2 为限制不确定性(相对预测误差)的阈值, 分别为 40% 和 100%。

基于 HSI 模型, 用色调值表示预测值, 白度值表示不确定性, 同步可视化表达预测图和预测误差图。用式(3)~ 式(5)进行 HSI 色彩计算:

$$\varphi_1 = -90 + z_r \times 300 \quad (3)$$

$$\varphi_2 = \begin{cases} \varphi_1 + 360 & \text{if } \varphi_1 \leq -360 \\ \varphi_1 & \text{if } \varphi_1 > -360 \end{cases} \quad (4)$$

式中, φ_1 为顺时针色调度数, φ_2 为 φ_1 转换到 [-360, 360] 范围的值。HSI 的值分别为:

$$H = (\varphi_2 + 360) \times \frac{240}{360} \quad (5)$$

$$S = (1 - u_r) \times 240 \quad (6)$$

$$I = (1 + u_r) \times 120$$

1.2.4 图例的设计 设计二维图例, 纵轴方向用不同的色调值表示预测结果, 横轴方向用白度值表示预测的不确定性。纵轴上最低色调值确定为带蓝色(色调值范围为 -90° 至 -150°), 最高值为带红色(色调值范围为 -330° 至 -30°)。中间值为青色、绿色和微黄色。横轴上, 白度值随饱和度值和亮度值从低到高呈线性变化(图 2b)。

应用 RGB-HSI 转化公式, 将 HSI 图转化为 RGB 图, 在 GIS 中显示。结合可视化能够表达的不确定性和给定阈值之间的关系, 原始 RGB 彩色图像(original RGB)表示最小的不确定性等于或低于 40% (随给定阈值变化而变化), 全白色表示最大的不确定性等于 100%, 其他颜色表示不确定性处于二者之间。为了避免混淆预测高值和低值区域, 图例中省去了色调圆轮上代表洋红色(-30° 至 -90°)的部分。

2 结果与讨论

2.1 描述性统计

从表层土壤全氮描述性统计分析(表 1)来看, 研究区表层土壤全氮变异性较大, 经单样本 K-S 检验, 数据符合正态分布。

2.2 地统计学分析

研究区表层土壤全氮的实验半方差函数与球状模型拟合最好, R^2 为 0.926。表层土壤全氮有效变程为 19.67 km, 块金系数为 40% (表 2), 说明全氮空间相关范围较大, 但在该范围内具有中等的空间相关性。

表 1 土壤表层全氮基本统计学特征

Table 1 Descriptive statistics for topsoil total N (g kg^{-1})

项目 Item	平均值 Mean	最小值 Minimum	最大值 Maximum	标准差 Std. deviation	变异系数 CV (%)	K-S 值 K-S value
(g kg^{-1})						
全氮 Total N	0.82	0.20	1.74	0.29	40	0.92

表 2 土壤表层全氮各向同性半方差函数理论模型及相关参数

Table 2 Theoretical semivariogram models of topsoil total N and their corresponding parameters

理论模型 Theory model	块金值 Nugget C_0 ($\text{g}^2 \text{kg}^{-2}$)	基台值 Sill $C_0 + C$ ($\text{g}^2 \text{kg}^{-2}$)	块金值/基台值 Nugget/Sill $C_0 / C_0 + C$ (%)	变程 Range (km)	R^2
球状模型 Spherical model	0.042	0.095	40	19.67	0.926

2.3 全氮空间预测结果和不确定性表达

采用普通克里格插值对全氮进行空间预测, 交叉验证结果显示, 标准预测误差(MSE)为 0.004 4, 均方根误差(RMSE) 0.246 2 和平均标准误差(ASE)

0.236 8 较接近, 均方根标准误差(RMSSE) 近似等于 1。平均预测误差(ME)为 0.001 3, 总体而言, 预测精度较高。

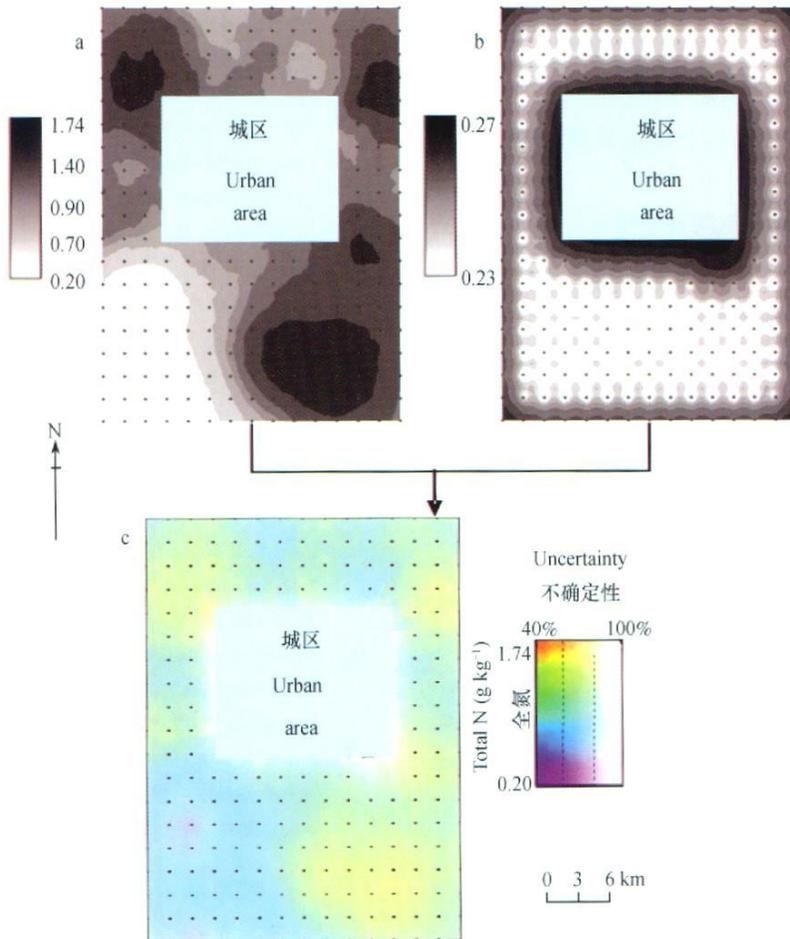


图 3 土壤表层全氮空间预测图(a)、预测标准误差图(b)和全氮空间预测结果及不确定性同步表达图(c)

Fig. 3 Prediction map of topsoil total N (a), prediction standard error map (b), simultaneous visualization of both prediction and prediction uncertainty of total N for topsoil (c)

土壤表层全氮空间分布呈明显的片状和斑块状特点,高值区主要分布在研究区东南、东北和西北角(图 3a)。制图结果可为指导区域合理施肥和有效控制农业面源氮污染提供参考。预测标准误差图表现出典型的粒状模式,离采样点越远,不确定性越大(图 3b)。同步可视化图(图 3c)显示带蓝色低预测值主要分布在样区的西南角,带青色和绿色的中预测值零散分布在样区中,带黄色高预测值有规律地分布于样区的东南部、东北角和西北角。预测结果分布图同图 3a。不确定性最高的发白区域位于城区边上,其他区域不确定性相对较高,基本处于 75% 左右。

3 结 论

交叉验证结果表明,本案例研究的空间预测输出结果精度较高,但同时不确定性也相对较高,相对误差基本处于 75% 左右。预测标准误差受采样密度和样点布设结构影响,采样密度越大,样点分布越规则,估值标准差越小,不确定性越低,预测结果越可靠。本案例区样点分布均匀,但样点间距大,相对误差较大,提高采样密度是降低预测结果不确定性的有效途径。

在实现不确定性与空间预测结果同步表达过程中,本研究采用的二维图例对可视化效果具有不容忽视的限制作用,其主要原因是图像中白色成分难以区别,导致图例颜色与 HSI 图像颜色之间直观的对对应关系难以建立。因此,相对误差在 $\pm 5\%$ 范围的实际不确定无法在可视化制图上区分。鉴于此,图例的完善应该成为下一步研究工作的重点。

土壤空间预测过程中不确定分析与表达不以削减不确定性为目标,而是利用现有的信息帮助判断模型输出结果的可信赖程度并直观地予以展示,这对于基于预测结果的实际应用决策而言非常重要。以本研究为例,如果基于土壤表层全氮空间预测结果进行施肥管理决策,那么首先应该决定在什么样的风险水平上或者不确定性区间内对研究区特定地块的土壤施用氮肥,当特定地块内土壤全氮高的不确定性特别高或者特别低,相应的针对性施肥决策

难度就较小;而对于中等不确定性的区域,制定施肥方案就会相对困难得多。

参 考 文 献

- [1] Mowrer H T, Congalton R G. Quantifying Spatial Uncertainty in Natural Resources: Theory and Applications for GIS and Remote Sensing. Hannover: Ann Arbor Press, 2000. 350
- [2] Goovaerts P. Geostatistical modelling of uncertainty in soil science. *Geodema*, 2001, 103: 3~ 26
- [3] Smith J L, Halvorson J J, Papendick R I. Using multiple variable indicator kriging for evaluating soil quality. *Soil Sci. Soc. Am. J.*, 1993, 57: 743~ 749
- [4] Webster R, Oliver M A. Optimal interpolation and isarithmic mapping of soil properties: VI. Disjunctive kriging and mapping the conditional probability. *J. Soil Sci.*, 1989, 40: 497~ 512
- [5] Goovaerts P, Journel A G. Integrating soil map information in modelling the spatial variation of continuous soil properties. *Eur. J. Soil Sci.*, 1995, 46: 397~ 414
- [6] 李艳, 史舟, 王人潮, 等. 海涂土壤剖面电导率的协同克立格法估值及不同取样数目的比较研究. *土壤学报*, 2004, 41(3):434~ 443. Li Y, Shi Z, Wang R C, *et al.* Estimates of electrical conductivity for coastal saline soil profile using cokriging under different sampling density (In Chinese). *Acta Pedologica Sinica*, 2004, 41(3): 434~ 443
- [7] Dutton G. Handling positional uncertainty in spatial databases. *In: Proceedings 5th International Symposium on Spatial Data Handling*. University of South Carolina, August 1992. 460~ 469
- [8] MacEachren A M. Visualizing uncertain information. *Cartographic Perspective*, 1992, 13: 1 019
- [9] Monmonier M. Strategies for the interactive exploration of geographic correlation. *In: Proceedings of the 4th International Symposium on Spatial Data Handling*, Vol. 1. IGU, July 1990. 512~ 521
- [10] Pang A, Fuman J, Nuss W. Data quality issues in visualization. *In: Robert J, Moorhead II, Deborah E, et al. eds. SPIE Vol. 2 178 Visual Data Exploration and Analysis*, SPIE, February 1994. 12~ 23
- [11] Hengl T, Heuvelink G B M, Stein A. A generic framework for spatial prediction of soil variables based on regression kriging. *Geodema*, 2004, 120: 75~ 93
- [12] Hengl T. Visualisation of uncertainty using the HSI colour model: Computations with colours. *In: 7th International Conference on Geocomputation (CD ROM)*. 2003. 8
- [13] Park S, Vlek P. Environmental correlation of three dimensional soil spatial variability: A comparison of three adaptive techniques. *Geodema*, 2002, 109 (1/2): 117~ 140

VISUALIZATION OF UNCERTAINTY ASSOCIATED WITH SPATIAL PREDICTION OF TOTAL N IN TOPSOIL USING HSI MODEL

Tan Manzhi^{1,2} Chen Jie[†]

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

(2 *Graduate School of the Chinese Academy of Sciences, Beijing 100049, China*)

Abstract Spatial prediction is always facing challenges of uncertainty because it is almost impossible for any model to represent exactly the real world. Since the 1990s, more and more attention has been given to analysis and evaluation of uncertainties in output of spatial soil prediction. Visualization can be a way to increase communication between uncertainty analysis and decision-making, since it can help expose and express reality more intuitively. First of all, different techniques or ways to visualize uncertainty were briefly reviewed. And a case study was conducted of the statistics-based spatial prediction of total N content in the topsoil of peri-urban Beijing. Relative errors were deemed as expression of uncertainty in the prediction. A Hue-Saturation-Intensity (HSI) color model, which is a psychologically appealing color model, was recommended to visualize simultaneously soil attribute prediction results and uncertainties therein. The HSI model uses hue to visualize prediction values and whiteness to visualize uncertainty. A two-dimensional legend was designed to supplement the visualization. This case study indicated that visualization of both prediction and prediction uncertainty offers a possibility to enhance visual exploration of data uncertainty and compare different prediction methods or predictions of totally different variables. The whitish area of the visualized output can be simply interpreted as unsatisfactory predictions, where additional samples may be needed for a better prediction. The limitation of using the two-dimensional legend is that it is not easy to match the colors between the HSI coded image and the legend because the pale colors are always difficult to distinguish.

Key words Soil spatial prediction; Uncertainty; Visualization; HSI model