

# 稀疏格网样点的国家尺度土壤属性制图方法与应用

余万里<sup>1,2</sup>, 李宝林<sup>1\*</sup>

(1. 中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室, 北京 100101;

2. 中国科学院研究生院, 北京 100049)

**摘要:** 国家尺度土壤属性数据是地球生物化学循环及水循环等领域研究的重要数据, 目前, 该尺度土壤属性数据的获取方法主要有两类: 土壤属性-空间数据连接法和空间插值。为了确定哪一类方法更适合稀疏样点的国家尺度土壤属性制图, 本文以中国吉林省的土壤有机质含量制图为例, 采用 8~32km 格网样点和 1:100 万土壤图, 对这两类方法进行对比分析。独立样本验证结果表明, 土壤属性-空间数据连接法的平均误差(ME)大于距离反比加权(IDW)插值, 而平均绝对误差(MAE)和均方根误差(RMSE)都小于 IDW 插值。IDW 插值获得的土壤属性图虽然能大致反映土壤属性空间分布的基本规律, 但出现了类似“牛眼睛”的空间结构, 且存在无样点区估计值不准确等问题; 土壤属性-空间数据连接法尽管忽略了同种土壤类型内部的差异, 保留了不同土壤类型边界处的属性值突变, 但获得的土壤属性图更能反映土壤属性分布的基本规律, 也具有比较详细的土壤属性空间结构。因此, 在基于稀疏样点的国家尺度土壤属性制图中, 土壤属性-空间数据连接法的制图效果要优于 IDW 空间插值法。

**关键词:** 土壤属性制图; 稀疏格网样点; 国家尺度; 土壤属性-空间数据连接法; 空间插值

**DOI:** 10.3724/SP.J.1047.2012.00049

## 1 引言

土壤在地球生物化学循环和水循环中具有重要作用, 全球、大陆或国家尺度土壤属性数据已被广泛用于碳储量估计<sup>[1-3]</sup>、作物产量模拟<sup>[4]</sup>和全球气候变化模拟<sup>[5]</sup>等研究。为获取可靠的土壤属性分布信息, 国内外开展了多项大陆或国家级的土壤调查(尤其是地球化学基线调查), 如英格兰和威尔士的国家土壤目录<sup>[6]</sup>(NSI)、英国地球化学基线调查(G-BASE)、法国国家土壤监测网<sup>[7]</sup>(RMQS)、欧洲地理调查<sup>[8]</sup>(FOREGS)、北美土壤地球化学景观项目, 以及近年来我国开展的全国土壤污染状况调查等。

这些调查大多具有区域大、样点稀疏以及格网采样方式为主的特点。目前, 以这种数据进行土壤属性制图的方法主要有两类: 土壤属性-空间数据连接法和空间插值法。土壤属性-空间数据连接法是将代表性的土壤剖面数据与土壤图中的具体土

壤类型连接起来<sup>[9-11]</sup>, 空间插值法利用样点的空间关系对土壤属性进行空间推测, 包括克里格插值<sup>[6]</sup>、IDW 插值<sup>[12]</sup>以及滑动加权中位数<sup>[8]</sup>等方法。

一些研究曾对这两类方法进行了比较分析<sup>[13-15]</sup>, 结果表明这两类方法的制图精度没有明显差异。其研究区域范围较小、样本比较密集, 而在研究区域范围较广、样点分布稀疏的条件下缺乏相关研究。因此, 为了确定哪一类方法更适合稀疏格网土壤采样数据条件下的国家尺度土壤属性制图, 本文以吉林省的土壤有机质含量制图为例, 采用 8~32km 格网采样土壤数据和 1:100 万土壤图, 对两类方法进行对比分析。

## 2 研究数据与方法

### 2.1 研究区及数据

#### 2.1.1 研究区

吉林省的气候是温带季风气候, 自东南向西

收稿日期: 2011-10-01; 修回日期: 2012-01-09.

基金项目: 资源与环境信息系统国家重点实验室创新项目。

作者简介: 余万里(1987-), 男, 湖北省宜昌人, 硕士研究生, 研究方向为资源详查与自动制图。E-mail: yuwl@reis.ac.cn

\* 通讯作者: 李宝林(1970-), 男, 博士, 研究员, 研究方向: 遥感环境变化检测、生态环境质量评估与土壤环境信息系统技术。E-mail: libl@reis.ac.cn.

北,从湿润半湿润区过渡为半干旱区。年平均气温 2.5~5℃,年降水量 400~1 300mm。研究区海拔范围 5~2 552m,地势从东南向西北降低,东部主要为山地和丘陵,中部和西部分别以台地和平原为主。研究区内的土壤主要为地带性土壤,自东向西分别为棕壤、暗棕壤、黑土、黑钙土和栗钙土;其他非地带性土壤主要包括草甸土、风沙土、水稻土、沼泽土、碱土和盐土等。

2.1.2 研究数据

(1)土壤属性数据

研究区布设了 1851 个土壤样点,包括 1765 个建模点和 86 个验证点。建模点为分层格网样点,格网大小由土地利用类型决定,其中,耕地为 8km×8km,草地和林地 16km×16km,未利用地为 32km×32km。样点选取在格网中心,若无法获取(如落在湖中或者离道路太近)则对样点位置做微调或者删除。研究区西部的建模点密度高于东部。验证点约每 30km×30km 格网内采集 1 个,样点确定原则为:(1)选择格网内面积最大的土类;(2)选取该土类的典型位置。每个调查样点采集 0~20cm 土样,测定土壤有机质含量、pH 值、全氮、全磷、全钾及土壤质地(粉粒、粘粒和砂粒),本文以土壤有机质含量为例。

(2)土壤图

土壤图采用全国 1:100 万土壤图<sup>[16]</sup>,其为第二次全国土壤普查期间(1979-1995)依据不同比例尺的地方土壤图及实验室分析数据编纂而成。土壤图的分类系统,研究区共包含 19 个土类、44 个亚类和 99 个土属,本文使用土类这一级别。

2.2 土壤属性-空间数据连接法

本文采用土壤属性-空间数据连接法中的转换函数法(Taxotransfer Function)确定 1:100 万土壤图中每个多边形(图斑)的土壤属性值<sup>[10]</sup>,共包括以下 4 条规则:

规则 1:如果 1 个多边形包含 10 个以上(含 10 个)同种土类的样点,则将这些样点的土壤属性值的中位数作为该多边形的土壤属性值。这里选择中位数是因为对于对数正态分布的土壤属性值,中位数优于算术均值<sup>[17]</sup>。

规则 2:如果 1 个多边形包含少于 10 个同种土类的样点,则按照 Shangguan 等<sup>[18]</sup>估计中国土壤质地分布图时所采用的方法,扩大搜索范围,使用

离多边形最近的 10 个同种土类样点土壤属性值的中位数作为该多边形的土壤属性值。

规则 3:如果整个研究区内某种土类的样点少于 10 个,则将所有该土类样点土壤属性值的中位数作为所有该土类多边形的土壤属性值。

规则 4:如果研究区内没有某种土类的样点,则使用研究区外相邻省份同种土类的样点土壤属性值的中位数来估计该种土类多边形的土壤属性值。

2.3 IDW 插值方法

空间插值利用样点间的空间关系来估计未知点的土壤属性值,Schloeder 等<sup>[19]</sup>研究认为稀疏样点条件下,IDW 插值与克里格插值的精度差异不大,而样条函数法精度相对较低;另外 IDW 插值对数据限制较少,因而本文以 IDW 插值法为空间插值法的代表,比较其与土壤属性-空间数据连接法对土壤属性制图的影响。IDW 插值计算如公式(1)、(2)所示:

$$P(x,y)=\sum_{i=1}^nW_i\cdot P_i \tag{1}$$

其中: $P(x,y)$ 表示待推测点 $(x,y)$ 的土壤属性估计值, $n$ 代表参与计算的相邻样点数目, $W_i$ 是土壤样点 $i$ 的权值, $P_i$ 代表土壤样点 $i$ 的属性值。权值 $W_i$ 由公式(2)计算:

$$W_i=d_i^{-a}/\sum_{k=1}^nd_k^{-a} \tag{2}$$

其中: $d_i$ 是公式(1)中点 $(x,y)$ 与土壤样点 $i$ 之间的欧氏距离, $a$ 是指数因子, $a$ 越大距离近的样点权值越高。

2.4 验证

本文使用平均误差(ME)、平均绝对误差(MAE)和均方根误差(RMSE)3 项指标评价制图精度。ME 反映方法的无偏性,MAE 和 RMSE 体现估计值的精度。公式如下:

$$\left\{\begin{aligned} ME &= \frac{1}{n} \sum_{i=1}^n (\hat{Z}_i - Z_i) \\ MAE &= \frac{1}{n} \sum_{i=1}^n |\hat{Z}_i - Z_i| \\ RMSE &= \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Z}_i - Z_i)^2} \end{aligned}\right. \tag{3}$$

其中: $Z_i$ 是验证点 $i$ 的土壤属性观测值, $\hat{Z}_i$ 是估计

值, $n$  为验证点数量。

3 土壤属性制图与结果分析

3.1 土壤属性-空间数据连接法制图与分析

图 1 为土壤属性-空间数据连接法得到的土壤有机质含量分布图。矩形框部分对应图 3(a),椭圆形勾画区域为一个土类为暗棕壤的图斑。

如图 1 所示,东部湿润森林区发育的棕壤和暗棕壤,其有机质含量较高,中部半湿润台地发育的黑土有机质含量中等,而西部半干旱草原区发育的黑钙土有机质含量较低,体现了土壤有机质含量自东向西递减的地带性分布规律。该图同时也反映了土壤的非地带性分布规律,例如,东部地区发育的白浆土虽然属于草原土壤,但有机质含量与周围发育的暗棕壤相当;发育在东部河谷平原的草甸土有机质含量明显低于周边的棕壤和暗棕壤,而在西部则明显高于周围的黑钙土;在西部沙丘上发育的风沙土,有机质含量则明显低于周边黑钙土等其它草原土壤。

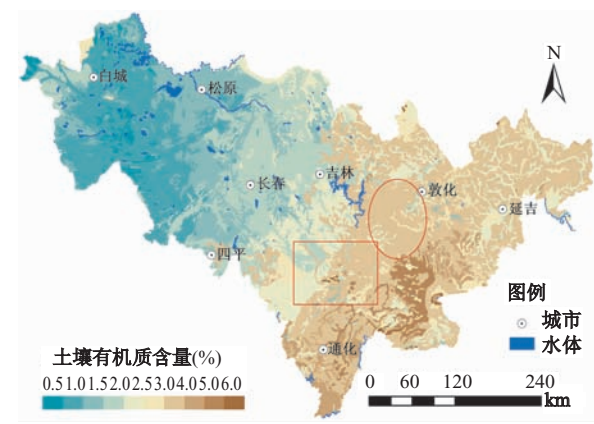


图 1 土壤属性-空间数据连接法得到的土壤有机质含量分布图  
Fig. 1 Map of soil organic matter (SOM) content estimated by the soil linkage method

3.2 IDW 插值法生成的土壤有机质含量分布图

图 2 为 IDW 插值法得到的土壤有机质含量分布图,图中矩形区域对应图 3(b),圆圈勾出区域为典型的“牛眼睛”结构(bull’s eye)<sup>[20]</sup>(“牛眼睛”是指由于建模点附近区域的值整体性突然升高或降低,从而呈现出的类似于一组同心圆的图案)。

该图大致反映了研究区土壤有机质含量分布

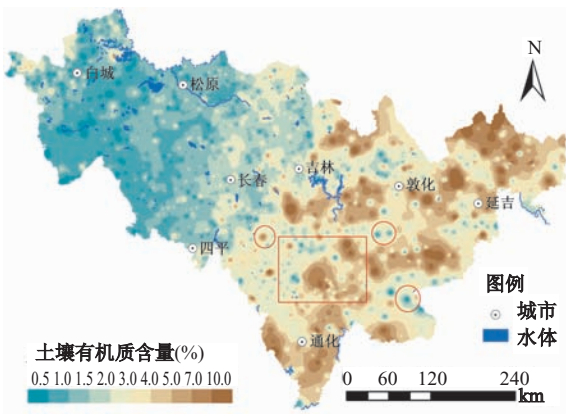


图 2 IDW 插值法得到的土壤有机质含量分布图  
Fig. 2 Map of SOM content derived from IDW interpolation

的总体规律,即土壤有机质含量自东向西递减。但土壤属性空间结构,尤其是相对较精细的土壤属性的空间结构没有能够得到准确的反映。

3.3 验证结果与分析

86 个独立样本点的验证结果如表 1 所示。土壤属性-空间数据连接法的 ME 为-0.58, IDW 插值为 0.13,表明 IDW 插值结果无偏性较好。土壤属性-空间数据连接法的 MAE 和 RMSE 分别为 1.12 和 1.76, IDW 插值分别为 1.43 和 2.12。所以,土壤属性-空间数据连接法的估计值比 IDW 插值结果更为可靠。

图 3 为两种制图方法的详细对比。(a)和(b)分别对应图 1 和图 2 中的矩形框区域,(c)为该区域的土类图。

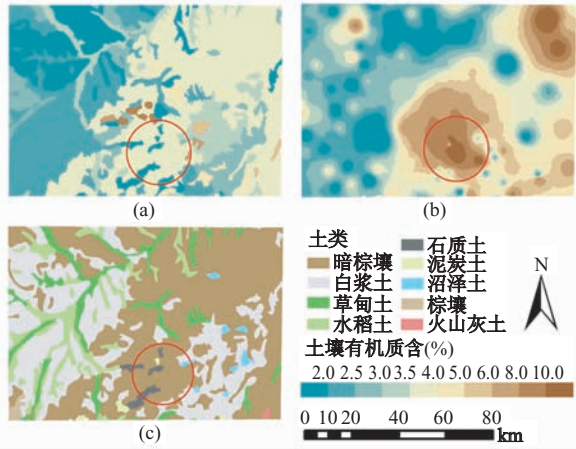


图 3 两种方法的制图效果  
Fig. 3 Mapping details by the two methods

表 1 两种制图方法验证结果

Tab. 1 Validation results for maps produced by these two methods

土壤属性-空间数据连接法	IDW 空间插值
ME	-0.58
MAE	1.12
RMSE	1.76

土壤属性-空间数据连接法获得的土壤属性图对样点稀少或者没有样点区域的估计值较为可靠。例如,图 3 中圆圈勾出的石质土分布区,尽管分布范围较小且没有采样点分布,但土壤有机质含量明显低于周围的暗棕壤,较准确地反映了土壤有机质的基本分布规律。同时,这种方法获得的土壤属性图可以反映这种分布范围较小且没有调查样点的土壤类型分布区的土壤属性特征,因此对土壤属性空间分布的描述也较精细,结果也比较可靠。

该方法获得的土壤属性图存在两个主要问题:(1)忽略了同种土类内的土壤属性差异;(2)不同土类边界处的土壤属性值存在突变。例如,在研究区东部,一个土类为暗棕壤的图斑包含 100 多个样点(图 1 中椭圆标识的区域),但其内部的土壤属性差异未能体现,因为土壤属性-空间数据连接法中,对图斑只赋予一个典型值。另外,明显可以看到不同土类边界处土壤有机质含量的突变,这也是由土壤图的硬分类加上每个图斑都只有一个典型值所导致。

IDW 插值方法获得的土壤属性图主要存在以下三个问题:(1)出现了类似“牛眼睛”的土壤属性空间分布结构;(2)不同土类边界处的属性值估计不准确;(3)样点稀少或无样点区域的估计值不可靠。

IDW 插值中,建模点的权值仅仅由与之的空间距离决定,所以,每个建模点对周围区域的影响范围是各向同性的。对于属性值明显高于或者低于周围背景值的建模点,与之临近的区域都受到该点的影响,从而导致类似同心圆一样的图案环绕在建模点周围,形成类似“牛眼睛”的空间结构。

土壤有机质含量的分布与土壤类型有密切关系,通过 IDW 插值获得的土壤属性图,在不同土壤类型交界处,有机质含量与土壤类型的空间结构已产生非常大的差异,如图 3 中的水稻土和白浆土。这种情况的出现主要是在 IDW 插值中,边界

附近区域的土壤属性值由周围各种不同土壤类型的样点插值得来,且样点的影响区域为各向同性,所以获得的土壤属性图不能反映不同土类边界处土壤属性分布的基本空间规律。

稀疏样点条件下,许多区域没有或者只包含较少样点。对于面积较小又没有样点分布的土壤类型,其有机质含量与其周边土壤类型已没有明显的差异。如图 3 所示,石质土的土壤有机质含量应该比相邻的暗棕壤低,但 IDW 插值的预测结果却与暗棕壤相似。这是因为 IDW 插值中,对于缺少或者没有样点的区域,仅依靠空间关系并不可靠,导致偏离典型值的不正常估计。

IDW 插值只利用了样点间的空间关系,而对于 8~32km 网格的建模样点而言,难以通过单纯的空间关系精确地反映土壤属性空间分布信息,所以,其估计值尽管无偏性较好但误差较大。土壤属性-空间数据连接法从土壤发生规律出发,制图结果符合土壤属性分布的基本规律,特别是对于样点稀少或者没有样点的区域,其估计值比 IDW 插值更为可靠,因而其 MAE 与 RMSE 比 IDW 空间插值低。

4 结 语

本文对土壤属性-空间数据连接法与空间插值法在稀疏网格样点数据条件下的国家尺度土壤属性制图进行了对比分析。独立样本验证结果中,土壤属性-空间数据连接法的 ME 为-0.58, IDW 插值为 0.13,表明 IDW 插值法无偏性较好。土壤属性-空间数据连接法的 MAE 和 RMSE 分别为 1.12 和 1.76,低于 IDW 插值(1.43 和 2.12),显示其制图精度高于 IDW 插值。对比两种方法所得到的土壤属性图,土壤属性-空间数据连接法的结果更符合土壤发生规律,具有更为详细的土壤属性空间结构,对样点稀少或没有样点区域的估计值也较为可靠;IDW 插值法所获得的土壤属性图,其空间结构的表达不如土壤属性-空间数据连接法精细,且不同土类的边界处,以及无样点区域的估计值不如前者可靠。对比两类方法,在稀疏网格样点数据的国家尺度土壤属性制图中,土壤属性-空间数据连接法的制图效果优于空间插值法。

在稀疏调查样点的情况下,土壤属性-空间数据连接法依据土壤发生原理,估计结果体现了土壤属性分布的空间分布规律,特别是对大尺度采样调

查中样点稀少或者没有样点的区域,估计值更为可靠;空间插值法由于只利用了土壤样本之间的空间关系,而在样点稀疏的条件下,难以精确反映土壤属性的空间变异,插值结果较为平滑,对样点稀少或无样点区域估计可靠性较差。

稀疏格网样点数据的国家尺度土壤属性制图中,依据土壤发生规律比依据空间自相关性制图更可靠,但土壤属性-空间数据连接法不能从原理上保证无偏,容易受到数据和方法的影响<sup>[3,21]</sup>。另外,基于点位数据的验证方法可能不完全适合粗分辨率土壤属性数据的验证,还有待进一步改善。

### 参考文献:

- [1] 王绍强,周成虎. 中国陆地土壤有机碳库的估算[J]. 地理研究,1999,18(4):349-356.
- [2] 解宪丽,孙波,周慧珍,等. 中国土壤有机碳密度和储量的估算与空间分布分析[J]. 土壤学报,2004,41(1):35-43.
- [3] 于东升,史学正,孙维侠,等. 基于 1:100 万土壤数据库的中国土壤有机碳密度及储量研究[J]. 应用生态学报,2005,16(12):2279-2283.
- [4] Gijsman A J, Thornton P K, Hoogenboom G. Using the WISE database to parameterize soil inputs for crop simulation models[J]. Computers and Electronics in Agriculture, 2007, 56(2): 85-100.
- [5] Maeda E E, Pellikka P K E, Siljander M, *et al.* Potential impacts of agricultural expansion and climate change on soil erosion in the Eastern Arc Mountains of Kenya [J]. Geomorphology, 2010, 123(3~4): 279-289.
- [6] Oliver M A, Loveland P J, Frogbrook Z L, *et al.* Statistical and geostatistical analysis of the national soil inventory of England and Wales[R]. The Technical Report, Defra. 2002.
- [7] Jolivet C, Boulonne L, Ratié C. Manuel du Réseau de Mesures de la Qualité des Sols, édition 2006[R], Unité InfoSol, INRA Orléans, France, 2006.
- [8] Salminen R, Batista M J, Bidovec M, *et al.* FOREGS Geochemical Atlas of Europe, Part 1: Background Information, Methodology and Maps [M]. Geological Survey of Finland, Otamedia Oy, Espoo, 2005.
- [9] Batjes N H. Total carbon and nitrogen in the soils of the world[J]. European Journal of Soil Science, 1996, 47(2): 151-163.
- [10] Batjes N H. A world dataset of derived soil properties by FAO-UNESCO soil unit for global modeling[J]. Soil Use and Management, 1997, 13(1): 9-16.
- [11] Cerri C E P, Easter M, Paustian K, *et al.* Predicted soil organic carbon stocks and changes in the Brazilian Amazon between 2000 and 2030[J]. Agriculture, Ecosystems & Environment, 2007, 122(1): 58-72.
- [12] British Geological Survey. G-BASE: Environmental geochemical atlas of central and eastern England[M]. Keyworth, Nottingham, UK. 2010.
- [13] Bregt A K, Bouma J, Jellinek M. Comparison of thematic maps derived from a soil map and from Kriging of point data[J]. Geoderma, 1987, 39: 281-291.
- [14] Brus D J, de Gruijter J J, Marsman B A, *et al.* The performance of spatial interpolation methods and choropleth maps to estimate properties at points: A soil survey case study[J]. Environmetrics, 1996, 7: 1-16.
- [15] Voltz M, Webster R. A comparison of kriging, cubic splines and classification for predicting soil properties from sample information[J]. European Journal of Soil Science, 1990, 41(3): 473-490.
- [16] 席承潘. 中国土壤[M]. 北京:中国农业出版社,1998.
- [17] Snedecor W, Cochran W. Statistical methods, 7th edition[M]. Iowa State University Press, Ames Iowa, 1980.
- [18] Wei S G, Dai Y J, Liu B Y, *et al.* A soil particle-size distribution dataset for regional land and climate modelling in China[J]. Geoderma, In Press, Corrected Proof, 2011.
- [19] Schloeder C A, Zimmerman N E, Jacobs M J. Comparison of methods for interpolation soil properties using limited data[J]. Soil Science Society of America Journal, 2001, 65: 470-479.
- [20] de Smith M J, Goodchild M F, Longley P A. Geospatial analysis: a comprehensive guide to principles, techniques and software tools[M], Matador, 2007.
- [21] 张勇,史学正,于东升,等. 属性数据与空间数据连接对土壤有机碳储量估算的影响[J]. 地球科学进展, 2008, 23(8):840-847.

# A Study on Soil Property Mapping at National Scale Based on Sparse Grid Soil Samples

YU Wanli<sup>1,2</sup> and LI Baolin<sup>1</sup>

- (1. *State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China;*
- 2. *Graduate University of Chinese Academy of Sciences, Beijing 100049, China*)

**Abstract:** Soil property data at national scale are necessary inputs for researches on biogeochemical and hydrological cycles. There are two commonly used methods to obtain accurate soil property data at this scale including soil linkage method and spatial interpolation method. In this paper, soil organic matter (SOM) content in Jilin Province was taken as a case study to evaluate these two methods based on 8~32 km grid soil samples and 1 : 1 million soil map, in order to provide suggestions about soil property mapping at national scale based on sparse grid soil samples. Independent validation indicated that the mean error (ME) of the soil linkage method was  $-0.58$ , higher than the inverse distance weighted (IDW) interpolation method ( $0.13$ ), but the mean absolute error (MAE) and root mean square error (RMSE) of the soil linkage method were  $1.12$  and  $1.76$ , respectively, lower than those of IDW interpolation method ( $1.43$  and  $2.12$ ). Although the soil property map through IDW interpolation revealed the general trend of soil property distribution, it had “Bull’s eye” pattern and the estimations in areas with few or no soil samples were unreliable. The soil property map obtained by soil linkage method ignored the variations inside the same soil type on the soil map and a break of soil property existed around different soil boundaries, but the soil distribution regularity was well represented in this map and the details could be obtained. According to the validation results of maps through these two methods, the soil linkage method is better than the spatial interpolation method in soil property mapping based on sparse grid soil samples at national scale.

**Key words:** soil property mapping; sparse grid samples; national scale; soil linkage method; spatial interpolation