

DOI: 10. 11766/trxb201309130415

# 应用 SGS 和 LHS 分析数字土壤质量评价中的不确定性\*

孙孝林<sup>1,2</sup> 王会利<sup>1†</sup> 曹继钊<sup>1</sup>

(1 广西优良用材林资源培育重点实验室, 广西林业科学研究院, 南宁 530002)

(2 中山大学地理科学与规划学院, 广州 510275)

**摘要** 土壤信息不确定性在空间分析中的传递是数字土壤评价中的关键问题。本文提出使用序贯高斯模拟(Sequential Gaussian simulation, SGS)和拉丁超立方抽样(Latin hypercube sampling, LHS)相结合的方法(即 SGS-LHS), 来应对该问题, 目的在于充分利用 SGS 和 LHS 各自的优点, 互补各自的缺点, 以提高不确定性传递分析的准确性和效率。这种方法(包括两种途径: SGS-LHS1 和 SGS-LHS2)和 SGS、LHS 一起被应用于香港农田土壤质量评价中, 并进行了比较。结果表明: (1) SGS-LHS 分析所得的不确定性结果与 SGS 接近一致, 与 LHS 则有一定的差别, 但差别不大; (2) SGS-LHS 估计不确定性的准确性与 SGS 接近一致, 且两者均高于 LHS, 尽管 LHS 估计的置信度区间平均宽度略显精确。

**关键词** 序贯高斯模拟; 拉丁超立方抽样; 土壤质量; 不确定性; 数字土壤评价

**中图分类号** S159; S158.9; P934

**文献标识码** A

精准土壤信息是土壤资源可持续利用和发展的基础。该方面的研究长期以来一直是土壤学、地理信息系统和其他相关学科的重要研究方向。尤其是最近的十多年, 数字土壤制图研究蓬勃发展, 显著地提高了土壤信息的准确性<sup>[1]</sup>。这是因为: 一方面, 数字土壤制图利用地理信息系统、数据挖掘技术和详细的环境变量数据, 提高了土壤信息的精确性; 另一方面, 数字土壤制图在提供土壤信息的同时, 还提供这些信息的准确性和不确定性(即误差), 避免土壤信息用户不恰当地使用实际并不准确的土壤信息而蒙受损失<sup>[1-2]</sup>。

准确分析土壤信息的不确定性对土壤信息的应用非常重要。首先, 土壤信息的不确定性直接决定土壤信息的质量和可用性。Malone 等<sup>[3]</sup>将不确定性和准确性同时作为数字土壤图的质量评价指标; Carré 等<sup>[2]</sup>保守地建议不使用误差较大的土壤信息; Sun 等<sup>[4]</sup>发现空间土壤质量评价中的不确定性较大, 认为有必要提高土壤信息的质量。其次,

准确分析土壤信息中的不确定性来源, 有助于通过进一步的研究来降低其中的不确定性, 从而提高信息的质量。Brown 和 Heuvelink<sup>[5]</sup>指出, 信息的不确定性是与生俱来、不可避免的, 分析不确定性来源的目的并不在于消除误差, 而在于识别较大的误差来源, 再通过进一步研究来降低较大误差来源中的误差, 最终降低总体误差。例如, 张淑杰等<sup>[6]</sup>根据不确定性设计土壤采样方案, 并证明了这种采样设计方法可以有效地提高土壤信息的准确性。最后, 土壤信息的不确定性具有传递性, 能够在土壤信息的进一步应用过程中被放大, 或缩小, 最终使得应用结果也具有未知的不确定性<sup>[2]</sup>。Goovaert<sup>[7]</sup>表明, 准确分析土壤信息的不确定性有助于评估决策制定带来的风险, 例如过高的土壤污染预测会造成土壤修复中的资源浪费。为此, Carré 等<sup>[2]</sup>特别强调数字土壤图应用过程中的不确定性分析问题, 并认为该问题是数字土壤评价中的关键问题。

尽管土壤信息的应用非常广泛, 例如空间土壤

\* 国家自然科学基金项目(41201216)、广西优良用材林资源培育重点实验室开放课题(12A0405)和香港研究资助局(2002-PPR-3)资助

† 通讯作者, E-mail: wanghuili6@163.com

作者简介: 孙孝林(1980—), 男, 湖北荆州人, 博士, 副教授, 主要研究数字土壤制图、土壤资源评价和土壤信息系统。E-mail: sun\_xiaolin@yahoo.com

收稿日期: 2013-09-13; 收到修改稿日期: 2014-04-15

质量评价 (Qi 等<sup>[8]</sup>)、土壤侵蚀评价 (Castrignano 等<sup>[9]</sup>)、碳储量估算 (Minasney 等<sup>[10]</sup>)、土壤水文动态模拟 (李润奎等<sup>[11]</sup>)、土壤转换函数 (廖凯华等<sup>[12]</sup>),但只有少数研究注意到了土壤信息中的不确定性及其在空间分析中的传递问题。例如, Sun 等<sup>[4]</sup>在香港农田土壤数字制图的基础上,将序贯高斯模拟 (Sequential Gaussian Simulation, SGS)<sup>[13]</sup>和蒙特卡罗 (Monte Carlo)方法估计的土壤信息进行随机配对,并输入到土壤质量评价模型中,从而分析土壤质量评价结果中的不确定性;Buttafuoco 等<sup>[14]</sup>采用同样的方法分析了土壤可侵蚀性评估中的不确定性;廖凯华等<sup>[12]</sup>利用拉丁超立方抽样 (Latin Hypercube Sampling, LHS)<sup>[15]</sup>分析了土壤饱和和导水率空间预测中的不确定性。土壤信息不确定性很少被注意的主要原因有两个:(1)主观上,大多数研究人员尚未充分认识信息的不确定性问题;(2)客观上,分析信息不确定性及其传递不容易实现。随着信息的广泛应用,信息所包含的不确定性会逐渐得到大范围的认识,因而第一个原因会得以解决。第二个原因则需要更多的研究。

事实上,空间不确定性研究由来已久,已形成了多种不确定性分析方法,例如蒙特卡罗、泰勒级数 (Taylor Series Method, 又称一阶分析 (First-order analysis))、拉丁超立方抽样法、自举法 (Bootstrap Method)、随机模拟等<sup>[16]</sup>。其中,SGS 和 LHS 已在土壤信息的应用中得以使用,即 Sun 等<sup>[4]</sup>、Buttafuoco 等<sup>[14]</sup>和廖凯华等<sup>[12]</sup>。比较这几种方法,可以发现:SGS 一般需要大量估计,例如 1 000 次模拟,因而数据量非常庞大,计算强度大,效率低;LHS 仅做少量抽样,例如 50 次,计算强度小,效率高,但在抽样过程中,完全忽略了空间信息的自相关性,准确性有所下降。为此,本文提出将这两种方法进行结合,即在土壤信息不确定性分析时使用 SGS,在信息应用过程中分析不确定性传递时使用 LHS。显然,这种方法的结合既考虑了信息的空间自相关性,又减轻了计算强度,因而既保持了准确性又提高了效率。本文以土壤质量评价为例,在 Sun 等<sup>[4]</sup>的基础上,将这两种方法的结合进行了应用,并与这两种方法的单独使用进行了比较。

## 1 材料与方 法

### 1.1 研究区概况

研究区位于香港北部 (图 1),总面积为 110.6

km<sup>2</sup>。该区地形极其复杂,被周边的山丘分隔成多个平坦的区域。气候属于亚热带气候,年均气温约 23℃,年降雨量约为 2 214 mm。该区的城市化程度较高,35% 的土壤被建筑物、道路、公园、设施、铁路等封闭,且这些封闭的土壤与剩下 65% 未封闭的土壤极其复杂地混杂在一起。约 1/2 的未封闭土壤目前处于耕种状态,主要种植蔬菜、水果和花卉;剩下的未封闭土壤则处于休耕状态。根据骆永明等<sup>[17]</sup>,该区的土壤类型主要有:铁铝土、淋溶土、人为土、锥形土和新成土;成土母质类型主要有:第四纪红土、花岗岩及河海相沉积物。

### 1.2 数字土壤制图

Sun 等<sup>[4]</sup>对土壤制图及不确定性分析作了详细介绍,本文仅概述其中的主要过程。土壤采样分三个阶段进行。第一阶段,以研究区的土系图为基础,根据土系在研究区内的分布状况,对每个土系进行采样。采样点的个数根据土系的面积来决定。该阶段共采集 41 个土壤剖面样品,如图 1 所示。第二阶段,考虑样点在研究区上的空间分布以及第一阶段的采样点的位置,利用 Walvoort 等<sup>[18]</sup>开发的 SPCOSA 包,在 R 程序中选择了包含第一阶段采样点在内的 150 个采样点。这种采样方法的目的在于最小化克里格制图方差。然而,在第二阶段的实际采样过程中,共有 17 个样点由于被建筑物覆盖或位于边境管制区域内而无法采样。因此,第二阶段实际采集了 92 个土壤表层样品,如图 1 所示。第三阶段,利用 SPCOSA 包和 R 程序,随机选择 35 个采样点,并采集了土壤表层样品,如图 1 所示。第一、二阶段的采样用于土壤制图;第三阶段则用于不确定性分析结果的验证。

对土壤剖面的 A 层样品和其他表层样品进行了多项指标的测定,即:pH、容重、机械组成、有机质、全氮、全磷、全钾、铵态氮、有效磷、速效钾、阳离子交换量、有效铜、有效铁、有效锰、有效锌、有效硼,测定方法依据鲁如坤<sup>[19]</sup>。各项指标的统计特征详见 Sun 等<sup>[4]</sup>。对这些指标采用普通克里格方法制图,并在制图过程中,对异常值 (Outlier) 采用了 Lark<sup>[20]</sup>建议的稳健估计量方法进行了剔除。然而,由于容重、阳离子交换量、有效铜、有效锰的空间异质性较强,不能模拟出合格的变异函数。因此,对这些指标采用了 Liu 等<sup>[21]</sup>提出的反距离加权与土壤图 (Inverse distance weighted combined with soil map-delineation, IDW-SMD) 相结合的方法进行制图。该方法首先根据采样结果计算出每个土壤类型

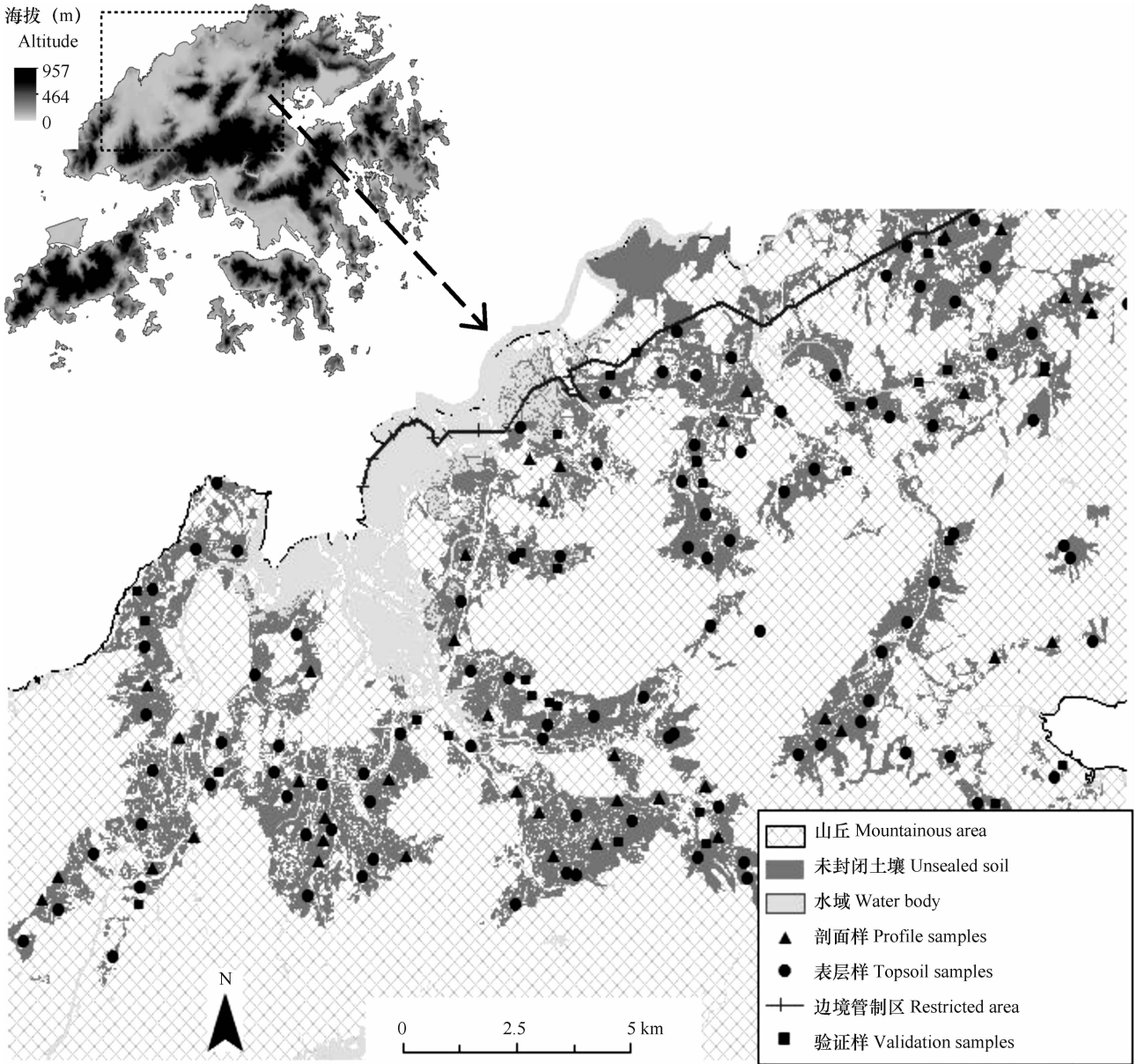


图 1 研究区及采样点分布  
 Fig. 1 Study area and sampling sites

的平均值并把该值对应地赋给研究区上的每个点, 再对采样点上的土壤属性值与对应的土壤类型平均值之间的差值进行 IDW 插值, 最后将插值结果与土壤类型平均值加和即得最后的制图值。此外, 由于土壤剖面的采样个数有限, 土壤 A 层厚度和阻滞层深度的制图采用了传统土壤制图方法, 即将每个土系的土壤 A 层厚度和阻滞层深度平均值赋给整个土系。

条件(Conditioned)SGS 用于分析普通克里格制图的不确定性, 对每个属性随机模拟 500 次; 随机高斯模拟用于分析 IDW-SMD 制图的不确定性, 模拟参数设为土系的平均值和方差, 对每个属性随机模

拟 500 次, 不包括不合理值, 例如负值、加和值超过 100% 等; 传统土壤制图方法的不确定性因没有合适的方法进行分析, 因而在下文中未考虑其不确定性。

### 1.3 土壤质量评价模型

根据研究区的实际情况, Sun 等<sup>[4]</sup>从土壤剖面、地形、水文、物理、植物营养五个方面, 选取了 22 个土壤质量指标, 它们是: A 层厚度、阻滞层深度、特定汇流面积、饱和导水率、pH、容重、砂粒、粉粒、黏粒、有机质、全氮、全磷、全钾、铵态氮、有效磷、速效钾、阳离子交换量、有效铜、有效铁、有效锰、有效锌、有效硼。其中, 特定汇流面积从研究区的数字高程模

型(图 1)中计算得到;饱和导水率利用 Vereecken 等<sup>[22]</sup>研究的转换函数,从有机质、容重、黏粒和粉粒中计算得到。

每个指标均根据研究文献和研究区的实际情况,采用模糊评分方法进行打分;随后,在土壤制图采样的基础上,通过因子分析计算得出每个指标在土壤质量中的权重;最后,根据打分结果和权重计算综合土壤质量指数。详细的评分函数及权重见 Sun 等<sup>[4]</sup>。

#### 1.4 LHS、SGS 和 SGS-LHS

LHS 是 McKay 等<sup>[15]</sup>在蒙特卡罗法的基础上发展形成的分层随机抽样方法,广泛应用于多维变量的实验设计、仿真模拟及不确定性分析中。与简单随机抽样相比,LHS 能更完全地利用样本空间,更显著地节省样本数。有研究表明,LHS 仅需要 1.3 倍~3 倍变量个数的抽样数,就能得到很好的分析结果<sup>[23]</sup>。LHS 的一般流程为:(1)根据变量的累计概率分布,将变量的域等概率地划分成  $n$  个互不重叠的区间,使得这些区间上的抽样概率相同,即为  $1/n$ ;(2)在每个区间上,随机抽取任一值;(3)将所有变量的  $n$  个抽样值随机配对,共构成  $n$  组变量;(4)将这  $n$  组变量输入到空间模型中进行运算,从而估计不确定性。本研究考虑了 pH、容重、砂粒、粉粒、黏粒、有机质、全氮、全磷、全钾、铵态氮、有效磷、速效钾、阳离子交换量、有效铜、有效铁、有效锰、有效锌、有效硼的不确定性,共 18 个变量。LHS 抽样数设为 50。

SGS 与克里格基本相似,不同之处在于 SGS 同时利用预测点上的预测值进行预测,并随机抽取以预测值、预测方差为正态分布的任意值为最终预测值<sup>[13]</sup>。可以看到,SGS 在模拟土壤属性值时,不仅利用了空间点之间的条件关系(Spatial conditioning),还利用了随机高斯抽样,使得空间点上的模拟值在每一次模拟中与其他点上的模拟值一一对应,且这种对应又具有一定的偶然性。因此,在利用 SGS 模拟值估计不确定性时,不考虑空间点之间的条件关系,而对每个点分别进行 LHS 抽样,可能会产生偏差。为此,本研究设计了两种途径来利用 SGS-LHS:

(1)考虑 SGS 中的条件关系(SGS-LHS1):随机选择研究区内任一点,仅对该点进行 LHS,研究区内其他点上的土壤属性模拟值为与该点 LHS 抽样值相对应的 SGS 模拟值,重复此过程 10 次,检验这种随机选点是否会产生较大差异;

(2)不考虑 SGS 中的条件关系(SGS-LHS2):对每个点上的 SGS 模拟值分别进行 LHS。这两种途径将在研究区的每个点上产生 50 个(前文 LHS 抽样数设为 50)土壤质量估计值。本研究计算出这 50 个估计值的平均值和标准差,前者表示土壤质量的条件均值(Conditional mean)<sup>[7]</sup>,后者表示土壤质量的不确定性。进一步地,本研究根据验证点上的观测值,用以下三个指标来评价不确定性估计的准确性:

(1)置信区间准确度(Accuracy of probability intervals,  $(\bar{\xi}(p))$ ):

$$\bar{\xi}(p) = \frac{1}{N} \sum_{i=1}^N \xi_i, \xi_i = \begin{cases} 1, & Q_{i,(1-p)/2} \leq O_i \leq Q_{i,(1+p)/2} \\ 0, & \text{其他} \end{cases}$$

(2)置信区间优度(Goodness of probability intervals,  $G$ ):

$$G = 1 - \int_0^1 (3a(p) - 2)(\bar{\xi}(p) - p) dp,$$

$$a(p) = \begin{cases} 1, & \text{if } \bar{\xi}(p) \geq p \\ 0, & \text{其他} \end{cases}$$

(3)置信区间平均宽度(Average wideness for a probability  $p$ ,  $\bar{W}(p)$ ):

$$\bar{W}(p) = \frac{1}{N\bar{\xi}(p)} \sum_{i=1}^N \xi_i (Q_{i,(1+p)/2} - Q_{i,(1-p)/2})$$

式中,  $Q_{i,(1-p)/2}$  和  $Q_{i,(1+p)/2}$  分别为 50 个 LHS 估计中的  $(1-p)/2$  和  $(1+p)/2$  分位数,  $p$  为置信度,  $O_i$  为验证点  $i$  上的观测值。这些指标从不同的角度反映了不确定性估计的准确性:准确度反映区间估计是否正确,取值 0 到 1 之间,值越大说明准确性越好,一般要求大于对应的置信度;优度反映的是区间估计与实际值的吻合程度,取值 0 到 1 之间,值越大说明准确性越好;平均宽度反映区间估计的精确性,值越小越精确<sup>[7]</sup>。

除了 SGS-LHS 外,本研究还分别使用 SGS 和 LHS 方法来分析土壤质量评价中的不确定性,并与 SGS-LHS 的分析结果进行了比较。

## 2 结果与讨论

### 2.1 四种方法估计的土壤质量平均值

图 2 列出了四种不确定性分析方法估计的土壤质量平均值;表 1 对这些估计的平均值进行了统计分析。由于 SGS-LHS1 方法在 10 个随机点上的结果几乎一致,因而在图 2 中仅举一例。这说明,SGS-LHS1 方法在估计土壤质量的平均值时不受随机选

点的影响。图 2 和表 1 从总体上反映出,SGS、SGS-LHS1 和 SGS-LHS2 三种方法估计的土壤质量大致相同;LHS 方法估计的土壤质量平均值与这三种方法有一定的差异,但差异不大。例如,在表 1 中,

LHS 方法估计的土壤质量平均值的平均值为 0.64,略高于其他方法估计的平均值都均为 0.62;标准差、变异系数和变程也均略大于其他方法。

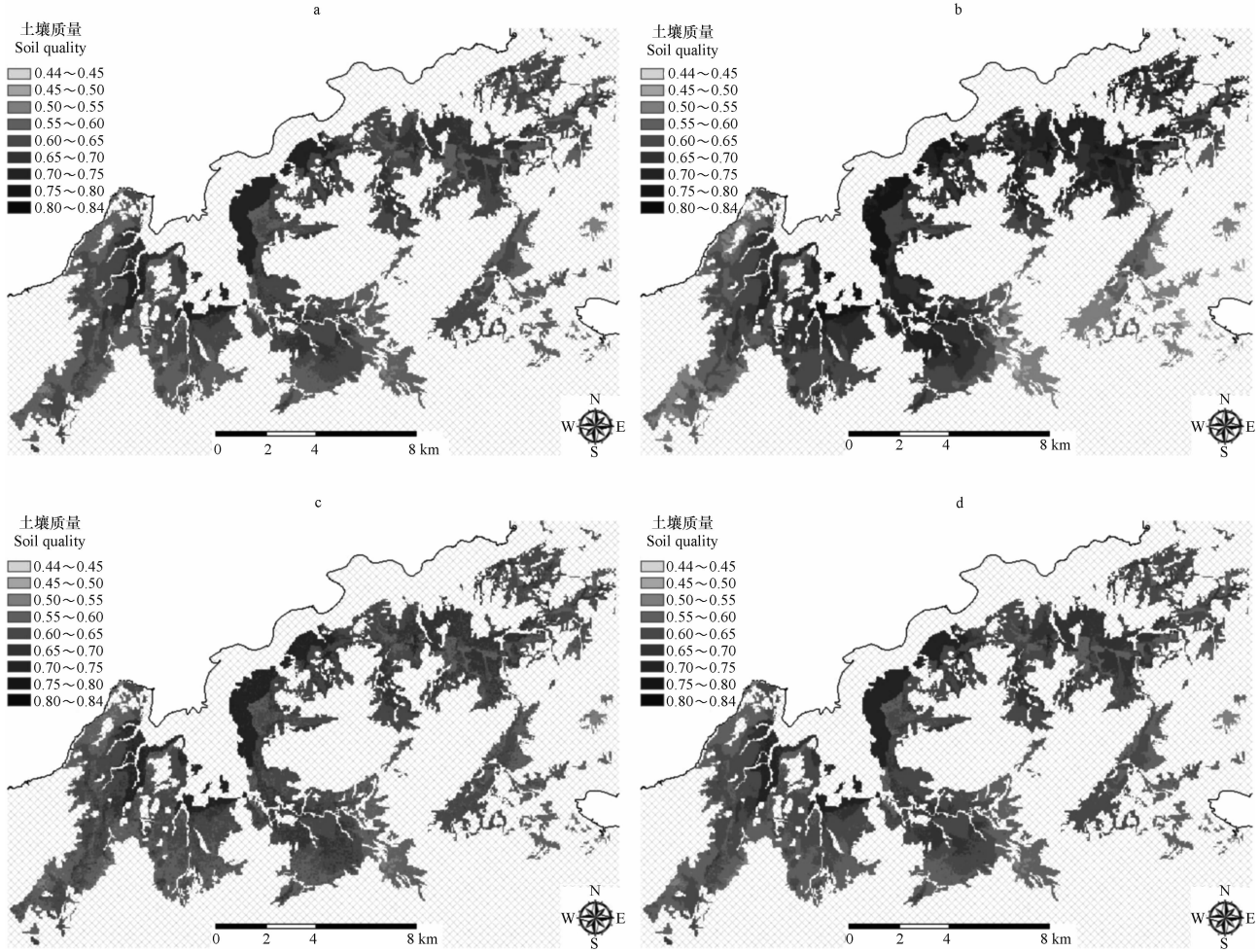


图 2 四种不确定性分析方法估计的土壤质量平均值:(a) SGS;(b) LHS;(c) SGS-LHS1(随机点 1);(d) SGS-LHS2

Fig. 2 Meanvalues of soil quality estimated using the four different uncertainty analysis methods: (a) SGS, (b) LHS, (c) SGS-LHS1 (random 1) and (d) SGS-LHS2

图 3 进而详细地显示了不同方法估计的土壤质量平均值在研究区内每一个空间点上的变异系数及其统计值。该图更加清楚地反映了上述结果,即:(1)四种方法估计的土壤质量平均值之间差异较小,变异系数的均值仅为 2.41%,最大值为 5.65% (图 3a);(2)四种方法之中,LHS 方法估计的土壤质量平均值与其他方法有一定的差异,而其他方法之间的差异则不明显(图 3a 的平均值约为图 3b、图 3c、图 3d 的 1.6 倍,且图 3b、图 3c、图 3d 的平均值几乎相等);(3)SGS-LHS1 方法的 10 次模拟之间差异很小,且略高于与 SGS、SGS-LHS2 两种方法

之间的差异(图 3d 的变异系数平均值略高于图 3b 和图 3c)。

### 2.2 四种方法估计的土壤质量标准差

图 4 列出了四种不确定性分析方法估计的土壤质量标准差(SGS-LHS1 仅举一例);表 2 对这些估计的标准差进行了统计分析;图 5 详细地显示了不同方法估计的土壤质量标准差在研究区内每一个空间点上的变异系数及其统计值。总体上,四种方法估计的土壤质量标准差大致相似,差异并不大。例如,图 5a 反映出这四种方法估计的土壤质量标准差的变异数平均为 9.74%,属于较小的变异。

表 1 四种不确定性分析方法估计的土壤质量平均值的统计值

Table 1 Statistics on the mean values of soil quality estimated using the four different uncertainty analysis methods

方法 Method	平均值 Mean	标准差 Std <sup>1)</sup>	最小值 Min	最大值 Max	变异系数(%) CV <sup>2)</sup>	
SGS	0.62	0.04	0.51	0.73	6.45	
LHS	0.64	0.07	0.44	0.83	10.94	
SGS-LHS1	随机 1 Random 1	0.62	0.04	0.50	0.77	6.45
	随机 2 Random 2	0.62	0.04	0.50	0.76	6.45
	随机 3 Random 3	0.62	0.04	0.49	0.76	6.45
	随机 4 Random 4	0.62	0.04	0.49	0.75	6.45
	随机 5 Random 5	0.62	0.04	0.48	0.75	6.45
	随机 6 Random 6	0.62	0.04	0.49	0.76	6.45
	随机 7 Random 7	0.62	0.04	0.48	0.76	6.45
	随机 8 Random 8	0.62	0.04	0.49	0.76	6.45
	随机 9 Random 9	0.62	0.04	0.50	0.76	6.45
	随机 10 Random 10	0.62	0.04	0.48	0.75	6.45
SGS-LHS2	0.62	0.04	0.51	0.74	6.45	

1) Standard deviation; 2) Coefficient of variation

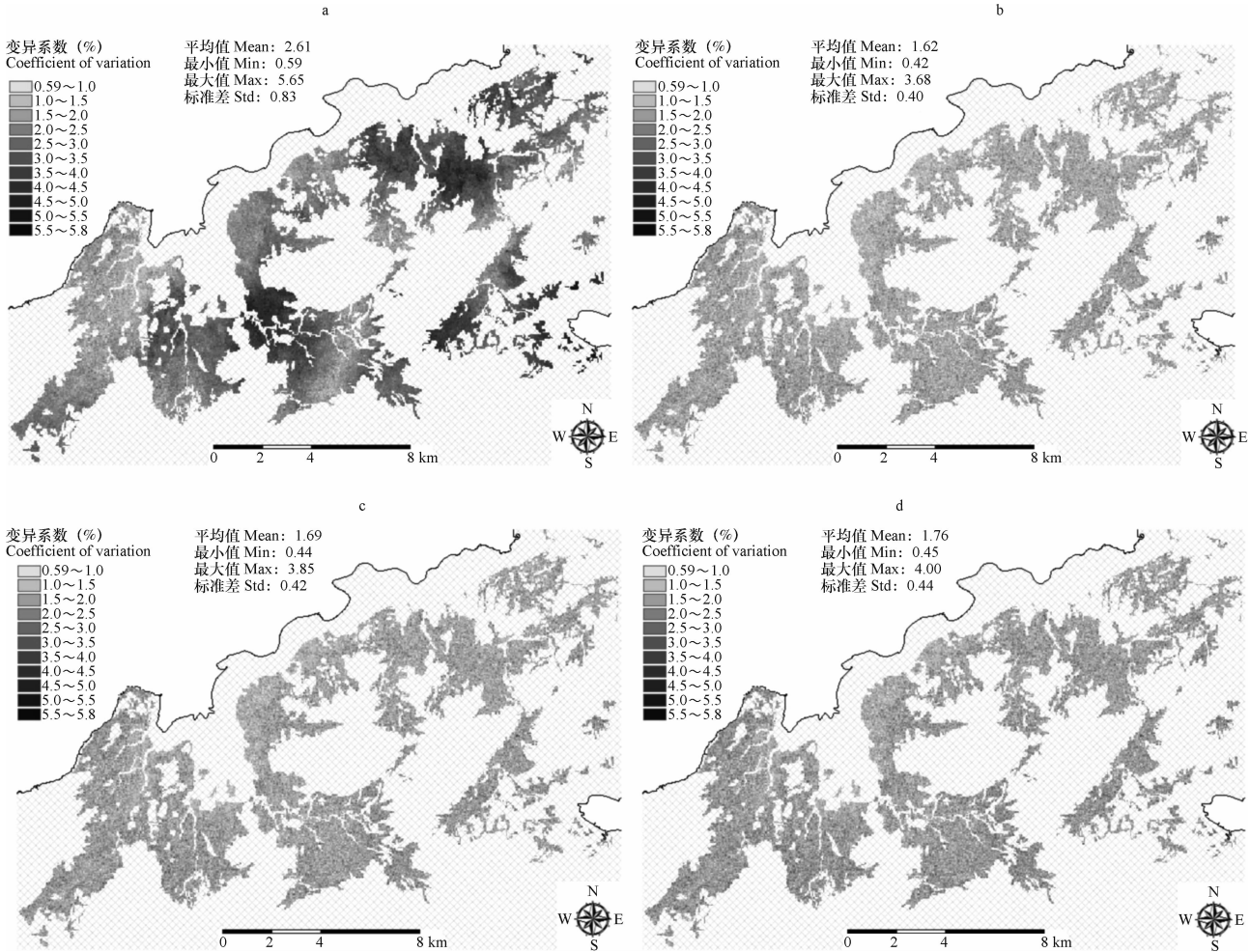


图 3 几种不确定性分析方法估计的土壤质量平均值变异系数: (a) SGS、LHS、SGS-LHS1 (10 次模拟)、SGS-LHS2; (b) SGS、SGS-LHS1 (10 次模拟)、SGS-LHS2; (c) SGS-LHS1 (10 次模拟)、SGS-LHS2; (d) SGS-LHS1 (10 次模拟). Std: 标准偏差

Fig. 3 Coefficients of variation of the mean values of soil quality estimated using different uncertainty analysis methods: (a) SGS, LHS, SGS-LHS1 (10 replicates), SGS-LHS2, (b) SGS, SGS-LHS1 (10 replicates), SGS-LHS2, (c) SGS-LHS1 (10 replicates), SGS-LHS2, and (d) SGS-LHS1 (10 replicates). Std: standard deviation



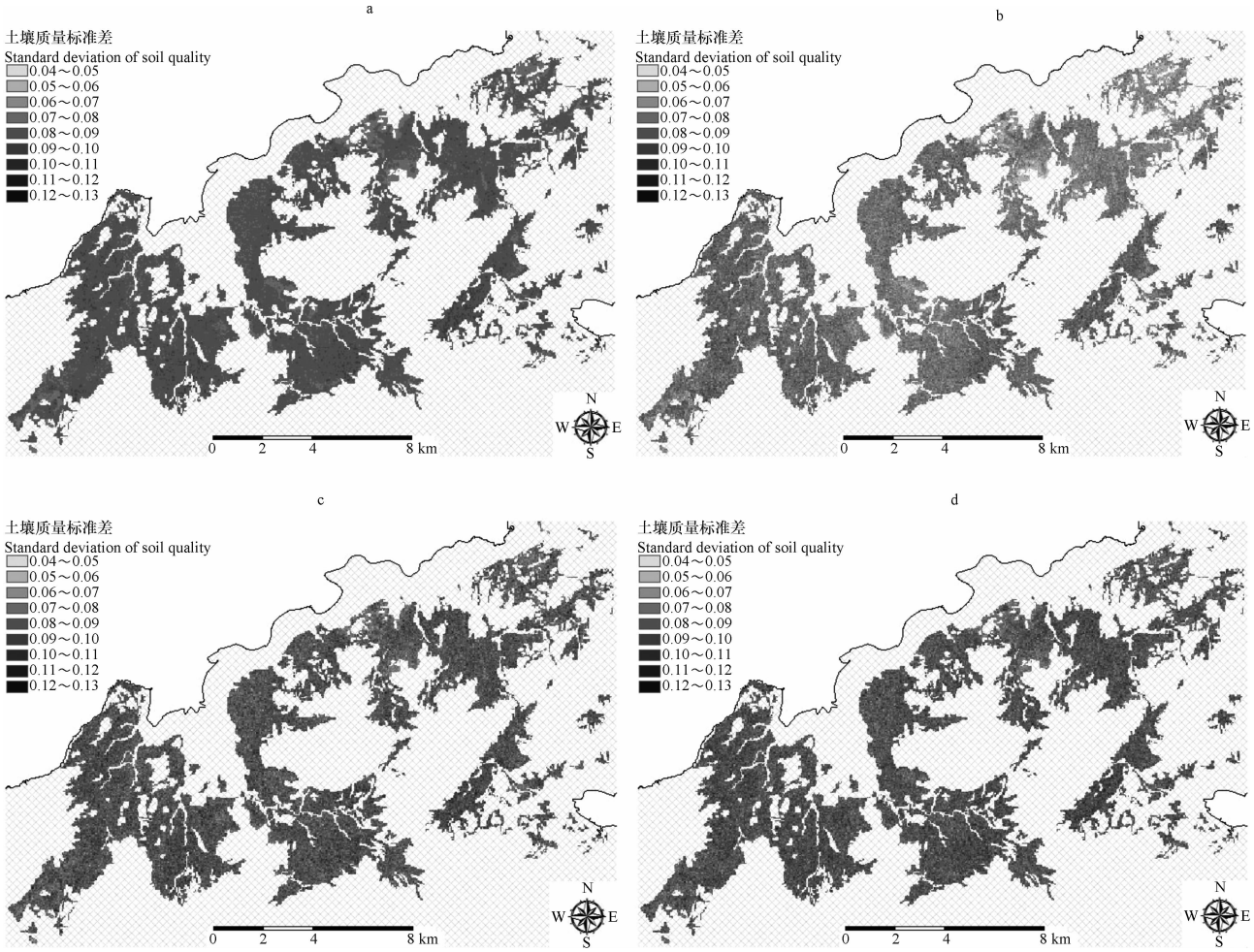


图 4 四种不确定性分析方法估计的土壤质量标准差:(a)SGS;(b)LHS;(c)SGS-LHS1(随机点 1);(d)SGS-LHS2

Fig. 4 Standard deviations of the mean values of soil quality estimated using the four different uncertainty methods: (a) SGS, (b) LHS, (c) SGS-LHS1 (random 1) and (d) SGS-LHS2

表 2 四种不确定性分析方法估计的土壤质量标准差的统计值

Table 2 Statistics of the standard deviations of the mean values of soil quality estimated using the four different uncertainty analysis methods

方法 Method	平均值 Mean	标准差 Std <sup>1)</sup>	最小值 Min	最大值 Max	变异系数 (%) CV <sup>2)</sup>	
SGS	0.08	0.004	0.07	0.10	0.05	
LHS	0.07	0.01	0.04	0.10	0.14	
SGS-LHS1	随机 1 Random 1	0.08	0.01	0.05	0.13	0.13
	随机 2 Random 2	0.08	0.01	0.05	0.13	0.13
	随机 3 Random 3	0.08	0.01	0.05	0.13	0.13
	随机 4 Random 4	0.08	0.01	0.05	0.13	0.13
	随机 5 Random 5	0.08	0.01	0.05	0.13	0.13
	随机 6 Random 6	0.08	0.01	0.05	0.13	0.13
	随机 7 Random 7	0.08	0.01	0.05	0.13	0.13
	随机 8 Random 8	0.08	0.01	0.05	0.13	0.13
	随机 9 Random 9	0.08	0.01	0.05	0.13	0.13
	随机 10 Random 10	0.08	0.01	0.05	0.13	0.13
SGS-LHS2	0.08	0.01	0.05	0.12	0.13	

1) Standard deviation; 2) Coefficient of variation

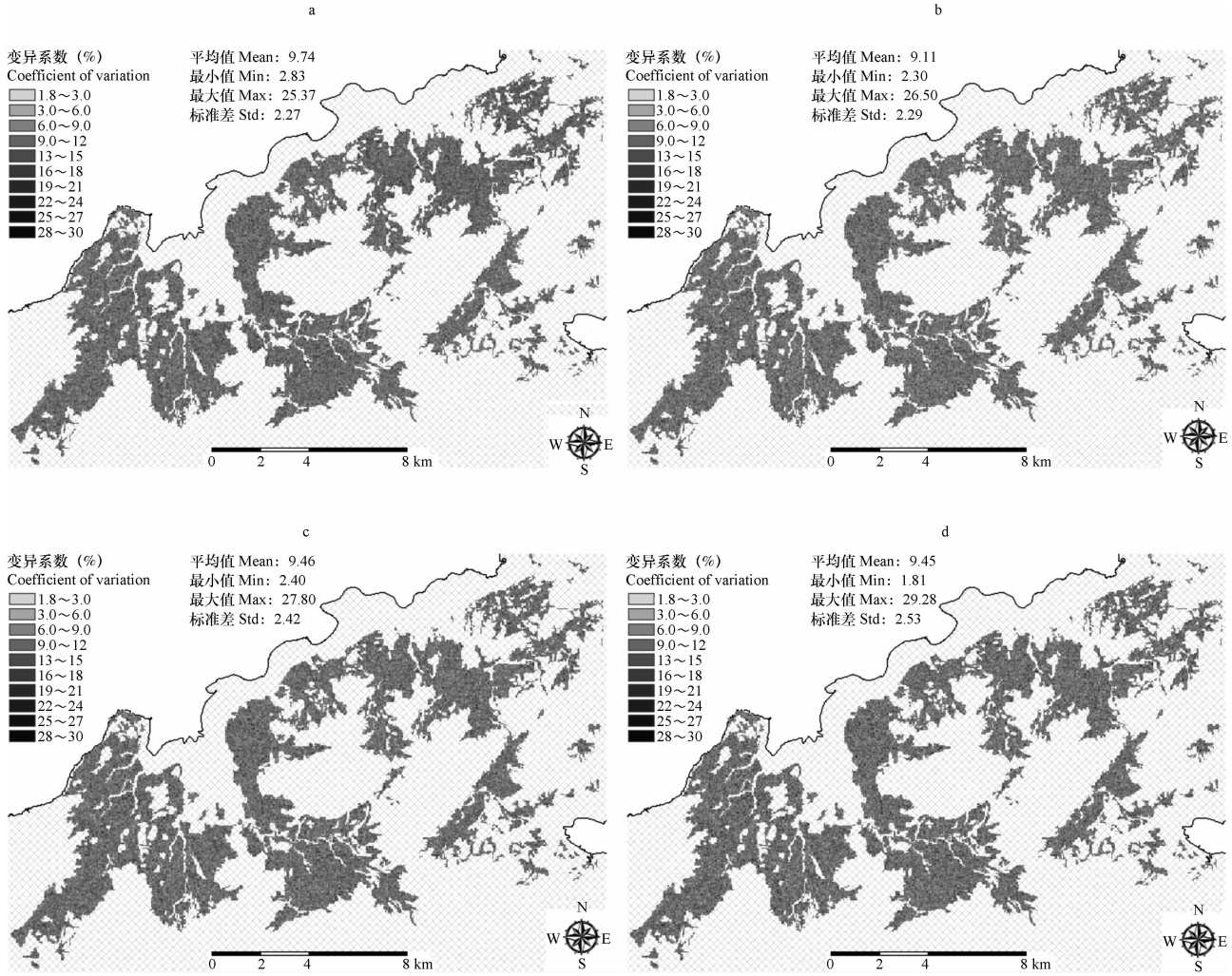


图 5 四种不确定性分析方法估计的土壤质量标准差变异系数: (a) SGS、LHS、SGS-LHS1 (10 次模拟)、SGS-LHS2; (b) SGS、SGS-LHS1 (10 次模拟)、SGS-LHS2; (c) SGS-LHS1 (10 次模拟)、SGS-LHS2; (d) SGS-LHS1 (10 次模拟). Std: 标准偏差  
 Fig. 5 Coefficients of variation of the standard deviations of the mean values of soil quality estimated using the four different uncertainty analysis methods; (a) SGS, LHS, SGS-LHS1 (10 replicates), SGS-LHS2, (b) SGS, SGS-LHS1 (10 replicates), SGS-LHS2, (c) SGS-LHS1 (10 replicates), SGS-LHS2, and (d) SGS-LHS1 (10 replicates). Std: standard deviation

SGS-LHS1 的 10 次随机选点模拟之间在总体上差异极小(表 2),但在空间点上有较小的差异,表现在图 5d 中的变异系数平均值为 9.45%。这表明随机选点对 SGS-LHS1 估计标准差的影响较小,反映出 LHS 对 SGS 模拟结果进行抽样后,数据的偶然性较强而空间点之间的条件关系较弱。因此,SGS-LHS2 与 SGS-LHS1 的 10 次随机选点模拟之间差异很小(见表 2 和图 5c,甚至小于 SGS-LHS1 的 10 次模拟内部的差异(图 5c 中变异系数的变程和标准差均大于图 5d)。

LHS 方法估计的标准差在总体上与 SGS-LHS1 和 SGS-LHS2 的差异较小,主要的差异表现在标准差的变程较窄(表 2),但在空间点上与 SGS-LHS1 和 SGS-LHS2 的差异较大(图 5a 的均值在图 5 的四

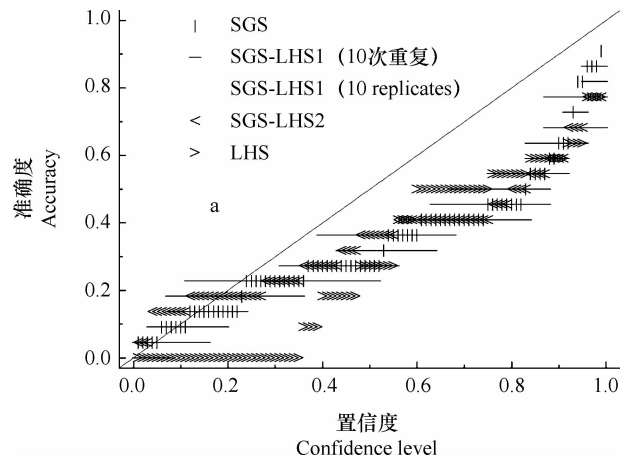
幅图中最大)。相反,SGS 方法估计的标准差在总体上与 SGS-LHS1 和 SGS-LHS2 的差异较大,主要表现在标准差和变异系数相对很小(表 2),但在空间点上,则与 SGS-LHS1 和 SGS-LHS2 的差异很小(图 5b 的均值在图 5 的四幅图中最小)。这是因为 SGS 模拟时考虑了空间点之间的条件关系,而 LHS 抽样则仅依据克里格方差。Sun 等<sup>[4]</sup>研究表明 SGS 模拟估计的方差小于克里格方差。因此,SGS 模拟的土壤信息值更加聚集于平均值两侧,而以克里格预测值和预测方差为正态分布进行 LHS 抽样的土壤信息值将更加分散于平均值两侧。所以,在本研究中,SGS 方法估计的标准差的变程和变异总体上较 SGS-LHS1、SGS-LHS2 和 LHS 方法估计的要分别窄和小(表 2),且在单个空间点上与 SGS-LHS1、SGS-



LHS2 方法差异较小。虽然在总体上 LHS 方法估计的标准差与其他三种方法之间没有较大差异,但在单个空间点上却差别较大。这可归因为 LHS 方法完全忽略了空间点之间的条件关系。

### 2.3 四种方法估计的土壤质量不确定性的准确性

图 6 列出了四种不确定性分析方法估计的土壤质量置信度区间的准确度和平均宽度。在图 6a 中,随着置信度的增加,四种方法估计的置信度区间的准确度逐渐提高,但绝大部分均低于对应的置信度(在 1:1 对照线以下),说明估计的不确定性的准确度有所欠缺。总体上看,四种方法之间的差别较小,在不同的置信度下各占优势,没有明显的规律。



在图 6b 中,除 LHS 方法外,其他三种方法之间的差别较小。LHS 方法与这三种方法之间的差别较为明显,其估计的置信度平均宽度一直略低于其他方法,说明这种方法估计的稍显精确。SGS-LHS1 方法 10 次重复模拟的置信度区间差别不大,说明随机选点对 SGS-LHS1 没有影响。

四种方法估计的不确定性的优度分别为:0.68 (SGS)、0.58 (LHS)、0.67 (SGS-LHS1 的 10 次重复平均值,这些值分别为 0.71、0.65、0.67、0.62、0.71、0.66、0.69、0.68、0.70、0.66) 和 0.68 (SGS-LHS2)。显然,SGS、SGS-LHS1 和 SGS-LHS2 几乎一致,而 LHS 方法则相对较差。

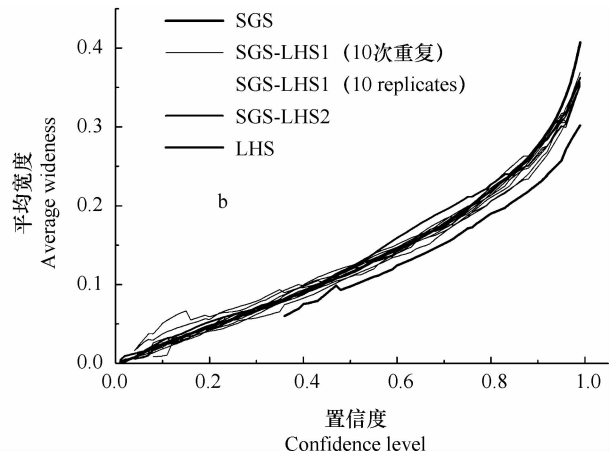


图 6 四种不确定性分析方法估计的土壤质量不确定性的准确性:(a)置信度区间准确度(图中直线为 1:1 对照线);(b)置信度区间平均宽度

Fig. 6 Comparison of the four analysis methods in accuracy: (a) accuracy of confidential interval (1:1 line), (b) average wideness of confidential interval

## 3 结 论

本文提出使用 SGS 和 LHS 相结合的方法(即 SGS-LHS)来分析土壤信息不确定性在空间分析中的传递,目的在于充分利用 SGS 和 LHS 各自的优点,互补各自的缺点,以提高不确定性传递分析的准确性和效率。这种方法在香港农田土壤质量评价中的应用结果表明,这种方法分析所得的不确定性结果与 SGS 很接近,而与 LHS 的结果则有一定的差别,尽管差别不大。并且,这种方法估计不确定性的准确性与 SGS 也很接近,且均高于 LHS 方法,尽管 LHS 估计的置信度区间平均宽度略显精确。本文没有对这些方法的计算效率进行比较,但从上文中的计算过程来看,LHS 方法的效率最高,其次是 SGS-LHS1,然后是 SGS,最后是 SGS-LHS2。然

而,由于现在的计算机处理速度一般均较快,即使高性能计算机也容易得到(例如本研究使用的曙光 5000A 高性能并行计算机,位于广西大学,隶属国家高性能计算中心南宁分中心),这几种方法的效率差别较小。因此,本文推荐使用 SGS-LHS1 方法进行空间不确定性传递分析。同时,本文研究的土壤质量评价分析是线性模型,对于其他类型的模型(如非线性)该方法一般也具有适用性,但仍需要研究。

### 参 考 文 献

- [1] 孙孝林,赵玉国,刘峰,等. 数字土壤制图及其研究进展. 土壤通报, 2013, 44(3): 752—759. Sun X L, Zhao Y G, Liu F, et al. Digital soil mapping and advance in research (In Chinese). Chinese Journal of Soil Science, 2013, 44(3): 752—759
- [2] Carré F, McBratney A B, Myar T, et al. Digital soil assess-

- ments: Beyond DSM. *Geoderma*, 2007, 142: 69—79
- [ 3 ] Malone B P, de Grujter J J, McBratney A B, et al. Using additional criteria for measuring the quality of predictions and their uncertainties in a digital soil mapping framework. *Soil Science Society of American Journal*, 2011, 75: 1032—1043
- [ 4 ] Sun X L, Wu S C, Wang H L, et al. Uncertainty analysis for the evaluation of agricultural soil quality based on digital soil maps. *Soil Science Society of American Journal*, 2012, 76(4): 1379—1389
- [ 5 ] Brown J D, Heuvelink G B M. Assessing uncertainty propagation through physically based models of soil water flow and solute transport//Anderson M G. *Encyclopedia of Hydrologic Sciences*. Chichester, UK: Wiley, 2005:1181—1195
- [ 6 ] 张淑杰, 朱阿兴, 刘京, 等. 整合已有土壤样点的数字土壤制图补样方案. *地理科学进展*, 2012, 31(10): 1318—1325. Zhang S J, Zhu A X, Liu J, et al. An integrative sampling scheme for digital soil mapping (In Chinese). *Progress in Geography*, 2012, 31(10): 1318—1325
- [ 7 ] Goovaerts P. Geostatistical modeling of uncertainty in soil science. *Geoderma*, 2001, 103: 3—26
- [ 8 ] Qi Y, Darilek J L, Huang B, et al. Evaluating soil quality indices in an agricultural region of Jiangsu Province, China. *Geoderma*, 2009, 149: 325—334
- [ 9 ] Castrignano A, Buttafuoco G, Canu A, et al. Modelling spatial uncertainty of soil erodibility factor using joint stochastic simulation. *Land Degradation & Development*, 2008, 19(2): 198—213
- [ 10 ] Minasny B, Sulaeman Y, McBratney A B. Is soil carbon disappearing? The dynamics of soil organic carbon in Java. *Global Change Biology*, 2011, 17(5): 1917—1924
- [ 11 ] 李润奎, 朱阿兴, Augello P C, 等. SWAT 模型对高精度土壤信息的敏感性研究. *地球信息科学*, 2007, 9(3): 73—90. Li R K, Zhu A X, Augello P C, et al. Sensitivity of SWAT model to detailed soil information (In Chinese). *Geo-Information Science*, 2007, 9(3): 73—90
- [ 12 ] 廖凯华, 徐绍辉, 吴吉春, 等. 土壤饱和和导水率空间预测的不确定性分析. *水科学进展*, 2012, 23(2): 200—205. Liao K H, Xu S H, Wu J C, et al. Uncertainty analysis for spatial prediction of soil saturated hydraulic conductivity (In Chinese). *Advances in Water Science*, 2012, 23(2): 200—205
- [ 13 ] Webster R, Oliver M A. *Geostatistics for environmental scientists*. Chichester, UK: Wiley, 2007: 273—274
- [ 14 ] Buttafuoco G, Conforti M, Aucelli P P C, et al. Assessing spatial uncertainty in mapping soil erodibility factor using geostatistical stochastic simulation. *Environmental Earth Sciences*, 2012, 66(4): 1111—1125
- [ 15 ] McKay M D, Beckman R J, Conover W J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 1979, 21: 239—245
- [ 16 ] 承继成, 金江军. 地理数据的不确定性研究. *地球信息科学*, 2007, 9(4): 1—4. Cheng J C, Jin J J. The uncertainty of geographic data (In Chinese). *Geo-information Science*, 2007, 9(4): 1—4
- [ 17 ] 骆永明, 等. 香港地区土壤及其环境. 北京: 科学出版社, 2007. Luo Y M, et al. *Hong Kong soils and environment* (In Chinese). Beijing: Science Press, 2007
- [ 18 ] Walvoort D J J, Brus D J, de Grujter J J, et al. An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means. *Computers and Geosciences*, 2010, 36(10): 1261—1267
- [ 19 ] 鲁如坤. 土壤农业化学分析方法. 北京: 中国农业科技出版社, 2000. Lu R K. *Analytical methods for soil and agro-chemistry* (In Chinese). Beijing: China Agricultural Science and Technology Press, 2000
- [ 20 ] Lark R M. Modelling complex soil properties as contaminated regionalized variables. *Geoderma*, 2002, 106: 173—190
- [ 21 ] Liu T L, Juang K W, Lee D Y. Interpolating soil properties using kriging combined with categorical information of soil maps. *Soil Science Society of American Journal*, 2006, 70(4): 1200—1209
- [ 22 ] Vereecken H, Maes J, Feyen J. Estimating unsaturated hydraulic conductivity from easily measured soil properties. *Soil Science*, 1990, 149(1): 1—12
- [ 23 ] Gurdak J J, Qi S L, Geisler M L. Estimating prediction uncertainty from geographical information system raster processing: A user's manual for the raster error propagation tool (REPTool). U. S. Geological Survey Techniques and Methods 11 - C3, 2009

## APPLICATION OF SGS AND LHS TO ANALYZING UNCERTAINTIES IN DIGITAL SOIL QUALITY ASSESSMENT

Sun Xiaolin<sup>1,2</sup> Wang Huili<sup>1†</sup> Cao Jizhao<sup>1</sup>

(1 Guangxi Key Laboratory of Superior Timber Tress Resource Cultivation, Guangxi Forestry Research Institute, Nanning 530002, China)

(2 School of Geography and Planning, Sun Yat-set University, Guangzhou 510275, China)

**Abstract** Transmission of uncertainty of soil information in spatial analysis is an issue critical to digital soil assessment. To cope with the issue, coupling of the sequential Gaussian simulation (SGS) and the Latin hypercube sampling (LHS) methods, i. e., SGS-LHS, was proposed with a view to making full use of the advantages of the two methods as

complementation to overcome their respective drawbacks, so as to improve accuracy and efficiency of the transmission and analysis of the uncertainty. The new method, (including two pathways: SGS-LHS1 and SGS-LHS2), SGS and LHS were tested in soil quality assessment of farmlands in Hong Kong for comparison. Results show that the uncertainty of the analysis using the SGS-LHS method was similar to that using SGS, but different to a certain, rather a big extent from that using LHS, and the combination method was approximate to SGS, but higher than LHS in accuracy, although LHS appeared to a little bit more accurate in terms of mean width of the confidential intervals. Therefore, the combined SGS-LHS method is recommended for analysis of transmission of soil information uncertainty in spatial analysis.

**Key words** Sequential Gaussian simulation; Latin hypercube sampling; Soil quality; Uncertainty; Digital soil assessment

(责任编辑:汪叔生)