

# 利用高程辅助进行土壤有机质的随机模拟

柴旭荣<sup>1,2</sup>, 黄元仿<sup>1\*</sup>, 苑小勇<sup>1</sup>, 高如泰<sup>3</sup>

(1. 中国农业大学土壤与水系, 教育部植物-土壤相互作用重点实验室, 农业部土壤与水重点实验室, 北京 100094;  
2. 山西师范大学城市与环境科学学院, 临汾 041000; 3. 河北农业大学资源与环境科学学院, 保定 071001)

**摘要:** 为了探讨在条件模拟计算环境下, 是否可以利用高程数据辅助提高土壤有机质空间变化的预测精度及相应的预测不确定性模拟的准确性, 该文在北京市平谷区内选取研究样区, 以土壤有机质作为目标变量, 一方面利用序贯高斯模拟法对土壤有机质的空间分布进行模拟, 另一方面以高程作为辅助信息, 利用序贯高斯协模拟法对土壤有机质的空间分布进行模拟, 然后对两种方法的模拟结果进行对比分析。结果表明, 在土壤有机质的空间预测精度、模拟预测结果的局部不确定性和模拟预测结果的空间不确定性三方面, 通过将高程数据考虑进有机质条件模拟过程中, 准确性都得到了提高。这对于农业可持续发展以及全球碳平衡研究都具有十分重要的意义。

**关键词:** 土壤有机质, 序贯高斯模拟, 序贯高斯协模拟, 不确定性, 高程

**中图分类号:** S158

**文献标识码:** A

**文章编号:** 1002-6819(2008)-12-0210-05

柴旭荣, 黄元仿, 苑小勇, 等. 利用高程辅助进行土壤有机质的随机模拟[J]. 农业工程学报, 2008, 24(12): 210—214.  
Chai Xurong, Huang Yuanfang, Yuan Xiaoyong, et al. Random simulation of soil organic matter using elevation as auxiliary information[J]. Transactions of the CSAE, 2008, 24(12): 210—214.(in Chinese with English abstract)

## 0 引言

准确预测土壤有机质空间分布是进行相关土壤有机质模拟和有机质田间管理的必要基础, 而描述和评价预测结果本身所具有的不确定性, 可以进一步帮助决策者做出合理的决策<sup>[1]</sup>。地统计学为分析空间分布数据提供了一系列工具和方法, 并且在土壤学中得到广泛的研究与应用<sup>[2-5]</sup>。使估计误差的方差最小是克里格方法的显著特点, 但是克里格法却不能很好地再现原始区域化变量的空间结构及变量之间的相互关系<sup>[6]</sup>。条件模拟方法可以克服克里格法的缺陷, 能够很好地再现原始变量的空间结构, 同时还可以对预测结果的不确定性进行定量的模拟, 所以条件随机模拟方法在土壤及其他学科中的应用受到越来越多的关注<sup>[7-11]</sup>。

在区域尺度上, 高程是影响土壤有机质空间分布的一个主要因素。如果数据量丰富的高程数据能够来辅助提高土壤有机质的空间预测精度, 这在经济方面和环境方面都具有应用价值。相关研究已经证实, 通过多元克里格方法利用数据量丰富的辅助变量可以提高土壤属性的空间预测精度<sup>[12-16]</sup>。但是, 因为克里格法具有上述的缺陷, 以及不能进行定量的不确定性分析, 使得克里格

与辅助变量结合的预测结果应用也受到一定的限制。然而, 将辅助变量与条件模拟方法相结合, 对土壤属性空间空间分布进行模拟, 并对模拟结果进行不确定性分析的相关文献还比较少。为此, 本文将高程辅助数据与序贯高斯协模拟方法相结合, 与只考虑单一有机质变量的序贯高斯模拟法进行比较分析, 来评价在条件模拟计算环境下高程是否有助于提高不确定性模拟的准确性, 是否有助于提高空间预测精度。

## 1 研究区概况与数据来源

研究区位于北京市平谷区中部 (116.93°E ~ 117.33°E, 40.14°N ~ 40.27°N), 如图 1 所示。研究区总面积为 489 km<sup>2</sup>。地貌类型主要包括山地、山前洪积扇和山前冲积平原, 山地主要分布在研究区的北部和东部, 中部和西南部依次为山前洪积扇和山前冲积平原。海拔变化从 20 m 到 910 m (图 1)。土壤类型主要包括棕壤、褐土、潮土等土类。

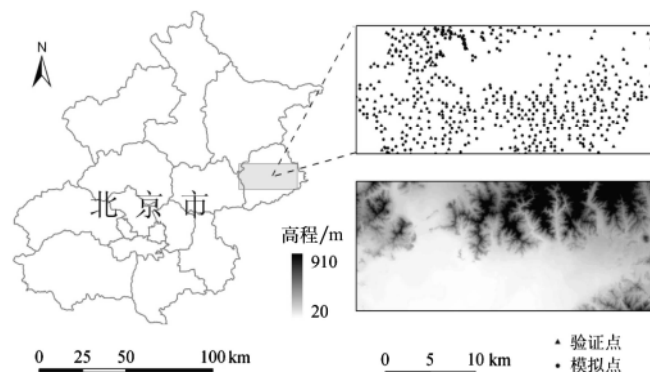


图 1 研究区位置、数字高程模型及采样点分布图  
Fig.1 Distribution map for study area and soil sampling sites and digital elevation model map

收稿日期: 2007-07-18 修订日期: 2008-11-08

基金项目: 国家自然科学基金项目 (40671084); 北京自然科学基金项目 (6072017); 国家科技支撑计划 (2006BAD10A01); 教育部新世纪优秀人才支持计划 (NCET-06-0107); 长江学者和创新团队发展计划资助 (IRT0412)  
作者简介: 柴旭荣 (1975—), 男, 山西曲沃人, 博士研究生, 主要研究方向土地利用与信息技术。北京 中国农业大学土壤与水系, 100094。

Email: rongchai@yahoo.com.cn

\*通讯作者: 黄元仿 (1968—), 男, 湖北仙桃人, 教授, 博士生导师, 研究方向为土壤水分、养分模型与水土资源管理。北京 中国农业大学土壤与水系, 100094。Email: yfhuang@china.com

土壤有机质数据来源于 2005 年平谷区土壤调查数据, 有机质采用重铬酸钾氧化容量法——外加热法测定。研究区内土壤有机质样点总数为 547 个, 为了对模拟结果进行验证, 本文将这些点随机分为模拟点 (382) 和验证点 (165) 两部分, 采样点分布如图 1 所示。高程数据采用栅格大小为 25 m 的数字高程模型。

## 2 研究方法与评价标准

### 2.1 序贯高斯模拟和序贯高斯协模拟

本文采用序贯高斯模拟 (SGS) 和序贯高斯协模拟 (SGCS) 两种方法对土壤有机质空间分布进行模拟。SGS 是应用高斯概率理论和序贯模拟算法产生连续变量空间分布的一种随机模拟方法。它主要是根据采样数据计算待模拟点值的条件累计概率分布, 从该分布中随机取一值作为模拟现实。在每个点上得出 1 个模拟值后, 就把它连同原始采样数据、此前得到的模拟数据一起作为条件数据, 沿着一条随机路线进入下一点的模拟, 循环上述过程直至所有待模拟点都被赋予一个模拟值, 就得到一个实现 (realization)。通过重复上述序贯过程沿着不同随机路线, 得到不同的实现。通过多个实现, 就可以对模拟结果的不确定进行评价<sup>[6,17]</sup>。

序贯高斯协模拟法 (SGCS) 的原理和模拟过程与 SGS 基本是一致的。不同之处在于, SGS 只用目标变量的自变异函数通过单变量克里格法来决定高斯条件累计分布函数的均值和方差, 而 SGCS 是用目标变量和辅助变量的自变异函数和协变异函数通过协克里格法来决定高斯条件累计分布函数的均值和方差。

### 2.2 评价标准

本文从空间预测精度、预测的局部不确定性和空间不确定三个方面对模拟结果进行评价, 评价标准具体如下。

#### 2.2.1 空间预测精度评价

空间预测精度通过预测值与实测值之间的误差大小来评价, 误差越小预测精度越高。本文用均方根误差 (RMSE) 作为验证标准来评价不同的预测结果。

$$RMSE = \sqrt{\frac{1}{N} \sum_{j=1}^N [z(u_j) - z^*(u_j)]^2} \quad (1)$$

式中  $z(u_j)$ ——预测值;  $z^*(u_j)$ ——实测值;  $N$ ——验证点的个数。

#### 2.2.2 局部不确定性模拟评价

局部不确定性评价主要是评价一系列实现对单一点条件概率分布再现 (reproduction) 得准确性和拟合度。

评价标准选用 Deutsch (1997 年) 和 Goovaerts (2001 年) 提出的标准<sup>[1]</sup>。在任一验证点  $u$  上, 从条件累计分布函数  $F(u_j, z(n))$  可以得到一系列以条件累计分布函数的  $(1-p)/2$  和  $(1+p)/2$  的分位数为边界的对称概率区间 (symmetric p-probability interval)。那么, 实测值落入对称概率区间的比例由公式 (2) 计算得到:

$$\bar{\xi}(p) = \frac{1}{N} \sum_{j=1}^N \xi(u_j, p) \quad \forall p \in [0, 1] \quad (2)$$

其中:

$$\xi(u_j, p) = \begin{cases} 1 & \text{if } F^{-1}\left(\frac{u_j, (1-p)}{2}\right) < z(u_j) \leq F^{-1}\left(\frac{u_j, (1+p)}{2}\right) \\ 0 & \text{otherwise} \end{cases}$$

实测值落入对称概率区间的比例与概率区间的散点图称为准确图, 如果  $\bar{\xi}(p) > p$ , 即散点落在平分线上方, 则表示所预测的条件概率分布是准确的, 如果落在平分线下方则表示所预测条件概率分布不准确; 也可以通过准确性统计 (公式 (3)) 和拟合度统计 (公式 (4)) 来量化它们的准确性和接近程度。

$$A = \sum_{k=1}^K a(p_k) \Delta p_k \quad (3)$$

$$G = 1 - \sum_{k=1}^K [3a(p_k) - 2][\bar{\xi}(p_k) - p_k] \Delta p \quad (4)$$

在公式 (2)、(3) 和 (4) 中,  $p$  为概率, 如果  $\bar{\xi}(p) \geq p$ ,  $a(p)$  等于 1, 否则为 0;  $N$  为验证点总的个数;  $K$  为对称概率区间的总数。

#### 2.2.3 空间不确定性模拟评价

与对局部不确定性模拟的评价不一样, 目前关于对空间不确定模拟的评价还没有具体量化标准。Bourennane (2007 年) 提出用标准差分布图代表空间不确定进行评价, 指出在保持模拟准确性的同时, 标准差越小, 那么空间不确定性模拟就越准确<sup>[18]</sup>。本文也采用这一标准。

## 3 结果与分析

### 3.1 数据统计与基本分析

表 1 给出了模拟点和验证点土壤有机质的基本统计结果。总的来看, 模拟点与验证点的各项统计结果基本相近。从这些统计结果可以看出, 研究区土壤有机质含量属于中等水平, 均值大约为 15 g/kg, 其变化幅度比较大, 最大值是最小值的 12 倍还多, 变异系数在 30% 附近, 属于中等变异强度。偏态值显示研究区数据基本服从正态分布。从线性相关系数 ( $r$ ) 看出, 研究区内土壤有机质与高程中等强度相关。

表 1 土壤有机质的基本统计特征

Table 1 Basic statistical characteristics of soil organic matter

|                                 | 样点数 | 平均值   | 标准差  | 最小值  | 最大值   | 变异系数(%) | 偏态值  | $r$  |
|---------------------------------|-----|-------|------|------|-------|---------|------|------|
| 有机质 (模拟点) /g · kg <sup>-1</sup> | 382 | 15.12 | 4.74 | 2.81 | 38.35 | 31.34   | 1.02 | 0.37 |
| 有机质 (验证点) g · kg <sup>-1</sup>  | 165 | 15.19 | 4.27 | 3.18 | 31.05 | 28.10   | 0.73 | 0.39 |

图 2 是模拟点土壤有机质数据经正态积分转换后计算得到的实验自变异函数及其拟合模型结果图。变程表

示观测变量间的最大空间自相关距。相对块金效应 (块金值/基台值=0.64) 说明研究区引起土壤有机质的具有中

等偏弱的空间自相关性。另外, SGCS 所需的土壤有机质与高程协变异函数参数, 根据 Markov-type 近似值得到。

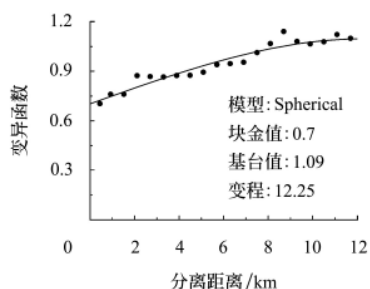


图 2 试验变异函数(实点)及其所拟合模型(实线)

Fig.2 Experimental variograms (dots) and their fitted models (solid lines)

### 3.2 模拟与评价

根据上述变异函数参数, 利用 SGS 法和 SGC 算法分别对研究区土壤有机质进行了 200 次模拟。图 3 列出了每种算法的第 1 次实现、第 100 次实现、第 200 次实现和全部 200 次实现的平均值结果图。从这些结果图可以看出, 所选的各次实现显示出相同的总的空间变化, 例如, 高的有机质含量主要分布在研究区的东北部, 有机质含量从西南角往东北角呈现出逐步上升的趋势。但是在细部空间变化上各次实现之间是有差别的。而全部实现的平均值结果图, 在表现有机质含量在研究区空间变化的总体趋势上比较清晰明显, 但局部变化比较平滑。

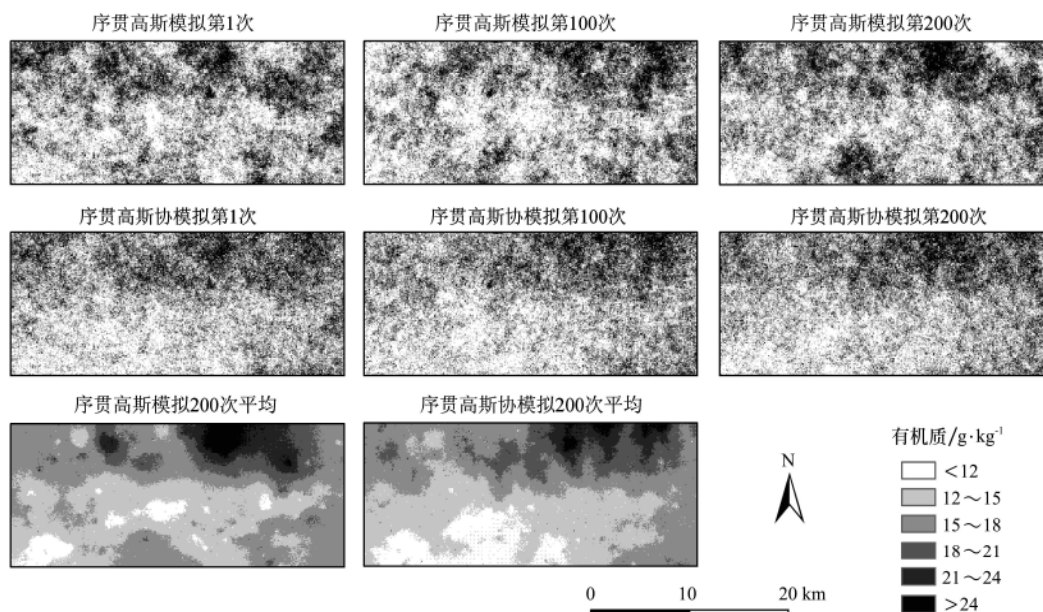


图 3 土壤有机质的模拟实现结果图

Fig.3 Realizations of soil organic matter generated by simulation approaches

#### 3.2.1 预测精度评价

根据验证点计算每种算法各次实现的均方根误差值 ( $RMSE$ ), 图 4 对 SGS 法和 SGCS 法前 50 次模拟实现的  $RMSE$  值进行了对比。从图中可以看出, SGCS 法各次实现的  $RMSE$  值总体上都低于 SGS 法各次实现的  $RMSE$  值。

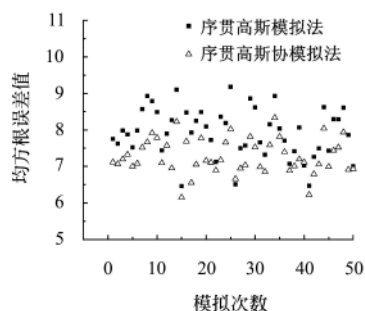


图 4 序贯高斯模拟法和序贯高斯协模拟法前 50 次均方根误差值对比图

Fig.4 Comparison chart of the root mean square error value from the 1-50 realizations of sequential Gaussian simulation and sequential Gaussian co-simulation

另外, 从全部实现的平均实现来看, SGCS 法平均实现的  $RMSE$  值 (3.99) 也明显小于 SGS 法平均实现的  $RMSE$  值 (4.09)。总的看来, SGCS 法将高程数据考虑到模拟过程中所得到模拟结果的精度明显的高于 SGS 法的预测结果。可见, 通过条件随机模拟方法, 可以利用辅助变量高程来提高土壤有机质的空间预测精度。

#### 3.2.2 局部不确定性模拟评价

利用验证点实测值和每种方法的 200 次模拟实现, 根据公式 (2) 分别计算得到两种算法的准确图 (图 5)。从图 5 中可以看出, 在概率区间  $0 < p < 0.3$  范围内, SGS 法的点落在平分线的下方, 而 SGCS 法的点基本上落在平分线的上方, 这指示出在这个概率区间 SGS 法在模拟预测局部不确定性方面的不准确, SGCS 法比较准确; 在概率区间  $0.3 < p < 1$  范围内, 结果恰恰相反, SGS 法的点都落在平分线的上方, 而 SGCS 法的大多数点都落在平分线的下方。从准确性统计值  $A$  可以看出, 在模拟预测局部不确定方面, SGS 法 ( $A=0.72$ ) 总体上比 SGCS ( $A=0.4$ ) 法准确。然而, 从点偏离平分线的情况来看, SGS 法的偏离程度要大于 SGCS 法, 对应于拟合度统计值  $G$  也可

以看出, SGS 的拟合度 ( $G=0.93$ ) 不如 SGCS 法的拟合度 ( $G=0.97$ ) 高。这就说明了在所模拟的实现重现单点条件概率分布方面, SGCS 法的拟合度高于 SGS 法。

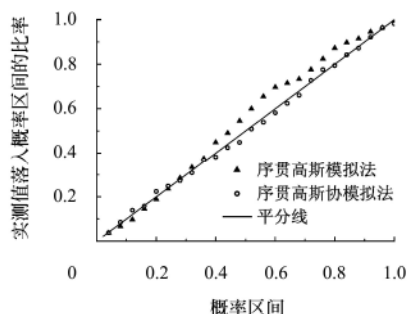


图 5 两种算法的准确图

Fig.5 Accuracy plots of the two simulation algorithms

### 3.2.3 空间不确定模拟评价

图 6 是两种算法的全部实现的标准差分布图。从图中可以看出, 两种算法得到的标准差在空间分布的总体布局上是一致的, 在研究区东北部都有一个高值区域, 在中部有一个条带状区域标准差值都比较小。但是, 二者差别也比较明显, 总体上, SGS 法的标准差值普遍高于 SGCS 法, 特别是在研究区南部有两个区域 SGS 法的值明显高于 SGCS 法。从这个结果可以看出, 在模拟空间不确定方面, SGCS 法比 SGS 法准确。

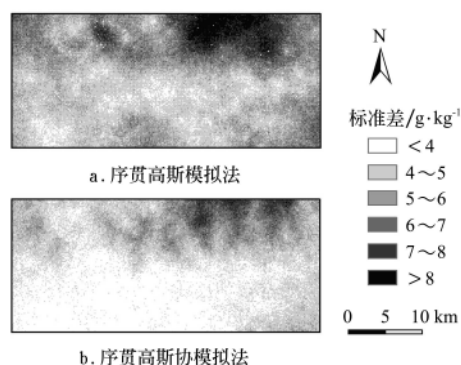


图 6 两种模拟算法得到土壤有机质 200 次实现的标准差分布图

Fig.6 Standard deviations of soil organic matter from 200 realizations obtained by the two simulation algorithms

## 4 结 论

- 1) 将丰富的高程数据考虑到条件高斯模拟过程中, 提高了土壤有机质的空间分布的模拟预测精度;
- 2) 将丰富的高程数据考虑到条件高斯模拟过程中, 在模拟结果的局部不确定性方面, 模拟的准确性虽然没有提高, 但是提高了条件概率分布的拟合度;
- 3) 将丰富的高程数据考虑到条件高斯模拟过程中, 对于模拟结果的空间不确定性来讲, 模拟的准确性也得到了相应地提高。
- 4) 通过运用条件高斯协模拟法, 可以利用丰富的高程数据辅助提高土壤有机质的空间预测精度, 也可以提

高预测结果不确定性模拟的准确性。这对于土壤肥力和农业可持续发展以及全球碳平衡研究都具有十分重要的意义。

### [参 考 文 献]

- [1] Goovaerts P. Geostatistical modelling of uncertainty in soil science[J]. Geoderma, 2001, 103(1-2): 3-26.
- [2] Goovaerts P. Geostatistics in soil science: state-of-the-art and perspectives[J]. Geoderma, 1999(1-2), 89: 1-45.
- [3] 孙英君, 王劲峰, 柏延臣. 地统计学方法进展研究[J]. 地球科学进展, 2004, 19(2): 268-274.
- [4] 李笑吟, 毕华兴, 张 志, 等. 晋西黄土区土壤水分有效性分析的克立格法[J]. 土壤学报, 2006, 43(6): 1004-1010.
- [5] Rodgers S E, Oliver M A. A Geostatistical Analysis of Soil, Vegetation, and Image Data Characterizing Land Surface Variation[J]. Geographical Analysis, 2007, 39(2): 195-216.
- [6] Goovaerts P. Geostatistics for Natural Resources Evaluation [M]. New York: Oxford University Press, 1997.
- [7] 史 舟, 李 艳, 程街亮. 水稻土重金属空间分布的随机模拟和不确定评价[J]. 环境科学, 2007, 28(1): 209-214.
- [8] 李保国, 胡克林, 陈 研, 等. 农田土壤表层饱和和导水率的条件模拟[J]. 水利学报, 2002, (2): 36-40.
- [9] 陈翠英, 江永真. 土壤养分空间变异性的随机模拟及其应用[J]. 农业机械学报, 2006, 37(12): 67-95.
- [10] 戴明新, 赵玉杰, 白志鹏, 等. 序贯高斯法模拟新沂、东海农业区土壤磷空间变异性[J]. 中国环境科学, 2007, 27(2): 221-225.
- [11] Chirico G B, Medina H, Romano N. Uncertainty in predicting soil hydraulic properties at the hillslope scale with indirect methods[J]. Journal of Hydrology, 2007, 334(3-4): 405-422.
- [12] 王 红, 刘高焕, 宫 鹏. 利用 Cokriging 提高估算土壤盐离子浓度分布的精度——以黄河三角洲为例[J]. 地理学报, 2005, 60(3): 511-518.
- [13] 杨劲松, 姚荣江, 刘广明, 等. 黄河三角洲地区土壤盐分的空间变异性及其 CoKriging 估值[J]. 干旱区研究, 2006, 23(3): 439-445.
- [14] Baxter S J, Oliver M A. The spatial prediction of soil mineral N and potentially available N using elevation[J]. Geoderma, 2005, 128(3-4): 325-339.
- [15] Zhao Y, Shi X, Yu D, et al. Uncertainty assessment of spatial patterns of soil organic carbon density using sequential indicator simulation, a case study of Hebei province, China[J]. Chemosphere, 2005, 59(11): 1527-1535.
- [16] 许红卫, 高玉蓉, 王 珂, 等. 基于水稻冠层光谱信息的稻田土壤速效 N 的 Cokriging 插值研究[J]. 农业工程学报, 2007, 23 (3): 13-17.
- [17] Deutsch C V, Journel A G. GSLIB: Geostatistical Software Library and User's Guide[M]. New York: Oxford University Press, 1998.
- [18] Bourennane H, King D, Couturier A, et al. Uncertainty assessment of soil water content spatial patterns using geostatistical simulations: An empirical comparison of a simulation accounting for single attribute and a simulation accounting for secondary information[J]. Ecological Modelling, 2007, 205(3-4): 323-335.

## Random simulation of soil organic matter using elevation as auxiliary information

Chai Xurong<sup>1,2</sup>, Huang Yuanfang<sup>1\*</sup>, Yuan Xiaoyong<sup>1</sup>, Gao Rutai<sup>3</sup>

(1. Key Laboratory of Plant-soil Interaction, Ministry of Education; Key Laboratory of Soil and Water, Ministry of Agriculture; Department of Soil and Water Science, China Agricultural University, Beijing 100094, China;

2. College of Urban and Environmental Science, Shanxi Normal University, Linfen 041000, China;

3. College of Resources and Environmental Science, Hebei Agricultural University, Baoding 071001, China)

**Abstract:** The object of this study is to determine whether elevation could be used to enhance the accuracy of prediction of soil organic matter spatial patterns and the accuracy of modeling the prediction uncertainty. The study was conducted on a area in Pinggu district of Beijing. Soil organic matter was used as target variable. The sequential Gaussian simulation accounting for single attribute and the sequential Gaussian co-simulation accounting for elevation secondary information were used for simulation of the soil organic matter. Thus, the results of the two methods were compared. The results showed that the accuracy of the spatial prediction, the accuracy of modeling the local uncertainty as well as the spatial uncertainty were increased by using the method of condition simulation accounting for the intensive elevation information. The findings are tremendously significant for sustainable agricultural production and global carbon change modeling.

**Key words:** soil organic matter, sequential Gaussian simulation, sequential Gaussian co-simulation, uncertainty, elevation