

不同随机模拟方法定量土壤镉含量预测的不确定性研究

赵永存¹, 孙维侠¹, 黄 标¹, 王志刚^{1,2}, 史学正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), 江苏 南京 210008; 2. 中国科学院研究生院, 北京 100039)

摘 要:定量土壤重金属空间预测的局部不确定性对于控制土壤空间数据的应用风险具有重要意义。描述统计分析表明,张家港 541 个土壤表层(0~15 cm)样品的 Cd 含量变异性较大,而 pH 的变异性较小,并且土壤 Cd 含量与 pH 存在极显著正相关($r=0.51$, $P<0.001$)。本文采用序贯指示模拟、序贯指示协同模拟和 270 个 Cd 模拟集数据对 Cd 含量进行了 1 000 次条件随机模拟。结果表明,序贯指示模拟的中值估计对 271 个验证点 Cd 含量预测的均方根误差为 0.080,增加 pH 后的序贯指示协同模拟预测的均方根误差降低为 0.069,预测精度有所提高。对定量土壤 Cd 含量预测的局部不确定性而言,随着临界概率值的增大,这两种随机模拟方法对验证点位置不同 Cd 阈值的局部不确定性定量的分类错误率也越来越大。如果选择临界概率值为 0.50 可满足要求,则序贯指示模拟的分类错误率要小于序贯指示协同模拟;如果要求较高的可信度、选择临界概率为 0.95,序贯指示协同模拟的分类错误率则比序贯指示模拟要低。

关键词:土壤制图;局部不确定性;随机模拟;序贯指示模拟;序贯指示协同模拟

中图分类号:X53 **文献标识码:**A **文章编号:**1672-2043(2008)01-0139-08

Comparison of Sequential Indicator Simulation and Sequential Indicator Co-simulation for Quantifying the Local Uncertainty of Soil Cd Content

ZHAO Yong-cun¹, SUN Wei-xia¹, HUANG Biao¹, WANG Zhi-gang^{1,2}, SHI Xue-zheng¹

(1. State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, the Chinese Academy of Sciences, Nanjing 210008, China; 2. Graduate School of the Chinese Academy of Sciences, Beijing 100039, China)

Abstract: The local uncertainties associated with the prediction of heavy metal contents at un-sampled locations are important for decision-makings in the heavy metal pollution management. A total of 541 topsoil (0~15 cm) samples were collected from Zhangjiagang city of China for modelling the local uncertainty of Cd contents. The descriptive statistics indicated that high variations existed in Cd contents whereas small variations for soil pH data, and the positive correlations between Cd and pH were significant ($r=0.51$, $P<0.001$). The 1 000 times of sequential indicator simulation (SIS) and sequential indicator co-simulation (SIcS) using 270 Cd data (the simulation set) and 541 pH soft indicator data suggested that the median estimates of SIcS at 271 validation points were of lowest root mean square error. As for the abilities of modeling the local uncertainty of Cd contents, the misclassification rate of Cd content at 271 validation points given by different threshold values increased with the increase of selected critical probability. If the selected critical probability was 0.50, the SIS would have smaller misclassification rate; if high critical probability was necessary, i.e. 0.95, the SIcS incorporating 541 soil pH soft data was of low misclassification rate compared with SIS.

Keywords: soil mapping; local uncertainty; stochastic simulation; sequential indicator simulation; sequential indicator co-simulation

收稿日期:2007-03-30

基金项目:江苏省自然科学基金(BK2006581);中国科学院南京土壤研究所知识创新工程领域前沿项目(ISSASIP0604);中国科学院“优秀博士学位论文、院长奖获得者科研启动专项资金”

作者简介:赵永存(1973—),男,博士,助理研究员,从事空间数据质量、尺度效应和不确定性建模方面的研究。E-mail: yczhao@issas.ac.cn

通讯作者:黄 标 E-mail: bhuang@issas.ac.cn

土壤属性制图的一个重要目的就是提供有关土壤性质空间变异的信息,但由于空间预测模型的限制等,使得由采样点位置的观测值来预测土壤性质空间分布的过程中通常包含一定的不确定性(Uncertainty),这种不确定性将通过随后的分析或决策过程加以传递,并对最终的应用结果产生深刻影响^[1]。

单一空间位置土壤性质预测的不确定性称为局部不确定性(local uncertainty)^[2]。在空间位置 u 的连续土壤属性 $z(u)$ 预测的不确定性可以通过概率模型来定量,即把 $z(u)$ 看作是随机变量 $Z(u)$ 的实现(realization)并获取 $Z(u)$ 的条件累积分布函数(ccdf)来建模不确定性。在建模土壤性质不确定性方面,目前最常用的就是随机模拟方法:Zhao 等^[1]采用序贯指示模拟方法评价了河北省土壤碳密度空间分布表达的不确定性;李保国等^[3]采用序贯高斯模拟对农田土壤表层饱和导水率进行了条件模拟;王学军^[4]应用转向带法对北京东郊污灌区土壤铜和铅含量进行了条件模拟;而徐英和陈亚新^[5]则运用条件模拟对土壤水盐特性空间变异的各向同性进行了探讨。在模拟过程中增加与模拟主变量相关的协同变量则称为协同模拟(co-simulation),Hendriks 等^[6]探讨了使用“软”数据(soft data,指间接测定的结果)来提高环境空间预测精度问题;Franco 等^[7]利用不同种类重金属间的相关性和序贯协同模拟的方法对西班牙南部瓜的亚纳河边缘地区土壤重金属进行了污染风险制图。与克里格(kriging)或指示克里格(indicator kriging)相比,随机模拟方法在建模土壤性质不确定性方面的优势,表现在它能够综合考虑结果的整体统计性质和模拟值的空间相关性,并且没有平滑效应。此外,随机模拟能够产生多个等概率实现,通过 ccdf 来建模未采样点位置土壤性质预测的局部不确定性和空间不确定性(spatial uncertainty)^[1,8]。

目前建模土壤性质局部不确定性最常用的随机模拟方法包括序贯高斯模拟、序贯高斯协同模拟、序贯指示模拟和序贯指示协同模拟^[9]。序贯高斯(协同)模拟在建模 ccdf 时采用的是参数化的方法,其前提是数据符合多高斯分布(Multi-Gaussian distribution),但在实践中很难确定多高斯分布是否存在,而序贯指示模拟则不考虑分布类型,在不能确定多高斯分布是否存在时序贯指示模拟是一个明智的选择^[2]。本研究以江苏省张家港市土壤 Cd 含量为例,探讨了序贯指示模拟和序贯指示协同模拟定量土壤 Cd 含量局部不确定性的优劣,这对于控制该区土壤重金属含量空

间分布信息的应用风险具有重要的理论和现实意义。

1 研究方法

1.1 研究区域概况及土样采集与分析

研究区域为江苏省张家港市(图 1)。该市地处长江三角洲冲积平原,地势平坦。该区属北亚热带季风气候,年平均气温 15.2 °C,年均降雨量 1 039 mm。全市总面积 999 km²,其中陆地面积 799 km²。该市土壤类型主要有潮土、水稻土两个土类,潮土主要分布于北部的沿江圩田地区,水稻土则主要分布于南部的平田地区。此外黄棕壤、堆叠土和园田土在该区亦有零星分布,但数量极少。土壤采样时间为 2004 年,共采集了 541 个土壤表层样品(0~15 cm),土壤全 Cd 含量采用 HNO₃-HClO₄ 消解、石墨炉原子吸收法测定(3510,安捷伦-上海分析仪器有限公司),土壤 pH 采用玻璃电极法测定(PHS-3C),土水比 1:2.5^[10]。541 个采样点中随机抽取 270 个作为模拟数据集(simulation set)进行 Cd 含量局部不确定性建模,其余 271 个作为验证数据集(validation set)进行建模的准确性检验(图 1)。

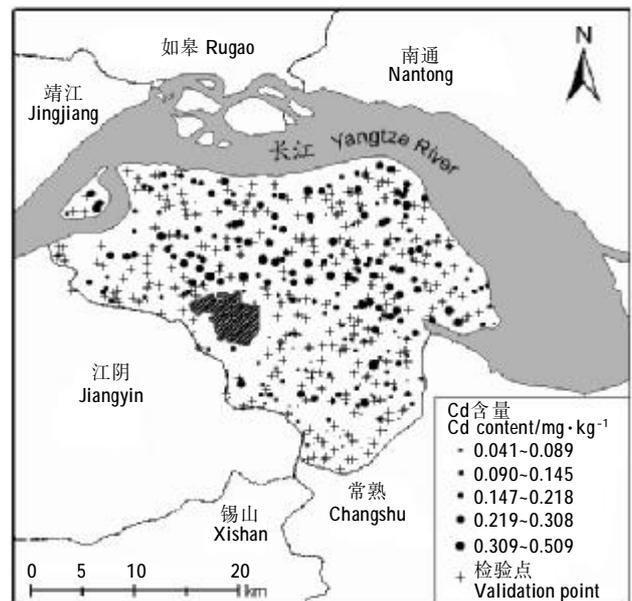


图 1 研究区域及采样点分布图

Figure 1 Geographic location of Zhangjiagang city and the distribution of sampling sites

1.2 序贯指示模拟和序贯指示协同模拟

1.2.1 序贯指示模拟 (Sequential indicator simulation, SIS)^[2]

(1) 选取 K 个阈值(cutoff value) Z_1, \dots, Z_K 对模拟集 Cd 含量 $Z(u)$ 进行二态指示变换,指示变量 i

(u, z_k) 的定义为:

$$i(u, z_k) = \begin{cases} 1 & \text{if } z(u) \leq z_k \\ 0 & \text{otherwise} \end{cases} \quad k=1, \dots, K \quad [1]$$

(2) 建立指示变量 $i(u, z_k)$ 的半方差模型。

(3) 定义一条遍历所有待模拟位置的随机路径, 通过以下步骤实现序贯指示模拟:

a) 对于随机路径上某一位置 u , 通过 kriging 估计指示随机变量 $i(u, z_k)$, 其估计值 $i^*(u, z_k) = \text{Prob}^*(Z(u) \leq z_k)$ 。

b) 纠正次序关系错误后, 估计 Cd 含量 $z(u)$ 的累积分布函数 $F_{z(u)}$, 从 $F_{z(u)}$ 随机获取一个模拟值并将该值赋予位置 u 。

c) 对随机路径上的每一位置都重复步骤 a) ~ b) 便获得一个随机模拟实现 (realization)。重复 L 次序贯模拟, 每次模拟均使用不同的随机路径则将产生 L 个等概率的模拟实现。

1.2.2 序贯指示协同模拟 (Sequential indicator co-simulation, SIcS) [11]

序贯指示协同模拟与序贯指示模拟类似, 主要区别在于在模拟过程中增加了与主变量相关的协同变量。假设土壤 pH 值 $Y(u)$ 是与 Cd 含量 $Z(u)$ 相关的变量, 如果条件先验概率函数 $F(Z(u) \leq z_k | Y(u))$ 已知, 该函数可用于土壤 pH 值 $Y(u)$ 对 Cd 含量 $Z(u)$ 的影响进行指示编码: $i_k(u) = F(Z(u) \leq z_k | Y(u))$ $k=1, \dots, K$, 其中 $0 \leq i_k(u) \leq 1$ 。由于土壤 pH 值 $Y(u)$ 为连续变量, 土壤 Cd 含量的条件先验概率函数 $F(Z(u) \leq z_k | Y(u))$ 可由下式估计:

$$F^*(Z \leq z_k | Y \in [y_1, y_{1+1}]) = \frac{1}{\sum_{u_\alpha \in u} i(u_\alpha, [y_1, y_{1+1}])}$$

$$\sum_{u_\alpha \in u} i(u_\alpha, z_k) i(u_\alpha, [y_1, y_{1+1}]) \quad [2]$$

式中: $y_1 = y_1, \dots, y_k$ 是用于将土壤 pH 变量 Y 进行离散化处理 (划分为不同的范围) 的界限值; u 是一系列 Cd 含量 Z 和 pH 值 Y 均已知采样点位置, 并且:

$$i(u_\alpha, [y_1, y_{1+1}]) = \begin{cases} 1 & \text{if } y(u_\alpha) \in [y_1, y_{1+1}] \\ 0 & \text{otherwise} \end{cases} \quad [3]$$

序贯指示协同模拟中 Cd 的条件分布函数是通过协同 kriging 估计的, 对于每个阈值 z_k 都需要两个协方差模型和一个交互协方差模型, 因此计算量是相当大的。然而, 在假定马尔可夫性 (Markov property) 的条件下, 计算量可以大大减轻。在假定下列马尔可夫性的条件下:

$$\text{Prob}(Z(u') \leq z_k | i(u, z_k), y(u, z_k)) = \text{Prob}\{Z(u') \leq z_k | i(u, z_k)\} \quad [4]$$

不同阈值 z_k 时, Cd 软指示数据 (soft indicator data) (由相同位置的 Cd 含量和 pH 数据计算获得) 的协方差 $C_Y(h, z_k)$ 及 Cd 含量 (硬) 指示数据 (hard indicator data) 与软指示数据的交互协方差 $C_{YV}(h, z_k)$ 可由下式计算 [12]:

$$C_Y(h, z_k) = \begin{cases} B(z_k) C_1(h, z_k) & \text{if } \|h\| = 0 \\ B(z_k)^2 C_1(h, z_k) & \text{if } \|h\| \neq 0 \end{cases}$$

$$C_{YV}(h, z_k) = B(z_k) C_1(h, z_k) \quad [5]$$

式中: h 为距离 (lag); $C_1(h, z_k)$ 为 Cd 含量的指示协方差; $B(z_k)$ 为条件期望值 $m_1(z_k)$ 与 $m_0(z_k)$ 之差, 即 $B(z_k) = m_1(z_k) - m_0(z_k) \in [-1, 1]$, 其中 $m_1(z_k) = E\{Y(u, z_k) | i(u, z_k) = 1\}$; $m_0(z_k) = E\{Y(u, z_k) | i(u, z_k) = 0\}$, 即通过校正系数 $B(z_k)$ 和土壤 Cd 含量指示协方差模型来推断软指示数据的协方差模型及 Cd 含量 (硬) 指示数据与软指示数据的交互协方差模型, 上述过程称为马尔可夫 - 贝叶斯算法。

本研究中共进行了 1 000 次序贯指示模拟和 1 000 次序贯指示协同模拟, 序贯指示协同模拟中将 541 个土壤 pH 值作为协同模拟变量, 使用模拟数据集中的 270 个采样点的土壤 Cd 含量、pH 值和公式 [2] ~ [3] 来建立验证数据点位置 Cd 含量的先验概率。验证结果表明马尔可夫假定 [5] 成立, 因此基于公式 [5] 计算不同阈值 z_k 对应的交互协方差。校正系数 $B(z_k)$ 和 Cd 软指示数据计算采用的是 GSLIB 软件包中的 “bicalib” 程序。本研究所进行序贯指示模拟和序贯指示协同模拟均为条件模拟, 模拟过程中模拟数据集中采样点位置的 Cd 含量值保持不变, 模拟所采用的程序为 GSLIB 软件包中的 “sisim” 程序 (协同模拟过程中, 由于 sisim 程序默认的模拟结点数量和协同模拟变量样本数量的限制, 本研究修改了部分 Fortran 程序源代码, 重新编译了 sisim 程序)。

1.3 模拟预测的准确性和局部不确定性

准确性评价: 序贯指示模拟和序贯指示协同模拟预测的准确性采用验证集采样点位置 Cd 含量的平均预测误差 (MPE) 和均方根预测误差 (RMSPE) 来评价, MPE 和 RMSPE 的计算方法如下:

$$MPE = \frac{1}{l} \sum_{j=1}^l [z(s_j) - \hat{z}(s_j)]$$

$$RMSPE = \sqrt{\frac{1}{l} \sum_{j=1}^l [z(s_j) - \hat{z}(s_j)]^2} \quad [6]$$

式中: l 为验证集中采样点的个数; $z(s_j)$ 为验证数据点

位置 j 上土壤 Cd 含量 S_j 的实际观测值, $\hat{z}(s_j)$ 为 $z(s_j)$ 的估计值。

局部不确定性:单一空间位置 x' 上 Cd 含量的不确定性,即局部不确定性(local uncertainty)可以用该点 Cd 含量 $z(x')$ 大于或小于等于某一给定 Cd 含量阈值 z_1 的概率 $P_g = \text{Prob}[z(x') > z_1]$ 或 $P_e = \text{Prob}[z(x') \leq z_1]$ 来定量:

$$P_g = \text{Prob}[z(x') > z_1] = \frac{n(x')}{1000} \quad P_e = 1 - P_g \quad [7]$$

式中:1 000 为随机模拟的次数; $n(x')$ 为 1 000 次模拟中在 x' 位置上 Cd 含量模拟值大于阈值 z_1 的次数。

分类错误率:给定不同临界概率值 p_c 和 Cd 含量阈值 z_1 (比如环境质量标准值),基于规则 $\text{Prob}[z(x') > z_1] \geq p_c$ 或 $\text{Prob}[z(x') \leq z_1] \geq p_c$ 对 Cd 含量大于 z_1 的点(比如污染点)或 Cd 含量不超过 z_1 的点(比如清洁点)进行划分时包含两类分类错误,即实际观测值大于给定阈值而模拟结果小于该阈值(假阴性错误),例如实际上为污染,模拟结果划分为清洁;或观测值小于等于阈值而模拟划分为大于该阈值(假阳性错误),比如实际为清洁,模拟结果划分为污染^[3]。因此随机模拟的分类错误率 SM (Site misclassification) 可由下式计算:

$$\text{SM}\% = \frac{1}{I} (N_g + N_e) \times 100 \quad [8]$$

式中: I 为验证数据集中采样点的个数; N_g 为实际观测值大于 z_1 而随机模拟划分为小于等于 z_1 的点的个数; N_e 为实际观测值小于等于 z_1 而随机模拟划分为大于 z_1 的点的个数。

2 结果与讨论

2.1 描述统计分析和指示半方差模型

研究区土壤表层 Cd 含量和 pH 平均值分别为 $0.173 \text{ mg} \cdot \text{kg}^{-1}$ 和 7.41, 极差分别为 $0.513 \text{ mg} \cdot \text{kg}^{-1}$ 和 3.95。Cd 含量的描述统计分析结果表明土壤 Cd 具有较强的变异性,541 个采样点计算的 Cd 变异系数为 52%,而土壤 pH 值的变异性相对较小,变异系数仅为 13%。土壤 Cd 含量的偏度为 0.716,频率分布表现为

右侧长尾分布(表 1)。土壤 Cd 含量与 pH 存在极显著正相关关系 ($P < 0.001$),即研究区土壤 pH 高的位置其土壤 Cd 含量也较高。为了比较两种随机模拟方法预测局部点位置土壤 Cd 含量大于或小于给定阈值的概率,土壤 Cd 含量数据被随机划分为模拟集和验证集,表 1 的结果表明,随机抽取的 270 个 Cd 模拟集数据及 271 个验证集数据与全部 541 个采样点 Cd 含量具有相似的统计特性,比如均值、标准差、偏度、Cd 与 pH 的相关性等,这说明本研究中模拟数据集和验证数据集的划分是合理的,具有较好的代表性。

本研究中使用了 9 个门槛值进行 Cd 含量的指示变换,门槛值为模拟集 Cd 含量的 9 个十分位数,相应指示编码的半方差图和模型拟合参数(图 2)均由统计计算软件 R^[14] 中的 geoR^[15] 模块计算获得,半方差函数拟合采用的是权重最小二次方法(WLS)。从图 2 中可以看出 10%分位数和 90%分位数半方差模型拟合的均方根误差(RMSE)相对稍大,这主要是由于极端阈值条件下数据点对数过少而造成的。本研究中的 9 个指示半方差图都具有明显的空间结构性,并且拟合的参数差异也比较明显,这表明使用这 9 个门槛值能够较好地体现研究区土壤 Cd 含量的整体变异性特征。

2.2 模拟预测结果的准确性分析

通过计算序贯指示模拟和序贯指示协同模拟的条件均值(亦称为 E 型均值, E-type estimate)或中值可以对验证点位置的土壤 Cd 含量进行预测,其中条件均值可以通过取 1 000 个模拟实现的平均值获得,中值可以通过取 1 000 个模拟实现的中值获得。从表 2

表 2 Cd 含量预测结果准确性分析
Table 2 Prediction accuracy of SIS and SlcS for soil Cd content at validation points

	序贯指示模拟		序贯指示协同模拟	
	均值估计	中值估计	均值估计	中值估计
平均预测误差	0.009	-0.007	0.009	-0.005
均方根预测误差	0.081	0.080	0.077	0.069

表 1 土壤 Cd 含量和 pH 的描述统计分析

Table 1 Descriptive statistics for soil Cd content and pH

	样本数	均值	标准差	最小值	中值	最大值	变异系数	偏度	与 pH 相关系数
		$\text{mg} \cdot \text{kg}^{-1}$					%		
Cd	541	0.173	0.091	0.020	0.163	0.533	52	0.716	0.51 ($P < 0.001$)
pH	541	7.41	0.94	4.40	7.94	8.35	13	-1.100	—
Cd 模拟集	270	0.176	0.092	0.041	0.163	0.509	53	0.707	0.49 ($P < 0.001$)
Cd 验证集	271	0.171	0.090	0.020	0.166	0.533	52	0.727	0.52 ($P < 0.001$)

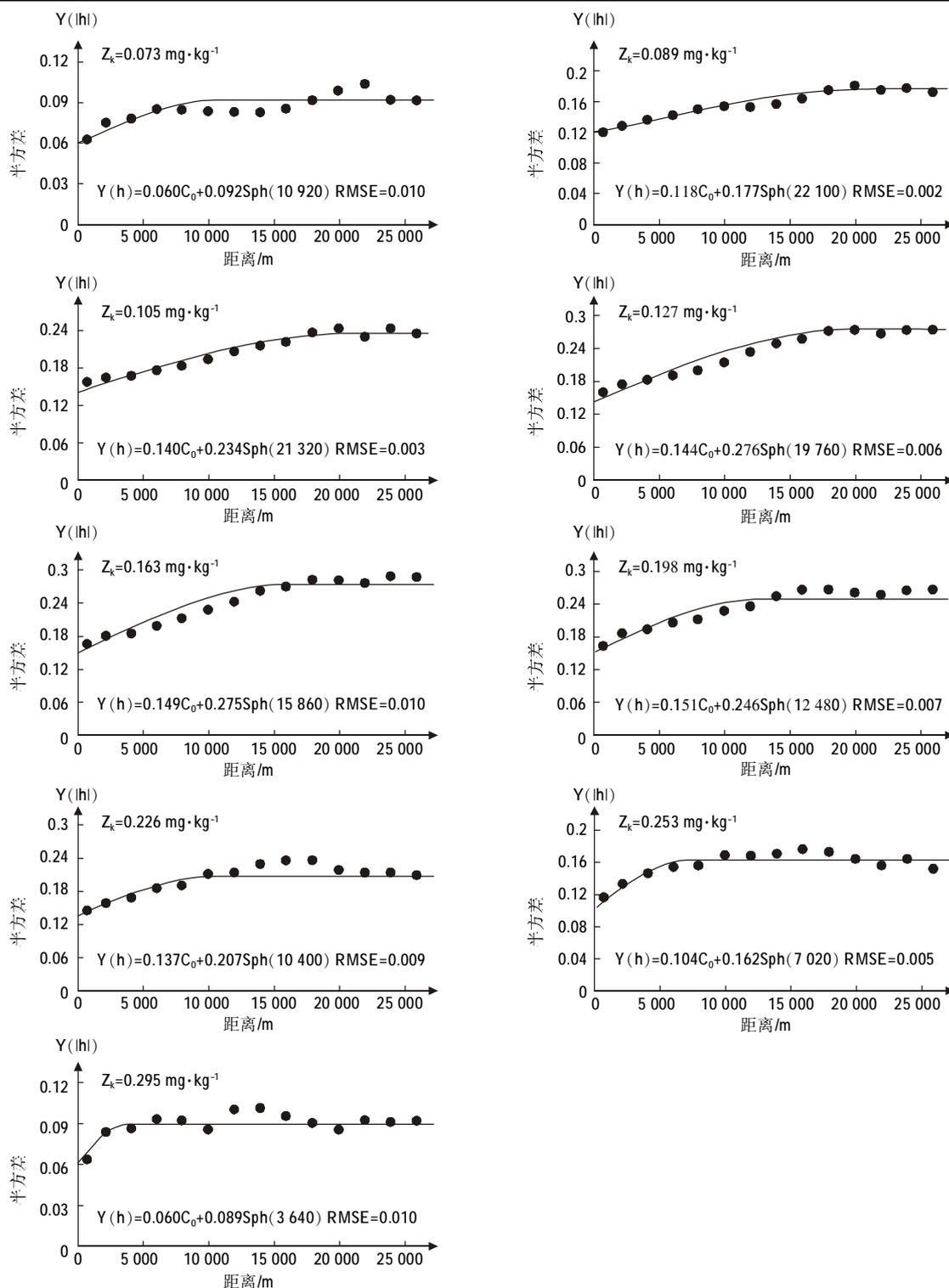


图2 不同门槛值 Z_k 时 Cd 含量的指示半方差图和拟合模型参数(半方差模型表示为 $aC_0+bSph(c)$, 其中 a 为块金常数, b 为基台值, c 为变程, Sph 表示球状模型, RMSE 为模型拟合的均方根误差)

Figure 2 Indicator variograms and fitted models of Cd for the nine cutoff values Z_k (The fitted models of the variograms were presented as ' $aC_0+bSph(c)$ ', where a , b and c represent nugget, sill and range, respectively. RMSE means root mean square error)

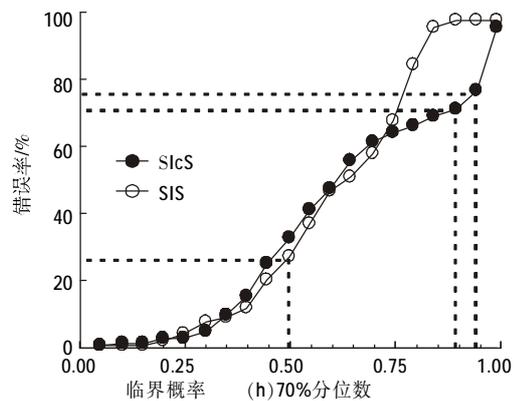
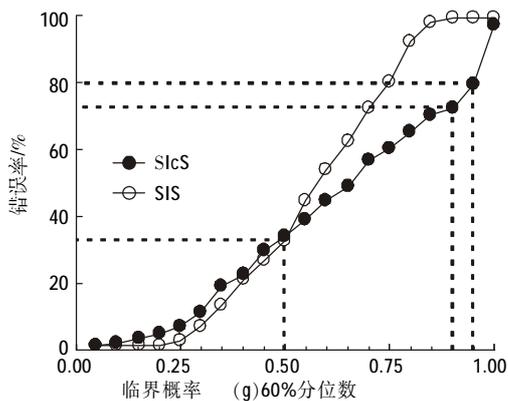
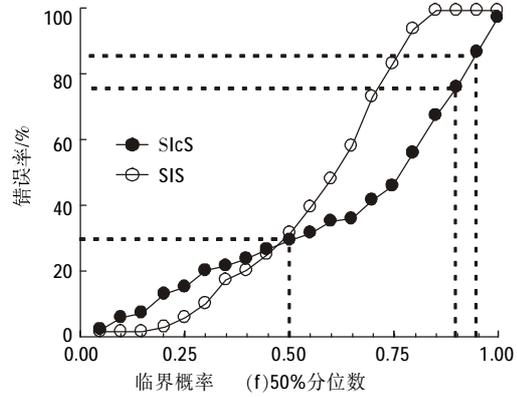
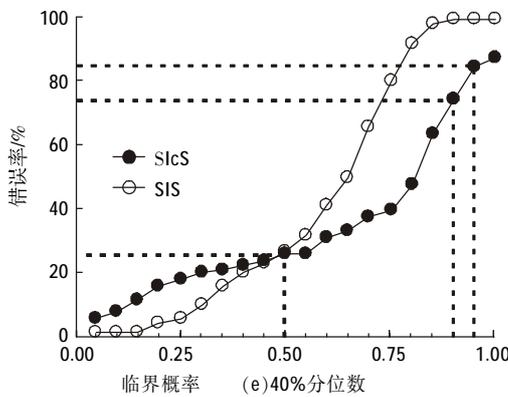
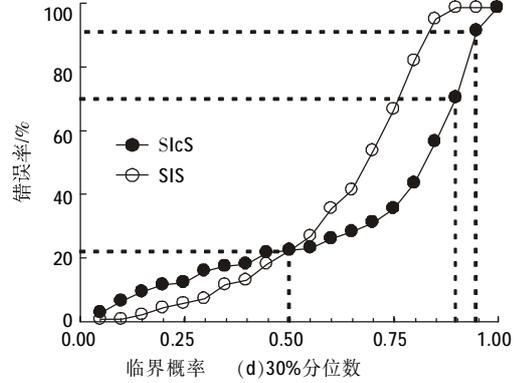
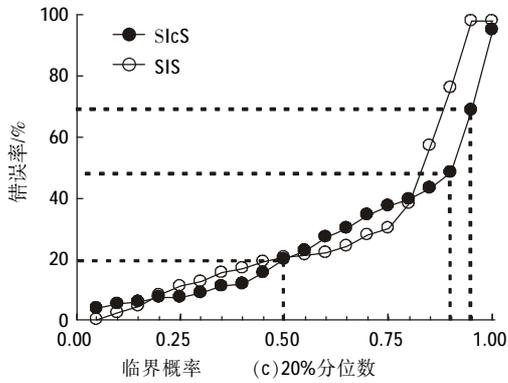
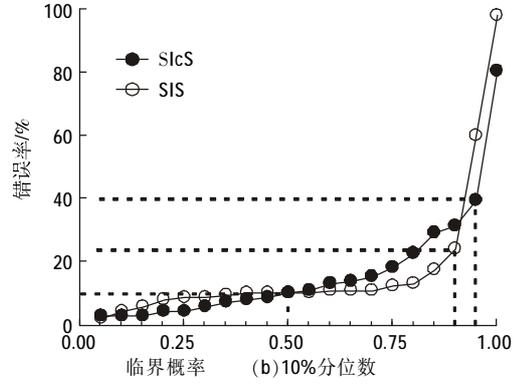
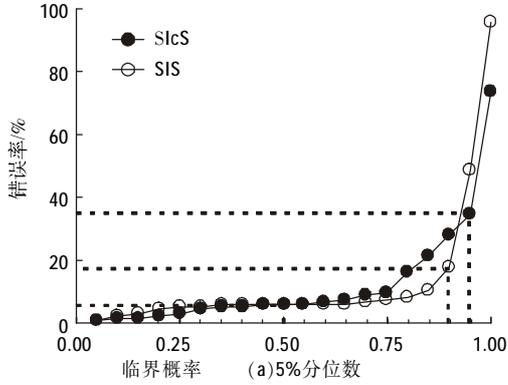
可以看出, 不论序贯指示模拟还是序贯指示协同模拟, 中值估计的预测平均误差和均方根预测误差均要低于条件均值估计的预测误差, 这主要是由于在非对

称分布条件下中值估计要比均值估计更稳健^[16]。同时序贯指示协同模拟的条件均值估计预测和中值估计预测的均方根预测误差都比序贯指示模拟方法低。由

此可见,序贯指示协同模拟的中值估计预测 Cd 含量的误差最低、在预测准确性方面序贯指示协同模拟的预测效果要好于序贯指示模拟。

2.3 模拟结果的分类错误率分析

使用空间模型来预测未采样点位置的污染物含量通常包含一定的不确定性,比如实际含量大于某一给定阈值而预测结果为小于该阈值从而导致假阴性误差 (False negative errors),或者是实际含量小于某



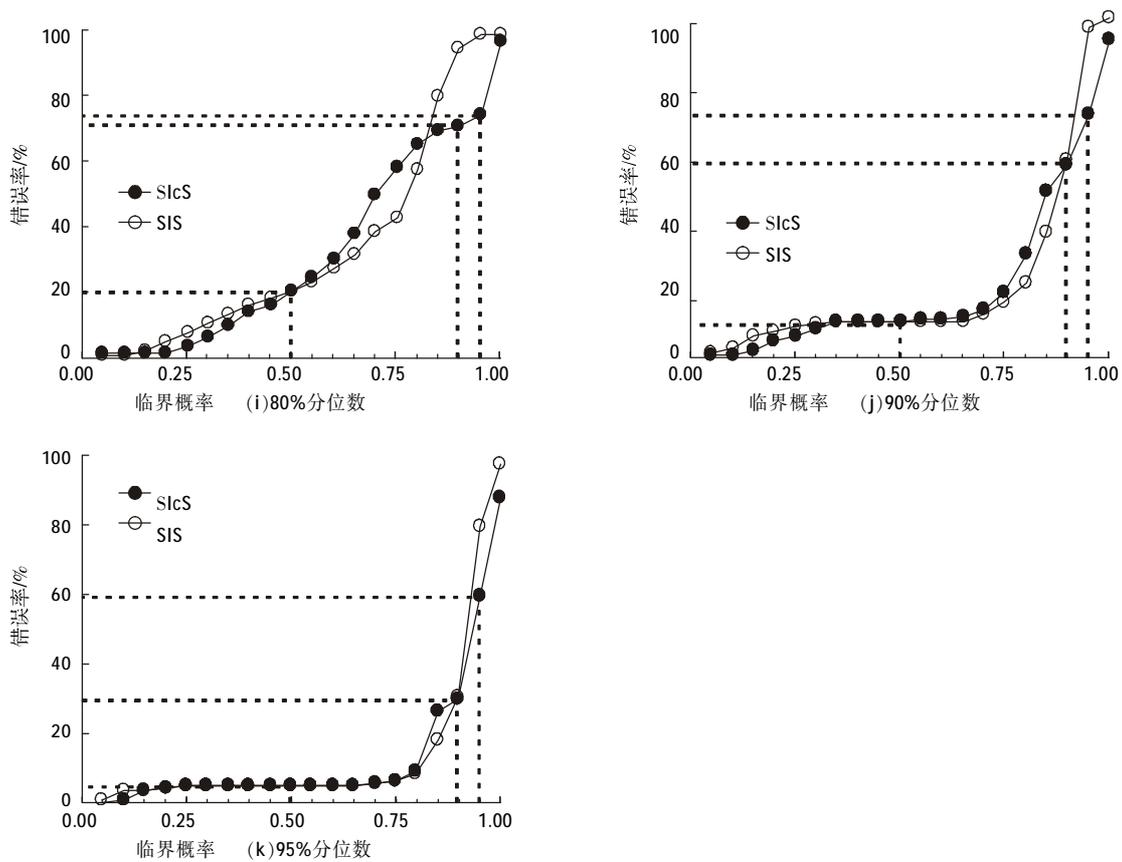


图 3 不同阈值 z_t 时序贯指示模拟(SIS)和序贯指示协同模拟(SlcS)的分类错误率

Figure 3 The misclassification rate of SIS and SlcS of Cd content for eleven threshold values

一给定阈值而预测结果为大于该阈值从而导致假阳性误差 (False positive errors), 这均会对污染风险评价结果带来潜在的影响, 因此必须评价预测结果的不确定性。序贯指示模拟和序贯指示协同模拟除了通过模拟均值、中值预测未采样点位置的 Cd 含量外, 还可以通过局部条件累积分布函数 (ccdf) 来定量 Cd 含量预测的局部不确定性, 即通过临界概率的方法 (公式[8]) 确定未采样点位置 Cd 含量大于 / 小于给定阈值 z_t 。

图 3 是不同临界概率 p_c 条件下序贯指示模拟和序贯指示协同模拟对验证集 11 个分位数阈值 (阈值即公式[7]中的 z_t) 建模局部不确定性的误差统计 (即模拟 Cd 含量大于 / 小于等于这 11 个分位数的分类错误率), 其中选择验证集 Cd 含量的 5% 和 95% 分位数作为阈值, 主要是为了比较极端阈值条件下这两种随机模拟方法建模 Cd 含量局部不确定性的误差。从图 3 中可以看出, 随着选择的临界概率值的增大, 序贯指示模拟和序贯指示协同模拟对于不同 Cd 阈值局部不确定性建模的分类错误率也越来越大, 而较小

的临界概率时 (比如 0.10) 模拟的分类错误率也较低。但临界概率小于等于 0.50 时的结果是不可靠的, 因为 Cd 含量大于给定阈值和 Cd 含量小于等于给定阈值的可信程度均低于 50%, 即 1 000 次模拟中只有不到 500 次的模拟值大于给定阈值, 就认为实际值大于给定阈值; 或者是 1 000 次模拟中有不到 500 次的模拟值小于等于给定阈值, 就认为实际值小于给定阈值, 所以临界概率的选择至少要大于 0.50。当临界概率为 0.90 时, 序贯指示协同模拟对于低阈值 (验证集的 5% 分位数和 10% 分位数) 的分类误差比序贯指示模拟要大, 但当临界概率超过 0.95 时, 序贯指示协同模拟对于验证集 11 个 Cd 分位数阈值建模的分类错误率均要低于序贯指示模拟。实际上对于 Cd 局部不确定性建模时, 要保证临界概率可信程度的基础上尽量降低分类的错误率, 因此, 图 3 可以作为选择研究区 Cd 含量不确定性建模随机模拟方法的依据。从图 3 中可以看出, 如果临界概率为 0.50 即可满足应用要求的话, 选择序贯指示模拟定量研究区 Cd 含量的局部不确定性即可; 如果需要较高的可信度, 选择临界

概率为 0.95,那么序贯指示协同模拟的分类错误率则比序贯指示模拟要低,更适合于定量研究区土壤 Cd 含量的局部不确定性。

3 结论

(1)在预测准确性方面,增加协同变量 pH 后,序贯指示协同模拟的中值估计对研究区 Cd 含量的预测精度比序贯指示模拟要高。

(2)对定量土壤 Cd 含量预测的局部不确定性而言,1 000 次的序贯模拟结果表明,随着选择的临界概率值的增大,序贯指示模拟和序贯指示协同模拟对于不同 Cd 阈值局部不确定性建模的分类错误率也越来越大,如果选择的临界概率值为 0.50 即可满足要求,则序贯指示模拟的分类错误率较小;如果要求较高的可信度、选择临界概率为 0.95,那么增加相关的协同变量 pH 后,序贯指示协同模拟的分类错误率则比序贯指示模拟要低,因此更适合于定量研究区土壤 Cd 含量的局部不确定性。

参考文献:

- [1] Zhao Y C, Shi X Z, Yu D S, 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.
- [2] Goovaerts P. Geostatistical modeling of uncertainty in soil science [J]. *Geoderma*, 2001, 103: 3–26.
- [3] 李保国,胡克林,陈德立,等.农田土壤表层饱和和导水率的条件模拟[J].水利学报,2002,2:36–40,46.
- [4] 王学军.应用转向带法进行土壤铜和铅含量的条件模拟[J].应用生态学报,2002,13(12):1667–1670.
- [5] 徐英,陈亚新.土壤水盐特性空间变异的各向同性近似探讨[J].灌溉排水学报,2003,22(4):14–17,24.
- [6] Hendriks L A M, Leumens H, Stein A, et al. Use of soft data in a GIS to improve estimation procedures for environmental purposes [J]. *Water Air and Soil Pollution*, 1998, 101: 217–234.
- [7] Franco C, Soares A, Delgado J. Geostatistical modelling of heavy metal contamination in the topsoil of Guadiana river margins (S Spain) using a stochastic simulation technique [J]. *Geoderma*, 2006, 136: 852–864.
- [8] Juang K W, Chen Y S, Lee D Y. Using sequential indicator simulation to assess the uncertainty of delineating heavy-metal contaminated soils[J]. *Environmental Pollution*, 2004, 127: 229–238.
- [9] Goovaerts P. Geostatistics in soil science: State-of-the-art and perspectives [J]. *Geoderma*, 1999, 89: 1–45.
- [10] 鲁如坤.土壤农业化学分析方法[M].北京:中国农业科技出版社,2000.
- [11] Goovaerts P, Journel A G. Integrating soil map information in modelling the spatial variation of continuous soil properties [J]. *Eur J Soil Sci*, 1995, 46: 397–414.
- [12] Zhu H, Journel A G. Formatting and integrating soft data: stochastic imaging via the Markov–Bayes algorithm [C]/Soares A. ed. *Geostat Troia*. Dordrecht: Kluwer Academic Publishers, 2000. 1–12.
- [13] Cattle J A, McBratney A B, Minasny B. Kriging method evaluation for assessing the spatial distribution of urban soil lead contamination [J]. *J Environ Qual*, 2002, 31: 1576–1588.
- [14] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing [M]. Vienna, Austria. ISBN 3–900051–07–0, URL <http://www.R-project.org>. 2005.
- [15] Ribeiro Jr P J, Diggle P J. GeoR: a package for geostatistical analysis [J]. *R-NEWS*, 2001, 1(2): 15–18.
- [16] Isaaks E H, Srivastava R M. An introduction to applied geostatistics [M]. New York: Oxford Univ. Press, 1989.