

水稻土重金属空间分布的随机模拟和不确定评价

史舟^{1,2}, 李艳¹, 程街亮¹

(1. 浙江大学环境与资源学院农业遥感和信息技术应用研究所, 杭州 310029; 2. 浙江大学环境与资源学院环境修复与生态环境教育部重点实验室, 杭州 310029)

摘要:以杭嘉湖平原为样区, 利用普通克里格法和序贯高斯模拟方法对土壤 Cu 的空间分布进行估值和模拟, 并利用序贯指示模拟进行不确定性评价. 结果表明, 样区北部区域土壤 Cu 含量较高. 由普通克里格法得到的土壤 Cu 的空间分布整体比较连续, 具有明显的平滑效应, 估值结果数据的分布频率趋于平缓. 序贯高斯模拟结果整体分布相对离散, 突出了原始数据分布的波动性, 其模拟结果数据的分布频率相对集中. 以 $35 \text{ mg} \cdot \text{kg}^{-1}$ 作为土壤 Cu 含量的阈值进行的序贯指示模拟结果显示, 在土壤 Cu 含量较高的北部地区超过阈值的概率超过 80%, 而南部很多相对含量低的地区该概率值都低于 10%. 以超阈值概率为 0.85 和 0.75 来选取 Cu 的高污染风险区进行空间不确定评价结果表明, 联合概率比单点统计的概率更为严格, 在划分较大范围污染区域时, 最好同时采用联合概率来进行信度评价.

关键词: 重金属; 空间分布; 随机模拟; 不确定性评价

中图分类号: X53 文献标识码: A 文章编号: 0250-3301(2007)01-0209-06

Stochastic Simulation of Spatial Distribution and Uncertainty Assessment of Heavy Metal in Paddy Soil

SHI Zhou^{1,2}, LI Yan¹, CHENG Jie-liang¹

(1. Institute of Agricultural Remote Sensing and Information Technology Application, Zhejiang University, Hangzhou 310029, China; 2. Ministry of Education Key Laboratory of Environmental Remediation and Ecosystem Health, Zhejiang University, Hangzhou 310029, China)

Abstract: Ordinary kriging and sequential Gaussian simulation (SGS) were used to estimate and simulate spatial distribution of heavy metal Cu in paddy soil in Hang-Jiang-Hu plain, and sequential indicator simulation (SIS) was used to assess the probability of Cu content exceeding the predefined threshold. The results indicated that north region in the study area has high Cu content. The spatial distribution of Cu obtained by ordinary kriging was continuous and smooth, whereas that obtained by the SGS was relatively discrete and fluctuant. The results of the SIS indicated that the conditional probability of Cu content exceeding the threshold of 35 mg/kg was greater than 80% in the north area, whereas in south area the probability of Cu content exceeding 35 mg/kg was less than 10%. Based on the joint probability statistics, the spatial uncertainty (multi-points uncertainty) obtained from the SIS was used to assess the reliability of delineation for regions suspected of contamination. The result showed that assessment of spatial uncertainty are more strict than that of single-point uncertainty. Therefore, it is more acceptable and convinced that these two approaches of uncertainty assessment were used simultaneously for delineating the contaminated area.

Key words: heavy metal; spatial distribution; stochastic simulation; uncertainty assessment

对土壤中重金属含量的空间分布进行估值、模拟和制图对进行土壤环境质量评估是很重要的. 目前, 地统计学的克里格插值法被广泛用于对土壤中重金属含量及其空间分布进行预测^[1-5]. 然而大量的研究发现, 克里格插值法具有一定的平滑效应^[6,7]. 在土壤重金属的研究中, 有时研究者关注的是某些重金属空间数据变化剧烈的区域, 克里格估值法对数据进行平滑处理后, 就会使这些变化剧烈, 甚至可能是异常区的重要信息丢失.

地统计学的空间随机模拟法被提出来克服克里格法的缺陷^[8]. 它将数据作为一个整体来复原其整体的空间结构, 追求的是模拟的真实性, 尽可能地接

近真实的空间分布, 不像克里格方法, 追求的是特定点位某个属性的局部最优估值. 同时, 空间随机模拟还可直接用于空间不确定性研究, 这也是克里格插值法欠缺的. 因此, 近年来空间随机模拟方法的应用逐渐增加, 成为地统计学研究主要趋势之一.

空间随机模拟最初主要应用于矿藏勘探评估, 后来才逐渐应用于土壤学、环境学等^[9]. 如对土壤表

收稿日期: 2006-02-08; 修订日期: 2006-03-21

基金项目: 国家自然科学基金项目(40001008); 浙江省科技计划项目(2004C32066)

作者简介: 史舟(1970~), 男, 博士, 副教授, 主要研究方向为遥感和地理信息系统、土壤特性空间变异和精确农业, E-mail: shizhou@zju.edu.cn

土含水量超阈值的概率分布评价^[10]、农田土壤表层饱和导水率的模拟^[11]、对水盐空间变异的进行模拟^[12]等.在研究受重金属污染的土壤方面,空间随机模拟也有大量的应用实例.如 Lin 等人利用条件模拟的序贯高斯模拟和模拟退火方法对土壤重金属 Pb 的空间分布进行制图,并与克里格方法估值结果进行比较^[13,14].Juang 等人利用序贯指示模拟方法对土壤重金属 Cu 空间分布进行描述,并采用概率分布图来评价污染风险程度^[15].

本研究以杭嘉湖平原水稻土重金属调查数据为例,利用序贯高斯模拟方法进行土壤重金属空间分布模拟,并利用序贯指示模拟进行土壤重金属超标的概率分布评价,以期在环境污染指标的确定以及土壤质量的评估等研究和实践领域,为管理决策者进行合理管理和决策提供一种定量方法.

1 样区与数据

杭嘉湖平原是我国传统高产综合性农业区之一.近年来,随着地区经济的快速增长、城市化进程的加快以及人口的不断增加,农业环境问题,特别是土壤非点源污染和土壤重金属污染等问题逐渐引起重视.本实例数据取自国土资源部与浙江省政府从 2002 年开始的合作项目“浙江省农业地质环境调查”的子项目“农业地质环境与农产品安全研究”的部分调查数据.样点分布于嘉兴市 5 个县市区,采样点的布设根据各地种植面积和分布来设计样点位置和数量,GPS 被用来进行样点定位,在 0~20 cm 深度进行土壤混合采样.共采集了 204 个土壤样品(图 1).



图 1 研究区土壤采样点分布

Fig.1 Spatial distribution of the soil sampling points

土样风干后过 2 mm 孔隙的尼龙筛. As 和 Cu 用王水水浴加热消解, As 采用氢化物-原子荧光法测定; Cu 采用原子吸收分光光度法测定; 其它重金属用通常所说的“四酸消解”, 即王水再加 HClO_4 、 HNO_3 、 HF 消解; Cd、Pb 均用等离子体发射光谱仪 (ICP) 测定; Cr 用等离子体发射光谱仪测定. 在重金属的分析测定中, 均用 GSS-1 标准参考土壤进行全程质量控制.

2 研究方法

2.1 序贯高斯模拟和序贯指示模拟

序贯高斯模拟方法主要根据现有数据计算待模拟点值的条件概率分布, 从该分布中随机取一值作为模拟现实. 每得出 1 个模拟值, 就把它连同原始数据、此前得到的模拟数据一起作为条件数据, 进入下一点的模拟. 该方法比较简单、灵活和计算效率高, 所以是条件高斯随机模拟中最常用的方法之一^[8].

序贯指示模拟方法的主要算法同序贯高斯模拟方法类似. 其主要区别是在模拟前根据一定的阈值对原始数据进行重新赋值. 同时, 序贯指示模拟方法对原始数据的分布没有严格要求, 而采用序贯高斯模拟方法进行空间不确定评价, 用于评价的数据要尊重多元高斯分布假设. 但很多研究表明, 土壤重金属含量的数据分布特征往往具有很大的变异系数和偏度值, 不满足多元高斯分布, 所以序贯高斯模拟方法会受到很多限制.

这里指示变量是二元变量, 仅取 0 或 1 两值来表示存在或不存在. 如设 z_c 为阈值, 这样原始数据 z 被划分为 2 部分, 一部分为 $z(x) \leq z_c$, 另一部分为 $z(x) > z_c$, 并分别赋值为 1 和 0. 这种转换特别适用于环境污染风险评价, 通过设定合理的阈值 z_c , 就可以将 1 个连续性的随机变量 $z(x)$ 转化为 1 个指示函数. 对这个指示函数而言, 1 表示没有受到污染可以被接受, 0 表示受到污染不能被接受^[16].

2.2 不确定评价

(1) 点位不确定评价 主要指每个模拟点位其本身超过阈值的概率评价. 其计算公式如下:

$$P_{\text{SIS}}[z(x') > z_c] = \frac{n(x')}{N_{\text{SIS}}} \quad (1)$$

式中, $z(x')$ 为模拟点位 x' 处的值; z_c 为评价阈值; N_{SIS} 为序贯指示模拟的全部次数; $n(x')$ 为全部 N_{SIS} 次模拟中, 该点位模拟结果超过阈值的次数. 该评价的方差可以表示为:

$$s_p^2(x') = \frac{p(1-p)}{N_{\text{sis}}} \tag{2}$$

式中 p 就是公式(1)计算的概率值 $P_{\text{sis}}[z(x') > z_c]$.

(2) 空间不确定或称多点集合不确定评价 是指对评价区内所有待模拟点进行不确定性联合概率评价.其计算公式如下:

$$\begin{aligned} &P_{\text{sis}}[z(x'_1) > z_c, z(x'_2) > z_c, \cdots, z(x'_m) > z_c] \\ &= \frac{n(x'_1, x'_2, \cdots, x'_m)}{N_{\text{sis}}} \end{aligned} \tag{3}$$

式中 $n(x'_1, x'_2, \cdots, x'_m)$ 为所有 m 个模拟点位其模拟结果值超过阈值的模拟次数.这个评价就是求所有待评价点在某次模拟中同时超过阈值的概率.而上式概率的方差值为:

$$s_{p_j}^2 = \frac{p_j(1-p_j)}{N_{\text{sis}}} \tag{4}$$

式中 p_j 为公式(3)计算的概率值.

本研究采用 SPSS12.0 和 Gslib 软件^[8] 分别进行常规统计和地统计的随机模拟分析.

3 结果与讨论

3.1 土壤重金属分布特征

按照国家 GB15618-1995 土壤环境质量的

二类土壤执行标准,即能保证农业生产和维护人体健康为标准.根据研究区土壤调查,该地区土壤 pH 值一般都小于 6.5,只有个别靠近钱塘江边的土样 pH 值超过 6.5,所以选择国家标准的第二级 pH<6.5 的阈值为标准进行评价.样区土壤中各主要重金属 Cd、Cu、As、Cr 和 Pb 平均值都未超过国家标准的限值(表 1).另外,所有土样中只有 Cd 和 Cu 有部分样超过规定限值,分别有 24 个和 5 个,占总样品的 11.8% 和 2.5%.因此,该地区总体土壤环境质量是好的.

下面单独以土壤 Cu 为例进行空间变异和不确定性评价.图 2 为研究区土壤 Cu 含量的频率分布图,其数据分布接近高斯正态分布.4 个百分位数(20%、40%、60%、80%)分别为 28.1、30.5、33.9 和 36.97 mg·kg⁻¹,这 4 个分位数将作为随后的序贯指示模拟方法的 4 个截值.另外,因为各级土壤 Cu 的背景值在 20 mg·kg⁻¹ 左右,如世界、中国、浙江省和杭嘉湖地区的背景值分别为 20、23.1、21、23.8 mg·kg⁻¹^[17].因此,这里取国家一类标准 35 mg·kg⁻¹ 为评价阈值进行土壤 Cu 污染潜力分析.这样,在全部 204 个样品中有 59 个超过该阈值,占总数的 29.9%.

3.2 土壤重金属 Cu 的空间变异特性

表 1 主要土壤重金属含量统计($n=204$)/mg·kg⁻¹

Table 1 Conventional statistics of some heavy metal in soil($n=204$)/mg·kg⁻¹

元素	最小值	最大值	平均值	国家标准 ¹⁾	标准差	偏度	峰度	K-S(Z)
Cd	0.08	1.26	0.22	0.3	0.12	4.5	33.06	2.97(0)
Cu	16.80	61.72	32.8	50	6.7	1.1	3.04	1.13(0.16)
As	4.07	12.47	8.66	30	1.57	-0.32	-0.49	0.66(0.78)
Cr	57.12	125.59	87.77	250	11.68	0.03	0.15	0.38(1.0)
Pb	17.6	49.41	33.89	250	4.86	-0.04	0.51	0.5(0.96)

1) 国家 GB15618-1995 土壤环境质量的二类土壤执行标准

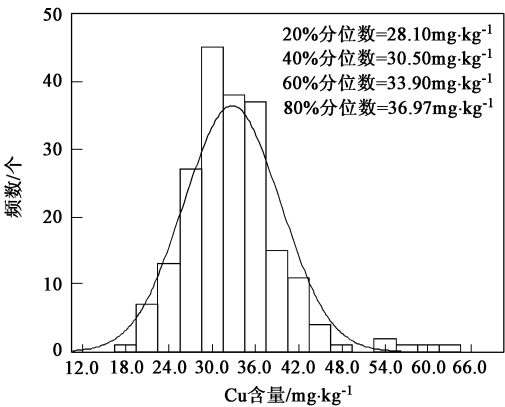


图 2 土壤 Cu 含量频率分布
Fig.2 Frequency distribution of soil Cu

首先将原始土壤 Cu 数据按照 4 个分位数进行指示码转化,并进行方差函数分析.其 4 个方差函数的具体参数见表 2 所示.可以看到 4 个半方差函数有很大的差异,块金值与基台值的比率变化很大.特别是块金值与基台值的比值,是用于表示空间相关性程度^[18],其变化从具有强烈的空间相关性(4.8%)到中等程度的空间自相关(42.9%),到空间相关性很弱(70.6%和 78.6%).所以,选择这 4 个分位数作为截值能够较好反映土壤 Cu 的整体空间变异特性.同时,这 4 个分位数的半方差函数将被用来构建序贯指示模拟方法 4 个截值的先验条件累积分布函数.另外,根据各向异性分析,在随机模拟中增加了主变程旋转角和变程长短比值参数.

表 2 土壤 Cu 含量不同分位数指示半方差拟和模型参数

Table 2 Indicator semivariograms of soil Cu and their fitted models for different quantiles

截值	拟和模型	块金值	基台值	块金值/基台值	变程/m	R ²
20%分位数	球状模型	0.060 4	0.140 8	42.9%	7 890	0.786
40%分位数	指数模型	0.010 9	0.225 8	4.8%	4 740	0.74
60%分位数	线性有基台	0.163	0.230 8	70.6%	14 756	0.802
80%分位数	线性有基台	0.133 4	0.169 8	78.6%	19 439	0.434

由于土壤 Cu 数据分布接近正态分布,所以利用普通克里格方法和序贯高斯模拟方法分别进行估值和模拟来生成其全区土壤 Cu 含量的空间分布图.从图 3 可见,土壤 Cu 的高含量地区主要分布在研究区北部,其含量接近或超过国家标准.另外从图 3(c)和 3(d)可以比较普通克里格估值方法和序贯高斯模拟方法得到数据的分布特征,明显看到普通克里格估值结果数据分布频率更趋于平缓,数据在 33 mg•kg⁻¹值分布趋于对称,数据整体接近正态分布.克里格估值的空间分布图[图 3(a)]整体比较连续,具有明显的平滑效应.而序贯高斯模拟结果在尊重原始实测数据值情况下,整体分布相对离散,突出

了原始数据分布的波动性,其模拟结果的数据分布频率相对集中.

3.3 土壤重金属 Cu 的空间不确定性分析

以评价标准35 mg/kg作为土壤 Cu 含量的阈值进行序贯指示模拟计算.对于各评价点超过该阈值的概率利用公式(1)来计算,这里模拟次数采用了500次.这样土壤 Cu 含量较高的区域,其每次模拟结果值高于该阈值的概率自然相对较高,相反含量低的区域其超过阈值的概率也低.从序贯指示模拟方法运算的结果来看[图 4(a)],土壤 Cu 含量较高的北部地区,超过阈值的概率超过 80%,而南部很多相对含量低的地区,超过阈值的概率值都低于

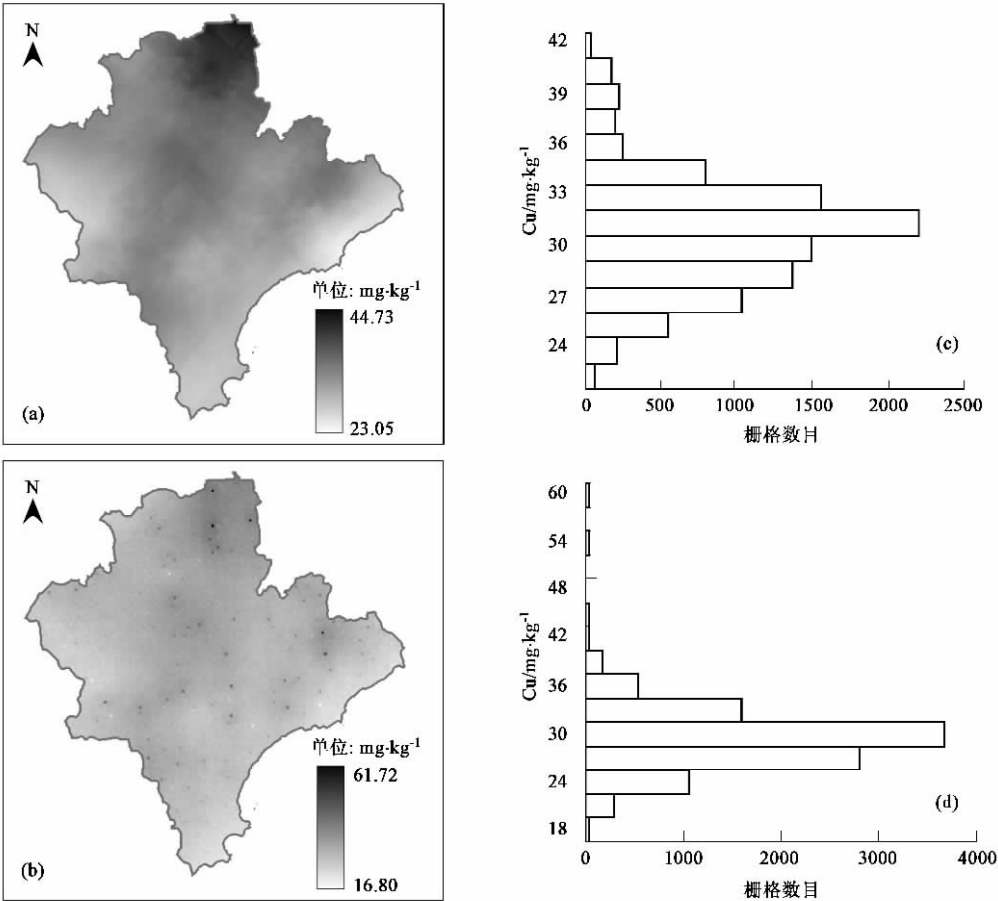


图 3 土壤 Cu 含量普通克里格插值(a)和序贯高斯模拟(b),以及空间分布统计频率(c,d)

Fig.3 Results of ordinary kriging interpolation and SGS for soil Cu and their spatial distribution frequency

10% . 因此,在高概率和低概率 2 类地区可以很确定给出其污染评价结果.但是对于那些概率值接近 50% 的区域,其评价结论就有很大的不确定性.所以这样的概率分布给各级环境管理或治理部门的决策提供更多的信息,而不是直接按某个阈值将整个区

域机械的划分为污染或未污染区域.

图 4(b)为每个模拟点超阈值概率的标准差.这里最大的标准差 0.022 出现在超阈值概率为 0.5 地方,两者的比值,也即相对最大误差值为 4.4%.这也表明经 500 次序贯指示模拟的概率分布结果是可靠的.

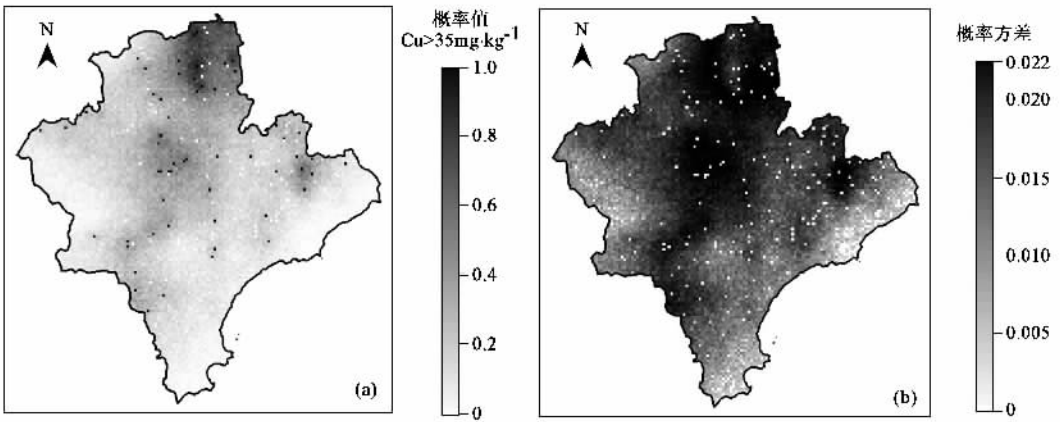


图 4 土壤 Cu 含量超过 $35\text{ mg}\cdot\text{kg}^{-1}$ 阈值概率的序贯指示模拟结果(a)和概率方差图(b)

Fig.4 Results of SIS of soil Cu exceeding $35\text{ mg}\cdot\text{kg}^{-1}$ threshold (a) and its prediction variance map (b)

空间不确定评价(多点集合不确定评价)按照公式(4)进行计算,就是在高污染风险区进行超阈值的联合概率评价.首先按照各模拟点位的超阈值概率来选择高风险区,这里分别选择超阈值概率为 0.85

和 0.75 2 个值来选取高风险区(图 5),分别选中了 34 个和 172 个栅格.由于序贯指示模拟对已知采样点不进行模拟,直接采用观测值,所以在不确定性评价时,其概率会直接取 0 或 1.本研究在所选的栅格

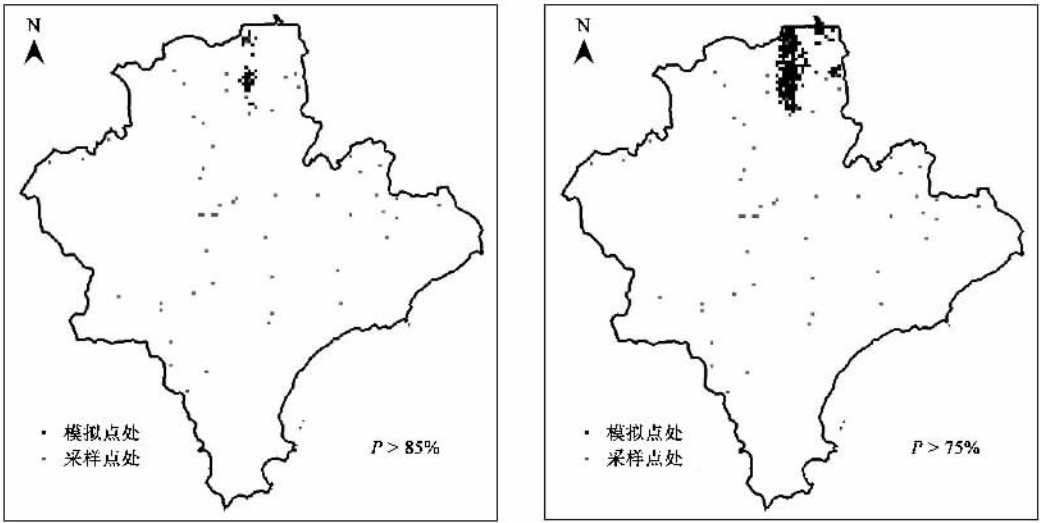


图 5 土壤 Cu 含量超过 $35\text{ mg}\cdot\text{kg}^{-1}$ 阈值概率按 0.85 和 0.75 截值的分布

Fig.5 Spatial distribution of soil Cu exceeding $35\text{ mg}\cdot\text{kg}^{-1}$ threshold as cutoff values were 0.85 and 0.75

中不包括这些样点.

然后在这些区域进行超阈值的联合评价,也就是统计在 500 次模拟中有多少次的模拟结果在这些区域内的所有栅格都超过阈值.很明显,这里的联合

概率比前面单点统计的概率更为严格,完全可以用它来衡量污染区划分的可信度.如果某区域单点位序贯高斯模拟结果的概率很高,同时其联合概率也高,那么将其划分为污染区的可信度是非常高的.从

表 3 可见,点位概率为 0.85 和 0.75 的区域其联合概率分别为 0.571 和 0.082. 可以看到尽管单点评价概率达到 0.75,但是采用更为严格的联合概率只有 0.082,所以如果将这个评价区整体划为污染区,其可信度还是不高,需要进一步调查确认. 所以,在划分较大范围污染区域时,最好同时采用联合概率来进行信度评价.

表 3 土壤 Cu 含量超过 35 mg·kg⁻¹ 阈值概率
按 0.85 和 0.75 划分区域的联合概率统计

Table 3 Joint probability statistics of soil Cu exceeding 35 mg·kg⁻¹
threshold as cutoff values were 0.85 and 0.75

概率值	栅格数目	联合概率	联合概率标准差
> 0.85	34	0.571	0.022
> 0.75	172	0.082	0.012

4 讨论

克里格估值方法在半方差函数和邻近收索规则不变的前提下,只能是 1 个结果. 而随机模拟的最大优点就是能在符合数据整体分布和方差函数前提下,产生多个模拟结果. 可以用来研究空间不确定问题,为评价不确定性提供了 1 个本性方法,用概率或其它相关量值来表征不确定性. 空间模拟和空间不确定性研究可以进一步服务于辅助决策规划,如划分污染区、作物土宜评价等^[19]. 特别是在环境研究领域,如大面积的非点源污染问题,是很难获得足够的实测点来客观表征实际的空间分布. 在这种复杂性和不确定性并举的问题面前,条件模拟的方法显然优于克里格估值法. 它提供的不确定性研究结果能够反映出这种环境风险性,更有助于决策者从“不确定性”出发做出合理的决策.

同样可以将随机模拟结果作为蒙特卡罗模拟或其它模型的输入数据. 通过随机模拟技术得出的各种模拟结果,可以进一步定量评价对环境所造成的各种可能影响. 而克立格估值仅给出 1 个单一的数值结果,很显然不能评价不确定性^[11].

随机模拟结果还可以被用来检测各种资源环境应用模型的稳固性,及其模型的误差传播等问题. 如将土壤随机模拟的空间分布结果输入到复杂的模型,如作物生长模型、污染物的流动或运移模型等. 这样土壤特性的不确定性将通过这些模型进行传播,而影响到这些模型的最后输出结果,如作物产量、污染物扩散浓度等,从而可以评价这些模型的稳固性.

5 结论

克里格法具有明显的“平滑”效应,减少了数据

的变异性,而条件随机模拟结果突出了原始数据分布的波动性,较适宜研究那些对环境具有不良影响的风险性变量,同时提供的不确定性研究结果有助于决策者从“不确定性”出发做出合理的决策并进行风险评价.

参考文献:

[1] Juang K W, Lee D Y. A comparison of three kriging methods using auxiliary variables in heavy-metal contaminated soils [J]. J. Environ. Qual., 1998, **27**: 355 ~ 363.

[2] Van Meirvenne M, Goovaerts P. Evaluating the probability of exceeding a site-specific soil cadmium contamination threshold [J]. Geoderma, 2001, **102**: 75 ~ 100.

[3] Von Steiger B, Webster R, Schulin R, *et al.* Mapping heavy metals in polluted soil by disjunctive kriging [J]. Environ. Pollut., 1996, **94**(2): 205 ~ 215.

[4] 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.

[5] 胡克林, 张凤荣, 吕盼忠, 等. 北京市大兴区土壤重金属含量的空间分布特征[J]. 环境科学学报, 2004, **24**(3): 463 ~ 468.

[6] Goovaerts P. Geostatistics for Natural Resources Evaluation [M]. New York: Oxford University Press. 1997.

[7] Wang G, Gertner G, Parysow P, *et al.* Spatial prediction and uncertainty analysis of topographic factors for the revised soil loss equation (RUSLE) [J]. J. Soil Water Conserv., 2000, **55**: 374 ~ 384.

[8] Deutsch C V, Journel A G. GSLIB: Geostatistical Software Library and User's Guide [M]. New York: Oxford University Press. 1998.

[9] Journel A G. Geostatistics for conditional simulation of ore bodies [J]. Econ. Geol., 1974, **69**(5): 673 ~ 687.

[10] Castrignano A, Buttafuoco G. Geostatistical stochastic simulation of soil water content in a forested area of south Italy [J]. Biosyst. Eng., 2004, **87**(2): 257 ~ 266.

[11] 李保国, 胡克林, 陈德立, 等. 农田土壤表层饱和导水率的条件模拟[J]. 水利学报, 2002, (2): 36 ~ 40, 46.

[12] 陈亚新, 史海滨, 田圃德, 等. 水盐空间变异性监测的条件模拟[J]. 水利学报, 2000, (6): 67 ~ 73.

[13] Lin Y P, Chang T K. Simulated annealing and kriging method for identifying the spatial patterns and variability of soil heavy metal [J]. J. Environ. Sci. Health, Part A, Toxic/Hazard. Subst. Environ. Eng., 2000, **35**(7): 1089 ~ 1115.

[14] Lin Y P, Chang T K, Teng T P. Characterization of soil lead by comparing sequential Gaussian simulation, simulated annealing simulation and kriging methods [J]. Environ. Geol., 2001, **41**(1-2): 189 ~ 199.

[15] Juang K W, Chen Y S, Lee D Y. Using sequential indicator simulation to assess the uncertainty of delineating heavy-metal contaminated soils [J]. Environ. Pollut., 2004, **127**(2): 229 ~ 238.

[16] Webster R, Oliver M A. Geostatistics for environmental scientists [M]. UK: John Wiley & Sons. 2001.

[17] 中国环境监测总站主编. 中国土壤元素背景值 [M]. 北京: 中国环境科学出版社, 1990.

[18] Chien Y J, Lee D Y, Guo H Y, *et al.* Geostatistical analysis of soil properties of mid-west Taiwan soils [J]. Soil Sci., 1997, **162**: 291 ~ 297.

[19] Goovaerts P. Geostatistical modelling of uncertainty in soil science [J]. Geoderma, 2001, **103**(1-2): 3 ~ 26.