

普通克里格评估土壤污染物空间分布精度控制研究

高 远¹, 赵玉杰², 杨海松³

(1.中国地质大学研究生院, 湖北 武汉 430074; 2.农业部环境保护科研监测所, 天津 300191; 3.泰达建设集团奥亚工程管理公司, 天津 300456)

摘 要: 从数据的统计特征分析、异常值筛选及处理、实验半变异函数计算及拟合等方面对普通克里格计算过程中的精度控制问题进行了初步研究, 并以天津市东南郊区土壤砷(As)空间分布评估为例, 分析了在多步骤精度控制下普通克里格法估计土壤污染物空间分布的可靠性。结果表明, 研究区土壤 As 表现出了明显的累积特征, 且监测值为正态分布, 异常值较少且为局部异常值, 实验半变异函数表现出明显的各向异性。在较严格的精度控制下, 普通克里格估值的最优无偏特性可以得到较好的体现, 其对土壤 As 的空间分布的估计精度可以达到较高的水平。

关键词: 普通克里格; 土壤; 砷; 精度控制

中图分类号: X53 文献标志码: A 文章编号: 1672-2043(2009)05-0871-06

Precision Control of Assessing Soil Contamination Spatial Distribution Based on Ordinary Kriging

GAO Yuan¹, ZHAO Yu-jie², YANG Hai-song³

(1. Graduate School of China University of Geosciences, Wuhan 430074, China; 2. Institute of Agro-Environmental Protection, Ministry of Agriculture, Tianjin 300191, China; 3. TCG AOYA Engineering Management Co., Tianjin 300456, China)

Abstract: In this paper, the statistical characteristics analysis, outliers screening and treatment, experimental semi-variogram calculation and fitting were studied preliminarily as the main calculation process to improve the precision of ordinary kriging interpolation. A case of assessing the reliability of using ordinary kriging under multi-step precision controls to evaluate the soil arsenic (As) pollution in Tianjin southeast suburb was studied. The results showed that the soil As in the study area was marked accumulation, and the monitoring values were normal distribution. There were few outliers in the monitoring values and they were all local anomaly values. The experimental semivariogram showed anisotropy. The optimal and unbiased characteristics of ordinary kriging had been brought into full play under the precision control and the estimation of spatial distribution of soil As could reach a higher level.

Keywords: ordinary kriging; soil; arsenic; precision control

土壤污染物砷(As)是一种类金属元素, 其化合物有三价和五价之分, 在淹水条件下, 五价 As 化合物易还原成三价 As 化合物, 后者的毒性比前者大, 且易为植物吸收进入食物链, 危害人体健康, 而在碱性土壤条件下, As 的活性也较酸性土壤强^[1]。因此土壤 As 污染已成为各国科研人员研究的热点和难点^[2-3]。As 在农田的分布是随机性及空间结构性的统一体, 正确分析 As 污染的空间分布特征, 是对农田安全性进行评价的前提。而地质统计学是研究土壤污染物空间分布随机性及结构性, 描述其分布特征的最重要的工具。近年来, 应用地质统计学描述土壤污染物分布特征的

研究越来越多, 而普通克里格法是这一技术中应用最多的方法, 且是美国材料与实验协会(ASTM)推荐的基础方法^[4]。如 Sven 等采用普通克里格法分析了德国一金属冶炼区周边农田超过当地环境质量标准的不安全区的分布情况^[5], 赵玉杰等应用此法研究了山东淄博玉米产区土壤 As 的分布情况^[6], 其他相似的研究如 Tavares 等对 Domingos 矿区的重金属污染的评价^[7], Liu 等对于浙江省农田的安全性评价等^[8]。普通克里格法虽然应用广泛, 但其计算步骤较多, 而在每步计算过程中如不采取严格的精度控制, 常会导致评估结果精度降低甚至产生错误的结果, 而有关普通克里格计算过程中精度控制的研究国内外还少见报道。本文从普通克里格应用过程中的数据统计分析、异常值筛选及处理、实验半变异函数计算及拟合等方面对提高普

收稿日期: 2008-08-05

作者简介: 高远(1964—), 女, 高级工程师, 博士研究生, 主要从事环境科学研究。E-mail: Gaoyuan@msa.gov.cn

通克里格法的评估精度问题进行了研究,并以天津市东南郊区土壤 As 含量的空间分布特征分析为研究案例,以期为提高普通克里格估计精度提供一定的参考。

1 材料与方法

1.1 研究区概况

研究区位于天津东南郊,西临天津市,东面是滨海新区,处于东经 $117^{\circ}14' \sim 117^{\circ}36'$,北纬 $38^{\circ}50' \sim 39^{\circ}5'$,属暖温带半湿润大陆性季风气候,年平均气温 12.1°C ,年平均降水量 554.1 mm 。区域内土壤类型为盐化潮土,土壤 pH 在 $8.09 \sim 8.35$,有机质含量为 $1.02\% \sim 1.58\%$ 。土地利用类型以设施农业用地为主,是天津市重要的优势农产品水稻、玉米、西芹、青萝卜生产基地。然而本区域毗邻城市,区域内排污河较多,污灌及大气降尘给本区域农田带来了严重的安全隐患^[9-11]。因此,采用地质统计学方法分析该区土壤 As 的分布情况,对保障本区域农产品安全至关重要。

1.2 样品采集及分析

2005 年 7 月,在水稻成熟前后采用格网布点法平均约 100 hm^2 采集一个土样,每个样品由 5 个子样品组成,子样间距根据样地大小略有不同,在 $3 \sim 6\text{ m}$ 。采取“X”法、“S”法、棋盘法中的任何一种布点方法,用木铲、竹铲均匀随机采集土壤耕层 $0 \sim 20\text{ cm}$ 样品 5 个,充分混合后以四分法留取 1 个 1.5 kg 的送检样,样点采用 GPS 定位,并记录经纬度坐标。采样点位分布如图 1 所示。在本区域共采集了 163 个土壤复合样品。

将采集的土壤样品自然风干后除去土样中石子和动植物残体,混匀后用四分法缩分至约 100 g ,用木棒碾压,通过 2 mm 尼龙筛,混匀,用玛瑙球磨仪将通过 2 mm 尼龙筛的土样研磨至全部通过孔径 0.149 mm 尼龙筛,混匀后备用。采用硼氢化钾-硼酸银分光光度法测定土壤总 As 含量^[10]。

1.3 数据分析

采用普通克里格法分析污染物的空间分布情况,首先需对数据进行统计分析,了解数据的集中分散性分布特征,而后进行空间结构性分析,主要是筛选异常值及计算实验半变异函数,再通过拟合实验半变异函数获得半变异函数。拟合效果不确定性分析可采用交叉验证法,评价标准为 (1)平均误差(ME)最好接近于 0,从而保证了评估结果的无偏性 (2)标准克里格方差(MSSE)代表了实际方差与理论方差的比例,其值应尽可能的接近于 1, Liu 等人认为, MSSE 值介

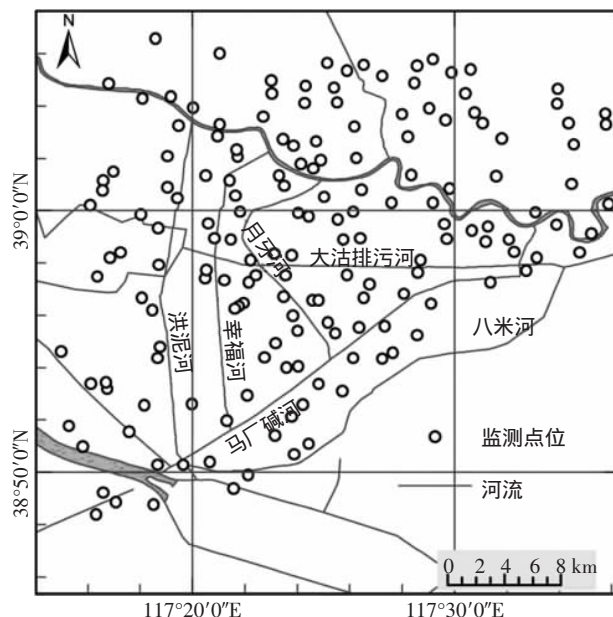


图 1 天津东南近郊土壤环境质量监测样点分布图

Figure 1 Sampling distribution sketch in south-east suburb, Tianjin

于 $1 \pm 3\sqrt{2/N}$ (N 为建模采用的数据量个数)时,符合评估精度的要求^[11]。本研究中 MSSE 值应界于 $0.67 \sim 1.33$ 。半变异函数确定后可根据普通克里格算法估计污染物的空间分布情况。具体方法可参见相关文献^[12]。

原始数据的处理及基本数理统计分析采用 SPSS (Version 16.0) 软件,实验半变异函数计算、拟合、普通克里格插值计算等采用 Geostatistics 公司的 ISATIS (Version 6.0) 完成,图件制作采用 ESRI 公司的 ArcGIS (Version 9.2) 完成。

2 结果与讨论

2.1 土壤 As 监测数据描述性统计特征

监测数据的描述性统计是对监测值的集中、分散性及分布状况的总体了解,是数据能否满足普通克里格插值应用条件的一般性认识。研究区土壤 As 监测值一般性统计特征见表 1。

表 1 原始数据统计结果表明,本监测区土壤 As 最大值达 $20\text{ mg}\cdot\text{kg}^{-1}$,达到《土壤环境质量标准》(GB15618—1995)二级标准有关土壤 As 最高限量 $20\text{ mg}\cdot\text{kg}^{-1}$ 的规定,原始监测数据的平均含量为 $(11.43 \pm 2.66)\text{ mg}\cdot\text{kg}^{-1}$,较本区域土壤 As 背景值 $9.06\text{ mg}\cdot\text{kg}^{-1}$ 高出 $2.37\text{ mg}\cdot\text{kg}^{-1}$,表明本区域土壤 As 已存在明显的累积^[13]。监测值的峰度偏度值表明,数据呈正偏尖峰分布,但幅度较小。K-S 值表明,虽然数据的峰度偏度值较小,但依然不符合正态分布的要求。

表 1 土壤 As 监测数据描述性统计结果

Table 1 Descriptive statistical results of soil As concentrations

项目	样点数	均值/ $\text{mg} \cdot \text{kg}^{-1}$	极小值/ $\text{mg} \cdot \text{kg}^{-1}$	极大值/ $\text{mg} \cdot \text{kg}^{-1}$	标准差/ $\text{mg} \cdot \text{kg}^{-1}$	峰度	偏度	K-S 值
原始数据	163	11.43	6.40	20.00	2.66	0.38	0.31	0.00
去异常值数据	161	11.40	6.40	17.20	2.26	0.42	0.04	0.09

2.2 监测数据异常值分析

监测数据含有异常值,易造成实验半变异函数失真,使半变异函数在短变程时翘尾、长变程时抬头或增加实验半变异函数的基台值^[16]。H-Scatter 是筛选监测数据异常值最有效的工具之一。H-Scatter 图展示了监测数据的空间相关性。根据地质统计学原理,位置越近的点监测值就越相似,数据的变异性就越小,数据点就越应靠拢于斜率为 1 的直线,形成延 1:1 直线分布的团聚点,而如果在某个矢量方向上,相邻点变异性较大,数据点就会偏离团聚区,形成相对的分散点,从而作为数据异常值判断的依据^[14]。本区域土壤 As 的 H-Scatter 图(图 2)中 A、B 相关的数据即为本研究区的异常值,与 A、B 相关的监测点见图 3 中的 A、B 两处。由于监测数据异常值并不一定是由于监测错误所致,因此一般不宜采用直接删除的方式,本文采用掩膜处理,即在计算实验半变异函数时暂不考虑这些值,而在最后插值时再将其考虑在内^[15]。

监测数据经异常值掩膜处理后,其统计特征如表 1 所示。K-S 检验结果表明数据为正态分布,满足普通克里格对数据分布的要求^[16],无需考虑数据偏态性引起的实验半变异函数的比例效应^[17]。

2.3 实验半变异函数计算及拟合

正确选择及计算实验半变异函数是提高研究区

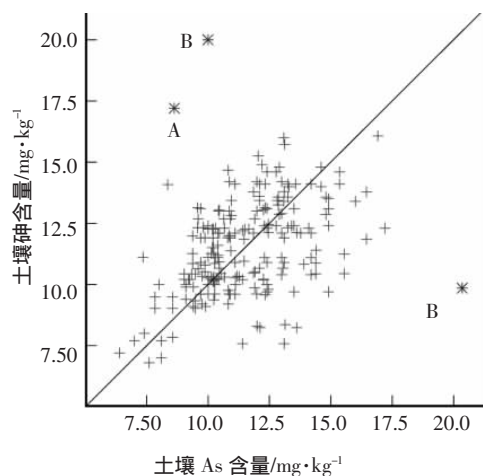


图 2 土壤 As 含量 H-Scatter 图

Figure 2 H-Scatter plot of soil As concentrations

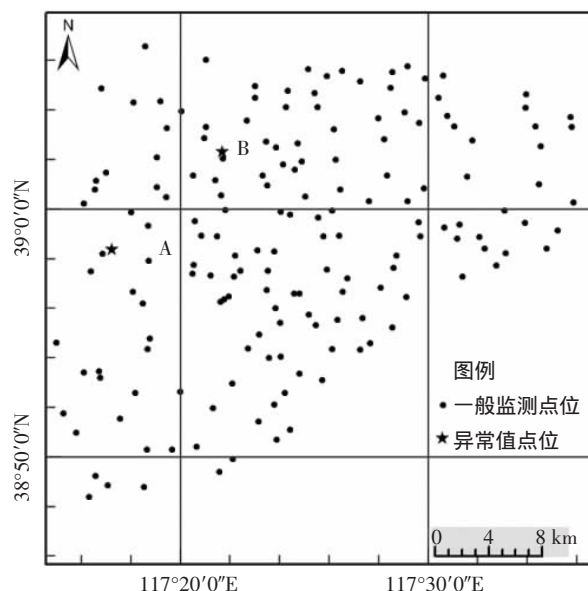


图 3 土壤 As 异常值分布示意图

Figure 3 Outliers distribution sketch of soil As concentrations

土壤 As 空间变异估计精度的重要一环。实验半变异函数有多种形式如 Matheron 提出的传统实验半变异函数,其他的有一阶变异函数、对偶相对变异函数、Genton 实验半变异函数等^[18-19],传统的 Matheron 实验半变异函数是应用最广泛的实验半变异函数,如没有特殊原因如数据存在明显的偏态分布,异常值较多,数据波动性强等,一般不使用其他形式的实验半变异函数^[20]。依据惯例本文采用 Matheron 实验半变异函数分析数据的空间结构性。实验半变异函数的计算非常复杂,ASTM 总结了实验半变异函数计算需把握的基本原则^[21],在计算实验半变异函数时,除把握以上原则外,还应考虑计算获得的实验半变异函数的实用性,即保障计算得到的实验半变异函数在充分反映样点小承载的同时,尽量做到兼顾全局,不强调局部特例,避免不同步长分组实验半变异函数值的大起大落,从而易于采用人工方法或计算机半自动化方法进行拟合。当实验半变异函数的基台值位于滞后距离的 2/3 处时,计算的实验半变异函数最稳健^[22]。同时由于区域化变量常表现为各向异性,因此在计算各向同性实验半变异函数的同时,也应计算各向异性实验半变

异函数。

图 4、图 5(a、b)分别为研究区各向同性及各向异性实验半变异函数及拟合模型图(步长=2 150 m,步长数=8,步长容限为步长值的 1/2,各向异性模型长轴方向为北偏东约 48°),实验半变异函数值在达到先验方法后并没有表现出继续增加趋势,表明区域化变量符合二阶平稳假设条件,从而不用考虑数据的全局趋势特征^[18]。图 4 中未掩膜实验半变异函数表现出明显的翘尾情况,严重歪曲了研究区土壤 As 的空间结构性。分别采用球形模型与球形模型套合结构拟合实验半变异函数,拟合方法为人工干预下的基台值自动拟合法^[20],采用交叉验证法评价拟合效果(表 2)。图 4 中的双球形模型套合结构虽然能精确地拟合实验半变异函数拟合,但 ME 及 MSSE 值却较球形模型拟合结果高,表明模型存在过拟合现象^[23]。由于土壤 As 存在各向异性,而且土壤 As 各向异性拟合模型的 MSSE 值最小,因此本研究采用各向异性模型描述区域化变量的空间变异特性。模型参数为长轴变程=10 159 m,短轴变程=6 843 m,拱高=1.83,块金值=2.88,步长容限为步长值的 1/2。

在拟合的各向异性球形模型中,块金值占模型基台值的 61.15%,表明本区域土壤 As 含量空间结构性较弱而微域变异性明显。本区域土壤类型为盐化湿潮土,北部区域有部分湿潮土分布,土壤质地为粘壤^[24],区域内土壤类型及质地没有微域性的变化,且盐化潮土及湿潮土 As 含量差异性亦不显著^[25],As 含量的微域变异不是来源于土壤本底 As 含量的差异。土壤 As 除本底含量以外主要来源于污灌、大气降尘、农业生

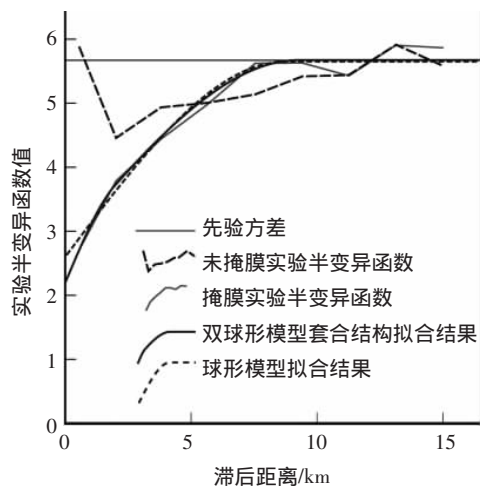


图 4 研究区土壤 As 各向同性实验半变异函数及拟合结果
Figure 4 Isotropic experimental semivariogram and fitting model of soil As concentrations

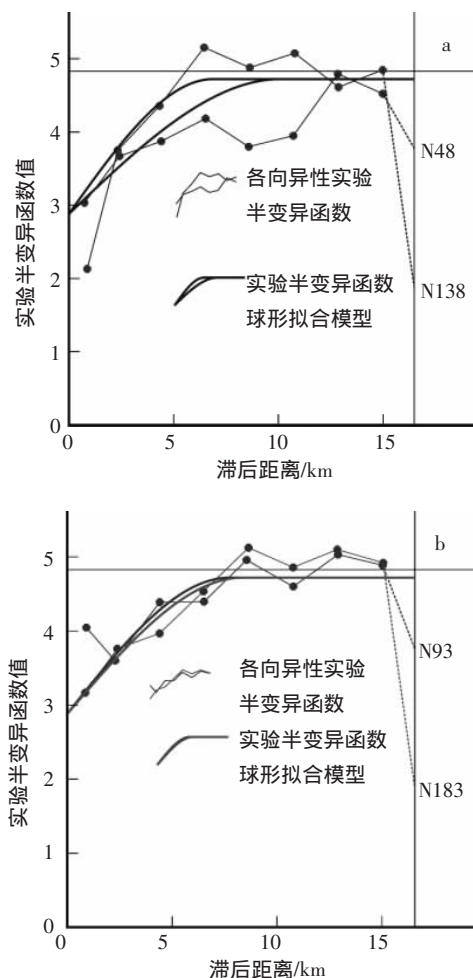


图 5 (a、b) 监测区土壤 As 各向异性实验半变异函数及拟合结果
Figure 5 (a、b) Anisotropic experimental semivariogram and fitting model of soil As concentrations

表 2 各向同性及各向异常模型交叉验证结果表

Table 2 Cross-validation results of isotropic and anisotropic models

项目	ME	MSSE
各向同性球形模型	-0.001 64	1.059
各向同性套合球形模型	-0.005 9	1.072
各向异性球形模型	-0.005 61	1.033

产过程中的含 As 农药、除草剂使用等,近年来本区域大气降尘中 As 含量有所增加^[26],但大气颗粒物具有远距离传输的能力,造成土壤 As 含量小面积局部增加的可能性较小。而污灌及农业生产过程中的含 As 农药、除草剂的使用造成本区域土壤 As 含量微域变异的可能性却无法排除。

3 As 元素分布情况估计

采用普通克里格法,以拟合的各向异性球形模型

为计算模型,对研究区土壤 As 含量空间分布情况进行了估计(图 6)。插值结果的统计特征(表 3)表明,插值结果的平均值与去异常值的土壤 As 平均值相同,进一步表明本插值结果是研究区土壤 As 分布的最优估计。插值结果的最小值较原始值提高了 $0.36 \text{ mg} \cdot \text{kg}^{-1}$,最大值降低了 $1.90 \text{ mg} \cdot \text{kg}^{-1}$ 。虽然较原始数据的统计特征有一定不同,但预测值与监测值的相关系数为 0.86,达极显著性相关水平。说明采用较严格的精度控制方法可以使普通克里格的最优无偏估计特性得到较好的实现。

研究区土壤 As 含量高值区主要分布在城市周边及区域内 2 条主要排污河(大沽排污河和月牙河)的交汇处以南部分区域,As 含量介于 $13 \sim 15 \text{ mg} \cdot \text{kg}^{-1}$,高于本区域土壤 As 背景值 $4 \sim 5 \text{ mg} \cdot \text{kg}^{-1}$ [27]。而土壤 As 含量低值区位于洪泥河中部以西区域,土壤 As 含量低于 $8 \text{ mg} \cdot \text{kg}^{-1}$ 。实际调查发现,土壤 As 含量偏高的区域一般为农田灌溉区,而洪泥河中部区域由于设施农田建设及采用水库水灌溉从而没有受到 As 污染。可

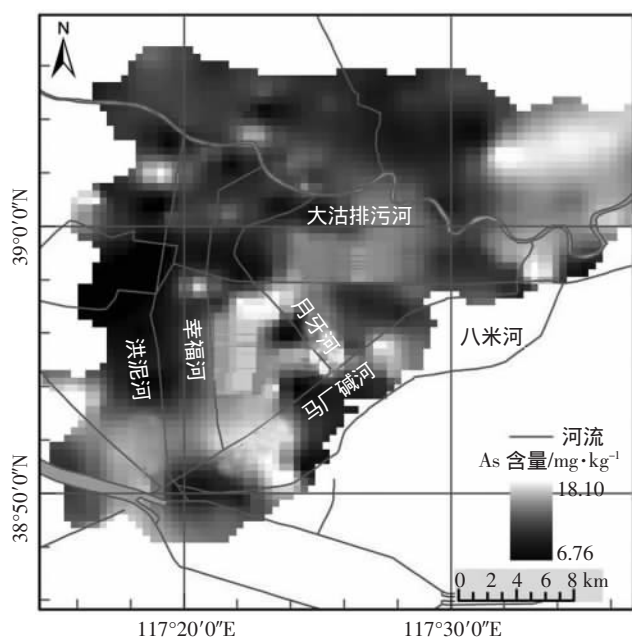


图 6 研究区土壤 As 含量空间分布图

Figure 6 Distribution sketch of soil As concentrations

表 3 土壤 As 含量空间插值结果统计

Table 3 Statistics results of spatial interpolation of soil As concentrations

预测点 数量	均值/ $\text{mg} \cdot \text{kg}^{-1}$	极小值/ $\text{mg} \cdot \text{kg}^{-1}$	极大值/ $\text{mg} \cdot \text{kg}^{-1}$	标准差/ $\text{mg} \cdot \text{kg}^{-1}$	峰度	偏度	相关 系数
163	11.40	6.76	18.10	1.64	0.16	0.10	0.86**

注:**表示在 0.01 水平(双侧)上显著相关。

见污水灌溉对本区部分农田环境质量已经产生了不利影响。

4 结论

(1)采用普通克里格方法研究了天津东南郊区土壤 As 含量的空间分布特征。结果表明,研究区土壤 As 平均含量比背景值有较明显的增加,虽然还未达到污染水平,但局部地区土壤 As 有明显的增加。异常值分析结果表明数据的异常值较少,数据经过掩膜处理后为正态分布。经掩膜处理后的监测值的实验半变异函数表现出明显的各向异性,采用人工干预下的基台值自动拟合法获得了实验半变异函数的拟合模型,并通过交叉验证后确定了模型参数,结果表明,各向异性球形模型效果最好。

(2)分别从监测数据的统计特征分析、异常值检验、实验半变异函数计算及拟合等方面分析了提高普通克里格方法估计土壤污染物空间分布精度的方法。结果表明,采用较严格的精度控制方法,能达到较满意的效果。

参考文献:

- [1] Smith E, Juhasz A L, Weber J, et al. Arsenic uptake and speciation in rice plants grown under greenhouse conditions with arsenic contaminated irrigation water[J]. *Science of the Total Environment*, 2008, 392(2-3): 277-283.
- [2] Lado L R, Polya D, Winkel L, et al. Modelling arsenic hazard in cambodia: a geostatistical approach using ancillary data[J]. *Applied Geochemistry*, 2008, 23(11): 3010-3018.
- [3] Lee J, Jang C, Liang C, et al. Assessing carcinogenic risks associated with ingesting arsenic in farmed smeltfish (*Ayu, Plecoglossus altivelis*) in arseniasis-endemic area of Taiwan[J]. *Science of the Total Environment*, 2008, 403(1-3): 68-79.
- [4] ASTM. D5923-96(2004) Standard guide for selection of kriging methods in geostatistical site investigations[S]. 2004.
- [5] Sven A C. Distribution of Cd in the vicinity of a metal smelter: interpolation of soil Cd concentrations with regard to regulative limits [J]. *Journal of Plant Nutrition and Soil Science*, 2002, 165(6): 697-705.
- [6] 赵玉杰, 师荣光, 白志鹏, 等. 山东淄博玉米产区土壤砷含量空间变异研究[J]. *环境科学*, 2006, 27(8): 1676-1682.
- ZHAO Yu-jie, SHI Rong-guang, BAI Zhi-peng, et al. Spatial variability analysis of soil arsenic in Zibo maize producing area, Shandong Province[J]. *Environmental Science*, 2006, 27(8): 1676-1682.
- [7] Tavares M T, Sousa A J, Abreu M M, et al. Ordinary kriging and indicator kriging in the cartography of trace elements contamination in domingos mining site (alentejo, portugal)[J]. *Journal of Geochemical Exploration*, 2008, 98(1-2): 43-56.
- [8] Liu X, Wu J, Xu J, et al. Characterizing the risk assessment of heavy met-

- als and sampling uncertainty analysis in paddy field by geostatistics and gis[J]. *Environmental Pollution*, 2006, 141(2) :257-264.
- [9] 吴光红, 苏睿先, 李万庆, 等. 大沽排污河污灌区土壤重金属富集特征和来源分析[J]. *环境科学*, 2008, 29(6) :1693-1698.
- WU Guang-hong, SU Rui-xian, LI Wan-qing, et al. Source and enrichment of heavy metals in sewage-irrigated area soil of Dagu sewage discharge channel[J]. *Environmental Science*, 2008, 29(6) :1693-1698.
- [10] GB/T 17135—1997, 土壤质量, 总砷的测定. 硼氢化钾-硝酸银分光光度法[S].
- GB/T 17135—1997, Soil quality—determination of total arsenic—spectrophotometric method with potassium borohydride and silver nitrate[S].
- [11] Liu C W, Jang C S, Liao C M, et al. Evaluation of arsenic contamination potential using indicator kriging in the yun-lin aquifer (Taiwan)[J]. *Science of the Total Environment*, 2004, 321(1-3) :173-188.
- [12] 张仁铎. 空间变异理论及应用[M]. 北京: 科学出版社, 2005.
- ZHANG Ren-duo. Spatial variability theory and its applications[M]. Beijing: Science Press, 2005.
- [13] 赵玉杰. 基于地统计的土壤重金属污染对人体危害风险评价方法研究[D]. 天津: 南开大学, 2007.
- ZHAO Yu-jie. Human health risk assessment of soil heavy metals contamination based on geostatistics[D]. Tianjin: Nankai University, 2007.
- [14] Jeffrey C Myers. Geostatistical error management: quantifying uncertainty for environmental sampling and mapping[M]. New York: John Wiley and Sons, 1997.
- [15] Geostatistics. Isatis case studies[R]. 2005 :362-369.
- [16] 王政权. 地统计学及在生态学中的应用[M]. 北京: 科学出版社, 1999.
- WANG Zheng-quan. Geostatistics and its ecological application[M]. Beijing: Science Press, 1999.
- [17] Roko R O, Daemen J J K, Myers D E. Variogram characterization of joint surface morphology and asperity deformation during shearing[J]. *International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts*, 1997, 34(1) :71-84.
- [18] 侯景儒. 实用地质统计学[M]. 北京: 地质出版社, 1998.
- HOU Jing-ru. Applied geostatistics[M]. Beijing: Geological Publishing House, 1998.
- [19] 张长波, 吴龙华, 骆永明, 等. 稳健变异函数在土壤污染物来源识别中的应用: 以某重金属污染场地为例 [J]. *环境科学*, 2008, 29(3) :804-808.
- ZHANG Chang-bo, WU Long-hua, LUO Yong-ming. Application of robust estimator in source identification of soil pollutants: a case study at a heavy metal contaminated site[J]. *Environmental Science*, 2008, 29(3) :804-808.
- [20] Geovariance. Isatis user handbook: structure identification in the intrinsic case[R]. 2005 :1-5.
- [21] ASTM. D5922-96 (2004) Standard guide for analysis of spatial variation in geostatistical site investigations[S]. 2004.
- [22] Kevin J, Jay M Ver H, et al. Using ArcGIS geostatistical analyst: ArcGIS 8[M]. New York: ESRI Press, 198-210.
- [23] Goovaerts P. Geostatistics for natural resource evaluation[M]. New York: Oxford Univ. Press, 1997.
- [24] 蒋德勤. 天津土种志[M]. 天津: 天津科技出版社, 1990.
- JIANG De-qin. Soil species records of Tianjin[M]. Tianjin: Tianjin Science and Technology Press, 1990.
- [25] 李玉浸, 高怀友. 中国主要农业土壤污染元素背景值图集[M]. 天津: 天津教育出版社, 2006.
- LI Yu-jin, GAO Huai-you. Background value Atlas of major agricultural soil pollution element[M]. Tianjin: Tianjin Education Press, 2006.
- [26] 姬亚芹, 朱 坦, 白志鹏, 等. 天津市土壤风沙尘元素的分布特征和来源研究[J]. *生态环境*, 2005, 14(4) :518-522.
- JI Ya-qin, ZHU Tan, BAI Zhi-peng, et al. Distribution characteristics and sources of the elements in soil dust[J]. *Ecology and Environment*, 2005, 14(4) :518-522.
- [27] 王祖伟, 张 辉. 天津污灌区土壤重金属污染环境质量与环境效应[J]. *生态环境*, 2005, 14(2) :211-213.
- WANG Zu-wei, ZHANG Hui. Environmental quality and biological effects of heavy metals in soils in the regions of sewage irrigation in Tianjin[J]. *Ecology and Environment*, 2005, 14(2) :211-213.