

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

张长波^{1,2}, 吴龙华¹, 骆永明^{1*}

(1. 中国科学院南京土壤研究所土壤环境与污染修复重点实验室 土壤与农业可持续发展国家重点实验室, 南京 210008;
2. 中国科学院研究生院, 北京 100049)

摘要:空间特异值可以提供污染源的位置信息. 以 Genton 稳健估计量为基础的变异函数为研究工具, 对某重金属冶炼场地土壤中 Cu、Zn、Pb、Cd、As、Se 和 Hg 含量进行了空间特异值分析, 并根据空间特异值的地理位置及土地利用方式等信息指示污染源. 结果表明, 7 种元素变异函数曲线的理论模型符合球状模型, 7 种元素都具有较小的块金方差, Cu、Zn 和 Pb 只为 0.000. 与 Matheron 估计量相比, 该研究得到的变异函数具有更小的块金效应, 说明 Genton 估计量能够消除或减弱特异值对变异函数的影响. 7 种元素的空间特异值多集中在个别小范围内, 以簇状形式存在, 与该区域冶炼高炉及尾渣堆放处的位置具有很好的相关性, 说明冶炼行为及其相关过程是该区重金属的重要来源.

关键词:稳健估计; Genton 估计量; 源识别; 土壤污染物

中图分类号: X53 文献标识码: A 文章编号: 0250-3301(2008)03-0804-05

Application of Robust Estimator in Source Identification of Soil Pollutants: A Case Study at a Heavy Metal Contaminated Site

ZHANG Chang-bo^{1,2}, WU Long-hua¹, LUO Yong-ming¹

(1. State Key Laboratory of Soil and Sustainable Agriculture, Key Laboratory of Soil Environment and Pollution Remediation, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China; 2. Graduate University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Spatial outliers can give us useful information on locations of pollution sources. A Genton's estimator based variogram was applied to analyse the concentration of Cu, Zn, Pb, Cd, As, Se and Hg in soils at a contaminated site. Spatial outliers were identified to imply the location of the pollution sources of heavy metals. The results showed that the semivariograms of 7 elements were best described by spherical model, and the nugget variances were very small, especially for Cu, Zn and Pb, which were 0.000. Compared with Matheron's estimator, Genton's estimator can damp the effect of outliers, and so get a smaller nugget effect. The spatial outliers existed in a cluster pattern, whose locations coincided with those of copper smelters in the area. This indicated that the copper smelters were the important sources of soil heavy metals.

Key words: robust estimator; Genton's estimator; source identification; soil pollutants

在人为活动干扰较为严重的区域, 土壤污染物含量可以看作 2 个过程共同作用的结果: 一是连续的背景过程 (continuous background process), 主要受到母质和扩散源污染 (diffuse pollution) 的影响; 另一是准点过程 (quasi point process), 主要受到点源污染 (point source pollution) 的影响^[1,2]. 主要受准点过程影响的土壤样品, 其重金属含量相对于整个研究区域而言往往表现为特异值或异常值 (outlier). 这也为研究者提供了一种研究的思路, 识别研究区域的特异值可以为点状污染源和其污染过程的认识提供有用的信息.

在目前的研究中, 特异值的识别有多种方法. 应用较广的是传统的探索性数据分析, 比较有代表性的是域值法, 即样本平均值加减 3 倍标准差, 在此区

间以外的数据均定为特异值^[3~5]. Dowd^[6]指出, 空间数据的特异值不仅仅指超过某个阈值的数据, 而是指与其周围的数据相比非常大或非常小的数据. Laslett 等^[7]称之为空间特异值 (spatial outliers). 该定义与非空间统计中简单的数据分析相比具有明显的优势, 后者只是鉴别出了对于整体数据分布来说不正常的的数据而忽略了其空间信息. 由此可以看出, 空间特异值的识别对理解污染源更加具有实际意义.

地统计学是空间信息分析的重要工具. 地统计学是以区域化变量理论为基础, 以变异函数为基本

收稿日期: 2007-04-13; 修订日期: 2007-06-10
基金项目: 国家重点基础研究发展规划 (973) 项目 (2002CB410810)
作者简介: 张长波 (1979 ~), 男, 硕士研究生, 主要研究方向为土壤污染物的来源解析, E-mail: cbzhang@issas.ac.cn
* 通讯联系人, E-mail: ymluo@issas.ac.cn

工具的方法, 广泛应用于空间数据的分析^[8]. 在空间估计过程中, 空间特异值点位的预测值与其真实值相差较大. 可以利用这一性质识别出空间特异值. 然而传统实验变异函数[Matheron 估计量 $\gamma_M(h)$]基于偏差的平方[见公式(1)], 对特异值非常敏感, 不具有稳健性, 因此当结果中出现空间特异值时, 可能会对变异函数的估计产生巨大的影响, 继而对空间特异值周围点位的估计产生较大的影响, 发生明显的“平滑效应”, 使得周围点位的预测值与实际值显著偏离^[8], 给上述方法鉴别空间特异值带来了困难.

$$2\gamma_M(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \quad (1)$$

把稳健统计和地统计学融为一体的稳健地统计学可以解决上述难题. 稳健变异函数可以减弱空间特异值对变异函数的影响, 增加空间估计的准确性与合理性, 进而为空间特异值和污染源的识别带来帮助.

目前, 空间特异值的研究主要集中在其对空间估计过程的影响, 特异值对污染源指示方面的研究相对较少. 本研究以稳健地统计学理论为基础, 选择一种稳健变异函数为工具, 对浙江省某典型重金属污染场地土壤中 Cu、Zn、Pb、Cd、As、Se 和 Hg 的含量进行分析, 分别判断其空间特异值, 分析其空间位置, 以期为污染源的识别提供有用信息, 并为后续的空间点格局分析奠定基础.

1 稳健变异函数的选择

为了更加适合于空间特异值的识别, 要求变异函数的估计量对空间特异值具有较好的稳健效果, 减弱“平滑效应”. 不同稳健估计量对不同类型的数据具有不同的稳健效果, 通常用以下性质来加以评判^[9-12].

(1) 失误敏感度 (gross error sensitivity) 用于估计引入一个任意大的特异值时所可能产生的最大影响. 该值越小, 说明稳健估计量对特异值具有较大的稳健效果.

(2) 崩溃点 (breakdown point) 崩溃点反映统计方法的全面可靠性, 是稳健性的全面度量, 它反映了要有多大比例的扰动才能使估计量破坏. 从稳健性角度, 希望有尽量高的崩溃点. 崩溃点是评价估计量稳健性最常用的指标.

(3) 渐近效率 (asymptotic efficiency) 表示样本容量无限增大时估计量效率的极限.

在稳健统计学中提出了多种稳健估计量, 但是

只有少数几种广泛应用于土壤科学的研究. Cressie 等^[13]利用偏差绝对值的平方根构建了稳健变异函数[见公式(2)], 并得到了广泛地应用. 但是该稳健变异函数具有一定的局限性. Genton^[14]指出, 公式(2)与(1)相比, 平方根转换虽然在一定程度上能减弱特异值的影响, 但是该稳健估计量的崩溃点是 0, 这就意味着即使只有一个极大的特异值就将使该稳健估计量毫无意义.

$$2\gamma_{CH}(h) = \frac{\left[\frac{1}{N(h)} \sum_{i=1}^{N(h)} |z(x_i) - z(x_i + h)|^{\frac{1}{2}} \right]^4}{0.457 + \frac{0.494}{N(h)} + \frac{0.045}{N^2(h)}} \quad (2)$$

Dowd 稳健变异函数是以中位绝对偏差 (median absolute deviation, MAD) 为基础建立起来的^[6], 见公式(3).

$$2\gamma_D(h) = 2.198 [\text{median } |y_i(h)|]^2 \quad (3)$$

式中, $y_i(h) = z(x_i) - z(x_i + h)$, $i = 1, 2, \dots, N(h)$.

MAD 的崩溃点是 50% (是所有稳健估计量崩溃点的理论最大值), 这就意味着即使有一半的成对差被极大或极小值所取代也不会影响估计量的效果; 失误敏感度为 1.166, 渐近效率为 0.37^[10].

Genton 发展了 Rousseeuw 等^[15,16]的 Q_n 估计量, 得到了 Genton 稳健变异函数.

$$2\gamma_G(h) = \{2.219 [|y_i(h) - y_j(h)|; i < j] (\frac{H}{2}) \}^2 \quad (4)$$

式中, $y_i(h)$ 的定义同公式(3), H 是 $[N(h)/2] + 1$ 的整数部分.

Q_n 稳健估计量的崩溃点是 50%, 失误敏感度为 2.069, 渐近效率为 0.82.

通过以上性质的对比, $\gamma_D(h)$ 和 $\gamma_G(h)$ 各有特点. 前期研究的结果表明, 研究区为典型重金属污染场地, 遭受到了较严重的重金属污染^[17], 因此有必要应用稳健估计量对其重金属的含量进行分析, 本研究主要选择 Genton 稳健变异函数作为分析工具.

2 处理思路及其过程

对 7 种元素含量执行适当的探索性数据分析, 去掉具有明显错误的的数据, 并对数据进行适当的变换. 前期研究结果表明, 7 种元素含量不符合正态分布, 进行对数转换后, 符合正态或近似正态分布, 而 Genton 估计量对近似正态分布数据具有更高的稳健效果, 所以本研究仍对数据进行对数转换. 分析数据的统计特征值见文献[17].

根据 Croux 等^[18]提供的计算方法,以 Genton 稳健估计量为工具,用距离组法计算每一种元素的实验变异函数,利用 Genstat 软件中的 Model Variogram 程序对实验变异函数进行模型模拟,选择最佳模型,得到变异函数的参数(变程、基台值和块金方差等)。

以理论变异函数模型的参数为基础,利用 GSLIB 软件系统中的 xvalid 程序对每一种元素含量进行克里格交叉检验,得到估计值和克里格方差 σ^2 等参数,进而根据公式(5)估计 7 种元素每一采样点位相对应的克里格标准误 $\epsilon(x_i)$ 。

$$\epsilon(x_i) = \frac{\hat{Z}(x_i) - z(x_i)}{\sigma}$$

(5)

式中, $z(x_i)$ 和 $\hat{Z}(x_i)$ 分别代表某一采样点某重金属含量的真实值与克里格估计值。

Bardossy 等^[19]建议,标准误能够用作识别特异值的标准,当标准误 < -1.96 时就把该数据当作空间特异值。本研究也采用这一方法。

把每一种重金属空间特异值点位的坐标分别代入 ArcGIS 8.2,得到研究区内各重金属空间特异值的空间分布图,并结合当地历史上及近些年的土地利用状况、企业分布图等综合分析污染源类型。

3 结果与分析

3.1 土壤重金属半方差函数的拟合模型及其参数

利用 Genstat 软件中的 Model Variogram 程序对实验变异函数进行模型模拟,选择最佳的模型,得到变异函数的变程、基台值和块金方差等 3 个重要参数,结果如表 1 所示。

块金方差反映了区域化变量内部随机性的可能程度,块金方差较大则表明较小尺度上的某种过程不可忽略。与以 Matheron 估计量为基础的传统变异函数得出的结果^[17]相比,该研究得到的块金方差 C_0 和基台值变小,Cu、Zn 和 Pb 甚至只为 0.000,这主要是由于 Genton 稳健变异函数对特异值的稳健效果所致。其他研究结果表明,传统变异函数所分析的数据不符合正态分布时,就可能存在比例效应,从而抬高块金方差和基台值,降低估计精度^[8,20,21]。本研究结果从事物的另一个方面很好地印证了这一点。块金方差与基台值之比表示随机部分引起的空间异质性占系统总变异的程度,如比值 $< 25\%$,说明系统具有强烈的相关性;而比例在 $25\% \sim 75\%$ 之间,表明系统具有中等的相关性。除 Se 外,其它重金属都具有较强的相关性,与文献[17]的结果相同。

表 1 7 种元素变异函数的拟合模型及其参数

Table 1 Semivariogram model parameters for 7 elements

元素	预测模型	块金方差 C_0	基台值 $C_0 + C$	$C_0 / C_0 + C$	有效变程 /m
Cu	球状	0.000	0.122	0.000	1 702
Zn	球状	0.000	0.106	0.000	828
Pb	球状	0.000	0.040	0.000	898
Cd	球状	0.017	0.192	0.089	1 747
As	球状	0.003	0.015	0.200	866
Se	球状	0.008	0.016	0.500	1 725
Hg	球状	0.002	0.040	0.050	1 514

3.2 土壤重金属含量的空间特异值的识别、空间分布图及其污染源识别指示意义

计算 7 种元素每一采样点位相对应的克里格标准误 $\epsilon(x_i)$,当该值 < -1.96 时就把该数据当作空间特异值,其空间分布如图 1 所示。

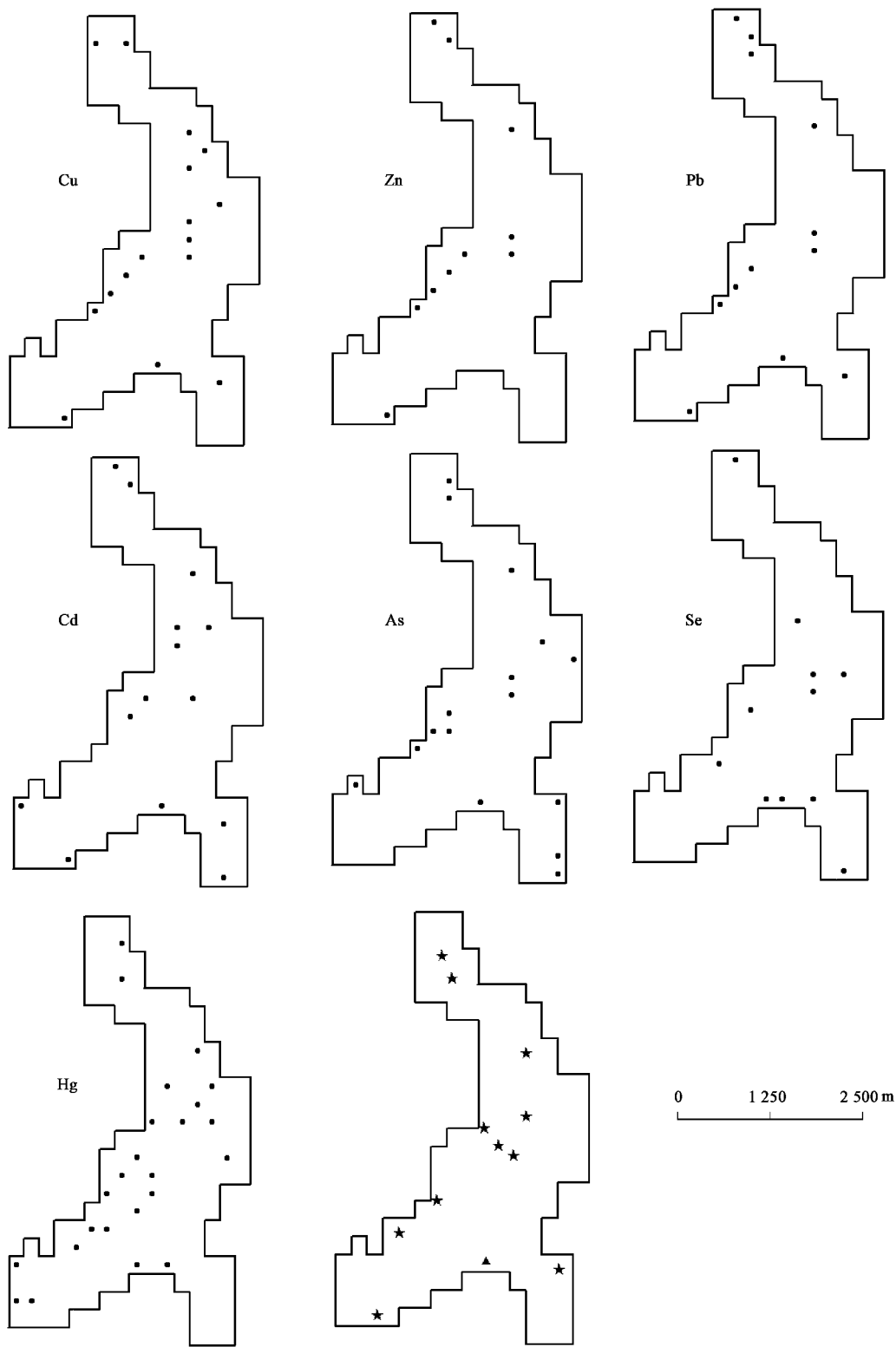
由图 1 可以看出,各重金属的空间特异值多集中在个别小范围内,以簇状形式存在,且与研究区内冶炼高炉及尾渣堆放处(三角星处)的位置具有很好的相关性。由此可以推断,冶炼行为及其相关过程是该区域土壤重金属的重要来源。需要注意的是,空间特异值的位置与污染源的位置并不具有一一对应的关系,因为采样点位并不一定分布在污染源的位置上。

前期研究表明,在该研究区域内,大气干湿沉降是重金属进入土壤的重要传播途径。虽然该研究区内冶炼高炉的主要产品是粗铜和粗锌,但由于冶炼原料是多元素的混合物,因此冶炼过程中也会带入其他的污染物。火法冶炼工艺中,为了得到粗铜,需要去除锌、铅、砷、镉及汞等元素,这些元素气化后形成氧化物微粒^[22~24],或聚集在煤炭燃烧所产生的飞灰颗粒的表面^[25,26],以灰尘的形式散发到大气中,进而通过干湿沉降等方式进入土壤^[27,28]。以上研究结果说明,重金属污染物从污染源向受体(土壤)传输的过程中,各元素可能存在着形态和数量上的联系,因而其空间特异值的空间分布图具有相似的形式。

各元素空间特异值的数量不尽相同,这可能与元素的性质有关。Cd、As 和 Hg 等元素具有较强的迁移性^[29~31],影响面积相对较大,可能导致更多空间特异值的产生。

4 结论

- (1)Genton 估计量对空间特异值具有很好的稳健效果,体现在得出的块金方差较小。
- (2)Genton 稳健变异函数能够识别研究区内的



图中圆形点表示各元素的空间特异值,五角星表示冶炼高炉位置,三角形表示冶炼尾渣堆放处

图 1 7 种元素的空间特异值和污染源分布

Fig.1 Distribution of spatial outliers of 7 elements and copper smelters

空间特异值,进而可对污染源的位置加以指示。

(3)空间特异值的分布与研究区内冶炼高炉的位置具有很好的相关性,说明冶炼高炉是该区域重金属污染的重要来源。

参考文献:

- [1] Lark R M. Modelling complex soil properties as contaminated regionalized variables[J]. *Geoderma*, 2002, **106**(3-4): 173-190.
- [2] Rawlins B G, Lark R M, O'Donnell K E, *et al.* The assessment of point and diffuse metal pollution of soils from an urban geochemical survey of Sheffield, England[J]. *Soil Use and Management*, 2005, **21**(3): 353-362.
- [3] 刘付程, 史学正, 于东升, 等. 基于地统计学和 GIS 的太湖典型地区土壤属性制图研究——以土壤全氮制图为例[J]. *土壤学报*, 2004, **41**(1): 20-27.
- [4] 路鹏, 彭佩钦, 宋变兰, 等. 洞庭湖平原区土壤全磷含量地统计学和 GIS 分析[J]. *中国农业科学*, 2005, **38**(6): 1204-1212.
- [5] 张朝生, 章申, 张立成, 等. 长江水系河流沉积物重金属元素含量的计算方法研究[J]. *环境科学学报*, 1995, **15**(3): 258-264.
- [6] Dowd P A. The variogram and kriging: robust and resistant estimators[A]. In: Verly G, David M, Journel A G, *et al.* *Geostatistics for Natural Resources Characterization*[C]. Dordrecht: Reidel, 1984. 91-106.
- [7] Laslett G M, McBratney A B. Further comparison of spatial methods for predicting soil pH[J]. *Soil Science Society of America Journal*, 1990, **54**(6): 1553-1558.
- [8] 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. 65-101.
- [9] Hampel F R. The influence curve and its role in robust estimation[J]. *Journal of the American Statistical Association*, 1974, **69**(346): 383-393.
- [10] Lark R M. A comparison of some robust estimators of the variogram for use in soil survey[J]. *European Journal of Soil Science*, 2000, **51**(1): 137-157.
- [11] 闫章更, 魏振军. 试验数据的统计分析[M]. 北京: 国防工业出版社, 2001. 313-333.
- [12] 周世健, 陈永奇. 污染分布下的 L_p 估计抗差测度[J]. *武汉测绘科技大学学报*, 1999, **24**(1): 71-74, 89.
- [13] Cressie N, Hawkins D M. Robust estimation of the variogram: I[J]. *Mathematical Geology*, 1980, **12**(2): 115-125.
- [14] Genton M G. Highly robust variogram estimation[J]. *Mathematical Geology*, 1998, **30**(2): 213-221.
- [15] Rousseeuw P J, Croux C. Explicit scale estimators with high breakdown point[A]. In: Dodge Y. *LI-Statistical Analysis and Related Methods*[C]. Amsterdam: North-Holland, 1992. 77-92.
- [16] Rousseeuw P J, Croux C. Alternatives to the median absolute deviation[J]. *Journal of the American Statistical Association*, 1993, **88**(424): 1273-1283.
- [17] 张长波, 李志博, 姚春霞, 等. 污染场地土壤重金属含量的空间变异特征及其污染源识别指示意义[J]. *土壤*, 2006, **38**(5): 525-533.
- [18] Croux C, Rousseeuw P J. Time-efficient algorithms for two highly robust estimators of scale[A]. In: Dodge Y, Whittaker J. *Computational Statistics Volume 1*[C]. Heidelberg: Physica-Verlag, 1992. 411-428.
- [19] Bardossy A, Kundzewicz Z W. Geostatistical methods for detection of outliers in groundwater quality spatial fields[J]. *Journal of Hydrology*, 1990, **115**(1-4): 343-359.
- [20] 郑袁明, 陈煌, 陈同斌, 等. 北京市土壤中 Cr, Ni 含量的空间结构与分布特征[J]. *第四纪研究*, 2003, **23**(4): 436-445.
- [21] 郭旭东, 杨福林. 基于 GIS 和地统计学的土壤养分空间变异特征研究: 以河北省遵化市为例[J]. *应用生态学报*, 2000, **11**(4): 557-563.
- [22] Martley E, Gulson B L, Pfeifer H R. Metal concentrations in soils around the copper smelter and surrounding industrial complex of Port Kembla, NSW, Australia[J]. *Science of the Total Environment*, 2004, **325**(1-3): 113-127.
- [23] 李文军. 铅烟治理技术探讨[J]. *矿山环保*, 2002, **2**(1): 23-24.
- [24] 周全法, 尚通明. 废电子元器件与材料的回收利用[M]. 北京: 化学工业出版社, 2004. 213-248.
- [25] Page A L, Elseewi A A, Straughan I. Physical and chemical properties of fly ash from coal-fired power plants with references to environmental impacts[J]. *Residue Reviews*, 1979, **71**(1): 83-120.
- [26] Sajwan K, Punshon T, Seaman J. Production of coal combustion products and their potential uses[A]. In: *Coal Combustion Byproducts and Environmental Issues*[C]. Now York: Springer, 2006. 3-9.
- [27] Payne G G, Martens D C, Kornegay E T, *et al.* Availability and form of copper in three soils following eight annual applications of copper-enriched swine manure[J]. *Journal of Environmental Quality*, 1988, **17**(4): 740-746.
- [28] Manceau A, Lanson B, Schlegel M L, *et al.* Quantitative Zn speciation in smelter-contaminated soils by EXAFS spectroscopy[J]. *American Journal of Science*, 2000, **300**(4): 289-343.
- [29] Martley E, Gulson B L, Pfeifer H R. Metal concentrations in soils around the copper smelter and surrounding industrial complex of Port Kembla, NSW, Australia[J]. *Science of the Total Environment*, 2004, **325**(1-3): 113-127.
- [30] 夏学齐, 陈骏, 廖启林, 等. 南京地区表土镉汞铅含量的空间统计分析[J]. *地球化学*, 2006, **35**(1): 95-102.
- [31] Cloquet C, Carignan J, Libourel G, *et al.* Tracing Source Pollution in Soils Using Cadmium and Lead Isotopes[J]. *Environmental Science and Technology*, 2006, **40**(8): 2525-2530.