

地统计学方法在环境污染研究中的应用^{*}

杨秀虹, 李适宇

(中山大学环境科学与工程学院环境科学系, 广东 广州 510275)

摘要: 空间自相关性是环境变量普遍具有的一个统计特征。地统计学方法是描述、模拟和预测这一类变量空间变异特征的有效工具。本文概述地统计学方法在环境污染研究中的应用情况, 主要包括: 通过变异函数分析来描述和模拟环境变量的空间分布结构; 由克立格插值技术来估算环境变量在未采样点处的取值; 采用多元地统计学方法研究多个环境变量间的空间相关结构; 通过指示地统计学方法评价空间插值的局部不确定性以及进行环境变量空间分布的随机模拟等。同时结合实例分析, 简要探讨了地统计学方法在描述、预测环境变量空间结构特征和空间相关分析方面的优势。

关键词: 地统计学; 空间自相关性; 环境变量; 污染

中图分类号: P628.2, X131 **文献标识码:** A **文章编号:** 0529-6579 (2005) 03-0097-05

研究污染物在环境中的空间分布特征、迁移转化规律是进行环境质量评价、环境污染修复的必要前提。在环境污染调查中, 通常对有限的污染样本进行常规统计分析(如均值、方差等统计特征)。经典统计学中一个重要假设是样本中个体的独立性。然而, 环境是一个时空连续体, 对具有空间自相关性的环境样本仅进行经典统计分析难以全面描述其特征, 而且不利于进一步考察环境变量的空间变异特征以及不同环境因素之间的空间相关关系。

源于采矿和石油地质勘探研究中的地统计学方法将变量观测值的空间坐标结合到数据处理过程, 通过半变异函数、多种克立格插值法来描述预测区域化变量在空间上变化的结构性, 预测未采样点处的取值以及评价预测结果的不确定性^[1]。鉴于地统计学方法在描述、模拟和预测自然现象空间变异方面所具有的优势以及在气象学、土壤学等领域中的成功应用, 近年来, 环境学研究也开始借鉴和引进地统计学的原理和方法, 并已成功应用到环境学各领域特别是环境污染研究中。本文主要介绍地统计学方法在环境污染研究中的应用情况, 并结合研究实例说明该方法在揭示环境变量空间分布特征以及空间相关分析方面的优势。

1 单变量的空间结构分析、模拟和估值

地统计学中的半变异函数分析、模拟以及各种克立格插值技术是进行环境变量空间结构分析、模拟和估值的主要工具。通过描述和模拟污染物的空

间分布特征以及估算未采样点处的取值, 从而揭示出污染物在空间上的分布迁移趋势; 此外, 结合现有的污染背景资料, 还可以识别出各种可能的污染来源。例如, 在沉积物污染调查中应用半变异函数分析和克立格插值方法揭示出污染物 DDT 的分布具有明显的空间依赖性, 同时阐明了其空间分布特点^[2]。Carlton 等^[3]通过球状模型的拟合分析得出不同种多环芳烃的含量的空间相关范围, 由克立格插值结果揭示出多环芳烃从东到西的迁移趋势。Facchinelli 等^[4]利用多变量分析和克立格插值等方法, 对土壤中重金属的不同来源(天然来源或人为来源)进行识别。在国内, 王学军、张朝生等对环境中污染物(如重金属)的空间变异性进行了不少研究^[5-8]。例如, 利用克立格方法对北京东郊污灌土壤表层重金属含量进行了空间分析和污染评价^[7]。张朝生等^[5, 9]采用地统计学方法研究了长江水系沉积物 11 种重金属元素含量的空间分布特征, 揭示出多数元素的空间相关尺度为 1000 km 左右, 而且空间相关性表现出一定的方向性。

2 多变量间的空间相关分析

研究单变量(如污染物的浓度)的空间分布固然重要, 但环境介质是个复杂的体系, 其中各种因素互相影响和作用。因此, 研究不同变量间的相关关系是否随空间尺度而改变以及如何改变等问题有助于深入了解驱动变量空间分布的机制。

多变量间的空间相关分析主要借助于多元地统

^{*} 收稿日期: 2004-05-01

基金项目: 教育部高等学校骨干教师资助计划项目

作者简介: 杨秀虹(1975年生), 女, 博士; 通讯联系人: 李适宇; E-mail: eeslsy@zsu.edu.cn

计学方法。常用的分析方法有交叉半变异函数分析、共离散系数（codispersion coefficient）分析、因子克立格分析等^[1,9]。在一项关于表土中铜和钴含量的空间相关研究中，由于多元地统计学方法的应用，使得隐含在两者总体弱相关关系（经典的 Pearson 相关分析）中的区域强相关关系得以显示^[10]。文献 [11] 应用交叉半变异函数分析和因子克立格分析揭示出土壤中多种重金属之间的空间相关性以及共同的污染释放源。此外，通过多元地统计分析（如协同克立格法）还能结合容易获得的辅助变量的信息来改善主变量的估值效果。Zhang 等^[12] 分别采用克立格和协同克立格法对土壤中硝酸盐的分布进行空间插值分析。结果表明，结合了 pseudo crossvariograms 的协同克立格法是确定非点源污染空间分布的一种既精确又经济的方法。Jeannée 等^[13] 充分利用各种土壤定性指标（如煤焦油的存在与否，土样是否有气味等）的信息来改善主变量苯并（a）芘含量的空间插值效果。

此外，地统计学分析和多元统计分析（如因子分析）相结合的研究方法也是描述多变量空间关系的有效工具。Bořuvka 等^[14] 通过结合地统计学方法和因子分析法，揭示了土壤中两种形态铝和土壤基本特性的空间分布特点以及它们间的相互关系。在另外一项研究中，克立格法和主成分分析法一起被用于描述多环芳烃污染的空间分布特点以及鉴别主要的和次要的污染源。作者认为这些方法的应用有助于从较小的样本数据中提取风险评价所需要的基本信息^[3]。

3 其它方面的应用

地统计学方法除了在上述两个方面的应用外，在空间插值的局部不确定性评价、区域化变量空间分布的随机模拟等方面的应用也逐渐得到重视。环境污染研究中对污染物浓度或土壤特性进行克立格插值预测并非最终目的。这通常是为进一步的污染评价（如污染物浓度超标的风险）、修复决策提供依据。因此，估计污染物的超标程度也是多数研究者的关注点。此外，利用克立格法进行估值不可避免存在一定的误差，在决策中必需考虑到这种附带误差的影响，以尽可能避免对污染区范围的错误判断^[9,15]。地统计学中的非参数方法（或称指示方法）可用来解决这些问题^[9]。Colin 等^[16] 利用具有详尽样本数据的电阻率与信息量不足的多环芳烃浓度之间的相关关系，通过简单指示克立格技术和概率模型对多环芳烃超标的风险进行评价。Goovaerts 等^[17] 利用指示地统计学方法对表土中铜、

铜和铅的污染风险进行预测并绘制污染超标概率分布图。

克立格插值法虽然对变量在空间的取值产生单一且统计学上“最优”的估值分布图，具有不少优点。但它存在平滑效应，从而造成低值高估和高值低估的现象。而且利用克立格估值的方差来估计真实值的方差，估计往往偏小^[9]。不同于克立格估值，地统计模拟关注的是区域化变量空间变异特征的重现。因此，它在诸如区域污染特征的刻画以及相关污染风险评价等方面显示出比克立格法更有优势^[8,18]。Mckenna^[19] 应用地统计模拟确定了在可接受超标概率条件下需要对铅进行修复的污染面积图以及相应的修复费用。张泽浦等^[8] 通过比较污灌土壤表层中微量元素含量的条件模拟结果与克立格插值结果发现，空间模拟方法可以更好地再现土壤微量元素的空间结构特征。此外，污染物分布的空间可变性会给土壤修复方案的成本预算带来不确定性，并往往是导致最终成本超出预算的重要原因。有鉴于此，Broos 等^[15] 通过地统计模拟技术估算出土壤修复方案成本预算中 8%~13% 的不确定性来自土壤污染物和岩性的空间变化。

4 实例分析

下文将对污染土壤中微生物量和多环芳烃含量进行地统计学分析，说明其中的半变异函数分析、克立格插值、空间相关分析等方法的具体应用，并简要讨论其在描述、预测环境变量的空间结构特征和空间相关分析方面的优势。

多环芳烃污染土样采自瑞典南部一个废弃的木材防腐场地。研究样地面积为 280 m×40 m。采用规则网格布点与局部加密布点相结合的采样方案。土壤中微生物量和多环芳烃含量分别由磷脂类脂肪酸总含量（totPLFA）和 16 种优控多环芳烃总含量（sumPAH）来表示。表 1 给出了这两个变量观测数据的基本统计特征。

由表 1 中的变异系数（大于 100%）可知，原数据样本具有强变异性。显然，对于具有高度变异性的随机变量而言，仅采用平均值、标准差、变异系数等基本统计量来反映该变量的总体水平以及取值分布情况是非常粗略的，而且因没有考虑空间坐标，从这些统计结果中无法得到研究场地中微生物量和多环芳烃总含量在空间上的总体分布格局，也无法给出每一空间点（除采样点外）的具体取值。这种粗略的统计结果不利于指导污染土壤微生物修复工程的开展。为此，下文采用地统计学方法对样本数据进行空间结构分析以及插值计算。

表 1 经典统计特征值¹⁾

Tab 1 Summary statistics

变量	平均值	标准偏差	最小值	最大值	偏斜度	峰度	变异系数/%
totPLFA	152.48	170.91	7.41	1001.82	2.21	5.99	112
sumPAH	100.73	212.67	0.36	1803.70	5.01	33.93	211

1) totPLFA 的单位是 nmol/g (干质量), sumPAH 的单位是 mg/kg (干质量)

4.1 土壤微生物量的空间结构特征分析

为了获得较为稳健的实验半变异函数，将具有偏斜分布的样本数据进行自然底对数转换。以下采用经对数转换的样本数据进行各种地统计学的分析。

因研究区域呈长方形，且变量 totPLFA 的空间各向异性不明显，所以只对其进行全方位的空间分析。实验半变异函数的计算和理论模型拟合使用 GS+ 软件 (Version 5.1.1, Gamma Design Software, USA)。采用有块金效应的球状模型进行拟合。模型表达式是：

$$\gamma(h) = C_0 + C[1.5(\frac{h}{A_0}) - 0.5(\frac{h}{A_0})^3], h \leq A_0$$

$$\gamma(h) = C_0 + C, h > A_0$$

式中， h 表示分隔空间两点的矢量，又称滞后距； C_0 是块金变差 (不小于 0)； C 是结构变差 (不小于 C_0)； A_0 表示变程的参数，球状模型中 A_0 等于有效变程 a 。各模型参数分别是： $C_0=0.286$ ， $C=1.143$ ，基台值 (C_0+C) = 1.429， $A_0=11.00\text{ m}$ 。模型的拟合效果较好，回归系数 R^2 是 0.62，残差平方和 RSS 是 1.07，见图 1。

实验半变异函数的计算和拟合结果说明：①该研究场地中土壤的 totPLFA 在变程内 (11.00 m) 具有明显的空间自相关性。只有超过该距离之后，totPLFA 不存在空间自相关性，经典统计方法的前提假设才得到满足。②块金系数不为零，说明除了采样或分析误差外，在小于最小采样间距 (0.625 m) 时还可能存在未被观测到的微观空间结构。如有学者观察到土壤中微生物在微小尺度上 (如厘米级、毫米级) 仍存在明显的空间分布结构^[20]。③基台值 (C_0+C) 表示变量达到的最大变异。结构变差与基台值的比值称为相对结构变差，用于反映空间自相关程度。本例中，该比值为 80%，表明总变异中的 80% 是源于空间自相关结构。这种显著的空间自相关性反映了土壤中微生物活动在变程内具有高度的空间连续性。

可见，如果仅从经典统计学的角度来解释具有明显空间结构的数据的话，不但可能损失大量有用的信息，而且可能得出模糊或不合理的结论，给实验结果的分析、应用带来困难。

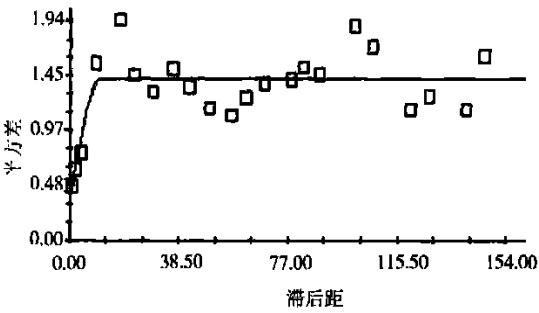


图 1 totPLFA 的实验半变异函数图及其拟合的球状模型

Fig 1 Experimental semivariogram for totPLFA and its fitted spherical model

4.2 克立格法制图

地统计学中的克立格法是一种求最优、线性、无偏内插估计量的局域插值法。本例采用块段普通克立格法，将研究区划分为 1 m×1 m 的块段进行插值计算并绘制分布图 (见图 2a)。由图可知，土壤中微生物量的空间分布并不是随机的，而是呈现明显的空间斑块状。斑块的平均延伸范围与由半变异函数模型所揭示的变程基本相符。在变程范围内，微生物量的空间分布具有自相关性，相近的取值点趋于集结在一起。利用克立格法进行估值的同时还可以计算出克立格估值标准差 (见图 2b)。克立格估值标准差 (SD) 的范围是 0.28~1.12，均小于实测值的标准差 (1.15)，这与克立格估值产生的平滑作用有关。图 2b 中 SD 的分布非常有规律，基本上反映了采样点的分布形式。这是因为这种“优先采样”的布点方式造成了局部地区信息样品量比其它地区要密集得多，所以 SD 也相应较小。影响 SD 大小的主要因素有：待估块段、信息样品 (实测样品) 的几何特征 (如大小、形状)、数量和空间分布以及半变异函数的特征^[1]。

4.3 多变量间的空间相关分析

利用地统计学中的共离散系数可以考察两个变量间的关联度随空间尺度的变化规律。共离散系数的计算公式如下：

$$\rho_{XY}(h) = \frac{\gamma_{XY}(h)}{\sqrt{\gamma_{XX}(h)\gamma_{YY}(h)}}$$

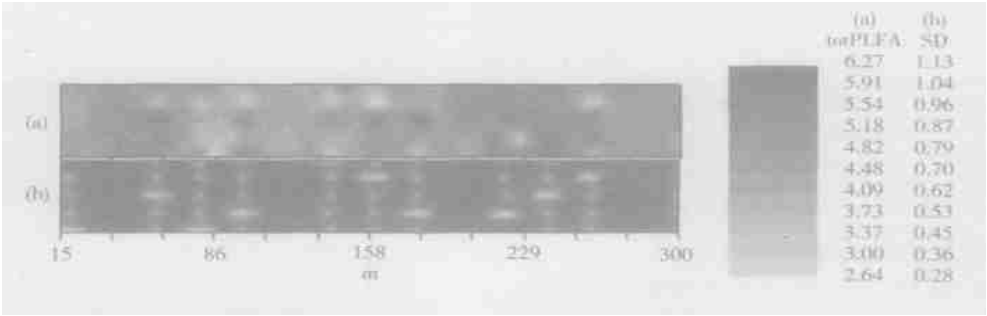


图 2 totPLFA 的块段克里格插值结果 (a) 以及插值标准差 (b) 分布图
Fig. 2 Contour maps of totPLFA (a) and its kriged standard deviation (b) by block ordinary kriging

其中， $\gamma_{xx}(h)$ ， $\gamma_{yy}(h)$ ，和 $\gamma_{xy}(h)$ 分别是变量 X 和 Y 在滞后距 h 的半变异函数以及交叉半变异函数。如果描述两个以上变量间的空间相关结构，可将两两变量间的共离散系数矩阵进行主成分分析，以获取多变量间的关联度如何随空间尺度而变化的信息。如果该系数不随空间间距而改变，说明变量间的空间相关结构不受空间尺度影响；系数为正时说明两变量以同一方式发生变化^[9]。本例应用该方法来考察 totPLFA 与 sumPAH 的空间相关关系（见图 3）。由图可知这两个变量在所有空间间距上均呈正相关关系，而且相关关系随空间间距而变化。比如滞后距处于 1~18 m 范围内共离散系数呈明显上升趋势，即在较小的空间尺度下 totPLFA 与 sumPAH 的相关性较弱，说明除了多环芳烃的影响外其它因素（如土壤团粒结构）对土壤微生物也存在一定程度的影响。而靠经典的 Pearson 相关分析是无法获得这些信息的。

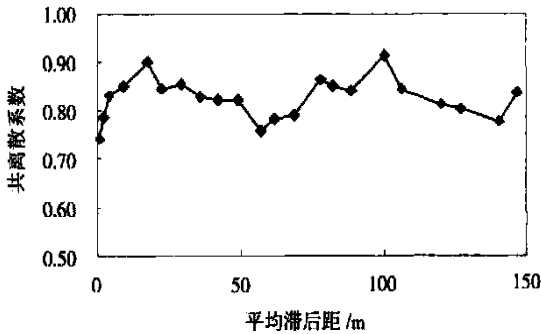


图 3 totPLFA 与 sumPAH 的共离散系数随不同滞后距的变化曲线图
Fig. 3 Codispersion coefficient between totPLFA and sumPAH against different lag distance

5 结 语

环境变量通常是具有空间自相关性的区域化变量。研究表明，经典统计方法无法提供这类变量在空间分布结构方面的信息。作为一种处理空间数据

的有效工具，地统计学方法被证明适合用于描述、模拟和预测环境变量的空间分布模式以及揭示多变量间的空间相关结构等方面的研究。此外，地统计学本身也在不断的发展中，近年来出现了不少新方法用于处理新的问题。如上文提到的变量空间插值的局部不确定性评价以及变量空间分布的随机模拟。需要指出的是，地统计学仅是一种处理数据的有效工具。它不能从有缺陷的样本数据中得出好的结果，但它有助于最大可能地提炼出有用信息。此外，在地统计学分析过程中往往需要结合诸如常识、专家的推断等方面的“软”信息，以保证结果的合理性。致谢：文中的实例研究得到瑞典隆德大学生态系 Göran Bengtsson 教授和 Niklas Törneman 博士的支持与合作，在此表示衷心感谢。

参考文献:

[1] WEBSTER R. Geostatistics for environmental scientists[M]. New York: John Wiley & Sons, Ltd 2001.
[2] YING O Y, NKEDI-KIZZA P, MANSELL R S, et al. Spatial distribution of DDT in sediments from estuarine rivers of central Florida[J]. Journal of Environmental Quality, 2003, 32(5): 1710—1716.
[3] CARLON C, CRITTO A, MARCOMINI A, et al. Risk based characterization of contaminated industrial site using multivariate and geostatistical tools [J]. Environmental Pollution, 2001, 111(3): 417—427.
[4] FACCHINELLI A, SACCHI E, MAILLEN L. Multivariate statistical and GIS-based approach to identify heavy metal sources in soils [J]. Environmental Pollution, 2001, 114: 313—324.
[5] 张朝生, 章申, 何建邦. 长江水系沉积物重金属含量空间分布特征研究——地统计学方法[J]. 地理学报, 1997, 52(2): 184—192.
[6] 张朝生, 章申, 何建邦. 长江水系沉积物重金属含量空间分布特征研究——空间自相关与分形方法[J]. 地理学报, 1998, 53(1): 87—96.
[7] 王学军, 席爽. 北京东郊污灌土壤重金属含量的克里

- 格插值及重金属污染评价[J]. 中国环境科学, 1997, 17(3): 225—228.
- [8] 张泽浦, 王学军. 土壤微量元素含量空间分布的条件模拟[J]. 土壤学报, 1998, 35(3): 423—429.
- [9] GOOVAERTS P. Geostatistics tools for characterizing the spatial variability in microbiological and physico-chemical soil properties[J]. Biology and Fertility of Soils, 1998, 27: 315—334.
- [10] GOOVAERTS P, WEBSTER R. Scale-dependent correlation between topsoil copper and cobalt concentrations in Scotland[J]. European Journal of Soil Science, 1994, 45: 79—95.
- [11] EINAX J W, SOLDT U. Multivariate geostatistical analysis of soil contaminations[J]. Fresenius' Journal of Analytical Chemistry, 1998, 361: 10—14.
- [12] ZHANG R D, SHOUSE P, YATES S. Estimates of soil nitrate distributions using cokriging with pseudo-crossvariograms[J]. Journal of Environmental Quality, 1999, 28(2): 424—428.
- [13] JEANNÉ E N, DE FOUQUET C. Contribution of auxiliary information for the estimation of grades in heterogeneous media: example in soil pollution[J]. Comptes Rendus Geoscience, 2003, 335: 441—449.
- [14] BORUVKA L, KOZÁK J. Geostatistical investigation of a reclaimed dumpsite soil with emphasis on aluminum[J]. Soil & Tillage Research, 2001, 59: 115—126.
- [15] BROOS M J, AARIS L, VAN TOOREN C F, et al. Quantification of the effects of spatially varying environmental contaminants into a cost model for soil remediation[J]. Journal of Environmental Management, 1999, 56: 133—145.
- [16] COLIN P, FROIDEVAUX R, GARCIA M, et al. Integrating geophysical data for mapping the contamination of industrial sites by polycyclic aromatic hydrocarbons: a geostatistical approach[A]. Geostatistics for environmental and geotechnical applications[C]. West Conshohocken, PA: American Society for Testing and Materials, 1996: 69—87.
- [17] GOOVAERTS P, WEBSTER R, DUBOIS J P. Assessing the risk of soil contamination in the Swiss Jura using indicator geostatistics[J]. Environmental and Ecological Statistics, 1997, 4: 31—48.
- [18] GOOVAERTS P. Geostatistics in soil science: state-of-the-art and perspectives[J]. Geoderma, 1999, 89: 1—45.
- [19] MCKENNA S A. Geostatistical approach for managing uncertainty in environmental remediation of contaminated soils: Case study[J]. Environmental & Engineering Geoscience, 1998, 4(2): 175—184.
- [20] ETTEMA C H, WARDLE D A. Spatial soil ecology[J]. TRENDS in Ecology & Evolution, 2002, 17(4): 177—183.

Geostatistical Methods for Research on Environmental Pollution

YANG Xiu hong, LI Shi yu

(Department of Environmental Science, School of Environmental Science and Engineering,
Sun Yat sen University, Guangzhou 510275, China)

Abstract: Autocorrelation is a general statistical property of environmental variables. Geostatistics provides useful tools to describe, model and predict spatial heterogeneity of environmental variables. The main applications of geostatistics to the research on environmental pollution include introducing semivariance analysis to characterize and model the spatial structures of environmental variables; using kriging methods to estimate values of environmental variables at unsampled locations; adopting multivariate geostatistics to detect the spatial correlation structures between different environmental variables; assessing the local uncertainty of predicted unsampled values by indicator geostatistics, and performing stochastic simulation for the spatial distributions of environmental variables. The advantages of geostatistics analysis in characterizing and predicting spatial structures and spatial correlations of environmental variables are briefly discussed using a case study.

Key words: geostatistics; spatial autocorrelation; environmental variable; pollution