

环境空间结构分析中模拟数据集的引入^{*}

齐 峰 王学军 黄泽春

(北京大学城市与环境学系, 北京 100871)

摘要 利用对高程数据的数学变换建立了包含 10000 个数据点的模拟土壤微量元素含量基线值大型数据集, 并在此基础上进行污染模型叠加产生污染影响大型数据集. 以之为基础进行模拟采样, 研究了采样方法对空间结构表达的影响. 通过对实验半变异函数的比较发现了各采样方法不同的特点和空间结构表达效果.

关键词 环境空间分析, 空间结构分析, 模拟数据集, 采样方法.

Application of Environmental Spatial Structure Analysis with Simulated Data Sets^{*}

Qi Feng Wang Xuejun Huang Zechun

(Department of Urban and Environmental Sciences, Peking University, Beijing 100871, China)

Abstract Large data sets of baseline trace metal contents in soil, each composed of 10000 spots, were simulated by mathematical transformation with elevation data. Pollution models were added to these data to simulate pollution data sets. Based on these data sets, different sampling designs were conducted and their efficiencies in interpreting spatial continuity were studied. By comparison of the sample semi-variograms, characteristics and influences on spatial analysis of each sampling design were discovered.

Keywords environmental spatial analysis, spatial structure analysis, simulated data sets, sampling designs.

自 70 年代末、80 年代初开始, 对于环境变量这些典型的区域化变量^[1]的统计分析不再只限于普通统计学范畴, 开始实施空间分析如空间插值与空间模拟等. 用于预测、评价及空间数据库的建立等方面, 空间结构分析是空间分析的基础步骤. 在空间分析的理论研究中, 可以利用实测数据, 也有大量利用各种模拟数据进行研究的经验^[2~5]. 由于环境变量的实际测量既需花费大量的人力、物力而又仅能提供有限的数量, 本文采用了 2 套大型模拟数据集研究采样方案对空间结构表达的影响.

1 空间结构分析与数据获取

空间结构分析一般指空间连续性描述或称空间自相关性描述, 是表征变量空间分布结构的一种方法^[1]. 协方差函数和半变异函数是定量表示空间自相关性的有效工具, 分别表示为:

$$C(x_1, x_2) = E\{[Z(x_1) - m(x_1)][Z(x_2) - m(x_2)]\} \quad (1)$$

$$\gamma(x_1, x_2) = \frac{1}{2} \text{Var}\{Z(x_1) - Z(x_2)\} \quad (2)$$

实验半变异函数是对半变异函数的最优估计, 设 h 是一个模 $r = |h|$, 方向为 α 的矢量, 如果有被矢量 h 所隔开的 N 对数据, 则在 α 方向上, 对于 h 的实验半变异函数值为:

$$\gamma^*(r, \alpha) = \frac{1}{2N} \sum_{i=1}^{N'} [z(x_i + h) - z(x_i)]^2 \quad (3)$$

以矢量 h 为横轴, 实验半变异函数值为纵

^{*} 国家教委跨世纪人才培养计划及国家自然科学基金资助项目 (Project Supported by Trans-Century Training Program Foundation for the Talents by the State Education Commission of China and the National Natural Science Foundation of China): 49771068
作者简介: 齐峰(1977~), 女, 硕士生
收稿日期: 1998-12-24

轴,可作实验半变异函数曲线,它通过起伏特征、原点处性状、趋势走向、方向差异性等形状特点提供丰富的空间结构信息.

在空间分析理论研究中,数据来源一般有2种.最初空间分析仅用于地质采矿,全球长期的采矿、勘探实践积累了丰富的数据,因而研究中一般利用实测数据集.但当地统计学为其它学科应用后,大型实测数据集往往难以得到,许多学者于是用模拟数据集进行研究. Isaaks 等利用高程数据的数学变换数据集进行了一系列地统计学理论研究^[4], Weber 等也用数字高程模型及其衍生数据进行了空间插值方法的比较研究^[5], 国内陶澍、冯泉等根据自然过程对土壤元素空间分布的影响模拟了土壤元素含量*. 在这些模拟方法中,应该说考虑成土过程的方法更能反映实际空间结构,但这种方法势必带来较繁重的工作量,且目前方法还不成熟.在实施条件有一定限制的情况下,空间分析理论研究中采取类似 Isaaks 等的方法,利用高程数据的衍生数据集作为研究的基础数据,既可以节约大量的人力物力,也仍可以得到有相当参考价值的理论研究结果.

2 模拟数据集的生成

2.1 模拟基线值数据集(baseline data sets, Bds)

应用1 10万地形图数字化的高程数据衍生出2个基本数据集:模拟土壤元素基线值数据集1(Bds1)和2(Bds2). 据 Isaaks 等^[5], 高程数据进行数学变换可产生不同空间分布类型的

数据,其中方差变换数据的变异性较大,分布类似土壤或沉积物中的微量元素含量. 本研究对高程数据的变换就以方差变换为基础,又与实际元素含量在数值尺度及分布特征上进行了一致化转换.

令 A 代表 Bds1 中的变量,每个点值由原始高程数据 4×4 栅格点阵的方差值经换算得出: $A = \frac{1}{1000}(1 + \sigma^2)$; (4)

其中, $\sigma^2 = \frac{1}{16} \sum_{i=1}^{16} [x_i - m(x)]^2$. 方差值加1以使所有数据均不为0,乘以比例系数 $\frac{1}{1000}$ 以符合土壤微量元素含量的大体尺度. 令 B 代表 Bds2 的变量,其值为方差的函数,空间变异性较小:

$$B = \frac{1}{1000}m(x) \cdot \sqrt{\sigma^2 + 1} \quad (5)$$

每个数据集包括 100×100 的规则栅格数据点10000个,基本统计量列于表1. 图1a和1b分别是2数据集的三维效果图,相应的直方图见图2. 由图2所示,直接方差数据 A 变量的分布呈显著的尖峰态和右偏态,空间分布连续性较弱,噪声大,变异系数较大,类似土壤中微量元素 Cd、Hg 等的分布特征,其数值通过比例系数的修正后与这2种元素含量监测值的空间尺度相吻合. B 变量的分布特征除象 A 一样依旧表现出土壤重金属元素分布的对数正态形式外,在空间分布上呈连续形态,起伏相对平缓,经尺度比例修正后,可代表 Cu、As、Pb 等元素的分布类型.

表 1 基线值数据基本统计量

数据集	均值	标准差	变异系数 / %	偏度系数	峰度系数	第一 四分位数	第二 四分位数	第三 四分位数	第四 四分位数
Bds1	0.077	0.089	116.66	2.66	12.00	0.020	0.044	0.098	1.205
Bds2	27.42	16.89	61.57	1.24	1.88	14.81	23.44	36.08	133.00

2.2 模拟污染土壤数据集(pollution data sets, Pds)

母质、地形、气候和各种生物地球化学作用决定了土壤中微量元素含量基线值的分异. 近

年工业生产的加速发展及人类活动的加剧导致

* 冯泉,陶澍.土壤表层元素含量二维空间分布形式模拟. 北京大学硕士学位论文,1994

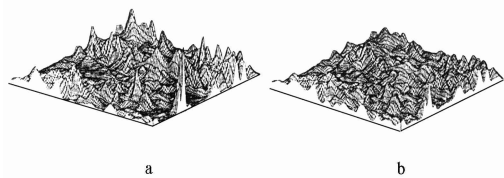


图 1 基线值数据集三维效果

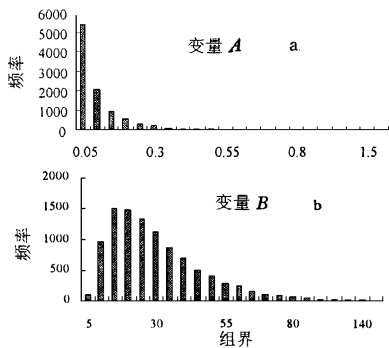


图 2 基线值数据直方图

各种污染物包括重金属元素在土壤中进一步累积。如重金属可通过金属冶炼的烟气排放过程进入土壤,经逐步积累达到危害农作物的程度,污泥和污水的农业利用也是土壤重金属污染的重要来源^[7]。本文中的污染数据模拟主要考虑了工业污染物的大气沉降和污灌 2 种因素,对各种污染因子设定了污染源位置、强度等条件,通过建立模型模拟进入土壤中的污染物量,再通过与原数据集 Bds1 的基线值数据迭加而得到模拟污染数据集。

(1) 大气沉降影响的污染数据(Pds1) 在本研究中,对于大气沉降污染主要考虑高架点源的高斯扩散引起地表大气中污染物浓度增高进而沉降到地面并为土壤固定而导致土壤中微量元素含量累积。方法是首先在待研究地块上设定污染源位置(见图 3a),根据不同研究目的给定模型参数和源强,利用高斯模型计算出地表大气污染数据,再以一定的比例系数确定沉降到土壤中的污染物含量,在原基线数据上迭加而得到模拟值。计算中模型参数的选取、外界条件如风向等的设定及数量级的确定等都依实际记录而定。图 4a 给出了一个模拟结果的三维效果,这是对照天津市环境监测站确定的土壤

重金属污染分级标准选取了符合较严重污染的一个数据集得到的,它可作为以后研究的基础数据集。

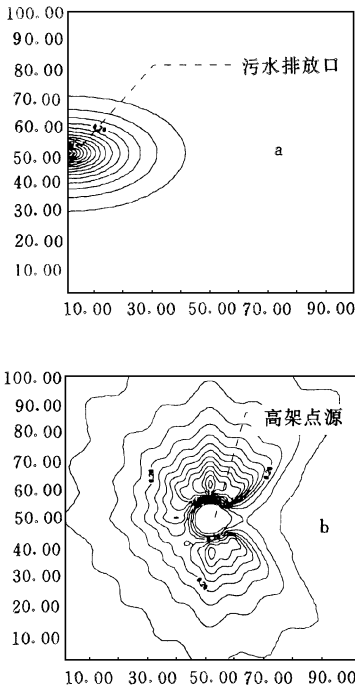


图 3 污染影响等值线图

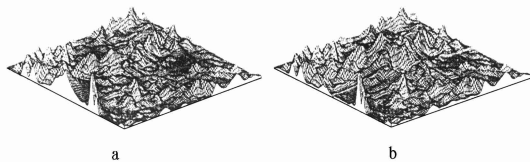


图 4 污染影响数据集三维效果

(2) 污灌影响的污染数据(Pds2) 对于污灌影响,考虑单一排放口污灌使污水中的重金属元素与土壤迅速吸附络合而引起的土壤中微量元素累积。污水在农田的流动中,其中的重金属等元素会因土壤的阻持、沉淀、吸附等物理、化学作用而留存于土壤中,对土壤造成污染。本文中这种污染模型设计参照了冯汉平等对污灌土壤中微量元素微域分异规律的研究^[8],基于该研究中建立的浓度与距污灌口距离的关系曲线,通过回归拟合出浓度随距离变化的函数关系。计算中设定了排污口位置,如图 3a 所示,考虑了一定

的地表性状和出水口水流特征, 根据实际经验确定了各方向迁移系数等参数, 进行二维污染模拟, 点 (x, y) 处的含量为:

$$\hat{c}_{ss}(x, y) = c_0 \cdot \exp \left[-0.0332 \cdot \frac{(D_x \cdot x)^2 + (D_y \cdot y)^2}{(D_x \cdot x)^2 + (D_y \cdot y)^2} \right] \quad (6)$$

其中, c_0 表示污灌口浓度, D_x 和 D_y 分别表示 x, y 方向的迁移系数. 污染数量级和污染浓度大小的确定仍旧参照与高斯影响数据同样的实测数据和经验分类标准. 图 4b 给出了一个根据较严重污染的情况建立的模拟污染数据集的三维效果.

3 结果与讨论

根据上述几个数据集得到的实验半变异函数表现出各异的空间结构特征. 2 个基线数据集都表现了显著的可迁性, 但二者基台值、空穴效应、各向异性特征明显不同. 污染数据集表现出不同于基线数据的空间特征, 2 种污染的叠加都因在原基线数据面上加入了趋势结构而使实验半变异函数出现漂移, 成非平稳性, 2 种污染方式还导致了截然不同的空间各向异性.

在模拟数据集上进行了采样方法对空间结构表达效果影响的比较研究. 以不同的采样方法、采样密度抽取样本集, 计算样本实验半变异函数, 与 100×100 大型数据集计算得到的实验半变异函数比较, 衡量采样方法的效果. 结果表明, 环境调查中几种常用的采样方法在空间结构表达上各有特点: 系统采样结果噪声小, 但在采样密度低时不可靠; 系统随机采样有噪声干扰, 但可通过有容差的计算予以改善, 而且在采样密度较小时可以准确反映半变异函数趋势; 随机采样噪声大, 虽可利用有容差的计算予以减弱, 但趋势不准确, 最不可取. 对污染数据平面还进行了分层采样和优先采样的试验, 结果表明, 在污染区适宜地加大采样布点密度可以较方便地同时反映小区域污染结构与大范围的综合变异性特点, 但却存在着如何事先准确划分污染区的问题.

目前常用的环境空间分析方法如克里格插值方法, 它作为一种最优估计方法, 在环境评价、环境污染范围、水平的确定以及空间数据库的建立和工程措施的采取等方面都具有重要意义. 本文讨论的空间结构分析是空间插值的基础, 它通过影响理论半变异函数拟合的合理性影响着插值效果. 模拟数据集的应用是这项空间分析理论研究的基础所在, 在普遍存在的数据缺乏条件限制下, 利用模拟数据进行理论研究, 在数据量、实施便捷性方面显示了强大优势. 正如前人在统计估值、方差预测及地统计学理论研究中利用模拟数据得到了有效结论一样, 利用以上所述的 2 套模拟数据集比较了采样方法对空间结构分析的影响, 发现了各采样方式的特点和适用环境, 对环境工作的实践采样具有指导意义.

参 考 文 献

- 1 Journal A G, Juijbrechts Ch J. Mining Geostatistics. London: Academic Press, 1978. 21 ~ 230
- 2 McArthur R D. An evaluation of sample designs for estimating a locally concentrated pollutant: Communications in Statistics Computation and Simulation, 1987, **16**(3): 735 ~ 759
- 3 Cordy C B, Thompson C M. An Application of Deterministic Variogram to Design-Based Variance Estimation: Math. Geology, 1995, **27**(2): 173 ~ 206
- 4 Isaaks E H, Srivastava R M. An Introduction to Applied Geostatistics. New York: Oxford University Press, 1989. 6 ~ 98
- 5 Weber D D, Englund E J. Evaluation and Comparison of Spatial Interpolators: Math. Geology, 1994, **26**(5): 589 ~ 603
- 6 Overton W S, Stehman S V. Properties of designs for sampling continuous spatial resources from a triangular grid: Communications in Statistics theory and Methods, 1993, **22**(9): 2641 ~ 2660
- 7 买永彬等. 农业环境学. 北京: 中国农业出版社, 1994. 50 ~ 83
- 8 冯汉平等. 污灌土壤中铜和锌含量及其微域分异. 环境化学, 1991, **11**(2): 33 ~ 38
- 9 桑建国等. 大气扩散的数值计算. 北京: 气象出版社, 1992. 23 ~ 150