

# 城市表层土壤重金属污染分析

周敏强<sup>1</sup>, 盛振峰<sup>1</sup>, 黄思<sup>1</sup>

指导教师: 曹春正<sup>2</sup>

(1. 南京信息工程大学 大气物理学院, 江苏 南京 210044; 2. 南京信息工程大学 数学与统计学院, 江苏 南京 210044)

**摘要:** 重金属污染是环境污染的主要指标之一。本文利用某城市实际观测数据, 对该城市的重金属污染状况进行推断。首先利用地累积污染指数衡量重金属污染程度, 进而基于克里金插值法和 ArcGIS 软件分析各重金属元素污染指数的空间分布特征, 据此对重金属污染离子分类, 并基于污染负荷指数比较不同功能区的污染程度。其次, 采用 Pearson 相关性分析和主成分分析对所得分类进行合理性检验, 并结合重金属来源分类的国家参考标准和污染物的分布特性, 得出各类重金属离子的污染来源。最后, 利用指数衰减模型, 对所有样本的高程信息进行订正, 并利用加权混合二元正态分布密度函数去拟合多污染源传播形成的浓度曲面, 估计位置参数, 确定出重金属污染源的具体位置。

**关键词:** 重金属; 污染分析; 地统计; 空间数据; 地累积指数; 混合正态分布

中图分类号: O241.3; X8

文献标志码: A

文章编号: 2095-3070(2012)01-0062-05

## 1 引言

随着城市经济的快速发展和城市人口的不断增加, 人类活动对城市环境质量的影响日益突出。对城市土壤地质环境异常的查证, 以及如何应用查证获得的海量数据资料开展城市环境质量评价, 日益成为人们关注的焦点。

本文利用某城市城区土壤地质环境的实测数据, 分析并推断该城市的重金属污染分布状况。首先利用地累积指数结合克里金插值法得到各重金属元素的空间分布特征, 并对重金属污染离子分类, 进而基于污染负荷指数比较不同功能区的污染程度。其次, 利用 Pearson 相关性分析和主成分分析对分类进行合理性检验, 并结合重金属来源分类的国家参考标准和污染物的分布特性, 推断各类重金属离子的污染来源。最后, 利用指数衰减模型, 对样本的高程信息进行订正, 并利用加权混合二元正态密度函数去拟合多污染源传播形成的浓度曲面, 由位置参数估计出污染源位置。

## 2 模型建立及求解

### 2.1 重金属污染程度分析

直接用重金属土壤浓度值刻画重金属污染, 会有以下缺点: (1) 不同元素之间量级相差大, 无法完成合成分析; (2) 原始浓度值没有区分自然影响因素和人为影响因素。为了避免上述缺点, 常常运用污染指数评价的方法来评估土壤的重金属污染程度<sup>[1]</sup>。在分析了几种指数的优缺点后, 选取地累积指数作为本文的评价指数:

$$I_{geo,i} = \log_2 [C_i / (A \times B_i)], \quad (1)$$

其中,  $I_{geo,i}$  为第  $i$  种重金属元素的地累积指数,  $C_i$  是第  $i$  种重金属元素的实测浓度值,  $B_i$  为当地无污染地区第  $i$  种重金属元素的环境背景,  $A$  为考虑各地沉积岩差异可能引起的背景值变动系数。根据  $I_{geo}$  的值, 可以规定相应的污染程度级别。

收稿日期: 2011-12-02

通讯作者: 曹春正, E-mail: caochunzheng@163.com

接下来利用 ArcGIS 软件进行克里金插值处理,画出重金属离子的地累积指数分布图和等级分布图,然后对各个金属离子逐一观测,找到一些污染关键区域并结合位置和属性特征进行进一步分析,同时按污染形态和强度将 8 种离子分为 4 类。

为了评定不同功能区的重金属综合污染程度,采用负荷指数:

$$\tilde{C}_i = \frac{C_i}{C_{i0}}, PLI = \sqrt[n]{\prod_{i=1}^n \tilde{C}_i}, PLI_{zone} = \sqrt[m]{\prod_{j=1}^m PLI_j}, \quad (2)$$

其中  $PLI_{zone}$  为区域的污染负荷指数,  $m$  为采样点个数,  $PLI$  为单个点的污染负荷指数,  $n$  为金属污染物种数,  $\tilde{C}_i$  为第  $i$  种金属的最高污染物指数,  $C_i$  为第  $i$  种金属浓度的实测值,  $C_{i0}$  为第  $i$  种金属的背景值。

通过定量计算,得出五大功能区的污染程度:工业区 > 主干道区 > 生活区 > 公园绿地区 > 山区。这一结果符合实际,工业区污染最强,其余四个功能区的污染程度都为中等污染,但是山区更接近于无污染,而主干道更接近于强污染。

## 2.2 重金属污染物的来源

假定同一类金属离子具有相同的来源和相同的扩散模式,对上述分类做进一步分析。为了验证分类的合理性,利用 Pearson 相关分析分别对 8 种金属元素的地累积指数和标准化浓度进行分析,可以得到大致相同的结论,即 Ni 和 Cr 高度相关, Pb、Zn、Cd、Cu 的相关关系比较显著, As 和 Hg 相关性较弱。此外我们还对 8 种重金属离子做了主成分分析,也得到了类似的离子聚类结论。以上不同分析方法的相互印证,充分说明了所建分类的合理性。

根据国家发布的中国土壤重金属的主要来源表,结合上述的分类和功能区的分布,给出 Pb 的来源主要有工业和尾气; Cu 的来源主要有工业和农药; Ni、Cr、Cd 和 Zn 主要来自于工业; As 主要来自于矿业开采、工业和农药; Hg 主要来自于工业废水和生活垃圾。

## 2.3 建立区域曲面拟合模型确定污染源位置

设有一污染源,沿  $x$ 、 $y$ 、 $z$  三个方向扩散,扩散系数分别为  $a^2$ 、 $b^2$ 、 $c^2$ , 设  $u(x, y, z, t)$  是  $t$  时刻点  $(x, y, z)$  处某物质的浓度。任取一个闭曲面  $S$ , 其围成的区域为  $\Omega$ , 由扩散理论中的费克第一定律, 在  $\Delta t$  时间内, 通过  $S$  流入  $\Omega$  的质量为

$$M_1 = \int_t^{t+\Delta t} \oint_S (a^2 \frac{\partial u}{\partial x} \cos \alpha + b^2 \frac{\partial u}{\partial y} \cos \beta + c^2 \frac{\partial u}{\partial z} \cos \gamma) dS dt, \quad (3)$$

由于衰减,  $\Delta t$  时间内  $\Omega$  中的质量减少为

$$M_2 = \int_t^{t+\Delta t} \iiint_{\Omega} k^2 u dV dt, \quad (4)$$

其中  $k^2$  为衰减系数。  $\Delta t$  时间内  $\Omega$  中由浓度变化引起的质量增加为

$$M_3 = \iiint_{\Omega} [u(t + \Delta t) - u(t)] dV = \int_t^{t+\Delta t} \iiint_{\Omega} \frac{\partial u}{\partial t} dV dt. \quad (5)$$

由扩散的连续性, 显然  $M_3 = M_1 - M_2$ , 于是衰减扩散过程的数学模型为

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2} + b^2 \frac{\partial^2 u}{\partial y^2} + c^2 \frac{\partial^2 u}{\partial z^2} - k^2 u. \quad (6)$$

设扩散源在点  $(x_0, y_0, z_0)$  处, 则此扩散问题满足 Cauchy 问题。用傅里叶变换可求得此 Cauchy 问题的解析解为

$$u(x, y, z, t) = \frac{M}{8\pi abc\sqrt{\pi t}} \exp \left\{ -\frac{(x-x_0)^2}{4a^2 t} - \frac{(y-y_0)^2}{4b^2 t} - \frac{(z-z_0)^2}{4c^2 t} - k^2 t \right\}, \quad (7)$$

其中  $M$  为扩散源的质量。

由于在无限大介质中扩散浓度集中在点源, 这类扩散问题都具有高斯分布的形式, 即“高斯解”。由此受到启发, 假设污染扩散浓度模式为多元的正态分布曲面。

多元正态分布是一元正态分布的直接推广<sup>[2]</sup>。设随机向量  $\mathbf{X} = (x_1, x_2, \dots, x_p)^T$  服从  $p$  维正态分布, 则其密度函数为

$$f(\mathbf{X}) = (2\pi)^{-p/2} \left| \sum \right|^{-1/2} \exp\left\{- (\mathbf{X} - \mu)^T \sum^{-1} (\mathbf{X} - \mu) / 2\right\}, \quad (8)$$

或可简化地表示为

$$g(u) = k \cdot \exp(-u/2), \quad (9)$$

其中,  $k$  为正则化因子,  $u$  为马氏距离。

设第  $i$  个观测点的污染物浓度指数为

$$I_i = I(x_i, y_i, h_i), i = 1, 2, \dots, n, \quad (10)$$

其中,  $x_i, y_i$  是观测点的水平位置,  $h_i$  是观测点的海拔高度。将位置函数  $f$  与海拔高度  $h$  做变量分离后得到

$$I_i = \tilde{I}(x_i, y_i) J(h_i) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \quad (11)$$

其中,  $\varepsilon_i$  是扰动误差,  $\tilde{I}(x_i, y_i)$  是浓度随水平位置的变化函数,  $J(h_i)$  是浓度随高度扩散函数。上式表明第  $i$  个观测点的污染物浓度可以近似等效为由水平扩散和垂直扩散的叠加作用造成的<sup>[3-5]</sup>。

根据比尔—伯格—朗伯定律, 假设重金属污染扩散随高度呈指数衰减, 于是用指数函数对污染传播高度因子函数进行估计<sup>[6]</sup>, 得到

$$J(h) = e^{-0.006h}, \quad (12)$$

拟合  $R$  方值达到了 0.983, 从而有充分的理由认为指数衰减能够较好地描述污染传播特征与地形高度的关系。用公式(12)去给之前的地累积指数做高度订正, 可将原始样本数据  $(x_i, y_i, h_i, I_i)$  换算成新的样本  $(x_i, y_i, \tilde{I}_i)$ , 为接下来的位置估计降维。

当只有一个污染源时,  $(x_i, y_i)$  处的污染浓度为

$$\tilde{I}_i = k \cdot \exp(-u_i/2). \quad (13)$$

当有  $N$  个污染源时,  $(x_i, y_i)$  处的浓度是  $N$  个污染源在此处污染效果的叠加, 即

$$\begin{cases} \tilde{I}_i = \sum_{j=1}^N k_j \exp(-u_{ij}/2) \\ u_{ij} = \begin{pmatrix} x_i - x_0^j \\ y_i - y_0^j \end{pmatrix}^T \sum_j^{-1} \begin{pmatrix} x_i - x_0^j \\ y_i - y_0^j \end{pmatrix} = (\mu_i - \mu_0^j)^T \sum_j^{-1} (\mu_i - \mu_0^j) \end{cases}, \quad (14)$$

其中,  $k_j, \sum_j, (x_0^j, y_0^j) (j = 1, 2, \dots, N)$  都是待估计的参数。基于极大似然估计, 计算出位置参数  $(x_0^j, y_0^j) (j = 1, 2, \dots, N)$ , 即为污染源位置。这里污染源的个数  $N$  依据 AIC 模型选择标准或交叉验证法选取<sup>[7]</sup>。

### 3 模型分析

基于空间数据的区域曲面拟合模型有如下优点:

(1) 基于重金属浓度的地累积指标, 而不是直接使用测到的浓度去分析, 消除了当地环境背景中金属元素的分布影响和不同金属污染量级差异;

(2) 由于给出的测量值是包含高程值的空间数据, 我们在对观测点污染物指标进行扩散分析时, 利用比尔吸收定理将含有高程信息的样本订正到平面上, 降低了模型维数;

(3) 结合污染源扩散污染物浓度的椭球型分布规律, 由分布的位置参数确定污染源位置, 并且可以处理多个污染源的污染叠加效应。

如果能获取到更多的样本量和更丰富的地理信息, 本文的重金属污染浓度指标分布的精度可以进一步提高; 如果能采集不同时刻点的样本浓度, 可以模拟污染物的扩散过程。

### 4 思路总结和体会

本文讨论的重金属污染状况推断属于空间数据分析研究范畴。以下给出我们的研究思路和体会。

#### 4.1 盲人摸象

在对问题不是很了解的情况下, 要大胆地去尝试, 试着理出头绪。

提供了分析的信息, 包括样本点的位置参数(水平和高程)、功能类型以及各种重金属离子的浓度分布, 很复杂, 开始并没有一个明确的模型。对此类问题, 挖掘数据的主要特征, 并据此选择合适的模型是我们解

题的经验。首先把整个区域的样本点分布和重金属离子浓度分布用图像表示,观察具体形式。得到直观的地形分布和重金属离子浓度分布的信息,为后续分析做好准备。

#### 4.2 找准标杆

找准所研究问题的背景知识和重要的指标性信息,对于问题的分析十分重要。

分析污染离子的分布以及不同区域重金属的污染程度时,为定量地刻画污染程度,需要消除背景土壤中的离子浓度的影响,并且要对不同金属元素建立统一的比较准则。因此,在分析分布情况的时候,并不能直接用源数据去处理。我们在比较了几个指数后,最终选择了用地累积指数去分析金属离子污染分布。该指数不仅考虑了人为污染因素、环境地球化学背景值,它还考虑到了自然成岩作用可能会引起的背景值变动。在刻画不同区域重金属污染程度时,我们选择了污染负荷指数以进行定量比较。该指数的优点是可以消除或降低不同金属离子、不同区域和不同环境的影响。

#### 4.3 顺藤摸瓜

对已经完成的工作,做到最大化的利用,总结线索,得到建模思路。

分析重金属污染离子的来源,直观来看,每种金属离子都不一样,是不是每种都要分析呢?我们根据已得出的各个重金属离子污染分布图和等级分布图,将重金属离子分为四类,然后用 Pearson 相关性分析和主成分分析对上述分类加以统计验证,结果合理。

在寻找污染源位置时,结合之前的分析可知山区并未达到污染水平,属无污染区,因此,在选择污染源的搜索范围时,只需考虑左侧区域,使高度订正方法得以适用的同时降低了工作量。

#### 4.4 环环相扣

文章要联系紧密,上下对应,结构完整,有理有据。

问题是层层递进、逐步加深的,所以模型及方法要做到上下一致,环环相扣。在分析污染程度时定义了地累积指数,并用它去进行了分析,那么之后对于污染源的位置推算中所用的数据都应该基于该指数来实现。

对寻源模型求解的结果分析时,根据进行过高度订正之后的重金属地累积指数分布拟合图,结合之前已经完成的图像,来分析结果的共同点和不同点。

#### 4.5 奇思妙想

加入自己的一些合理性猜测,往往会得到意外的惊喜。

对于寻源问题,首先想到污染源的扩散方程,在三维的情况下获取所建方程的解析解,但由于题目中只给出了无时间节点信息重金属离子浓度,这对扩散方程的求解带来难度,也会因“维数灾难”使得后续的参数估计置信度降低。考虑到在无限大介质中扩散浓度集中在点源,这类扩散问题都具有高斯分布的形式,即“高斯解”。据此,可假设污染扩散浓度曲面是二元的正态分布,大大降低了点源分析的难度。

#### 4.6 专业结合

联系所学专业知识,灵活应用,发挥特长。

天气学专业知识在建模时发挥了重要作用。根据连续方程结合数理方程的扩散问题,分析了扩散方程的解析解的形式。扩充模型中,结合大气环境中一些大气污染物受盛行风向、城市热岛效应、卡门涡街等因素作用,进一步联想到河流、道路等地理信息分布对于污染源扩散的影响。

寻源模型的困难之处在于所给的信息是空间数据,由于样本容量的限制,高度参数必须进行处理。联系气象中温度随高度递减变化的订正方法,但重金属离子浓度的衰减很明显不符合线性衰减的特征,结合辐射的比尔吸收定理,我们采用指数衰减模式去订正高程对浓度的影响,达到了非常好的效果。

#### 参考文献

- [1] 柴世伟. 地累积指数法在土壤重金属污染评价中的应用[J]. 同济大学学报:自然科学版, 2006, 34(12): 1657-1661.
- [2] 李卫东. 应用多元统计分析[M]. 北京: 北京大学出版社, 2008.
- [3] 林艳. 基于地统计学与 GIS 的土壤重金属污染评价与预测[D]. 长沙: 中南大学, 2009.
- [4] 徐理超. 阜新市农田土壤重金属污染的空间分析及污染评价[D]. 重庆: 西南大学, 2007.

- [5] 李月清. 选煤厂产尘环节粉尘颗粒扩散方程的建立[J]. 北京工业职业技术学院学报, 2008, 7(1): 84-89.
- [6] 陈杰. MATLAB 宝典[M]. 3 版. 北京: 电子工业出版社, 2011.
- [7] 滕葳. 重金属污染对农产品的危害与风险评估[M]. 北京: 化学工业出版社, 2010.

## Analysis of Heavy Metal Pollution in Urban Topsoil

Zhou Minqiang<sup>1</sup>, Sheng Zhenfeng<sup>1</sup>, Huang Si<sup>1</sup>

Advisor: Cao Chunzheng<sup>2</sup>

(1. School of Atmospheric Physics, Nanjing University of Information Science & Technology, Nanjing Jiangsu 210044, China;  
2. School of Math & Statistics, Nanjing University of Information Science & Technology, Nanjing Jiangsu 210044, China)

**Abstract:** Heavy metal pollution (HMP) is one of the main indicators of environmental pollution. In this paper, we focus on the inference of a city's HMP status based on a real data set. Firstly, the extent of HMP is assessed by the geoaccumulation index (GI). Then the spatial distribution feature of each metal's GI is analyzed by Kriging interpolation via ArcGIS. Thus, a classification of heavy metal ions is concluded, and the pollution extents of different functional areas are ranked by the pollution load index (PLI). Secondly, the Pearson correlation analysis and the principal component analysis are used to examine the rationality of the classification. Combining the national standard for the classification of HMP's source and the distributed characteristics of the pollution, the pollution source is inferred for each HMP. Finally, following the revision of the sample's elevation information through the exponential decay model, the pollutant concentration surfaces formed by multi-sources are fitted by the density of weighted mixtures of bivariate normal distributions. The estimates of the location parameters can be regarded as the HMP's sources.

**Key words:** heavy metal; pollution analysis; geostatistical analysis; spatial data; geoaccumulation index; mixtures of normal distributions

### 作者简介

曹春正, 男, 汉, 江苏人, 副教授, 研究方向: 统计建模和统计诊断。