



## 城市表层土壤化学元素污染传播的核回归模型

谢中华, 夏国坤

(天津科技大学理学院, 天津 300457)

**摘要:** 以某城市城区主要功能区表层土壤 0~10 cm 中 As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 8 种化学元素在采样点处的含量和背景值为基础, 利用插值方法, 给出主要观测区域内 8 种主要化学元素的空间分布; 并通过单因素方差分析及聚类分析探究污染原因, 建立了城市表层土壤化学元素污染传播的核回归模型. 利用该模型对化学元素的污染趋势进行模拟仿真, 形象化展示了化学污染传播的动态过程.

**关键词:** 表层土壤; 化学元素; 污染; 核回归; 仿真

**中图分类号:** X53      **文献标志码:** A      **文章编号:** 1672-6510(2012)06-0074-05

## Kernel Regression Model of the Spread of Chemical Elements Pollution in Urban Surface Soil

XIE Zhonghua, XIA Guokun

(College of Science, Tianjin University of Science & Technology, Tianjin 300457, China)

**Abstract:** Based on the soil values and the concentration of chemical elements of As, Cd, Cr, Cu, Hg, Ni, Pb and Zn in about 0~10 cm surface soil from some sample points of a certain city, spatial distribution of the 8 major chemical elements in the observed region were discovered with the interpolation method. Causes of contamination were analyzed with both single and cluster factor variance analyses, and a kernel regression model of the spread of chemical elements pollution in urban surface soil was established. By means of this model, chemical elements pollution trends were simulated and dynamic process of the spread of chemical elements pollution was demonstrated.

**Key words:** surface soil; chemical elements; pollution; kernel regression; simulation

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渐明显, 由于人类活动造成的城市表层土壤化学元素污染问题更是尤为突出. 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, 研究人类活动影响下城市地质环境的演变模式, 日益成为人们关注的焦点<sup>[1-4]</sup>. 本文将建立核回归模型, 对城市表层土壤化学元素污染趋势进行分析.

### 1 数据来源

2011 年高教社杯全国大学生数学建模竞赛 A 题

中给出了某城市城区土壤地质环境调查数据, 包括采样点的位置、海拔高度及其所属功能区等信息数据, 以及 8 种主要化学元素 (As、Cd、Cr、Cu、Hg、Ni、Pb、Zn) 在采样点处的含量及背景值数据. 具体调查方式: 按照功能划分, 将城区分为生活区、工业区、山区、主干道区及公园绿地区, 分别记为 1 类区、2 类区、……、5 类区, 不同的区域环境受人类活动影响的程度不同. 将所考察的城区划分为间距 1 km 左右的网格子区域, 按照每平方千米 1 个采样点对表层土 (0~10 cm 深度) 进行取样、编号, 并用 GPS 记录采样点的位置. 应用专门仪器测试分析, 获得了每个样本所含的 8 种主要化学元素的含量数据. 另一方面,

收稿日期: 2012-04-05; 修回日期: 2012-05-07

基金项目: 天津市应用基础及前沿技术研究计划重点资助项目 (08JCZDJ15000)

作者简介: 谢中华 (1978—), 男, 河南人, 副教授, xiezhh@tust.edu.cn.

按照 2 km 的间距在那些远离人群及工业活动的自然区取样, 将其作为该城区表层土壤中元素的背景值.

## 2 主要化学元素的空间分布

利用调查数据中给出的采样点坐标和 8 种主要

化学元素的含量数据, 可以绘制出主要化学元素的空间分布图. 它是一种四维图, 采样点的  $x, y, z$  坐标为前三维, 采样点处化学元素的含量为第四维. 为了直观, 用颜色表示第四维, 利用插值方法, 绘制 8 种主要化学元素的空间分布图, 如图 1 所示. 将空间分布图投影到  $xoy$  平面, 得到 8 种主要化学元素的平面分布图, 如图 2 所示.

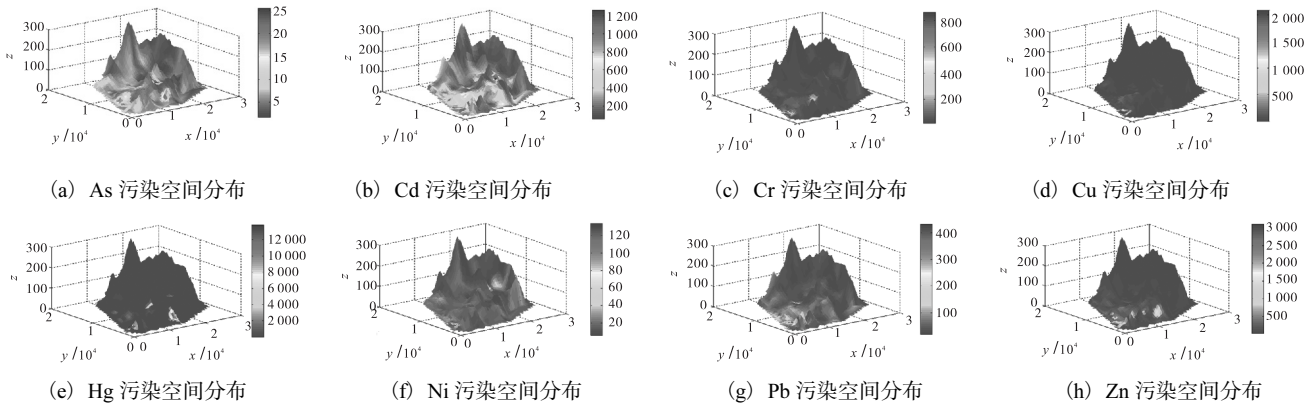


图 1 8种主要化学元素的空间分布图

Fig. 1 Spatial distribution of eight major chemical elements

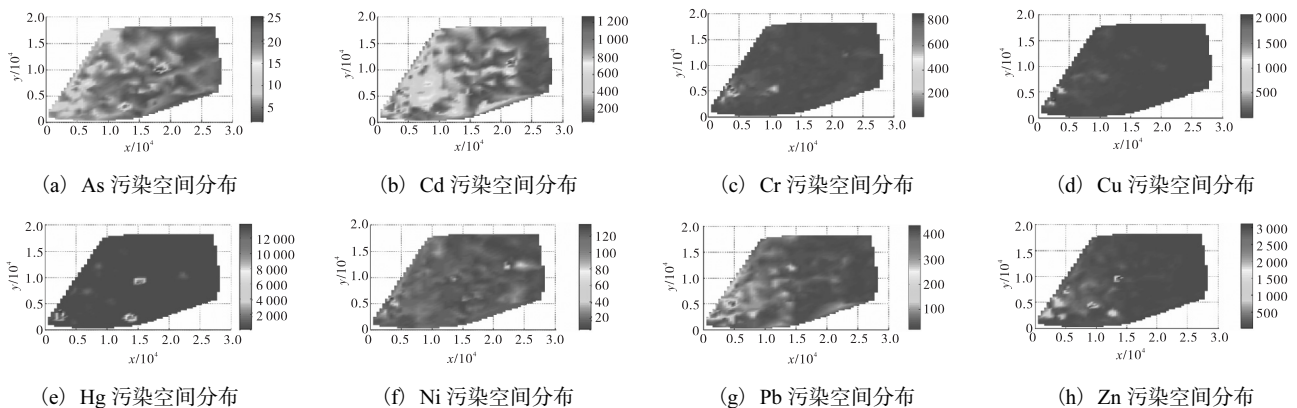


图 2 8种主要化学元素的平面分布图

Fig. 2 Plane distribution of eight major chemical elements

从 8 种主要化学元素的空间分布图和平面分布图可以看出, As、Cd、Ni 和 Pb 污染的扩散性比较强, Zn 和 Cr 污染的扩散性次之, Hg 和 Cu 污染的扩散性较弱. 每种元素对应的污染较为严重区域的位置和数量也是一目了然的. 图 2 表明: Cd 和 Zn 均有多处污染较为严重区域, As 和 Hg 均有 3 处, Cu 和 Pb 均有 2 处, Cr 和 Ni 均有一处.

## 3 污染原因分析

8 种主要化学元素的空间分布是不尽相同的, 它们的污染原因也可能是不同的, 下面研究化学元素污

染主要是由什么造成的.

### 3.1 方差分析

弄清楚该城区不同功能区域化学元素的含量是否有显著区别, 将有助于确定各种主要化学元素的污染原因. 为此, 对 8 种主要化学元素含量的观测数据进行单因素方差分析, 结果见表 1 和表 2.

表 1 方差分析的 P 值

Tab. 1 P-value of variance analysis

元素	检验的 P 值	元素	检验的 P 值
As	$P_1 = 9.7391 \times 10^{-7} < 0.05$	Hg	$P_5 = 0.2368 > 0.05$
Cd	$P_2 = 1.0620 \times 10^{-9} < 0.05$	Ni	$P_6 = 0.1567 > 0.05$
Cr	$P_3 = 0.1784 > 0.05$	Pb	$P_7 = 1.0242 \times 10^{-6} < 0.05$
Cu	$P_4 = 0.0180 < 0.05$	Zn	$P_8 = 0.0054 < 0.05$

方差分析结果表明在显著性水平  $\alpha = 0.05$  下, 各区域污染程度差异显著的化学元素有: As、Cd、Cu、Pb 和 Zn, 从各区域化学元素含量平均值还可以看出 8 种主要化学元素污染的主要原因. 各元素污染程度较重区域基本都在工业区, 可以认为工业污染是主因. 主干道路区的 Cd、Cr、Hg 和 Zn 元素含量均比较

高, 道路交通(例如汽车尾气)造成的污染也是不容忽视的. 生活区的 Cr 含量比较高, As、Cd、Cu、Ni、Pb、Zn 含量处于中等水平, 因此生活垃圾造成的污染也是值得重视的. 总体来看, 山区和绿地区的污染程度是比较轻的.

表 2 各区域主要化学元素含量平均值

Tab. 2 Regional average concentrations of major chemical elements

( $\mu\text{g}\cdot\text{g}^{-1}$ )

区域	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
生活区	6.27	289.96	69.02	49.40	93.04	18.34	69.11	237.01
工业区	7.25	393.11	53.41	127.54	642.36	19.81	93.04	277.93
山区	4.04	152.32	38.96	17.32	40.96	15.45	36.56	73.29
主干道路区	5.71	360.01	58.05	62.21	446.82	17.62	63.53	242.85
公园绿地	6.26	280.54	43.64	30.19	114.99	15.29	60.71	154.24

### 3.2 聚类分析

各观测点的污染程度是否有一定相似性? 具体哪些观测点的污染程度是比较重的? 这些也是需要进一步研究的问题. 下面对 8 种主要化学元素含量的观测数据进行聚类分析, 聚类树形图如图 3 所示. 将聚类树形图从虚线位置截断, 全部的 319 个观测点被聚为 4 类. 其中第 22、36 和 61 号观测点聚为一类, 类序号为 2; 第 8 号观测点单独为一类, 类序号为 3; 第 9、182 和 257 号观测点聚为一类, 类序号为 4; 其他观测点聚为一类, 类序号为 1.

各类中每种化学元素平均含量见表 3. 从表 3 可以看出第 3 类(即第 8 号观测点)是所有观测点中综合污染程度最重的区域, 第 2 类主要是 Zn 污染, 第 4 类主要是 Hg 污染, 第 1 类中的各观测点是综合污染

程度较轻区域.



图 3 各观测点聚类树形图

Fig. 3 Cluster tree of the observation points

表 3 各类别主要化学元素含量平均值

Tab. 3 Average concentrations of major chemical elements in each category

( $\mu\text{g}\cdot\text{g}^{-1}$ )

类序号	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
1	5.67	293.23	49.68	42.59	118.78	16.66	59.65	169.06
2	6.08	637.87	371.74	496.56	116.07	68.92	121.31	2 824.30
3	9.62	1 066.20	285.58	2 528.50	13 500.00	41.70	381.64	1 417.90
4	5.05	665.17	56.43	81.07	14 900.00	20.27	113.51	515.51

## 4 污染趋势分析

### 4.1 核回归模型

化学元素的污染是一个动态的过程, 文献[5-6]中提出了化学元素污染传播的偏微分方程模型, 这些模型的缺点是求解非常困难, 并且本研究中的观测数据没有时间信息, 无法应用偏微分方程模型. 本研究提出核回归模型, 该模型便于求解, 利用该模型可以

对化学元素污染传播的趋势进行模拟仿真.

在污染物传播过程中, 污染物由污染源不断地向周围渗透和扩散, 在污染范围内, 每一点都可以看成是一个小的污染源, 对周围各点会造成一定的影响, 污染范围内任一点处的污染物含量均可表示为该点附近一定范围内的所有点处污染物含量的加权. 污染作用范围示意图如图 4 所示, 选取地形区面上的任意一点  $A(x_0, y_0, z_0)$ , 设此点处第  $k$  种化学元素含量为  $f_k(x_0, y_0, z_0)$ ,  $k = 1, 2, \dots, 8$ , 污染的作用范围为

$r$ . 以  $A$  点为球心, 以  $r$  为半径作球面, 落入此球面的所有点对球心处的污染物含量都有一定的贡献, 并且贡献随距离的增大而减小.

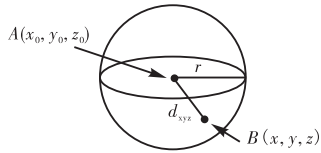


图 4 污染作用范围示意图

Fig. 4 Diagram of pollution scope

设  $B(x, y, z)$  为球面内的任意一点, 它到球心的距离记为  $d_{xyz}$ ,  $B$  点处第  $k$  种化学元素含量记为  $f_k(x, y, z)$ ,  $k = 1, 2, \dots, 8$ , 则

$$f_k(x_0, y_0, z_0) = \sum_{B \in D} K(d_{xyz}/r) f_k(x, y, z) = \iint_D K(d_{xyz}/r) f_k(x, y, z) ds \quad (k = 1, 2, \dots, 8) \quad (1)$$

其中

$$D = \left\{ (x, y, z) \left| \begin{array}{l} (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 \leq r^2 \\ z = g(x, y) \end{array} \right. \right\}$$

$z = g(x, y)$  为地形曲面方程.

式 (1) 中的  $K(\cdot)$  称为核函数,  $K(d_{xyz}/r)$  反映了  $B$  点对  $A$  点处化学元素含量的贡献,  $A, B$  两点的距离越近, 贡献越大 ( $K(d_{xyz}/r)$  越大), 反之贡献越小.

#### 4.2 参数估计

模型 (1) 中的核函数  $K(\cdot)$  可以有多种不同的表示形式<sup>[7]</sup>, 这里选取高斯核函数如下:

$$K(u) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{u^2}{2\sigma^2}} I(|u| \leq 1)$$

其中:  $\sigma$  为高斯核函数的参数,  $I(\cdot)$  为示性函数, 当  $|u| \leq 1$  时,  $I(|u| \leq 1) = 1$ , 否则  $I(|u| \leq 1) = 0$ .

由于化学元素污染的作用范围  $r$  是未知的, 基于高斯核函数的核回归模型 (1) 中含有两个待估参数:  $\sigma$  和  $r$ . 以 As 元素为例, 根据 319 个观测点的 As 含量数据得出参数估计值  $\hat{\sigma} = 2.9943$ ,  $\hat{r} = 369.5809$ , 拟合值  $\hat{y}_i$  与真实值  $y_i$  的平均相对误差为  $ARE = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} = 0.1883$ , 模型拟合效果图如图 5 所示.

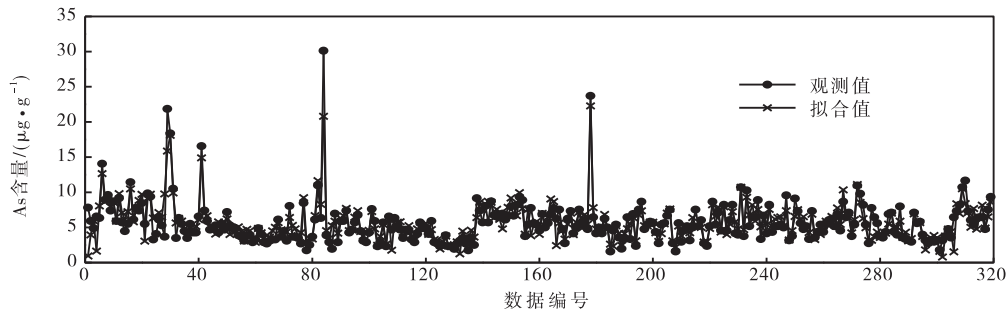


图 5 模型拟合效果图(As)

Fig. 5 Schematic diagram of model fitting(As)

8 种主要化学元素污染传播模型(核回归模型)的拟合结果见表 4.

表 4 核回归模型拟合结果

Tab. 4 Fitting results of kernel regression model

元素	拟合结果		
	$\hat{\sigma}$	$\hat{r}$	ARE
As	2.99	369.58	0.188 3
Cd	2.97	370.26	0.154 7
Cr	2.31	348.42	0.176 4
Cu	1.53	344.33	0.590 2
Hg	1.92	344.33	0.423 2
Ni	2.83	365.68	0.125 9
Pb	2.97	367.49	0.111 9
Zn	2.50	363.14	0.212 4

从表 4 也可以看出, As、Cd、Ni 和 Pb 污染的扩

散性比较强, Zn 和 Cr 污染的扩散性次之, Hg 和 Cu 污染的扩散性较弱.

#### 4.3 污染趋势分析

将现有的观测点  $(x, y, z)$  处第  $k$  种化学元素的含量数据作为  $t = 0$  时刻的数据, 记为  $f_k^0(x, y, z)$ , 由模型 (1) 得到的拟合数据作为  $t = 1$  时刻的数据, 记为  $f_k^1(x, y, z)$ , 将  $f_k^1(x, y, z)$  代入拟合模型得到  $t = 2$  时刻的数据  $f_k^2(x, y, z)$ , 依次类推, 可以得到任意时刻 8 种主要化学元素的含量数据, 从而可以看出每种化学元素的污染趋势.

这里以 As 元素为例, 根据式 (1) 中的核回归模型计算  $t = 0$ 、 $t = 5$ 、 $t = 10$  和  $t = 15$  时刻各观测点的 As 含量, 绘制 As 元素的平面分布图, 如图 6 所示.

从图 6 可以清楚地看出 As 元素在整个观测区域

内的动态扩展过程,在不加以人为控制情况下,As 元素从最初的零星分布迅速扩展到整个观测区域. 正

如核回归模型所描述,污染传播方式是以污染较重区域为中心向周边渗透和扩散,从而造成大面积污染.

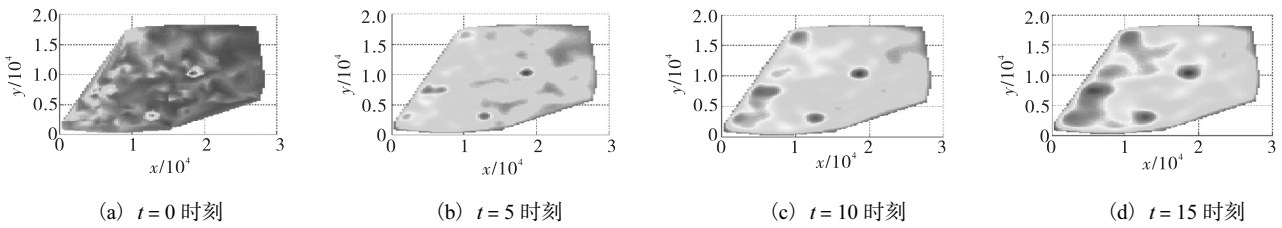


图 6 As 污染趋势图

Fig. 6 Schematic diagram of pollution trends(As)

### 5 结 语

针对 2011 年高教社杯全国大学生数学建模竞赛 A 题中给出的某城市城区土壤地质环境调查数据,本研究对观测区域内 8 种主要化学元素的空间分布和污染原因进行了分析,结果表明各元素污染程度较重区域基本都在工业区,可以认为工业污染是主因. 主干道路区的 Cd、Cr、Hg 和 Zn 元素含量均比较高,道路交通(例如汽车尾气)造成的污染也是不容忽视的. 生活区的 Cr 含量比较高,As、Cd、Cu、Ni、Pb、Zn 含量处于中等水平,因此生活垃圾造成的污染也是值得重视的. 本研究还建立了城市表层土壤化学元素污染传播的核回归模型,并以 As 元素为例,利用该模型对化学元素的污染趋势进行模拟仿真,形象展示了化学元素污染传播的动态过程.

#### 参考文献:

[1] 王海东,方凤满,谢宏芳,等. 芜湖市土壤化学污染

评价及来源分析[J]. 城市环境与城市生态, 2010, 23(4): 36-40.

[2] 郑喜坤,鲁安怀,高翔,等. 土壤中化学污染现状与防治方法[J]. 土壤与环境, 2002, 11(1): 79-84.

[3] 王济,张凌云. 贵阳市表层土壤化学污染元素之间的相关分析[J]. 贵州师范大学学报:自然科学版, 2006, 24(3): 33-36.

[4] 段雪梅,蔡焕兴,巢文军. 南京市表层土壤化学污染特征及污染来源[J]. 环境科学与管理, 2010, 35(10): 31-35.

[5] 李宏艳,王金生,滕彦国,等. 土壤中化学迁移数值仿真与参数灵敏度分析[J]. 系统仿真学报, 2007, 19(4): 720-724.

[6] 隋红建,吴璇,崔岩山. 土壤化学迁移模拟研究的现状与展望[J]. 农业工程学报, 2006, 22(6): 197-200.

[7] 谢中华. MATLAB 统计分析与应用: 40 个案例分析[M]. 北京:北京航空航天大学出版社, 2010: 179-181.

责任编辑:周建军