

# 对于城市表层土壤重金属污染的建模与分析

## 摘要

本文对城市表层土壤重金属污染分析问题进行了建模、求解以及相关分析。

针对问题一，利用插值的方法，首先做出了所讨论区域的地形图，采用相似的方法，得到了 8 种金属污染空间的分布图，然后分别讨论了重金属的平面分布、垂直分布及在不同功能区域的分布。利用附录中的元素浓度和背景值建立了金属污染评价模型，计算得到每种金属在不同功能区的污染等级和综合等级，例如工业区的重金属综合污染负荷指数为 2.41，等级为 II 级（强污染）。

针对问题二，首先对元素浓度进行了标准化处理，构造了特征区域块，建立了因子分析模型。利用 SPSS 软件求解，得到前 5 个主因子，累计贡献率达到 88%，然后通过分析重金属元素（主因子）与污染成因（功能区）的相关系数矩阵，得到造成重金属元素污染的主要原因，例如造成 Cu 污染的主要原因是工业污染。最后，利用聚类的方法对模型进行了检验，结果表明因子分析模型是正确的。

针对问题三，在第二问将 8 种元素分为 5 类的基础上，分传播介质各向同性和各向异性建立了重金属污染物扩散的偏微分方程模型。为了简化模型，通过引入占优主方向和等效速度等概念，分别将两个偏微分方程模型离散化为第一类和第二类掩膜，利用二维卷积原理（空间滤波）确定出 5 种因子的污染物的位置。最后，我们通过讨论污染源的可辨识性对问题进行了进一步分析。

针对问题四，首先对前三问中所建的模型进行了评价，然后在缺少重金属在空气和土壤中的扩散系数、降解系数等信息的情况下，建立了空气和土壤的二房室模型，对城市地质环境的演变模式进行了研究。

最后，根据模型所得结果，从搬迁工业区、增加绿化面积等方面向有关部门提出了治污建议。

关键词      等高线      污染负荷指数      因子分析      偏微分方程      图像处理

# 一、问题重述

## 1.1 问题背景

随着社会经济的发展、科学技术的进步以及人口的迅猛增长，人类在大自然中所处的主导地位越发稳定。从好的方面来说，人类的生活远远比过去舒适，但从不好的角度来看，人类作为主导者，对大自然造成的损害也是巨大的，例如人类活动对城市环境质量的破坏。对城市环境进行考察，并应用查证得到的海量数据进行分析对城市环境质量进行综合评价，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。本题就是这类问题中比较典型的一种——城市表层土壤重金属污染分析。

城市是一个功能极其多样化的地域，粗略地看，城市一般可分为以下 5 类：生活区、工业区、山区、主干道路和公园绿地区，分别记为 1 类区、2 类区、3 类区、4 类区和 5 类区。显然，不同的功能区对于城区而言具有不同的功能，反之，不同的功能区受人类活动影响的程度也不同。

现在要求对某一个城区表层的土壤重金属污染进行调查分析。首先，将所要调查的城区划分成间距为 1 千米的网格状子区域，接着按照每平方公里 1 个采样点的规则，对表层土壤（大约为 0~10cm 深度）进行采样并编号，然后用 GPS 对采样点定位，记录采样点的位置，包括海拔。通过专门的仪器测试分析，进一步获得每个样本点所含的多种化学元素的浓度。此外，还在远离城区和人类活动的地方，按照每两公里 1 个采样点的规则对表层土进行采样，由于人类很少涉足，土壤几乎没有被污染，故可以把其作为城区表层土壤中化学元素的背景值。

## 1.2 需要解决的问题

现有采样点的位置、海拔高度及其所属功能区等信息和 8 种主要重金属元素在采样点处的浓度信息，以及 8 种主要重金属元素的背景值信息。

要求通过数学建模来完成以下任务：

- 1) 给出 8 种主要重金属元素在该城区的空间分布，并分析该城区内不同区域重金属的污染程度。
- 2) 通过对数据进行分析，说明造成重金属污染的主要原因。
- 3) 分析重金属污染物的传播特征，并基于此建立数学模型，确定污染源的位置。
- 4) 分析所建立模型的优缺点。为更好地研究城市地质环境的演变模式，确定应该收集的数据信息。基于这些信息，建立一个可以研究地质环境演变模式的模型。

## 二、问题分析

对于城市表层土壤金属污染分析的问题，本文就此建立了多个模型，求解并进行了详细的分析。

首先对于第一问的土壤重金属元素的空间分布问题，通过插值得到了该城区的地形图，再从不同角度刻画了 8 种金属污染空间的分布图，并对每个角度进行了详细的分析。然后，借助附录中提供的 8 种金属元素的背景值，建立了一个评价重金属污染程度的模型，从而得到了每种金属在不同区域的污染等级和综合等级。

对于第二问分析各个污染物形成的主要原因的问题，在对原始数据进行了标准化之后，构造了子区域，建立了因子分析模型，用软件求解得到主因子的个数，然后再用系统聚类法对 8 种金属元素的浓度原始数据进行因子分析的验证性分析。接着再求得主因子与功能区的相关系数矩阵，作为分析成因的一个依据，经过详尽的分析，最终得到重金属元素污染的主要原因。

对于第三问，在第二问对 8 种金属进行因子分析的基础上，分传播介质各向同性和各向异性建立了重金属污染物扩散模型。为了简化模型，通过引入占优主方向和等效速度等概念，分别将两个偏微分方程模型离散化为第一类和第二类掩膜，利用二维卷积原理（空间滤波）确定出 5 种因子的污染物的位置。最后，我们通过讨论污染源的可辨识性对问题进行了进一步分析。

对于第四问，先是对前三个模型分别进行了优缺点评价，然后，确定了需要的数据，建立了比较完善的城市环境地质演变模型。

## 三、模型假设与符号系统

### 3.1 模型假设

针对本问题，建立如下合理假设：

简化性假设：造成重金属污染的主要原因只考虑工业、交通、燃煤、矿区和人类生活。

概括性假设：因子分析中累计贡献率超过 85%即认为是涵盖了源数据的全部信息量

惟一性假设：每一种重金属元素属于且仅属于一个因子。

合理性假设：气体扩散和液体渗透的差别仅在于弥散系数和对流速度不同。

3.2 符号系统

符号	意义	单位
$C_{ij}$	第 $i$ 个采样点第 $j$ 种元素的浓度	ug/g 或 ng/g
$F_{ij}$	金属 $j$ 在采样点 $i$ 的污染系数金属	\
$R_j$	第 $j$ 种元素的惩罚因子 $F_{ij}$	\
$P_i$	某个采样点的污染负荷指数	\
$I$	某个功能区的污染负荷指数	\
$D$	元素在土壤中的弥散系数	cm/天
$v$	元素在土壤中的对流速度	cm/天

四、问题一的建模与求解

4.1 八种主要重金属元素在该城区的空间分布

由采样所得的数据可以得到八种重金属在该城区的平面分布，海拔分布及在各个不同功能区的分布。

4.1.1 八种重金属的平面分布

由采样点的  $x,y$  坐标和相应采样点的重金属元素的浓度，用 matlab 软件的 `contourf` 函数做出重金属浓度的等高线图。图 1 和图 2 分别显示了 Cd 元素和 Pb 元素浓度平面分布情况：

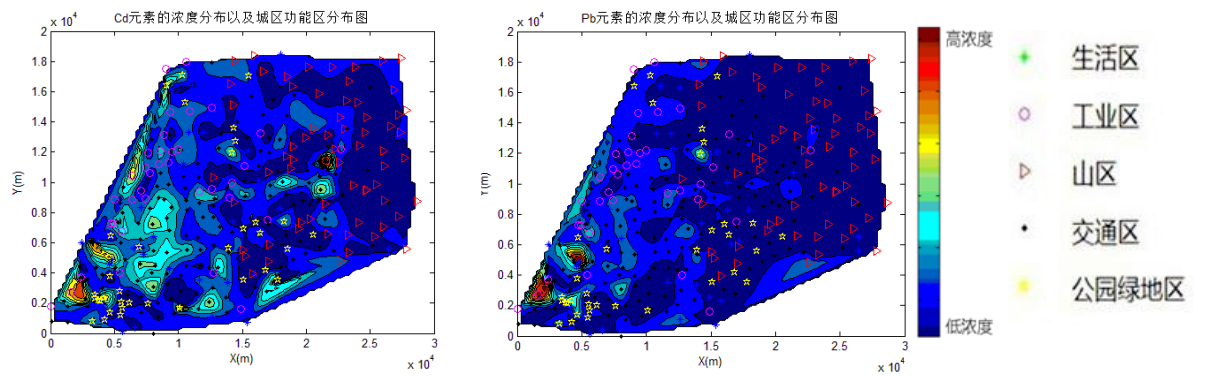


图 1 Cd 元素浓度分布图

图 2 Pb 元素浓度分布图

由图 1 和图 2 可知：

- A. 重金属具有富集的特性，即一种金属在某些较小的范围内浓度比较高，而在其他地区浓度相对较低；
- B. Cd 元素和 Pb 元素都在区域坐标原点附近浓度较高，结合功能区分布可知，此区域为工业区和生活区的交界处；
- C. 元素 Cd 除了分布在原点附近外，还主要分布在工业区，而 Pb 元素基本上全部集中在远点附近

#### 4.1.2 八种重金属的海拔分布

将海拔每 10 米分一段，假设在某段中共有  $n$  个采样点，第  $i$  个采样点的重金属  $j$  的实测含量为  $C_{ij}$ ，则该段中重金属  $j$  的平均含量  $W_j$  为

$$W_j = \frac{1}{m} \sum_{i=1}^m C_{ij} \quad (4.1)$$

根据上式，利用采样所得的数据分别计算得到每段中 8 种重金属的平均含量，结果见表 1：

表 1 不同海拔高度重金属含量表

海拔	As(ug/g)	Cd(ng/g)	Cr(ug/g)	Cu(ug/g)	Hg(ng/g)	Ni(ug/g)	Pb(ug/g)	Zn(ug/g)
0~9	7.1	381.2	86.3	122.8	444.9	21.1	87.8	305.1
10~19	5.8	340.5	48.9	47.6	405.3	17.0	67.3	239.3
20~29	5.7	320.0	50.3	50.5	391.5	17.5	59.4	259.1
30~39	5.7	316.3	57.3	54.6	645.9	16.6	67.8	164.1
40~49	6.6	328.1	43.0	39.5	117.2	17.0	52.3	155.0
50~59	5.1	260.6	41.0	43.5	66.1	14.8	55.2	124.8
60~69	4.4	190.2	38.3	22.0	33.8	13.1	40.0	101.6
70~79	4.7	264.5	55.8	30.9	60.7	22.7	59.8	205.6
80~89	3.8	129.4	37.7	15.5	45.7	15.4	36.0	72.8
90~99	4.7	180.8	31.8	16.3	24.8	13.0	32.4	72.0
≥100	3.4	320.0	35.1	12.9	37.4	13.6	38.1	73.1

由表 1 可知：

- A. 八种重金属元素随海拔分布的大致趋势是：海拔越高，土壤中重金属元素的浓度越低；
- B. 在海拔 70 至 79 米这一段中，所有八种重金属元素的浓度均比附近海拔浓度高。进一步分析数据可知，这一海拔的区域大多为山区，造成元素浓度异常的原因可能是重金属随工厂排出的大气运动的结果；另一个可能的原因是，这一海拔区域建有采煤矿。

#### 4.1.3 八种重金属在不同功能区的分布

假设某区域中共有  $m$  个采样点，第  $i$  个采样点的重金属  $j$  的实测含量为  $C_{ij}$ ，则该区域中重金属  $j$  的平均含量  $W_j$  为

$$W_j = \frac{1}{m} \sum_{i=1}^m C_{ij} \quad (4.2)$$

根据上式，利用 matlab 编程求解分别得到 8 种元素在 5 种不同功能区中平均含量，如表 2 所示：

表 2 各重金属元素在不同功能区的平均含量表

功能区 重金属	生活区	工业区	山区	主干道区	公园绿地区
As(ug/g)	6.3	7.3	4.0	5.7	6.3
Cd(ng/g)	290.0	393.1	152.3	360.0	280.5
Cr(ug/g)	69.0	53.4	39.0	58.1	43.6
Cu(ug/g)	49.4	127.5	17.3	62.2	30.2
Hg(ng/g)	93.0	642.4	41.0	446.8	115.0
Ni(ug/g)	18.3	19.8	15.5	17.6	15.3
Pb(ug/g)	69.1	93.0	36.6	63.5	60.7
Zn(ug/g)	237.0	277.9	73.3	242.9	154.2

由表 2 可知：

- A. 元素 Cr 在生活区的含量最高，浓度为 69ug/g，其余元素均为在工业区的浓度最高，含量远超过各种元素背景值的平均值，其中，元素 Hg 的含量为背景值的平均值的 18 倍；
- B. 元素 Ni 在公园绿地区的含量最低，浓度为 15.3ug/g，其余元素均为在山区的浓度最低，但含量略微高于背景值的平均值；

## 4.2 不同区域重金属的污染程度的评价

### 4.2.1 评价指标的确定和污染等级的划分

考虑到每个采样点测量了 8 种重金属的含量，所以分析每个点的污染情况时需要考虑每种金属相对于背景值的情况，同时分析不同区域的污染情况时，要考虑一个区域内所有采样点的污染情况。

#### 1) 金属 $j$ 在采样点 $i$ 的污染系数 $F_{ij}$

假设在第  $i$  个采样点的重金属  $j$  的实测含量为  $C_{ij}$ ，则金属  $j$  在采样点  $i$  的污染系数  $F_{ij}$  为：

$$F_{ij} = \frac{C_{ij}}{C_{0j}} \quad (4.3)$$

其中， $C_{0j}$  为金属  $j$  在该城区表层土壤中元素的背景值。

## 2) 在第 $i$ 个采样点重金属的污染负荷指数 $P_i$

由于每种金属的背景值是按照 2 公里的间距在那些远离人群及工业活动的自然区取样得到的，存在一定的误差，而且每种重金属元素背景值测量的误差不同，在均值相同的情况下，方差越大，背景值的准确度越低，本文定义背景值的测量准确度  $R_j$  为

$$R_j = \frac{\mu_j}{\sigma_j} \quad (4.4)$$

其中， $\mu_j, \sigma_j$  分别为第  $j$  种金属背景值的测量均值和测量方差。定义惩罚因子（即标准化的背景值的测量准确度）为：

$$R_j' = \frac{R_j}{\sum_{i=1}^m R_i} \quad (4.5)$$

其中， $m$  是采样的元素种数。

为了使背景值的测量误差对污染程度的估计产生较小的影响，所以可以定义污染负荷指数  $P_i$  为：

$$P_i = \prod_{j=1}^m (F_{ij})^{R_j'} \quad (4.6)$$

由  $P_i$  的定义可知，背景值准确度高的元素的污染系数  $F_{ij}$  对采样点污染程度的估计影响更大。

## 3) 一个区域内重金属的污染负荷指数 $I$ <sup>[1]</sup>

$$I = \sqrt[n]{\prod_{i=1}^n P_i} \quad (4.7)$$

## 4) 污染等级划分

由文献[1]可知，土壤重金属污染负荷指数一般分为四个级别，如表 3 所示：

表 3 土壤重金属污染负荷指数级别划分表

污染负荷指数 $I$	<1	1~2	2~3	≥3
污染等级	0	I	II	III
污染程度	无污染	中等污染	强污染	极强污染

## 4.2.2 不同区域污染程度的计算

### (1) 不同功能区域单种元素的污染程度

根据上述区域污染负荷指数的定义，在惩罚因子中取  $m=1$ （即每次只考虑一种元素的污染程度），用 **matlab** 软件编程求解得到生活区、工业区、山区、主干道路区及公园绿地区五个区域中不同元素的污染负荷指数和污染级别，如表 4 所示：

表 4 不同重金属元素在不同功能区的污染负荷指数和污染级别表

功能区 重金属	生活区		工业区		山区		主干道路区		公园绿地区	
As	1.64	I	1.74	I	1.04	I	1.44	I	1.65	I
Cd	1.90	I	2.61	II	1.04	I	2.27	II	1.70	I
Cr	1.66	I	1.47	I	1.10	I	1.53	I	1.34	I
Cu	2.75	II	3.90	III	1.14	I	3.16	III	1.97	I
Hg	1.72	I	3.73	III	0.99	0	2.07	II	1.72	I
Ni	1.42	I	1.47	I	1.09	I	1.32	I	1.19	I
Pb	1.77	I	2.39	II	1.09	I	1.82	I	1.64	I
Zn	2.08	II	2.72	II	0.99	0	2.36	II	1.56	I

由表 4 可知：

- 五种功能区污染最严重的金属均为 Cu，例如生活区 Cu 的污染负荷指数为 2.75，污染级别为 II；
- 山区中元素 Hg 和 Zn 的污染负荷指数皆为 0.99，属于无污染级别；
- 除山区外的其余功能区污染程度最轻的金属皆为 Ni，例如公园绿地区的污染负荷指数为 1.19，污染级别属于 I。

### (2) 不同功能区域重金属元素的综合污染程度

根据上述区域污染负荷指数的定义，在惩罚因子中取  $m=8$ （即综合考虑 8 种元素），用 **matlab** 软件编程求解得到生活区、工业区、山区、主干道路区及公园绿地区五个区域的重金属综合污染负荷指数和污染级别，如表 5 所示：

表 5 不同区域重金属综合污染负荷指数和污染级别表

区域名称	生活区	工业区	山区	主干道路区	公园绿地区
污染负荷指数	1.84	2.41	1.06	1.95	1.59
污染级别	I	II	I	I	I
污染程度	中等污染	强污染	中等污染	中等污染	中等污染

由表可知：

- 五种不同的区域污染程度从重到轻依次为工业区、主干道路区、生活区、公园绿地区、山区；
- 在五种不同的区域中，工业区的污染级别为强污染，其余区域为中等污染；
- 山区的污染负荷指数为 1.06，处于无污染与污染的边界上，这说明山区远离了污染源；主干道路区的污染负荷指数为 1.95，此区域有变为强污染区的危险。



## 五、问题二的建模与求解

### 5.1 问题二的分析

重金属污染的原因与工业活动、人类生活和交通运输等因素密不可分。同一种工业活动可以产生多种重金属元素，同一种重金属元素的来源也不尽相同。为了更好地鉴别该城区土壤中重金属污染的主要原因，采取多种统计分析的因子分析对 8 中重金属元素以及可能的污染源进行统计分析。

### 5.2 数据的预处理

#### (1) 重金属元素浓度标准化

由于所给采样点不同金属元素浓度的单位不同，且数量级有较大差别，所以需要首先对 8 种金属元素浓度进行标准化处理。第  $i$  个观测点第  $j$  种元素的浓度  $C_{ij}$  标准化为：

$$C'_{ij} = \frac{C_{ij} - \bar{C}_j}{\sqrt{S_{ij}}} \quad (5.1)$$

$$\bar{C}_j = \frac{1}{n} \sum_{i=1}^n C_{ij}, \quad \sqrt{S_{ij}} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (C_{ij} - \bar{C}_j)^2}$$

其中， $\bar{C}_j$  和  $\sqrt{S_{ij}}$  分别为元素浓度的平均值和方差。

根据上述标准化方法，利用 matlab 软件的 `zscore` 函数将 8 种元素的浓度标准化。

#### (2) 特征块的构造

构造一个区域  $S_{all}$ ，使其可以包括整个城区，取  $S_{all} = 30000 \times 20000(m^2)$ ，根据已知条件，功能区有五种  $f = 1, 2, 3, 4, 5$ ，其中 1, 2, 3, 4, 5 分别代表生活区、工业区、山区、交通区和公园绿地区，我们称属于上述功能区的取样点为有效点；对于不属于任意功能区 ( $f = 1, 2, 3, 4, 5$ ) 的点，我们将其设为  $f = 0$ ，我们称其为无效点。再构造子区域  $S_{part}$ ，取  $S_{part} = 5000 \times 2000(m^2)$ ，然后在  $S_{all}$  上依次无重叠地取面积大小为  $S_{part}$  的子区域，子区域个数总计  $n_{part} = 60$ 。

统计每一个子区域  $S_{part}$  中有效采样点的个数，依次记为  $n_k (k = 1, 2, 3, 4, 5)$ ，并定义子区域内有效点的总数：

$$n = n_1 + n_2 + n_3 + n_4 + n_5$$

定义  $p_k (k=1,2,3,4,5)$  分别表示每个子区域内分属于生活区、交通区、公园绿地区、山区和工业区的有效点的个数占每个区域内总有效点的比率：

$$p_k = \frac{n_k}{n}$$

不同的功能区会造成不同的重金属污染，在该模型中我们用  $p_k$  表示各个功能区在子区域中对重金属污染做出的“贡献”大小，即每个功能区产生的污染与总污染量之比。

### 5.3 因子分析模型的建立

因子分析<sup>[2]</sup>是利用降维的思想，把一个原始变量分解成两部分，一部分是少数几个公共因子的线性组合，另一部分为该变量所独有的特殊因子。

设  $p$  维总体  $x = (x_1, x_2, \dots, x_p)'$  的均值为  $\mu = (\mu_1, \mu_2, \dots, \mu_p)'$ ，协方差矩阵为  $\Sigma = (\sigma_{ij})_{p \times p}$ ，相关系数矩阵为  $R = (\rho_{ij})_{p \times p}$ 。则因子分析的一般模型为：

$$\begin{cases} x_1 = \mu_1 + a_{11}F_1 + a_{12}F_2 + \dots + a_{1m}F_m + \varepsilon_1 \\ x_2 = \mu_2 + a_{21}F_1 + a_{22}F_2 + \dots + a_{2m}F_m + \varepsilon_2 \\ \vdots \\ x_p = \mu_p + a_{p1}F_1 + a_{p2}F_2 + \dots + a_{pm}F_m + \varepsilon_m \end{cases} \quad (5.2)$$

其中， $F_1, F_2, \dots, F_m$  为  $m$  个公共因子， $\varepsilon_i$  是变量  $x_i (i=1, 2, \dots, p)$  所独有的特殊因子，它们都是不可观测的隐变量。

式 (5.2) 可以写成矩阵形式

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{A}\mathbf{F} + \boldsymbol{\varepsilon}$$

其中， $\mathbf{A} = (a_{ij})_{p \times m}$  为因子载荷矩阵， $\mathbf{F} = (F_1, F_2, \dots, F_m)'$  为公共因子向量， $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m)'$  为特殊因子向量。

### 5.4 模型的求解

由于同一种活动可以产生多种重金属元素，即不同种类的重金属元素之间可能彼此相关，故首先对 8 中重金属污染物进行因子分析。因子分析的关键就在于凭借 8 种重金属元素浓度的相关系数矩阵求出相应的因子的特征值和累计贡献率，利用 SPSS 17.0 软件进行因子分析，结果如表 6 所示：

表 6 特征值和累计贡献率

因子	旋转前			旋转后		
	总的特征值	占总量的百分率/%	累计贡献率/%	总的特征值	占总量的百分率/%	累计贡献率/%
F1	3.560	44.500	44.500	2.059	25.733	25.733

<b>F2</b>	1.150	14.377	58.877	1.767	22.087	47.820
<b>F3</b>	0.965	12.063	70.941	1.217	15.215	63.035
<b>F4</b>	0.768	9.596	80.537	1.026	12.824	75.858
<b>F5</b>	0.578	7.220	87.756	0.952	11.898	87.756

从表 X 中可以看出，当提取 5 个主因子时，累计贡献率达到了 87.756%，即包含了原始数据中 87.756% 的信息量，满足因子分析的原则。此外，可以看到，旋转先后累计贡献率没有发生变化，可见数据量没有丢失。

然后，我们还可以得到旋转前因子载荷矩阵以及旋转后因子载荷矩阵，如下表所示：

表 7 旋转前因子载荷矩阵

指标	F1	F2	F3	F4	F5
As	0.426	-0.200	0.681	0.551	-0.026
Cd	0.711	0.281	0.282	-0.322	-0.254
Cr	0.735	-0.444	-0.303	-0.046	-0.110
Cu	0.756	0.125	-0.365	0.137	-0.155
Hg	0.408	0.673	-0.297	0.449	0.154
Ni	0.723	-0.515	-0.190	0.137	-0.014
Pb	0.764	0.314	0.237	-0.248	-0.158
Zn	0.699	-0.037	0.123	-0.241	0.654

表 8 旋转后因子载荷矩阵

指标	F1	F2	F3	F4	F5
<b>As</b>	0.131	0.154	0.023	0.970	0.076
<b>Cd</b>	0.170	0.877	0.082	0.102	0.132
<b>Cr</b>	0.882	0.209	0.002	0.014	0.146
<b>Cu</b>	0.614	0.362	0.505	-0.022	0.029
<b>Hg</b>	0.013	0.134	0.953	0.030	0.083
<b>Ni</b>	0.864	0.089	0.019	0.222	0.196
<b>Pb</b>	0.195	0.832	0.191	0.121	0.208
<b>Zn</b>	0.271	0.285	0.101	0.091	0.904

从表 X 和表 X 中可以看出，旋转前后，因子载荷的变量结果基本一致。变量与某一个因子的联系系数绝对值（载荷）越大，则该因子与变量关系越近。经过分析可知，Cr 和 Ni 元素都与因子 F1 联系密切，故因子 F1 可以理解为 Cr 和 Ni 的组合，同理，Cd 和 Pb 元素都与因子 F2 联系密切，故因子 F2 可以理解为 Cd 和 Pb 元素的组合；因子 F3 为 Hg 元素，因子 F4 为 As 元素，因子 F5 为 Zn 元素，其中要特别注意到 Cu 元素，Cu 元素与因子 F1 关系密切，同时与因子 F3 也有相当多的联系，但由假设 3 可知 Cu 只能属于因子 F1。

定义每一个有效点的因子浓度为  $f_m$ ，其中  $m=1,2,3,4,5$ ，分别对应主因子

$F1, F2, F3, F4, F5$ 。

定义  $p_f$ ，表征每个子区域内的平均因子浓度：

$$p_f = \frac{f_m}{n_k}$$

其中， $n_k (k=1,2,3,4,5)$  为子区域内有效点的总数。

最后我们就可以建立  $F1, F2, F3, F4, F5$  和  $p_k$  的相关系数矩阵，以确定 5 个主因子分别对应哪一种或是哪些污染成因。相关系数矩阵如表 9 所示：

表 9 重金属元素（主因子）与污染成因（功能区）的相关系数矩阵

相关系数	$p_1$ (生活区)	$p_2$ (工业区)	$p_3$ (山区)	$p_4$ (交通区)	$p_5$ (公园绿地区)
$f_1$	0.207	0.1658	-0.0837	0.2703	-0.2019
$f_2$	-0.0284	0.2164	0.0866	0.318	-0.1495
$f_3$	0.1366	0.0986	0.215	0.2926	-0.0331
$f_4$	0.2539	0.4156	-0.0569	-0.5689	-0.3178
$f_5$	0.1922	0.2121	-0.0542	0.3898	-0.1344

图 3 反映了因子 4 和因子 5 的分布和不同功能区的分布：

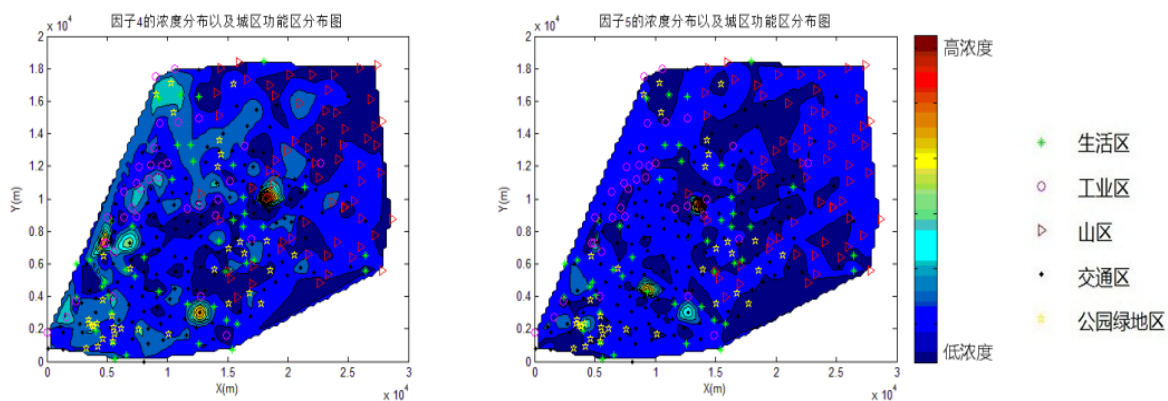


图 3 因子 4 和因子 5 的分布和不同功能区的分布图

通过上述因子分析我们已经确定，因子 1 主要是 Cr 和 Ni 的组合，外加部分 Cu。由相关系数矩阵可知，因子 1 与交通区最相关，同时与生活区和工业区也有一定的相关度。由图可知，因子 1 在该城区的大部分区域浓度都是很低的，主要富集在该城区图的左下角。综上所述我们可以得出因子 1 污染来源于某种工厂。由于 Cr 的污染主要来源于金属加工、电镀、制革、冶金、水泥等工业，Ni 污染的主要来源是不锈钢和抗腐蚀合金，以及镀镍、铸币行业，Cu 污染的原因在于铜锌矿的开采和冶炼、金属加工、机械制造、钢铁生产等，由此我们可以推断，因子 1 污染的主要原因是在城区的左下角有从事金属加工钢铁生产的工厂，属于工业污染。

因子 2 包含了 Cd 和 Pb 元素，由相关系数矩阵可知，因子 2 与交通区的相关度最大，所以我们可以判断因子 2 污染的原因的交通。再从因子浓度分布图中得到，因子 2 浓度较高的地带分布比较广泛，与交通区遍布大部分城区的情况相吻合，又与汽车尾气中含有较多的 Cd 和 Pb 元素这一事实相结合，可以确定因子 2 污染的原因是汽车尾气，属于交通污染。

因子 3 为元素 Hg，由相关系数矩阵可知，Hg 的浓度与山区和交通区相关性较大，与生活区也有一定的相关性。再由因子浓度分布图可得，Hg 主要分布在临近生活区和工业区的区域。可以断定，Hg 的污染主要是由于燃煤造成的，无论是工业用煤还是居民用煤，都会造成地表土受到 Hg 的污染。而且，Hg 的浓度与交通区相关，Hg 又大量存在于汽车尾气当中，所以 Hg 污染也属于交通污染。此外，Hg 的浓度也与山区相关，很可能是由于某山区上存在汞矿。所以，因子 3，即 Hg 污染的原因既有燃煤污染，亦有交通污染，也可能有矿区污染。

因子 4 为元素 As，由相关系数矩阵可知，As 的浓度主要与生活区和工业区相关。从图中亦可看出，As 主要分布在工业区和生活区比较密集的地方。在工业区附近，As 污染主要是经各种工业生产产生，在生活区附近，则有可能是砷农药的使用导致了砷污染，同时，工业区和生活区都可能经煤燃产生 As 污染。所以，因子 4，即 As 污染的原因是工业污染、燃煤污染和生活污染。

因子 5 是元素 Zn。由相关系数矩阵可知，Zn 元素的浓度与生活区、工业区和交通区十分相关。再可以通过图看出，一部分 Zn 分布在工业区，另一些则分布地较无规律。由于 Zn 元素的主要污染源有冶炼加工、机械制造以及镀锌等工业的排放，可以断定，分布在工业区的 Zn 污染属于工业污染。此外，考虑到无规律分布的 Zn 的污染，可能是锌矿的开采导致的。所以，Zn 元素的污染原因既是工业污染，又可能是矿区污染。

另外，注意到 5 个因子与公园绿地区全部都是负相关，其中的 3 个因子与山区也呈负相关，说明了绿化带，例如山区和公园绿地区，对重金属污染物的扩散有抑制作用。

根据上述分析可得到 8 种金属不同的产生原因，结果如表 10 所示：

表 10 造成重金属污染的主要原因表

重金属元素	造成该类重金属污染的主要原因
Cr, Ni, Cu	工业污染
Cd, Pb	交通污染
Hg	燃煤污染，交通污染，矿区污染
As	工业污染，燃煤污染，生活污染
Zn	工业污染，矿区污染

### 5.5 模型的检验

为了进一步验证因子分析的合理性，可以使用聚类分析来验证因子分析的结果。聚类分析是通过使用一种距离相似测度方法不断把性质最为接近的两类合并为一个大类，我们对 8 重金属元素进行聚类分析，得到其系统树：

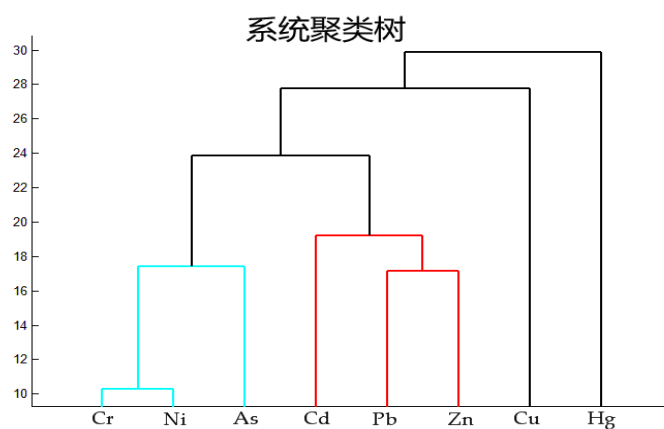


图 4 8 种重金属元素的系统聚类树

由上图可知，根据因子分析所确定的主因子数为 5，故将这 8 种重金属元素聚成 5 类。在系统聚类树上容易看出，Cr、Ni 和 As 元素为一类，Pb 和 Zn 元素为一类，其余元素皆是自成一类。还可以看到，Cd 与包含 Pb 和 Zn 的一类元素相距很近，即相似度较高，说明 Pb 和 Cd 相对而言还是较为相似的。而 As 与包含 Cr 和 Ni 的一类元素相距较远，说明 As 与 Cr 和 Ni 相似度相对不高，也有自成一类的可能。由此可以看到，聚类分析的结果与前面因子分析的分类结果基本一致，因子分析是合理且可靠的。

## 六、问题三的建模与求解

### 6.1 模型三的建立

#### 6.1.1 建模思路

在问题二中我们已经按照重金属污染的主要原因，将重金属污染物分为五类，所以在本问题中我们首先画出五种因子在所考察城区的分布图，以便进行分类别的污染源确定。

这里，传播特征理解为不同重金属污染物在空间上的不同分布模式。这种相对稳定的空间分布模式是由于重金属元素随着时间的推移而不断扩散的结果，故能够反映重金属污染物的传播特征。首先，我们通过建立各向同性介质下的重金属污染物扩散偏微分方程，得出了第一种分布模式——；接着，我们在类比物理学中多普勒效应的基础上，建立各向异性介质下的重金属污染物扩散的偏微分方程，得出了第二种和第三种分布模式——。为了通过分析分布模式找到污染源的具体位置，本文将分布模式离散化为数字图像处理<sup>[3]</sup>中的掩膜形式，然后与原城区矩阵做卷积，找出响应最大的地方，则此时掩膜的中心即为污染源的位置。

#### 6.1.2 重金属传播的基本偏微分方程

根据文献[4]，重金属污染物在土壤中渗透的基本方程为

$$R_d \cdot \frac{\partial C}{\partial t} + v_x \cdot \frac{\partial C}{\partial x} + v_y \cdot \frac{\partial C}{\partial y} + v_z \cdot \frac{\partial C}{\partial z} = D_x \cdot \frac{\partial^2 C}{\partial x^2} + D_y \cdot \frac{\partial^2 C}{\partial y^2} + D_z \cdot \frac{\partial^2 C}{\partial z^2} \quad (6.1)$$

式中， $R_d$ 代表的是时间变化系数， $D_x$ ， $D_y$ ， $D_z$ 分别代表的是 $x$ ， $y$ ， $z$ 方向上的弥散系数， $v_x$ ， $v_y$ ， $v_z$ 分别代表的是 $x$ ， $y$ ， $z$ 方向上的对流速度。

由于气体的扩散和液体的渗透均满足内斯特实验定律，所以上述偏微分方程对于重金属污染物在各向同性空气中扩散的情况也适用。为了简化模型，假设气体扩散和液体渗透的差别仅在于弥散系数和对流速度不同。

#### 6.1.3 各向同性介质下的重金属传播稳态分布模式

各向同性介质下的重金属传播是一个比较普遍的现象，例如重金属在平原均匀土壤中的渗透和重金属在无风空气中的扩散均属于此种情况。为了建立各向同性介质下的重金属传播稳态偏微分方程，我们做出如下简化：

- 1) 由于本问考虑的是稳态情形，故 $\frac{\partial C}{\partial t}$ 可以忽略不计。
- 2) 对偏微分方程中的对流项 $\frac{\partial C}{\partial x}$ ， $\frac{\partial C}{\partial y}$ 和 $\frac{\partial C}{\partial z}$ 进行数据分析及其定性分析，容易

发现这三项对整个等式影响很小，可以忽略不计。

- 3) 由于本问考虑的是重金属污染物的二维分布模式，故不需考虑重金属污染物浓度在  $z$  方向上的变化。
- 4) 由于本问考虑的是各向同性介质下的情形，故  $D_x = D_y$ 。

经过以上简化后，可建立各向同性介质下的重金属传播稳态偏微分方程为

$$\Delta C = 0, 0 < x < \infty, 0 < y < \infty$$

这是一个二维拉普拉斯方程，将其变换到极坐标系（以污染源为极点），得：

$$\frac{\partial^2 C}{\partial r^2} + \frac{1}{r} \cdot \frac{\partial C}{\partial r} = 0, 0 < r < \infty \quad (6.2)$$

易求得其解析解为：

$$C(r) = \frac{Q}{r}$$

其中， $Q$  为待定常数。

将  $C(r)$  离散化为  $50 \times 50$  的矩阵  $M_{50 \times 50}$ 。 $M_{50 \times 50}$  即为反映各向同性介质下的重金属传播稳态分布的第一类掩膜，如下图所示：

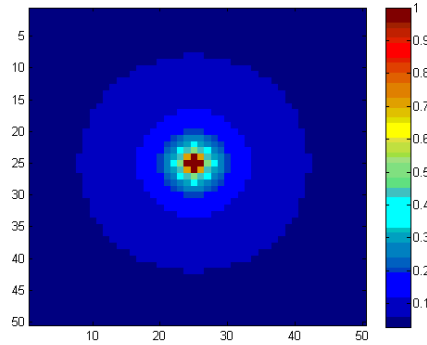


图 5：第一类掩膜

由图可知：第一类掩膜为以污染源为圆心的同心圆形式。

#### 6.1.4 各向异性介质下的重金属传播稳态分布模式

各向异性介质下的重金属传播也是一种不能忽略的情形，例如重金属在存在坡度变化的土壤中的渗透、重金属在有风向的空气中的扩散以及当污染源自身处于运动状态时。在这里，为了给出一个统一的模型，首先，我们将上述三种情况均转化为污染源的运动：

- 1) 引入占优正方向  $A^+$  和等效速度  $\bar{v}$ ：
  - a) 对于重金属在存在坡度变化的土壤中的渗透：



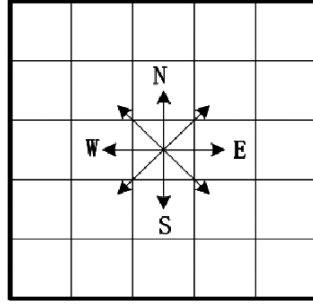


图 6: 8 个子区域划分的示意图

将一块区域按面积均匀划分为 N,NE,E,SE,S,SW,W,NW 这 8 个子区域，设每个子区域内的采样点数为  $z^{(n)}, n=1,2,\dots,8$ ，该区域的占优正方向为  $A^+$ 、占优负方向为  $A^-$ 。与占优方向正交的两个方向分别为  $B^+$  和  $B^-$ 。

这 8 个子区域内任意两个采样点之间的坡度  $s_{ij}^{(n)}$  为

$$s_{ij}^{(n)} = \frac{h_i - h_j}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}, n=1,2,\dots,8 \quad (6.3)$$

每个子区域的平均坡度  $\bar{s}^{(n)}$  和坡度分布方差  $\sigma_s$  分别为

$$\begin{aligned} \bar{s}^{(n)} &= \frac{\sum_{i,j} s_{ij}^{(n)}}{z^{(n)}}, n=1,2,\dots,8 \\ \sigma_s &= \frac{\sum_{n=1}^8 (\bar{s}^{(n)} - \bar{s})^2}{8}, \text{其中 } \bar{s} = \frac{\sum_{n=1}^8 \bar{s}^{(n)}}{8} \end{aligned} \quad (6.4)$$

该区域的占优正、负方向及等效速度为

$$A^+ = \arg \max_n (\bar{s}^{(n)})$$

$$A^- = -A^+$$

$$v = \frac{\bar{s}^{(A^+)}}{\bar{s}^{(A^-)}}$$

则可将此种情况下的稳态分布模式等效为平原均匀土壤中污染源以速度  $\bar{v}$  运动的稳态分布模式。

b) 对于重金属在有风向的空气中的扩散：

假设风速为  $\bar{v}$ ，则这里的占优方向即为风的运动方向。此种情况的稳态分布模式等效为无风时污染源以速度  $\bar{v}$  运动的稳态分布模式。

2) 确定第二类掩膜  $N_{50 \times 50}$ ：

在引入占优方向和等效速度的基础上，通过修正各向同性介质下的重金属传播稳态分布模式，给出各向异性介质下的重金属传播稳态分布模式。

在这里，类比第一类掩膜的同心圆模式和多普勒效应，我们设计出了共焦点的椭圆来模拟各向异性介质下的重金属传播稳态分布模式：

a) 占优正方向  $A^+$ ：

即为椭圆的长轴方向，且远离主焦点的一测。

b) 等效速度  $\bar{v}$ ：

即为椭圆的长轴长与短轴长之比， $\bar{v} = \frac{b}{a}$

将椭圆离散化为  $50 \times 50$  的矩阵  $N_{50 \times 50}$ 。 $N_{50 \times 50}$  即为反映各向异性介质下的重金属传播稳态分布的第二类掩膜，如下图 4 所示：

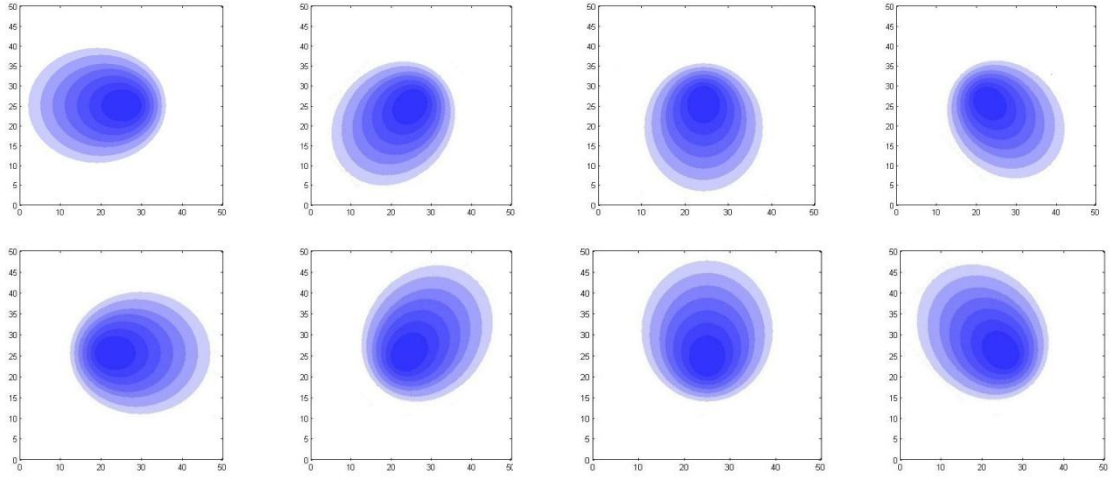


图 7：第二类掩膜（共 8 种）

### 6.2.5 利用空间滤波确定污染源的位置

空间滤波器又成为掩膜，原图像  $f(x, y)$  经过空间滤波器  $h$  的运算，得到滤波后的图像  $g(x, y)$ ：

$$g(x, y) = h(i, j) * f(i, j)$$

这一运算为卷积。卷积的运算模式为：从左向右、从上到下移动滤波器  $h$ ，针对每一次滤波器视窗所涵盖的原图像的区域进行相乘，最后累加所有乘积，得到滤波后的图像上一个像素的值。假设空间滤波器的大小是  $M \times N$ ，则我们可以写成

$$g(i, j) = \sum_{m=-M/2}^{M/2} \sum_{n=-N/2}^{N/2} h(m, n) \cdot f(i-m, j-n)$$

示意图如下：

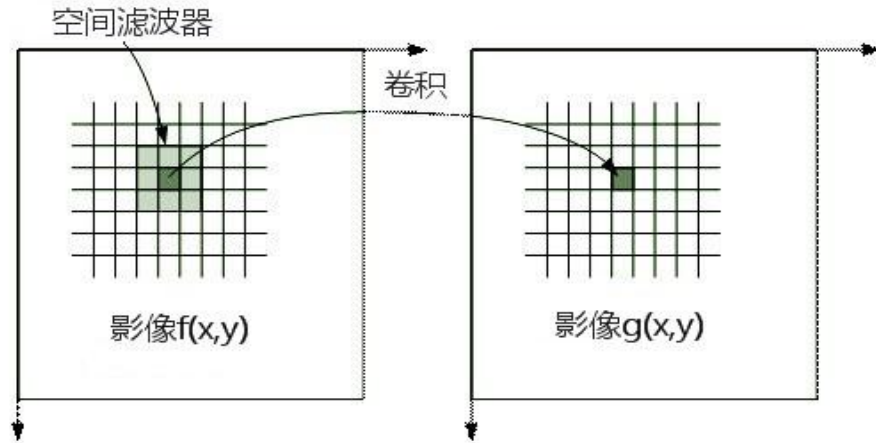


图 8：空间卷积示意图

由卷积的定义易得：若  $f(x, y)$  某处的模式与滤波器  $h$  的模式越接近，则该处的响应  $g(x, y)$  越大。那么，我们可以利用这条有用的性质来确定污染源的位置：

$$\text{污染源的位置 } (x_0, y_0) = \arg \max_{(x, y)} g(x, y) = \arg \max_{(x, y)} h(x, y) * f(x, y)$$

### 6.3 模型三的求解

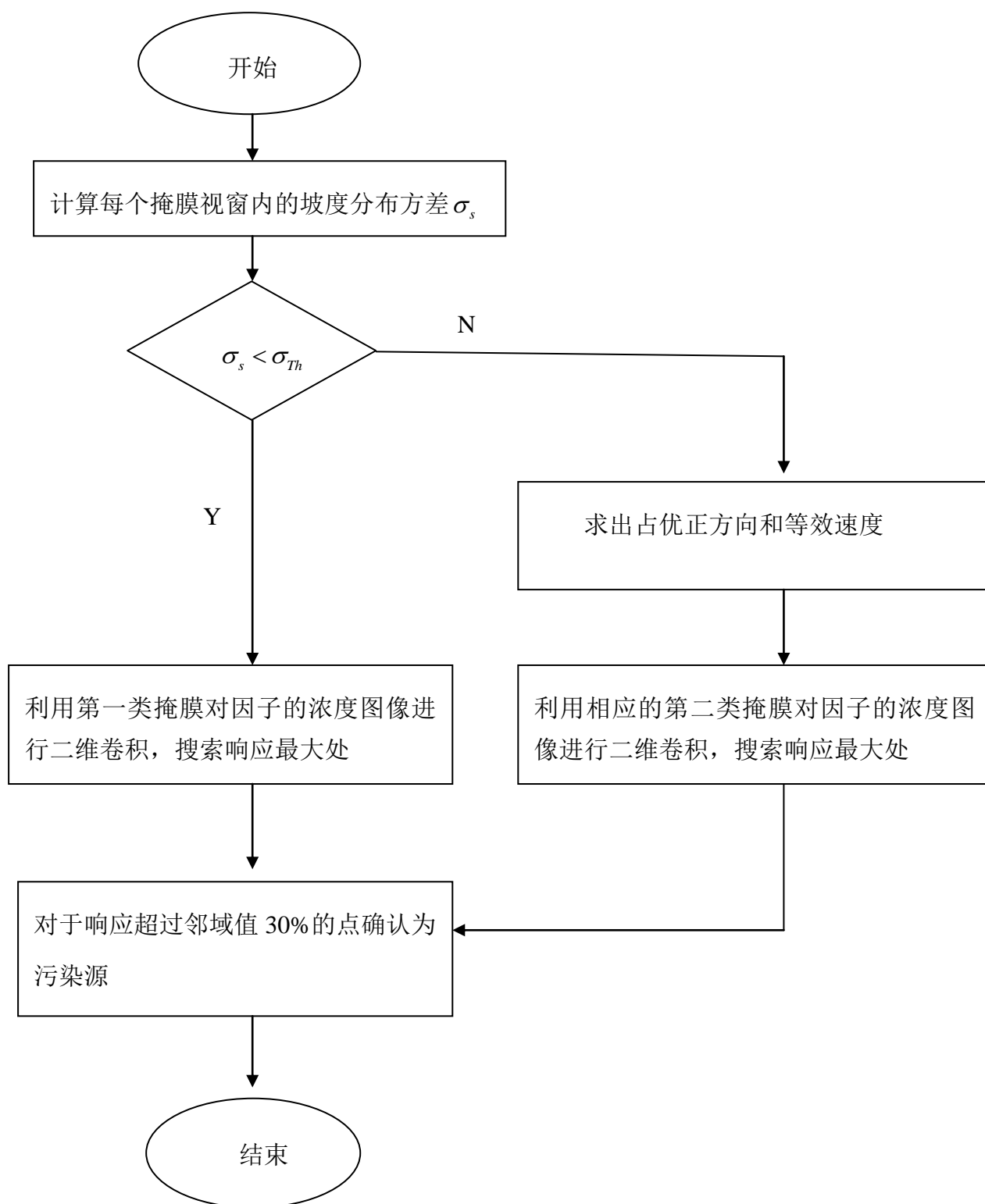
#### 6.3.1 求解的具体步骤

STEP 1：计算每个掩膜视窗内的坡度分布方差  $\sigma_s$

STEP 2：若  $\sigma_s < \sigma_{Th}$ ，则利用第一类掩膜对因子的浓度图像进行二维卷积，搜索响应最大处；若  $\sigma_s \geq \sigma_{Th}$ ，则先求出相应视窗内的占优正方向和等效速度。再利用相应的第二类掩膜对因子的浓度图像进行二维卷积，搜索响应最大处。

STEP 3：对于响应超过邻域值 30% 的点确认为污染源。

其具体流程图如下：



6.3.2 求解结果

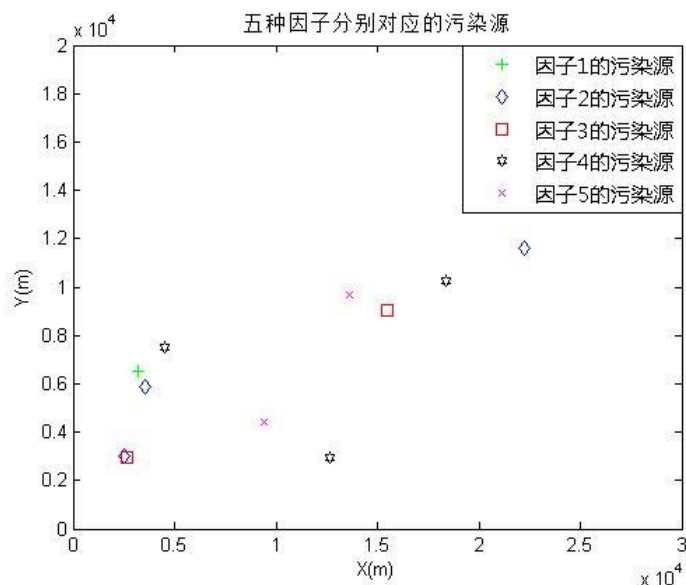


图 9：5 种因子对应的污染源

表 11 五种因子对应的污染源坐标

	污染源
因子 1	(3194,6357)
因子 2	(2520,3008), (3554,5859), (22215,11595)
因子 3	(15449,9019), (13610,24231), (2646,2935)
因子 4	(12653,2945), (18355, 10280), (4490,7540)
因子 5	(13636,9696), (9394,4444)

结论：

- 1 由第二问的结果可知，Pb 和 Cd 污染属于交通污染，所以通过因子 2 的污染源散点图，可以推测，因子的污染源极有可能是交通线枢纽。
- 2 由该城区的海拔图可以得知，该城区的西南方海拔较低且平缓，进一步分析可知，工业往往都会建立在平原上。
- 3 由上图可以看出，大部分污染源都处在该城区图的西南角，可以推测，该城区西南方是工业集中地，基于工业要建立在城市的下风带这一原理，可以推断，西南方是该城区的下风带。

### 6.3.3 利用计算机仿真进行污染源的可辨识性判断

当两个污染源相隔太近时，会由于两个污染源之间的相互影响，而导致本模型对于污染源无法辨识。这里，我们利用计算机仿真，确定在本模型下两污染源可辨识最小距离  $d_{\min}$ 。

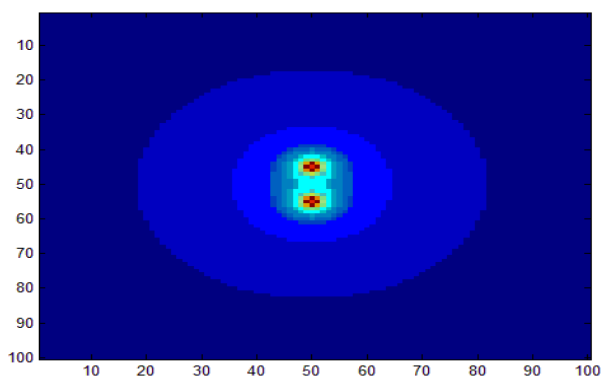


图 10 计算机仿真图

通过计算机多次仿真，得： $d_{\min} = 13m$

## 七、问题四的建模

### 7.1 前三问中模型的优缺点评价

#### 第一个模型

优点：1 通过三次多项式插值得到了重金属元素连续的分布；

2 分别从平面分布，海拔分布，不同功能区分布三个角度建立空间分布模型；

3 定义了多个指数，例如污染系数，污染负荷指数等等，使模型更具有说服力。

缺点：1 数据量不够充足，针对 Cd 和 Pb 元素浓度分布的结论可能会与事实有偏差；

2 定义指数时有一定的主观因素，不一定很准确。

#### 第二个模型

优点：1 统计方法采用因子分析而不是主成分分析，更全面地保留了源数据的信息；

2 利用系统聚类法去验证因子分析的结果的正确性；

3 除因子分析外，还考虑了因子和功能区的相关性，分析更加透彻。

- 缺点：1 假设每种重金属元素只能属于一个因子，不够合理；
- 2 造成重金属污染的因素很多，该模型仅考虑了 5 种，可能不够全面；
- 3 定义的平均因子浓度仅对区域内的所有点的因子浓度进行了普通的求和平均，不一定合理。

第三个模型

- 优点：模型简单，易于实现，且能依据模式较准确地找到污染源；
- 缺点：对于重金属污染物的传播过程过于简化，忽略了其中可能隐藏的特征。

7.2 模型四的建立

7.2.1 建模思路

土壤中重金属的来源是多途径的，首先是成土母质本身含有重金属，不同的母质、成土过程所形成的土壤含有重金属量差异很大。此外，人类工农业生产活动，也造成重金属对大气、水体和土壤的污染，而大气和水体中的重金属又会反过来对土壤的污染造成影响。考虑到水体与土壤的密切相关性，所以我们建立了空气和土壤的二房室模型，对城市地质环境的演变模式进行了研究，并对未来城市地质环境的发展进行了预测。

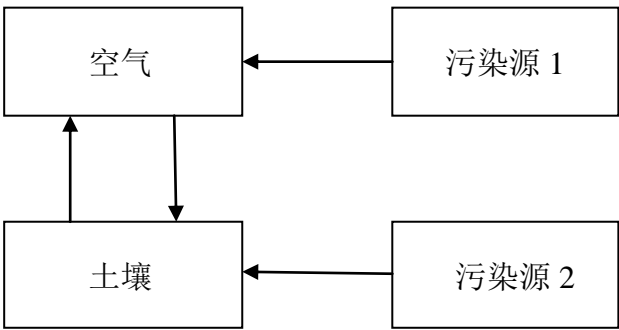


图 11：二房室模型示意图

7.2.2 所需收集的数据

表 12 所需收集数据表

土壤	$\mathbf{k} = (k_x, k_y)$ 为土壤中重金属元素的扩散
----	--

	系数; $\lambda_s$ 为土壤中的降解系数
空气	$\mathbf{v} = (v_x, v_y)$ 为空气速度; $\mathbf{E} = (E_x, E_y)$ 为空气中的扩散系数; $\lambda_a$ 为空气中的降解系数
土壤与空气的交界面	$\kappa$ 为空气中重金属元素的沉降系数
污染源	$f_1(x, y, t)$ 为污染源 1 所释放的重金属元素浓度; $f_2(x, y, t)$ 为污染源 2 所释放的重金属元素浓度

### 7.2.3 城市地质环境的演变模型

#### 1) 土壤中的重金属渗透模型:

$$\begin{cases} \frac{\partial C_s(x, y, t)}{\partial t} + \nabla \mathbf{q} = f_1(x, y, t) - \lambda_s \cdot C_s(x, y, t) \\ \mathbf{q} = -\mathbf{k} \cdot C_s(x, y, t) \cdot \nabla h(x, y) \end{cases} \quad (7.1)$$

其中,  $C_s(x, y, t)$  为土壤中的重金属元素浓度,  $\mathbf{k} = (k_x, k_y)$  为土壤中的扩散系数,  $f_1(x, y, t)$  为污染源 1 所释放的重金属元素浓度,  $\lambda_s$  为土壤中的降解系数,  $h(x, y)$  为该处土壤的海拔分布。

#### 2) 空气中的重金属扩散模型:

$$\frac{\partial C_a(x, y, t)}{\partial t} + \nabla(\mathbf{v} \cdot C_a(x, y, t)) = \mathbf{E} \cdot \nabla^2 C_a(x, y, t) + f_2(x, y, t) - \lambda_a \cdot C_a(x, y, t) \quad (7.2)$$

其中,  $C_a(x, y, t)$  为空气中的重金属元素浓度,  $\mathbf{v} = (v_x, v_y)$  为空气速度,  $\mathbf{E} = (E_x, E_y)$  为空气中的扩散系数,  $f_2(x, y, t)$  为污染源 2 所释放的重金属元素浓度,  $\lambda_a$  为空气中的降解系数。

#### 3) 土壤与空气相互影响的二房室模型:

$$\begin{cases} \frac{\partial C_s(x, y, t)}{\partial t} + \nabla \mathbf{q} = f_1(x, y, t) + \kappa \cdot C_a(x, y, t) - \lambda_s \cdot C_s(x, y, t) \\ \mathbf{q} = -\mathbf{k} \cdot C_s(x, y, t) \cdot \nabla h(x, y) \\ \frac{\partial C_a(x, y, t)}{\partial t} + \nabla(\mathbf{v} \cdot C_a(x, y, t)) = \mathbf{E} \cdot \nabla^2 C_a(x, y, t) + f_2(x, y, t) - \lambda_a \cdot C_a(x, y, t) \end{cases}$$



## 七、对有关部门的建议

自改革开放以来，社会各个方面都有了长足的进步，尤其是进入到 21 世纪以后，我国发展的迅猛之势前所未有，科技的发展使得周围的一切事物不断更新换代，人们生活舒适度也大幅度提高。

但是，不能忽略的是，换取安逸的代价是很惨重的，我们身边的自然一点点被蚕食，环境一点点变得恶劣，就如本文中的城区，由第一问的结果可以看出，大部分区域都处于中等污染的程度，而工业区则是强污染，这严重危害了城区的形象和居民的身体健康。

为此，就该城区而言，向当地政府提出几点建议。

(1) 从功能分区图上可以看出，部分工业区和生活区距离过近，甚至有产生共同区域的情况，这对人们的日常生活和身体健康都是有害的，故建议把工厂全部迁到城区的下风带，或是将生活区全部迁至城区的上风带且远离工业区的的地方。

(2) 由第二问的结果我们知道，绿化带有抑制重金属元素富集或传播的作用，故建议大力开展植树造林。不仅美化了城区形象，还可以平衡重金属污染物的破坏效应。

(3) 由于该城区交通污染较为严重，建议推广电瓶车的使用或倡导居民出行时步行或自行车，尽量减少汽车尾气的排放。

(4) 该城区的燃煤污染也较为严重，建议大力推广天然气或者石油气的使用。

(5) 该城区也存在矿区污染的问题，建议当地政府应做好环境保护措施，并加大检查力度，确保矿区的酸性废水达到可以排放的标准之后再排放。

## 八、参考文献

- [1] 徐争启等，应用污染负荷指数法评价攀枝花地区金沙江水系沉积物中的重金属[J]，四川环境，2004，23 Vol，3 期
- [2] 王雄军，赖健清，等 基于因子分析法研究太原市土壤重金属污染的主要来源 生态环境 2008,17(2): 671-676
- [3] 覃邦余，重金属污染物在土壤环境系统中运移的建模与仿真[D] 2009
- [4] 冈萨雷斯, 数字图像处理（第二版） 北京：电子工业出版社 2004 年