**A题 城市表层土壤重金属污染分析**

**摘 要**

随着工业迅速发展，重金属污染物对环境产生了巨大的影响，现对某一城市进行实际观测，从而分析该城区内不同区域重金属的污染程度，推测重金属污染的主要原因，确定污染源的位置。

第一问：用地累积污染指数衡量重金属污染程度，利用克里金插值法，在ArcGIS软件中画出各个重金属污染指数的等值线分布图像，分析每种重金属元素污染指数在空间的大致分布特征，并计算了每个功能区的污染负荷指数，得到污染程度为工业区>主干道区>生活区>公园绿地区>山区；

第二问：按污染指数分布特征将重金属归为四类，用重金属之间两两Pearson相关性分析和主成分分析加以检验，证实分类合理，结合国家对重金属来源分类和污染物的分布特性，得出Pb主要来自于工业、尾气；Cu主要来自于工业、农药；Ni、Cr、Cd、Zn主要来自于工业；As主要来源于工业、农药；Hg主要来自于工业废水和生活垃圾；

第三问：建立基于空间数据的区域曲面拟合模型，考虑到样本的取样信息中包括高程值，我们利用指数衰减分析，对所有样本的高程信息进行订正，把所有的地累积指数订正为只跟平面位置有关的参数，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，根据最小二乘法，对给定区域进行曲面拟合，由位置参数的估计值，确定出重金属污染源的具体位置。Hg的污染分布由四个位于（2379，3685）、（8460，11190）、(6867，7290)、（1369，2357）的单污染源形成，As的污染是由于位于（4777,4897），（4944,7294）的一个双污染源和位于（9120，16378）的单污染源共同形成，Zn 、Pb、 Cu、 Cd是由位于（2383,3692），（3297,6016）的一个双污染源和位于（9436，8375）的单污染源共同形成，Ni和Cr的污染是由位于（4592，4604）的单污染源形成的；

第四问：本文所建模型均只基于所给样本数据集，若增加样本信息，可将所建模型进行推广。如在补充河流位置、流向等地表信息下，就可以利用偏斜椭球分布密度模拟出非点状污染源的扩散分布；如果能获取该城区所有工厂或垃圾场等所有可能污染源位置，即可降低污染源误判概率，并提高污染源位置估计精度。

关键词：重金属 污染分析 地统计分析 空间数据 地累积指数 混合正态分布

**目 录**

[**一、 问题的重述 3**](#_Toc7937)

[**二、 模型假设 3**](#_Toc23409)

[**三、 符号说明 3**](#_Toc12375)

[**四、 对问题一的分析和处理 4**](#_Toc18735)

[**4.1 污染指数的定义 4**](#_Toc21886)

[**4.2 对8种重金属元素的污染指数在城区中的空间分布分析 4**](#_Toc5053)

[**4.2.1．对下面图形的说明 4**](#_Toc14440)

[**4.2.2 研究区域的地形分布和功能区分布分析 5**](#_Toc20832)

[**4.2.3 对8种重金属元素污染指数分布图解析 5**](#_Toc16894)

[**4.3 用污染负荷指数的分析各功能区的污染程度 9**](#_Toc3144)

[**4.4 重金属污染分布和功能区污染强度总结 10**](#_Toc11500)

[**五、 对问题二的分析和处理 11**](#_Toc4263)

[**5.1 重金属污染来源 11**](#_Toc2035)

[**5.2 利用pearson相关性分析对重金属污染分类的评价 12**](#_Toc8262)

[**5.2.1 地累积指数对重金属污染的分类 12**](#_Toc27543)

[**5.2.2 利用相关性分析对5.2.1结论的检验 13**](#_Toc24565)

[**5.3 利用主成分分析分析同类型金属 15**](#_Toc1629)

[**5.3.1 主成分分析原理 15**](#_Toc13642)

[**5.3.2 利用主成分分析对同源金属的分类 16**](#_Toc31748)

[**5.4 分析重金属污染的来源 16**](#_Toc16266)

[**六、对问题三的分析和处理 17**](#_Toc14156)

[**6.1 污染源衰减时的扩散问题 18**](#_Toc11446)

[**6.2 基于空间数据的区域曲面拟合模型的建立 19**](#_Toc2902)

[**6.2.1 扩散模式 19**](#_Toc29707)

[**6.2.2 某固定点污染浓度方程 19**](#_Toc10400)

[**6.2.3 污染传播随高度变化分析 20**](#_Toc2422)

[**6.2.4 搜索污染源的方法介绍 22**](#_Toc19232)

[**6.2.5 搜索结论 23**](#_Toc31188)

[**七、 对问题四的分析和处理 26**](#_Toc29204)

[**7.1 模型的优缺点分析 26**](#_Toc7662)

[**7.2 新增信息 27**](#_Toc14210)

[**7.3 改进的模型 27**](#_Toc17039)

[**八、 模型评价和推广 27**](#_Toc19579)

[**8.1 模型的分析和评价 27**](#_Toc22064)

[**8.2 模型的推广 28**](#_Toc2197)

[**九、 参考文献 28**](#_Toc28933)

[**十、 附录 29**](#_Toc11849)

[**附录1 29**](#_Toc29299)

[**附录2 30**](#_Toc21257)

1. **问题的重述**

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显突出。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。

按照功能划分，城区一般可分为生活区、工业区、山区、主干道路区及公园绿地区等，不同的区域环境受人类活动影响的程度不同。

现对某城市城区土壤地质环境进行调查。为此，将所考察的城区划分为间距1公里左右的网格子区域，按照每平方公里1个采样点对表层土（0~10 厘米深度）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多种化学元素的浓度数据。另一方面，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，将其作为该城区表层土壤中元素的背景值。

根据空间观测数据的特点，以及污染物浓度的椭球分布，建立基于空间数据的区域曲面拟合模型，给出 8种主要重金属元素在采样点处的浓度并分析该城区内不同区域重金属的污染程度；说明重金属污染的主要原因；确定污染源的位置。

1. **模型假设**

1．题目所给的8种元素的扩散方式相同，每种元素的扩散是相互独立的

2．假设只考虑表层以内的污染，不考虑重金属污染在地面以下向地心方向的扩散

1. **符号说明**

：重金属种类

：地累积指数

：第种重金属的实测浓度

：第种重金属的环境背景浓度

：背景值变动系数

：第种金属实测值浓度的标准差

：衰减因子

：浓度随高度扩散的函数

：浓度随平面位置变化函数

1. **对问题一的分析和处理**

**4.1 污染指数的定义**

在定量刻画某地的重金属污染，如果直接用重金属的土壤浓度值表示，会有以下缺点：（1）不同元素之间量级相差大，做合成分析很不方便；（2）原始浓度值不能有效地突出某些因素的作用，比如人类活动对重金属污染的影响。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。常用的指数评价有综合评价法、地累积指数法，污染负荷评价法、聚类法，毒理评价法等等，不同的指数有不同的侧重点，本文采用地积累指数作为刻画污染程度空间的物理量。

地累积指数法是Miller1949年提出的针对沉积物种的重金属污染程度的评价方法。除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的变动因素[2]。这种评价方法也可以用来评价土壤中重金属的污染程度及其分级情况。地累积指数的计算式如下：

 （4.1）

式（4.1）中，为地累积指数，是第种重金属元素的实测浓度值，为当地无污染地区第种重金属元素的环境背景，为考虑各地沉积岩差异可能引起的背景值变动系数，本文取1.5。

根据的值，规定了相应的污染程度级别划分标准见表4-1。

表4-1 Muller地累积指数污染物等级划分：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 指数 | <0 | 0~1 | 1~2 | 2~3 | 3~4 | 4~5 | >5 |
| 级数 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 累计污染程度 | 无 | 无~中 | 中 | 中~强 | 强 | 强~极强 | 极强 |

**4.2 对8种重金属元素的污染指数在城区中的空间分布分析**

由于不同的重金属因素的分布不同，可能产生的来源也有所不同，因此下面对8中元素土壤浓度的空间分布分别做讨论。

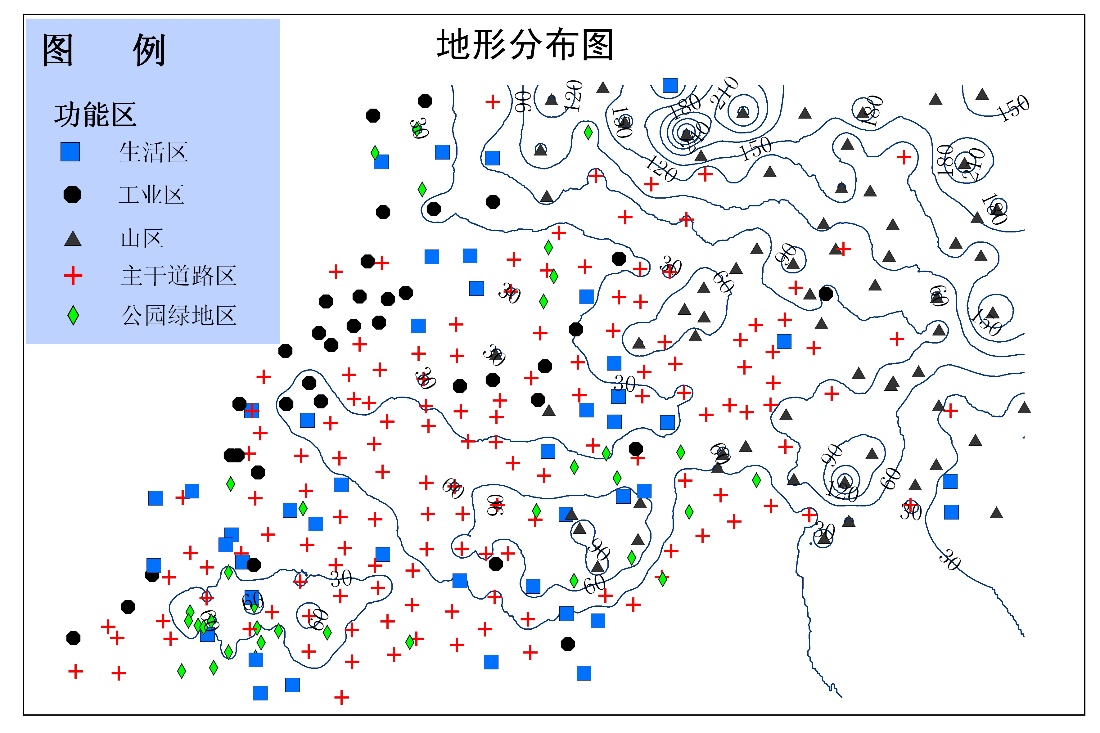
**4.2.1．对下面图形的说明**

下面的图形中的是海拔等高线和污染指数等值线均是利用ArcGIS9.3中的克里金插值法作出的。克里金插值法也叫空间局部插值法，克里金插值以样品的加权平均值求估计值，即对任意待估点或块段的实际值，其估计值是通过该待估点或块段影响范围内的个有效样品值的线性组合得到，即：，式中为权重系数，是各已知样品在估计时影响大小的系数，而估计的好坏主要取决于怎样计算或选择权重系数，估计量称为的克里金估计量。

由于题目给出的数据区域并不是规则的矩形，下面图形的左上角和右下角没有数据，即使插值之后的值也不可信，所以该区域不考虑。

**4.2.2 研究区域的地形分布和功能区分布分析**

根据题目给的数据，利用ArcGIS9.3画出的城区海拔高度和功能区分布图4-1，假设图中箭头所示方向为北，可以看出该城区东部和北部边缘海拔较高，主要为丘陵山区。西部和中南部海拔较低，为广阔的平原地区，中部主要是生活住宅区和主干道路区，西部是比较密集的工业区，西南部有比较集中地公园绿化带，中部有比较分散的公园绿化带。圆圈标出了在城区的中心有一座高约90m左右的山丘，在城区东南角有一座高约60m左右山丘。



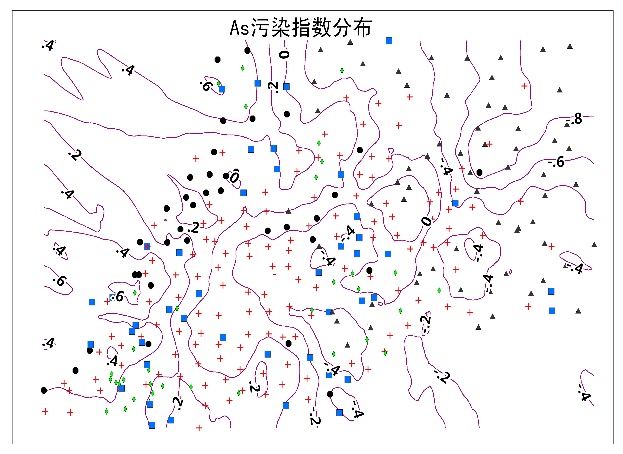
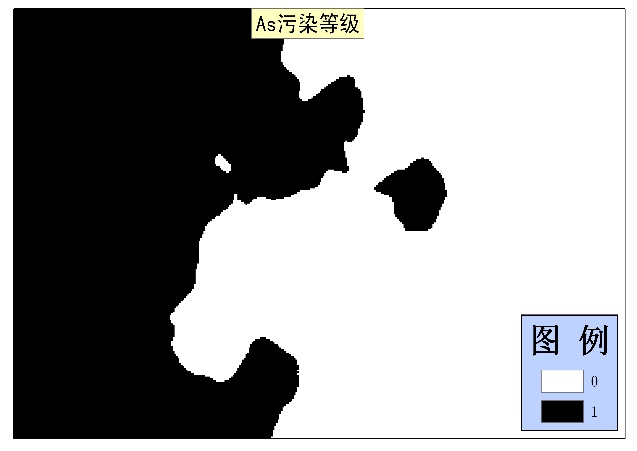
北

图4-1 研究城区的海拔高度和功能区分布情况

**4.2.3 对8种重金属元素污染指数分布图解析**

1、As污染指数分布

将As的土壤浓度按照式（4.1）换算成地累积指数，得到图4-2：



**A**

图4-2(1) 图4-2(2)

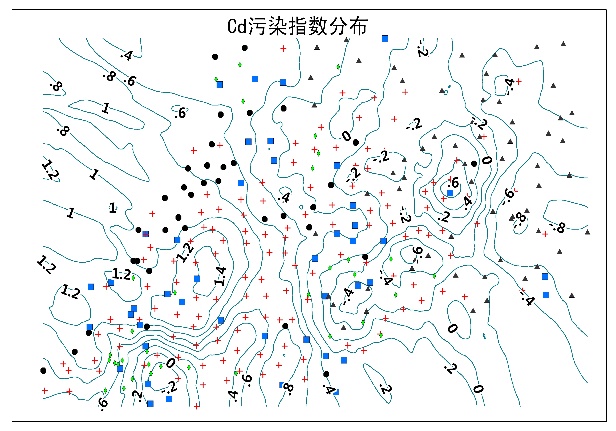
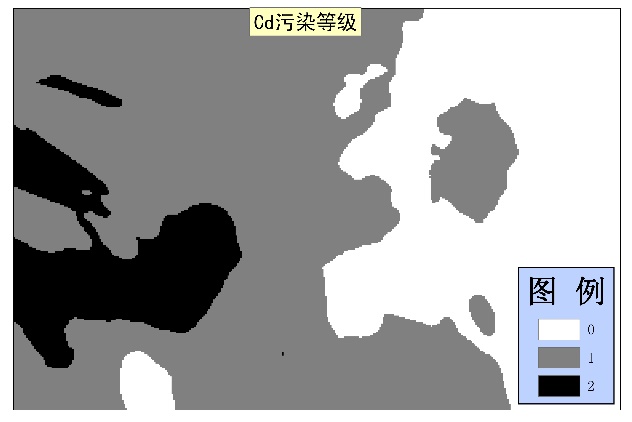
图4-2 As的污染指数分布图

分析As污染指数分布图，发现整张图的As污染指数值范围为,由表4-1可知As在整个城区的污染级别为轻度-中等级别。结合图4-1的海拔高度来说，可以发现西侧的平原地带污染指数较高，大部分地区为污染指数在0.4~0.6范围内，东侧山区的污染指数普遍为负，说明山区无As污染。对于同是平原地带的城区西侧和中部，西侧的工业区周围As污染属于轻中度污染，而中部主干道路区、公园绿地区都是相对低值区。

异常区：根据我们的分析，山区对重金属浓度的分布具有阻碍作用，海拔高度高的地区重金属浓度较低。然而图4-2中方框所示的一座海拔约60m的山的As污染指数相比周围的污染指数高0.6左右，将该处标记为A。

2、Cd污染指数分布

将Cd的土壤浓度按照式（4.1）换算成地累积指数，得到图4-3：



**A**

图4-3(1) 图4-3(2)

图4-3 Cd的污染指数分布图

Cd污染指数在-0.8到1.2之间，分布大致趋势是东侧（山地）的污染指数为负，属于零污染地区，仅在前文标记的A区出现了异常增高，中部的住宅区为中轻度污染地区，在西南侧的生活区和主干道区域污染指数较高，属于中度污染。从功能区的角度分析，发现Cd污染指数在主干道路区和西部的生活区较大，而在其他工业区、中部的生活区、公园绿地区污染指数较小。

异常区域：如4-3（2）方框所示，在前文标记的A区出现了类似与As的土壤浓度异常区。

1. Cr污染指数分布

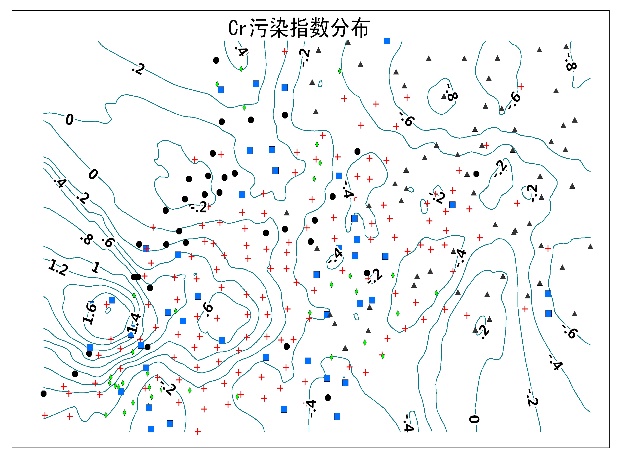
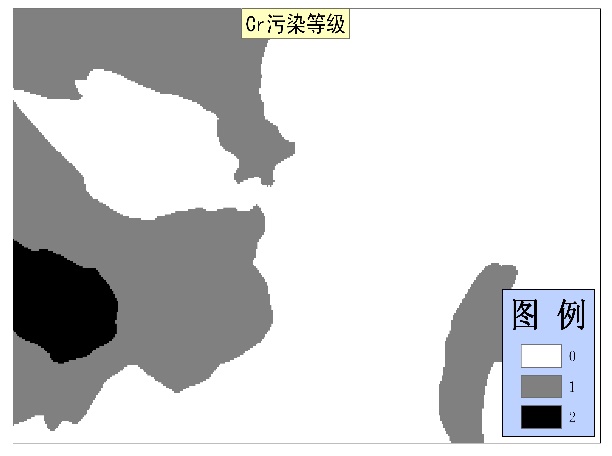
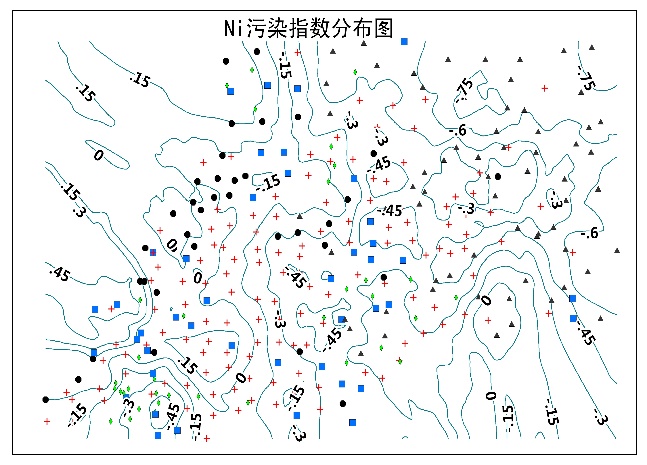
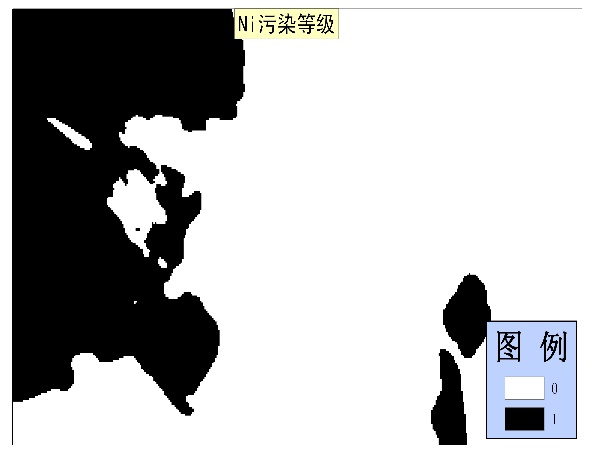


图4-4(1) 图4-4(2)

图4-4 Cr的污染指数分布图

计算可得Cr污染指数值范围为，图4-4(1)中显示图分为东部和西部，西部为白色，说明西部无污染；东部分为轻中度污染（灰色）和中等污染（黑色）两个区域。结合图4-4（1）从功能区的角度分析，中等级别污染的黑色区域位于黑色圆圈之内，该地区主要是一些居民生活区和主干道地区，而在工业区、中部的生活区、公园绿地区污染指数较小。

1. Ni污染指数分布



**A**

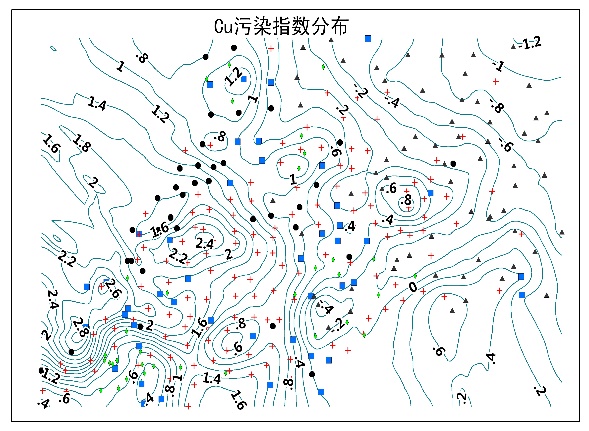
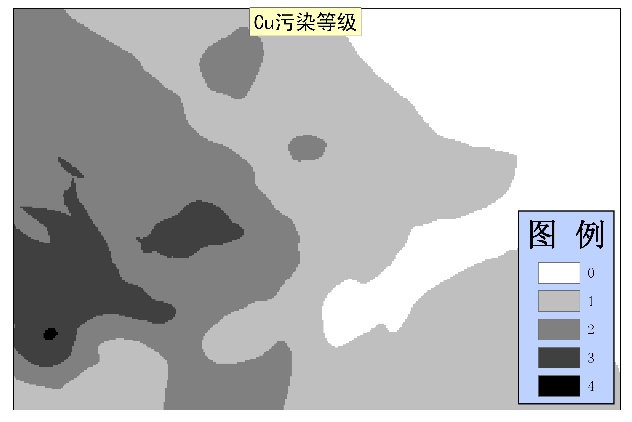
图4-5(1) 图4-5(2)

图4-5 Ni的污染指数分布图

Ni污染指数值范围为，就大部分的城区来说，Ni的地积累指数为负数，即城区的大部分地区都没有Ni污染，仅城区东南角的部分区域有轻度污染。如4-5(2)所示，从功能区的角度分析，该轻度污染区域包含大量的居民住宅区。而在工业区、中部的生活区、公园绿地区、山区污染指数较小。

异常区域：与As类似，在山区A地也出现了一个随海拔高度增加，Ni土壤浓度增高的异常区，如4-5（2）方框所示，该异常区的浓度值较低，并没有造成重金属污染。

1. Cu污染指数分布



**A**

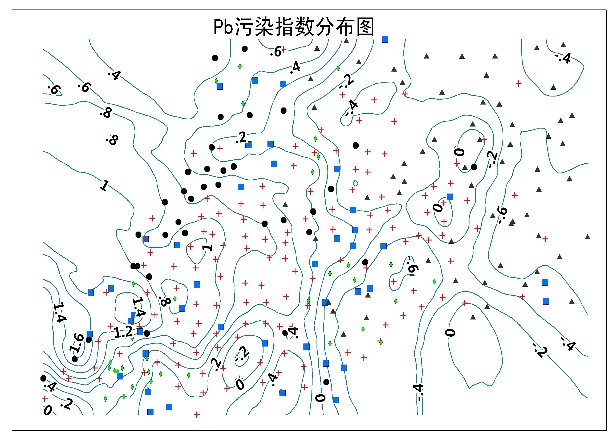
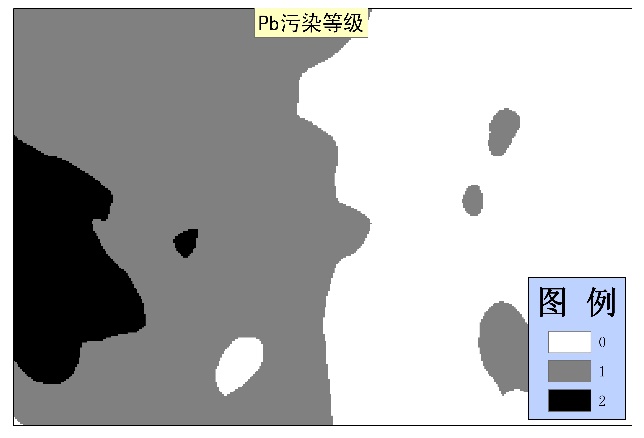
图4-6(1) 图4-6(2)

图4-6 Cu的污染指数分布图

Cu污染指数值范围为，由图4-6(1)可以看出，污染程度由西南到东北呈阶梯型分布，东北角无Cu污染，中部和东南角为轻中度污染，西部为中等程度污染，中间嵌套一些中强污染。根据4-6(2)可以得出，污染严重的区域和污染梯度主要来源是西南部的主干道路区、某些工厂区。

异常区域：如图4-6(2)，方框标志的地区随着海拔升高重金属土壤含量增加，造成了该地区的轻度污染。

1. Pb污染指数分布



**A**

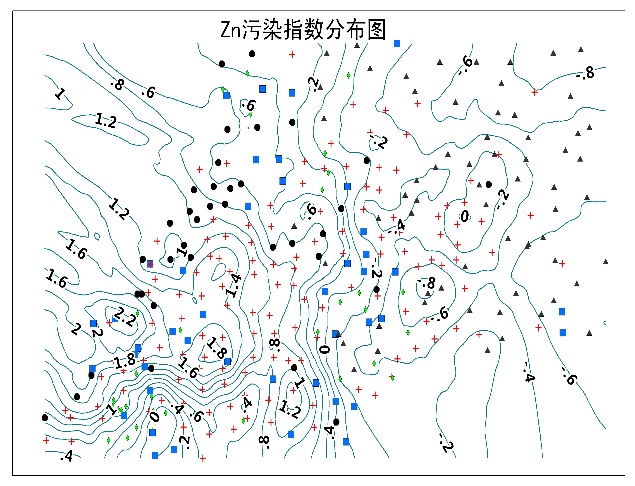
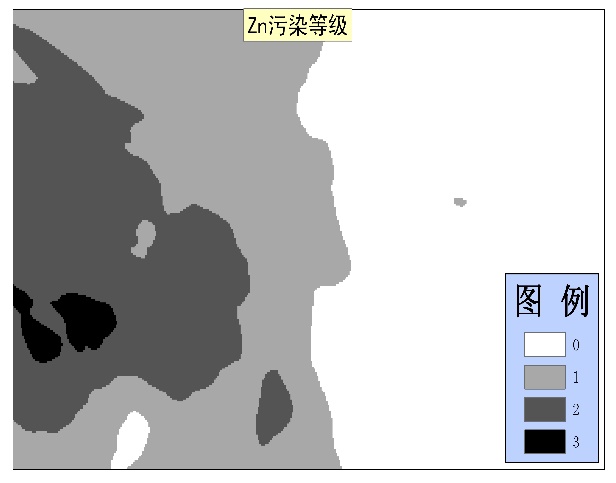
图4-7(1) 图4-7(2)

图4-7 Pb的污染指数分布图

Pb的污染指数在-0.6到1.6之间，如图4-7(1)，城区东部为无污染区，西部为轻度到中等污染区，结合图4-7(2)，轻度污染区污染梯度最大的区域往往分布大量的工厂，城区西南部为中等污染区，该方位为工业区和生活区交汇处。

异常区域：如图4-7(2)，方框标志的A地区仍然出现Ni元素随着海拔升高重金属土壤含量增加的异常区，造成了该地区的轻度污染。

7、Zn污染指数分布



**A**

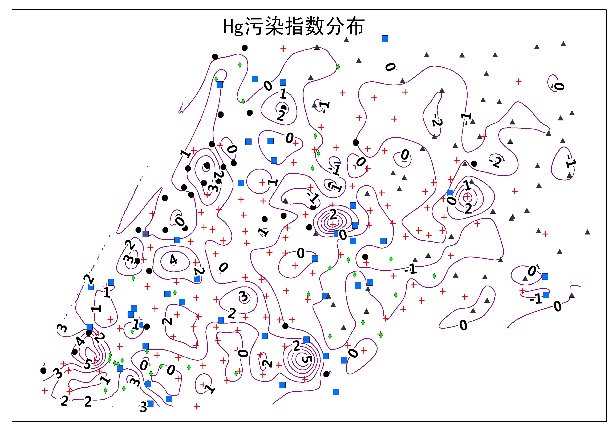
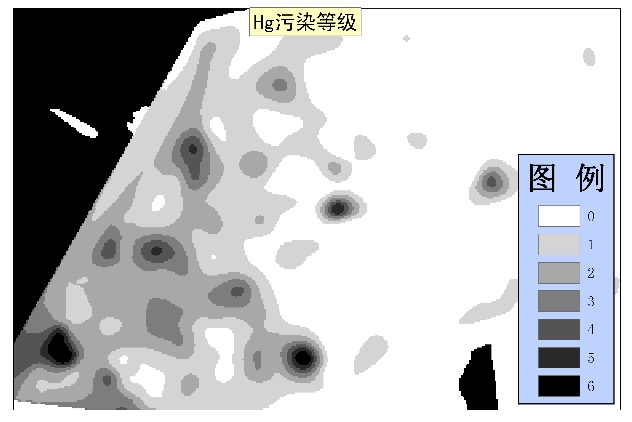
图4-8(1) 图4-8(2)

图4-8 Zn的污染指数分布图

Zn的污染指数分布与Pb的分布趋势比较类似，呈现出较为明显的阶梯状分布，城区东部为无污染区，中部和南部边缘为轻中度污染，西部为中等程度污染，中间嵌套一些中强污染。从功能区来看，工厂带主要位于轻中度污染区和中等污染区，而污染最严重的地方仍然是西南部的主干道区和工厂区。

异常区：如图4-7(2)，方框标志的地区随着海拔升高重金属土壤含量增加。位置与Pb、Cu、As的污染异常区非常相近。浓度没有造成该地区的Zn污染。

1. Hg污染指数分布



**A**

图4-9(1) 图4-9(2)

图4-9 Hg的污染指数分布图

从4-9(1)来看，Hg的污染分布峰值分布比较散乱，东部主要还是无污染地区，自中部到西部大大致趋势是又轻中度污染到中度污染过度，但是与前面几种元素的阶梯状分布不同的是，城区中部和东部出现了许多小的污染严重区域，对比图4-9(2)，这些区域存在与生活区和工业区内，初步估计高污染地区是由工业垃圾或生活垃圾形成。

异常区：Hg含量随海拔升高而升高的异常区如如图4-7(2)的方框标志所示，其位置也与前面分析出的A地区很相近。

**4.3 用污染负荷指数的分析各功能区的污染程度**

地积累指数分析了每种重金属的污染分布情况，为了评定不同功能区的重金属的综合污染程度，定义最高污染物指数：

 （4.2）

式中，为第i种金属浓度的实测值，第i种金属的背景值。

如果某点由多种重金属共同污染，该点的污染负荷指数为：

 （4.3）

为单个点的污染负荷指数，n为金属污染物种数。

由式（4.3），对有单点组成的区域来说，污染负荷指数为：

 （4.4）

为区域的污染负荷指数，m为采样点个数。

根据的值，规定了相应的污染负荷指数等级划分标准见表4-2。

表4-2 污染负荷指数等级划分

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| PLI值 | <1 | 1~2 | 2~3 | >3 |
| 污染等级 | 0 | 1 | 2 | 3 |
| 污染程度 | 无污染 | 中等污染 | 强污染 | 极强污染 |

利用式（4.4）对5个功能区计算，得到各功能区污染负荷和污染程度如表4-3。

表4-3 各功能区污染程度

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 功能区 | 生活区 | 工业区 | 山区 | 主干道区 | 公园绿地区 |
| 污染负荷指数 | 1.83 | 2.35 | 1.06 | 1.92 | 1.58 |
| 污染程度 | 中等污染 | 强污染 | 中等污染 | 中等污染 | 中等污染 |

由上表可以得出五大功能区得污染程度为：工业区>主干道区>生活区>公园绿地区>山区，这一结果符合生活实际，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，但是山区更接近无污染，主干道更接近于强污染。

**4.4 重金属污染分布和功能区污染强度总结**

1．重金属污染指数分布的结论

根据4.2.3分析的结果，总结出污染指数分布的大致趋势如表格4-5所示。

表4-5 污染指数分布的大致趋势

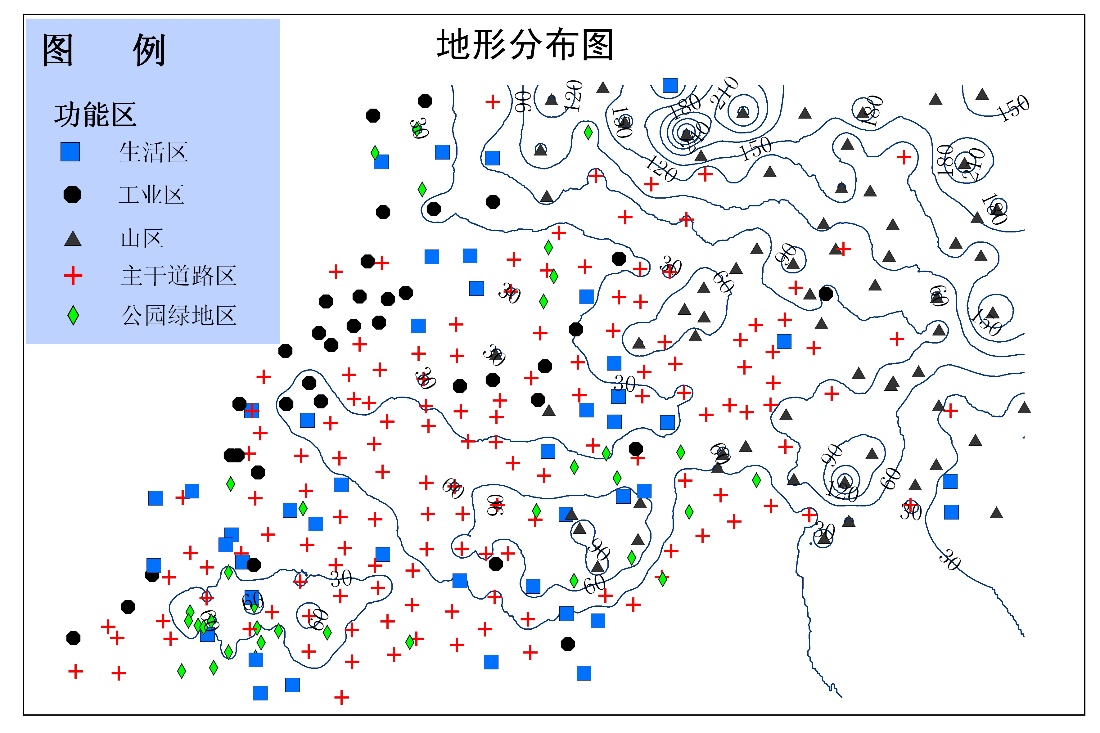
|  |  |  |
| --- | --- | --- |
| 重金属 | 地累积指数 | 污染分布 |
| Ni | （-0.75,0.15） | 东部和中部无污染，西南地区轻度-中级污染 |
| Cr | （-0.8,1.6） | 东部和中部无污染，西南地区轻度-中级污染到中度污染 |
| As | （-0.8，0.6） | 东部和中部无污染，西部轻度-中级污染，但污染区域均匀且范围较大 |
| Cd | （-0.8,1.2） | 东部无污染，中部和西部轻度-中级污染，西南角中度污染 |
| Pb | （-0.6,1.6） | 东部无污染，中部和西部轻度-中级污染，西南角中度污染 |
| Zn | （-0.8，2.2） | 东部无污染，中部轻度-中级污染，西部中度污染，西南角中强度污染 |
| Cu | （-0.12,2.8） | 东北角无污染，东部和中部轻度-中级污染，西部中度污染，西南角中强度污染 |
| Hg | （-0.2,5） | 东部无污染，中部有散布的轻度-中级污染，西部中度污染，呈现污染多峰值散乱分布 |

2. 功能区综合污染强度

根据4.3分析的结果，五大功能区得污染程度为：工业区>主干道区>生活区>公园绿地区>山区。

3. 对关键区域的分析

另外，由以上分析，还可以发现城区中的2个关键区域如图中方框表示的A区和B区。



**B**

**A**

图4-10 关键区域示意图

A区：该区域为污染异常区，由前面的分析可得，除Cr以外，其他七种元素在该地都出现了随海拔增加，重金属土壤浓度增高的异常情况。其中As、Ni、Pb、Zn的浓度小于0，未达到污染水平，Cd、Cu、Hg的浓度达到了污染水平，鉴于多种金属在这里的异常分布，可以初步推断该地区可能有矿区存在，元素浓度增大可能是矿区本身富含比较丰富的重金属元素，也可能是采矿作业中提炼出得矿渣所致。

B区：该区域位于城区的东南角，为污染最严重的区域，每种重金属在这里的浓度都达到了最大值，区域内包含一些工业区域、一些生活区和大量的主干道区域，考虑到该区域海拔较低，其北部有大量的工业集中区，那么B区的重金属的严重污染可能是由于其低地势导致对周围重金属的汇集作用引起。

1. **对问题二的分析和处理**

**5.1 重金属污染来源**

造成土壤重金属污染的原因是多种多样的，土壤重金属污染分为天然污染和认为污染两大类，天然污染包括：某些元素富集中心或者矿床附近等地质因素造成的地区性土壤污染，天气因素引起的土壤淹没、冲刷流失。风蚀以及地震、火山爆发等。人为污染则包括：固体废物的污染，如人类生活垃圾、工业渣土、矿山开采等；农药、肥料在土壤中的残留、积累；污水灌溉以及大气重金属污染物通过降水，沉降方式进入土壤。人为因素造成的重金属污染往往比天然污染更为直接地影响人类的生存环境。表格5-1总结了中国土壤重金属的主要人为来源。

表5-1 中国土壤重金属的主要来源

|  |  |
| --- | --- |
| 来源 | 重金属 |
| 矿产开采、冶炼、加工排放的废气、废水和废渣 | Cr、Hg、As、Pb、Ni、Mo |
| 煤和石油烧烤过程中排放的飘尘 | Cr、Hg、As、Pb |
| 电镀工业废水 | Cr、Cd、Ni、Pb、Cu、Zn |
| 塑料、电池、电子工业排放的废水 | Hg、Cd、Pb、Ni、Zn |
| Hg工业排放的废水 | Hg |
| 染料、化工制革工业排放的废水 | Cr、Cd |
| 汽车尾气 | Pb |
| 农药、化肥 | As、Cu、Cd |

由上表可以得到，不同的重金属污染可以共享同一种污染来源。由同一种人类活动产生得多种重金属污染，称为不同重金属的同源污染。根据本文的假设，所有的重金属扩散模式相同，那么如果两种重金属由同一污染源产生，其空间分布上必然存在某种相似性。为了更好地分析该地区8种重金属的污染成因，下面利用统计中的相关分析和主成分分析对它们的来源进行分类解析。

**5.2 利用pearson相关性分析对重金属污染分类的评价**

**5.2.1 地累积指数对重金属污染的分类**

根据问题一中对每种元素污染指数分布的分析，将分布相同或类似的元素合并，最终得到4个大类，表5-2给出了4个大类所包含的元素，以及每个大类的分布特点。

表5-2 研究区域的重金属污染分布总结和分类

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 类型 | 污染程度 | 重金属 | 地累积指数 | 污染分布 |
| 1 | 低 | Ni | （-0.75,0.15） | 东部和中部无污染，西南地区轻度-中级污染 |
| Cr | （-0.8,1.6） | 东部和中部无污染，西南地区轻度-中级污染到中度污染 |
| 2 | 较低 | As | （-0.8，0.6） | 东部和中部无污染，西部轻度-中级污染，但污染区域均匀且范围较大 |
| 3 | 中 | Cd | （-0.8,1.2） | 东部无污染，中部和西部轻度-中级污染，西南角中度污染 |
| Pb | （-0.6,1.6） | 东部无污染，中部和西部轻度-中级污染，西南角中度污染 |
| 较高 | Zn | （-0.8，2.2） | 东部无污染，中部轻度-中级污染，西部中度污染，西南角中强度污染 |
| Cu | （-0.12,2.8） | 东北角无污染，东部和中部轻度-中级污染，西部中度污染，西南角中强度污染 |
| 4 | 高 | Hg | （-0.2,5） | 东部无污染，中部有散布的轻度-中级污染，西部中度污染，呈现污染多峰值散乱分布 |

注：上述第3类型中，Cd和Pb污染等级较低，最严重为中度污染，Zn和Cu污染等级较高，最严重为中强污染，尽管污染等级有差别，但是由于污染形态均成阶梯状分布。

根据表5-2，可以总结出每个大类的污染特征如表5-3.

5-3 每一大的类型重金属污染分布特征

|  |  |  |
| --- | --- | --- |
| 类型 | 元素 | 特征 |
| 1 | Ni，Cr | 污染程度最低，东部和中部没有污染，西部少量地区有轻微污染，污染范围小 |
| 2 | As | 污染程度较低，东部和中部无污染，西部轻度-中级污染，但污染区域均匀且范围较大 |
| 3 | Cd，Pb  Zn，Cu | 污染程度中等偏高，污染等级呈阶梯分布，受中等污染的范围较大，个别地区出现强污染 |
| 4 | Hg | 污染程度较高，污染分布散乱，似奶牛斑，总体污染程度较低，但在某些点出现极强污染 |

**5.2.2 利用相关性分析对5.2.1结论的检验**

5.2.1根据对各种元素的空间分布图对不同元素的进行了分类，为了在理论上验证上述分类的合理性，利用简单相关分析分别对地积累指数和标准化浓度进行分析，并结合5.2.1的结论进行比较评价。

1. 相关分析原理简介

相关分析是研究两组变量之间整体线性相关关系的一种多元统计分析方法。所研究两组变量既可以某种因果关系，也可以处于同等地位。描述两个变量线性相关关系的统计量为简单相关系数，也即Pearson相关系数，用表示，它也作为两总体相关系数的估计。相关系数计算方法如下：

设有两个变量，观测值分别为和，那么相关系数可以表示为：  （4.1）

其中为样本数量，分别为样本均值。

（1）相关系数的性质： 1）

2）绝对值越大，表示变量之间关系越密切。如果，表明变量之间成正相关关系，如果，表明变量之间成负相关关系。

（2）相关系数的检验

由于在假设总体相关系数成立的条件下，样本相关系数的概率密度函数正好是分布的密度函数，于是可利用检验法来检验。

在原假设的条件下，统计量符合自由度为的分布，给定信度和样本系数，根据自由度查得，当, 那么否定，总体相关，否则总体非相关。

2. 对地积累指数的相关分析

利用SPSS18.0计算8种元素的地累积指数的相关系数矩阵如表5-4所示：

表5-4 地累积指数的简单相关系数矩阵

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | As | Cd | Cr | Cu | Hg | Ni | Pb | Zn |
| As | 1 |  |  |  |  |  |  |  |
| Cd | 0.199\*\* | 1 |  |  |  |  |  |  |
| Cr | 0.440\*\* | 0.420\*\* | 1 |  |  |  |  |  |
| Cu | 0.416\*\* | 0.682\*\* | 0.632\*\* | 1 |  |  |  |  |
| Hg | 0.359\*\* | 0.503\*\* | 0.374\*\* | 0.567\*\* | 1 |  |  |  |
| Ni | 0.538\*\* | 0.306\*\* | 0.759\*\* | 0.538\*\* | 0.316\*\* | 1 |  |  |
| Pb | 0.281\*\* | 0.774\*\* | 0.491\*\* | 0.754\*\* | 0.587\*\* | 0.329\*\* | 1 |  |
| Zn | 0.306\*\* | 0.754\*\* | 0.587\*\* | 0.792\*\* | 0.580\*\* | 0.482\*\* | 0.818\*\* | 1 |

注：表中带\*\*的相关系数表示在信度为0.01时，通过了双边检验显著性相关

可以看到相关性比较大的元素有：Zn和Pb（0.818），Zn和Cd（0.754），Pb和Cd（0.774），Cu和Pb（0.754），Cu和Cd（0.682），Cu和Cr（0.792），也就是说，Zn、Pb、Cd、Cu之间的两两相关非常密切，相关系数均在0.6以上。Cr和Ni的相关性比较大，达到了0.759。As和 Hg是两种比较独立的元素，与其他几种因素的指数的空间分布相关性不是非常显著。

3. 对标准化浓度数据的相关分析

本题给出的数据中，不同的金属浓度变化范围相差很大，为了使不同的数据能在同一标准上比较，需要对原始的浓度数据进行标准化。标准化是将变量中的观察值减去该变量的平均数，然后除以该变量的标准差。即

 （4.2）

式中下标表示元素种类，表示金属土壤浓度，表示平均浓度，表示标准差。

对变量进行的标准差标准化消除了量纲影响和变量自身变异的影响。使得分析的结果更加精确。

将原始浓度进行标准化后，导入到SPSS18.0中计算出相关系数矩阵如表5-5：

表5-5 标准化浓度简单相关系数矩阵

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | As | Cd | Cr | Cu | Hg | Ni | Pb | Zn |
| As | 1 |  |  |  |  |  |  |  |
| Cd | 0.255\*\* | 1 |  |  |  |  |  |  |
| Cr | 0.189\*\* | 0.352\*\* | 1 |  |  |  |  |  |
| Cu | 0.160\*\* | 0.397\*\* | 0.532\*\* | 1 |  |  |  |  |
| Hg | 0.064 | 0.265\*\* | 0.103 | 0.417\*\* | 1 |  |  |  |
| Ni | 0.317\*\* | 0.329\*\* | 0.716\*\* | 0.495\*\* | 0.103 | 1 |  |  |
| Pb | 0.290\*\* | 0.660\*\* | 0.383\*\* | 0.520\*\* | 0.298\*\* | 0.307\*\* | 1 |  |
| Zn | 0.247\*\* | 0.431\*\* | 0.424\*\* | 0.387\*\* | 0.196\*\* | 0.436\*\* | 0.494\*\* | 1 |

注：表中带\*\*的相关系数表示在信度为0.01时，通过了双边检验显著性相关

分析上表可得，Cr和Ni之间有显著的相关性，相关系数为0.716， Pb和Cd（0.660），Pb和Cu（0.520），Pb和Zn（0.494）之间，相关性也比较显著，而As和Hg与其他元素的相关性相对较小，属于两种比较独立的元素。以上结论与前文中对地累积指数的分析比较一致。然而表5-5也反映出，就标准化数据来说，Cu与Cr，Cu与Ni的相关性也比较好，这与前文结论有比较大地差异，根据标准化数据的定义来看，造成这种差异的原因有可能是由于标准化数据没有考虑到不同重金属的背景值差异而引起的。但总体看来，标准化浓度和地累积指数的相关分析的结论是比较一致的。

1. 结论

标准化数据浓度是根据原始浓度标准化后得到的，它比较能反映重金属浓度空间分布的实际情况。地累积指数则是考虑到了重金属浓度的背景值信息，广泛应用在评定重金属污染等级的问题中。这两种数据类型各有侧重。

应用这两种数据类型分别分析8种元素的两两相关关系，可以得到大致相同的结论，即: Ni和Cr相关非常显著，Pb、Zn、Cd、Cu的相关关系比较显著，As和Hg为比较独立的两种元素。这两种数据类型相关分析的结论的相互印证，说明了地累积指数在该问题中能够较好的反映实际的污染情况。

另外这个结论和表5-2分析出的实际情况也是大致吻合的。

**5.3 利用主成分分析分析同类型金属**

**5.3.1 主成分分析原理**

在实际问题种，搜集到得资料通常包含相当多的变量，并且变量间存在较强的相关性，主成分分析就是一种利用原始变量之间的相关性，通过原来变量的上述几个线性组合解释原来变量实现降维的多元统计方法。

假设个样本，个变量的数据，则原始的资料矩阵为：

，用向量形式表示为。

对进行线性变换，形成新的综合变量，用表示：

，其中。

系数要满足以下限定：（1）和不相关 

(2) 依次为所有线性组合中方差第一大，第二大···第P大者

假设的协方差矩阵有非零特征根，各个特征根对应的特征向量，以为系数向量得到的，分别为随机向量的第一主成分、第二主成分、··· 、第主成分。

**5.3.2 利用主成分分析对同源金属的分类**

利用SPSS18.0，对8种重金属空间分布采用主成分分析法分析污染成因，选择特征值大于1作为主成分的抽取条件，分析结果如下表5-6：

表5-6 主成分分析法得到的特征值、累积方差以及因子载荷矩阵

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 特征值 | 方差 | 累计方差 | 因子载荷矩阵 | | | | | | | |
| As | Cd | Cr | Cu | Hg | Ni | Pb | Zn |
| 第一主成分 | 3.56 | 44.5 | 44.5 | 0.426 | 0.711 | 0.735 | 0.756 | 0.408 | 0.723 | 0.764 | 0.699 |
| 第二主成分 | 1.15 | 14.377 | 58.877 | -0.2 | 0.281 | -0.444 | 0.125 | 0.673 | -0.515 | 0.314 | -0.037 |

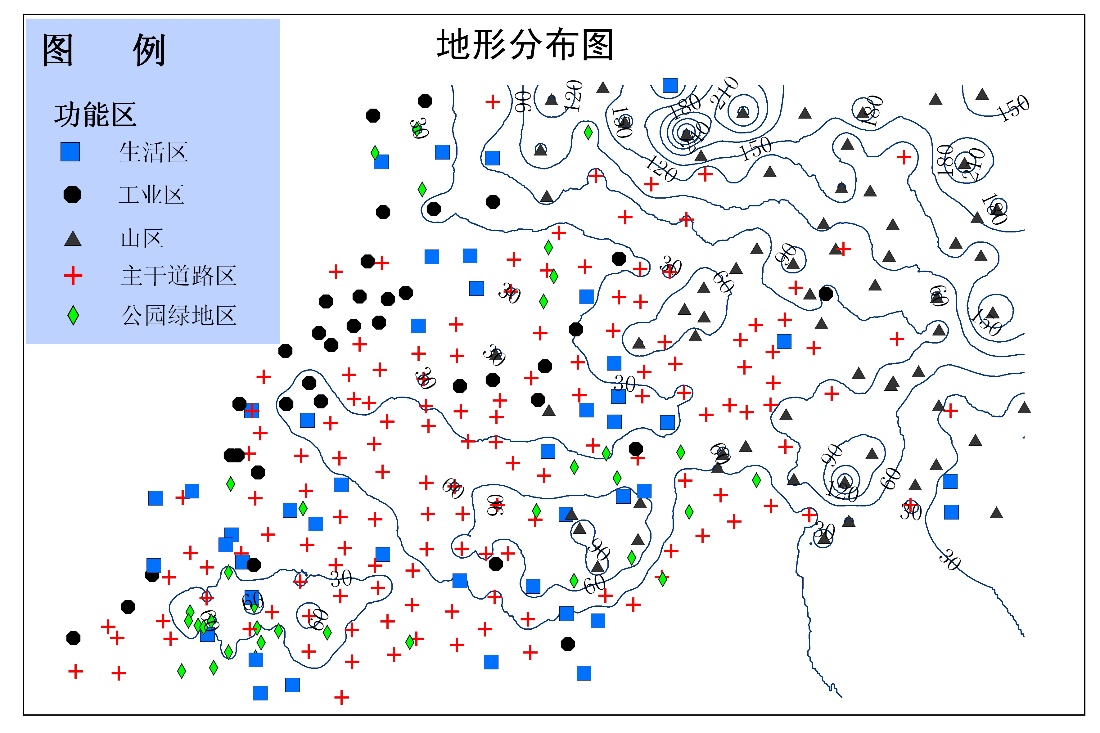
由上表可以看出，特征值大于1的只有两个主成分：第一主成分的方差贡献为44.5%，根据因子载荷矩阵，第一主成分可以很好地代表Cd,Cr,Cu,Ni,Pb,Zn六种元素空间分布，但只能在很少的程度上代表As元素，这说明这六种元素的空间分布相似性比较大，可能来自同一的污染源；第二成分的方差贡献率为14.377%，最能代表的元素只有Hg。进一步可以看出Pb、 Cu在第一主成分中具有相对高的载荷，Ni、Cr、Cd、Zn有相近的载荷值。

**5.4 分析重金属污染的来源**

结合5.3的相关性分析，Ni和Cr相关非常显著，Pb、Zn、Cd、Cu的相关关系比较显著，As和Hg比较独立。综上我们可以得出Cd,Cr,Cu,Ni,Pb,Zn有一个相同的来源，除此之外，Pb、Cu在某些区域又有各自独特的来源；Hg和As独立性表明其来源特殊。

根据表5.2中国土壤重金属来源，可以看出Pb的来源主要有工业、尾气；Cu的来源主要有工业、农药；Ni、Cr、Cd、Zn主要来自于工业；As主要来源于矿业开采、工业、农药；Hg的来源主要来自于工业废水和生活垃圾。

根据第一问对于污染物分布的分析，该城区的污染主要集中在图所示的A、B、C、D、E五的区域。



**D**

**B**

**A**

**E**

**C**

图5-1污染集中分布区域图

综上分析得：

①在A区每种重金属都有相当高的污染指数，我们猜测，该区的污染源主要有A区得工业造成，在该区肯定有大型的重工企业排放Cd、Cr、Cu、Ni、Pb、Zn、Hg等污染物；

②在B区，Pb、Cr、Cu、Zn有高污染值，B区左上角处为大片的工业区，但各重金属污染工业区的污染相对B区而言却反而不大，而其余金属污染高值的来源是由于B左上角出的工业区，我们猜测在B区有一集中的垃圾场；B主要是道路交通主干道，我们确定，大片Pb高值区很大一部分来自于汽车尾气；

③C区Zn、Pb、Hg有很高的值，C区周围是集中的生活区，其污染来源为生活垃圾，Hg、Zn、Pb主要来源于生活垃圾（比如废电池）；

④D区为主要有生活区和工业区，各种金属污染相对而言都比较高，其污染源为工业废物和生活垃圾；

⑤E区为山区，其地势相对周围很高，根据污染指数指示，在该区各重金属相对周围都有一个异常的极大值，但其污染指数相对重污染地区而言污染很小，为轻污染区。该处交通发达，又有一个孤立的工业区，我们猜测此处有一个矿点。据文献资料显示，矿点采矿历史越长，矿井周田土壤中重金属的富集越高，尤其是Cu、Pb、Ni、Zn等5种污染物，这在一定与本文的的计算和判断相符；

⑥在D、E直接Cu有个高值区，该区为大片的公园绿地，农药是Cu的一个重要来源，我们猜测，该区Cu高值来源于公园使用的农药。

⑦A区和D区As的污染指数都比较高，这些区域都是生活密集区，猜测As的来源主要源自化石燃料的使用；

⑧A区、C区和E区的Hg的污染指数都非常高，这些区域都是工业密集区、生活区密集区猜测Hg主要源自工业废水和生活垃圾。

**六、对问题三的分析和处理**

**6.1 污染源衰减时的扩散问题**

设有一污染源，某物质从此扩散源向四周扩散，沿 *x*，*y*，*z* 三个方向的扩散系数分别为常数，衰减（例如吸收、代谢等）使质量的减少与浓度成正比，扩散前周围空间此物质的浓度为零，估计物质的分布。

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*，它所围的区域是 Ω，由于扩散，从 *t* 到 *t* + Δ*t* 时刻这段时间内，通过 *S*流入 Ω的质量为



其中分别是沿 *x*，*y*，*z* 方向的扩散系数。

由高斯公式



由于衰减，Ω 内的质量减少为



其中为衰减系数。

由物质不灭定律，在 *t* 到 *t* + Δ*t* 时刻间 Ω 内由于扩散与衰减的合作用，积存于 Ω 内的质量为。

换一个角度看，在 *t* 到 *t* + Δ*t* 时刻间 Ω 内由于浓度的变化引起的质量增加为



显然，，即



由 Δ*t*，*t*，Ω 的任意性得：



上述方程是常系数线性抛物型方程，它就是有衰减的扩散过程的数学模型。

设扩散源在点处，则此扩散问题满足 Cauchy 问题：



其中 *M* 为扩散源的质量。用傅立叶变换可求得Cauchy 问题的解析解为



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，假设污染扩散是多元的正态分布，为使得模型更符合实际，本文假设为椭球分布。

**6.2 基于空间数据的区域曲面拟合模型的建立**

**6.2.1 扩散模式**

设一个复合函数。根据6.1的假设，扩散模式呈椭球分布的形态。是应符合多元正态分布，多元正态分布是一元正态分布的直接推广。设随机向量服从维正态分布，则有：



在本问题中， 可以简化地表示为：

 （6.1）

式中为污染源的权重系数。

**6.2.2 某固定点污染浓度方程**

设第个观测点的污染物浓度指数为：

 （6.2）

式中，,是观测点的水平位置，是观测点的海拔高度。

如果污染是呈椭球分布，那么式（6.2）可以改写为：

 （6.3）

式中，是位置函数。

对上述模型做进一步简化，将位置函数与海拔高度做变量分离后得到：

 （6.4）

式中是扰动误差，符合正态分布； 是浓度随水平位置的变化，是浓度随高度扩散。式（6.4）表明第个观测点的污染物浓度可以近似等效为水平扩散和垂直扩散的叠加作用造成。

根据式（6.4），污染物的水平扩散可以表示为：

 （6.5）

其中应大于题目所给数据中金属污染背景值得方差。

**6.2.3 污染传播随高度变化分析**

（1）一般的衰减定理

Beer-伯格-朗伯定律描述了一般的衰减作用。根据该定律，设为初始的污染浓度，为考虑高度衰减后海拔处的浓度，根据衰减的一般规律，有：

 为衰减因子 （6.6）

假设升高单位高度上的衰减系数是，则升高距离上所衰减的浓度是：

 （6.7）

对式（6.7）进行变换积分：



得到积分结果为：

 （6.8）

带入（6.6）式，得到衰减因子为：

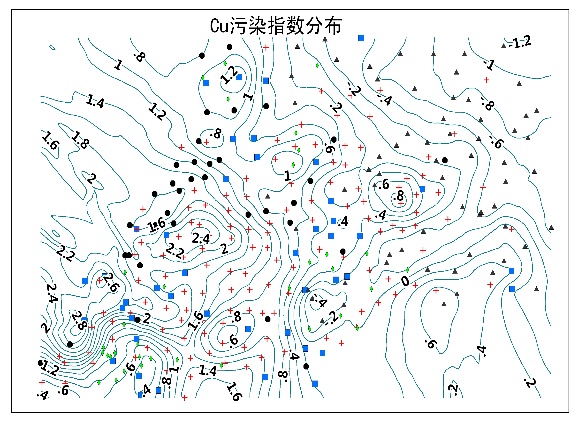
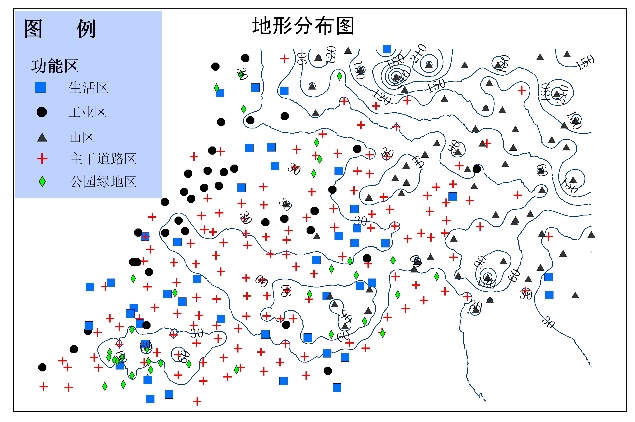


Beer-伯格-朗伯定律描述一般衰减都是成指数分布的，因此本文假设金属浓度扩散随高度的衰减也是成指数分布。

（2）浓度扩散随高度衰减函数的确定

为得到污染传播与地形高度之间的关系，必须消除污染指数随水平传播距离变化的影响。为尽可能的减少水平传播衰减的影响，在较小的水平尺度上寻找高度梯度最大处对进行求解，在水平差异很小而垂直差异很大的区域内，我们有理由认为传播的衰减主要由高度梯度引起。

在地形分布图6-1（a）中，等高线最密集即地形最陡峭处，主要有两处，一处在城区东南部（标示为A），另一处在最北边（标示为B）。



**B**

**A**

**A**

**B**

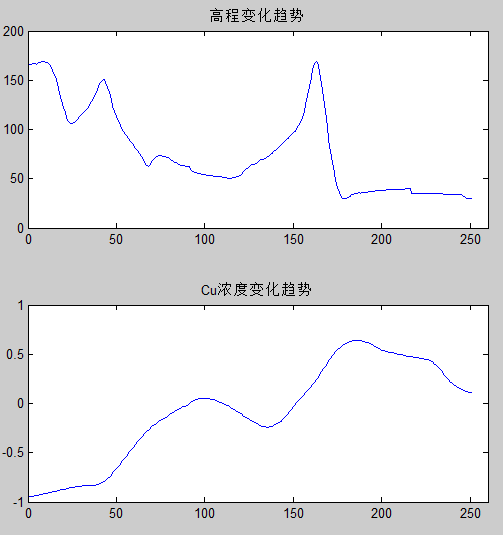
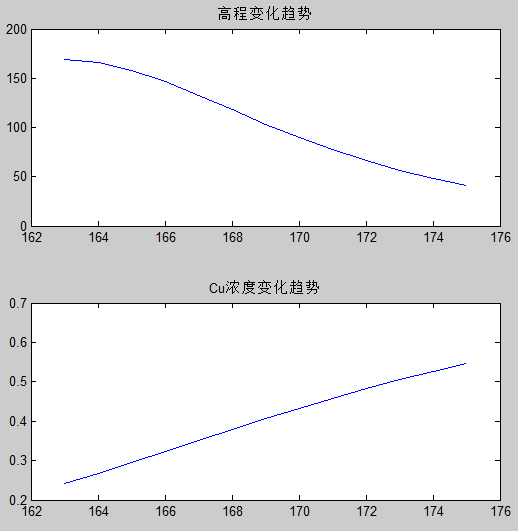
(a) (b)

图6-1 Cu的污染指数与海拔的对比关系

根据假设1，各种污染物扩散模式相同、相互独立，为使结果更具代表意义，取在A、B两处污染都较为适中的铜的污染分布去求解污染随高度衰减。

6-1(a)图显示，B处为密集的山区，A处是孤立的山区（等高线次密集区）南边为大片坡度缓和的平原区，对应于6-1(b)Cu的污染分布，可以较为直观的看出，随着海拔的增高，Cu的浓度降低，说明山地对污染的传播有阻隔作用；6-1(b)图中，在A南部缓和的坡面和平原区，污染指数衰减缓慢，A的北端由于等高线密集，地形陡峭，这里的污染相比与周围平原区在扩散时衰减更加厉害。相对于B处，A处涵盖了地形陡峭和缓和两种情况，更有利于做比较分析。因此取A处研究高度对污染的影响更为合理。

在6-1(a)图中等高线最密集处做垂直剖面，如6-1(a)中的虚线所示，得到6-2组图：

(a) 陡峭处 (b)放大图

图6-2 剖面的高程变化和浓度变化对比分析

图6-2(a)中横坐标为网格序数，单位网格距为73.796米。由图6-2(a)的高程变化趋势可以看出，在第160个网格至175个网格处海拔高度的急剧下降即为图6-1中的A点等高线最密处，截取该区间得到放大后的图像，如图6-2(b)。可以看到，随海拔高度下降，Cu的浓度上升。

根据Beer-伯格-朗伯定律，假设重金属污染扩散随高度呈指数衰减，将得到图6-2(b)的海拔高度和污染指数数据用指数函数进行拟合和参数估计结果如下：



图6-3 拟合曲线效果图

由图6-3看出，用指数函数拟合的效果非常理想， R方值达到了0.983，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，拟合得到海拔高度与污染指数的关系为：



与（6.6）式比较得，衰减因子的表达式为：

因此，式（6.5）中浓度随高度扩散的函数为：

 （6.9）

**6.2.4 搜索污染源的方法介绍**

根据（6.5）和（6.9）式，得到浓度随平面位置的分布函数为：

 （6.10）

这样就将高度为的浓度根据指数衰减规律订正到了平面上，从而根据题目所给的数据样本利用（6.10）换算得到新的样本。

若扩散模式呈椭球分布的形态，扩散函数应符合多元正态分布，当只有一个污染源时，处的浓度分布为：

 （6.11）

当有个污染源时，处的浓度是N个污染源在此处污染效果的叠加：

 （6.12）

其中都是待估计的参数。

目标函数为：，表示估计的污染分布与实际分布偏差最小。

结合问题二中重金属空间分布相似性的分类，将Ni、Cr归为一类；As为一类；Zn 、Pb、 Cu、 Cd 为一类；Hg为一类。由于扩散模式相同，同一类型的重金属认为它们具有相同的污染源。虽然山区存在一些相对高值区，根据假设3，这些区域往往与人为因素无关。可以认为这些高值区都是由于自然界自身形成的，不是污染区。因此以下的高度订正和对污染源的搜索都只在城区中部和东部海拔较低的地区进行，不考虑右侧的山区。

搜索步骤：

Step1：由高度订正公式（6.10），将高度为的浓度根据指数衰减规律订正到了平面后，可得到订正后的重金属污染等值线图:。

Step2：分析订正后的重金属污染物指标的等值线图，根据AIC准则，找到可能的污染源位置，选择搜索区域，把落在该区域中的观测点取出来；

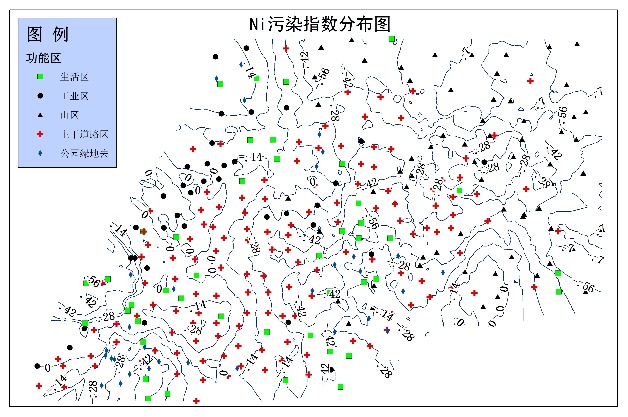
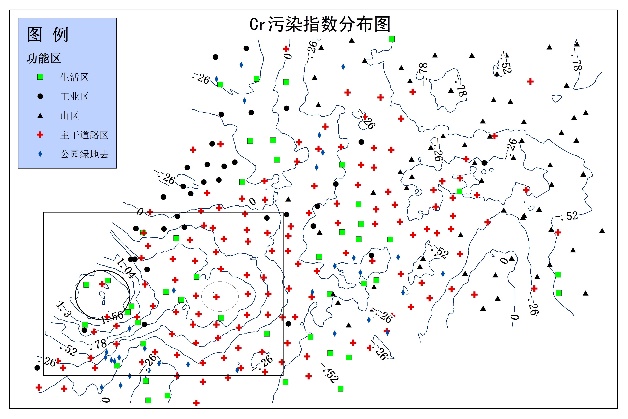
Step3：找出Step2中观测点内金属污染指标的最大值的那一点，或者极大值的那几点，作为污染源位置参数估计的初始位置；

Step4：进行区域搜索，根据最小二乘法找出最优的目标函数所对应的污染源位置。

**6.2.5 搜索结论**

1. 第一类（包含Cr和Ni）

Ni、Cr类的金属，因为Cr污染严重，所以选取Cr作分析，高度订正后的污染指数分布图如组图6-4，图中的圈代表可能的污染源位置，框表示搜素区域（下面不再赘述）。



(a) (b)

图6-4 第一类金属元素的污染分布及污染源搜索

根据等值线分布图像， Cr可能仅有一个单污染源形成或者一个双污染源形成。

单污染源：搜索区域：X~[1000，9000]，Y~[1000,9000]，无论带入哪个测量值作为初值，经过计算后，找到污染源的位置为（4592，4604）。

双污染源：搜索区域：X~[1000，12000]，Y~[1000,9000]，初值取该搜索区域内污染物含量指数的极大值所在位置，经过计算后，找到污染源的位置为

分析：双污染源得到两个污染源的位置，发现（15291，3208）处已经位于城市的山区，而且在污染指标等值线分析图中是表现为一个低值，认为这种分析结果并不合理，因此我们认为Cr污染源应该为一个单污染源

总结：Cr的污染是由位于（4592，4604）的单污染源形成的。Ni的污染源位置类似。

2. 第二类（包含As）

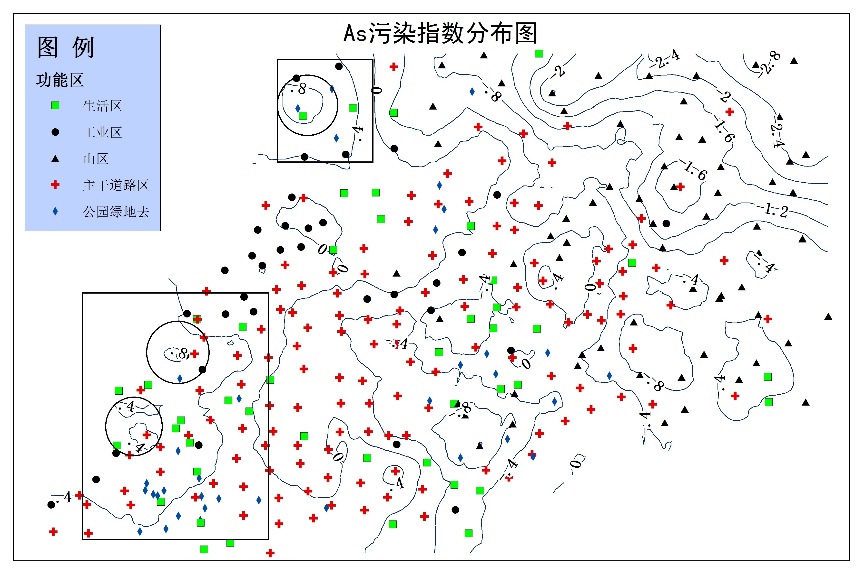


图6-5第二类金属元素的污染分布及污染源搜索

根据等值线分析的极大值中心位置，推断As由一个双污染源和一个单污染源形成。

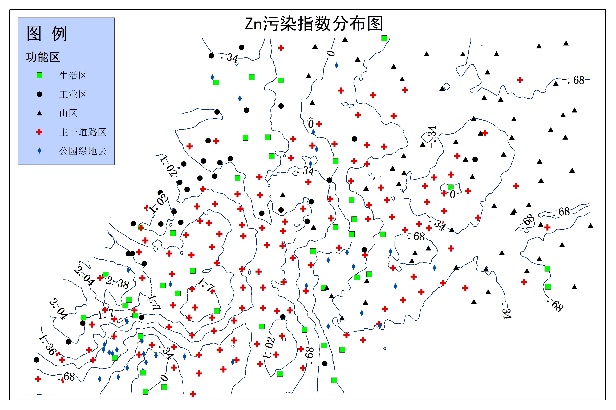
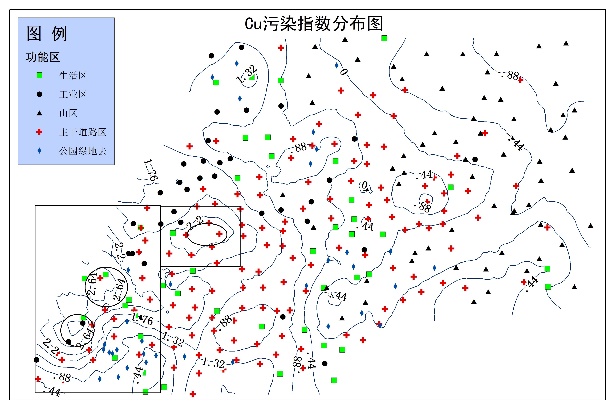
单污染源：搜索区域：X~[7000，11000]，Y~[5000,20000]，无论带入哪个测量值作为初值，经过计算后，找到污染源的位置为（9120，16378）

双污染源：搜索区域：X~[2000，6000]，Y~[2000,8000]，初值取该搜索区域内污染物含量指数的极大值所在位置，经过计算后，找到污染源的位置为

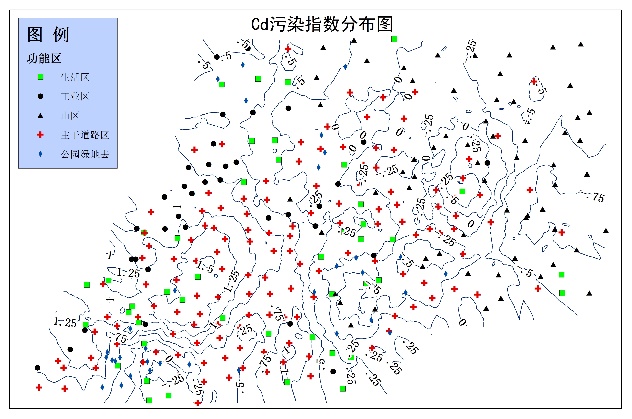
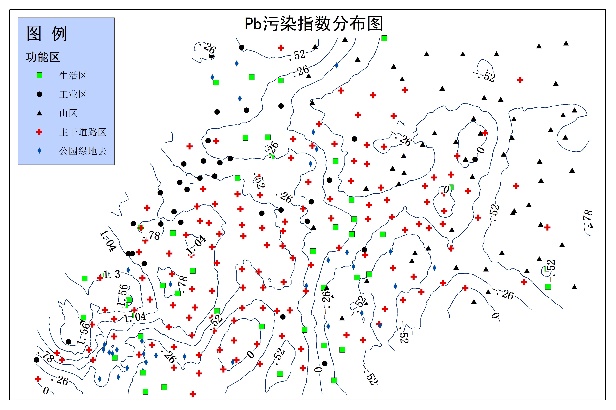
总结：As的污染是由于位于的一个双污染源和位于（9120，16378）的单污染源共同形成的。

1. 第三类（包含Cu、Zn、Pb、Cd）

Zn 、Pb、 Cu、 Cd类的金属，因为Cu的污染最严重，所以选取Cu作分析。



(a) (b)



(c) (d)

图6-6 第三类金属元素的污染分布及污染源搜索

根据污染指数等值线的极大值中心，推断Cu由一个双污染源和一个单污染源组成。

单污染源：搜索区域：X~[6000，10000]，Y~[6000,9000]，无论带入哪个测量值作为初值，经过计算后，找到污染源的位置为（9436，8375）

双污染源：搜索区域：X~[0，6000]，Y~[0,9000]，初值取该搜索区域内污染物含量指数的极大值所在位置，经过计算后，找到污染源的位置为

总结：Cu的污染是由于位于的一个双污染源和位于（9436，8375）的单污染源共同形成的。Zn、Pb、Cd由相同的污染源造成。

1. 第四类（包含Hg）

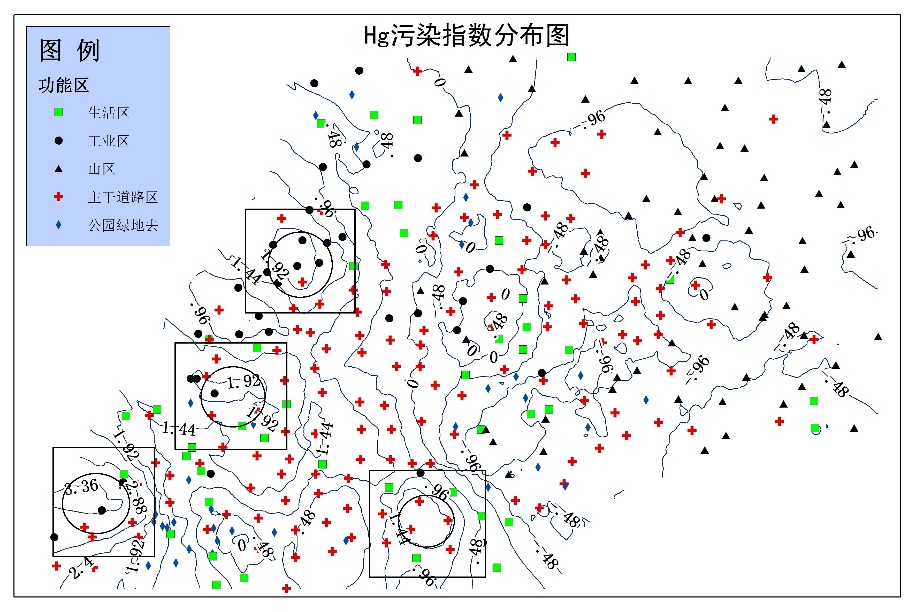


图6-7 第四类金属元素的污染分布及污染源搜索

Hg：根据等值线分布图像，由4个单污染源模型形成。

第一个单污染源：搜索区域：X~[0，3500]，Y~[1000，4000]，无论带入哪个测量值作为初值，经过计算后，找到污染源的位置为（2379，3685）

第二个单污染源：搜索区域：X~[7000，10000]，Y~[9000，12000]，无论带入哪个测量值作为初值，经过计算后，找到污染源的位置为（8460，11190）

第三个单污染源：搜索区域：X~[5000，8000]，Y~[4500，8000]，无论带入哪个测量值作为初值，经过计算后，找到污染源的位置为(6867，7290)

第四个单位污染源：搜索区域：X~[12000，15000]，Y~[0,3500]，无论带入哪个测量值作为初值，经过计算后，找到污染源的位置为（1369，2357）

总结：Hg的污染是由于四个位于（2379，3685）、（8460，11190）、（8460，11190）、（1369，2357）污染源形成的。

1. **对问题四的分析和处理**

**7.1 模型的优缺点分析**

1、基于空间数据的区域曲面拟合模型优点：

(1) 该模型中所有分析的对象都是重金属浓度的地累积指标，而不是直接用测到的浓度去分析，这样做的好处消除了当地环境背景中金属元素的分布影响；

(2) 根据污染等级的划分，得到整个山区都不是污染区，在污染源考虑时不考虑右侧的山区，这就大大简化了问题；

(3) 对高程信息进行处理。由于给出的测量值是一个空间数据，其中包括高程信息，我们在对观测点污染物指标进行扩散的分析中要去除高程的影响，所以要对高程先进行订正。对于如何订正，我们利用比尔吸收定理的衰减原理去拟合样本，能够得到令人满意的效果，消除高程的影响。

(4) 结合污染源扩散污染物的椭球型分布规律，并且考虑了多个污染源共同产生的影响，从图形中找到一些可能的污染源位置，对污染源的位置进行最优的搜索，最终得到较为合理的污染源个数和它们的位置信息。

2、基于空间数据的区域曲面拟合模型缺点：

(1) 模型只适用于点污染源的概率分布，没有考虑线污染源和其它形式污染源的扩散概率特点；

(2) 建立该模型的样本信息量不足，离真实的重金属污染物浓度指标分布还是有所偏差。

**7.2 新增信息**

1、继续采样，增加样本信息，得到更为真实的重金属污染物浓度指标分布

2、城市的地表信息，如河流、农田、风向等的分布状况。

3、城市中每个工厂、垃圾场及可能的重金属污染源的类型，排放的污染物信息等。

**7.3 改进的模型**

掌握城市的地表信息之后，可以估计一些非电源分布的污染源，如沿着河流分布的条带状污染源，那么污染物分布就不会符合椭球模型，要用到偏斜二元正态分布，来拟合区域曲面；

根据基于空间数据的区域曲面拟合模型寻找到的一些污染源位置，结合当地一些可能的污染工厂或者垃圾场等，了解其排放污染重金属类型，确定出该种重金属污染的正真来源。

1. **模型评价和推广**

**8.1 模型的分析和评价：**

第一问：该模型从重金属污染物分布出发，利用克里金插值，把1Km分辨率的观测值查到73m分辨率的空间网格点中，然后画出各个重金属浓度指标的等值线分布图像，结合各个功能区的空间分布，可以得到各个功能区每种重金属的污染程度。该方法简单易行，克里金插值中同时考虑了地形的要素，插值精度高，得到的等值线分布较准确；

第二问：利用重金属之间两两pearson相关性分析和主成分分析结合地累积指数，把重金属污染物归为四类，每一类都有类似的分布，可以认为它们具有相同的污染源，再根据国家对于各种重金属污染来源的分类，确定出重金属污染的主要来源。对于该问题的求解，用到了多种统计分析方法，并都通过了显著性检验，对于图像上的实际分布也是吻合的很好；

第三问：建立基于空间数据的区域曲面拟合模型，考虑到样本的取样信息中包括高程值，利用指数衰减拟合分析，对所有样本的高程信息进行订正，把所有的浓度化为只跟平面位置有关的参数，最后结合多污染物传播具有二元正态分布概率密度叠加的特征，利用最小二乘法，对给定区域的样本进行参数估计，确定出重金属污染源的具体位置。该模型没有直接出扩散方程的复杂形势去求解，而是转化到了统计学中椭球分布的模型中，可以得到同样的解，但是计算复杂度就大大降低。

**8.2 模型的推广：**

在基于空间数据的区域曲面拟合模型的基础上，如果掌握城市的地表信息之后，可以估计一些非点源分布的污染源，把椭球模型，用到偏斜二元正态分布替代之后来拟合区域曲面；根据原模型寻找到的污染源位置，结合当地一些可能的污染工厂或者垃圾场等，在掌握其排放污染重金属类型后，确定出正真产生该种重金属污染的来源；

从扩散方程的角度出发，如果我们可以得到某一地点不同时刻的污染物浓度测量值，可以想方法对扩散方程进行数值求解，来更好的模拟整个污染物扩散的过程及扩散规律。

1. **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，同济大学学报（自然科学版），Vol.34(12)：1657-1661,2006。

[2] 李卫东，应用多元统计分析，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，中南大学硕士学位论文，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，西南大学硕士学位论文，2007。

[5] 陈杰，MATLAB宝典（第三版），北京，电子工业出版社，2011。

[6] 滕葳，重金属污染对农产品的危害与风险评估，北京，化学工业出版社，2010。

1. **附录**

附录1：

ni1=dem;

d=0;

m=0;

n=0;

for i=125:251

for j=1:389

if ni1(i,j)>=d

d=ni1(i,j);

m=i;

n=j;

end

end

end

% d,m,n

clear x

% x=163:251;

x=1:89

% sx=ni1(163:251,n);

% subplot(2,1,1);

% plot(x,sx);

%

% cu=cu1(163:251,n);

% subplot(2,1,2)

% plot(x,cu)

sx=ni1(1:251,n);

subplot(2,2,1);

plot(1:251,sx);

title('高程变化趋势')

cu=cu1(1:251,n);

subplot(2,2,2)

plot(1:251,cu)

title('Cu浓度变化趋势')

clear sx cu

sx=ni1(163:175,n)

subplot(2,2,3);

plot(163:175,sx)

title('高程变化趋势')

cu=cu1(163:175,n);

subplot(2,2,4);

plot(163:175,cu)

title('Cu浓度变化趋势')

附录2：

clc

clear all

close all

%单个污染源的位置估计

data1=xlsread('cumcm2011A附件\_数据',1);

data2=xlsread('cumcm2011A附件\_数据',5);

x=data1(:,2);%横坐标

y=data1(:,3);%纵坐标

h=data1(:,4);

I=data2(:,1);%订正后的地累积指数

%选择某一区域

k=1;

for i=1:319

if x(i)>=7000&&x(i)<=11000

if y(i)>=15000&&y(i)<=20000

x(k)=x(i);

y(k)=y(i);

I(k)=I(i);

k=k+1;

end

end

end

x=x(1:k-1);y=y(1:k-1);I=I(1:k-1); %选取研究区域

k-1

location=[x y];

%-------------------------------------------------------

% 参数包括权重w,源的个数N,源的位置(X,Y)，方差参数cov

N=3;%%%污染的个数

mum1=1;

mum2=1;

% mum3=71;

% mum4=278;

X10=x(mum1);Y10=y(mum1);%源的初始位置

k10=I(mum1);%权重初始值

X20=x(mum2);Y20=y(mum2);%源的初始位置

k20=I(mum2);%权重初始值

% X30=x(mum3);Y30=y(mum3);%源的初始位置

% k30=I(mum3);%权重初始值

% X40=x(mum4);Y40=y(mum4);%源的初始位置

% k40=I(mum4);%权重初始值

%-------------------------------------------------------

theta0=[k10 X10 Y10]';%;k20 X20 Y20;k30 X30 Y30;k40 X40 Y40

% ssig=10^5;

% 估计函数：采用混合加权椭球函数一

fun=@(theta,location) theta(1,1)./(1+log(1+(location(:,1)-theta(2,1)).^2+(location(:,2)-theta(3,1)).^2))...

% +theta(1,2)./(1+log(1+(location(:,1)-theta(2,2)).^2+(location(:,2)-theta(3,2)).^2))...

% +theta(1,3)./(1+log(1+(location(:,1)-theta(2,3)).^2+(location(:,2)-theta(3,3)).^2))...

% +theta(1,4)./(1+log(1+(location(:,1)-theta(2,4)).^2+(location(:,2)-theta(3,4)).^2)) ;

% fun=@(theta,location) theta(1,1)\*exp(-1/(2\*ssig)\*((location(:,1)-theta(2,1)).^2+(location(:,2)-theta(3,1)).^2))...

% + theta(1,2)\*exp(-1/(2\*ssig)\*((location(:,1)-theta(2,2)).^2+(location(:,2)-theta(3,2)).^2))...

% + theta(1,3)\*exp(-1/(2\*ssig)\*((location(:,1)-theta(2,3)).^2+(location(:,2)-theta(3,3)).^2))...

% +theta(1,4)\*exp(-1/(2\*ssig)\*((location(:,1)-theta(2,4)).^2+(location(:,2)-theta(3,4)).^2));

% 估计函数：采用混合加权椭球函数二，即为混合二元正态

% ssigma10=100000;ssigma20=100000;%方差初始值

% theta0=[w10 X10 Y10 ssigma10;w20 X20 Y20 ssigma20]';

% fun=@(theta,location) theta(1,1)\*exp(-1./(2\*theta(4,1)).\*((location(:,1)-theta(2,1)).^2+(location(:,2)-theta(3,1)).^2))...

% +theta(1,2)\*exp(-1./(2\*theta(4,2)).\*((location(:,1)-theta(2,2)).^2+(location(:,2)-theta(3,2)).^2));

theta\_hat=lsqcurvefit(fun,theta0,location,I);

para\_int=[theta0,theta\_hat,theta\_hat-theta0];%前k列为源的初始位置，中间k列为源的估计位置,后k列为估计误差

para\_int=para\_int(2:3,:)

[(1:length(I))',I,fun(theta\_hat,location),I-fun(theta\_hat,location)];