

城市表层土壤重金属污染分析的数学模型

摘要

为研究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。本文通过处理和分析已给数据,给出金属的空间分布说明污染程度和主要原因;建立数学模型确定污染源位置;最后收集其他信息讨论城市地质环境的演变模式。

问题一,利用 matlab 软件作出位置坐标 x 、 y 与八种总金属元素浓度的空间分布图;分析采集的重金属元素浓度所在区域的大致情形。对采集的重金属元素浓度的数据进行分析,并计算单因子和多因子污染指数,根据土壤污染分级标准判断出不同重金属元素在各功能区的污染程度和各功能区的综合污染程度,其中工业区中铜是所有元素在不同功能区中污染程度最严重的,而工业区和交通区的综合污染程度是最严重的。

问题二,首先利用 SAS 软件对八种重金属元素在五个城区的含量进行主成分分析,得到八种重金属对各功能区的贡献率,可初步推断出工业生产、交通设施和生活垃圾造成重金属污染。再利用 SAS 软件对各城区的重金属进行因子分析,进一步判断八种不同重金属污染的原因,如汞污染的原因为工业生产中三废的排放、交通运输业中汽油的燃烧和汽车轮胎磨损产生的粉尘等。

问题三,根据所给数据,分析重金属污染传播特征,即分别是介质的迁移运动、污染物的分散运动、污染物的累积与转化、污染物被环境介质吸收或吸附、污染物的沉淀,然后利用 Matlab 软件,采用多元二次项式回归分析方法,分别得到每种重金属元素浓度与坐标的回归方程,并根据该方程利用多元函数求极值的方法确定出污染源的可能位置分别为: As (1878.2634 , 6003.7263 , 4.5846), Cd (970.5835 , 3946.7518 , 6.5891), Cr (1235.1956 , 2658.3427 , 8.5402), Cu (138.4682 , 6223.4521 , 3.2461), Hg (1231.5782 , 2561.5483 , 5.2478), Ni (12234.2587 , 5865.1656 , 23.2461), Pb (2310.6891 , 4145.2674, 3.2651), Zn (3015.4341 , 8642.2365 , 5.0543);

问题四,基于前三问,分析所建模型的优缺点。优点: 1) 通过作出浓度与坐标的分布图,可以很直观的看出各种重金属在不同位置的浓度不同,可以定性的分析金属聚集区; 2) 主成分分析、因子分析可在众多因素中找出各个变量最佳的子集合,从子集合所包含的信息描述多变量的系统结果及各个因子对系统的影响。抓住主要矛盾,把握主要矛盾的主要方面,舍弃次要因素,以简化系统的结构,认识系统的内核。主成分分析和因子可以进行数值分类,构造分类模式。以便找出它们之间的联系和内在规律性。所以第二问中我们采用主成分分析、因子分析可以准确的定量说明各种重金属在各功能区的污染程度和各功能区受不同重金属的污染;。同时参考相关文献,提出需要增加土壤水溶液的 PH、地下水重金属浓度、植物中重金属浓度等,在用指数法和层次分析法给出研究城市地质环境的演变模式。3) 第三问我们采用回归模型,简单易懂,便于实现。缺点: 1) 没有考虑到海拔考虑到这些模型中; 2) 第三问中回归模型不能很好的反映重金属浓度与地理位置的关系。

关键词: 土壤重金属污染 主成分分析 二项式回归 因子分析

一、问题重述

土壤是环境的重要组成部分，是人类赖以生存的自然环境和工农业生产的重要资源。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，尤其是对城市土壤地质环境的影响日显突出。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。

随着工业、城市污染的加剧和农用化学物质种类、数量的增加，土壤重金属污染日益严重，污染程度不断加剧，污染面积逐年扩大。由于重金属污染物在土壤中移动性差、滞留时间长、不能被微生物降解，并可经水、植物等介质最终影响人类健康。因此，土壤重金属污染问题已经成为当今环境科学研究的重要内容。

城区按照功能进行划分，一般可分为生活区（1类区）、工业区（2类区）、山区（3类区）、交通区（4类区）及公园绿地区（5类区）等，人类活动对不同区域环境的影响程度一般会有所不同。因此，分区域对影响土壤地质环境的8种主要重金属元素进行采样，本文要求解决如下几个问题：

- 1、分析8种主要重金属元素在该城区的空间分布，评价不同区域重金属的污染程度；
- 2、利用所给数据，探讨城区土壤重金属污染的主要原因；
- 3、根据重金属污染物的传播特征，建立并求解相应的数学模型，确定污染源的位置，以便为采取有效措施防治土壤重金属污染提供依据；
- 4、分析上述模型的优点和缺点，探讨需要补充的其它信息，重新建立模型研究城市地质环境的演变模式，以便更好地解决土壤重金属污染问题。

二、模型假设

- a. 因取样点海拔差异较小，认为海拔对重金属元素分布的影响较小，可以忽略；
- b. 认为采样点的数据测量准确，并且能较好的反映该地区的污染物浓度；
- c. 短期内重金属浓度不再增加；
- d. 因为采样点在地表表层，不考虑地下水对污染物传播的影响。

三、符号说明

- a. I_i 表示第*i*种污染物，其中 I_1 表示第1种污染物砷； I_2 表示第2种污染物镉； I_3 表示第3种污染物铬； I_4 表示第4种污染物铜； I_5 表示第五种污染物汞； I_6 表示第6种污染物镍； I_7 表示第7种污染物铅； I_8 表示第八种污染物锌
- b. I_i 为第*i*种污染物的污染分指数（ $i=1,2\cdots 8$ ）
- c. C_i 为第*i*种污染物的实测浓度（ $i=1,2\cdots 8$ ）
- d. S_i 为第*i*种污染物的评价标准（ $i=1,2\cdots 8$ ）

四、问题分析

重金属污染指由重金属或其化合物造成的环境污染，主要由采矿、废气排放、污水灌溉和使用重金属制品等人为因素所致。一些人类活动会导致环境中的重金属含量增加，超出正常范围，从而导致环境质量恶化。因此，对土壤中重金属污染的研究对缓解严峻的重金属污染现状有重大意义。

a. 问题一是根据采样点的重金属位置和浓度，通过软件作图说明 8 种重金属元素在 5 个功能区的空间分布和浓度分布，再对数据进行分析判断，以确定不同区域重金属的污染程度；

b. 问题二分析重金属污染的主要原因是一个相关分析的问题，首先我们应考虑这些重金属污染是在哪些或那个功能区较严重，再联系功能区的实际找其可能原因，再进一步分析各功能区中这八种重金属污染物的贡献率，两者结合推断出重金属污染的主要原因。

c. 问题三是在前两问的基础上，分析重金属污染物的传播特征；再结合所给数据，建立污染物浓度和坐标直接相关的数学模型，根据此模型确定污染源的大致位置；

d. 在分析总结前三个问题模型的基础上，查阅相关的地质环境资料，对模型做进一步改进。

五、模型建立与求解

5.1 重金属的空间分布及其在各城区的浓度分布图

5.1.1 空间分布

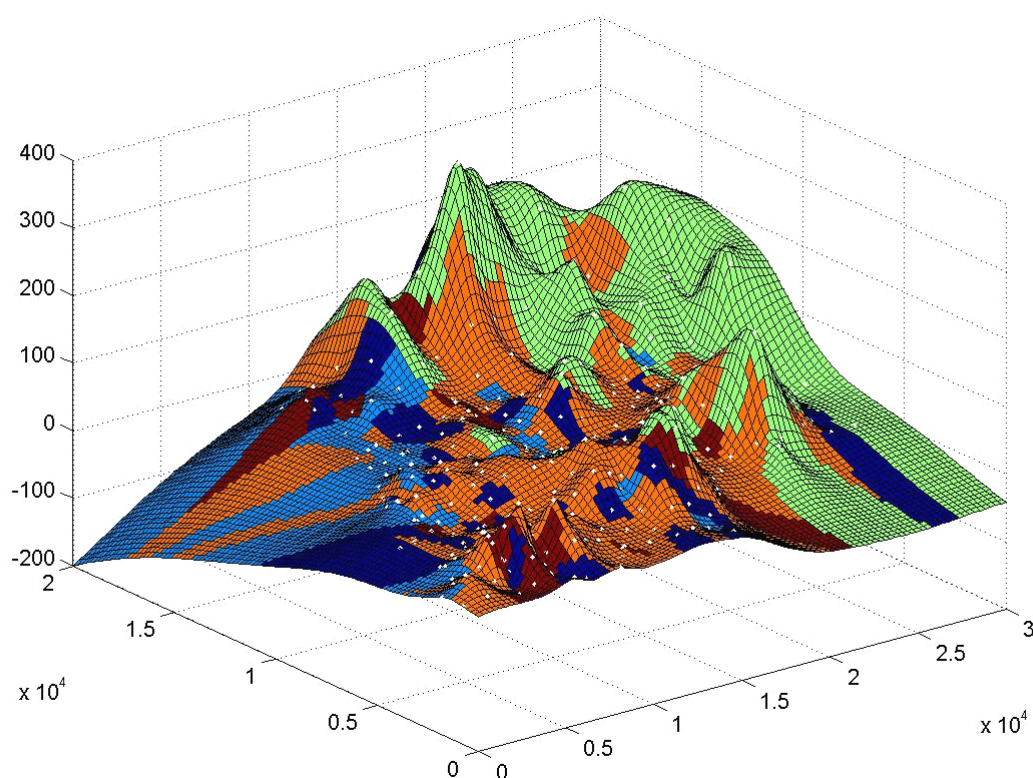


图1 各功能区空间分布图

注：绿色代表山区，橙色代表工业区，蓝色代表生活区，红色代表交通区，青色代表公园绿地区

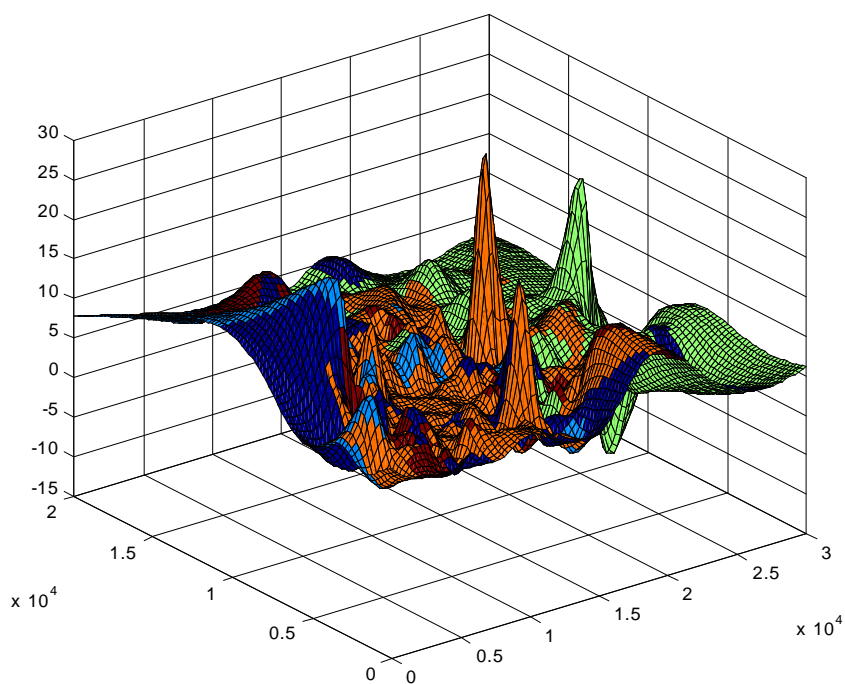


图2 砷 (As) 浓度与坐标x,y的曲面图

从空间分布图可看出，砷含量较高的基本都集中在几个小的区域内，从砷含量的地区分布和采样点的数据来看，其最高含量基本都出现在工业区，少数出现在交通区、公园绿地区、山区和生活区。

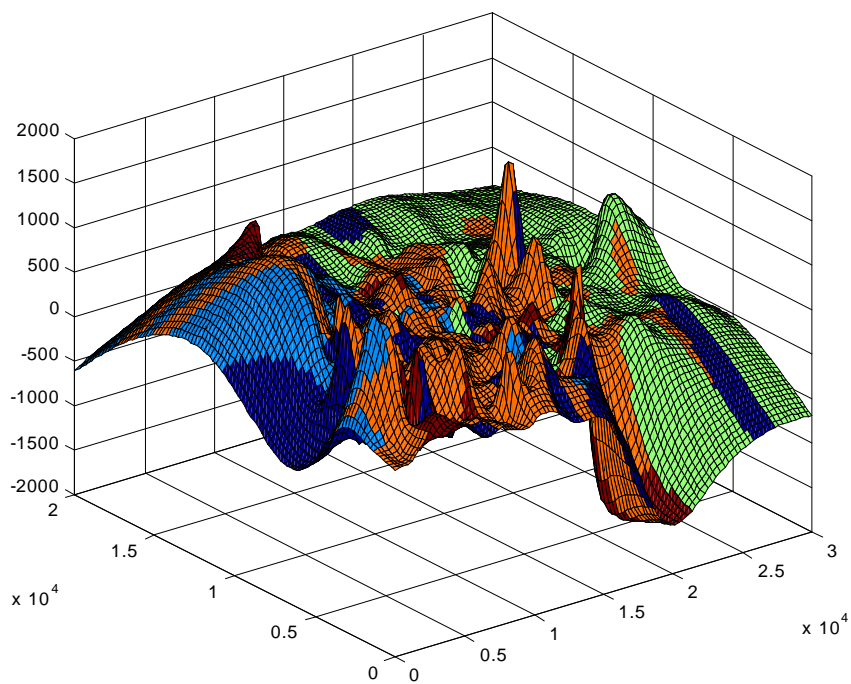


图3 镉 (Cd) 浓度与坐标x,y的曲面图

从空间分布图可看出，镉含量较高的区域比较分散，从镉含量的地区分布和采样点

的数据来看，其最高含量基本都出现在交通区和工业区，少数出现在生活区，极少数出现在山区和公园绿地区。

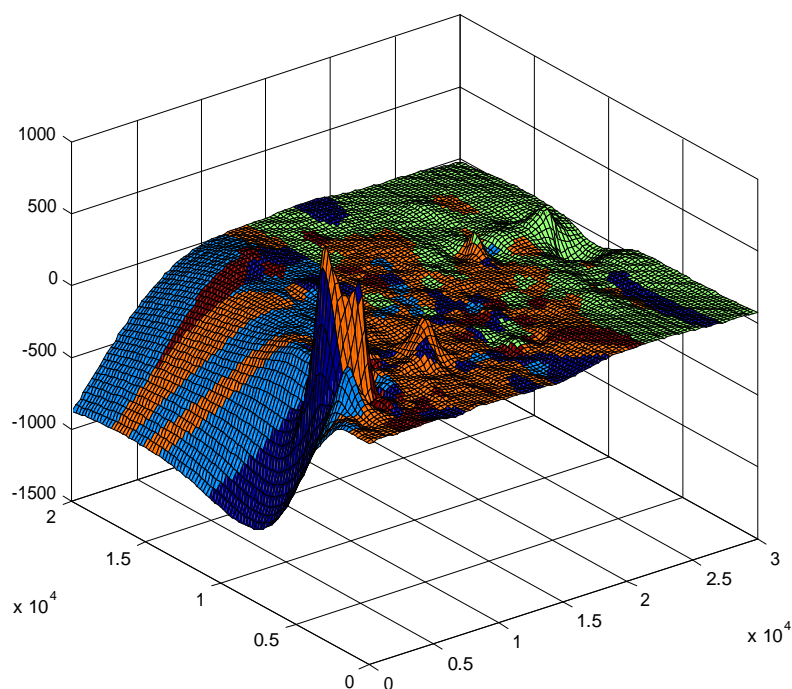


图4 铬（Cr）浓度与坐标x,y的曲面图

从空间分布图可看出，铬含量较高的基本都集中在一个小的区域内，从铬含量的地区分布及其采样点的数据来看，其最高含量基本都出现在工业区，少数出现在交通区和生活区，极少数出现在山区和公园绿地区。

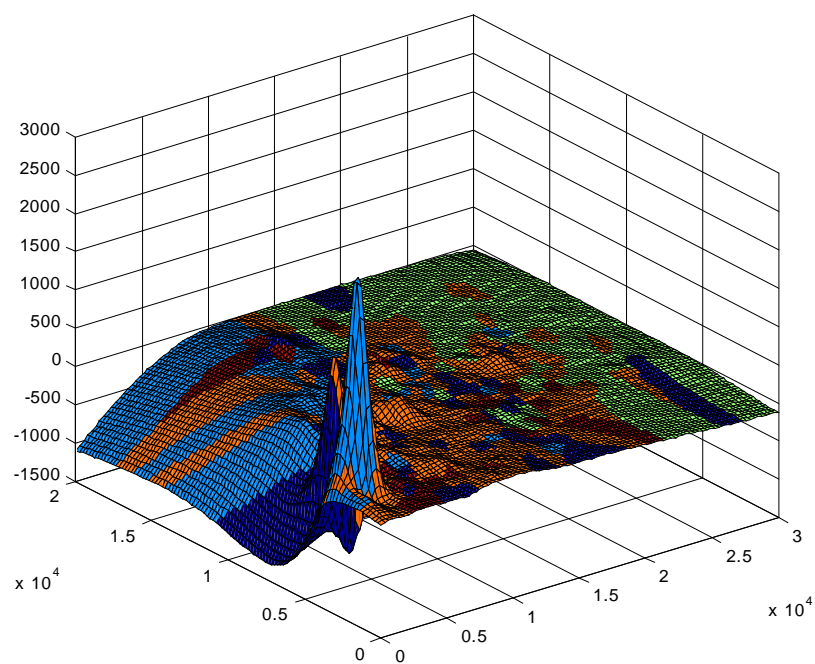


图5 铜（Cu）浓度与坐标x,y的曲面图

从空间分布图可看出，铜含量较高的基本都集中在两个小的区域内，从铜含量的地区分布及采样点的数据来看，其最高含量基本都出现在交通区和工业区，极少数出现在生活区、山区和公园绿地区。

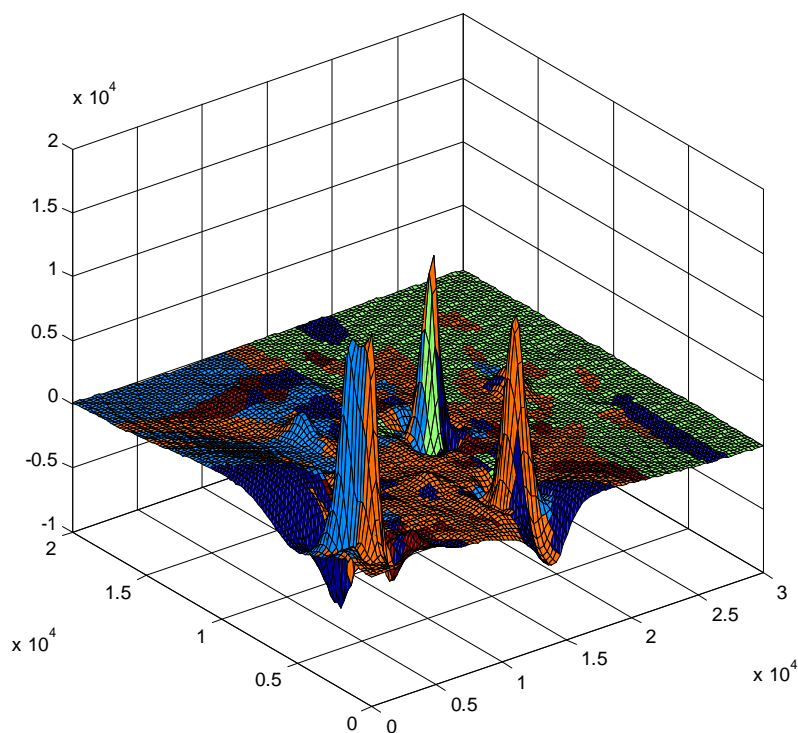


图6 汞 (Hg) 浓度与坐标 x,y 的曲面图

从空间分布图可看出，汞含量较高的基本都集中在三个小的区域内，从汞含量的地区分布及采样点的数据来看，其最高含量基本都出现在交通区和工业区，少数出现在生活区和公园绿地区，极少数出现在山区。

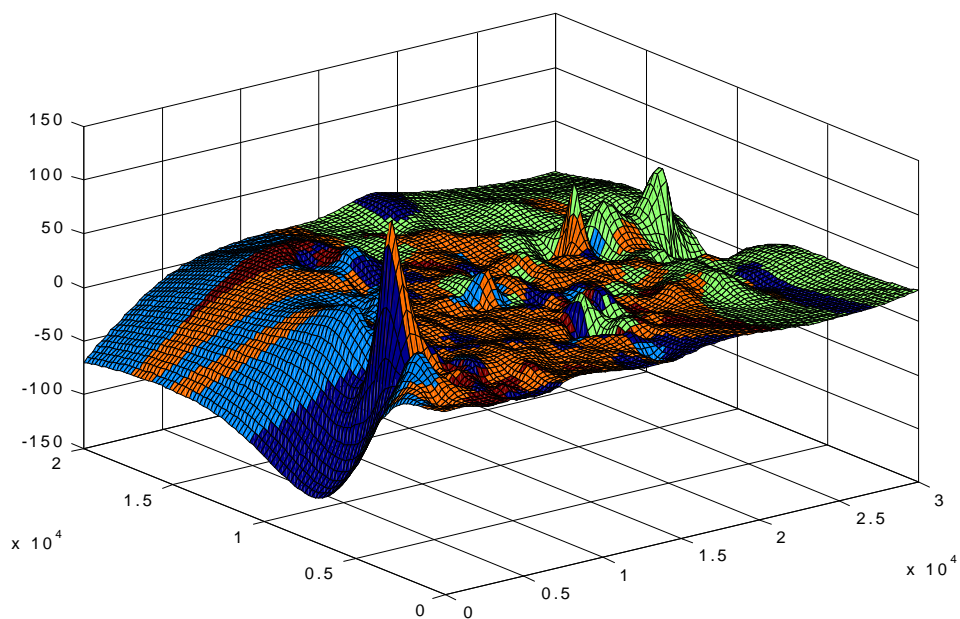


图7 镍 (Ni) 浓度与坐标x,y的曲面图

从空间分布图可看出，镍含量较高的基本都集中在一个小的区域内，从镍含量的地区分布和采样点数据来看，其最高含量基本都出现在交通区和工业区，少数出现在公园绿地区，极少数出现在山区和生活区。

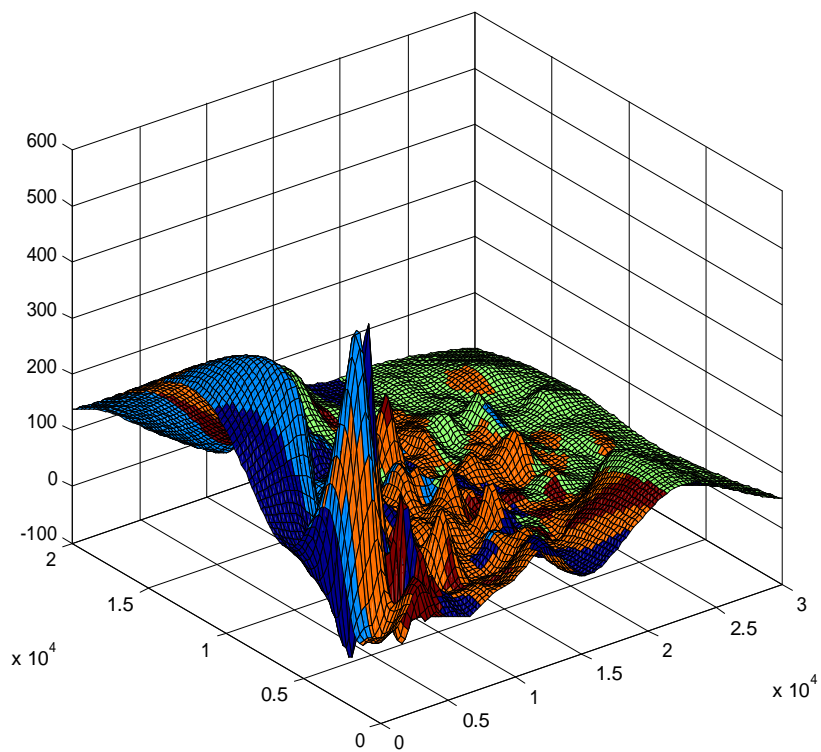


图8 铅 (Pb) 浓度与坐标x,y的曲面图

从空间分布图可看出，铅含量较高的不是很集中，但从铅含量的地区分布和采样点数据来看，其最高含量基本都出现在交通区和工业区，少数出现在公园绿地区，极少数出现在山区和生活区。

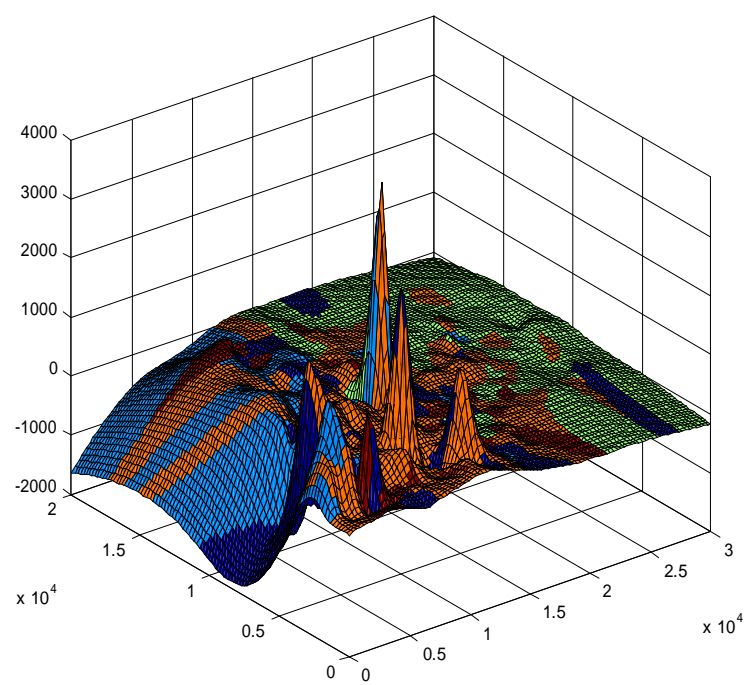


图9 锌（Zn）浓度与坐标x,y的曲面图

从空间分布图可看出，锌含量较高的基本都集中在几个小的区域内，从锌含量的地区分布和采样点的数据来看，其最高含量基本都出现在交通区和公园绿地区，少数出现在生活区和工业区，极少数出现在山区。

5.1.2 各功能区土壤重金属污染程度的评价

(1) 评价方法

a、评价因子

按照该地区的结合该区重金属的背景值，选取了在该区影响较大的 8 种重金属污染元素 (As, Cd, Cr, Cu, Hg, Ni, Pb, Zn) 作为该区的土壤环境质量评价因子。

b、评价标准

评价方法采用指数法。首先根据该区土壤重金属含量的实测值分别求出各污染因子的分指数，然后再根据分指数计算评价区域内土壤重金属的单因子综合污染指数和多因子综合污染指数 (见①、②和③)，最后根据分指数、单因子综合污染指数和多因子综合污染指数来判断土壤重金属污染现状。

表 1 8 种主要重金属元素的背景值

8种主要重金属元素的背景值			
元素	平均值	标准偏差	范围
As (μg/g)	3.6	0.9	1.8~5.4
Cd (ng/g)	130	30	70~190
Cr (μg/g)	31	9	13~49

Cu (μg/g)	13.2	3.6	6.0~20.4
Hg (ng/g)	35	8	19~51
Ni (μg/g)	12.3	3.8	4.7~19.9
Pb (μg/g)	31	6	19~43
Zn (μg/g)	69	14	41~97

a) 单因子污染指数：以土壤单项污染物的实测值与评价标准相比，比值即为分指数，用以表示土壤中该污染物的污染程度。 I_i 为第*i*种污染物的污染分指数， C_i 为其实测浓度， S_i 为其评价标准。

$$I_i = \frac{C_i}{S_i} \tag{①}$$

$$I = \frac{1}{n} \sum_{i=1}^n I_i \tag{②}$$

b) 单因子综合污染指数：区域内单个污染元素的所有分指数的算术平均值，见②式。

c) 多因子综合污染指数：由于土壤同时被多种重金属元素污染，因而土壤污染评价要运用综合指数法进行评价。综合指数的算法有多种，本文采用内梅罗(N. C. Nemerow)指数法来计算综合污染指数。见③式，其中为 P_N 土壤污染综合指数； $\max(I_i)^2$ 为土壤污染物中最大的污染分指数； \bar{I}_i 为各污染分指数的算术平均值。

$$P_N = \sqrt{\frac{\max(I_i)^2 + (\bar{I}_i)^2}{2}} \tag{③}$$

表2 土壤污染分级标准

等级划分	综合污染指数	单项污染指数	污染等级
1	$P_{综} \leq 0.7$	$P_i \leq 0.7$	安全
2	$0.7 < P_{综} \leq 1.0$	$0.7 < P_i \leq 1.0$	警戒级
3	$1.0 < P_{综} \leq 2.0$	$1.0 < P_i \leq 2.0$	轻度污染
4	$2.0 < P_{综} \leq 3.0$	$2.0 < P_i \leq 3.0$	中度污染
5	$P_{综} > 3.0$	$P_i > 3.0$	重度污染

d) 污染等级划分标准

综合污染指数全面反映了各污染物对土壤污染的不同程度，同时又突出高浓度对土壤环境质量的影响，因此用综合污染指数评定和划分土壤质量等级更为客观。本文结合此城市土壤重金属的含量情况和对作物危害的程度，按表2对该区土壤污染等级进行划分。

(2) . 结果与讨论

a、城市表层土壤重金属含量特征

表3 城市表层土壤重金属含量调查结果统计表

采样地区	样本 (个)	项目	As(μg/g)	Cd(μg/g)	Cr(μg/g)	Cu(μg/g)	Hg(μg/g)	Ni(μg/g)	Pb(μg/g)	Zn(μg/g)
生活区	44	最大值	11.45	1044.5	744.46	248.85	550	32.8	472.48	2893.47
		最小值	2.34	86.8	18.46	9.73	12	8.89	24.43	43.37
		平均值	6.27	289.96	69.02	49.4	93.04	18.34	69.11	237.01
工业区	36	最大值	21.87	1092.9	285.58	2528.48	13500	41.7	434.8	1626.02
		最小值	1.61	114.5	15.4	12.7	11.79	4.27	31.24	56.33

		平均值	7.25	393.11	53.41	127.54	642.36	19.81	93.04	277.93
山区	66	最大值	10.99	407.6	173.34	69.06	206.79	74.03	113.84	229.8
		最小值	1.77	40	16.2	2.29	9.64	5.51	19.68	32.86
		平均值	4.04	152.32	38.96	17.32	40.96	15.45	36.56	73.29
交通区	138	最大值	30.13	1619.8	920.84	1364.85	16000	142.5	181.48	3760.82
		最小值	1.61	50.1	15.32	12.84	8.57	6.19	22.01	40.92
		平均值	5.71	360.01	58.05	62.21	446.82	17.62	63.53	242.85
公园绿地区	35	最大值	11.68	1024.9	96.28	143.31	1339.29	29.1	227.4	1389.39
		最小值	2.77	97.2	16.31	9.04	10	7.6	26.89	37.14
		平均值	6.26	280.54	43.64	30.19	114.99	15.29	60.71	154.24

从表3可以看出，该城市表层土壤重金属的含量变化差异较大，As的含量范围在1.61~30.13ug/g之间，平均值为5.9ug/g，最高含量为最低含量的4.9倍，最高值出现在交通区；Cd的含量范围在40~1619.8ng / g之间，平均值为295.19ng/g，最高含量为最低含量的40.5倍，最高值出现在交通区；Cr的含量范围在15.32~920.84ug / g之间，平均值为52.6ug / g，最高含量为最低含量的60.1倍，最高值出现在交通区；Cu的含量范围在2.29~2528.28ug / g之间，平均值为57.33ug/g，最高含量为最低含量的1104倍，最高值出现在工业区；Hg的含量范围在8.57~ 13500.00ng / g之间，平均值为267.63ng / g，最高含量为最低含量的1575倍，最高值出现在工业区；Ni的含量范围在4.27~142.5ug / g之间，平均值为17.3ug / g，最高含量为最低含量的33.4倍，最高值出现在交通区；Pb的含量范围在19.68~ 472.48ug / g之间，平均值为64.59ug / g，最高含量为最低含量的24倍，最高值出现在生活区；Zn的含量范围在32.86~ 3760.82ug / g之间，平均值为197.06ug / g，最高含量为最低含量的114.45倍，最高值出现在交通区。

从各重金属元素平均含量的地区分布来看，As，Cd，Cr，Cu，Hg，Ni，Pb，Zn的最高含量分别出现在交通区，交通区，交通区，工业区，工业区，交通区，生活区，交通区，这与该城区各重金属元素单点最高值分布基本相符。其中，As的平均含量在各区的顺序为工业区>生活区>公园绿地区>交通区>山区；Cd的平均含量在各区的顺序为工业区>交通区>生活区>公园绿地区>山区；Cr的平均含量在各区的顺序为生活区>交通区>工业区>公园绿地区>山区；Cu的平均含量在各区的顺序为工业区>交通区>生活区>公园绿地区>山区；Hg的平均含量在各区的顺序为工业区>交通区>公园绿地区>生活区>山区；Ni的平均含量在各区的顺序为工业区>生活区>交通区>山区>公园绿地区；Pb的平均含量在各区的顺序为工业区>生活区>交通区>公园绿地区>山区；Zn的平均含量在各区的顺序为工业区>交通区>生活区>公园绿地区>山区。从这些先后顺序可以看出，交通区、工业区各重金属元素的含量相对较高，而山区、公园绿地区则相对偏低。

表4 城区表层土壤重金属污染评价指数

采样地区	样本（个）	单因子综合污染指数								多因子综合污染指数
		I(As)	I(Cd)	I(Cr)	I(Cu)	I(Hg)	I(Ni)	I(Pb)	I(Zn)	
生活区	44	1.74	2.23	2.23	3.74	2.66	1.49	2.23	3.43	3.97
工业区	36	2.01	3.02	1.72	9.66	18.35	1.61	3	4.02	16.96
山区	66	1.12	1.17	1.26	1.31	1.17	1.57	1.18	1.06	1.57
交通区	138	1.59	2.77	1.87	4.71	12.77	1.43	2.05	3.51	11.81
公园绿地区	35	1.74	2.16	1.4	2.29	3.29	1.24	1.96	2.23	3.42

从表4可以看出生活区多因子综合污染指数为3.97，根据表2的土壤污染分级标准，生活区污染等级应为重度污染。按此评判标准，山区的污染等级为轻度污染，公园绿地区、生活区、工业区和交通区都是重度污染，其中工业区>交通区>生活区>公园绿地区。

从单因子综合污染指数可判断出在生活区，砷、镍这两种重金属的污染等级为轻度；

镉、铬、汞、铅这四种重金属的污染等级为中度；铜、锌这两种重金属的污染等级为重度。在工业区，铬、镍的污染等级为轻度；砷的污染等级为中度；镉、铜、汞、铅、锌的污染等级为重度。在山区，砷、镉、铬、铜、汞、镍、铅、锌八种重金属都是轻度污染。在交通区，砷、铬、镍的污染等级为轻度；镉、铅的污染等级为中度；铜、汞、锌都是重度污染。在公园绿地区，砷、铬、镍、铅的污染等级为轻度；镉、铜、锌的污染等级为中度；汞是重度污染。

5.2、重金属污染的主要原因

模型首先从八个重金属在五个功能区的分布情况来考虑，采用主成分分析法对八种金属分别进行讨论。然后，再对各个功能区的污染情况进行分析，用因子分析法观察每个功能区中重金属的含量。

5.2.1主成分分析法

(1) 主成分分析法相关知识

主成分分析也称主分量分析，旨在利用降维的思想，把多指标转化为少数几个综合指标。主成分分析法是一种数学变换的方法，它把给定的一组相关变量通过线性变换转换成另一组不相关的变量，这些新的变量按照方差依次递减的顺序排列。在数学变换中保持变量的总方差不变，使第一变量具有最大的方差，称为第一主成分，第二变量的方差次大，并且和第一变量不相关，称为第二主成分。依次类推，I个变量就有I个主成分。

主成分分析法的计算步骤为：

a) 原始指标数据的标准化采集p维随机向量 $x = (x_1, x_2, \dots, x_p)^T$,

n个样品 $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$, $i=1, 2, \dots, n$, $n > p$, 构造样本阵，对样本阵元素进行如下标准化变换：

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, i = 1, 2, \dots, n; j = 1, 2, \dots, p$$

其中

$$\bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n}, s_j^2 = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}, \text{得标准化阵} Z.$$

b)、对标准化阵Z 求相关系数矩阵

$$R = [r_{ij}]_{p \times p} = \frac{Z^T Z}{n-1}, \text{其中 } r_{ij} = \frac{\sum z_{kj} \cdot z_{ki}}{n-1}, i, j = 1, 2, \dots, p$$

c)、解样本相关矩阵R 的特征方程 $|R - \lambda I_p| = 0$, 得p个特征根, 确定主成分

按 $\frac{\sum_{j=1}^m \lambda_j}{\sum_{j=1}^p \lambda_j} \geq 0.85$ 确定m值, 使信息的贡献率达85%以上, 对每个 λ_j , $j=1, 2, \dots, m$, 解方程

组 $Rb = \lambda_j b$ 得单位特征向量 b_j^0 。

d)、将标准化后的指标变量转换为主成分

$$U_{ij} = z_i^T b_j^0, j = 1, 2, \dots, m$$

U1称为第一主成分, U2 称为第二主成分, ..., Up 称为第p 主成分。

e)、对m个主成分进行综合评价

对m 个主成分进行加权求和, 即得最终评价价值, 权数为每个主成分的方差贡献率。

(2) 八种重金属元素的主成分分析结果

a、砷 (As)

表5.1 砷在各功能区的相关性分析及各功能区贡献率

	生活区	工业区	山区	交通区	公园绿地区
生活区	1.0000	0.2345	0.1194	0.1285	-0.4555
工业区		1.0000	0.5239	0.2237	-0.3600
山区			1.0000	0.4266	-0.1359
交通区				1.0000	-0.2389
公园绿地区					1.0000
贡献率	15.49%	43.08%	7.11%	11.31%	23.01%

输出给出了相关系数，如X2和X3呈现正相关，且相关系数高达0.5239。也给出了相关阵的特征值。特征值越大，它所对应的主成分变量包含的信息就越多。第1个至第5个主成分的贡献率分别为15.49%，43.08%，7.11%，11.31%和23.01%，由此数列值可知：第二个主成分的贡献率最大。

b、镉 (Cd)

表5.2 镉在各功能区的相关性分析及各功能区贡献率

	生活区	工业区	山区	交通区	公园绿地区
生活区	1.0000	0.4018	0.2035	-0.0098	0.1518
工业区		1.0000	-0.0989	-0.0266	-0.2781
山区			1.0000	0.1367	0.0279
交通区				1.0000	-0.0825
公园绿地区					1.0000
贡献率	22.63%	24.86%	7.66%	28.65%	16.20%

输出给出了原始变量的简单统计量的值、相关系数和相关阵的特征值。按八种重金属元素背景值判断五个区中的砷含量都超出了正常值，说明五个区都有砷污染，但程度不同。第1个至第五个主成分的贡献率分别为22.63%，24.68%，7.66%，28.65%和16.20%，由此数列值可知：第四个主成分的贡献率最大。

c、铬 (Cr)

表5.3 铬在各功能区的相关性分析及各功能区贡献率

	生活区	工业区	山区	交通区	公园绿地区
生活区	1.0000	0.1790	0.0804	0.0079	-0.0091
工业区		1.0000	0.0717	-0.0290	-0.1748
山区			1.0000	0.1850	-0.2257
交通区				1.0000	-0.1633
公园绿地区					1.0000
贡献率	23.10%	18.14%	13.72%	29.22%	15.83%

输出给出了原始变量的简单统计量的值、相关系数和相关阵的特征值。第1个至第五个主成分的贡献率分别为23.10%，18.14%，13.72%，29.22%和15.83%，由此数列值可知：第四个主成分的贡献率最大。

d、铜 (Cu)

表5.4 铜在各功能区的相关性分析及各功能区贡献率

	生活区	工业区	山区	交通区	公园绿地区
生活区	1.0000	0.3371	0.3783	0.6785	0.0103
工业区		1.0000	0.0212	-0.0214	-0.0743
山区			1.0000	0.0188	0.0577
交通区				1.0000	0.0496
公园绿地区					1.0000
贡献率	19.94%	37.04%	3.12%	22.01%	17.88%

输出给出了原始变量的简单统计量的值、相关系数和相关阵的特征值。第1个至第五个主成分的贡献率分别为19.94%，37.04%，3.12%，22.01%和17.88%，由此数列值可知：第二个主成分的贡献率最大。

e、汞 (Hg)

表5.5 汞在各功能区的相关性分析及各功能区贡献率

	生活区	工业区	山区	交通区	公园绿地区
生活区	1. 0000	0.2345	0.1194	0.1285	-0.4555
工业区		1.0000	0.5239	0.2237	-0.3600
山区			1.0000	0.4266	-0.1359
交通区				1.0000	-0.2389
公园绿地区					1.0000
贡献率	19.17%	21.80%	11.97%	27.15%	19.91%

输出给出了原始变量的简单统计量的值、相关系数和相关阵的特征值。第1个至第五个主成分的贡献率分别为19.17%，21.80%，11.97%，27.15%和19.91%，由此数列值可知：第四个主成分的贡献率最大。

f、镍 (Ni)

表5.6 镍在各功能区的相关性分析及各功能区贡献率

	生活区	工业区	山区	交通区	公园绿地区
生活区	1. 0000	0.1021	0.1578	0.0312	-0.0167
工业区		1.0000	-0.0424	0.1602	-0.4772
山区			1.0000	0.1457	-0.1537
交通区				1.0000	-0.1374
公园绿地区					1.0000
贡献率	22.76%	19.10%	16.74%	32.20%	9.20%

输出给出了原始变量的简单统计量的值、相关系数和相关阵的特征值。第1个至第五个主成分的贡献率分别为22.76%，19.10%，16.74%，32.20%和9.20%，由此数列值可知：第四个主成分的贡献率最大。

g、铅 (Pb)

表5.7 铅在各功能区的相关性分析及各功能区贡献率

	生活区	工业区	山区	交通区	公园绿地区
生活区	1. 0000	0.5655	0.0704	0.0362	0.0919
工业区		1.0000	-0.1112	0.0171	0.0423
山区			1.0000	0.0064	-0.1047
交通区				1.0000	0.2574
公园绿地区					1.0000
贡献率	20.60%	31.96%	8.02%	24.90%	14.52%

输出给出了原始变量的简单统计量的值、相关系数和相关阵的特征值。第1个至第五个主成分的贡献率分别为20.60%，31.96%，8.02%，24.90%和14.52%，由此数列值可知：第二个主成分的贡献率最大。

h、锌 (Zn)

表5.8 锌在各功能区的相关性分析及各功能区贡献率

	生活区	工业区	山区	交通区	公园绿地区
生活区	1. 0000	0.1950	-0.0296	-0.0044	0.8641
工业区		1.0000	-0.0526	-0.0425	-0.0561
山区			1.0000	-0.0491	-0.1124
交通区				1.0000	-0.0173
公园绿地区					1.0000

贡献率	20.76%	21.08%	8.02%	37.76%	18.49%
-----	--------	--------	-------	--------	--------

输出给出了原始变量的简单统计量的值、相关系数和相关阵的特征值。第1个至第五个主成分的贡献率分别为20.76%，21.08%，8.02%，37.76%和18.49%，由此数列值可知：第四个主成分的贡献率最大。

(3) 综合分析

由上述八个重金属元素的主成分分析可知，工业区和交通区对这八种重金属元素的贡献率是最大的，生活区对这八种重金属元素的贡献率稍低一点，公园绿地区和山区对其贡献率最低。由此，我们可推断重金属污染的最主要的原因是工业生产过程中的废气、废水和固体排放物以及各种交通工具在行驶过程中排放的污染物，如烟尘、氮氧化物、一氧化碳、二氧化硫、碳氢化合物、铅化合物、石油和石油制品以及有害有毒运输品。另外，工业区和交通场所排放的生活垃圾也可能是一个因素。造成重金属污染的主要原因居民区的生活垃圾，特别是废电池、废日光灯管、废水银温度计、过期药品等有害垃圾。

5.2.2 因子分析法

(1) 因子分析法相关知识

因子分析的基本目的就是用少数几个因子去描述许多指标或因素之间的联系，即将相关比较密切的几个变量归在同一类中，每一类变量就成为一个因子（之所以称其为因子，是因为它是不可观测的，即不是具体的变量），以较少的几个因子反映原资料的大部分信息。因子分析法(Factor Analysis)就是寻找这些公共因子的模型分析方法，它是在主成分的基础上构筑若干意义较为明确的公因子，以它们为框架分解原变量，以此考察原变量间的联系与区别。

因子分析法的步骤为：

- 输入原始数据 $X_{n \times p}$ ，计算样本均值和方差，进行标准化计算；
- 求样本相关系数矩阵 $R = (r_{ij})_{p \times p}$ ；
- 求样本相关矩阵的特征根 $\lambda_i (\lambda_1, \lambda_2, \dots, \lambda_p > 0)$ 和相应的标准正交特征向量 I_i ；
- 确定公共因子数
- 计算公共因子的共性方差 h_i^2 ；
- 对载荷矩阵进行旋转，以求能更好地解释公共因子；
- 对公共因子作出专业性的解释

因子分析的常用旋转法有方差最大正交预旋转和斜交旋转，方差最大正交预旋使公共因子的相对负荷的方差之和最大，且保持原公共因子的正交性和公共方差总和不变，可使每个因子上的具有最大载荷的变量数最小，因此可以简化对因子的解释。因子斜交旋转后，各因子负荷发生了较大变化，出现了两极分化。各因子间不再相互独立，而彼此相关。各因子对各变量的贡献的总和也发生了改变。

在 factor 过程语句中没有规定提取因子的方法。规定选项 rotate=p 要求产生方差最大正交预旋转后进行斜交旋转。选项 reorder 要求变量按它们的最大因子载荷重新排序。

(2) 各区域的因子分析结果

表6 各区域的因子结构

	生活区		工业区		山区		交通区		公园绿地区	
	金属代号	载荷系数	金属代号	载荷系数	金属代号	载荷系数	金属代号	载荷系数	金属代号	载荷系数
主要污染因子	Y2	0.42459	Y5	0.35090	Y6	0.41168	Y3	0.30880	Y7	0.30565
	Y7	0.38182	Y4	0.33948			Y4	0.30673	Y4	0.30245
	Y5	0.36442	Y3	0.30782	Y3	0.40199	Y6	0.29895	Y8	0.29924
							Y8	0.11433	Y2	0.25126
次要污染因子	Y1	0.46287	Y2	0.49396	Y2	0.38725	Y5	0.49164	Y1	0.38715
	Y6	0.40180	Y7	0.38020	Y7	0.37428	Y7	0.40786	Y6	0.37597
	Y4	0.28614	Y8	0.28703	Y8	0.24530	Y2	0.38442	Y3	0.32529
弱污染因子	Y8	0.58639	Y1	0.57387	Y4	0.48923	Y1	0.92895	Y5	0.85768
					Y5	0.41510				
	Y3	0.46769	Y6	0.44283	Y1	0.39643				

由表6输出的因子结构可以看出，生活区主要污染因子与Y2、Y7、Y5有大于0.8的载荷系数绝对值，次要污染因子与Y1、Y6、Y4有大于0.8的载荷系数绝对值，弱污染因子与Y8、Y3有大于0.9的载荷系数绝对值。工业区主要污染因子与Y5、Y4、Y3有大于0.8的载荷系数绝对值，次要污染因子与Y2、Y7、Y8有大于0.8的载荷系数绝对值，弱污染因子与Y1、Y6有大于0.9的载荷系数绝对值。山区主要污染因子与Y6、Y3有大于0.8的载荷系数绝对值，次要污染因子与Y2、Y7、Y8有大于0.8的载荷系数绝对值，弱污染因子与Y4、Y5、Y1有大于0.9的载荷系数绝对值。交通区主要污染因子与Y3、Y4、Y6、Y8有大于0.8的载荷系数绝对值，次要污染因子与Y5、Y7、Y2有大于0.8的载荷系数绝对值，弱污染因子与Y1有大于0.9的载荷系数绝对值。公园绿地区主要污染因子与Y7、Y4、Y8、Y2有大于0.8的载荷系数绝对值，次要污染因子与Y1、Y6、Y3有大于0.8的载荷系数绝对值，弱污染因子与Y5有大于0.8的载荷系数绝对值。

从表6中可分析除土壤中重金属污染物来源与分布是多途径的，主要可分为三个方面。

1) 大气沉降：污染物来源主要为能源、运输、冶金和建筑材料生产产生的气体和粉尘。例如，汽车运输对土壤造成的污染主要以铅、锌、镉、铬、铜等的污染为主，它们主要来自于含铅汽油的燃烧和汽车轮胎磨损产生的粉尘。除汞以外，重金属基本是以气溶胶的形态进入大气，经过自然沉降和降水进入土壤。大气汞通过干湿沉降进入土壤后，被土壤中的粘土矿物和有机物的吸附或固定，富集于土壤表层，或为植物吸收而转入土壤，造成土壤汞的浓度的升高。

2) 污水排放：城市污水包括生活污水、商业污水和工业废水。由于城市工业化的迅速发展，大量的工业废水涌入河道，使城市污水中含有的许多重金属离子，随着污水灌溉而进入土壤。在分布上，往往是靠近污染源头和城市工业区土壤污染严重，远离污染源头和城市工业区，土壤几乎不污染。污灌导致土壤重金属Hg、Cd、Cr、As、Cu、Zn、Pb等含量的增加。污水中的汞被土壤矿质胶体和有机质迅速吸附，一般累积在土壤表层。污水中的镉进入土壤后被粘土矿物吸附，而砷很容易被水中的悬浮物吸附，水中镉的含

量随着距排污口距离的增加而迅速下降，污染的范围较少。铅很容易被土壤有机质和粘土矿物吸附，迁移性弱，其累计分布特点是离污染源近土壤含量高。另外，金属矿山的开采、冶炼、重金属尾矿、冶炼废渣和矿渣堆放等，可以被酸溶出含重金属离子的矿山酸性废水，随着矿山排水和降雨使之带入水环境(如河流等)或直接进入土壤，都可以间接或直接地造成土壤重金属污染，特别是Hg, Cd, As, Pb、Cu、Cr含量高。矿山酸性废水重金属污染的范围一般在矿山的周围或河流的下游，在河流中不同河段的重金属污染往往受污染源(矿山)控制，河流同一污染源的下段自上游到下游, 由于金属元素迁移能力减弱和水体自净化能力的适度恢复，金属化学污染强度逐渐降低。

3) 固体废弃物：矿业和工业固体废弃物污染最严重，其次，动物粪便进入土壤可增加了土壤中Zn等的含量。还有，污水处理厂产生的污泥重金属含量高，而土壤又成为其处理的主要场所。污泥中Cd、Hg、Cr、Cu、Zn、Ni、Pb、砷含量易超过控制标准。污染的范围一般以废弃堆为中心向四周扩散。城市生活垃圾场及车辆废弃场附近土壤中的重金属Cd、Hg、Cr、Cu、Zn、Ni、Pb、As、Sb、V、Co、Mn的含量高于当地土壤背景值，重金属在土壤中的含量和形态分布特征受其垃圾中释放率的影响，且随距离的加大重金属的含量而降低。由于废弃物种类不同，各重金属污染程度也不尽相同。

同一区域土壤中重金属污染物的来源途径可以是单一的，也可以是多途径的。

5.3 重金属污染物传播特征及污染源位置

1) 重金属污染物传播特征

虽然污染物在进入不同的环境介质之后做着复杂的运动、变化，但都是由一下几种基本形式组成的：随着介质的迁移运动、污染物的分散运动、污染物的累积与转化、污染物被环境介质吸收或吸附、污染物的沉淀。

2) 回归分析的基本原理

回归分析(regression analysis)是研究一个变量关于另一个(些)变量的具体依赖关系的计算方法和理论。 从一组样本数据出发，确定变量之间的数学关系式对这些关系式的可信程度进行各种统计检验，并从影响某一特定变量的诸多变量中找出哪些变量的影响显著，哪些不显著。利用所求的关系式，根据一个或几个变量的取值来预测或控制另一个特定变量的取值，并给出这种预测或控制的精确程度。其用意：在于通过后者的已知或设定值，去估计和(或)预测前者的(总体)均值。

3) 回归分析的数学模型：

相关变量之间的关系可以是线性的，也可以是非线性的。这里只讨论多元非线性回归中纯二次二项式回归的数学模型一般可写为：

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \sum_{j=1}^m \beta_{jj} x_j^2$$

4)使用matlab软件对数据进行回归分析，结果如下：

As浓度与空间三维坐标的关系：

$$Z=8.3767-8.0123x+1.1972y-4.5233z+4.0439x^2+0.1195y^2+0.8855z^2$$

相应污染源位置：(1878.2634 , 6003.7263 , 4.5846)

Cd浓度与空间三维坐标的关系：

$$Z=448.4121-516.4672x+420.2077y-302.9962z+151.7293x^2-389.3995y^2+373.6502z^2$$

相应污染源的位置为：(970.5835 , 3946.7518 , 6.5891)

Cr浓度与空间三维坐标的关系：

$$Z=103.7937-272.9697x+146.6641y-40.4686z+189.3841x^2-137.1496y^2+55.9338z^2$$

相应污染源的位置为：(135.1956 , 2658.3427 , 8.5402)

Cu浓度与空间三维坐标的关系：

$$Z=175.9051-618.9608x+281.6080y-103.1799z+440.3064x^2-258.0317y^2+136.8568z^2$$

相应污染源的位置为：(138.4682 , 6223.4521 , 3.2461)

Hg浓度与空间三维坐标的关系：

$$Z=1220.9-2366.4x-755.1y-744.0z+1785.8x^2+432.8y^2+672.4z^2$$

相应污染源的位置为：(1231.5782 , 2561.5483 , 5.2478)

Ni浓度与空间三维坐标的关系：

$$Z=22.1605-26.9739x+18.3643y-7.2216z+16.7898x^2-16.1063y^2+5.7843z^2$$

相应污染源的位置为：(12234.2587 , 5865.1656 , 23.2461)

Pb浓度与空间三维坐标的关系：

$$Z=122.2430-268.9077x+109.5923y-12.5899z+159.2630x^2-97.5740y^2+26.8362z^2$$

相应污染源的位置为：(2310.6891 , 14145.2674 , 3.2651)

Zn浓度与空间三维坐标的关系：

$$Z=395.7462-881.9409x+667.1951y-166.2394z+420.3654x^2-684.8832y^2+322.4877z^2$$

相应污染源的位置为：(3015.4341 , 8642.2365 , 5.0543)

由上面的数据可知污染源位置误差较大，我们认为这是由于采集的点有奇异点和较多的比背景值小的数据，所以回归曲线方程效果较差，程序还有待于改进。

5.4 地质环境的演变模式研究

本题中所给附件数据包括采样点的位置、海拔高度及其所属功能区、8种主要重金属元素在采样点处的浓度和相应的背景值。为了更好地研究地质环境的演变模式，所给数据太少不足以做这方面的研究，所以我们可以考虑收集如下信息：

1) 土壤溶液的PH

就决定土壤中重金属的生物有效性而言，土壤的pH比土壤矿物学更具重要地位。土壤溶液的pH不仅决定了各种土壤矿物的溶解度，而且影响着土壤溶液中各种离子在固相上的吸附程度。首先，随土壤溶液pH升高，各种重金属元素在土壤固相上的吸附量和吸附能力加强。对酸性砂土中Cd的吸附现象进行研究，发现pH每增加0.5个单位Cd的吸附就增加一倍；廖敏等人的研究则表明，随pH的升高，Cd的吸附量和吸附能力急剧上升，最终发生沉淀。其次，土壤溶液的pH影响土壤溶液中重金属元素离子活度。当土壤溶液的pH由3.9升高到6.6时，溶液中的有机铜由30%上升到99%还多，极大的降低了 Cu^{2+} 的活度。当土壤溶液的pH由较低变为近中性时，溶液中的有机Pb也急剧增高。研究也证明了pH是决定 Cu^{2+} 和 Pb^{2+} 的重要因素之一。最后，土壤溶液pH对土壤中重金属元素的生物有效性影响可能不是单一的递增关系。经过对水-土壤体系中pH对Cd生物有效性影响的研究，在pH6以下镉的生物有效性随pH的升高而增加，而在pH6以上镉生物有效性则随pH升高而降低。一定pH范围内，植物Radish叶子中的重金属含量与土壤溶液pH呈正相关关系。然而大多数的研究都表明植物中的重金属含量与土壤溶液pH呈负相关关系。

2) 地下水重金属的浓度

地下水重金属污染是污染物在地下水系统中进行各种物理、化学和生物过程的结果。在一定条件下，土壤中的重金属可随水的移动而发生迁移，从而威胁人类的健康与其它动物的繁衍生息，因此积累在土壤中重金属的移动性受到人们广泛关注。一般认为，新土层对重金属有一定的吸附能力，重金属在土壤中的移动性很有限，通过田间实地采集剖面土样的分析结果也表明，重金属主要积累在表层。土壤重金属的生物毒性不仅与其总量有关，更大程度上由其形态分布所决定，不同的形态产生不同的环境效应，直接影响到重金属的毒性、迁移及在自然界的循环。

3) 植物中重金属浓度

植物对金属的吸收主要取决于植物从土壤中吸取金属以及向地上部运输金属的能力。植物对金属的吸收主要取决于自由态离子活度。许多螯合剂能诱导植物对重金属的吸收。金属离子在液泡中的区域化分布是植物耐重金属的主要原因。同时,细胞内的金属硫蛋白、植物螯合肽等蛋白质以及有机酸、氨基酸等在金属贮存和解毒方面也起重要作用。所以重金属浓度与植物体内中重金属浓度有正相关性,通过植物中重金属浓度来评价污染程度。

根据以上所收集的数据,先用指数法分别评价金属元素在土壤、PH、地下水、植物中的污染程度,然后采用层次分析法综合评价土壤污染程度,以此说明地址环境的演变模式。

层次分析法,是指将一个复杂的多目标决策问题作为一个系统,将目标分解为多个目标或准则,进而分解为多指标(或准则、约束)的若干层次,通过定性指标模糊量化方法算出层次单排序(权数)和总排序,以作为目标(多指标)、多方案优化决策的系统方法。

分析系统中各因素间的关系,对同一层次各元素关于上一层次中某一准则的重要性进行两两比较,构造两两比较的判断矩阵;由判断矩阵计算被比较元素对于该准则的相对权重,并进行判断矩阵的一致性检验;计算各层次对于系统的总排序权重,并进行排序,得到各方案对于总目标的总排序。

层次分析法的一个重要特点就是用两两重要性程度之比的形式表示出两个方案的相应重要性程度等级。如对某一准则,对其下的个方案进行两两对比,并按其重要性程度评定等级。记为第*i*和第*j*个因素的重要性之比,表7列出了Saaty给出的9个重要性等级及其赋值。按两两比较结果构成的矩阵称作判断矩阵。

表7 9个重要性等级及其赋值

因子比	量化值
同等重要	1
稍微重要	3
较强重要	5
强烈重要	7
极端重要	9
两相邻判断的中间值	2, 4, 6, 8

对判断矩阵的一致性指标检验的步骤如下:

(1) 计算一致性指标:

$$CI = \frac{\lambda_{\max} - n}{n - 1};$$

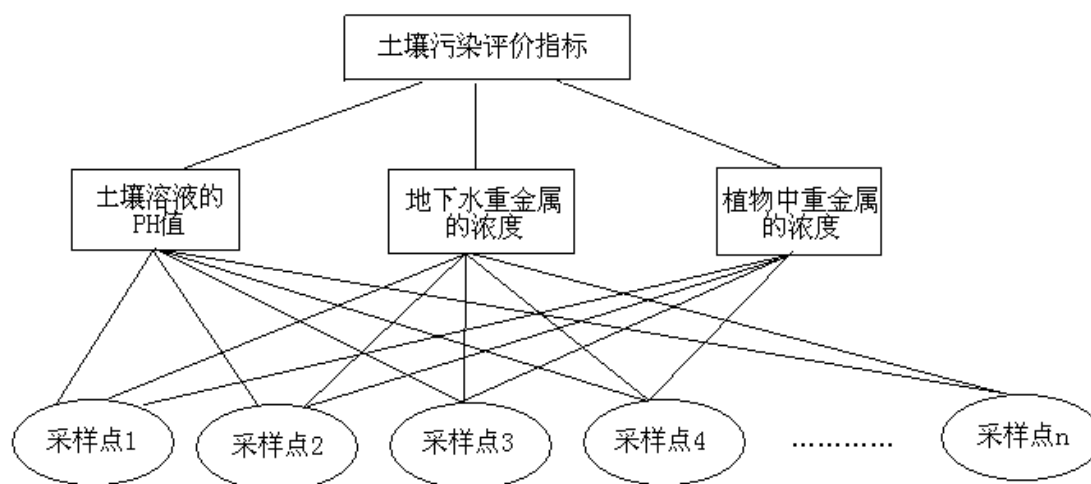
(2) 查找相应的平均随机一致性指标 *RI*。对 $n=1, \dots, 9$, Saaty给出了 *RI* 的值,如表8所示。

表8 一致性指标

矩阵阶数	3	4	5	6	7	8	9
RI	0.5149	0.8931	1.1185	1.2494	1.345	1.42	1.4616

(3) 计算一致性比例CR

$$CR = \frac{CI}{RI}$$



查阅历史文献，对比研究即可得出不同历史时间段土壤污染程度，以此可以说明城市的地质环境变化模式。

六、模型评价

模型的优缺点

优点：1) 问题一中通过作出浓度与坐标的分布图，可以很直观的看出各重金属在不同位置的浓度不同，可以定性的分析金属聚集区；

2) 主成分分析、因子分析可在众多因素中找出各个变量最佳的子集合，从子集合所包含的信息描述多变量的系统结果及各个因子对系统的影响。抓住主要矛盾，把握主要矛盾的主要方面，舍弃次要因素，以简化系统的结构，认识系统的内核。主成分分析和因子可以进行数值分类，构造分类模式。以便找出它们之间的联系和内在规律性。所以问题二中我们采用主成分分析、因子分析可以准确的定量说明各种重金属在各功能区的污染程度和各功能区受不同重金属的污染；

3) 问题三我们采用回归模型，简单易懂，便于实现。

缺点：

1) 没有考虑到海拔考虑到这些模型中；

2) 问题三中回归模型不能很好的反映重金属浓度与地理位置的关系。

七、参考文献

- [1]雷功炎, 数学模型讲义, 北京大学出版社, 2000 年。
- [2]陈超, MATLAB 应用实例精讲, 电子工业出版社, 2010 年。
- [3]曹金亮, 应用模糊综合评判方法对太原、晋中、阳泉地区地质环境质量评价, 《华北地质矿产杂志》, 第 13 卷第 3 期: 1-2, 1998 年 9 月。
- [4]史贵涛, 上海城市公园土壤及灰尘中重金属污染特征, 《环境科学》, 第 28 卷第 2 期: 1-5, 2001 年 2 月。
- [5]梅长林, 范金城等, 数据分析方法, 高等教育出版社, 2006 年。
- [6]姜芝萍, 杨俊衡, 城市重点污染区土壤重金属污染评价标准探讨 --以衡阳市某区为例, 《安全与环境工程》, 第 17 卷第 1 期: 1-5, 2010 年 6 月。
- [7]王海东 方凤满 谢宏芳 王翔 黄春海, 芜湖市区土壤重金属污染评价及来源分析, 《城市环境与城市生态》, 第 23 卷第 4 期: 2-5, 2010 年 8 月。
- [8]邹芬生, 董军义, 空间数据建模中的内插值方法讨论, 《中国钨业》, 第 19 卷第 4 期: 1-5, 2004 年 8 月。
- [9]张云, 张宇峰, 胡忻, 南京不同功能区街道路面积尘重金属污染评价与源分析, 《环境科学研究》, 第23卷第11期: 1-6, 2010年10月。

附录

第一题程序:

```
1.1 clear,clc
a=load('浓度.txt');
b=load('坐标及功能区.txt');
x=b(:,1);
y=b(:,2);
z=b(:,3);
quyu=b(:,4);
[x1,y1]=meshgrid(0:300:30000,0:200:20000);
z1=griddata(x,y,z,x1,y1,'v4');
[m1,n1]=size(z1);
c=zeros(m1,n1);
n=length(x);
for k2=1:m1
    for k3=1:n1
        nn0=1;
        mind=inf;
        for k1=1:n
            if norm([x1(k2,k3)-x(k1),y1(k2,k3)-y(k1)])<=mind
                nn0=k1;
                mind=norm([x1(k2,k3)-x(k1),y1(k2,k3)-y(k1)]);
            end
        end
        c(k2,k3)=quyu(nn0);
    end
end
figure(1)
surf(x1,y1,z1,c) %绘制曲面图
hold on;
plot3(x,y,z,'w.')
hold off;
figure(2)
contour(x1,y1,z1,30) %绘制等高线
1.2 clc;
clear;
data=load('data.txt'); %导入原始数据
x=data(:,1);
y=data(:,2);
z=data(:,7); %As(μg/g)浓度数据,其它7种重金属元素数据只要将5分别换成6~12即可
fenqu=data(:,4);
[x1,y1]=meshgrid(0:300:30000,0:200:20000);
z1=griddata(x,y,z,x1,y1,'v4');
[m1,n1]=size(z1);
c=zeros(m1,n1);
n=length(x);
for k1=1:m1
    for k2=1:n1
        k=1;
        mind=inf;
        for k3=1:n
            if norm([x1(k1,k2)-x(k3),y1(k1,k2)-y(k3)])<=mind
                k=k3;
                mind=norm([x1(k1,k2)-x(k3),y1(k1,k2)-y(k3)]);
            end
        end
        c(k1,k2)=fenqu(k);
    end
end
```

```

        c(k1,k2)=fenqu(k);
    end
end
surf(x1,y1,z1,c) %绘制曲面图
hold on;
b = bar(rand(10,5),'stacked'); colormap;
legend(b,'生活区','工业区','山区','交通区','公园绿地区');
hold off

```

第二题程序：

2.1 主成分分析

砷的主成分分析程序：

```

data As;
input X1-X5 @@;
cards;
9.17    6.56    4.09    7.84    5.93
5.72    14.08   2.72    5.93    3.3
11.45   9.62    3.69    4.93    3.69
7.84    21.87   6.14    6.35    4.49
8.518.38 10.99   8.94    3.69
5.51    10.53   7.41    7.41    7.41
9.39    2.34    4.29    8.72    5.72
4.09    9.35    7.63    4.49    8.06
6.35    4.79    2.91    5.51    5.72
3.51.61   4.69    6.14    8.72
4.29    4.79    6.35    7.41    6.77
7.41    8.67    5.19.84   6.56
2.91    7.12    4.69    6.14    6.77
3.34.58   3.55.31   6.98
6.14    8.23    4.49    3.69    6.14
8.06    4.58    3.33.56.69
4.69    8.23    2.91    5.51    3.37
2.34    3.97    4.09    4.49    6.91
6.56    10.74   4.95.51   5.62
8.23    3.77    4.94.29   2.77
8.910.27 4.09    6.56    3.17
5.41    8.95.93   16.58   4.79
7.78    3.37    2.91    5.93    5.83
5.62    6.69    2.72    4.69    7.34
6.05    5    2.34    4.98.23
4.17    9.58    2.53    5.31    10.74
6.26    3.17    3.89    4.29    11.68
5    3.77    2.34    5.51    7.34
4.58    3.37    2.34    4.69    6.05
5.41    6.05    2.72    7.25.41
7.56    3.77    3.11    5.31    6.26
5.41    5.41    3.54.96.47
5.83    8.23    2.72    4.96.47
5.25.41   1.77    3.89    4.79
5.27.78   3.53.69   9.35
4.38    6.47    2.53    3.11    .
5.41    .    9.17    3.89    .
8.67    .    2    3.89
6.47    .    2.38    2.91    .
7.12    .    2.77    3.3    .
10.97   .    5.62    4.9    .
9.81    .    2.97    2.72    .

```

2.77	.	4.58	3.11	.
6.47	.	6.05	3.3	.
.	.	2.77	3.11	.
.	.	2.38	3.69	.
.	.	5.2	3.5	.
.	.	7.34	8.5	.
.	.	3.77	1.77	.
.	.	3.57	2.53	.
.	.	4.38	6.35	.
.	.	3.97	30.13	.
.	.	8	3.89	.
.	.	3.57	2.91	.
.	.	3.17	1.96	.
.	.	2.97	6.98	.
.	.	3.97	2.91	.
.	.	3.17	5.93	.
.	.	2.77	5.93	.
.	.	3.17	4.29	.
.	.	2.97	5.72	.
.	.	3.17	6.77	.
.	.	1.8	4.49	.
.	.	3.17	3.11	.
.	.	4.79	2.91	.
.	.	3.57	5.93	.
.	.	.	5.72	.
.	.	.	4.69	.
.	.	.	5.72	.
.	.	.	1.96	.
.	.	.	8	.
.	.	.	6.69	.
.	.	.	3.77	.
.	.	.	5.41	.
.	.	.	4.58	.
.	.	.	6.91	.
.	.	.	5	.
.	.	.	5.62	.
.	.	.	6.91	.
.	.	.	7.78	.
.	.	.	6.26	.
.	.	.	7.56	.
.	.	.	4.79	.
.	.	.	2.77	.
.	.	.	6.26	.
.	.	.	7.34	.
.	.	.	4.17	.
.	.	.	5	.
.	.	.	7.56	.
.	.	.	5.62	.
.	.	.	23.72	.
.	.	.	6.47	.
.	.	.	5	.
.	.	.	4.17	.
.	.	.	3.57	.
.	.	.	6.47	.
.	.	.	3.37	.
.	.	.	8.67	.
.	.	.	4.38	.
.	.	.	5.41	.


```
. . . 5.62 .
. . . 6.69 .
. . . 7.56 .
. . . 1.61 .
. . . 5.2 .
. . . 5 .
. . . 7.56 .
. . . 5.2 .
. . . 5.41 .
. . . 6.05 .
. . . 4.17 .
. . . 5.2 .
. . . 6.47 .
. . . 6.47 .
. . . 4.17 .
. . . 8.23 .
. . . 4.38 .
. . . 5.2 .
. . . 5.41 .
. . . 9.13 .
. . . 5.62 .
. . . 7.34 .
. . . 3.57 .
. . . 4.38 .
. . . 5.62 .
. . . 6.26 .
. . . 5.2 .
. . . 6.91 .
. . . 4.58 .
. . . 5.62 .
. . . 6.91 .
. . . 7.12 .
. . . 4.38 .
. . . 3.37 .
. . . 7.12 .
. . . 5.62 .
. . . 3.77 .
. . . 7.56 .
```

```
;
proc princomp data=As prefix=Z out=O91;
var X1-X5;
run;
（其它七种重金属元素与其类似，只需修改数据行即可）
```

2.2 因子分析

生活区因子分析程序：

```
data ch21;
input number X1-X8 @@;
cards;
12 9.17 287 43.94 45.77 168 19.7 62.74 223.16
13 5.72 193.7 80.35 26.57 111 19.8 57.64 89.08
16 11.45 1044.5 94.78 136.97 202 22.3 472.48 602.04
18 7.84 347.9 57.65 97.14 213 19.6 70.82 307.24
20 8.5 614 744.46 130.55 156 32.8 228.64 1013.47
21 5.51 257.2 54.64 29.01 104 13.2 87.68 223.27
23 9.39 325.8 172.29 104.89 82 31.5 90.9 429.29
25 4.09 90.5 35.02 11.82 16 10.4 29.09 46.84
33 6.35 532 57.51 83.76 191 19.5 73.46 297.14
36 3.5 396.3 138.37 58.97 170 24.2 91.76 2893.47
```

```

39 4.29 449.1 67.22 51.64 315 15.7 106.97 294.69
42 7.41 337.3 77.27 248.85 90 20.1 99.58 210
63 2.91 265 35.66 29.39 24.64 9.23 60.54 122.96
67 3.3 488 51.18 34.55 37.5 10.8 54.62 125.92
68 6.14 227 42.15 67.04 49.29 16.31 34.28 82.96
72 8.06 113.1 52.4 20.81 65.36 19.69 29.56 62.24
94 4.69 132.8 36.11 17.28 20.36 15.53 37.03 82.76
106 2.34 87.6 18.46 9.73 13.93 8.89 24.43 43.37
107 6.56 245.6 36.73 61.3 55.71 14.18 47.24 114.29
152 8.23 756.4 42.73 87.52 63 19.26 88.74 184.69
154 8.9 307.3 54.39 57.21 326 25.72 131.93 256.94
156 5.41 178.9 29.54 23.73 52 9.89 49.84 118.88
157 7.78 315.5 49.76 28.03 550 18.95 45.73 109.29
158 5.62 134.6 25.33 19.1 45 11.66 40.5 87.14
176 6.05 365 35.92 30.91 110 17.91 44.75 147.35
180 4.17 310.7 40.7 42.64 58 14.99 115.59 177.76
183 6.26 387.2 38.03 30.06 85 21.97 54.98 142.55
184 5 196.5 50.03 18.56 29 25.82 33.9 84.8
186 4.58 129 31.09 18.93 38 15.1 29.76 69.8
187 5.41 204.9 40.16 25.86 41 16.24 33.28 80
195 7.56 206.6 55.79 24.44 22 28.63 29.73 81.73
198 5.41 250.1 34.32 16.25 15 18.53 41.29 90.2
199 5.83 89.2 54.9 23.28 19 21.76 24.95 53.57
243 5.2 334.3 47.05 40.37 76 18.4 61.03 211.47
254 5.2 113.7 41.85 20.73 12 19 31.87 57.27
258 4.38 246.3 29.9 21.84 84 14.1 59 95.17
259 5.41 158 46.86 24.02 31 19.3 36.27 79.24
267 8.67 245.7 47.79 27.55 35 18.4 53.42 98.81
268 6.47 86.8 41.12 15.46 23 15.9 37.53 70.18
269 7.12 367.8 92.02 49.8 97 16.3 41.26 321.12
272 10.97 248.5 40.61 61.52 81 17.2 76.19 168.05
273 9.81 171.8 75.38 163.2 30 26.3 45.27 125.16
276 2.77 236.2 42.67 16.35 62 9.4 41.88 149.52
306 6.47 282.9 52.68 20.34 25 22.9 32.53 103.5

```

```

;
proc factor data= ch21 rotate=p n=3 score
reorder out= out21;
var X1-X8;
proc sort data=out914;
by factor1 factor2 factor3;
proc print;
var factor1 factor2 factor3;
id number;
run;

```

（其它四个功能区与其类似，只需修改数据行即可）

问题三程序

回归分析：

X=load('浓度.txt');

Y=load('坐标及功能区.txt');

maxs=zeros(1,3);

mins=zeros(1,3);

A=zeros(319,3);

for i=1:1:3

maxs(1,i)=max(Y(:,i));

mins(1,i)=min(Y(:,i));

end

for i=1:1:3

A(:,i)=[Y(:,i)-mins(1,i)]/(maxs(1,i)-mins(1,i));

end

```

A;
x1=A(:,1);
x2=A(:,2);
x3=A(:,3);
y=X(:,1);%其它七个重金属与其类似,修改相应矩阵的列标值
x=[ones(319,1) x1 x2 x3 (x1.^2) (x2.^2) (x3.^2)];
[b,bint,r,rint,stats]=regress(y,x);
b,stats;

```