

城市表层土壤重金属污染分析

摘 要

本文针对城市表层土壤重金属污染分析问题,建立了三次样条插值模型、潜在生态危害指数模型、地累积指数模型、因子分析评价模型、基于神经网络的污染源定位模型、基于傅里叶变换的重金属扩散模型和基于稀疏系统辨识的空气污染传播模型,分别解决了污染物空间分布及污染程度的分析、重金属污染主要原因的分析、污染源位置的确定以及城市地质环境演变模式的分析问题。

针对问题一,我们针对不同区域人类活动不同的特点,分别建立了三次样条插值模型、潜在生态危害指数模型和地累积指数模型,解决了污染物空间分布及污染程度的分析问题。首先我们对剔除了数据的特异值并对有效数据进行标准化处理以消除量纲影响;然后我们在 292 条有效数据的基础上采用三次样条插值模型做出污染物的空间分布图;接着我们按照人类活动范围的不同对生活区、公园绿化区、交通区建立了潜在生态危害指数模型,而对工业区和山区建立了地累积指数模型。最终得出生活区、公园绿化区、交通区、的污染程度分别为中等、中等、较高、中等、较低。

针对问题二,建立了因子分析评价模型,解决了重金属污染主要原因的分析问题。首先,我们对标准化后的数据进行 KMO 和 Bartlett 球形检验,判断其能够进行因子分析;其次我们选择分析的变量并提取公共因子;最后根据其成份得分系数矩阵推断主要因子的可能来源。最终我们得出生活区、工业区、山区、交通区、公园绿化区的主要污染源分别为生活垃圾和汽车尾气、工业污染、工业污染和汽车尾气、汽车尾气,工业污染和交通尾气。

针对问题三,我们建立了基于神经网络的污染源定位模型,解决了重金属污染源位置的确定问题。首先我们分析了重金属污染物的传播特性;其次,根据分析过后的复杂特性以及采样特点建立了基于神经网络的污染源定位模型;然后,我们将前 282 组数据作为训练集、后 10 组数据作为测试集、地理位置信息作为自变量、对应的浓度作为因变量创建并训练神经网络;最后随机生成地图范围内的地理位置信息数据对其进行浓度预测得出 As、Cr 的污染源的大致范围分别为 (3827, 184500)、(4116, 145400);而 Cd、Zn、Hg、Ni、Pb、Cu 的污染源较分散,主要有 2~3 个,其坐标分别为 (2095, 16400), (6714, 7454)、(1806, 175200), (4116, 169600), (3538, 16400), (2961, 108100)、(7580, 184500), (243200, 154700)、(1806, 109900), (74, 4659)和 (5848, 184500), (5848, 6150)。

针对问题四,在收集采样步长信息、污染源特征信息的条件下,针对固体污染和气体污染分别建立基于傅里叶变换的重金属扩散模型和基于稀疏系统辨识的空气污染传播模型,以研究城市地质环境演变模式的分析问题。首先我们提出需要收集采样步长信息、污染源特征信息的条件;其次基于收集的信息针对固体污染和气体污染分别建立基于傅里叶变换的重金属扩散模型和基于稀疏系统辨识的空气污染传播模型;最后将各信息代入求解,即可研究城市每个坐标点地质环境随时间的变化情况。

关键词: 潜在生态危害指数 地累积指数 神经网络 傅里叶变换 稀疏系统辨识

一、问题重述

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,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。

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。附件1列出了采样点的位置、海拔高度及其所属功能区等信息,附件2列出了8种主要重金属元素在采样点处的浓度,附件3列出了8种主要重金属元素的背景值。

现要求你们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)通过数据分析,说明重金属污染的主要原因。

(3)分析重金属污染物的传播特征,由此建立模型,确定污染源的位置。

(4)分析你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,还应收集什么信息?有了这些信息,如何建立模型解决问题?

二、问题分析

2.1 问题一的分析

利用所给数据画出各金属元素在该城区的空间浓度分布图,实际上就是要求我们把实测点拟合成曲线,对于此类问题通常选用三次样条插值法为宜。利用样条函数进行插值,画出浓度分布图。为了分析该城区内不同区域重金属的污染程度,我们需要根据其受人为因素影响程度的大小进行分类,分别建立潜在生态危害指数模型和地质累积指数模型。最后通过将其综合潜在生态危害指数(RI)和地质累积指数的平均值与评价标准相对比,得出各功能区污染程度的大小。

2.2 问题二的分析

想要找出重金属污染的主要原因,我们首先应该确定产生重金属污染的主要因素并查阅相关文献了解各重金属的主要来源。为了能够用少数的几个因子反映原始数据的大部分信息,便于进行分析,选用在多元统计分析的问题中常见的因子分析模型为宜。该模型需要检验所给数据是否符合因子分析的条件,再利用其方差主成分分解表中的特征值和累计贡献率确定因子,最后通过分析旋转成份矩阵的各项数据并结合相关资料,找出重金属污染产生的主要原因。

2.3 问题三的分析

想要确定污染源所处的地理位置，首先我们需要确定其传播特性；其次处理题目所给数据。处理数据通常采用数据规格化的方法；接着我们可以通过分析其浓度分布的特征和采样的特点，建立基于神经网络的污染源定位模型，并选取合适的值作为测试集和训练集；最后利用此模型的算法画出各重金属元素的浓度分布图，确定污染源的具体位置。

2.4 问题四的分析

想要更加细致的研究城市每个坐标点地质环境随时间的变化情况，现有的信息量是不够的，必须收集更多的数据已进行更细致的分析建模。因此首先我们要大量调研考虑应收集哪些信息，在此基础上根据收集的信息的特点以及不同污染物的传播特性分别针对气体、固体建模。最后利用信息对模型进行求解，即可研究城市每个坐标点地质环境随时间的变化情况。

三、基本假设

- 1、经过剔除后的有效数据是可信的；
- 2、远离人群及工业活动的自然区不受人为活动的影响；
- 3、土壤的 PH 值不对污染物的传播造成影响；
- 4、连续稳定的污染源释放的污染物的时间足够长，此时污染物的浓度视为稳定；
- 5、污染物沿各个方向的扩散速度在一定时间内是恒定的，不考虑随时间衰变的可能；
- 6、污染物只因为植物的富集作用、污水及大气环流作用自由运动而发生扩散；
- 7、各个区域的土壤种类及特性大致相同。

四、符号说明

$S(x)$	插值函数
E_r^i	潜在生态危害指数
CF	污染物指数
RI	综合潜在生态危害指数
C_i	元素的实际测量值
B_i	元素的背景值
T_i	元素的毒性系数
I_{geo}	地累积指数
k	修正指数
R	相关系数矩阵
r_{ij}	相关系数的值

五、模型的建立与求解

5.1 数据分析

5.1.1 剔除特异值

剔除特异值就是剔除所给数据中的无效数据,对于一些明显不符合逻辑的数据,如果保留会在一定程度上影响结果的准确性。因此我们首先删除所给数据中明显不符合逻辑的数据 24 个,最后得到有效数据 292 个,占总数据的 92.41%。

5.1.2 对有效数据的分析

对数据进行分类处理,得到生活区的各重金属含量结果如表 1 所示:(其他功能区的重金属含量见附录)

表 1 生活区各重金属含量

类型	As($\mu\text{g/g}$)	Cd(ng/g)	Cr ($\mu\text{g/g}$)	Cu ($\mu\text{g/g}$)	Hg (ng/g)	Ni ($\mu\text{g/g}$)	Pb ($\mu\text{g/g}$)	Zn ($\mu\text{g/g}$)
平均值	6.11	248.68	50.41	43.99	87.57	17.71	53.98	143.37
极大值	10.97	532.00	172.29	248.85	550.00	31.50	131.93	429.29
极小值	2.34	86.80	18.46	9.73	12.00	8.89	24.43	43.37
方差	3.86004475	11984.6	607.7292	1966.969	10786.61	27.57318	667.84	7915.893

其次我们统计了各重金属在各功能区的含量的平均值,如表 2 所示:

表 2 各重金属在各功能区的平均含量

	As ($\mu\text{g/g}$)	Cd (ng/g)	Cr ($\mu\text{g/g}$)	Cu ($\mu\text{g/g}$)	Hg (ng/g)	Ni ($\mu\text{g/g}$)	Pb ($\mu\text{g/g}$)	Zn ($\mu\text{g/g}$)
生活区	6.1055	248.6775	50.41175	43.99325	87.56975	17.7125	53.9765	143.3678
工业区	6.04	306.9964	43.92321	46.60607	121.8854	18.0375	69.09	158.9318
山区	4.043594	152.1281	37.01844	16.49938	38.35188	14.59719	36.39	71.03266
交通区	5.484574	336.476	51.63279	52.44535	100.6795	16.48341	61.55643	189.0246
公园绿地区	6.047742	231.1323	41.76323	24.81677	69.76935	14.7171	48.34129	99.16258

通过对数据的整理我们发现,各个重金属在各个功能区的含量大有不同,如生活区 Zn 的平均含量最高,约为 $143.37 \mu\text{g/g}$;而 Hg 的平均含量最低,约为 87.57ng/g 。生活区各重金属平均含量顺序为: $\text{Zn}>\text{Pb}>\text{Cr}>\text{Cu}>\text{Ni}>\text{As}>\text{Cd}>\text{Hg}$;工业区各重金属平均含量顺序为: $\text{Zn}>\text{Pb}>\text{Cu}>\text{Cr}>\text{Ni}>\text{As}>\text{Cd}>\text{Hg}$;山区各重金属平均含量顺序为: $\text{Zn}>\text{Cr}>\text{Pb}>\text{Cu}>\text{Ni}>\text{As}>\text{Cd}>\text{Hg}$;交通区各重金属平均含量顺序为: $\text{Zn}>\text{Pb}>\text{Cu}>\text{Cr}>\text{Ni}>\text{As}>\text{Cd}>\text{Hg}$;公园绿地区各重金属平均含量顺序为: $\text{Zn}>\text{Pb}>\text{Cr}>\text{Cu}>\text{Ni}>\text{As}>\text{Cd}>\text{Hg}$ 。

接着我们将处理后生活区各重金属含量与处理前的数据进行对比,结果如表 3 所示:

表 3 重金属在各功能区的对比

功能区	项目	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
生活区	范围	2.34~ 10.97	86.80~ 532.00	18.46~ 172.29	9.73~ 248.8 5	12~ 550.0 0	8.89~ 31.50	24.43~ 131.93	43.37~ 429.29
	平均值	6.11	248.68	50.41	43.99	87.57	17.71	53.98	143.37
工业区	范围	1.61~ 10.74	114.50~ 551.10	15.40~ 87.90	12.70 ~132. 05	11.79 ~561. 00	4.27~ 36.00	31.24~ 178.88	56.33~ 303.06
	平均值	4.04	152.13	37.02	16.5	38.35	14.6	36.39	71.03
山区	范围	1.80~ 10.99	40.00~ 407.60	16.20~ 92.76	2.29~ 69.06	9.64~ 85.71	5.51~ 45.45	19.68~ 113.84	32.86~ 172.53
	平均值	6.04	307	43.92	46.61	121.8 9	18.04	69.09	158.93
交通区	范围	1.61~ 30.13	50.10~ 886.60	15.32~ 306.02	12.34 ~227. 82	8.57~ 900.0 0	6.19~ 27.28	22.01~ 171.14	40.92~ 853.96
	平均值	5.48	336.48	51.63	52.45	100.6 8	16.48	61.56	189.02
公园 绿地 地区	范围	2.77~1 0.74	97.20~ 907.00	16.31~96. 28	9.04~ 38.62	10.00 ~259. 00	7.60~2 5.80	26.89~1 28.60	37.14~2 21.22
	平均值	6.05	231.13	41.76	24.82	69.77	14.72	48.34	99.16
背景 值	范围	1.8~ 5.4	70~190	13~49	6.0~ 20.4	19~51	4.7~ 19.9	19~43	41~97
	平均值	3.6	130	31	13.2	35	12.3	31	69

通过表 3 我们发现，经过剔除后的各功能区的重金属含量的平均值、范围与背景值之间有一定的差距，这说明各功能区确实存在重金属污染，并且各种重金属的污染程度不同。

5.1.3 数据标准化

鉴于原始数据多而复杂,所以在这里我们将数据采用统一的标准,即数据标准化。本文选取了对城市表层土壤影响较大的 8 个重金属元素,考虑实际应用,在计算前,首先要消除量纲的影响,解决各指标不具综合性的问题,原始数据标准化的结果见附录二。

5.2 问题一的模型建立与求解

要想得出 8 中重金属元素在该城区的空间分布,就要把实际测量的点拟合成曲线作图。因此我们选用三次样条插值法,画出各重金属在该地区的空间浓度分布图。根据人为因素对各功能区污染程度的影响大小的不同,受人为因素影响较大的生活区、交通区和公园绿地区,我们选用潜在生态危害指数模型;而受人为因素影响较小的工业区和山区,选用地质累积指数模型进行分析。通过分析我们得知,生活区和公园绿化区的污染程度较低,而交通区的污染程度较高;工业区为中度污染,而山区污染程度较低,为轻度污染。

5.2.1 模型的建立

因为生活区、交通区、公园绿化区处于人流量较多的地区,因此这三个区域受人为因素的影响较大。根据此特征,我们选用潜在生态危害指数法进行污染程度的测量。同理,受人为因素影响较小的工业区和山区,我们选用地质累积法进行分析。

(1) 三次样条插值

利用样条函数进行插值,即取插值函数为样条函数,称为样条插值。已知函数 $y = f(x)$ 在区间 $[a, b]$ 上的 $n+1$ 个节点

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$$

上的值 $y_i = f(x_i) (i = 0, 1, \dots, n)$, 求插值函数 $S(x)$, 使得

① $S(x_i) = y_i (i = 0, 1, \dots, n)$;

② 在每个小区间 $[x_i, x_{i+1}] (i = 0, 1, \dots, n-1)$ 上 $S(x)$ 是三次多项式, 记为 $S_i(x)$;

③ $S(x)$ 在 $[a, b]$ 上二阶连续可微。

函数 $S(x)$ 称为 $f(x)$ 的三次样条插值函数。

由三次样条插值法,我们画出各重金属在该地区的空间浓度分布。

(2) 潜在生态危害指数模型

由于生活区、交通区、公园绿地区的人流量相对较大,其污染程度受人为因素的影响也较大。因此,我们建立潜在生态危害指数模型。潜在生态危害指数结合重金属毒性系数,是一种可以全面反映重金属元素对生态镜的影响的指数。重金属元素的潜在生态危害指数 E_r^i 、污染物指数 CF 以及综合潜在生态危害指数 (RI) 可表示为:

$$E_r^i = T_i \times CF$$

$$CF = \frac{C_i}{B_i}$$

$$RI = \sum E_r^i$$

其中， C_i 是该元素的实际测量值， B_i 是该元素的背景值， T_i 是该元素的毒性系数。各重金属的毒性系数如表 4 所示：

表 4 毒性系数

元素	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
毒性系数	10	30	2	5	40	5	5	1

RI 的评价标准如表 5 所示：

表 5 RI 的评价标准

RI	<110	110~220	220~440	≥ 440
程度	低	中等	较高	很高

最后，通过计算的数据与评价标准进行比较，得出三个功能区的污染程度。

(3) 地累积指数模型

工业区和山区处于远离人群的偏远地区，其受重金属污染的程度较浅，因此我们选用地质累积指数进行分析。

地质累积指数期初被广泛应用于定量研究水环境沉积物中重金属污染程度，近几年被广泛应用于土壤重金属污染评价，其表达式为：

$$I_{geo} = \log_2[C_n/(k \times B_n)]$$

其中， I_{geo} 为地累积指数， C_n 为元素 n 的实际测量值， B_n 为元素 n 的背景值，k 为修正指数，一般取 1.5。

为了综合分析重金属元素对功能区总污染程度，我们决定采用 8 种元素地累积指数的均值作为最后评价的实际测量值。其公式为：

$$I'_{geo} = \sum_{i=1}^n I_{geo_i}$$

得到地累积指数的平均值。 I_{geo} 可分为 7 个等级，如表 6 所示：

表 6 地质累积指数评价标准

地质累积指数 I_{geo}	<0	0~1	1~2	2~3	3~4	4~5	>5
等级	无	较轻	轻度	中等	较重	重度	严重

最后我们通过对两个功能区地质累积指数的计算，判断其污染程度。

5.2.2 模型的求解

(1) 各元素空间分布图

首先，我们利用三次样条插值方法画出该地区的地形图和等高线图，如图 1

所示：

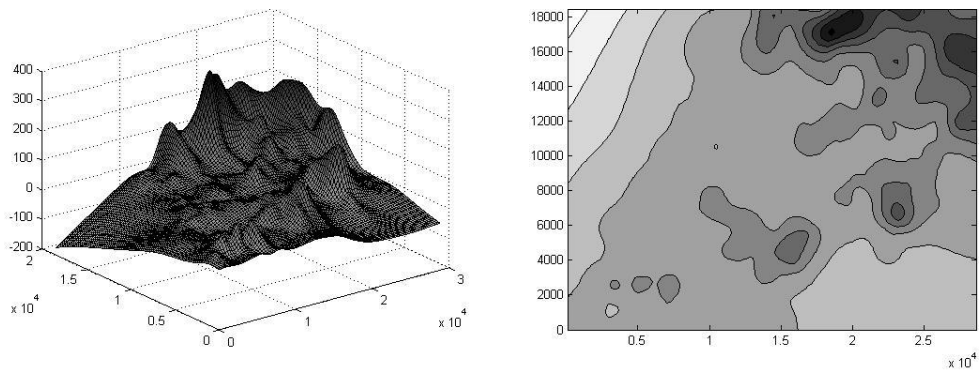


图 1 该地区三维地形图和等高线地形图

其次我们画出这 8 中主要重金属元素在该地区的空间分布，如图 2-图 9 所示：

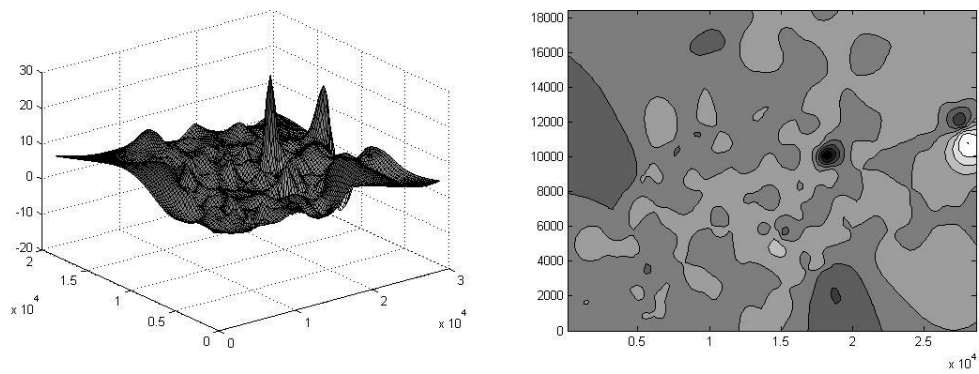


图 2 As 三维浓度分布图和平面浓度分布图

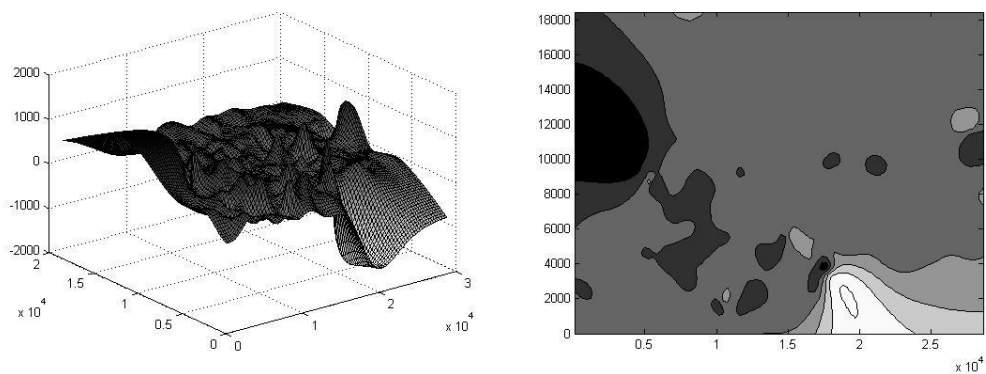


图 3 Cd 三维浓度分布图和平面浓度分布图

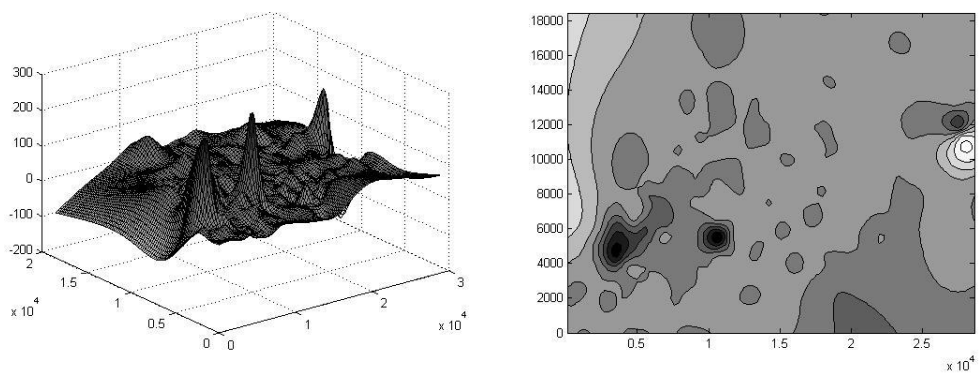


图 4 Cr 三维浓度分布图和平面浓度分布图

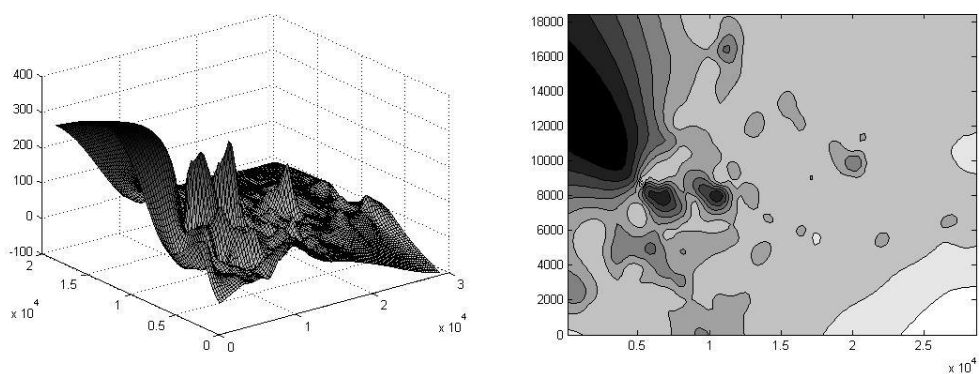


图 5 Cu 三维浓度分布图和平面浓度分布图

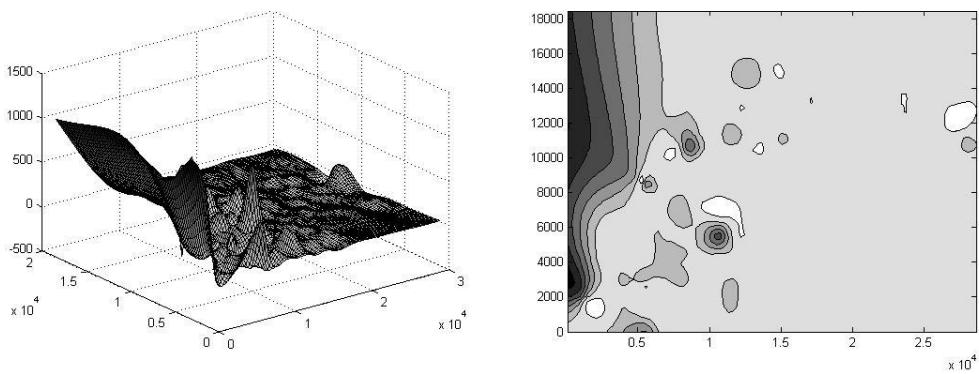


图 6 Hg 三维浓度分布图和平面浓度分布图

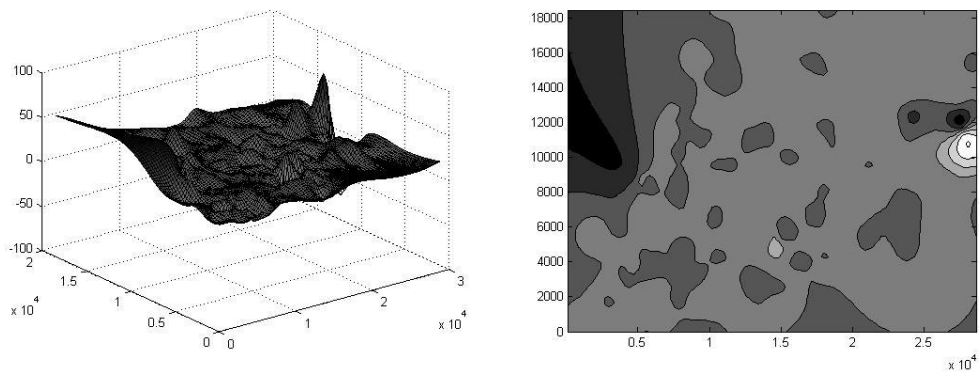


图 7 Ni 三维浓度分布图和平面浓度分布图

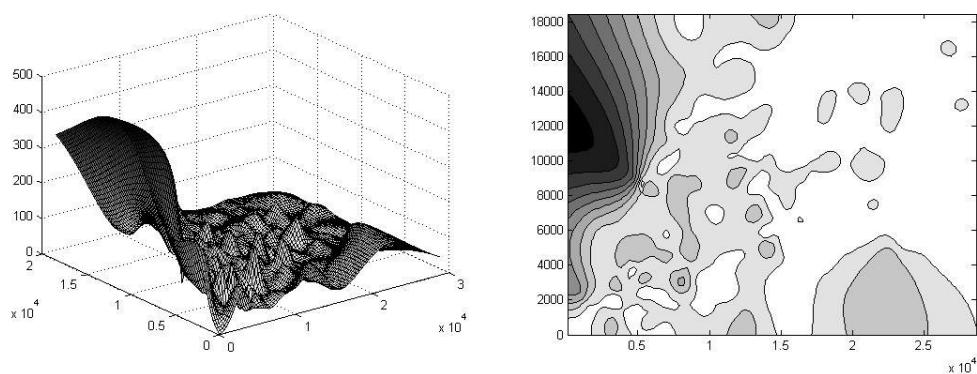


图 8 Pb 三维浓度分布图和平面浓度分布图

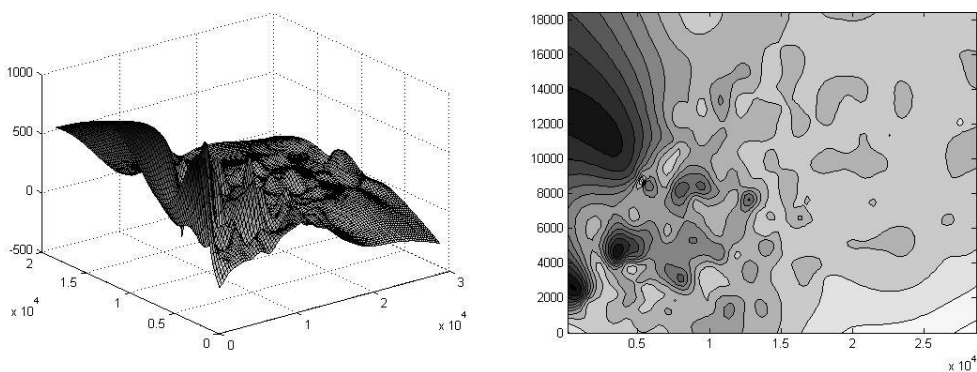


图 9 Zn 三维浓度分布图和平面浓度分布图

最后根据对该地区 8 种重金属元素浓度分布图的分析,我们发现该地区各重金属含量随地形的不同而不同。As 和 Ni 主要分布于海拔较低的地区,且 As 在不同位置浓度相差不大,我们可以近似的认为高海拔地区暂时未受 As、Ni 重金属的污染;Cd、Cr、Cu、Hg 和 Zn 在低海拔和高海拔地区都有分布,范围较广,但 Cd 浓度较小,受 Hg 轻微影响的地区面积较大,而 Cu 元素浓度含量较高的区域分布相对集中;Pb 在该地区内的不同位置含量相差较大。因此我们得到,同一城市的不同区域主金属污染存在差异,主要与该区域的功能有关。

(2) 潜在生态危害指数模型的求解

首先,根据计算三个功能区的重金属元素的潜在生态危害指数 E_r^i 和污染物指数 CF,得到潜在生态危害指数 RI,如表 7 所示:

表 7 潜在生态危害法结果

	生活区	交通区	公园绿地区
As ($\mu\text{g/g}$)	6.1055	5.484574	6.047742
CF	1.695972	1.523493	1.679928
Er	16.95972	15.23493	16.79928
Cd (ng/g)	248.6775	336.476	231.1323
CF	1.912904	2.588277	1.77794
Er	57.38712	77.6483	53.33821
Cr ($\mu\text{g/g}$)	50.41175	51.63279	41.76323

CF	1.626185	1.665574	1.347201
Er	3.252371	3.331148	2.694402
Cu (μg/g)	43.99325	52.44535	24.81677
CF	3.332822	3.973132	1.880059
Er	16.66411	19.86566	9.400293
Hg (ng/g)	87.56975	100.6795	69.76935
CF	2.501993	2.876556	1.99341
Er	100.0797	115.0622	79.73641
Ni (μg/g)	17.7125	16.48341	14.7171
CF	1.440041	1.340115	1.196512
Er	7.200203	6.700574	5.98256
Pb (μg/g)	53.9765	61.55643	48.34129
CF	1.741177	1.985691	1.559396
Er	8.705887	9.928457	7.796982
Zn (μg/g)	143.3678	189.0246	99.16258
CF	2.077793	2.739487	1.437139
Er	2.077793	2.739487	1.437139
RI	212.3269	250.5108	177.1853

接着，我们根据表 7 所得数据对三个功能区的污染程度进行分析，结果如表 8：

表 8 三个功能区的污染程度

功能区	RI	所属范围	污染程度
生活区	212.3269	110~220	中等
交通区	250.5108	220~440	较高
公园绿化区	177.1853	110~220	中等

通过表 8 我们发现，生活区和公园绿化区的污染程度较低，而交通区的污染程度较高。因为交通区处于人流量较多的繁华地带，其排放的汽车尾气较多，RI 值为 250.5108。而汽车尾气的排放是环境污染的主要因素之一，因此交通区的污染程度较为严重。相反，生活区和公园绿化区的绿化相对较好，其 RI 值较交通区较低，分别为 212.3269 和 177.1853，都属于中度污染。

(3) 地质累积模型的求解

首先，我们对工业区和山区的各重金属含量进行统计分析，得到其各元素地累积指数 I，结果如表 9 所示：

表 9 地质累积法的结果

	工业区	山区
As (μg/g)	6.04	4.043594
I	0.746552	0.167641
Cd (ng/g)	306.9964	152.1281
I	1.23971	0.226775
Cr (μg/g)	43.92321	37.01844
I	0.502715	0.255976

Cu ($\mu\text{g/g}$)	46.60607	16.49938
I	1.81998	0.321873
Hg (ng/g)	121.8854	38.35188
I	1.800098	0.131942
Ni ($\mu\text{g/g}$)	18.0375	14.59719
I	0.552341	0.247032
Pb ($\mu\text{g/g}$)	69.09	36.39
I	1.156209	0.231274
Zn ($\mu\text{g/g}$)	158.9318	71.03266
I	1.203739	0.041886
均值	1.127668	0.20305

其次，通过公式？我们算出功能区 8 个元素地累积指数的均值，与其评价标准进行对比分析：

表 10 两个功能区的污染程度

功能区	地累积指数	所属范围	污染程度
工业区	1.127668	1~2	中等
山区	0.20305	0~1	较低

我们得知，工业生产过程中的废水、废气和固体排放物等都能对环境质量造成很大程度的影响，大量工业污染物的排放使得工业区污染程度逐渐加深。而山区基于其独特的地理位置和海拔，空气质量较好且绿色植被较多，其污染程度较低。最后我们得出结论：工业区为中度污染，而山区污染程度较低，为轻度污染。

5.3 问题二的模型建立与求解

分析重金属污染的主要原因，需要我们先确定产生污染的主要因素，再找出这些因素产生污染的主要原因。根据多元统计的原理我们选用因子分析模型。首先通过对数据的检验，我们判断其是否能够进行因子分析。其次通过方差主成分分析找出产生污染的主要因子。最后对这几个主要因子的旋转成份矩阵进行分析，找出重金属污染产生的主要原因。

5.3.1 模型的建立

因子分析法是在许多变量中找出隐藏的具有代表性的因子，用找出的因子去描述许多指标或因素之间的联系，即将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，以较少的几个因子反映原始资料的大部分信息。利用这种模型，我们可以方便地找出影响各功能区污染程度的主要因素，以及这些因素的影响力。因此我们选用因子分析找出重金属污染的主要原因。

(1) 首先，我们对之前准备的标准化后的数据进行 KMO 和 Bartlett 球形检验，判断标准化后的数据是否能够进行因子分析。

当 KMO 的检验系数 >0.5 并且 Bartlett 球形检验的 P 值 <0.05 时，该数据才能够进行因子分析。

(2) 第二步，选择分析的变量。选择变量时需要用定性分析和定量分析的方法，

因子分析的前提条件是观测变量之间有较强的相关性，因为如果变量之间无相关性或相关性较小，则它们之间就不会存在共享因子。因此我们计算所选变量的相关系数矩阵 R 来判断各元素之间的相关系数，其计算公式如下：

$$R = (r_{ij})_{m \times m}$$

$$r_{ij} = \frac{\sum_{k=1}^n \widetilde{a_{ki}} \times \widetilde{a_{kj}}}{n-1}, i, j = 1, 2, \dots, m$$

其中， $r_{ii} = 1, r_{ij} = r_{ji}, r_{ij}$ 是第 i 个指标与第 j 个指标的相关系数。

- (3) 第三步，提取公共因子。在方差主成分分解表中取其特征值大于 1 的因子，再按照累计方差贡献率来确定因子，一般认为累计贡献率达到 60% 才能符合要求。
- (4) 第四步，根据其成份得分系数矩阵推断主要因子的可能来源。
- (5) 最后，由于我们的最终目的是知道每个因子的含义以便进行下一步的分析，所以我们需要知道其旋转成份矩阵。这样因子的实际含义更容易解释，并根据旋转成份矩阵找出影响该功能区的主要重金属元素，分析其可能来源。

5.3.2 模型的求解

下面我们以生活区为例，进行因子分析（其他功能区分析见附录）。

(1) KMO 和 Bartlett 球形检验

对生活区的各数据进行 KMO 和 Bartlett 球形检验的结果如表 11 所示：

表 11 KMO 和 Bartlett 球形检验

取样足够度的 Kaiser-Meyer-Olkin 度量。		.656
近似卡方		154.715
Bartlett 的球形度检验	df	28
	Sig.	.000

通过分析表 11 的数据，KMO 检验系数为 0.656>0.5；且 Bartlett 球形检验的 P 值为 0.000<0.05。说明生活区标准化后的数据可以进行因子分析。

(2) 相关系数矩阵的计算

为了判断各元素之间是否存在相关关系，我们计算其相关系数矩阵如表 12：

表 12 相关矩阵

	As	Cd	Cr (μg/g)	Cu	Hg	Ni (μg/g)	Pb	Zn
	(μg/g)	(ng/g)		(μg/g)	(ng/g)		(μg/g)	(μg/g)
As (μg/g)	1.000	.109	.444	.458	.285	.665	.241	.360
Cd (ng/g)	.109	1.000	.310	.338	.477	.125	.578	.703

Cr (μg/g)	.444	.310	1.000	.492	.159	.604	.353	.679
Cu (μg/g)	.458	.338	.492	1.000	.140	.348	.461	.478
Hg (ng/g)	.285	.477	.159	.140	1.000	.156	.470	.435
Ni (μg/g)	.665	.125	.604	.348	.156	1.000	.158	.298
Pb (μg/g)	.241	.578	.353	.461	.470	.158	1.000	.707
Zn (μg/g)	.360	.703	.679	.478	.435	.298	.707	1.000

通过表 12 我们发现其相关系数的值 $r_{ij} \in [0,1]$ ，其值越大，相关程度越高。

As 与 Ni、Cd 与 Zn、Cr 与 Zn、Pb 与 Zn 之间存在显著的正相关，相关系数分别为 0.665、0.703、0.679、0.707，则我们可以判定这六种重金属(As、Cd、Cr、Ni、Pb、Zn)的来源可能相同；而 Cd 与 Ni、Cr 与 Hg、Cu 与 Hg、Hg 与 Ni、Ni 与 Pb 之间的相关系数相对较低，分别为 0.125、0.159、0.140、0.156、0.158，说明生活区中这六种元素(Cd、Cr、Cu、Hg、Ni、Pb)的污染源可能并不唯一。

(3) 提取因子

为了提取具有代表意义的因子，我们进行方差主成分分解，结果如表 13 所示：

表 13 方差主成分分解

成份	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差 的 %	累积 %	合计	方差 的 %	累积 %	合计	方差 的 %	累积 %
1	3.821	47.760	47.760	3.821	47.760	47.760	2.790	34.881	34.881
2	1.555	19.439	67.199	1.555	19.439	67.199	2.585	32.318	67.199
3	.851	10.640	77.839						
4	.631	7.887	85.726						
5	.400	5.006	90.732						
6	.347	4.342	95.074						
7	.297	3.712	98.786						

8	.097	1.214	100.000						
---	------	-------	---------	--	--	--	--	--	--

提取方法：主成份分析。

由方差分解主成分提取表可以看出,特征值在某些程度上可以反映指标的影响力,大的特征值也表示相关矩阵中某一成分的方差影响大,因此适用因子分析法。而表中前 2 个指标的特征值都大于 1,分别为 3.821 和 1.555,且累计贡献率约为 67.199%,即可以解释全部数据的 67.199%。因此对前 2 个成份进行分析已经能够反映全部数据的大部分信息,所以我们可以提取 2 个指标作为确定污染源的主要因子。

(4) 计算成份得分系数矩阵

表 14 成份得分系数矩阵

	成份	
	1	2
As (μg/g)	.623	.552
Cd (ng/g)	.676	-.534
Cr (μg/g)	.750	.335
Cu (μg/g)	.681	.179
Hg (ng/g)	.544	-.397
Ni (μg/g)	.585	.655
Pb (μg/g)	.740	-.417
Zn (μg/g)	.875	-.246

提取方法：主成份。

a. 已提取了 2 个成份。

接着我们对 2 个因子进行分析,由表 14 可以得知, Cr、Pb、Zn 在成份 1 上系数较高,分别为 0.750、0.740、0.875,与其他成分相比载荷较高,是该因子产生的主要污染物。而 Cr、Pb、Zn 这三种元素主要都来自于工业生产活动中的废水废气排放,其中 Pb 也来自于尾气的排放,Zn 也来自于化学物质的燃烧产生的废气。因此因子 1 极有可能是工业生产中排放的废水废气。而 As、Cd、Ni 在成份 2 上系数较高,分别为 0.552、-0.534、0.655。查阅相关资料我们发现,As 的主要来源是矿业开采和农药使用产生的废气

烟尘，化学工业中排放的废气废水以及汽车排放的尾气中含有较多的 Cd，而 Ni 主要来源于化石燃料的燃烧所产生的尾气。因此成份 2 极有可能是各种化学燃料燃烧产生的废气和汽车排放的尾气。

为了进行更加深入的分析，我们计算接着计算元素的旋转成份矩阵。

(5) 旋转成份矩阵

表 15 旋转成份矩阵

	成份	
	1	2
As (µg/g)	.088	.827
Cd (ng/g)	.859	.061
Cr (µg/g)	.328	.753
Cu (µg/g)	.382	.591
Hg (ng/g)	.669	.074
Ni (µg/g)	-.010	.878
Pb (µg/g)	.827	.191
Zn (µg/g)	.812	.409

提取方法 :主成份。

旋转法 :具有 Kaiser 标准化的正交旋转法。

a. 旋转在 3 次迭代后收敛。

由表 13 和表 15 可以看出，两个主成分中第一主成分的贡献率约为 47.760%，该主成分在 Cd、Hg、Pb、Zn 上有较大的载荷。由于 Cd 主要来源于工业源及化肥的使用；Pb 主要来源于工业和尾气；而交通污染是 Zn 含量较高的主要原因。由此不难推断出，该地区可能处于交通要道，并且附近建有工厂，因此汽车排放的尾气和工厂排放的废气废水等，都可能是生活区的主要污染源。

同理，我们通过分析发现，工业区的重金属污染源可以由一个主成分反映得到。这一主成分的贡献率较高，约为 80.057%。它在 Cd、Cr、Cu、Hg、Pb、Zn 上的载荷较高，几乎接近于 1。其中 Cd、Cr 和 Zn 主要来自于工业生产的废水、废气、废渣；Cu 的主要来源是工业和农药；而 Pb 来源于工业和尾气，特别是冶炼厂排除的废气中 Pb 的含量较高；Hg 来源工业废水和生活垃圾。由此我们可以近似的认为，该工业区的主要污染源就是工厂排放的污染物，即废气、废水等，工厂的生产化活动、化学物质的燃烧对工业区的污染，就是该地区的主要污染源。

其次，我们对山区的污染源进行研究。采用 1 个主成分来反映山区的主要污染源，其累计贡献率为 70.856%。该主成分在 Cd、Cr、Hg、Pb、Zn 载荷较高。根据之前的分析结果，我们得出其污染源主要来自于工业污染，但由于山区所处的地理位置较偏远，我们可以推断出，该地区山区的主要污染源可能来自于附近工厂污染物的传播，并且此山区可能临近交通繁忙的路段，由于汽车尾气的传播而在成污染。

接着，交通区也只需要 1 个主成分就可以反映多数信息，其贡献率达到了 72.774%。Cd、Cr、Cu、Hg、Pb、Zn 是影响该功能区受污染的主要元素，除了交通区的汽车排放的尾气一定是交通区的主要污染源之外，我们可以推断出，附近的工业污染也是交通区的主要污染源。

最后，选取 2 个主成分对公园绿化区分析污染的主要来源。其累计贡献率为 90.786%，主要污染来源为 As、Cd、Cr、Hg、Ni、Pb、Zn 元素。除了之前的分析说明，As 主要来自于矿业开采、工业和农药；Ni 和 Cd、Cr 同样来源于工业生产活动。因此该公园绿化区可能临近于交通繁忙的路段，有助于改善汽车尾气排放造成的污染，即该地区主要受工业污染和交通污染。

综合上述对 5 个功能区的污染源的分析，我们发现 As、Ni、Cu 在这些地区中对其污染的贡献较小，而 5 个功能区的污染源几乎都来源于 Cd、Cr、Hg、Pb、Zn 这 5 类重金属，即该地区受工业污染和交通污染较为严重。其污染来源极有可能是工业生产活动排放的废水废气、化学燃料燃烧放出的废气烟尘和汽车排放的尾气。

5.4 问题三的模型建立与求解

为了确定污染源的位置，我们首先要确定各重金属元素在该地区的浓度分布图，结合其传播特性和取样特征，我们选用基于神经网络的污染源定位模型来确定污染源的位置。首先我们将数据规格化，使得其便于分析；其次选取合适的数

5.4.1 模型的建立

根据题目所给出的信息我们发现，该地区各重金属元素的性质随着空间位置的不同而发生着变化，也正是因为这种高度复杂的空间异质性，使得我们对该地区污染源位置的确定变得十分困难。除了上述自然状态下的空间变化之外，人类活动也能导致空间变化，结合重金属的浓度分布图，我们发现其传播特性：即重金属在空间分布上呈现出明显的集中，这也就说明了人类活动对重金属分布有着直接的作用。而这些因素往往使空间动态变得极为复杂，这时对重金属元素的空间分布做出定量的分析是较为困难的。结合该地区的地形图以及各重金属的浓度分布图，我们总结其传播特征如下：

- ①大部分重金属浓度异常区域呈现焦点式分布；
- ② 各重金属元素的分布区域并不完全相同；
- ③某些元素的分布与其地形有很大的关联。

根据上述重金属污染物的传播规律，可知求取污染源的位置即为求取金属浓度最大值得点对应的坐标值，重金属污染浓度高的地方一般为污染源。但是因为

我们按照每平方公里 1 个采样点对表层土进行取样，采样点较稀疏，数量比较少，利用传统的基于拟合、插值的思路往往会出现因点少而误差较大的问题，即重金属元素浓度与坐标(x, y)之间的函数关系很难用解析式来表达，利用传统优化算法的思路，往往会出现局部最优而不能达到全局最优。因此，我们选用基于神经网络的污染源定位模型，这样既考虑了污染物的传播特征，又避免了因采样点过少而误差较大的问题，且相对简单，便于实际操作，具有容错容差能力和一定的推广能力。

(1) 神经网络概述

①BP 神经网络是模拟人类大脑处理和分析问题的方式方法来研究实际问题，从本质上说，它是一种黑箱建模工具，它能够通过“学习”来仿真真实系统里的输入和输出之间的定量关系。它是一种具有三层或三层以上的多层神经网络，每一层都由若干个神经元组成，如图 10 所示，左层的每一个神经元与右层的每一个神经元都有连接，而上下神经元之间无连接。BP 神经网络按有导师学习方式训练，当一对学习模式提供给网络后，其神经元的激活值将从输入层经各隐含层向输出层传播，在输出层的各神经元输出对应于输入模式的网络相应。然后，按减少希望输出与实际输出误差的原则，从输出层经各隐含层，最后回到输入层逐层修正各连接权。由于这种修正过程是从输出到输入层逐层进行的，所以称它为“误差逆传播算法”。随着这种误差逆传播训练的不断修正，网络对输入模式响应的正确率也将不断提高。

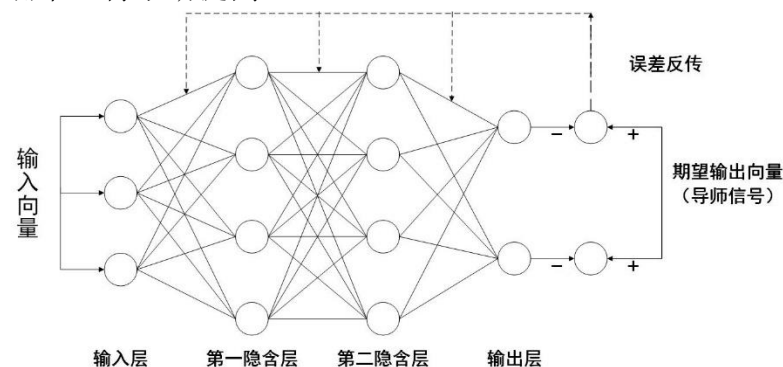


图 10 BP 网络模型结构

为了使神经网络具有某种功能，完成某项任务，必须调整层间连接权值和节点阈值，使所有样品的实际输出和期望输出之间的误差稳定在一个较小的值以内。一般地，可将 BP 网络的学习算法描述为如下步骤：

步骤一：初始化网络及学习参数；

步骤二：提供训练模式、训练网络，直到满足学习要求；

步骤三：前向传播过程：对给定训练模式输入，计算网络的输出模式，并与期望模式比较，若有误差，则执行步骤④，否则返回步骤②；

步骤四：反向传播过程：计算同一层单元的误差，修正权值和阈值，返回步骤二。

网络的学习是通过用给定的训练集训练而实现的。通常用网络的均方差误差来定量地反映学习的性能。一般地，当网络的均方差误差低于给定值时，则表明对给定训练集学习已经满足要求了。

②RBF 神经网络模型

RBF 神经网络具有很强的逼近能力、分类能力和学习速度。其工作原理是把网络看成对未知函数的逼近，任何函数都可以表示成一组基函数的加权和，也即

选择各隐层神经元的传输函数，使之构成一组基函数来逼近未知函数。RBF 神经网络由一个输入层、一个隐含层和一个输出层构成。设输入层的输入为 $X = [x_1, x_2, \dots, x_n]$ ，实际输出为 $Y = [y_1, y_2, \dots, y_p]$ 。输入层实现从 $X \rightarrow R_i(X)$ 的非线性映射，输出层实现从 $R_i(X) \rightarrow y_k$ 的线性映射，输出层第 k 个神经网络输出为

$$\hat{y}_k = \sum_{i=1}^m w_{ik} R_i(X), k = 1, \dots, p$$

其中： n 为输入层节点数； m 为隐含层节点数； p 为输出层节点数； w_{ik} 为隐含层第 i 个神经元与输出层第 k 个神经元的连接权值； $R_i(X)$ 为隐含层第 i 个神经元的作用函数，即：

$$R_i(X) = \exp\left(-\|X - C_i\|^2 / 2\sigma_i^2\right), i = 1, \dots, m$$

其中： X 为 n 维输入向量； C_i 为第 i 个基函数的中心，与 X 具有相同维数的向量； σ_i 为第 i 个基函数的宽度； m 为感知单元的个数（隐含层节点数）； $\|X - C_i\|$ 为向量 $X - C_i$ 的范数，它通常表示 X 与 C_i 之间的距离。对于给定的输入，只有一小部分靠近 X 的中心被激活。当确定了 RBF 网络的聚类中心 C_i 、权值 w_{ik} 以后，就可以求出给定某一输入时，网络对应的输出值。

在 RBF 网络中，隐层执行的是一种固定不变的非线性变换， C_i 、 σ_i 、 w_{ik} 需通过学习和训练确定，一般分三步进行：

步骤一：确定基函数的中心 C_i ；

步骤二：确定基函数的宽度 σ_i ；

步骤三：确定从隐含层到输出层的连接权值 w_{ik} ，RBF 连接权的修正可以采用最小均方误差测度准则进行。

(2) 原始数据预处理

对样本的输入、输出进行规格化处理，采用公式如下：

$$\tilde{t} = \frac{2(t - t_{min})}{t_{max} - t_{min}} - 1$$

其中， t 为规格化前的变量； t_{max} 、 t_{min} 分别为 t 的最大值和最小值， \tilde{t} 为规格化后的变量。

采用此方法，我们将数据化为 $[-1, 1]$ 之间的任意数值。

(3) 基于神经网络的污染源定位模型

根据(1)中对神经网络的描述，首先我们把该地区全部地域范围内的元素浓度当作训练集，其次选出测试集和训练集，对神经网络进行创建、训练、仿真测试，随机自动生成地图上的坐标，最后再预测此坐标的值作图。

5.4.2 模型的求解

结合上述污染物的传播特征和基于神经网络的污染源定位模型，我们将前 282 组数据作为训练集，后 10 组数据作为测试集，生成各元素在该地区浓度分布地形图，如图 11-18 所示：

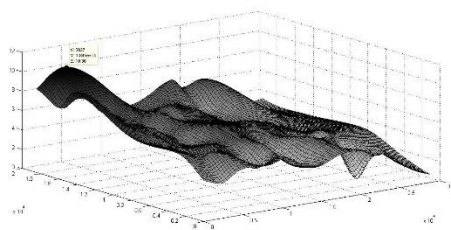


图 11 As 浓度分布地形图

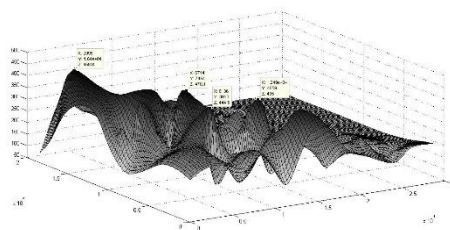


图 12 Cd 浓度分布地形图

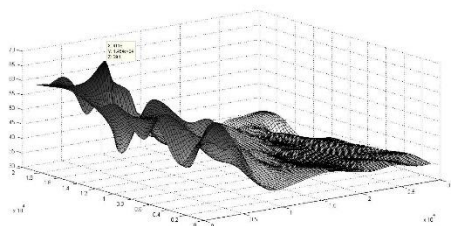


图 13 Cr 浓度分布地形图

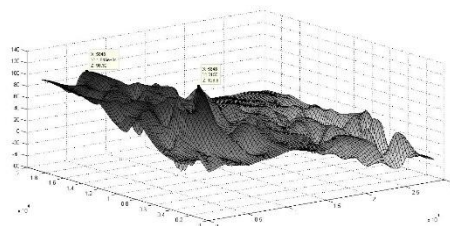


图 14 Cu 浓度分布地形图

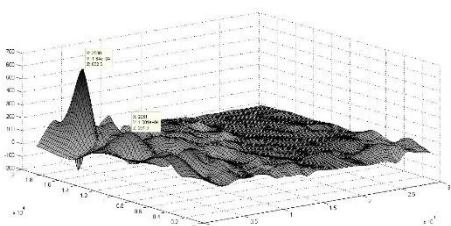


图 15 Hg 浓度分布地形图

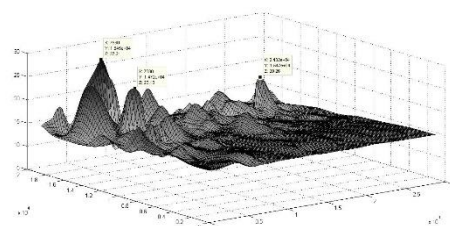


图 16 Ni 浓度分布地形图

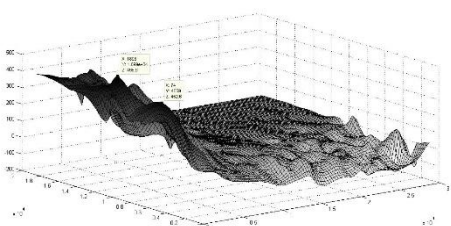


图 17 Pb 浓度分布地形图

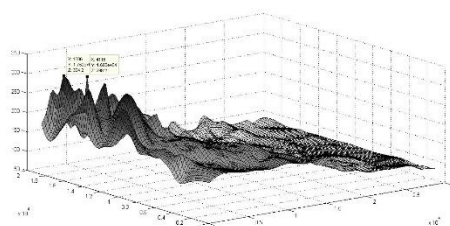


图 18 Zn 浓度分布地形图

由此，我们得出该地区各元素污染源的位置如表 16 所示：

表 16 各元素污染源位置

元素	污染源位置(x, y)
As	(3827, 184500)
Cd	(2095, 16400), (6714, 7454)
Cr	(4116, 145400)
Cu	(5848, 184500), (5848, 6150)
Hg	(3538, 16400), (2961, 108100)
Ni	(7580, 184500), (243200, 154700)

Pb	(1806, 109900), (74, 4659)
Zn	(1806, 175200), (4116, 169600)

由表 16 分析可知, As 和 Cr 污染源很集中, 坐标分别为(3827, 184500)和(4116, 145400); Cd 和 Zn 污染源主要分布在一片区域, 主要污染源坐标分别为(2095, 16400), (6714, 7454)和(1806, 175200), (4116, 169600); Cu 污染源分布呈直线, 主要污染源坐标为(5848, 184500), (5848, 6150); 而 Hg、Ni、Pb 的污染源较分散, 主要有 2~3 个, 其坐标分别为(3538, 16400), (2961, 108100)、(7580, 184500), (243200, 154700)和(1806, 109900), (74, 4659)。

5.5 问题四的模型建立与求解

为了更好的研究城市地质环境的演变模式, 我们需要搜集更多的信息进行更加准确的研究。因为空间的复杂多样性, 使得该城区的地质环境变得极为复杂, 因此, 我们选择根据该城区污染源的传播特性对其进行分类研究, 即分为在大气中传播的气体污染物和在土壤中传播的污染物。再搜集我们缺少的信息进行分析。

5.5.1 模型的评价

要想得出 8 种重金属元素在该城区的空间分布, 就要把实际测量的点拟合成曲线作图。因此我们选用三次样条插值法, 画出各重金属在该地区的空间浓度分布图。根据人为因素对各功能区污染程度影响大小的不同, 受人为因素影响较大的生活区、交通区和公园绿地区, 我们选用潜在生态危害指数模型; 而受人为因素影响较小的工业区和山区, 选用地质累积指数模型进行分析。通过分析我们得知, 生活区和公园绿化区的污染程度较低, 而交通区的污染程度较高; 工业区为中度污染, 而山区污染程度较低, 为轻度污染。

(1) 模型的优点

①问题一中的三次样条插值既保留了分段低次插值多项式准确拟合曲线的优点, 又提高了插值函数的光滑性, 能够较好的逼近测量数值, 使得拟合出的浓度图有较高的可信度;

②问题一中的潜在生态危害指数模型和地质累积指数模型通过区分人为因素对环境污染程度的影响分别计算各功能区的污染程度。潜在生态危害指数模型从不同重金属具有不同毒性的角度出发, 利用其毒性系数较为全面的评价了功能区的受污染程度; 而地累积指数不仅考虑了自然地质过程造成的背景值的影响, 还充分注意到了人为因素对重金属污染的影响, 两个方面相结合较为全面客观的评价了功能区的污染程度;

③问题二中的因子分析模型从较多的指标中找出产生污染的主要因子, 它不是对原始变量的取舍, 而是根据原始变量的信息进行重新组合, 简化了原始变量, 实际上是起降维的作用, 为我们处理数据降低了难度; 通过旋转使因子变量更具有可解释性, 使得因子分析模型的可信度更高。

④问题三中的基于神经网络的污染源定位模型, 有很强的非线性拟合能力, 可以映射非常复杂的线性关系, 而且学习规则简单, 避免了因采样点过少而误差较大的问题。

(2) 模型的缺点

问题三中的基于神经网络的污染源定位模型, 由于训练样本不够多, 因此所

得的预测结果与实际值有一定偏差，所以其学习过程还有待进一步的提高。

5.5.3 模型的建立和求解

为了更好的研究城市地质环境的演变模式，我们还需要搜集更加全面的信息，对城区地质环境的状况及演变模式做进一步的分析。还需要搜集的信息如下：

(1) 采样步长信息；

(2) 污染源特征信息，包括污染源传播方式以及污染扩散系数。

根据题目和我们对重金属污染源传播特性的分析，我们得知其污染物大致可以分为在大气中传播的气体污染物、流入土壤的废水和本身在土壤中传播的污染物。因此，我们建立基于稀疏系统辨识的气体扩散模型；而在土壤中传播的污染物，我们选择建立基于傅里叶变换的重金属扩散模型。

(1) 基于稀疏系统辨识的气体扩散模型

空气污染扩散模型可以抽象为连续时间随机动力系统，接下来我们描述气体污染物扩散的过程：

$$dx(t) = Ax(t)dt + db(t), t \in [0, T]$$

其中： $x(t) \in N$ 代表待观测空间中的 N 维污染物的浓度分布； $b(t)$ 为 N 维布朗运动，表示污染物浓度的随机波动；矩阵 $A \in N \times N$ 为扩散系数矩阵，代表系统的扩散特征。由扩散过程的空间性质可知，扩散系数矩阵 A 应为稀疏矩阵，其所有特征值满足：

$$z \in C: Re(z) \leq 0$$

因为观测过程是对该扩散过程的离散采样，为方便进一步分析，引入时间步长参数 $\eta > 0$ ，将扩散过程方程改写成离散时间的随机动力系统：

$$x(t) = x(t-1) + \eta Ax(t-1) + u(t) + w(t), t \in [1, n]$$

其中， $u(t)$ 是一个 N 维非负稀疏矢量，表示在空间上稀疏分布的污染源。 $u(t)$ 中非零元素的位置表示污染源的空间分布，非零元素的取值代表污染源的强度。

(2) 基于傅里叶变换的重金属模型

取采样区域闭合曲面 S 所围成地域为 Ω ， $C(x, y, z, t)$ 表 t 时刻 (x, y, z) 位置点的污染物浓度。

由于污染物因为植物的富集作用、污水以及大气环流作用自由运动而发生扩散，所以我们用 D_x ， D_y ， D_z 分别表示 x, y, z 方向的扩散系数。

$\theta(x, y, z, t)$ 是 x, y, z 位置点单位体积单位时间污染物的排放量，是污染源特征信息的一个参数。

由此：

①三维浓度函数 C 进行傅里叶变换可得：

$$C(w_x, w_y, w_z, t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} C(x, y, z, t) e^{-i(xw_x + yw_y + zw_z)} dx dy dz = F(C)$$

②有限差分。在 i 点附近把函数 $C = f(t)$ 展开为泰勒级数：

$$C = C_i + \left[\frac{dC}{dt} \right] (t - t_i) + \frac{1}{2!} \left[\frac{d^2C}{dt^2} \right]_i (t - t_i)^2 + \frac{1}{3!} \left[\frac{d^3C}{dt^3} \right]_i (t - t_i)^3 + \frac{1}{4!} \left[\frac{d^4C}{dt^4} \right]_i (t - t_i)^4 + \dots$$

其中， $\left[\frac{dC}{dt} \right]$ ， $\left[\frac{d^2C}{dt^2} \right]_i$ ， \dots 分别表示各阶导数分别表示各阶导数在 $t = t_i$ 处的数值。

③通过闭曲面 S 从时刻 t 到时刻 $t + \Delta t$ 流入 Ω 的污染物质量为:

$$M_1 = \int_t^{t+\Delta t} \oint_S \left[D_x \frac{\partial C}{\partial x} \cos \alpha + D_y \frac{\partial C}{\partial y} \cos \beta + D_z \frac{\partial C}{\partial z} \cos \gamma \right] dS dt$$

其中, $\cos \alpha$ 、 $\cos \beta$ 、 $\cos \gamma$ 为 S 的外法向余弦。由于重金属离子可以通过大气环流进行扩散, 污染物从 t 时刻到 $t + \Delta t$ 时刻流出 S 的质量为:

$$M_2 = \int_t^{t+\Delta t} \oint_S \left[u_x C(x, y, z, t) \cos \alpha + u_y C(x, y, z, t) \cos \beta + u_z C(x, y, z, t) \cos \gamma \right] dS dt$$

所以空间内重金属污染的 4D 数学模型即四维重金属污染物扩散模型为:

$$\frac{\partial C}{\partial t} = D_x \frac{\partial^2 C}{\partial x^2} + D_y \frac{\partial^2 C}{\partial y^2} + D_z \frac{\partial^2 C}{\partial z^2} - u_x \frac{\partial C}{\partial x} - u_y \frac{\partial C}{\partial y} - u_z \frac{\partial C}{\partial z} - KC + \theta(x, y, z, t)$$

初始条件: $C(x, y, z, 0) = \lambda(x, y, z)$

对扩散模型左右两边同时进行傅里叶变换得:

$$\begin{aligned} C(x, y, z, t) = & \frac{1}{(2\pi)^3} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(D_x w_x^2 + D_y w_y^2 + D_z w_z^2 + i w_x u_x + i w_y u_y + i w_z u_z + K)t} \\ & \times \left[\int_0^t \theta(w_x, w_y, w_z, t) \times e^{(D_x w_x^2 + D_y w_y^2 + D_z w_z^2 + i w_x u_x + i w_y u_y + i w_z u_z + K)t} dt \right. \\ & \left. + const \right] dw_x dw_y dw_z \end{aligned}$$

即可对模型进行求解。

由于山区的污染程度最低, 所以对于一些重金属污染物来说, 海拔对于污染物的扩散影响几乎可以忽略。为了简化处理, 我们对此模型进行简化并分成 3 种情况讨论:

①瞬时污染点源四维扩散模型。

考虑到重金属污染物扩散速度问题, 此处的“瞬时”指的是时间频率为 1d:

$$\frac{\partial C}{\partial t} = D_x \frac{\partial^2 C}{\partial x^2} + D_y \frac{\partial^2 C}{\partial y^2} + D_z \frac{\partial^2 C}{\partial z^2} - u_x \frac{\partial C}{\partial x} - u_y \frac{\partial C}{\partial y} - u_z \frac{\partial C}{\partial z}$$

当污染源排放量为 M , 该数学模型的解析解为:

$$C(x, y, z, t) = \frac{M}{8(\pi t)^{\frac{1}{2}} \sqrt{D_x D_y D_z}} \exp \left[-\frac{(x - u_x t)^2}{4D_x t} - \frac{(y - u_y t)^2}{4D_y t} - \frac{(z - u_z t)^2}{4D_z t} \right] \exp(-Kt)$$

②连续污染点源四维扩散模型

如果污染物释放是以连续污染点源的方式进行, 则形成的浓度相当于单位时间内连续释放瞬时点源的积分, 实际上相当于对上式进行时间区间上的积分:

$$C(x, y, z, t) = \int_0^t \frac{M}{8(\pi t)^{\frac{1}{2}} \sqrt{D_x D_y D_z}} \exp \left[-\frac{(x - u_x t)^2}{4D_x t} - \frac{(y - u_y t)^2}{4D_y t} - \frac{(z - u_z t)^2}{4D_z t} \right] \exp(-Kt) dt$$

③连续污染点源三维扩散稳态模型

当连续稳定的污染点源释放污染物的时间足够长, 这时污染物浓度可以看作将不再随时间的变化而变化, 而只会随三维空间位置的不同而发生改变, 其三维稳态的分析式是:

$$C(x, y, z) = \frac{M}{4\pi x \sqrt{D_x D_y}} \exp \left[-\frac{u_x y^2}{4D_y x} - \frac{u_x z^2}{4D_z x} \right] \exp \left[-K \frac{x}{u_x} \right]$$

将我们搜集的采样步长信息、污染源特征信息和其他各信息量代入基于稀疏系统辨识的气体扩散模型和基于傅里叶变换的重金属模型的公式, 即可得到污染

源的空间分布和污染源的强度。

六、模型的推广

本文运用神经网络模型对重金属污染源进行了预测,为一些用常规方法难以达到理想效果的问题的解决提供了一条崭新的途径。虽然模型建立过程中需要进行大量训练样本的学习,以及测试样本的检验,但是在先进的计算软件的支持下,我们仍然可以很方便地进行网络的设计和运算。在大大地降低了建模难度和建模时间的同时,减少了人为因素的干扰,得出的结论客观准确地反映了研究区的实际情况,可以为整个地区污染物的信息化监测与分析提供科学依据。通过本文的研究可以使当地政府能够对全地区重金属污染物的分布及其污染状况 有一个全面准确的了解,对当地质量的状况做出科学的判断,从而更好地服务于当地的土宜性评价及监控等。

参考文献

- [1]刘兆庆,徐方建,田旭,徐丰,赵永芳,李安春,江祖州,殷学博,胶州湾潮间带表层沉积物重金属污染评价[J],中国环境科学,2017,(06):2239-2247.
- [2]林承奇,胡恭任,于瑞莲,杨秋丽,余伟河,九龙江近岸表层沉积物重金属污染评价及来源解析[J],中国环境科学,2016,(04):1218-1225.
- [3]张慧,郑志志,马鑫鹏,杨欢,章桂芳,路中,乐容潮,哈尔滨市土壤表层重金属污染特征及来源辨析[J],环境科学研究, :1-13.
- [4]杨军,林岩龙,李龙杰,王小鹏,基于快速 Delaunay 三角化的散乱点曲面重建算法[J],计算机工程与科学,2015,(06):1189-1195.
- [5]胡大伟,卞新民,李思米,冯金飞,王书玉,基于神经网络的农田土壤重金属空间分布分析[J],农业环境科学学报,2007,(01):216-223.
- [6]王宇游,夏新,米方卓,陈纯,马芳,《土壤环境质量标准》中六种重金属测定精密度控制指标研究[J],土壤通报,2014,(06):1500-1504.
- [7]常静,城市地表灰尘—降雨径流系统污染物迁移过程与环境效应[D],华东师范大学,2007.
- [8]张慧敏,人类强烈影响地区土壤与灰尘中重金属的污染特征及风险评价[D],浙江大学,2017, [1]龚佃选,秦法宝,张梓婷,高晔.城市表层土壤重金属污染分析[J],资源节约与环保,2016,(04):183-184.
- [9]余潇潇,张林,李安国,基于稀疏系统辨识的空气污染源定位算法[J],清华大学学报(自然科学版),2013,(08):1077-1081.
- [10]马志宏,穆安冬,宫明,基于傅里叶变换的重金属扩散模型对污染源位置的研究[J],安徽农业科学,2013,(11):5014-5015+5118.

附录

附录一：第一问插值画图 Matlab 代码

```
A=load('F:\数学建模\暑假培训\11A\数据\地形.txt');
x=A(:,1);y=A(:,2);z=A(:,3);
[X,Y,Z]=griddata(x,y,z,linspace(74,28654)',linspace(0,18449)',v4');%插值
figure,contourf(X,Y,Z),colormap(flipud(gray)) %等高线图
figure,surf(X,Y,Z),colormap(flipud(gray))%三维曲面
```

附录二：第二问标准化表

As ($\mu\text{g/g}$)	Cd (ng/g)	Cr ($\mu\text{g/g}$)	Cu ($\mu\text{g/g}$)	Hg (ng/g)	Ni ($\mu\text{g/g}$)	Pb ($\mu\text{g/g}$)	Zn ($\mu\text{g/g}$)
-1.13899	-0.14655	-0.31445	-0.54258	-0.24554	-1.02954	-0.1155	0.234023
-1.33065	-0.17665	0.140581	1.156115	-0.05957	-0.90676	0.018751	0.117366
0.227172	-0.18001	-0.299	-0.44057	-0.18387	0.758008	-0.26749	0.056805
-0.1414	-0.15485	0.371223	1.018579	-0.06088	0.306439	-0.05975	0.011608
0.654713	-0.19069	-0.07682	-0.5271	-0.23525	0.216958	-0.31822	-0.01086
0.689113	-0.17436	-0.55944	1.145185	-0.05377	-1.31047	0.676885	-0.05675
1.008539	-0.18068	-0.16621	0.501228	0.063427	-0.28039	-0.21045	-0.07885
-0.29374	-0.02205	-0.11875	0.961197	-0.27337	-0.09519	-0.25852	-0.09935
0.3746	-0.1142	-0.31077	0.881955	-0.05377	-0.32617	0.522337	-0.10675
0.654713	-0.16177	0.18178	0.184259	-0.21702	1.569583	-0.14979	-0.11115
0.4434	-0.19423	-0.21072	0.611438	-0.24519	1.546693	-0.33381	-0.11935
2.325561	-0.11723	1.838212	0.300845	-0.12752	2.152253	0.161656	-0.12885
-0.0382	-0.17903	-0.04114	-0.09354	0.139318	-0.07438	-0.31341	-0.16118
-1.5911	-0.17963	-1.03728	-0.96521	-0.25845	-1.09821	-0.33189	-0.18361
1.642479	-0.18913	0.406537	0.467527	-0.19216	2.256301	-0.1516	-0.18368
-0.19054	-0.12295	-0.33431	-0.36497	-0.22365	0.333491	-0.25328	-0.21217
0.477799	-0.17743	-0.19527	0.238909	-0.17807	0.0484	0.305523	-0.22137
-0.08734	-0.15276	-0.10993	-0.00337	-0.20044	0.175338	-0.25894	-0.22534
-1.13899	-0.18818	-0.29495	0.013934	-0.24733	-0.13889	-0.26802	-0.22714
-0.74585	-0.19358	-0.35859	-0.26023	-0.25442	5.37E-04	-0.33146	-0.24007
-0.1414	-0.18717	0.115199	0.186992	-0.0804	-0.28039	-0.29462	-0.2496
-0.59842	-0.19734	-0.38802	-0.56626	-0.21702	-1.05243	-0.27304	-0.2798
0.065002	-0.19194	-0.17062	-0.30577	-0.14255	-0.51346	-0.28928	-0.28463
0.335286	-0.18867	-0.35675	-0.88414	-0.23027	-0.42606	-0.35838	-0.29143
0.123973	-0.19427	-0.13678	-0.85135	-0.24519	0.258577	-0.31726	-0.2923

0.020773	-0.1878	-0.3711	-0.73021	-0.23193	-0.15762	-0.30775	-0.29993
0.227172	-0.20144	-0.3722	-1.1492	-0.22033	-0.71948	-0.32986	-0.3159
-1.39453	-0.20119	-0.51787	-1.38874	-0.27005	-0.44687	-0.34204	-0.33293
0.3746	-0.20221	-0.28098	-0.71109	-0.2491	-0.04524	-0.38935	-0.337
1.092082	-0.19685	-0.11103	0.519445	-0.14741	-0.74029	-0.34193	-0.34713
-1.29625	-0.2001	-1.10349	-1.53812	-0.26176	-1.53105	-0.40943	-0.37953
0.020773	-0.17242	-0.16731	-0.10083	-0.17431	-0.04941	-0.18033	-0.17281
2.325561	-0.02205	1.838212	1.156115	0.139318	2.256301	0.676885	0.234023
-1.5911	-0.20221	-1.10349	-1.53812	-0.27337	-1.53105	-0.40943	-0.37953
-1.3648	5.741619	4.615517	3.617692	5.711852	0.874459	5.502824	5.675145

附录三：第二问因子分析表

表 1 工业区各重金属含量

类型	As (μg/g)	Cd (ng/g)	Cr (μg/g)	Cu (μg/g)	Hg (ng/g)	Ni (μg/g)	Pb (μg/g)	Zn (μg/g)
平均值	6.04	307.00	43.92	46.61	121.89	18.04	69.09	158.93
极大值	10.74	551.10	87.90	132.05	561.00	36.00	178.88	303.06
极小值	1.61	114.50	15.40	12.70	11.79	4.27	31.24	56.33
方差	6.527941	12427.8	284.1547	881.4663	16641.44	51.98749	1025.317	4738.96

表 2 山区各重金属含量

类型	As (μg/g)	Cd (ng/g)	Cr (μg/g)	Cu (μg/g)	Hg (ng/g)	Ni (μg/g)	Pb (μg/g)	Zn (μg/g)
平均值	4.04	152.13	37.02	16.50	38.35	14.60	36.39	71.03
极大值	10.99	407.60	92.76	69.06	85.71	45.45	113.84	172.53
极小值	1.80	40.00	16.20	2.29	9.64	5.51	19.68	32.86
方差	3.173744	6266.891	332	88.60942	357.0711	56.75862	317.4422	591.6472

表 3 交通区各重金属含量

类型	As (μg/g)	Cd (ng/g)	Cr (μg/g)	Cu (μg/g)	Hg (ng/g)	Ni (μg/g)	Pb (μg/g)	Zn (μg/g)
----	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------

平均值	5.48	336.48	51.63	52.45	100.68	16.48	61.56	189.02
极大值	30.13	886.60	306.02	277.82	900.00	27.28	171.14	853.98
极小值	1.61	50.10	15.32	12.34	8.57	6.19	22.01	40.92
方差	7.426707	39410.84	1226.884	2008.81	18699.53	20.37986	890.5014	19955.02

表 4 公园绿地区各重金属含量

类型	As (μg/g)	Cd (ng/g)	Cr (μg/g)	Cu (μg/g)	Hg (ng/g)	Ni (μg/g)	Pb (μg/g)	Zn (μg/g)
平均值	6.05	231.13	41.76	24.82	69.77	14.72	48.34	99.16
极大值	10.74	907.00	96.28	38.62	259.00	25.80	128.60	221.22
极小值	2.77	97.20	16.31	9.04	10.00	7.60	26.89	37.14
方差	3.230495	26814.03	171.7809	65.64565	3621.283	19.1596	580.4447	1853.691

工业区：

相关矩阵

	Zscore:	Zscore:	Zscore:	Zscore:	Zscore:	Zscore:	Zscore:	Zscore:
	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
Zscore:	1.000	.036	.276	.134	.053	.638	.118	.056
As								
Zscore:	.036	1.000	.924	.983	.998	.622	.983	.998
相关								
Cd								
Zscore:	.276	.924	1.000	.941	.922	.795	.951	.929
Cr								

Zscore:								
Cu	.134	.983	.941	1.000	.981	.702	.989	.989
Zscore:								
Hg	.053	.998	.922	.981	1.000	.622	.980	.997
Zscore:								
Ni	.638	.622	.795	.702	.622	1.000	.690	.642
Zscore:								
Pb	.118	.983	.951	.989	.980	.690	1.000	.987
Zscore:								
Zn	.056	.998	.929	.989	.997	.642	.987	1.000

KMO 和 Bartlett 的检验

取样足够度的 Kaiser-Meyer-Olkin 度量。		.866
近似卡方		684.492
Bartlett 的球形度检验	df	28
	Sig.	.000

解释的总方差

成份	初始特征值	提取平方和载入	旋转平方和载入
----	-------	---------	---------

	合计	方差 的 %	累积 %	合计	方差 的 %	累积 %	合计	方差 的 %	累积 %
1	6.405	80.057	80.057	6.405	80.057	80.057	6.056	75.695	75.695
2	1.354	16.931	96.987	1.354	16.931	96.987	1.703	21.293	96.987
3	.154	1.922	98.910						
4	.056	.702	99.612						
5	.020	.251	99.863						
6	.009	.110	99.973						
7	.001	.014	99.987						
8	.001	.013	100.000						

提取方法：主成份分析。

成份矩阵^a

	成份	
	1	2
Zscore: As	.212	.951
Zscore: Cd	.977	-.194
Zscore: Cr	.972	.087
Zscore: Cu	.990	-.086
Zscore: Hg	.977	-.182
Zscore: Ni	.764	.569

Zscore: Pb	.989	-.102
Zscore: Zn	.982	-.172

提取方法 :主成份。

a. 已提取了 2 个成份。

旋转成份矩阵^a

	成份	
	1	2
Zscore: As	-.045	.974
Zscore: Cd	.994	.070
Zscore: Cr	.915	.340
Zscore: Cu	.977	.177
Zscore: Hg	.990	.081
Zscore: Ni	.588	.750
Zscore: Pb	.981	.162
Zscore: Zn	.993	.092

提取方法 :主成份。

旋转法 :具有 Kaiser 标准化的正交旋转法。

a. 旋转在 3 次迭代后收敛。

成份得分系数矩阵

	成份	
	1	2
Zscore: As	-.153	.686
Zscore: Cd	.185	-.098
Zscore: Cr	.129	.102
Zscore: Cu	.166	-.021
Zscore: Hg	.182	-.090
Zscore: Ni	.005	.437
Zscore: Pb	.169	-.032
Zscore: Zn	.181	-.082

提取方法 :主成份。

旋转法 :具有 Kaiser 标准化的正交旋转法。

构成得分。

山区:

相关矩阵

	As (μ g/g)	Cd (ng/g)	Cr (μ g/g)	Cu (μ g/g)	Hg (ng/g)	Ni (μg/g)	Pb (μ g/g)	Zn (μ g/g)
相关 As (μg/g)	1.000	-.063	.154	.494	-.007	.340	-.010	.006

Cd								
(ng/g)	-.063	1.000	.885	.635	.913	.536	.904	.947
Cr (μg/g)	.154	.885	1.000	.754	.823	.826	.809	.896
Cu (μg/g)	.494	.635	.754	1.000	.680	.732	.685	.713
Hg								
(ng/g)	-.007	.913	.823	.680	1.000	.512	.883	.905
Ni (μg/g)	.340	.536	.826	.732	.512	1.000	.517	.643
Pb (μg/g)	-.010	.904	.809	.685	.883	.517	1.000	.940
Zn (μg/g)	.006	.947	.896	.713	.905	.643	.940	1.000

KMO 和 Bartlett 的检验

取样足够度的 Kaiser-Meyer-Olkin 度量。		.806
近似卡方		773.774
Bartlett 的球形度检验	df	28
	Sig.	.000

解释的总方差

成份	初始特征值	提取平方和载入	旋转平方和载入
----	-------	---------	---------

	合计	方差 的 %	累积 %	合计	方差 的 %	累积 %	合计	方差 的 %	累积 %
1	5.668	70.856	70.856	5.668	70.856	70.856	5.221	65.264	65.264
2	1.406	17.572	88.428	1.406	17.572	88.428	1.853	23.164	88.428
3	.480	6.003	94.431						
4	.182	2.277	96.708						
5	.126	1.573	98.282						
6	.080	.997	99.278						
7	.039	.488	99.766						
8	.019	.234	100.000						

提取方法：主成份分析。

成份矩阵^a

	成份	
	1	2
As (µg/g)	.160	.921
Cd (ng/g)	.930	-.278
Cr (µg/g)	.954	.046
Cu (µg/g)	.831	.423
Hg (ng/g)	.914	-.225
Ni (µg/g)	.753	.404

Pb (μg/g)	.918	-.227
Zn (μg/g)	.965	-.179

提取方法 :主成份。

a. 已提取了 2 个成份。

旋转成份矩阵^a

	成份	
	1	2
As (μg/g)	-.147	.923
Cd (ng/g)	.970	.038
Cr (μg/g)	.888	.352
Cu (μg/g)	.649	.670
Hg (ng/g)	.938	.083
Ni (μg/g)	.581	.626
Pb (μg/g)	.942	.082
Zn (μg/g)	.971	.143

提取方法 :主成份。

旋转法 :具有 Kaiser 标准化的正交旋转法。

a. 旋转在 3 次迭代后收敛。

成份得分系数矩阵

	成份	
	1	2
As (μg/g)	-.186	.629
Cd (ng/g)	.219	-.134
Cr (μg/g)	.149	.085
Cu (μg/g)	.041	.332
Hg (ng/g)	.204	-.099
Ni (μg/g)	.032	.315
Pb (μg/g)	.206	-.101
Zn (μg/g)	.202	-.066

提取方法 :主成份。

旋转法 :具有 Kaiser 标准化的正交旋转法。

构成得分。

交通区：

相关矩阵

	As (μg/g)	Cd (ng/g)	Cr (μg/g)	Cu (μg/g)	Hg (ng/g)	Ni (μg/g)	Pb (μg/g)	Zn (μg/g)
Zscore: 相关 As	1.000	.057	.191	.107	.075	.445	.119	.067

Zscore:	.057	1.000	.933	.968	.996	.088	.932	.997
Cd								
Zscore:	.191	.933	1.000	.940	.943	.192	.919	.943
Cr								
Zscore:	.107	.968	.940	1.000	.968	.159	.953	.975
Cu								
Zscore:	.075	.996	.943	.968	1.000	.102	.935	.996
Hg								
Zscore:	.445	.088	.192	.159	.102	1.000	.182	.103
Ni								
Zscore:	.119	.932	.919	.953	.935	.182	1.000	.942
Pb								
Zscore:	.067	.997	.943	.975	.996	.103	.942	1.000
Zn								

KMO 和 Bartlett 的检验

取样足够度的 Kaiser-Meyer-Olkin 度量。	.901
近似卡方	2422.882
Bartlett 的球形度检验	
df	28

Sig.	.000
------	------

解释的总方差

成份	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差 的 %	累积 %	合计	方差 的 %	累积 %	合计	方差 的 %	累积 %
1	5.822	72.774	72.774	5.822	72.774	72.774	5.759	71.993	71.993
2	1.422	17.780	90.554	1.422	17.780	90.554	1.485	18.560	90.554
3	.557	6.962	97.515						
4	.088	1.096	98.611						
5	.074	.923	99.534						
6	.032	.396	99.930						
7	.004	.048	99.977						
8	.002	.023	100.000						

提取方法：主成份分析。

成份矩阵^a

	成份	
	1	2
Zscore: As	.142	.840

Zscore: Cd	.986	-.106
Zscore: Cr	.967	.048
Zscore: Cu	.985	-.030
Zscore: Hg	.989	-.087
Zscore: Ni	.181	.828
Zscore: Pb	.965	-.003
Zscore: Zn	.991	-.092

提取方法 :主成份。

a. 已提取了 2 个成份。

公园绿化区：

相关矩阵

	Zscore: As (μ g/g)	Zscore: Cd (ng/g)	Zscore: Cr (μg/g)	Zscore: Cu (μ g/g)	Zscore: Hg (ng/g)	Zscore: Ni (μg/g)	Zscore: Pb (μ g/g)	Zscore: Zn (μ g/g)
Zscore: As (μg/g)	1.000	-.223	.241	.225	-.187	.655	-.097	-.199
Zscore: Cd (ng/g)	-.223	1.000	.817	.647	.994	.169	.963	.990
Zscore: Cr (μg/g)	.241	.817	1.000	.768	.832	.578	.852	.838

Zscore:								
Cu (µg/g)	.225	.647	.768	1.000	.677	.406	.749	.693
Zscore:								
Hg (ng/g)	-.187	.994	.832	.677	1.000	.173	.969	.990
Zscore:								
Ni (µg/g)	.655	.169	.578	.406	.173	1.000	.215	.201
Zscore:								
Pb (µg/g)	-.097	.963	.852	.749	.969	.215	1.000	.970
Zscore:								
Zn (µg/g)	-.199	.990	.838	.693	.990	.201	.970	1.000

KMO 和 Bartlett 的检验

取样足够度的 Kaiser-Meyer-Olkin 度量。		.826
近似卡方		505.470
Bartlett 的球形度检验	df	28
	Sig.	.000

解释的总方差

成份	初始特征值	提取平方和载入	旋转平方和载入
----	-------	---------	---------

	合计	方差 的 %	累积 %	合计	方差 的 %	累积 %	合计	方差 的 %	累积 %
1	5.385	67.311	67.311	5.385	67.311	67.311	5.286	66.072	66.072
2	1.878	23.475	90.786	1.878	23.475	90.786	1.977	24.714	90.786
3	.386	4.824	95.610						
4	.239	2.986	98.596						
5	.071	.893	99.489						
6	.028	.351	99.840						
7	.009	.113	99.953						
8	.004	.047	100.000						

提取方法：主成份分析。

旋转成份矩阵^a

	成份	
	1	2
Zscore: As (µg/g)	-.166	.919
Zscore: Cd (ng/g)	.985	-.076
Zscore: Cr (µg/g)	.865	.434
Zscore: Cu (µg/g)	.751	.389
Zscore: Hg (ng/g)	.989	-.049
Zscore: Ni (µg/g)	.217	.884

Zscore: Pb (μg/g)	.983	.035
Zscore: Zn (μg/g)	.992	-.038

提取方法 :主成份。

旋转法 :具有 Kaiser 标准化的正交旋转法。

a. 旋转在 3 次迭代后收敛。

成份得分系数矩阵

	成份	
	1	2
Zscore: As (μg/g)	-.085	.490
Zscore: Cd (ng/g)	.197	-.096
Zscore: Cr (μg/g)	.144	.177
Zscore: Cu (μg/g)	.124	.160
Zscore: Hg (ng/g)	.196	-.082
Zscore: Ni (μg/g)	-.008	.450
Zscore: Pb (μg/g)	.190	-.038
Zscore: Zn (μg/g)	.196	-.077

提取方法 :主成份。

旋转法 :具有 Kaiser 标准化的正交旋转法。

构成得分。

附录四：第三问基于神经网络的污染源定位的 Matlab 代码

```
clc, clear
a=load('F:\数学建模\暑假培训\11A\数据\新建文件夹\As.txt'); %根据重金属种类修改
a=a'; %注意神经网络的数据格式，不要把矩阵搞转置了。
P=a([1:3],[1:end-1]);
[PN,PS1]=mapminmax(P); %自变量数据规格化到[-1,1]
T=a(4,[1:end-1]);
[TN,PS2]=mapminmax(T); %因变量数据规格化到[-1,1]
net1=newrb(PN,TN) %训练 RBF 网络
x=a([1:3],end);
xn=mapminmax('apply',x,PS1); %预测样本点自变量规格化
yn1=sim(net1,xn);
y1=mapminmax('reverse',yn1,PS2) %求预测值，并把数据还原
delta1=abs(a(4,292)-y1)/a(4,292) %计算 RBF 网络预测的相对误差
net2=feedforwardnet(3); %初始化 BP 网络，隐含层的神经元取为 4 个（多次试验）
net2 = train(net2,PN,TN); %训练 BP 网络
x1=round(30000*rand(100,1));
x2=round(20000*rand(100,1));
x3=round(200*rand(100,1));
x=[x1 x2 x3]
x=x';
xn=mapminmax('apply',x,PS1); %预测样本点自变量规格化
yn2=net2(xn);
y2=mapminmax('reverse',yn2,PS2)%求预测值，并把数据还原
y2=y2';
[X,Y,Z]=griddata(x1,x2,y2,linspace(74,28654)',linspace(0,18449),'v4');%插值
figure,contourf(X,Y,Z),colormap(flipud(gray)) %等高线图
figure,surf(X,Y,Z),colormap(flipud(gray))%三维曲面
```