

基于统计分析和 BP 网络的土壤重金属污染研究

孙娓娓,黄映雪,张 杰,胡业刚

阜阳师范学院数学与统计学院, 安徽 阜阳 236041

摘要: 土壤中重金属不仅对植物生长造成影响, 还通过食物链危害人类健康, 并可能造成大气和水污染等问题, 影响和改变城市土壤的生态功能。因此, 为了改善生活环境, 彻底解决污染问题, 就必须找出重金属污染的原因。本文通过污染指数评价、统计相关分析、BP网络空间插值等方法建立数学模型, 得到了土壤重金属污染程度、原因及污染源位置, 为环境保护和城市发展提供重要依据。

关键词: 重金属污染; 污染指数; 相关分析; BP 网络

中图分类号: O29

文献标识码: A

文章编号: 1000-2324(2016)04-0536-04

Study on the Pollution of Heavy Metal in Soil Based on Statistical Analysis and BP Network

SUN Wei-wei, HUANG Ying-xue, ZHANG Jie, HU Ye-gang

School of Mathematics and Statistics/Fuyang Normal College, Fuyang 236041, China

Abstract: Heavy metals in soil not only affect the growth of plants but also harm to people's health through a food chain, and may cause the problem such as air and water pollution to influence the ecological function of urban soil. Therefore, in order to improve the living environment and solve the pollution problem thoroughly, we must find out the cause of the heavy metal pollution. In this article, the mathematical model was established through pollution index evaluation, statistical analysis, BP network spatial interpolation. The soil heavy metal pollution degree, reason and pollution source location are obtained. It provides important basis for environmental protection and urban development.

Keywords: Heavy metal pollution; pollution index; correlation analysis; BP network

城市土壤是城市生态系统的重要组成部分。对土壤地质环境异常的查证, 以及应用已获取的海量数据资料对城市环境质量进行评价是很有必要的。研究人类活动影响下城市土壤环境的演变模式, 对于城市的生态建设、提高农产品食品安全和人民身体健康、以及可持续发展都具有十分重要的意义。本文对某城区土壤重金属元素 Cu、Zn、Pb、Cd、Ni、Cr、As、Hg 的含量和空间分布进行统计分析探讨, 利用 BP 网络确定污染源的位置, 查明重金属污染的影响因素及对环境的潜在危害, 以期了解不同的城区活动对土壤重金属形态分布的作用及影响, 为评价重金属的环境效应和治理保障城市居民的身心健康提供依据。

1 采集样本

由于人类活动对不同环境的影响程度不同, 一般将城区化为 5 个功能区: 生活区、工业区、山区、主干道路区、公园绿地区^[1]。为了全面地分析城市土壤重金属污染问题, 首先要进行土壤采样。为使样点具有较好的代表性, 采用近似网格法 (1 km×1 km) 布点, 在不同功能区共采集土样 319 个。根据 GPS 记录采样点的位置, 利用 MATLAB 软件绘制出样点分布如图 1 所示。然后利用专门

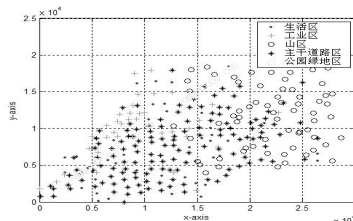


图 1 样点分布图

Fig.1 Distribution of samples

收稿日期: 2016-03-30

修回日期: 2016-06-03

基金项目: 安徽高校自然科学基金项目(2015KJ003,KJ2015A161);安徽省自然科学基金(1508085MA12);全国统计科学研究项目(2014LY088);安徽省质量工程项目(2015jxt023; 2015jxt0121);阜阳师范学院自然科学研究项目(2015FSKJ07)

作者简介: 孙娓娓(1981-),女,硕士,讲师,主要从事智能计算, 数据挖掘工作. E-mail:93692849@qq.com

仪器测试, 获得每个样本所含 Cu、Zn、Pb、Cd、Ni、Cr、As、Hg 8 种重金属元素的浓度数据。此外, 按照 2 km 的间距在远离人群和工业的自然区取样, 将其作为该城区土壤中元素的背景值。

2 重金属的污染程度

目前, 对土壤重金属污染主要使用的评价方法, 包括地积累指数法、污染指数法、潜在生态危害指数法^[2]等等。本文采用最普遍的单因子指数法和内梅罗综合污染指数法^[3], 可以评价不同类型的土壤受到重金属污染的强度, 具体公式为:

$$p_i = \frac{c_i}{S_i}, \quad p_{综} = \sqrt{\frac{\left(\max \frac{c_i}{S_i}\right)^2 + \left(\frac{1}{n} \sum_{i=1}^n \frac{c_i}{S_i}\right)^2}{2}}$$

结合数据, 得到该城市 5 个分类区土壤中 8 种重金属的单因子污染指数 p_i 和内梅罗综合污染指数 $p_{综}$, 如表 1 所示。

表 1 土壤重金属污染指数
Table 1 Pollution index of heavy metals in soil

功能区 Function region	单因子指数 Single factor index								内梅罗综合指数 Nemerow comprehensive index
	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn	
生活区	1.53	2.28	1.70	3.57	5.96	1.39	1.93	2.84	4.516
工业区	1.59	2.40	1.80	4.59	9.69	1.44	2.08	3.16	7.248
山区	1.48	2.01	1.34	2.58	5.92	1.31	1.68	1.97	4.487
主干道区	1.57	2.33	1.73	4.17	8.58	1.40	1.99	2.92	6.448
公园绿地区	1.56	2.27	1.70	3.49	5.80	1.39	1.91	2.78	4.498

为了说明该城区不同区域重金属的污染程度, 依据中国《绿色食品产地环境质量现状评价纲要(试行)》规定^[4], 相关标准如表 2、3 所示。

表 2 土壤中各元素污染程度分级标准
Table 2 Pollution classification standard of elements in the soil

等级划分 Level classification	1	2	3	4
单因子污染指数	$p_i \leq 1$	$1 < p_i \leq 2$	$2 < p_i \leq 4$	$p_i \geq 4$
污染程度	非污染	轻度污染	中度污染	重度污染

表 3 土壤综合污染程度分级标准
Table 3 Classification standard of comprehensive pollution in soil

等级划分 Level classification	1	2	3	4	5
综合污染指数	$p_{综} \leq 1$	$1 < p_{综} \leq 2$	$2 < p_{综} \leq 3$	$3 < p_{综} \leq 5$	$p_{综} \geq 5$
污染程度	安全	警戒线	轻度污染	中度污染	重度污染

通过研究对比可知, 5 个功能区土壤重金属污染存在一定差异。具体如下:

1) 重金属 Hg 在 5 个分类区均属于重度污染, 尤其工业区和主干道区的污染指数分别为 9.69、8.58, 明显严重超标。As、Cr、Ni、Pb 这 4 种元素在 5 个分类区均属于轻度污染。Cd 在 5 个分类区均属于中度污染。Zn 除了在山区属于轻度污染, 在其他 4 个区均属于中度污染。Cu 除了在工业区和主干道区属于重度污染, 在其他 3 个区均属于中度污染。

2) 通过综合污染指数可以看出, 工业区和主干道区属于重度污染, 其余 3 个区都为中度污染, 且其污染程度从大到小依次为工业区、主干道区、生活区、公园绿地区、山区。

3) 综合所有数据, 在工业区和主干道区所有重金属浓度都明显高于其他功能区, 表明工业污染与交通污染已经超过其他污染, 成为该城市的主导污染源。

3 污染的原因

土壤中重金属不仅对植物生长造成影响, 还通过食物链危害人类健康, 并可能造成大气和水污染等问题。因此, 为了改善生活环境, 彻底解决污染问题, 就必须找出重金属污染的原因。

首先利用 Excel 计算出该城市土壤中 8 种重金属的均值、标准差、变异系数等统计量，具体结果如表 4 所示。

表 4 土壤中重金属含量统计
Table 4 Statistics of heavy metals in soil

项目 Item	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
最小值	1.61	40.00	15.32	2.29	8.57	4.27	19.68	32.86
最大值	30.13	1619.80	920.84	2528.48	16000	142.50	472.48	3760.82
平均值	5.68	302.40	53.51	55.02	299.71	17.26	61.74	201.20
标准差	3.02	225.27	8.36	12.75	42.30	9.93	49.98	22.39
变异系数(%)	53	74	11	23	14	57	81	11
超标率(%)	77.4	79.6	80.6	88.4	66.5	75.2	81.5	79.0
背景值	3.6	130	31	13.2	35	12.3	31	69

通过表 4 数据可知：

(1) 该城市土壤中各种重金属的平均含量均高于背景值；As、Cd、Cr、Cu、Hg、Ni、Pb、Zn 分别是背景值的 1.58、2.33、1.73、4.17、8.56、1.40、1.99、2.92 倍，说明 Cu、Hg 都有一定程度上累积，主要来源于人类的工业生产活动以及交通活动。

(2) 依据变异系数大小粗略分级规律，土壤重金属含量变异系数在 28.8%~60.62%之间，属于中等程度变异^[5]。则该城市土壤中 As、Cd、Pb 为强变异，Ni 为中等变异，Cr、Cu、Hg、Zn 均为弱变异；8 种元素变异系数范围为 11%~81%，变异性很大，可见土壤来源受外界影响很大，可能受人为活动的不合理布局及企业和道路交通的影响。

其次，元素地球化学研究表明，成因相似的元素之间常有较好的相关性^[5]。因此，统计上相关性较高的重金属元素之间在成因上也有相似之处。运用 SPSS 统计软件进行因子分析，可以得出表 5 的重金属之间的相关系数。

表 5 土壤中各重金属含量的相关系数
Table 5 Correlation coefficient of heavy metal content in the soil

元素 Elements	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
As	1							
Cd	0.2547	1						
Cr	0.1890	0.3524	1					
Cu	0.1597	0.3967	0.5316	1				
Hg	0.0644	0.2647	0.1032	0.4167	1			
Ni	0.3166	0.3294	0.7158	0.4946	0.1029	1		
Pb	0.2899	0.6603	0.3828	0.5200	0.2981	0.3068	1	
Zn	0.2469	0.4312	0.4243	0.3873	0.1958	0.4364	0.4937	1

由表 5 可见，几种重金属含量都与至少 1 种或其他几种重金属呈正相关；Cr、Ni 相关性最强，说明来源大致相同，并且相关系数最大为 0.7158，说明土壤中 Cr、Ni 的关系最为密切，他们的含量彼此影响较大。其次 Cd、Pb 呈显著的正相关，说明来源极为相近。Zn、Cd、Cr、Cu、Ni、Pb 呈正相关，可能暗示了在研究区域内存在相似的过程控制着土壤重金属元素的含量的分布特征。但 Pb、Zn、Cd、Cu 在不同功能区中取的最大值，因此，Pb、Zn、Cd、Cu 没有产生复合污染物，可能具有相同的来源。

4 污染源的位置

若某种重金属的浓度在某一处取得最大值，就可以认为该处是污染源的位置。按照这个思想，我们试图建立重金属的浓度 ρ 与样本点三维空间坐标分量 (x, y, h) 的函数关系式，利用 319 个数据进行三元函数 $\rho(x, y, h)$ 的插值或拟合，然后求该函数的最大值。但由于 Matlab 软件无法实现四维散点数据普通的插值或拟合，于是我们考虑统计回归模型。为了找出重金属浓度 ρ 与坐标 x, y, h 的关系式，首先利用数据分别作出 ρ 对 x, y, h 的散点图，如元素 As 效果如图 2 所示。

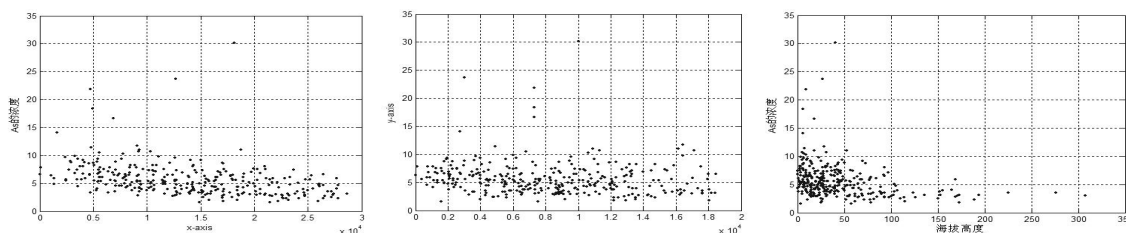


图 2 As 元素的散点图

Fig.2 Scattered points of As element

但由图 2 无法看出 ρ 与 x 、 y 、 h 具有明显的函数关系, 因此很难进一步建立 ρ 与 x 、 y 、 h 的回归模型。事实上, 土壤中各种重金属的含量与其空间位置之间存在着高度复杂的非线性映射关系, 因此不宜用常规的建模方法来解决此类问题, 也难以得到重金属含量与空间位置的定量关系。

仅根据样本中的数据显然不能求出整个城区的污染源的位置, 为了能够充分有效地利用样本数据的信息量, 确定某重金属在该城区的最大浓度的空间位置, 我们仍然利用插值法求解, 但本文采用 BP 神经网络的方法进行空间加密插值^[6], 它能学习和存贮大量的输入—输出模式映射关系, 而无需事前揭示描述这种映射关系的数学方程。因此, 可以有效地将研究区域内的海拔高程融入网络, 提高网络的稳定性和精度。具体算法如下:

(1) 确定 BP 网络拓扑结构。输入节点与输出节点是由问题的本身决定的, 该问题采用有 3 个输入节点, 分别表示样本点的三维坐标 x 、 y 、 h ; 输出节点为 8 个, 分别表示该样本点中 8 种重金属的浓度; 隐层节点数凭经验取 6, 因此这里 BP 网络选用 3-6-8 的拓扑结构。

(2) 随机采用 250 个样本数据作为训练集, 剩余的 69 个样本作为测试数据。对网络在 $[-1, 1]$ 范围内的初始权值随机取值, 学习率 $\eta=0.9$, 误差精度设置为 0.0001, 最大迭代步数为 10000, 当满足设置的误差精度或最大迭代步数, 算法就停止训练。从而确定重金属浓度与空间位置的映射关系。

(3) 用训练好的 BP 网络去进行空间插值。将所有样本点缩小其取样间隔 (如 10 m), 将所有加密点的 x 、 y 、 h 数据一一输入该 BP 网络, 网络系统就会自动算出该点 8 种重金属的浓度。

(4) 由所有插值点和样本点的重金属浓度的最大值, 求出坐标确定污染源的位置(表 6)。

表 6 重金属污染源位置

Table 6 Locations of pollution sources of heavy metals

	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
$x(m)$	18134	18101	18034	18014	17985	17934	18253	19002
$y(m)$	10046	10012	9946	9926	9883	10085	10093	9887
$h(m)$	41	43	42	43	44	41	45	46
区域	4	4	4	2	2	4	1	4

由表 6 可以看出, As、Zn、Cr、Ni、Cd 的污染源都在 4 区, 即主干道路区; Cu、Hg 的污染源都在 2 区, 即工业区; Pb 的污染源在 1 区, 即生活区。

5 结论

本文就土壤重金属污染问题, 分别利用统计分析、BP 网络高维插值等方法, 得到了污染程度、原因、污染源位置及传播特征相关结论。该模型不仅可以用于本文未提及的其他重金属污染问题的讨论, 而且还可用于其他地区的土壤, 甚至大气、水流污染等问题, 具有一定的推广性。

参考文献

- [1] 李春明.城市表层土壤重金属空间分布特性及影响因素[J].晋中学院学报,2014,31(3):24-27
- [2] 汤洁,陈初雨,李海毅,等.大庆市建成区土壤重金属潜在生态危害和健康风险评价[J].地理科学,2011,31(1):118-122
- [3] 尹俊,柳云龙.上海市城郊土壤重金属空间分布及其污染评价[J].现代农业科技,2010(10):251-255
- [4] 王海东,方凤满,谢宏芳,等.芜湖市区土壤重金属污染评价及来源分析[J].城市环境与城市生态,2010,23(14):36-40
- [5] 钱翌,张玮,冉德超.青岛城市土壤重金属的形态分布及影响因素分析[J].环境化学,2011,30(3):652-654
- [6] 胡大伟,卞新民,王书玉,等.基于BP模型的南通市农田土壤重金属空间分布研究[J].安全与环境学报,2007,7(1):91-95
- [7] 陈俊坚,张会化,刘鉴明,等.广东省区域地质背景下土壤表层重金属元素空间分布特征及其影响因素分析[J].生态环境学报,2011,20(4):646-651