

基于拉依达检验法的场地环境调查土壤 重金属异常值识别的研究

董聪慧^{1,2}, 袁一¹, 刘小喜¹, 梅祖明^{2*}

- (1. 生态环境部太湖流域东海海域生态环境监督管理局生态环境监测与科学研究中心, 上海 200125;
2. 上海市岩土工程检测中心有限公司, 上海 200436)

摘要: **目的** 研究一种场地环境调查土壤重金属异常值识别的方法。**方法** 首先对土壤重金属样本进行正态分布检验, 其次采用拉依达检验法识别出重金属的异常值, 最后频率直方图验证异常值识别结果。**结果** 案例研究表明, 采用拉依达检验法能够有效的识别出场地环境调查土壤重金属含量的异常值。**结论** 采用拉依达检验法识别场地环境调查土壤重金属含量的异常值, 有助于识别低于筛选值但是受人类生产生活影响较大的土壤, 能够大大降低土壤累积污染的风险。

关键词: 拉依达检验法; 正态分布; 频率直方图; 土壤重金属; 异常值

Research on the method to identify the outliers of the heavy metal in soil of site environmental investigation based on the PauTa criterion

DONG Cong-Hui^{1,2}, YUAN Yi¹, LIU Xiao-Xi¹, MEI Zu-Ming^{2*}

- (1. *Research Center for Monitoring and Environmental Sciences, Taihu Basin & East China Sea Ecological Environment Supervision and Administration Authority, Ministry of Ecology and Environment, Shanghai 200125, China;*
2. *Shanghai Geotechnical Engineering Detecting Centre Co., Ltd., Shanghai 200436, China*)

ABSTRACT: Objective A method to identify heavy metal outliers in soil of site environmental investigation was studied. **Methods** First, the normal distribution test of soil heavy metal samples was carried out, then the heavy metal outliers were identified by the PauTa criterion, and the frequency histogram was used to verify the outliers identification results. **Results** The results of the case study show that the PauTa criterion can effectively identify the outliers of heavy metal content in the soil of the site environmental investigation. **Conclusion** The use of PauTa criterion to identify the outliers of heavy metal content in soil of site environmental investigation is helpful to identify the soil that is lower than the screening value but greatly affected by human production and life, and can greatly reduce the risk of soil cumulative pollution.

KEY WORDS: PauTa criterion; normal distribution; frequency histogram; heavy metal in soil; outliers

*通信作者: 梅祖明, 高级工程师, 研究方向为环境污染防治、环境科学研究。E-mail: 985812141@qq.com

*Corresponding author: MEI Zu-Ming, Senior Engineer, Shanghai Geotechnical Engineering Detecting Centre Co., Ltd., Shanghai 200436, China. E-mail: 985812141@qq.com

0 引言

场地环境调查是指通过对场地土壤和地下水等的调查确定场地是否被污染,及污染程度和范围,土壤重金属是衡量土壤是否受到污染的重要指标。在场地环境调查工作中,一般将土壤重金属含量与《土壤环境质量 农用地土壤风险管控标准》^[1]进行比对,低于筛选值则认为对农产品质量安全、农作物生长或土壤生态的风险低,或与《土壤环境质量建设用土壤污染风险管控标准》^[2]进行比对,未超过筛选值则认为对人体健康的风险可以忽略^[3-4]。但是,根据2024年1月《中共中央国务院关于全面推进美丽中国建设的意见》的要求,开展土壤污染源头防控行动,严防新增污染,逐步解决长期积累的土壤和地下水严重污染问题^[5]。国家对土壤污染的管控提出了更高的要求,按照管控标准判定是否被污染则会忽略一些虽然没有超过筛选值,但明显高于土壤背景值的数据,这种数据称为异常数据、异常值或离群值。这些土壤重金属含量异常可能是人类的生产和生活活动造成的^[6-7],存在污染环境的风险,或存在为环境带来累积污染的风险^[8-9]。在场地用地类型变更前,快速筛选出这些异常数据,及时溯源并进行处理可以大大降低环境污染的风险。

《正态样本离群值的判断和处理》^[10]和相关研究^[11-13]中给出的异常值检验方法包括拉依达检验法、箱式图法、奈尔检验法、格拉布斯检验法和狄克逊检验法等。众多学者也对环境数据异常值识别的方法进行了研究,王景云^[14]采用数理统计方法、土壤背景值和空间自相关方法对北京市土壤重金属As、Cd和Pb元素中的异常值识别进行了研究;刘庚等^[15]运用稳健变异函数和平尾化方法对污染场地土壤重金属的空间异常值的识别进行了研究;何朝红^[16]研究了一种运用格拉布斯检验法剔除土壤半挥发性有机物的离群值的方法,提升检测结果的准确度;Weijun Fu等^[17],采用直方图、箱式图和莫兰指数相结合的方法识别土壤磷素的异常值;Xiao-Lin Sun等^[18]进行了中位值克里格法识别土壤污染异常值和空间预测的研究,取得了较好的效果;Chaosheng Zhang等^[19]通过莫兰指数的局部空间相关指标、克里格交叉验证和地理加权回归来识别土壤中重金属的异常值。近年来部分学者运用拉依达检验法进行环境异常值的识别,侍建国等^[20]运用拉依达检验法剔除区域水文数据异常值;王磊等^[21]分别运用拉依达准则、水化学图法+拉依达准则、信息筛选法+拉依达准则识别地下水化学组分的异常值,异常识别后的剩余数据阈值范围一致;仝伟亮等^[22]运用拉依达准则原则计算

出了区域垃圾产生量的分布范围;周秘^[23]采用拉依达准则研究排除基坑监测数据中粗大误差的方法,方法快速、方便、有效。

莫兰指数和克里格法更倾向于空间的相关性;格拉布斯检验法是通过逐个判定样本的最大值和最小值是否离群来依次识别样本异常值,计算过程缓慢;拉依达检验法又名 3σ 检验法,是指先假设一组检测数据只含有随机误差,按一定概率确定一个区间,认为凡超过这个区间的误差,就不属于随机误差而是粗大误差^[24],含有该误差的数据应予以剔除。拉依达检验法适用于未知标准差的情况,该准则检验数据基于正态分布,计算简单,当样本量充分大时,运用拉依达检验法能够快速筛查异常数据^[25]。为了能够快速识别场地环境调查数据中的异常值,本研究将采用拉依达检验法,快速对土壤重金属数据进行异常数据识别,并验证其异常值的识别结果。

1 异常值的识别过程

1.1 分布属性的判定

首先判定检测结果的分布类型是否服从正态分布。运用SPSS软件计算偏度值和偏度标准误差、峰度值和峰度标准误差,按照公式(1)和公式(2)分别计算偏度Z-评分和峰度Z-评分。在 $\alpha=0.05$ 的检验水平下,若偏度Z-评分和峰度Z-评分均在 ± 1.96 之间,则可认为服从正态分布,否则判定为服从偏正态分布。

$$\text{偏度 Z-评分} = \text{偏度值} / \text{偏度标准误差} \quad (1)$$

$$\text{峰度 Z-评分} = \text{峰度值} / \text{峰度标准误差} \quad (2)$$

对于服从正态分布的检测结果样本,若样本重金属含量值大于99%置信概率的临界值上限则判定数据为异常数据,判定公式见式(3):

$$X > \bar{x} + 3\sigma \quad (3)$$

式中,X表示样本重金属含量值; \bar{x} 表示样本重金属含量的平均值; σ 表示样本重金属含量的标准偏差。

如数据服从偏正态分布,则将所有检测结果取对数,将取对数后的结果重新进行分布类型判定,若服从正态分布则选择几何平均值法进行统计分析。判定公式见式(4):

$$X > 10^{\lg G + 3\sigma_g} \quad (4)$$

式中X表示样本重金属含量值;G表示样本重金属含量的几何平均值; σ_g 表示样本重金属含量的几何标准偏差。

1.2 异常值的判定

土壤重金属含量频率直方图能够直观地表示出样品中各含量的主要分布情况,以分组含量为直方图的横坐标,以频数为纵坐标绘图,观察离散情况,以判断异常值识别的效果。

结合人类生产生活情况,对识别出的异常值产生的原因进行分析,并应慎重处理,减少土壤累积污染的风险。

2 结果与分析

2.1 案例概况

案例为上海某区场地,用地历史为工业用地,共采集了

表1 重金属含量平均值和标准偏差

Table 1 Mean value and standard deviation of heavy metal content

元素	砷	镉	铬	铜	汞	铅	镍	锌
平均值/几何平均值 (mg/kg)	9.00	0.16	78.0	29.2	0.113	29.2	33.0	94.2
标准偏差/几何标准偏差	0.0944	0.197	7.32	0.0795	0.253	0.0967	4.92	0.110

2.3 异常值识别

运用拉依达检验法对样本中的所有重金属数据分别进行异常值识别的研究,结果见图1,位于Xmax线上的数据点即为异常值,共识别出土壤镉异常值3个,砷、铜、铅、锌异常值各1个。其中tr62点位镉和铅2项重金属被识别为异常值,见表2所示。

表2 重金属异常值对应的点位列表

Table 2 The points to abnormal heavy metal outliers

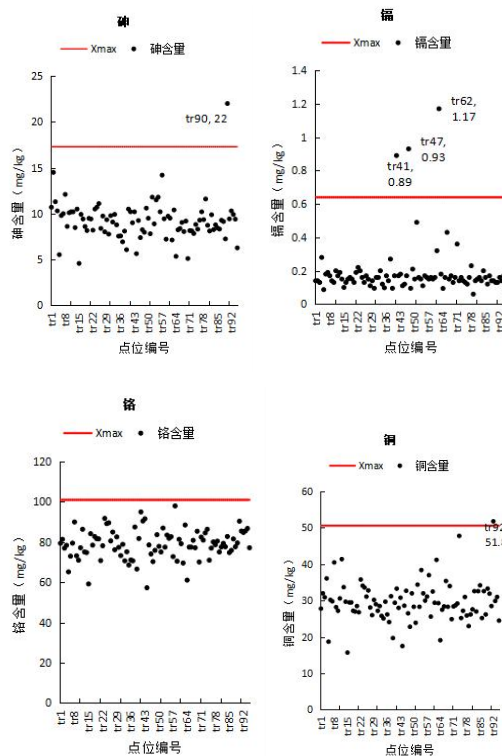
异常点位	异常金属元素
tr22	锌
tr41	镉
tr47	镉
tr62	镉、铅
tr90	砷
tr92	铜

95个土壤样品,测试的重金属指标包括砷、镉、铬、铜、汞、铅、锌和镍,重金属含量均未超用地类型对应的筛选值。样本数据量充足,采用拉依达检验法对案例的重金属含量异常值进行分析研究。

2.2 数据预处理

分别计算8项重金属的偏度Z-评分和峰度Z-评分,结果铬和镍两项Z-评分均小于1.96,判定铬和镍数据样本服从正态分布,其它重金属服从偏正态分布,对砷等6项重金属含量取对数,重新进行验证后服从正态分布,8项重金属含量的样本均满足拉依达检验法的要求。

计算重金属含量的均值和标准偏差,见表1。



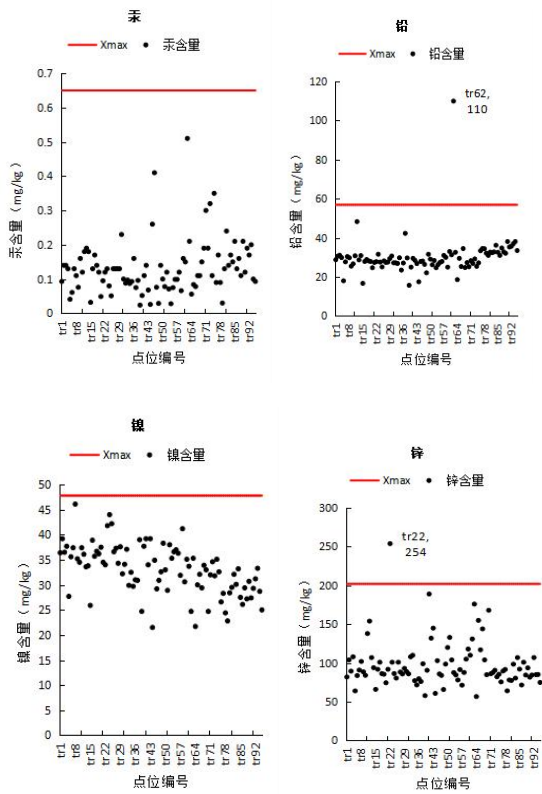


图 1 重金属异常值识别

Fig.1 Heavy metal outlier recognition

2.4 异常值识别结果验证

对比运用拉依达检验法识别异常值前后土壤重金属的频率直方图, 如图 2 所示。由图可知, 运用拉依达检验法识别出了重金属检测结果中离群的异常值。这些异常值点位受人类生产生活影响较大, 对识别出的异常值产生的原因进行分析, 追本溯源包括调查污染源和迁移途径等, 结合分析结果针对偶发性输入和长期必然性输入分别采用不同的处理办法, 减少产生土壤累积污染的风险。

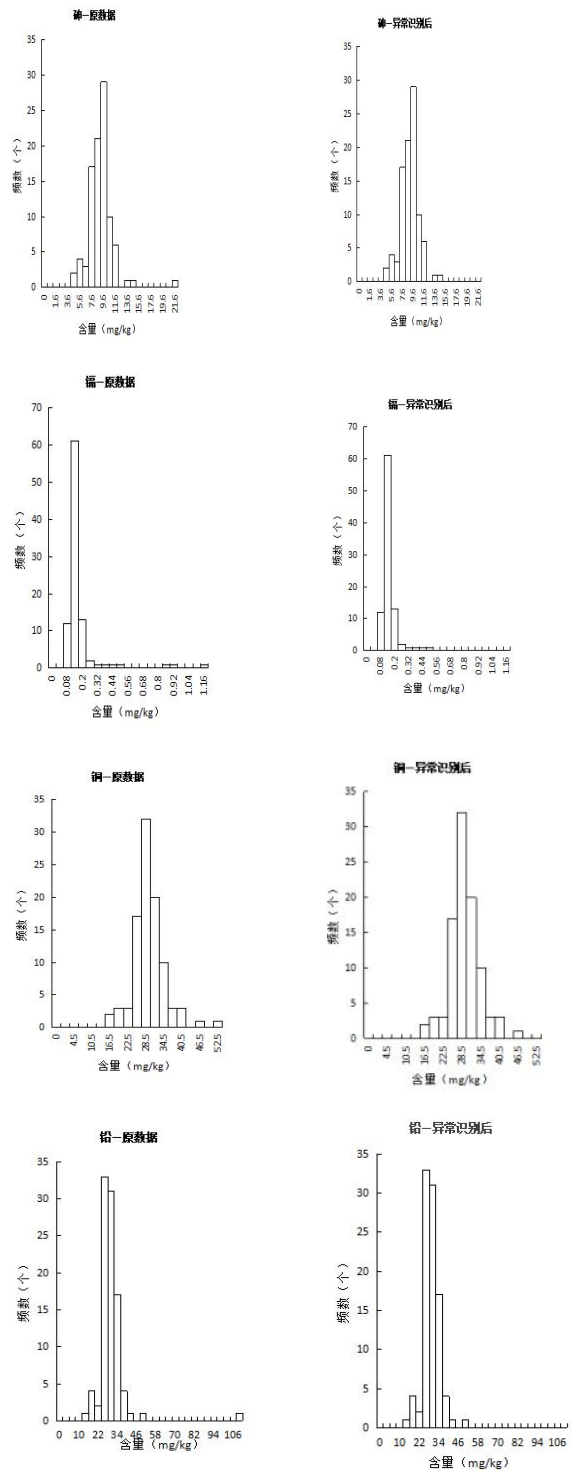
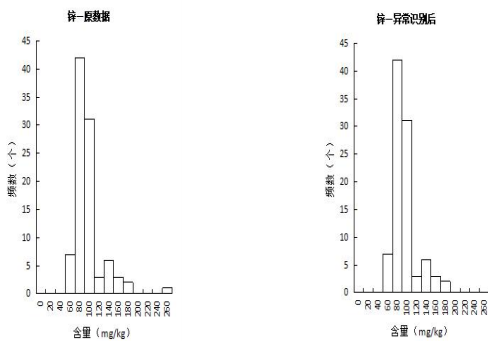


图 2 重金属异常值识别前后频率直方图

Fig.2 Frequency histogram of heavy metal outliers before and after identification

3 结论

(1) 案例研究结果表明, 采用拉依达检验法能够有效的识别出场地环境调查土壤重金属含量的异常值。将研究结果应用到实际工作中, 有助于识别低于筛选值但是受人类生产生活影响较大的土壤, 降低土壤累积污染的风险。

(2) 对识别出的异常值产生的原因进行分析, 可追本溯源包括调查污染源和迁移途径等, 结合分析结果针对偶发性输入和长期必然性输入分别采用不同的处理办法。

参考文献

- [1] GB 15618-2018 土壤环境质量 农用地土壤风险管控标准[S]. 生态环境部, 国家市场监督管理总局, 2018.
- [2] GB 36600-2018 土壤环境质量 建设用地土壤污染风险管控标准[S]. 生态环境部, 国家市场监督管理总局, 2018.
- [3] 孙述海, 李鹏飞. 疑似污染场地土壤环境调查方法研究[J]. 吉林地质, 2018, 37(04): 67-70.
- [4] 欧阳婕. 污染场地环境调查现状及存在问题探讨[J]. 环境与发展, 2018, 30(08): 31-32.
- [5] 中共中央国务院关于全面推进美丽中国建设的意见[N]. 人民日报, 2024-01-12 (20).
- [6] 余斐许, 窈孜, 叶彩红, 等. 城市化发展对林地土壤重金属砷含量的影响[J]. 林业与环境科学, 2022, 38(04): 1-6.
- [7] 张一鸣. 人类活动对沉积环境的影响研究综述[J]. 城市地质, 2022, 17(02): 184-192.
- [8] 彭锐, 吴坚, 谢亚莉, 等. 土壤污染风险评价及重金属积累情况研究——以长沙地区为例[J]. 环境生态学, 2023, 5(01): 35-38.
- [9] 张又文. 天津市不同土地利用类型土壤重金属的积累特征及污染评价[D]. 沈阳: 沈阳农业大学, 2020.
- [10] GB/T 4883-2008 正态样本离群值的判断和处理[S]. 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会, 2008.
- [11] 何少华. 试验设计与数据处理[M]. 长沙: 国防科技大学出版社, 2002.
- [12] 田嘉禹, 刘俐, 汪群慧, 等. 中美土壤元素背景值调查研究中数理统计方法运用及影响环境科学研究[J]. 环境科学研究, 2020, 33(03): 718-727.
- [13] 赵云丽. 利用 Python 实现计量检定中异常值快速剔除[J]. 工业计量, 2020, 30(6): 47-51.
- [14] 王景云. 土壤重金属污染异常值识别方法及其应用[D]. 山东: 山东科技大学, 2018.
- [15] 刘庚, 牛俊杰, 张朝, 等. 某铅酸蓄电池污染场地表层土壤重金属 Pb 空间分布预测研究[J]. 环境科学, 2014, 35(12): 4712-4719.
- [16] 何朝红. 回收率控制图在土壤半挥发性有机物替代物测定的应用[J]. 当代化工研究, 2021, 24(23): 65-67.
- [17] FU W, ZHAO KL, ZHANG CS, *et al.* Outlier identification of soil phosphorus and its implication for spatial structure modeling [J]. *Precis. Agric.*, 2016, 17(2): 121-135.
- [18] SUN XL, WU YJ, ZHANG CS, *et al.* Performance of median kriging with robust estimators of the variogram in outlier identification and spatial prediction for soil pollution at a field scale [J]. *Sci. Total Environ.*, 2019, 666: 902-914.
- [19] ZHANG CS, TANG Y, LUO L, *et al.* Outlier identification and visualization for Pb concentrations in urban soils and its implications for identification of potential contaminated land [J]. *Env. Pollution*, 2009, 157(11): 3083-3090.
- [20] 侍建国, 张亦飞. 拉依达准则在处理区域水文数据异常值中的应用[J]. 海河水利, 2016, 35(5): 49-51.
- [21] 王磊, 何江涛, 张振国, 等. 基于信息筛选和拉依达准则识别地下水主要组分水化学异常的方法研究[J]. 环境科学学报, 2018, 38(3): 919-929.
- [22] 全伟亮, 马慧民. 城市生活垃圾收集点垃圾产生量的概率分布研究——以上海市某区为例[J]. 环境卫生工程, 2018, 26(2): 59-63.
- [23] 周秘. 基于拉依达准则的基坑监测数据的粗大误差处理研究[J]. 华东六省一市地学科技论坛文集, 2023: 343-347.
- [24] 常锦才, 马义茹, 李菁月, 等. 基于灰色关联度分析法的合金成本最优化[J]. 华北理工大学学报(自然科学版), 2020, 42(02): 29-41.
- [25] 熊艳艳, 吴先球. 粗大误差四种判别准则的比较和应用[J]. 大学物理实验, 2010, 23(1): 66-68.

作者简介



董聪慧, 硕士, 工程师, 研究方向: 环境监测实验室分析、环境质量综合分析与评价和环境科学研究。

梅祖明, 高级工程师, 研究方向为环境污染防治、环境科学研究。