

区域农田土壤重金属污染源解析方法研究现状及展望

丁海元^{1*}, 高秉博^{2#}

¹对外经济贸易大学统计学院, 北京

²中国农业大学土地科学与技术学院, 北京

收稿日期: 2023年12月10日; 录用日期: 2024年1月15日; 发布日期: 2024年1月22日

摘要

重金属污染源精细解析是区域农田土壤开展污染消减与阻控措施的首要措施。农田土壤重金属来源复杂多样, 且污染物的迁移与累积过程受到地形、气象、植被等多种因素的影响, 导致污染源对区域不同位置农田产生差别化污染贡献。为精细解析该差别化污染贡献, 需使用多个类别型和连续型代理变量表征各污染源空间输出强度及影响因素, 并建模代理变量与农田土壤重金属含量之间的关系。而在区域尺度这一关系往往存在空间异质性, 现有方法能否有效解决该问题。对此, 本文在分析农田土壤重金属污染来源及其空间异质性污染贡献的基础上, 梳理了区域农田土壤重金属污染源解析方法的研究现状及其发展趋势, 并展望了新的源解析方法, 为我国各级政府和相关组织制定既经济又环保的农田土壤重金属污染风险防控措施提供技术支撑。

关键词

农田土壤, 重金属污染, 源解析方法, 空间异质性关系

Research Status and Prospects of Apportionment Methods for Heavy Metal Pollution Sources of Farmland Soil at Regional Scale

Haiyuan Ding^{1*}, Bingbo Gao^{2#}

¹School of Statistics, University of International Business and Economics, Beijing

²College of Land Science and Technology, China Agricultural University, Beijing

*第一作者: 丁海元, 对外经济贸易大学统计学院在职人员高级课程研修班学员。

#通讯作者: 高秉博, 中国农业大学土地与科学技术学院副教授。

Abstract

Fine apportionment of the heavy metal pollution sources is the prerequisite for pollution reduction and prevention for farmland soil. The sources of heavy metals in farmland soil are complex and diverse, and the migration and accumulation process of pollutants are affected by multiple factors such as topography, meteorology, vegetation, etc., resulting in different pollution contributions from pollution sources to farmland soil in different places in the region. In order to analyze the different pollution contribution accurately, it is necessary to use multiple categorical and continuous proxy variables to characterize the emission intensity of each pollution source and corresponding influencing factors, and to model the relationship between proxy variables and heavy metal content in farmland soil. The relationships between the heavy metal contents and proxy variables are spatially heterogeneous, thus whether existing methods can solve this problem deserves investigation. To answer this question, we first analyze the sources of heavy metal pollution in farmland soil and their spatially heterogeneous pollution contribution, and then summarize the research status and prospects of apportionment methods of heavy metal pollution sources of farmland soil at a regional scale. This study aims to provide technical support for the government and related organizations to take measures to prevent and control heavy metal pollution risks in farmland soil that are both economical and environmentally friendly.

Keywords

Farmland Soil, Heavy Metal Pollution, Source Apportionment Method, Spatial Heterogeneous Relationship

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

农田土壤是农业生产的重要基础。然而,我国农田土壤重金属污染形势严峻[1],由于土壤重金属具有隐蔽性、滞后性、累积性以及不可逆性等污染特点,农田土壤一旦被重金属污染便难以修复,会长期影响农产品数量和质量安全等,最终危害人类健康和社会经济发展。对此,国务院于2016年印发《土壤污染防治行动计划》,制定了全国土壤污染防治工作的行动纲领,重点部署了截止2050年的土壤污染防治工作任务和目标,要求严控新增污染。而农田土壤重金属风险防控的关键在于源头阻控[2]。因此,为了制定既经济又环保的污染消减与阻控措施,首先需要精细解析污染源及其对区域不同位置农田的差别化污染贡献。

农田土壤重金属源于多种自然源与人为源的叠加,同时各类污染物的迁移与累积受到地形、气象、植被等多种因素的影响。为了精细解析污染源对区域不同位置农田的差别化污染贡献,需要使用多个类别型和连续型代理变量(本文称为“多维混合类型代理变量”)表征各污染源空间输出强度及影响因素,并建模代理变量与农田土壤重金属含量之间的关系。由于污染源的空间位置和尺度的差异,以及污染物迁移与累积特征的空间相关性和异质性,导致这种关系在局部地理空间范围内存在相关性,但是在区域不同部分间经常存在显著差异[3][4],即存在空间异质性关系。定量解析这种空间异质性关系不仅能够获得

污染源对区域不同位置农田的差别化污染贡献, 还能提高对区域总体污染贡献解析的准确性, 而忽略这种关系会导致全局建模的失败[5]。

通常, 土壤重金属污染源解析方法按照分析对象可分为污染源分析、受体分析、污染源与受体联合分析三类, 能够对污染来源进行定性或定量研究。但上述被广泛应用的方法无法有效处理空间异质性关系。对此, 污染源与土壤重金属之间的空间位置关系已被越来越多的学者认识到其对污染源解析的重要意义。

空间异质性关系建模是空间统计领域的基本问题。目前针对该问题的研究已经取得了部分进展, 如地理加权回归与空间多水平建模。但是这些方法无法有效处理多维混合类型代理变量。而机器学习领域的基于结构化算子的基函数扩展模型(如提升树、SVM、神经网络等)在处理多维混合类型代理变量方面具有显著的优势, 但不能建模空间异质性关系。因此, 需要结合两种方法的优势, 研发一种能够处理多维混合类型代理变量的空间异质性关系模型, 用于精细解析具体污染源对区域不同位置农田的差别化污染贡献。

本文首先分析了农田土壤重金属污染来源及其空间异质性污染贡献, 在此基础上梳理了区域农田土壤重金属污染源解析方法的研究现状, 并探讨了其发展趋势, 最后展望了新的源解析方法。

2. 农田土壤重金属污染来源及其空间异质性污染贡献

土壤中重金属元素的来源分为自然源和人为源两大类[6] [7] [8] [9] [10]。自然源包括成土母质、岩石风化、火山喷发等, 它们在气候、生物、地形等因素作用下, 形成了土壤重金属背景值[11] [12]。人为源主要包括工业生产、农业生产、交通运输和居民生活等, 通过大气沉降、灌溉与径流、固废与堆肥向农田输入重金属, 其输入和累积过程受到气象、地形和植被等因素的影响[13] [14] [15]。污染源的尺度差异及影响因素的空间相关性和分异性, 导致污染源对土壤的污染贡献同时呈现空间相关性和异质性。对于自然源贡献, 除了全国尺度上的地带性分布与东西共轭关系外, 在市县以及更小的尺度上, 由于局部地形和生物等因素的影响, 形成了垂向分异性规律和斑块式的空间异质性分布格局[16]。人为源中, 点源污染的污染物以企业或生产场地为中心沿风向和水向扩散, 污染贡献在空间上呈现非线性渐变趋势[4], 而非点源污染则存在明显的斑块特征(如污灌、农药化肥等)或条带化分布(如交通) [17] [18]。因此, 农田土壤重金属污染源多样且影响因素复杂, 且空间尺度差异较大, 污染源对区域农田土壤重金属的污染贡献既存在局部的空间相关性又存在全局上的空间异质性[19] [20]。

传统基于产排污系数的污染源强计算方法只能得到排放总量, 无法反映污染源对区域不同位置的污染输出量。因此一般使用代理变量反映其输出强度的空间分布, 如采用土壤母质、土壤类型代理自然源, 使用与污染源间距离的倒数反映点源污染强度, 使用污灌分布、土地利用类型、种植模式、距离道路距离或道路密度等反映非点源污染强度[17] [21] [22]。在受体分析方法中, 还会使用主成分得分表征对应类型污染源的强度[23]。虽然代理变量能够反映污染源输出的空间差异, 但它并不是直接的污染输出量, 还需要结合污染迁移与累积影响因素, 并与土壤重金属污染含量进行建模校正, 才能定量计算各自的污染贡献[24]。而在区域尺度上, 代理变量与真实污染输出分布之间难以存在全局一致的线性比例关系, 而且未知历史污染源的贡献以及未能被考虑的影响因素均存在空间异质性, 导致各类代理变量与土壤重金属含量之间的关系在区域不同部分存在差异, 即存在空间异质性关系(统计学上称为空间非平稳关系或随空间变化的关系) [5] [24]。

因此, 各类污染源的污染贡献存在局部的空间相关性与全局的空间异质性, 且空间变异尺度差别较大。可以使用多维混合类型代理变量表征污染源在区域不同空间位置的污染输出量及其影响因素变化, 但是代理变量与土壤重金属含量之间存在空间异质性关系。在区域农田土壤重金属污染源解析模型构建

中, 需要妥善处理这种空间异质性关系[25]。

3. 农田土壤重金属污染源解析方法及其发展趋势

污染源解析是对污染源进行定性或定量研究的方法。土壤重金属源解析方法按照分析对象可分为污染源分析、受体分析、污染源与受体联合分析三类(见表 1)。

Table 1. Apportionment methods for heavy metal pollution sources of farmland soil

表 1. 农田土壤重金属污染源解析方法

| 方法类别 | 具体方法 | 定性/定量 | 是否精确匹配污染源 | 解析空间差别化贡献 | 是否考虑空间异质性关系 |
|------------------|-------------|--------------------------------------|-----------|-----------|-------------|
| 污染源分析 | 源排放清单法 | 定量 | 是 | 不能 | 否 |
| 受体分析 (传统多元统计) | 降维方法 | RA、PCA、CA | 定性 | 否 | 否 |
| | 分解方法 | PMF、UNMIX | 定量 | 否 | 样点处 |
| | 组合方法 | PCA-MLR、APCA-MLR | 定量 | 否 | 样点处 |
| | 化学质量平衡 | 元素比值法、同位素比值法、重金属形态分析 | 定量 | 是 | 样点处 |
| 污染源与受体联合分析 | 空间模式匹配 | Kriging、SGCS、MFK | 定性 | 是 | 不能 |
| | 空间模式匹配+多元统计 | SGCS+PMF、Kriging+PCA+CA、Kriging+PMF、 | 定量或定性 | 是 | 样点处 |
| | 空间回归 | MLRD、GWR | 定量 | 是 | 能 |
| | 机器学习 | CIT、Cubist、SGBT、RF | 定量 | 是 | 不能 |

污染源分析方法主要是源排放清单法。该方法在对各类污染源状况调查和统计的基础上, 依据污染源的活动水平与排放因子模型估算污染源排放量[26]。详细的污染源清单建立难度较大, 且污染源对区域土壤的污染贡献存在空间异质性, 不同来源的重金属在土壤中的累积能力也有差异, 因此该方法不能估算污染源对区域不同位置农田的差别化污染贡献, 难以用于指导精准防控。

受体分析法以污染受体(土壤)作为主要研究对象, 目前以传统多元统计方法为主。主要包括相关性分析(Correlation Analysis, CoA)、主成分分析(Principal Component Analysis, PCA)、因子分析(Factor Analysis, FA)、聚类分析(Cluster Analysis, CA)等降维方法, PMF 和 UNMIX 等分解方法, 以及主成分分析-多元线性回归(Principal Component Analysis-Multiple Linear Regression, PCA-MLR)、绝对主成分分析-多元线性回归(Absolute Principal Component Analysis-Multiple Linear Regression, APCA-MLR)等组合方法[27][28][29]。该类方法基于土壤中各类重金属之间的线性关系对土壤重金属进行归类, 然后依据先验知识判断每一类土壤重金属可能的污染源, 不能精确匹配具体的污染源。分解方法和组合方法虽然能给出每类污染源在样点处的贡献, 但是在解析过程中采用全局线性模型, 没有考虑污染源与土壤重金属之间的空间异质性关系。

污染源与受体联合分析法通过建立具体污染源与土壤重金属含量之间的关系模型, 定量解析各污染源对土壤重金属污染的贡献, 主要包括化学质量平衡法、空间模式匹配法、空间回归法和机器学习法。化学质量平衡法通过对应污染源排放成分谱与受体中各污染物的浓度, 建立线性比例关系解析污染源贡献, 如元素比值法、同位素比值法和形态分析法[26]。该方法能够定量解析具体污染源对样点处的局部贡献, 但是假设各污染源排放物的化学组成稳定且存在明显差异, 不适用于复杂污染源的情况, 同时还需要较为完整的污染源指纹图谱, 代价较高。空间模式匹配法则只需要污染源位置、类型等简单信息, 通过匹配由克里金(Kriging)、序贯高斯模拟(Sequential Gaussian Simulation, SGCS)等空间插值方法获得的土

壤重金属空间分布模式, 或匹配由多变量因子克里金(Multivariate Factorial Kriging, MFK)得出的各尺度因子空间分布特征, 定性识别各尺度污染源[30] [31]。

近些年, 越来越多的学者认识到污染源与土壤重金属之间的空间位置关系对污染源解析的重要意义, 除了将空间模式匹配方法与传统多元统计方法相结合, 定性或定量解析污染源[32]-[37], 还采用距离线性回归模型(Multiple Linear Regression with Distance Model, MLRD)和地理加权回归模型(Geographically Weighted Regression, GWR)定量解析污染源与土壤重金属含量之间的空间异质性关系[11] [25] [38]。但是MLRD 只能以距离倒数的方式建模点源污染贡献, 忽略了其他类型源的污染贡献, 容易过高估计点源的污染贡献。GWR 能够定量建模污染源与土壤重金属含量之间的空间异质性关系, 但是难以处理表征众多污染源及影响因素的多维混合类型代理变量。机器学习方法在多维混合类型变量处理方面具有明显的优势, 逐渐被引入土壤重金属污染源解析领域, 如条件推断树(conditional inference tree, CIT) [39]、Cubist模型[40]、随机梯度提升树(Stochastic Gradient Boosting Tree, SGBT) [24]、随机森林(Random Forest, RF) [41]。但是这些机器学习方法不能处理空间异质性关系, 且不能解析空间差别化污染贡献。最常用的RF在处理多维混合类型变量时得出的全局重要性严重有偏, 会误导对污染贡献的判断[42] [43]。

因此, 传统的污染源分析方法、受体分析方法以及化学质量平衡法, 不能匹配具体的污染源或者解析其对区域不同位置农田的差别化污染贡献。利用农田污染大数据以及污染源和受体之间的空间位置关系进行源解析, 已成为新的发展趋势。但目前采用的空间模式匹配法、空间回归法和机器学习法, 由于不能建模多维混合类型代理变量与土壤重金属含量间的空间异质性关系, 难以提高解析精度。

4. 新方法展望

地理学第一定律和第二定律不仅阐述了单个目标变量在地理空间上的相关性与异质性, 还包括变量之间关系在地理空间上的相关性与异质性[44] [45]。除依照先验知识分区外, 监督学习模型是目前最主要的空间异质性关系建模工具。它将土壤重金属污染源及影响因素代理变量作为预测变量, 将土壤重金属含量作为响应变量, 通过对样点数据的学习构建预测变量与响应变量之间的关系模型, 解析土壤重金属污染源贡献[24] [46]。目前, 主要的空间异质性关系监督学习模型都是从非空间的监督学习模型经过空间化改进而来, 按照模型的构建方式可以分为空间全局模型、空间多水平模型和空间局部模型三类(如表2)。

Table 2. Supervised learning models for spatial heterogeneous relationships

表 2. 空间异质性关系监督学习模型

| 类型 | 空间化方式 | 具体模型 | 先验知识要求 | 多维混合变量处理能力 | 解析的空间异质性关系 |
|---------|---------------|-----------------|------------------|------------|------------|
| 空间全局模型 | 直接加入地理空间坐标变量 | 多元线性回归 | 无 | 弱 | 不能解析 |
| | | 广义相加模型 | 模型形式 | 中 | 不能解析 |
| | | 基于结构化算子的基函数扩展模型 | 无 | 强 | 不能解析 |
| | 地理空间坐标扩展 | 空间扩展模型 | 关系的全局变化模式 | 中 | 预设的全局变化模式 |
| 空间多水平模型 | 属性水平转化为地理空间水平 | 随机效应模型、随机系数模型 | 划分地理空间水平, 同水平内分区 | 弱 | 多水平异质性关系 |
| 空间局部模型 | 属性空间转化为地理空间 | 贝叶斯空间变系数模型 | 先验概率 | 弱 | 全局异质但局部相关 |
| | | GWR、GWRR、GWL | 无 | 弱 | 全局异质但局部相关 |

空间全局模型是从非空间全局线性模型及其扩展模型中通过两种空间化方式改进而来。第一种方式直接将地理空间坐标作为独立的预测变量加入非空间模型, 而不改变方法本身。常见模型包括多元线性回归、广义相加模型以及基于结构化算子的基函数扩展模型。其中基于结构化算子的基函数扩展模型, 如梯度提升树(Gradient Boosting Decision Tree, GDBT)、支持向量机(Support Vector Machine, SVM), 采用结构化基函数(或希尔伯特空间的核函数)并使用正则化处理方法, 能够快速有效地处理多维混合类型变量并防止过拟合[46]。但是这种方法只能获得响应变量与预测变量之间的全局关系及响应变量的空间依赖, 不能给出除坐标外其他预测变量与响应变量之间的空间局部关系。其中 GDBT 在提升学习过程中, 虽然会重点拟合局部样点, 但其损失函数依然是全局的, 是在全局损失函数控制下的局部重点拟合, 因此同样不能得出预测变量与响应变量之间的空间局部关系[47]。第二种方式将基函数扩展模型中的参数扩展为空间坐标的函数, 得到空间扩展模型(Spatial Expansion Method, SEM) [48]。该类模型能够解析各预测变量与响应变量之间的空间异质性关系, 但是所解析的非平稳关系只能是某种预先设定的变化模式, 而且这种变化模式必须是全局稳定的, 因此在实际中难以应用。

空间多水平模型通过将非空间多水平模型的属性水平转化为地理空间水平, 将水平内的属性分组转化为空间分区, 并基于不同空间水平与分区的方差分解将空间异质性关系分解为多层线性关系的组合[49][50], 主要包括随机效应模型、随机系数模型。它们能够考虑预测变量的空间尺度差异, 建模多水平异质性关系。但是该类方法需要预先确定关系的空间水平和同水平内关系差异的空间分区, 而关系差异的空间分区本身就是一个空间异质性关系探索的过程[51]。对于政治经济现象, 由于其存在行政区内的一致性和行政区之间的差异性, 可以按照行政区划进行分区。但是对于土壤重金属污染这种自然与人为混合作用的现象, 难以在建模前完成空间关系差异的分区。同时, 这类方法难以处理土壤重金属污染源解析所涉及的多维混合类型代理变量[52]。

空间局部模型通过将非空间局部线性模型中的多维属性空间邻近性转化为二维(或三维)地理空间邻近性, 避免了由于变量数目增加而带来的“维度灾难”, 并利用空间局部相关性关系解析全局空间异质性关系, 如贝叶斯空间变系数模型(Bayesian Spatially Varying Coefficient, BSVC) [53][54]和 GWR 模型[48]。BSVC 基于地理空间邻近关系所产生的自相关性构建空间局部条件概率分布函数, 获得地理空间局部位置上预测变量系数的概率分布。但是该方法需要预先设定各参数的先验分布, 且难以处理多维混合类型代理变量。GWR 基于地理空间距离核函数与加权最小二乘法, 通过为每个空间位置建立局部线性模型, 进行地理空间异质性关系的解析[55]。但是由于它在局部模型构建时采用传统的加权最小二乘法, 容易产生共线性问题; 同时在处理类别型变量时, 会引入过多的哑变量, 加剧共线性问题, 严重影响模型的准确性与稳定性[56]。为了缓解 GWR 的共线性问题, Wheeler [57][58]结合多元线性回归的正则化技术, 分别提出了地理加权岭回归(Geographically Weighted ridge Regression, GWRR)与地理加权 Lasso 回归(Geographically Weighted Lasso, GWL)。它们能够在一定程度上解决 GWR 的共线性问题, 但是由于采用线性基函数, 会增加拟合参数进而增大模型方差, 而且难以处理类别型变量, 依然不能较好地处理农田土壤重金属污染源解析中的多维混合类型代理变量。另外, 空间局部模型通过在区域内每个点上单独构建局部相关性模型的方式分解区域的空间异质性关系, 不仅带来大量的冗余计算, 而且无法处理不同污染源及影响因素的空间变异尺度差别较大的问题。

总之, 现有的空间异质性关系监督学习模型, 虽然在处理多维混合类型变量和空间异质性关系方面各有优势, 但是均难以同时解决。可结合地理空间局部加权与多水平分解, 将全局异质性关系转化为能适配污染源及影响因素空间变异尺度的局部相关性关系, 同时结合基于结构化算子的基函数扩展模型, 增强多维混合变量处理能力, 研发能够精细解析污染源对区域不同位置农田差别化污染贡献的模型工具。

5. 结论

农田土壤重金属污染源多样且影响因素复杂, 污染源对区域不同位置农田土壤重金属存在空间异质性污染贡献。作为区域农田土壤开展污染消减与阻控的首要措施, 精细解析重金属污染源需使用多维混合类型代理变量, 并建模代理变量与农田土壤重金属含量之间的空间异质性关系。当前常用的区域农田土壤重金属污染源解析方法无法有效处理这一空间异质性关系, 因此, 利用农田污染大数据以及污染源和受体之间的空间位置关系进行源解析, 已成为新的发展趋势。为实现污染源对区域不同位置农田差别化污染贡献的精细解析, 可结合地理空间局部加权与多水平分解, 将全局异质性关系转化为能适配污染源及影响因素空间变异尺度的局部相关性关系, 同时结合基于结构化算子的基函数扩展模型, 研发新的源解析方法工具, 为我国各级政府和相关部门制定既经济又环保的农田土壤重金属污染风险防控措施提供技术支撑。

基金项目

国家自然科学基金项目(42271428)。

参考文献

- [1] 环境保护部, 国土资源部. 全国土壤污染调查公报[J]. 中国环保产业, 2014(5): 2.
- [2] 曾希柏, 徐建明, 黄巧云, 等. 中国农田重金属问题的若干思考[J]. 土壤学报, 2013, 50(1): 186-194.
- [3] 瞿明凯, 李卫东, 张传荣, 等. 地理加权回归及其在土壤和环境科学上的应用前景[J]. 土壤, 2014, 46(1): 15-22.
- [4] 李鹏. 县域尺度耕地土壤重金属的时空变异特征及污染源识别研究[D]: [硕士学位论文]. 杭州: 浙江大学, 2016.
- [5] Fernández, S., Cotos-Yáñez, T., Roca-Pardiñas, J. and Ordóñez, C. (2018) Geographically Weighted Principal Components Analysis to Assess Diffuse Pollution Sources of Soil Heavy Metal: Application to Rough Mountain Areas in Northwest Spain. *Geoderma*, **311**, 120-129. <https://doi.org/10.1016/j.geoderma.2016.10.012>
- [6] Sawut, R. (2017) Using Regression Model to Identify and Evaluate Heavy Metal Pollution Sources in an Open Pit Coal Mine Area, Eastern Junggar, China. *Environmental Earth Sciences*, **73**, Article No. 822. <https://doi.org/10.1007/s12665-017-7035-5>
- [7] 魏复盛, 陈静生, 吴燕玉, 等. 中国土壤环境背景值研究[J]. 环境科学, 1991, 12(4): 12-19.
- [8] 齐杏杏, 高秉博, 潘瑜春, 等. 基于地理探测器的土壤重金属污染影响因素分析[J]. 农业环境科学学报, 2019, 38(11): 2476-2486.
- [9] 蔡立梅, 马瑾, 周永章, 等. 东莞市农田土壤和蔬菜重金属的含量特征分析[J]. 地理学报, 2008, 63(9): 994-1003.
- [10] Liu, J., Liu, Y.J., Liu, Y., Liu, Z. and Zhang, A.N. (2018) Quantitative Contributions of the Major Sources of Heavy Metals in Soils to Ecosystem and Human Health Risks: A Case Study of Yulin, China. *Ecotoxicology and Environmental Safety*, **164**, 261-269. <https://doi.org/10.1016/j.ecoenv.2018.08.030>
- [11] 黄颖. 不同尺度农田土壤重金属污染源解析研究[D]: [博士学位论文]. 杭州: 浙江大学, 2018.
- [12] 何纪力, 徐光炎, 朱惠民. 土壤环境背景值的影响因素综合分析[J]. 中国环境监测, 1992, 8(3): 69-73.
- [13] 赵彦锋, 郭恒亮, 孙志英, 等. 基于土壤学知识的主成分分析判断土壤重金属来源[J]. 地理科学, 2008, 28(1): 45-50.
- [14] 谢小进, 康建成, 李卫江, 等. 上海宝山区农用土壤重金属分布与来源分析[J]. 环境科学, 2010, 31(3): 768-774.
- [15] 倪润祥. 中国农田土壤重金属输入输出平衡和风险评价研究[D]: [博士学位论文]. 北京: 中国农业科学院, 2017.
- [16] 樊文华, 张毓庄, 郭新波, 等. 五台山土壤环境背景值及其垂直分异规律[J]. 山西农业大学学报, 1995(2): 142-146, 212-213.
- [17] Ha, H., Olson, J.R., Bian, L. and Rogerson, P.A. (2014) Analysis of Heavy Metal Sources in Soil Using Kriging Interpolation on Principal Components. *Environmental Science & Technology*, **48**, 4999-5007. <https://doi.org/10.1021/es405083f>
- [18] 吕悦风, 孙华. 浙北某县域耕地土壤重金属空间分异特征、污染评价及来源分析[J]. 农业环境科学学报, 2019, 38(1): 95-102.

- [19] Marchant, B.P., Tye, A.M. and Rawlins, B.G. (2011) The Assessment of Point-Source and Diffuse Soil Metal Pollution Using Robust Geostatistical Methods: A Case Study in Swansea (Wales, UK). *European Journal of Soil Science*, **62**, 346-358. <https://doi.org/10.1111/j.1365-2389.2011.01373.x>
- [20] 胡克林, 李保国, 吕贻忠, 等. 非平稳型区域土壤汞含量的各种估值方法比较[J]. 环境科学, 2004, 25(3): 132-137.
- [21] Hou, D., O'Connor, D., Nathanail, P., et al. (2017) Integrated GIS and Multivariate Statistical Analysis for Regional Scale Assessment of Heavy Metal Soil Contamination: A Critical Review. *Environmental Pollution*, **231**, 1188-1200. <https://doi.org/10.1016/j.envpol.2017.07.021>
- [22] Lin, Y.P., Teng, T.P. and Chang, T.K. (2002) Multivariate Analysis of Soil Heavy Metal Pollution and Landscape Pattern in Changhua County in Taiwan. *Landscape and Urban Planning*, **62**, 19-35. [https://doi.org/10.1016/S0169-2046\(02\)00094-4](https://doi.org/10.1016/S0169-2046(02)00094-4)
- [23] Lv, J. (2019) Multivariate Receptor Models and Robust Geostatistics to Estimate Source Apportionment of Heavy Metals in Soils. *Environmental Pollution*, **244**, 72-83. <https://doi.org/10.1016/j.envpol.2018.09.147>
- [24] Wang, Q., Xie, Z. and Li, F. (2015) Using Ensemble Models to Identify and Apportion Heavy Metal Pollution Sources in Agricultural Soils on a Local Scale. *Environmental Pollution*, **206**, 227-235. <https://doi.org/10.1016/j.envpol.2015.06.040>
- [25] Qu, M., Wang, Y., Huang, B. and Zhao, Y.C. (2018) Source Apportionment of Soil Heavy Metals Using Robust Absolute Principal Component Scores-Robust Geographically Weighted Regression (RAPCS-RGWR) Receptor Model. *Science of the Total Environment*, **626**, 203-210. <https://doi.org/10.1016/j.scitotenv.2018.01.070>
- [26] 陈雅丽, 翁莉萍, 马杰, 等. 近十年中国土壤重金属污染源解析研究进展[J]. 农业环境科学学报, 2019, 38(10): 2219-2238.
- [27] 林燕萍, 赵阳, 胡恭任, 等. 多元统计在土壤重金属污染源解析中的应用[J]. 地球与环境, 2011, 39(4): 536-542.
- [28] 李娇, 吴劲, 蒋进元, 等. 近十年土壤污染源解析研究综述[J]. 土壤通报, 2018, 49(1): 232-242.
- [29] Song, H., Hu, K., An, Y., Chen, C. and Li, G.D. (2018) Spatial Distribution and Source Apportionment of the Heavy Metals in the Agricultural Soil in a Regional Scale. *Journal of Soils and Sediments*, **18**, 852-862. <https://doi.org/10.1007/s11368-017-1795-0>
- [30] Qu, M.K., Li, W.D., Zhang, C.R., et al. (2013) Source Apportionment of Heavy Metals in Soils Using Multivariate Statistics and Geostatistics. *Pedosphere*, **23**, 437-444. [https://doi.org/10.1016/S1002-0160\(13\)60036-3](https://doi.org/10.1016/S1002-0160(13)60036-3)
- [31] Xu, X., Zhao, Y., Zhao, X., et al. (2014) Sources of Heavy Metal Pollution in Agricultural Soils of a Rapidly Industrializing Area in the Yangtze Delta of China. *Ecotoxicology and Environmental Safety*, **108**, 161-167. <https://doi.org/10.1016/j.ecoenv.2014.07.001>
- [32] Mahmoudabadi, E., Sarmadian, F. and Nazary Moghaddam, R. (2015) Spatial Distribution of Soil Heavy Metals in Different Land Uses of an Industrial Area of Tehran (Iran). *International Journal of Environmental Science and Technology*, **12**, 3283-3298. <https://doi.org/10.1007/s13762-015-0808-z>
- [33] Qu, M., Li, W., Zhang, C., et al. (2016) Assessing the Pollution Risk of Soil Chromium Based on Loading Capacity of Paddy Soil at a Regional Scale. *Scientific Reports*, **5**, Article No. 18451. <https://doi.org/10.1038/srep18451>
- [34] 陈丹青, 谢志宜, 张雅静, 等. 基于 PCA/APCS 和地统计学的广州市土壤重金属来源解析[J]. 生态环境学报, 2016, 25(6): 1014-1022.
- [35] 陈秀端, 卢新卫. 基于受体模型与地统计的城市居民区土壤重金属污染源解析[J]. 环境科学, 2017, 38(6): 2513-2521.
- [36] 瞿明凯, 李卫东, 张传荣, 等. 基于受体模型和地统计学相结合的土壤镉污染源解析[J]. 中国环境科学, 2013, 33(5): 854-860.
- [37] Rodríguez, J.A., Nanos, N., Grau, J.M., et al. (2008) Multiscale Analysis of Heavy Metal Contents in Spanish Agricultural Topsoils. *Chemosphere*, **70**, 1085-1096. <https://doi.org/10.1016/j.chemosphere.2007.07.056>
- [38] Harris, P., Clarke, A., Juggins, S., et al. (2015) Enhancements to a Geographically Weighted Principal Component Analysis in the Context of an Application to an Environmental Data Set: Enhancements to GWPCA. *Geographical Analysis*, **47**, 146-172. <https://doi.org/10.1111/gean.12048>
- [39] Hu, Y. and Cheng, H. (2016) A Method for Apportionment of Natural and Anthropogenic Contributions to Heavy Metal Loadings in the Surface Soils across Large-Scale Regions. *Environmental Pollution*, **214**, 400-409. <https://doi.org/10.1016/j.envpol.2016.04.028>
- [40] 孙慧, 毕如田, 郭颖, 等. 广东省土壤重金属溯源及污染源解析[J]. 环境科学学报, 2018, 38(2): 704-714.
- [41] 刘斌, 郭星, 朱宇恩. 基于随机森林模型的土壤重金属源解析——以晋中盆地为例[J]. 干旱区资源与环境, 2019, 33(1): 106-111.

- [42] Shih, Y.S. and Tsai, H.W. (2004) Variable Selection Bias in Regression Trees with Constant Fits. *Computational Statistics & Data Analysis*, **45**, 595-607. [https://doi.org/10.1016/S0167-9473\(03\)00036-7](https://doi.org/10.1016/S0167-9473(03)00036-7)
- [43] Strobl, C., Boulesteix, A.L., Zeileis, A. and Hothorn T., (2007) Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution. *BMC Bioinformatics*, **8**, Article No. 25. <https://doi.org/10.1186/1471-2105-8-25>
- [44] Goodchild, M.F. (2004) The Validity and Usefulness of Laws in Geographic Information Science and Geography. *Annals of the Association of American Geographers*, **94**, 300-303. <https://doi.org/10.1111/j.1467-8306.2004.09402008.x>
- [45] Li, S., Zhao, Z., Xie, M.M. and Wang, Y.L. (2010) Investigating Spatial Non-Stationary and Scale-Dependent Relationships between Urban Surface Temperature and Environmental Factors Using Geographically Weighted Regression. *Environmental Modelling & Software*, **25**, 1789-1800. <https://doi.org/10.1016/j.envsoft.2010.06.011>
- [46] Hastie, T., Tibshirani, R. and Friedman, J. (2010) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *Biometrics*, **66**, 1315. <https://doi.org/10.1111/j.1541-0420.2010.01516.x>
- [47] Friedman, J.H. (2002) Stochastic Gradient Boosting. *Computational Statistics & Data Analysis*, **38**, 367-378. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- [48] Fotheringham, A., Brunson, C. and Charlton, M. (2002) Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. John Wiley & Sons, New York.
- [49] Jones, K. (1991) Specifying and Estimating Multi-Level Models for Geographical Research. *Transactions of the Institute of British Geographers*, **16**, 148-159. <https://doi.org/10.2307/622610>
- [50] Duncan, C. and Jones, K. (2000) Using Multilevel Models to Model Heterogeneity: Potential and Pitfalls. *Geographical Analysis*, **32**, 279-305. <https://doi.org/10.1111/j.1538-4632.2000.tb00429.x>
- [51] Lupi, L., Bertrand, L., Monferrán, M.V., et al. (2019) Multilevel and Structural Equation Modeling Approach to Identify Spatiotemporal Patterns and Source Characterization of Metals and Metalloids in Surface Water and Sediment of the Ctlamochita River in Pampa Region, Argentina. *Journal of Hydrology*, **572**, 403-413. <https://doi.org/10.1016/j.jhydrol.2019.03.019>
- [52] Gelman, A. (2006) Multilevel (hierarchical) Modeling: What It Can and Can't Do. *Technometrics*, **48**, 432-435. <https://doi.org/10.1198/004017005000000661>
- [53] Wheeler, D.C. and Calder, C.A. (2007) An Assessment of Coefficient Accuracy in Linear Regression Models with Spatially Varying Coefficients. *Journal of Geographical Systems*, **9**, 145-166. <https://doi.org/10.1007/s10109-006-0040-y>
- [54] Sharkey, P. and Winter, H.C. (2019) A Bayesian Spatial Hierarchical Model for Extreme Precipitation in Great Britain: A Bayesian Hierarchical Model for Extreme Precipitation. *Environmetrics*, **30**, e2529. <https://doi.org/10.1002/env.2529>
- [55] Li, X., Zhou, Y., Asrar, G.R. and Zhu, Z.Y. (2018) Creating a Seamless 1km Resolution Daily Land Surface Temperature Dataset for Urban and Surrounding Areas in the Conterminous United States. *Remote Sensing of Environment*, **206**, 84-97. <https://doi.org/10.1016/j.rse.2017.12.010>
- [56] Wheeler, D.C. (2014) Geographically Weighted Regression. In: Fischer, M. and Nijkamp, P., Eds., *Handbook of Regional Science*, Springer, Berlin, 1435-1459. https://doi.org/10.1007/978-3-642-23430-9_77
- [57] Wheeler, D.C. (2009) Simultaneous Coefficient Penalization and Model Selection in Geographically Weighted Regression: The Geographically Weighted Lasso. *Environment and Planning A: Economy and Space*, **41**, 722-742. <https://doi.org/10.1068/a40256>
- [58] Wheeler, D.C. (2007) Diagnostic Tools and a Remedial Method for Collinearity in Geographically Weighted Regression. *Environment and Planning A: Economy and Space*, **39**, 2464-2481. <https://doi.org/10.1068/a38325>