邹心颖,闫庆武,吴子豪,等.基于 GeoShapley 算法的长春市城郊土壤砷含量局部主控因子探究 [J]. 中国环境科学, 2026,46(1):289-300.
Zou X Y , Yan Q W, Wu Z H, et al. Exploring key local influencing factors of arsenic content in suburban soils of Changchun city using the GeoShapley approach [J]. China Environmental Science, 2026,46(1):289-300.

基于 GeoShapley 算法的长春市城郊土壤砷含量局部主控因子探究

邹心颖[1],闫庆武[1],吴子豪[1*],祝元丽[1],陈奕云[2] (1.中国矿业大学公共管理学院,江苏 徐州 221116；2.武汉大学资源与环境科学学院,湖北 武汉 430079)

摘要：基于长春市城郊 200 个土壤砷样本,结合极限梯度提升树与 GeoShapley 算法,探究砷与环境因子间的异质性、非线性关系,并识别影响土壤砷含量的局部主控因子及来源.结果表明:①极限提升树模型优于地理加权回归和多尺度地理加权回归模型,其拟合和预测 R^2 分别为 0.958 和 0.729;②GeoShapley 的结果显示,锰含量的平均贡献率居首位,是影响土壤砷最显著的特征变量,而空间位置的特征重要性排名第二,证实了考虑空间坐标的必要性;③结合影响因子的 Geoshapley 空间分布图集发现,锰含量、碱解氮和硫含量是 99%样本点砷含量的首要影响特征,表明母质和农业生产是砷的主要来源;④非线性关系图显示砷含量与各环境变量之间均存在阈值效应,特别是当锰含量高于 2g/kg、碱解氮高于 175mg/kg、硫含量低于 20mg/kg 时,砷含量达到峰值,在污染防控时应予以特别关注.本文证实了 Geoshapley 算法的优越性和考虑空间位置因素的必要性.

关键词：土壤砷含量；城郊地区；GeoShapley；极限提升树；局部影响因子
中图分类号：X53 **文献标识码**：A **文章编号**：1000-6923(2026)01-0289-12

Exploring key local influencing factors of arsenic content in suburban soils of Changchun city using the GeoShapley approach.

ZOU Xin-ying[1] , YAN Qing-wu[1], WU Zi-hao[1*], ZHU Yuan-li[1], CHEN Yi-yun[2] (1.School of Public Policy & Management, China University of Mining and Technology, Xuzhou 221116, China；2.School of Resource and Environmental Science, Wuhan University, Wuhan 430079, China). *China Environmental Science*, 2026,46(1)：289~300

Abstract：sing 200soil As samples collected from suburban Changchun, this study combined the extreme gradient boosting tree (XGBoost) and the GeoShapley algorithm to explore the heterogeneity and nonlinear relationship between As and environmental factors, and to identify the local influencing factors and sources of As. The main conclusions are as follows: ① The XGBoost model had a fitted R^2 of 0.958, which was better than the geographically weighted regression and multiscale geographically weighted regression models, with fitted and predicted R^2 values of 0.958 and 0.729, respectively. ② The results of GeoShapley showed that the average contribution rate of manganese content ranked first and was the most significant characteristic variable affecting soil As, while the spatial location characteristic ranked second in importance, confirming the necessity of considering spatial coordinates. ③ The spatial distribution of GeoShapley values of influencing factors revealed that manganese content, alkali-hydrolyzable nitrogen, and sulfur content were the primary influencing factors for local As content at most sampling points, accounting for 99% of the samples. This indicated that parent material and agricultural production are the main sources of As. ④ The nonlinear relationship plots showed that there was a threshold effect between As content and various environmental variables, especially when the manganese content was higher than 2g/kg, the alkali-hydrolyzable nitrogen was higher than 175mg/kg, and the sulfur content was lower than 20mg/kg, As reached the peak, which should be paid special attention to in pollution prevention and control. This paper confirms the superiority of GeoShapley algorithm and the necessity of considering the spatial location, and provides data support for the governmental departments to formulate the policy of localized prevention and control of soil As pollution.

Key words：soil arsenic content；suburban area；GeoShapley；extreme gradient boosting tree；local influencing factor

砷是一种天然存在的有毒准金属,为我国第一类重点防控污染物[1].砷具有显著的生物富集性,能在农作物中累积[2],通过膳食途径对人体健康产生威胁.城郊地区土地利用方式复杂,受自然条件、工业及农业生产等多重影响,砷来源复杂.不同来源的砷污染物治理措施差异大,增加防治的难度.因此,准确识别土壤砷的局部主控因子及来源,是城郊地区砷污染防治的关键前提,有利于降低生态与健康风险.

收稿日期：2015-05-12
基金项目：国家自然科学基金资助项目(42201447)
* 责任作者, 副教授, 6247@cumt.edu.cn

地理加权回归模型(GWR)是一种通过建立局部回归方程探索空间异质关系的模型[3],能有效捕捉区域空间非平稳性特征,已被广泛应用于土壤重金属污染领域[4-6].然而,GWR 属于线性模型,无法揭示土壤污染与影响因子间复杂的非线性空间现象.

近年来,可解释的机器学习算法与沙普利加性解释(SHAP)结合的形式受到学者们广泛关注[7-9].一方面,学者们使用随机森林、极限梯度提升树和支持向量机等机器学习模型等探究土壤重金属污染与环境变量之间复杂非线性关系[9-11].相比于其他机器学习模型,极限梯度提升树模型(XGBoost)计算速率快,预测精度高,模型性能好.另一方面,SHAP 是一种基于博弈论的局部加性解释算法,根据 shapley 绝对值大小关系确定因变量的局部主要影响因子[12-13],例如,郑家桐等[14]使用 CatBoost-SHAP 耦合模型,以平均 SHAP 值量化了各因素的影响,从而确定土壤重金属污染的关键因子,Zhou 等[15]基于 XGBoost- SHAP 方法揭示了各环境变量与生态系统服务之间非线性关系.然而在土壤砷污染影响因素的识别研究中,重金属积累与环境变量的空间变化关系普遍存在[3],由于 SHAP 未考虑空间位置特征对模型的贡献,因此易高估其他变量的重要性,从而产生偏差.

针对上述问题,Li[16]提出了 GeoShapley 解释方法,该方法首次将空间坐标嵌入 Shapley 值计算过程,实现了空间特征重要性和空间交互效应的量化.它创新地采取了联合的方式来单独评估二维地理坐标重要性,提供空间显性解释,并表征非空间特征与地理坐标的非线性交互作用.该方法在满足空间匿名性公理的同时,又保证了预算平衡性,有效弥补 SHAP 解释方法在空间公平性分配上的不足,获得更真实的评估结果.

长春市宽城区是快速城镇化的典型区域,其东南部和西南部为工业区,北部和东部多为农业区,西北部多绿地,受功能区分异影响,砷污染来源存在明显空间异质性,需要针对性地进行防治.本研究采集宽城区 200 个土壤砷样本,使用 XGBoost 模型探究土壤砷含量与环境变量的非线性关系,并结合 GeoShapley 算法确定土壤砷的局部主控因子及来源,旨在为该地区砷污染防控提供科学支撑.

1 材料与方法

1.1 研究区概况

研究区位于长春市宽城区中南部的城郊地区(43°57′~44°7′ N,125°11′~125°24′ E)(图 1),占地面积约为 245km^2,属于温带大陆性半湿润季风气候,年平均降水量为 550mm,年均气温 4°C,海拔在 187~246m 之间,是东北重要的粮食生产区,土地利用类型以农用地和建设用地为主,少部分是绿地和水体,该地地质多为砂石、卵石夹砂、砾石,土壤类型主要是草甸土、黑土和黑钙土.由于近年来宽城区工业化、城镇化快速发展,加重了该区域土壤重金属污染的风险.

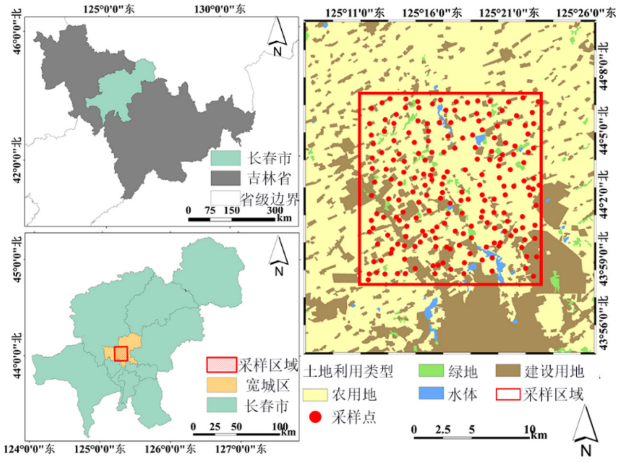


图 1 研究区位置和采样点分布

Fig.1 Location and SOC spatial distribution of the study area

该图基于自然资源局标准地图服务网站下载的审图号为 GS(2019)1822 号的标准地图制作,底图无修改

1.2 数据收集与预处理

1.2.1 土壤采样和实验室分析 基于土地利用类型,采用分层随机采样方法,在 2017 年 9 月采集了 200 个表土样本(深度为 0~20cm),其中耕地样点 175 个,建设用地样点 25 个.每个样品通过在 1m^2 区域内混合 5 个子样品制备而成.采样过程中使用木铲,每次采样后以净水清洗并用纸巾擦干,避免样品间的交叉污染.收集的土壤样品风干至恒重,过 2mm 筛,研磨并储存于实验室备用.取 0.25g 等分试样干燥土壤样品,用王水(65% HNO_3 与 37% HCl,体积比为 1:3)将其进行消解,随后使用原子荧光光谱法测定土壤中总砷的含量;碱解氮含量通过 Kjeldahl 方法测定;在土壤与水的比例为 1:2.5 的环境下,使用 pH 计测算土壤 pH 值;土壤

全量锰含量的测定采用硝酸-盐酸-氢氟酸高压消解法,消解液中的锰浓度通过火焰原子吸收光谱仪测定,并据此计算土壤中锰的全量;土壤硫含量则通过磷酸盐浸提-硫酸钡比浊法进行测定.

1.2.2　环境变量的获取及预处理　本研究从土壤砷的污染来源和迁移累积过程两方面选取了 17 个环境变量(表 1).污染来源方面包括农业生产、城镇化、工业生产和自然来源.其中,农业生产来源包括作物类型、速效钾含量、有效磷含量和碱解氮含量;城镇化来源包括道路密度和人口密度;工业生产来源包括到建设用地距离和 $PM_{2.5}$ 年均含量;自然来源包括土壤母质、锰含量和硫含量.在迁移累积方面,选取有机质、pH 值、高程、地形湿度指数、黏粒含量和粉粒含量 6 个环境变量.

表 1　潜在影响因子及数据源

Table 1　Potential impact factors and data sources

项目	类型	影响因子	数据来源	数据精度
主要来源	农业生产	作物类型	国家地球系统科学数据中心 http://www.geodata.cn	30m
		速效钾含量	国家地球系统科学数据中心 http://www.geodata.cn	90m
		有效磷含量	实测数据	—
		碱解氮含量	实测数据	—
	城镇化	道路密度	Open Street Map https://www.openstreetmap.org/	60m
		人口密度	LandScan 人口数据集 https://landscan.ornl.gov/	30m
	工业生产	到建设用地距离	根据土地利用数据计算得出	30m
		$PM_{2.5}$ 年均含量	国家青藏高原科学数据中心 https://data.tpdc.ac.cn/	30m
	自然来源	土壤母质	地质科学数据出版系统 http://dcc.ngac.org.cn	10m
		锰含量	实测数据	—
		硫含量	实测数据	—
迁移累积过程	环境因子	有机质	实测数据	—
		pH 值	实测数据	—
		高程	NASA 地球科学数据 https://nasadaacs.eos.nasa.gov/	30m
		地形湿度指数	NASA 地球科学数据 https://nasadaacs.eos.nasa.gov/	30m
		黏粒含量	SoilGrids 数据库 https://soilgrids.org/	30m
		粉粒含量	SoilGrids 数据库 https://soilgrids.org/	30m

高程数据(图 2(a))通过 NASA 地球科学数据(https://nasadaacs.eos.nasa.gov/)下载得到;通过公开信息,收集吉林省土壤环境重点监管企业名单和 2023 年宽城区污染源抽查企业名单,从名单中筛选具体的污染企业,再结合百度地图坐标拾取系统和坐标转换工具获得地理空间坐标;粉粒含量空间分布图(图 2(b))通过 SoilGrids 数据库(https://soilgrids.org/)下载得到;$PM_{2.5}$ 年均含量(图 2(c))获取自国家青藏高原科学数据中心(https://data.tpdc.ac.cn/),空间分辨率为 30m;道路密度(图 2(d))通过 Open Street Map(https://www.openstreetmap.org/)下载得到,其中包括高速公路、城市道路、郊区乡村道路等.以上数据均在 ArcGIS10.2 软件中处理.

1.3　统计分析模型

1.3.1　地理加权回归模型　地理加权回归模型(GWR)是一种基于线性回归框架的局部回归模型.它采用最优固定带宽,对研究范围内的各样本点建立局部回归方程[17],能有效捕捉空间非平稳性,揭示自变量与因变量间空间异质关系,适用于分析不同位置的驱动因子对土壤砷含量的影响.GWR 模型表

达公式如下:

$$Y_i = \beta_0(u_i,v_i) + \sum_{i=1}^{m}\beta_j(u_i,v_i)x_{i,j} + \varepsilon_i \quad (1)$$

式中:Y_i为因变量;(u_i,v_i)为点 i 处的 j 个解释变量的地理坐标;$\beta_0(u_i,v_i)$和$\beta_j(u_i,v_i)$为其截距和斜率,斜率是模型的回归系数;ε_i 为随机误差项.

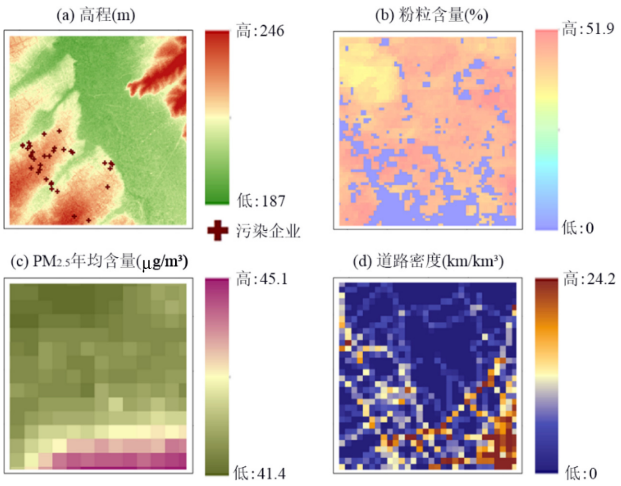


图 2 潜在影响因子空间分布

Fig.2 Spatial maps of potential influencing factors

1.3.2 多尺度地理加权回归模型 多尺度地理加权模型(Multi-scale GWR,MGWR)是地理加权回归模型的改进模型.它打破了 GWR 中关于变量的单一带宽假设,允许自变量以差异化的空间尺度影响因变量[18],有利于减少误差,提高模型的灵活性和可靠性,更接近现实空间过程.MGWR 模型表达式如下:

$$Y_i = \beta_0(u_i,v_i) + \sum_{i=1}^{m}\beta_{bw,j}(u_i,v_i)x_{i,j} + \varepsilon_i \quad (2)$$

式中:$\beta_{bw,j}$ 表示局部回归过程中不同自变量的差异化带宽;其他变量的含义与式(1)保持一致.

1.3.3 极限梯度提升树模型 极限梯度提升树模型(XGBoost)是以梯度提升决策树为基础的集成学习算法[19],能捕捉自变量和因变量之间复杂的非线性关系.XGBoost 模型集成多个弱学习器,并考虑前一轮预测的误差[20],大幅提升了模型预测性能.此外,XGBoost 对梯度提升决策树模型进行有效改进,采用并行计算的方式,显著提高计算效率[21].为避免模型过于复杂,引入正则化项(包括 L1 正则化和 L2 正则化),通过惩罚模型的复杂度,有效避免过拟合现象.目标函数公式如下所示:

$$f_{\text{obj}} = -\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j+\lambda} + \gamma T \quad (3)$$

$$G_i = \sum_{i\in I_j} g_i \quad H_j = \sum_{i\in I_j} h_i \quad I_j=\{i|q(x_i)=j\} \quad (4)$$

式中:g_i 和 h_i 分别为 1 阶和 2 阶导数;I 定义为每片叶子上的样本集合;I_j 为第 j 片叶子上的样本集合;结构函数 q 表示输入样本到叶子题注索引号之间的映射关系;T 为树 f 中叶子节点的总数目;γ 和 λ 为 XGBoost 自定义参数,γ 为 $L1$ 正则的惩罚项,λ 为 $L2$ 正则的惩罚项,f_{obj} 越小,说明树结构越优.

1.3.4 GeoShapley 算法 GeoShapley 是一种基于博弈论的空间效应测量方法,能够量化空间位置的重要性和考虑与其他特征的交互贡献[16].作为局部评估的解释方法,与 SHAP 算法类似,都与模型无关,并均基于 Shapley 值框架进行扩展评估,能测量每个特征的重要性以及对结果带来的正负效益[22],可视化自变量与因变量之间的非线性关系.然而,相比于 SHAP, GeoShapley 进一步考虑了空间自相关关系,以模型不可知的方式单独获取空间系数变化,通过计算空间位置贡献及其与其他特征变量的交互贡献,从而更准确地捕捉各点位上不同特征变量与因变量之间的关系,得出更真实的结论.毫无疑问,对于需要精细考虑空间作用的研究来说,GeoShapley 具有显著的优势.

计算 GeoShapley 值的公式具体如下:

$$\hat{y} = \phi 0 + \phi\text{GEO} + \sum_{j=1}^{p}\phi j + \sum_{j=1}^{p}\phi(\text{GEO},j) \quad (5)$$

式中:$\hat{y}$ 表示 GeoShapley 值的大小,$\hat{y}>0$ 时,表明在某个采样点上,某特征与砷含量呈正效益关系,$\hat{y}<0$ 时,表明在某个采样点上,某特征与砷之间是负效益关系,ϕ_0是常量基线值,表示在没有任何特征贡献的情况下,模型的基础输出,ϕ_{GEO} 表示空间位置特征对模型预测的边际贡献,测量了模型的空间固定效应,ϕ_j 表示非空间位置特征 x_j 对模型的独立贡献,这种贡献不受空间位置的影响,$\phi_{(\text{GEO},j)}$表示非空间位置特征 x_j 的空间交互效应.

$\phi_{(\text{GEO},j)}$计算公式如下所示:

$$\phi(\text{GEO},j) = \sum_{s\subseteq M\backslash\backslash\{\text{GEO},j\}}\frac{s!(p-s-g-1)!}{(p-g+1)!}\Delta\{\text{GEO},j\} \quad (6)$$

式中:GEO 表示为一组大小为 g 的位置要素,当地理坐标为(x,y)时,g 为 2;p 是特征总数;S 是特征组合的

集合;s 是 S 中特征组合的数量;$M\backslash\backslash\{GEO,j\}$是除了 $\{GEO, j\}$ 之外的特征集合.

1.3.5　模型实现　XGBoost 拥有出色的模拟空间非线性趋势能力,与 GeoShapley 结合表现良好.故本研究选择这两种模型进行土壤砷含量局部影响因子探究,所有模型的构建和可视化均在 Python 平台处理,主要使用 XGBoost 包和 Scikit-learn 包. XBGoost 模型中树的数量是 996,深度为 9,学习率为 0.01,行采样比例和列采样比例分别是 0.776 和 0.672,训练集的 R^2 为 0.958,均方根误差为 0.391,平均绝对误差为 0.308,残差平方和是 24.456.通过构建 GeoShapleyExplainer 解释模型的预测结果,将所有样本点作为背景样本,用于计算 GeoShapley 值.令 random_state=0 确保每次运行时抽样的结果一致,设置 n_jobs=-1 以加快计算速度.

为提高预测性能,避免模型过拟合,研究使用了遗传算法从 17 个环境因子中识别出对模型影响最优的特征变量组合,组合包括土壤锰含量、碱解氮、硫含量、粉粒含量、高程、道路密度、$PM_{2.5}$ 年均含量和有效磷,并将它们输入 MGWR2.2 软件中,用于 GWR 和 MGWR 建模.在软件中,输入自变量、因变量和坐标,Spatial Kernel 选择为 Adaptive Bisquare, Bandwidth searching 设置 Golden Section,Model Type 默认为 Gaussian,选择了 AICc 指标为 Optimization Criterion,Advanced Options 进行默认设置.最后得到 GWR 和 MGWR 两个模型的模拟结果.

1.3.6　模型精度评估　在本研究中,训练集和测试集划分比例为 8:2,其中训练集有 160 个点,测试集包含了 40 个点.研究采用决定系数(R^2)、均方根误差(RMSE)、残差平方和(RSS)评估并选择最优模型.具体公式如下所示:

$$R^2 = 1 - \sum_{1}^{n}(O_i - P_i)^2 / \sum_{1}^{n}(O_i - \overline{O})^2 \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum\nolimits_{1}^{n}(O_i - P_i)^2} \tag{8}$$

$$\mathrm{RSS} = \sum_{1}^{n}(O_i - P_i) \tag{9}$$

式中:n 代表了样本的数量;O_i 表示第 i 个样本的实际观测值;$\overline{O}$ 表示所有观测值的平均值;P_i 表示第 i 个样本的预测值.R^2 范围在 0~1,越接近 1,表示拟合效果越好;RMSE 值越大,误差也越大;RSS 越小,说明拟合程度越好,结果越可靠.

2　结果与分析

2.1　描述性统计

如表 2 所示,砷含量在 6.24~15.97mg/kg 之间,均值和中位数分别是 10.24 和 10.09mg/kg,本区域土壤中砷的背景值为 10.25mg/kg[12],平均砷浓度略低于背景值,尚未超出土壤砷污染风险筛选值(30mg/kg).然而,鉴于在土壤砷含量实测过程中,有 49%的样本点含量超过背景值,说明该地区会面临较大砷含量超标风险.变异系数为 18.26 %,表明土壤中砷含量的离散程度和空间异质性低,锰含量范围处在 374.4~2135.6mg/kg 之间,其均值和中位数分别是 728.5 和 717.1mg/kg,变异系数为 2.69%,变异程度小,碱解氮含量、硫含量和有效磷含量分别是 30~717.5mg/kg、10.75~232.71mg/kg 和 53~64mg/kg,变异系数分别为 39.79%,68.24%和 3.42%.根据描述性统计表内容显示,在实测的土壤性质中,硫的变异程度最高,空间异质性最强;锰的变异程度最轻,空间异质性和离散程度最低.

表 2　描述性统计

Table 2　Descriptive statistical results

实测元素	均值±标准差	最小值	中位数	最大值	变异系数(%)
砷(mg/kg)	10.24±1.87	6.24	10.09	15.97	18.26
锰(mg/kg)	728.5±19.55	374.4	717.1	2135.6	2.69
碱解氮(mg/kg)	149.53±59.50	30.00	145.90	717.50	39.79
硫(mg/kg)	31.50±21.50	10.75	26.11	232.71	68.24
有效磷(mg/kg)	59.41±2.03	53.00	59.33	64.00	3.42

2.2　模型比较

对土壤砷含量数据集进行空间自相关分析,发现其全局莫兰指数为 0.471(P<0.05),说明砷存在空间自相关性.根据遗传算法特征筛选结果,利用 SPSS 软件进行方差膨胀因子检验(VIF),诊断结果通过了 VIF 检验,表明特征之间均不存在多重共线性现象(VIF<5),因此将它们输入地理加权回归(GWR)和多尺度地理加权回归(MGWR)模型进行建模.

基于 80%数据集划分的训练集,各模型训练结

果如表 3 所示,三种模型中 XGBoost 的 R^2 最高,GWR 的 R^2 最低,说明 XGBoost 拟合效果最好.RMSE 衡量预测值与真实值之间差异,其中 MGWR 和 GWR 的 RMSE 均高于 XGBoost,说明在预测误差方面,XGBoost 准确性高.在残差平方和上,GWR 的数值最大,说明使用该模型后,数据的变异或者不确定性最强,MGWR 次之,而 XGBoost 的残差平方和最小,表明其预测值与实际观测值之间的总误差最小.

表 3 模型比较

Table 3 Model comparison results

模型	R^2	RMSE	RSS
GWR	0.637	0.603	58.113
MGWR	0.746	0.504	40.644
XGBoost	0.958	0.391	24.456

同时,根据 XGBoost 测试集的结果可知,测试 R^2 为 0.729,结合训练集评价指标情况,模型没有出现显著性能下降,未观察到过拟合现象.

综上分析,XGBoost 模型是最优模型,因此,后续只展示 XGBoost 模型的结果.

2.3 各影响因子全局和局部相对重要性

2.3.1 影响因子重要性 基于 GeoShapley 值揭示 XGBoost 模型中各环境变量的相对重要性.图 3 结果表明,对土壤砷含量产生首要、第二影响的特征分别是锰含量和空间位置.空间位置特征的高影响性证实了 GeoShapley 的有效性和必要性.碱解氮、土壤硫含量和粉粒含量的 GeoShapley 相对平均值分别位列第三、第四和第五,它们对土壤砷含量的影响虽然不及锰含量和空间位置特征显著,但作用仍不可忽视.高程、$PM_{2.5}$ 年均含量和有效磷含量对土壤砷含量也存在一定影响.

在交互效应上,碱解氮、土壤锰含量和硫含量与空间位置交互效应最高,说明碱解氮、土壤锰含量和硫含量在不同的空间位置上的表现有显著差异,进一步凸显空间异质性的重要性.粉粒含量、有效磷和道路密度的交互效应最低,说明这些特征在不同空间位置上的变化对土壤砷含量的影响相对一致.

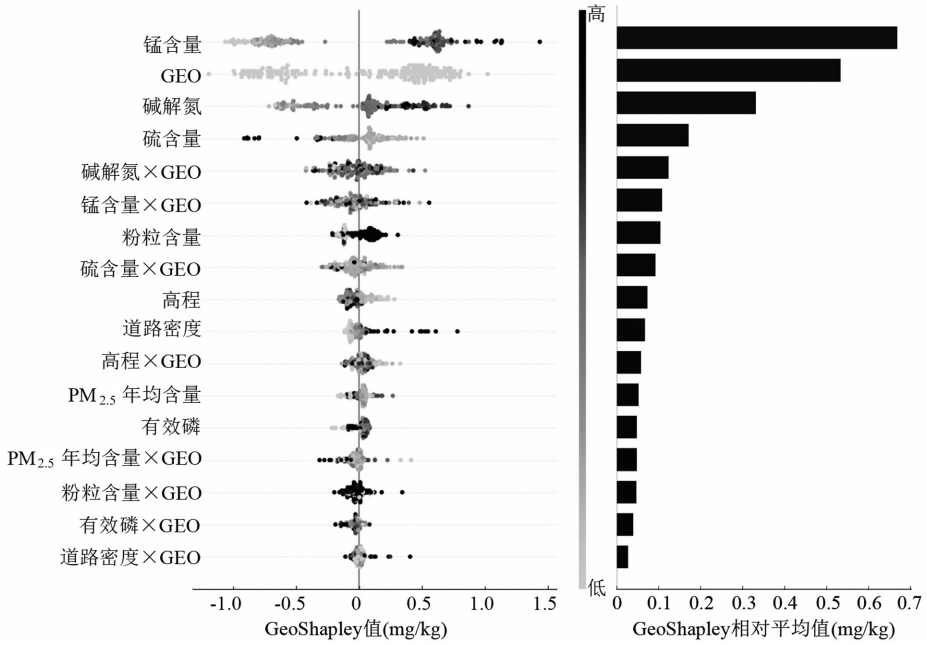


图 3 影响因子相对重要性图

Fig.3 Relative importance of impact factors

表 4 是影响因子的 GeoShapley 值情况.标准差能够衡量数据在数值空间中的离散程度.锰的标准差最大,说明数据点空间分布越分散,其 GeoShapley 值波动和变异程度也最大.有效磷的标准差最小,其数据波动较平缓,变异程度最小,数据点空间分布越集中.中位数会忽略极端数值的影响,分析数据集中趋势,按照绝对值从大到小排序依次是锰含量、碱解氮、硫含量、粉粒含量、道路密度、$PM_{2.5}$ 年均含量、有效磷和高程.结合图 4 和表 4 的结果,对土壤砷含量贡献度最大的前 3 个影响因子是锰、碱解氮和硫,这 3 个影响因子的污

染来源分别是自然来源和农业生产.

表 4　影响因子 GeoShapley 值的描述性统计结果

Table 4　Descriptive statistical results of the GeoShapley values of impact factors

影响因子	标准差 (mg/kg)	中位数 (mg/kg)	最小值 (mg/kg)	最大值 (mg/kg)
锰含量	0.71	0.34	−1.15	1.72
碱解氮	0.42	0.06	−0.93	1.09
硫含量	0.26	0.05	−1.00	0.69
粉粒含量	0.12	0.03	−0.28	0.48
高程	0.11	−0.01	−0.17	0.44
道路密度	0.14	−0.03	−0.15	0.99
$PM_{2.5}$年均含量	0.09	0.02	−0.25	0.38
有效磷	0.07	0.02	−0.24	0.13

2.3.2　影响因子 GeoShapley 值空间分布　图 4 为各影响因子 GeoShapley 值空间分布图,其中,锰含量的 GeoShapley 数值跨度最大,负值大部分分布在研究区北部采样点中,而南部采样点中的锰含量 GeoShapley 多为正值.在碱解氮的 GeoShapley 空间分布图中,数值大于 0 的采样点以东部、中部和西南方向为主,数值小于 0 的采样点在北部和东南部分布密集.硫含量和有效磷的正负 GeoShapley 值分布都较离散.粉粒含量和高程的 GeoShapley 空间分布相似,数值小于 0 的采样点多在研究区的正北、中部、西南和东南部.道路密度 GeoShapley 值以负向居多,说明该特征与土壤砷含量之间负效应明显.$PM_{2.5}$年均含量中 GeoShapley 值大于 0 的采样点沿着东北-西南方向块状分布.

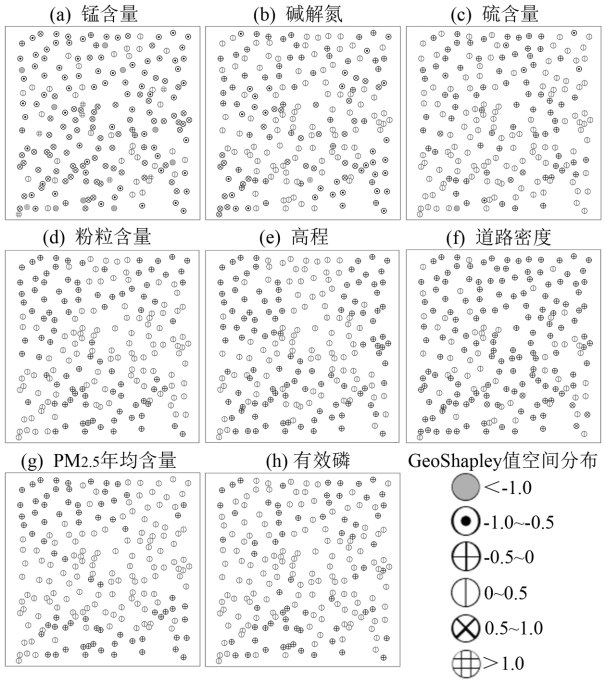


图 4　影响因子 GeoShapley 值空间分布

Fig.4　Spatial distribution of the GeoShapley values of the impact factors

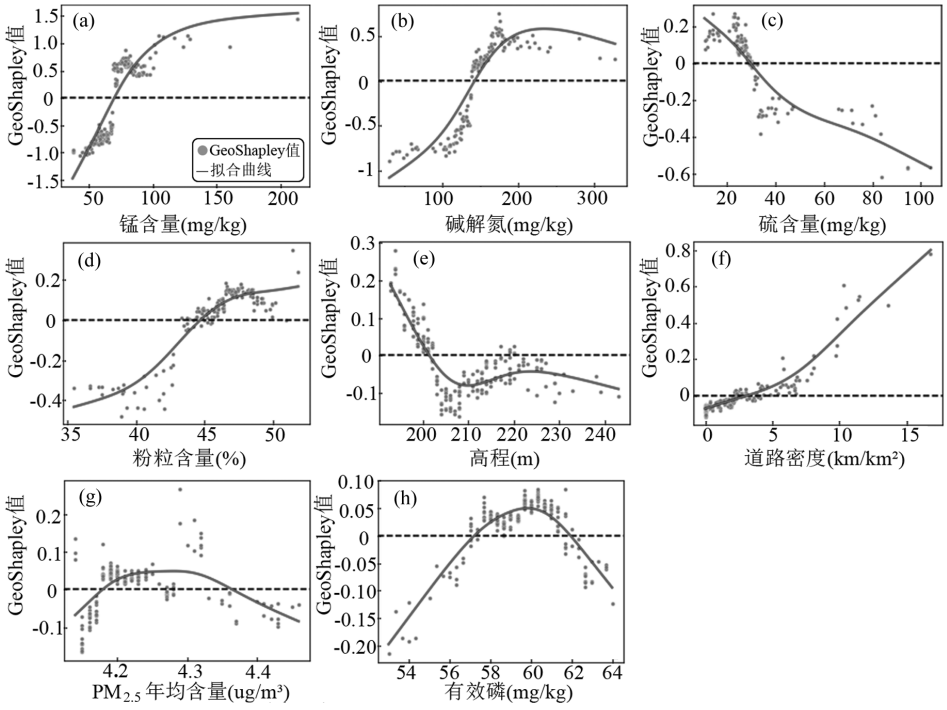


图 5　影响因子与总砷的非线性关系

Fig.5　Nonlinear relationship between impact factor and total As

2.3.3 土壤砷含量与影响因子的非线性关系分析 根据影响因子与总砷含量的非线性关系图(图5),土壤砷含量和锰含量、碱解氮、粉粒含量和道路密度均总体呈正相关关系,说明从整体上看,砷含量随着这些影响因子的含量增加而增加.不同的是,锰含量超过 1g/kg,砷增长趋势放缓,而粉粒含量达到47 %的阈值后,土壤砷含量达到顶峰,之后呈轻微下降趋势.当碱解氮含量达到 175mg/kg,此时砷含量最高,随后缓慢下降.道路密度超过 6km/km² 时,砷的上升趋势由平缓转变为陡峭,说明道路密度越大,越有利于砷的积累.

土壤砷与硫含量、高程的关系总体呈负相关.其中,砷与高程呈现复杂的波动关系,高程超过210m 时,砷含量由下降趋势转为上升趋势,然而,当高程超过 223.5m,趋势由递增转为递减.自然来源的硫与土壤中的砷含量为负相关关系,这是因为土壤中的硫元素被作物大量消耗却无法得到及时补充,同时农户又施用含砷的高浓度肥料,导致两者出现负相关关系[23].因此,硫无法作为污染来源判断的依据.

土壤中砷含量随着 $PM_{2.5}$ 年均含量递增,呈现先上升后下降的变化趋势,总体表现为正相关关系.有效磷含量为 59.9mg/kg 时,土壤中砷的含量增长至最高值,但随着有效磷含量超过 59.9mg/kg 后,砷与有效磷转而呈负相关关系.

2.4 砷的局部主控因子及来源

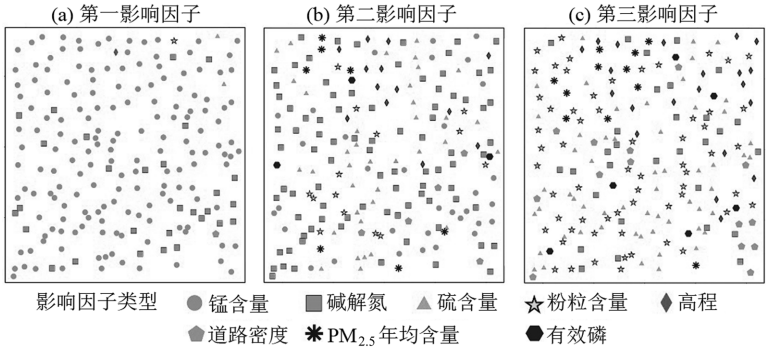


图 6 砷的局部主要影响因子

Fig.6 The main local influencing factors of As

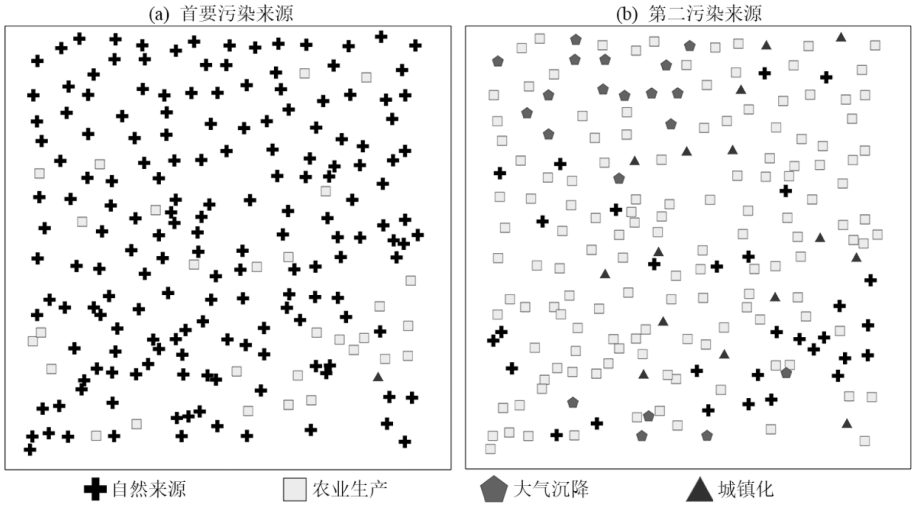


图 7 砷的局部来源

Fig.7 Local sources of As

根据采样点的影响因子 GeoShapley 绝对值大小分别确定了首要、第二和第三影响因子.如图 6(a)示,在 200 个采样点中,首要影响土壤砷含量的主导因子是土壤锰含量,而受到碱解氮、硫含量、粉粒含量和高程等因子首要影响的样本点仅有 34 个.对于第二影响因子(图 6(b)),碱解氮和硫含量作为第二影响因子的样本点的比例分别为44%和18.5%,锰含量影响的点有33个,占比16.5%.第三影响因子(图6(c))

中,采样点的砷含量主要受到硫含量和碱解氮的影响,其中,以硫含量为第三影响因子的点有 56 个,占比 28%,受到碱解氮影响的点有 30 个.综上,研究区采样点砷含量的高低会显著受到土壤锰、碱解氮和硫含量影响.

结合砷的局部主要影响因子与表 1 中的来源对应表,得出每个点的土壤砷来源(已剔除环境因子).图 7(a)中,以锰含量为主要特征的自然来源是研究区最主要来源.在第二来源(图 7(b))中,有 135 个样本点砷来源是以碱解氮为主的农业生产;在工业生产中,将 $PM_{2.5}$ 年均含量细化为大气沉降来源,受该来源影响的点多分布在研究区域北部;以道路密度为主要特征的城镇化源,多分布在中部.

综上所述,土壤砷含量的来源以农业生产和自然来源为主,受土壤自然属性影响最明显.

3　讨论

3.1　GeoShapley 有效性分析

GWR 是一种以空间非平稳性为前提的局部回归模型[24].基于空间异质性,该模型对传统回归方法进行有效扩展,通过将采样点的空间坐标纳入回归参数[25],来探究不同空间位置上的特征变量对土壤砷含量的影响程度.但 GWR 模型采用单一的核函数和带宽计算权重,忽略了现实中不同变量间差异化作用尺度[26],导致产生偏差.为解决这一问题,MGWR 在 GWR 的基础框架上进行扩展,允许不同变量的空间作用带宽发生变化[26-28],兼顾局部效应和尺度效应,模型更加灵活稳健,符合真实的空间过程.但究其二者本质,两类模型都以自变量与因变量之间的线性关系为前提.根据目前的研究[29-31],发现土壤重金属污染与环境变量之间存在复杂非线性关系和交互效应,导致 GWR 和 MGWR 模型在该领域的应用中存在局限性.而 GeoShapley 解释方法与机器学习模型结合后,可以展示非线性关系[32],并提供更加合理可靠的解释.

SHAP 以加法特征归因为假设,基于联盟博弈论理论,对 shapley 值进行加性组合来解释复杂的黑箱模型,以此合理公平地分配每个特征对模型的贡献[22].其优点在于能够同时提供全局和局部层面的解释,有效揭示自变量和因变量之间的非线性关系.而它的局限性在于难以单独评估空间位置特征的重要性,无法有效捕捉空间交互效应.同时,由于 SHAP 在评估特征主要效应时,会包含空间特征与其他特征的交互效应,导致高估某些特征的贡献,进而在空间公平性分配中存在偏差.而 GeoShapley 在 shapley 的价值框架上引入地理权重矩阵,并采用 Joint Shapley 方法解释空间坐标组合特征的贡献,量化空间依赖关系[16], 实现了空间公平性.据研究结果可知,GEO 特征贡献度排名第二,证明必须将地理坐标纳入考虑范畴,所以更应该使用 GeoShapley,而非 SHAP 作为 XGBoost 的解释方法.除此之外,GeoShapley 算法能区分特征的主要效应和交互效应,更加真实地反映每个非空间特征的实际重要性.

综上所述,在分析土壤砷含量局部影响因子过程中,GeoShapley 算法的解释效果最好,有利于精准探究不同位置的具体污染来源,从而制定针对性的措施和砷污染防治战略.

3.2　影响因子对土壤砷的作用分析

GeoShapley 的结果表明,锰含量、碱解氮和硫含量是研究区砷含量的局部主控因子,土壤母质和农业生产是砷的主要来源.

这一结论既有区域特殊性,又存在一定共性.Zhu 等[33]利用 MGWR 探究长春土壤砷主控因子发现,道路密度、黏粒含量、砂粒含量和氮元素对砷有显著影响,这与本文结论有一定重合;Zhang 等[34]基于随机森林模型分析了湖南西部地区砷污染的影响因子,研究结果表明高程、道路和距离河流距离为前三位关键因子.这与本研究确定的主导因子不同,可能是研究区的地域差异导致.Zhang 等[34]以南方稻田土壤为重点监测对象,因地形原因,湖南水稻广泛种植于梯田,而高程会通过影响地表径流、沉积物积累等,直接或间接对水稻土的砷产生极大影响,而本研究区属于松辽平原,地形平坦,且处于城郊地带,工农业交织,高程作用不明显.同时,湖南以双季稻种植为主,种植区需要频繁灌溉,在此过程中可能将河流周边的含砷物质带入土壤.但长春种植单季稻,灌溉类型也非河流依赖型灌溉,且研究区水域面积小,因此距离河流距离因素对本研究区的砷含量影响小.总结可知,砷含量的变化具有环境依赖性,因此在解释主导因子对砷的影响机制时,要结合当地实际情况具体分析.

长春宽城区土壤中的锰主要来源于锰矿石风

化、成土母质分解以及其他含锰矿物的地球化学过程.作为一种重要的非生物氧化剂,锰在土壤中广泛存在,其氧化物(如 MnO_2)及铁锰复合氧化物(如 Fe−Mn 氧化物)对砷具有显著的吸附固定能力[35-37],通过表面配位、离子交换等机制与砷形成稳定的共沉淀物,从而导致锰矿物中易富集砷,体现为锰与土壤中砷含量呈正相关关系.李月芬等人[38]研究吉林省西部土壤中 7 种不同形态砷与锰元素的相关性,发现锰和多形态砷均呈正相关关系,与水溶态、腐植酸结合态和残渣态砷的相关性尤为显著.Zhai 等[39]研究表明,中国东北松嫩平原地区的铁、锰元素与砷呈现协同富集现象,同样,Yang 等[40]进行长春地区表土重金属污染评价时,同样发现该区域的土壤砷与铜、锌、铅和锰等重金属存在显著正相关.这与本研究区域高度吻合,结论也保持一致.故针对因土壤锰的高含量而导致砷超标的地区,可以适当施用锰改性生物炭和种植锰高效作物[41-42],通过主动吸收锰元素的方式来辅助降低土壤的砷含量.

碱解氮作为植物生长发育必须的元素之一,在农业生产过程中,常通过施用氮肥或者有机肥料增加土壤碱解氮的含量,以提升作物产量.而动物饲料包含诸如罗沙松等砷化合物,使得动物粪便有砷的存在[33].当农户持续施用动物粪便这类有机肥时,土壤砷含量也会随之增加.在探讨氮和土壤砷的关系中,理论分析与本研究结论皆论证为正相关关系,这与贾正雷[43]的研究结果一致.因此,对氮肥施用进行优化管理是必要的,要根据作物生长的动态需要调整施氮量,以实现精准施肥,还能通过使用无砷的动物饲料添加剂、定期监测土壤中的氮元素和养殖场饮水的砷含量等方式,达到降低土壤砷含量的目的.

土壤中硫符合该区域的背景值(70mg/kg)[44],因此来源于自然环境.砷作为一种亲硫元素,易与硫生成不溶性硫化矿物,常见的有砷黄铁矿、雄黄和毒砂等[35].含砷的硫化矿物通过风化或分解等地球化学过程聚积在土壤中,两者关系应体现为正相关关系.然而,本研究发现硫和砷呈负相关,这一现象可能与研究区长期的农业管理方式密切相关.研究区土壤长期连续种植高产作物,导致硫元素被作物大量吸收消耗,与此同时,农户多施用复合肥和高浓度肥料,这些肥料含有砷元素,还无法有效补充土壤中缺失的硫,造成硫含量逐渐降低[23].因此,随着硫的输出和砷的持续输入,两者呈负相关关系.需要指出的是,硫本身不能对砷产生有效影响,故无法将硫作为污染来源的判断标准和直接依据,这一结果强调了在农业生产中平衡养分输入的重要性,以减少土壤中潜在污染物的积累.农户要及时为土壤补充硫肥,但要注意避免与硝态氮肥混施.

粉粒由于其小粒径,拥有较大的比表面积,能够为砷等污染物提供更多吸附点,因此,土壤中粉粒含量越高,砷的含量也随之增加[45].本研究发现土壤中粉粒含量与砷含量之间的正相关关系,这与 Zhu 等[33]研究结果一致.

道路密度多受城镇化影响,交通活动贡献尤为显著.道路密度越大的区域,交通流量相应也越高.车辆在行驶过程中,燃料燃烧会释放含砷的尾气,同时轮胎磨损产生的含砷颗粒物会沉积在道路周边的土壤中,导致砷含量升高.因此,道路密度与土壤砷含量呈现正相关关系,这与本研究的结果一致.此外,燃煤发电厂、金属冶炼厂、化工厂等在其工业活动过程产生的 $PM_{2.5}$ 等颗粒物通常含有砷及其他污染物,会通过大气沉降作用进入土壤,加剧土壤砷污染.本研究结果表明土壤中 $PM_{2.5}$ 年均含量与砷含量之间总体呈正相关关系,这与现实的污染机制相符.

综上所述,基于砷与环境变量之间复杂非线性关系的研究,得出研究区内土壤砷含量高低受到土壤自然特征影响最大.通过探讨砷的局部影响因子和不同点位的污染来源,提出针对性解决措施,对未来砷污染防控提供思路和指导.

4 结论

4.1 与 GWR、MGWR 相比, XGBoost 模型具有更高的精度和拟合优度,训练集 R^2 为 0.958,测试集 R^2 为 0.729.

4.2 GeoShapley 是基于 shapley 价值框架扩展的解释方法,根据其相对重要性图可知,锰含量的平均贡献率居首位,是影响土壤砷最显著的特征变量,而空间位置特征的第二重要性表明了自变量与土壤砷污染间的空间异质性,并证实了考虑空间坐标的必要性.

4.3 结合影响因子的 Geoshapley 空间分布图集发现,锰含量、碱解氮和硫含量是 99%样本点土壤砷含量的首要影响因子,但硫无法直接影响土壤中砷含量,不能成为判断来源的依据.这表明了母质和农业

生产是砷的主要来源,其中自然环境的影响最明显.

4.4　非线性关系图显示砷含量与各环境变量之间均存在阈值效应,特别是当锰含量高于 2g/kg、碱解氮高于 175mg/kg、硫含量低于 20mg/kg 时,砷含量达到峰值,在污染防控时应予以特别关注.

参考文献：

[1] 纪冬丽,孟凡生,薛　浩,等.国内外土壤砷污染及其修复技术现状与展望 [J]. 环境工程技术学报, 2016,6(1):90−99.
Ji D L, Meng F S, Xue H, et al. Situation and prospect of soil arsenic pollution and its remediation techniques at home and abroad [J]. Journal of Environmental Engineering Technology, 2016,6(1):90−99.

[2] 梁富威,肖丽莎,李镇城,等.中国土壤砷污染与植物修复的研究进展 [J]. 生态科学, 2024,43(6):250−258.
Liang F W, Xiao L S, Li Z C, et al. Research progress of arsenic contamination in soil and phytoremediation in China [J]. Ecological Science, 2024,43(6):250−258.

[3] Chen Z, Zhang S, Geng W, et al. Use of geographically weighted regression (GWR) to reveal spatially varying relationships between Cd accumulation and soil properties at field scale [J]. Land, 2022,11(5):635.

[4] Wang Z, Xiao J, Wang L Q, et al. Elucidating the differentiation of soil heavy metals under different land uses with geographically weighted regression and self−organizing map [J]. Environmental Pollution, 2020, 260:114065.

[5] Yang L X, Meng F H, Chen M, et al. Elucidating the spatial determinants of heavy metals pollution in different agricultural soils using geographically weighted regression [J]. Science of the Total Environment, 2022,853:158628.

[6] Song X D, Brus D J, Liu F, et al. Mapping soil organic carbon content by geographically weighted regression: A case study in the Heihe River Basin, China [J]. Geoderma, 2016,261:11−22.

[7] Jas K, Dodagoudar G R. Explainable machine le−arning model for liquefaction potential assessment of soils using XGBoost−SHAP [J]. Soil Dynamics and Earthquake Engineering, 2023,165:107662.

[8] Wang J, Deng Y, Huang Z, et al. Identification of driving factors for heavy metals and polycyclic aromatic hydrocarbons pollution in agricultural soils using interpretable machine learning [J]. Science of the Total Environment, 2025,960:178384.

[9] Song X, Sun Y, Wang H, et al. Uncovering soil heavy metal pollution hotspots and influencing mechanisms through machine learning and spatial analysis [J]. Environmental Pollution, 2025,370:125901.

[10] Huang H, Huang S, Du Q. Evaluation of soil heavy metal pollution based on K−means and SVM [J]. International Journal of Environmental Science and Technology, 2023,20(11):12015−12024.

[11] Xu X, Liu W, Jia T, et al. Investigation of relationship between polycyclic aromatic hydrocarbons and human activities in urban soils of China using machine learning methods [J]. Journal of Cleaner Production, 2023,404:136839.

[12] Yan Y, Yang Y. Revealing the synergistic spatial effects in soil heavy metal pollution with explainable machine learning models [J]. Journal of Hazardous Materials, 2025,482:136578.

[13] Alam S M K, Li P, Rahman M, et al. Key factors affecting groundwater nitrate levels in the Yinchuan Region, Northwest China: Research using the eXtreme Gradient Boosting (XGBoost) model with the SHapley Additive exPlanations (SHAP) method [J]. Environmental Pollution, 2025,364:125336.

[14] 郑家桐,王　鹏,石航源,等.基于 CatBoost 模型和 SHAP 解释方法的土壤重金属影响因素与程度定量分析 [J]. 环境科学学报, 2023, 43(4):448−456.
Zheng J T, Wang P, Shi H Y, et al, Quantitative analysis of the influence factors and extent of soil heavy metals based on CatBoost model and SHAP interpretation method [J]. Acta Scientiae Circumstantiae, 2023,43(4):448−456.

[15] Zhou B, Chen G, Yu H, et al. Revealing the nonlinear impact of human activities and climate change on ecosystem services in the karst region of southeastern Yunnan using the XGBoost–SHAP model [J]. Forests, 2024,15(8):1420.

[16] Li Z Q. GeoShapley: A game theory approach to measuring spatial effects in machine learning models [J]. Annals of the American Association of Geographers, 2024,114(7):1365−1385.

[17] 刘旭峰,陈文明,杨益民,等.基于 GWR 模型的湖南省耕地非粮化驱动因子及空间分异特征研究 [J]. 湖南师范大学自然科学学报, 2025,48(2):93−101.
Liu X F, Chen W M, Yang Y M, et al. Research on driving factors and differentiation characteristics of the spatial pattern of non−grain cultivated land in hunan province based on GWR model [J]. Journal of Natural Science of Hunan Normal University, 2025,48(2):93−101.

[18] 祝新明,宋小宁,冷　佩,等.多尺度地理加权回归的地表温度降尺度研究 [J]. 遥感学报, 2021,25(8):1749−1766.
Zhu X M, Song X N, Leng P, et al, Spatial downscaling of land sur−face temperature with the multi−scale geographically weighted reg−ression [J]. National Remote Sensing Bulletin, 2021,25(8):1749−1766.

[19] 李春泽,张　超,张皓源,等.基于 XGBoost 与地理加权回归的吉林省西部土壤盐渍化反演 [J]. 中国农业大学学报, 2024,29(2):1−10.
Li C Z, Zhang C, Zhang H Y, et al, Inversion of soil salinization in western Jilin Province based on XGBoost and Geographically Wei−ghted Regression [J]. Journal of China Ag−ricultural University, 2024,29(2):1−10.

[20] 罗　妍,王　枞,叶文玲.基于 XGBoost 和 SHAP 的急性肾损伤可解释预测模型 [J]. 电子与信息学报, 2022,44(1):27−38.
Luo Y, Wang C, Ye W L. An interpretable prediction model for acute kidney injury based on XGBoost and SHAP [J]. Journal of Electronics & Information Technology, 2022,44(1):27−38.

[21] 王桂兰,赵洪山,米增强.XGBoost 算法在风机主轴承故障预测中的应用 [J]. 电力自动化设备, 2019,39(1):73−7+83.
Wang G L, Zhao H S, Mi Z Q. Application of XGBoost algorithm in prediction of wind motor main bearing fault [J]. Electric Power Automation Equipment, 2019,39(1):73−77,83.

[22] 刘瑞雪,李佳轩,李　云.水库消落带植物多样性空间格局预测模型及环境解释——基于 XGBoost−SHAP 模型框架 [J]. 生态学报, 2024,44(21):9652−9669.
Liu R X, Li J X, Li Y. Prediction model and environmental interp−retation for the spatial pattern of plant diversity in the water−level fluctuation zone of reservoir based on XGBoost and SHAP [J]. Acta Ecologica Sinica, 2024,44(21):9652−9669.

[23] 崔　帅,刘烁然,王　寅,等.吉林省旱地土壤有效硫含量及其与土壤有机质和全氮的关系 [J]. 中国农业科学, 2022,55(12):2372−2383.
Cui S, Liu S R, Wang Y, et al. Soil available sulfur content in Jilin province and Its correlation with soil organic matter and soil total nitrogen [J]. Scientia Agricultura Sinica, 2022,55(12):2371−2383.

[24] 李桂华,王俊松,齐　洁.上海住宅价格时空分异格局及其影响因素——基于地理加权回归模型的分析 [J]. 地域研究与开发, 2024,43(6):74−81.

Li G H, Wang J S, Qi J. Spatio-temporal differentiation pattern of housing prices in Shanghai and its influencing factors: Analysis based on GWR model [J]. Areal Research and Development, 2024,43(6):74-81.

[25] 赵 慧,李新国,靳万贵,等.基于地理加权回归模型的博斯腾湖湖滨绿洲土壤盐分离子含量高光谱估算 [J]. 土壤, 2021,53(3):646-653.
Zhao H, Li X G, Qi W G, et al. Hyperspectral estimation of soil salt ion contents in lakeside oasis of bosten lake based on Geographical weighted regression model [J]. Soils, 2021,53(3):646-653

[26] 栗珂珂,周詹杭,王 真.中国土壤侵蚀的社会经济空间驱动因素研究:基于多尺度地理加权回归模型的分析 [J]. 华中农业大学学报, 2024,43(6):29-38.
Li K K, Zhou Z H, Wang Z. Socioeconomic spatial driving factors of soil erosion in China: An analysis based on a multi-scale geographically weighted regression model [J]. Journal of Huazhong Agricultural University, 2024,43(6):29-38.

[27] 吴子豪,刘耀林,冯向阳,等.基于多尺度地理加权回归的土壤镉污染局部影响因子分析 [J]. 地球信息科学学报, 2023,25(3):573-587.
Wu Z H, Liu Y L, Feng X Y, et al. Analysis of local influencing factors of cadmium pollution in Soil by using multi-scale geographically weighted regression [J]. Journal of Geoinformation Science, 2023,25(3):573-587.

[28] 姚雯雯,赵新正,马海涛,等.基于多尺度地理加权回归模型的中国城市网络外部性的空间异质性比较研究 [J]. 地理研究, 2024,43(5):1073-1087.
Yao W W, Zhao X Z, Ma H T, et al. Ancomparative study on spat-ial heterogeneity of China's urban network externality based on Multiscale Geographically Weighted Regression Model [J]. Geographical Research, 2024,43(5):1073-1087.

[29] 胡贵贵,杨粉莉,杨联安,等.基于主成分和机器学习的土壤有机质含量空间预测建模 [J]. 干旱区地理, 2021,44(4):1114-1124.
Hu G G, Yang F L, Yang L A, et al. Spatial prediction modeling of soil organic matter content based on principal c-omponents and m-achine learning [J]. Arid Land Geography, 2021,44(4):1114-1124.

[30] 马国林,丁建丽,韩礼敬,等.基于变量优选与机器学习的干旱区湿地土壤盐渍化数字制图 [J]. 农业工程学报, 2020,36(19):124-131.
Ma G L, Ding J L, Han L J, et al. Digital mapping of soil saliniz-ation in arid area wetland based on variable optimized selection and machine learning [J]. Transactions of the Chinese Society of Agricultural Engineering, 2020,36(19):124-131.

[31] 沈佳丽,陈颂超,胡碧峰,等.基于机器学习的江汉平原土壤有机碳预测及制图 [J]. 农业资源与环境学报, 2023,40(3):644-650.
Shen J L, Chen S C, Hu B F, et al. Prediction and mapping soil organic carbon in Jianghan Plain by machine learning [J]. Journal of Agricultural Resources and Environment, 2023,40(3):644-650.

[32] Ke E, Zhao J, Zhao Y. Investigating the influence of nonlinear spatial heterogeneity in urban flooding factors using geographic explainable artificial intelligence [J]. Journal of Hydrology, 2025,648:132398.

[33] Zhu Y L, Liu B, Jin G, et al. Identifying the local influencing factors of arsenic concentration in suburban soil: A multiscale geographically weighted regression approach [J]. Toxics, 2024,12(3):229.

[34] Zhang R J, Jiang L H, Dong T H, et al. Effects of geographical and soil factors on soils arsenic levels: a case study in typical arsenic-contaminated paddy fields based on machine learning [J]. Environmental Management, 2025,75(9):2272-2286.

[35] 安礼航,刘敏超,张建强,等.土壤中砷的来源及迁移释放影响因素研究进展 [J]. 土壤, 2020,52(2):234-246.
An L H, Liu M C,Zhang J Q, et al. Sources of Arsenic in Soil and Affecting Factors of Migration and Release: A Review [J]. Soils, 2020,52(2):234-246.

[36] 黄文妍,杨文俊,汤叶涛,等.土壤环境中铁锰氧化物对重金属的吸附络合机制及其影响因素研究进展 [J]. 地球环境学报, 2024,15(11):1-13.
Huang W Y, Yang W J, Tang Y T, et al. Research progress on the mechanism for the adsorption and complexation of heavy metals onto iron and manganese oxides and its influencing factors in soil environment [J]. Journal of Earth Environment, 2024,15(11):1-13.

[37] 王鑫浩,范 芳,张洛红,等.锰氧化物对环境中重金属吸附及影响因素研究进展 [J]. 应用化工, 2018,47(1):155-9+64.
Wang X H, Fan F, Zhang L H, et al. Research progress on adsor-ption of heavy metals by manganese oxides and its influencing factors [J]. Applied Chemical Industry, 2018,47(1):155-9+64.

[38] 李月芬,王冬艳,汤 洁,等.吉林西部土壤砷的形态分布及其与土壤性质的关系研究 [J]. 农业环境科学学报, 2012,31(3):516-522.
Li Y F, Wang D Y, Tang J, et al. Speciation of soil arsenic and its correlation with soil properties in western Jilin province,China [J]. Journal of Agro-Environment Science, 2012,31(3):516-522.

[39] Zhai Y, Cao X, Xia X, et al. Elevated Fe and Mn concentrations in groundwater in the Songnen Plain, Northeast China, and the Factors and Mechanisms Involved [J]. Agronomy, 2021,11(12):2392.

[40] Yang Z P, Lu W X, Long Y Q, et al. Assessment of heavy metals contamination in urban topsoil from Changchun City, China [J]. Journal of Geochemical Exploration, 2011,108(1):27-38.

[41] 张彦儒,孙思佳,王雅君,等.$MnSO_4$ 与钝化材料联合施用对镉污染稻田的修复效果 [J]. 农业环境科学学报, 2025,https://link.cnki.net/urlid/12.1347.S.20250529.1642.004.
Zhang Y R, Sun S J, Wang Y J, et al. Effect of combined application of $MnSO_4$ and passivation materials on the remediation performance of cadmium-contaminated rice fields [J]. Journal of Agro- Environment Science, 2025, https://link.cnki.net/urlid/12.1347.S.20250529.1642.004.

[42] 郭炳跃,杨锟鹏,张 璟,等.二氧化锰/氨基改性生物炭对铅、镉复合污染土壤的钝化修复研究 [J]. 生态与农村环境学报, 2023,39(3):422-428.
Guo B Y, Yang K P, Zhang J, et al. Research on the remediation effect of manganese dioxide/aminomodified biochar on Pb and Cd composite contaminated soil [J]. Journal of Ecology and Rural Environment, 2023,39(3):422-428.

[43] 贾正雷.土壤砷和氮含量的空间变异及其相互关系研究 [D]. 广州:华南农业大学, 2016.
Jia Z L. Study on spatial variability and relationship of soil arsenic and soil nitrogen [D]. GuangZhou: South China Agricultural University, 2016.

[44] 成杭新,李 括,李 敏,等.中国城市土壤化学元素的背景值与基准值 [J]. 地学前缘, 2014,21(3):265-306.
Cheng H X, Li K, Li M, et al. Geochemical background and base-line value of chemical elements in urban soil in China [J]. Earth Science Frontiers, 2024,21(3):265-305.

[45] Barkhordari M S, Qi C. Prediction of soil arsenic concentration in European soils: A dimensionality reduction and ensemble learning approach [J]. Journal of Hazardous Materials Advances, 2025,17:100604.

致谢：感谢国家科技基础条件平台-国家地球系统科学数据中心(http://www.geodata.cn)提供数据支撑.感谢国家青藏高原科学数据中心(http://data.tpdc.ac.cn)提供数据支撑

作者简介：邹心颖(2002-),女,江西宜春人,中国矿业大学硕士研究生,主要从事数字土壤制图和空间分析.发表论文 1 篇.ZouXinying@cumt.edu.cn.