



文章栏目：土壤污染防治

DOI 10.12030/j.cjee.202308001

中图分类号 X53

文献标识码 A

包海宁, 熊杰, 张超艳, 等. 污染地块土壤砷与苯并[a]芘生物可给性影响因素研究与模型预测[J]. 环境工程学报, 2023, 17(10): 3392-3399. [BAO Haining, XIONG Jie, ZHANG Chaoyan, et al. Effect factors and model prediction of arsenic and benzo[a]pyrene bioaccessibility in the soil of contaminated sites[J]. Chinese Journal of Environmental Engineering, 2023, 17(10): 3392-3399.]

污染地块土壤砷与苯并[a]芘生物可给性影响因素研究与模型预测

包海宁^{1,2}, 熊杰², 张超艳², 郭晓欣², 赵莹², 周友亚^{2,✉}

1. 中国环境科学研究院, 北京 100012; 2. 生态环境部土壤与农业农村生态环境监管技术中心, 北京 100012

摘要 利用模型预测生物可给性比现场采样测试获取生物可给性检测结果耗时短、成本低。为探究砷(As)和苯并[a]芘(BaP)生物可给量与土壤性质之间的关系,统计了12篇文献和3份风险评估报告中As和BaP生物可给性、生物可给量和土壤性质数据,分析了生物可给量与土壤性质之间的关系,并基于逐步回归分析法和随机森林回归法构建了生物可给量的预测模型。结果表明:土壤中As的生物可给量与总量呈极显著正相关($P<0.01$),与土壤pH和CEC呈极显著负相关,BaP的生物可给量与总量呈极显著正相关,与土壤pH和粘粒质量分数呈极显著负相关性;分别采用逐步回归法和随机森林回归法构建了As和BaP生物可给量预测模型,综合比较2种模型训练集和测试集的 R^2 大小,发现随机森林回归预测模型对生物可给量的预测结果优于逐步回归预测模型,且随机森林预测模型特征重要性与相关性分析结果一致;采用随机森林回归预测模型进行案例地块验证,验证结果表明,随机森林回归预测模型对6个典型污染地块As和BaP的生物可给量预测效果较好($R^2=0.97$)。研究结果可为重金属和半挥发性有机物污染地块中生物可给性的应用提供技术支持。

关键词 污染地块; 砷; 苯并[a]芘; 生物可给性; 逐步回归; 随机森林

重金属和多环芳烃是冶炼^[1]、钢铁^[2]和焦化^[3-4]等行业污染地块的主要特征污染物,经口摄入是土壤中重金属和多环芳烃对人体健康产生风险的主要暴露途径^[5]。目前我国主要以重金属和多环芳烃的总量计算健康风险^[6],但是重金属和多环芳烃随土壤经口摄入人体后,只有部分污染物会经过一系列的过程从土壤中解吸出来进入消化系统和血液循环系统^[7-8]。因此基于土壤中重金属和多环芳烃的总量进行风险评估往往过于保守,导致修复目标值过于严格,进而造成过度修复^[9-10]。2022年12月28日,生态环境部发布了《建设用地土壤污染修复目标值制定指南(试行)》,该指南对土壤中重金属和多环芳烃等以经口摄入为主要暴露途径的污染物建议结合开展生物可给性测试推导土壤污染物的修复目标值^[11]。不少研究也基于生物可给性开展风险评估^[12-13]。但因生物可给性存在区域性差异,每个污染地块在开展风险评估时均需测定生物可给性,这一过程既耗时,又会增加测试成本。通过探究生物可给性的影响因素,构建可靠的生物可给性预测模型能够有效解决地块开展生物可给性测试导致的成本高和耗时长的问题。

重金属和多环芳烃生物可给性的影响因素包括污染物赋存形态、土壤性质和老化时间等^[14-16]。GIROUARD等^[17]对加拿大土壤中As的生物可给性进行了研究,结果表明土壤有机质及粘粒的质量分数与As的生物可给性有很好的相关性;范婧婧^[18]研究了焦化场地中PAHs生物可给性的影响,结果显示土壤粘粒质量分数与PAHs生物可给性呈负相关;黄淑婷等^[19]研究了无机盐厂中土壤粒径对Cr生物可给性的影响,结果表明粒径大的土壤中Cr的生物可给性较高。李继宁等^[20]探究了株洲市某农田土壤中Cr、Cu、

收稿日期: 2023-08-01; 录用日期: 2023-09-30

基金项目: 国家自然科学基金资助项目(41977377)

第一作者: 包海宁(1999—),女,硕士研究生,1726197768@qq.com;✉通信作者: 周友亚(1969—),女,博士,研究员, zhou_yy69@163.com

Zn、As、Cd 和 Pb 生物可给量与重金属总量、土壤 pH 和有机质等土壤性质的关系，采用逐步回归法构建了重金属生物可给量预测模型 (R^2 范围为 0.662~0.983)，但构建模型的数据仅来源于研究区域的农田土壤，预测模型具有局限性；武慧君等^[21]利用土壤重金属 (Co、Cr、Cu、Ni、Pb 和 Zn) 总量、pH 和有机质构建了某煤矿型城市小学土壤中重金属生物可给量逐步回归预测模型，预测模型 R^2 为 0.006~0.523，只能在一定程度上预测土壤中重金属生物可给量；XIE 等^[22]采用随机森林回归法，利用重金属赋存形态与土壤理化性质构建了采煤矿和冶炼厂 As、Pb 和 Cd 胃相生物可给性的预测模型，预测模型 R^2 为 0.70~0.98，表明采用随机森林法能构建可靠的生物可给性预测模型。

国内研究大多采用逐步线性回归法构建可给性预测模型^[23-24]，该法构建的模型为线性模型，对异常值比较敏感^[25]，容易造成预测模型欠拟合或过拟合，而随机森林模型容忍多重共线性且容错性高，可以避免这个问题^[26]，因此被广泛应用于预测植物、土壤和沉积物中重金属^[27-28]等方面的研究，但在预测实际污染场地土壤重金属和 PAHs 生物可给性方面应用较少。本研究以我国东北老工业基地、长三角、西南等区域典型污染地块为研究对象，探究土壤中砷 (As) 和典型多环芳烃苯并 [a] 芘 (BaP) 的生物可给量、污染物总量与土壤性质的影响规律，并采用逐步回归方法和随机森林回归方法构建生物可给量预测模型，以期为我国污染场地土壤重金属和多环芳烃风险评估工作提供技术支持。

1 材料与方法

1.1 数据获取与处理

通过全国建设用地土壤环境管理信息系统^[29]、Web of Science、中国知网等数据库收集污染地块土壤 As 和 BaP 的体外试验生物可给性数据和对应的土壤性质数据，包括 As 和 BaP 总量、生物可给量，以及土壤 pH、阳离子交换量 (CEC)、土壤有机质 (SOM)、粒径分布等土壤性质数据。初步筛选了 200 篇文献和 6 份风险评估报告，最终选生物可给性测试规范、土壤性质数据详实且研究区域为我国污染地块的 12 篇论文和 3 份风险评估报告，其中 As 污染地块有 19 个，BaP 污染地块有 7 个。对数据进行预处理，去除异常值，采用基于随机森林的多重插补的方法处理土壤性质存在缺失值问题^[30]。

1.2 生物可给量与土壤性质相关性分析

首先对 As 和 BaP 的生物可给量与总量和各土壤性质之间进行相关性分析，统计分析发现 As 和 BaP 的生物可给量与其土壤中总量和土壤理化性质等数据不完全服从正态分布，因此对其进行 Spearman 相关性分析来评估各因素之间的相关性，显著性水平取 $P < 0.05$ 和 $P < 0.01$ 。

1.3 生物可给量预测模型构建

逐步回归方法能够筛选和识别出主要影响的变量。随机森林回归方法是通过组合多个决策树，最终结果取均值，构建的模型精确度高且泛化能力强。本研究分别采用逐步回归方法和随机森林回归方法构建土壤 As 和 BaP 生物可给性预测模型。为了构建可靠的预测模型，对 As 和 BaP 生物可给量、总量以及土壤 CEC、SOM、粘粒、粉粒和砂粒等非正态分布的数据进行对数转换。通过“statsmodels”包进行构建逐步回归模型，构建随机森林模型步骤如下：①将数据集随机拆分为训练集 (75%) 和测试集 (25%)；②通过“RandomForestRegressor”包进行建模，并通过“GridSearchCV”包进行网格搜索，调整树的数量、树的最大深度等参数寻找最优参数，同时用该包进行五折交叉验证 ($cv=5$)；③根据最优参数进行训练，输入污染物总量和土壤性质数据得到预测的生物可给量。为了更好地理解随机森林回归预测模型，通过计算内置随机森林重要性和 SHAP 值来描述随机森林回归模型中特征重要性。其中 SHAP 属于模型事后解释方法，对于每个预测样本，每个特征都有 1 个 SHAP 值，它表示特征对该预测样本行为的影响^[31]。使用决定系数 (R^2) 评估 2 种预测模型的性能，选择 R^2 高的预测模型进行案例地块验证。

1.4 预测模型案例验证

为了验证预测模型在实际应用中的可靠性，结合区域地质条件的影响，选取大连化工厂、白银冶炼厂、六盘水冶炼厂、福州冶炼厂、湖南冶炼厂和北京焦化厂 6 个未用于构建预测模型的典型污染地块，对构建的 As 和 BaP 生物可给量预测模型进行验证。这 6 个地块分别位于我国东北老工业基地、长三角、西南等区域，且影响生物可给量的主要土壤性质 SOM、pH 等差异较大，具有研究的代表性。

1.5 数据处理与分析方法

使用 R 4.2.2 软件中的“mice”包进行缺失值处理, 使用 SPSS 26.0 软件进行数据统计分析和相关分析, 使用 Origin 2021 软件实现数据可视化。逐步回归模型和随机森林回归模型的构建使用 Python 3.11 软件中 Scikit-Learn 包实现。

2 结果与讨论

2.1 数据集构建

整理筛选后的 12 篇文献和 3 份风险评估报告中的数据, 如表 1 所示。由表可知, As 污染地块主要涉及冶炼厂, 土壤 As 总量范围为 1.28~15 218 mg·kg⁻¹, As 生物可给量范围为 0.56~6 088.31 mg·kg⁻¹, 其生物可给性测试方法包括 PBET、IVG、UBM 和 SBRC 等, 生物可给性范围为 0.9~89.7%。BaP 污染地块主要涉及焦化厂和钢铁厂, 土壤 BaP 的总量范围为 1.20~320.1 mg·kg⁻¹, 生物可给量范围为 0.03~9.13 mg·kg⁻¹, 生物可给性范围为 1.4%~52.79%, 测试方法主要为 DIN。土壤性质数据整体情况如图 1 所示。由图可知,

表 1 污染地块土壤性质和生物可给量

Table 1 Soil properties and bioaccessibility content of contaminated sites

污染物	污染地块	样本个数	总量/(mg·kg ⁻¹)	测试方法	生物可给量/(mg·kg ⁻¹)	数据来源	
As	太原冶炼厂	3	5.76~7.61	IVG	0.96~1.24	文献[22]	
	福州冶炼厂	2	1.28~3.53	IVG	0.56~1.76		
	六盘水冶炼厂	3	11.50~345.56	IVG	1.41~17.0		
	太原冶炼厂	3	5.26~9.79	IVG	1.89~2.73		
	韶关冶炼厂	1	275.01	IVG	40.98		
	重庆冶炼厂	6	3.50~10.04	IVG	1.14~2.82		
	湖南锌业公司	2	9.40~79.54	IVG	1.74~22.71		
	山西冶炼厂	3	10.81~75.73	IVG	2.85~23.81		
	白银冶炼厂	3	6.13~722.62	IVG	1.78~247.61		
	大庆钢铁厂	4	11.75~15.66	IVG	4.52~6.16		
	大连化工厂	1	225.80	PBET	44.26		文献[29]
	采矿冶炼厂	11	36.00~4 172.00	IVG	12.30~3 183.24		文献[32]
	广州工业区	1	26.00	PBET	6.86		文献[33]
	重庆钢铁厂	8	7.00~205.00	PBET	0.85~32.25		文献[34]
	湖南采矿冶炼厂	11	28~15 218.00	UBM	0.95~495.52		文献[35]
	石门采矿厂	20	7.40~96.60	PBET	2.96~49.81		文献[36]
	衡阳冶炼厂	1	107.48	SBRC	96.41		文献[37]
湖南冶炼厂	1	7 063.00	PBET	6 088.31	文献[38]		
某工业园区	3	5.94~10.28	IVG	4.25~7.31	文献[39]		
BaP	河北焦化厂	13	8.21~21.41	DIN	4.00~9.13	文献[40]	
	北京焦化厂	1	320.10	DIN	6.40	文献[18, 41]	
	山东钢铁厂	1	88.90	DIN	6.05		
	北京钢铁厂	1	46.60	DIN	6.89		
	大连农药厂	1	1.20	DIN	0.48		
	重庆焦化厂	1	2.10	DIN	0.03		
	昆明钢铁厂	8	3.40~15.45	DIN	0.13~2.87		文献[29]

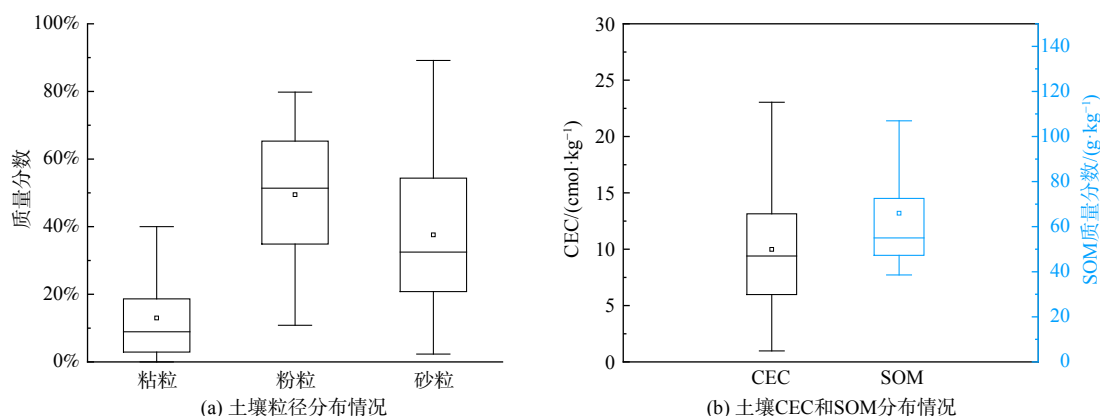


图 1 污染地块土壤性质箱线图

Fig. 1 Boxplot of soil properties data in contaminated sites

土壤性质分布差异显著，如砂粒 (2.33%~89.17%)、粉粒 (10.83%~79.80%)、粘粒 (0~53.5%)、CEC (0.98~36.18 $\text{cmol}\cdot\text{kg}^{-1}$)、SOM (1.72~343.08 $\text{g}\cdot\text{kg}^{-1}$)。土壤 pH 为 2.20~10.90。土壤性质的差异为研究 As 和 BaP 生物可给性与土壤性质之间关系的提供了基础。

2.2 生物可给量相关性分析

采用 Spearman 相关性分析方法，研究土壤中污染物可给量、污染物总量和土壤性质之间的相关性 (表 2)。由表可知，土壤中 As 和 BaP 生物可给量均与总量呈极显著的正相关性 ($P<0.01$)，与钟茂生等^[42]的研究结论相符。土壤 pH 也是影响生物可给性的重要因素之一，分析结果显示 As 和 BaP 的生物可给量均与 pH 呈极显著负相关性。CEC 与 As 的生物可给量呈极显著负相关性，这可能是由于土壤 CEC 增大更容易吸附重金属^[43]，与 CUI 等^[44]的研究结果一致。此外，粘粒质量分数与 BaP 生物可给量呈显著负相关，原因可能是土壤粘粒质量分数高，粒径小，比表面积大，更容易吸附污染物，使其生物可给性降低^[28]。众多研究表明其他土壤性质也会对生物可给性产生影响，如 BaP 生物可给性与 SOM 呈负相关性^[45-46]，SOM 对 BaP 等亲脂性的高分子量的 PAHs 吸附能力强，可以使土壤中 BaP 的生物可给性降低^[47]。

2.3 生物可给量预测模型构建

基于逐步回归方法，对 As 和 BaP 生物可给量构建预测模型如式 (1) 和 (2) 所示。由回归方程可知，土壤中 As 和 BaP 可给量主要由其总量解释。如表 3 所示，As 和 BaP 可给量训练集预测模型效果较好， R^2 分别为 0.818 和 0.917。As 测试集 R^2 为 0.802，但 BaP 测试集 R^2 仅为 0.306，这说明采用逐步回归法构建的 BaP 生物可给性预测模型对测试集预测结果较差。

$$\lg B_{\text{As}} = -2.67 + 0.89 \lg T_{\text{As}} + 0.53 \lg \omega(\text{Sand}) + 0.73 \lg \omega(\text{Silt}) + 0.04 \text{pH} - 0.15 \lg \text{CEC} \quad (1)$$

$$\lg B_{\text{BaP}} = -4.06 + 0.82 \lg T_{\text{BaP}} + 0.90 \lg \text{CEC} + 0.04 \text{pH} + 1.38 \lg \omega(\text{Sand}) \quad (2)$$

表 2 As 和 BaP 生物可给性与总量和土壤性质之间相关性

Table 2 Correlation between As and BaP bioaccessibility and total amount and soil properties

考察项目	As生物可给量	BaP生物可给量
污染物总量	0.888**	0.868**
pH	-0.295**	-0.612**
有机质	0.102	0.009
CEC	-0.277**	0.222
粘粒	0.113	-0.498**
粉粒	-0.104	0.347
砂砾	0.046	0.257

注：**、*分别代表1%、5%的显著性水平。

表 3 As 和 BaP 生物可给性逐步回归模型性能

Table 3 Stepwise regression models performance for bioaccessibility content of As and BaP

污染物	训练集 R^2	测试集 R^2
As	0.818	0.802
BaP	0.917	0.306

式中： B_{As} 表示 As 的生物可给量， $mg \cdot kg^{-1}$ ； T_{As} 表示 As 的总量， $mg \cdot kg^{-1}$ ； B_{BaP} 和 T_{BaP} 同； $\omega(Sand)$ 表示砂粒质量分数，%； $\omega(Silt)$ 表示粉粒质量分数，%。

基于随机森林回归法，以土壤性质和总量为输入特征对 As 和 BaP 可给量进行预测，采用网格搜索设置最优参数其中 As 预测模型树的数量为 86，其它参数为默认值，BaP 预测模型树的数量为 31，其它参数为默认值。预测模型如图 2 所示，由表 4 可知训练集 R^2 分别为 0.961 和 0.894，均方根误差 (RMSE) 分别为 0.158 和 0.220，测试集 R^2 分别为 0.743 和 0.793，RMSE 分别为 0.390 和 0.169，说明采用随机森林回归法构建的生物可给量模型有良好的泛化能力。通过计算内置随机森林重要性和 SHAP 值来表示随机森林回归模型中特征重要性^[48]。如图 3 所示，As 总量特征值高的样本点分布于 SHAP 值 1.0 至 1.5 区间，因此 As 的总量对预测 As 生物可给量有正向影响，与相关性分析结果一致，在构建预测模型中起到了关键作用。如图 4 所示，BaP 总量特征值高的样品点分布于 SHAP 值 1.0 至 1.5 区间，说明 BaP 的总量对预测生物可给量

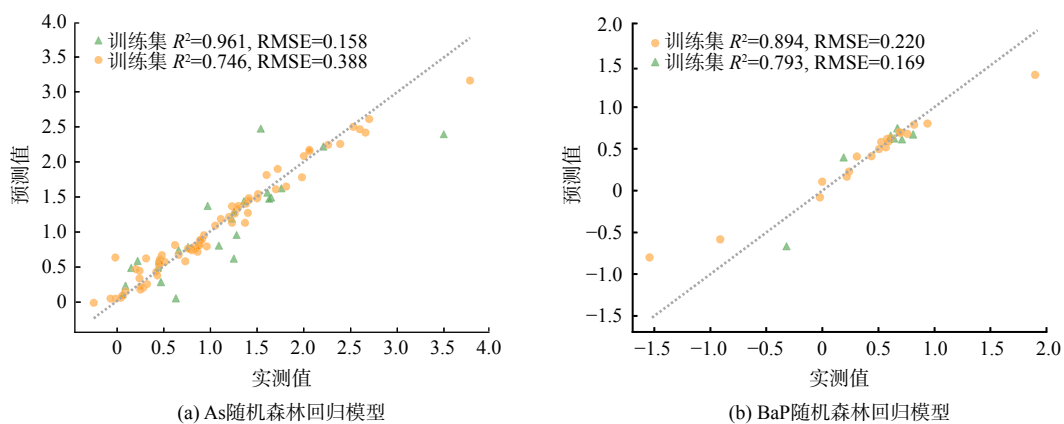


图 2 As 和 BaP 生物可给量随机森林回归模型

Fig. 2 Random forest regression models for bioaccessibility content of As and BaP

有正向影响，是随机森林回归模型中最重要的特征，pH 和粘粒质量分数特征值高的样本点分布于 -0.2 至 0 区间，因此 pH 和粘粒质量分数对 BaP 生物可给量的预测有负向影响，与相关性分析结果一致。

对 2 种方法构建的预测模型进行比较，随机森林回归方法构建的 As 和 BaP 生物可给量预测模

表 4 As 和 BaP 随机森林预测模型评价指标

Table 4 Evaluation index of random forest prediction model As and BaP

污染物	训练集 R^2	训练集 RMSE	测试集 R^2	测试集 RMSE
As	0.961	0.158	0.743	0.390
BaP	0.894	0.220	0.793	0.169

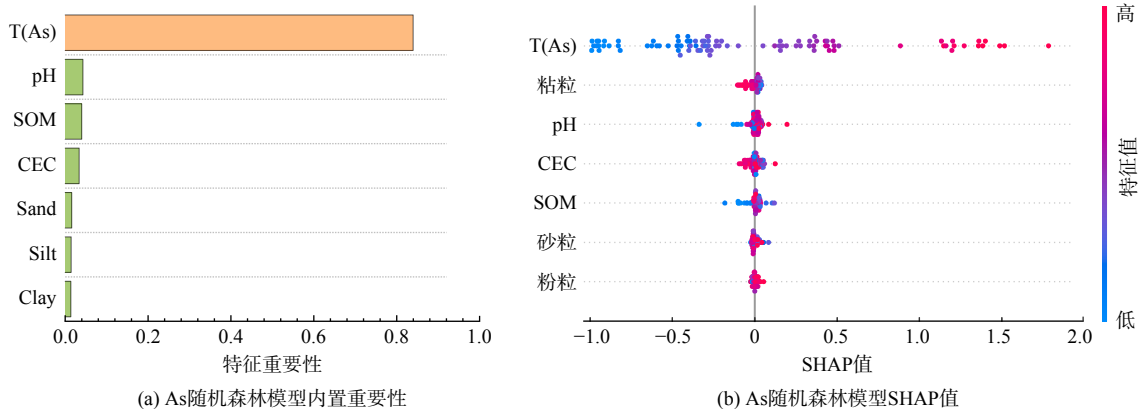


图 3 As 生物可给量随机森林特征重要性

Fig. 3 Random forest feature importance of As bioaccessibility content

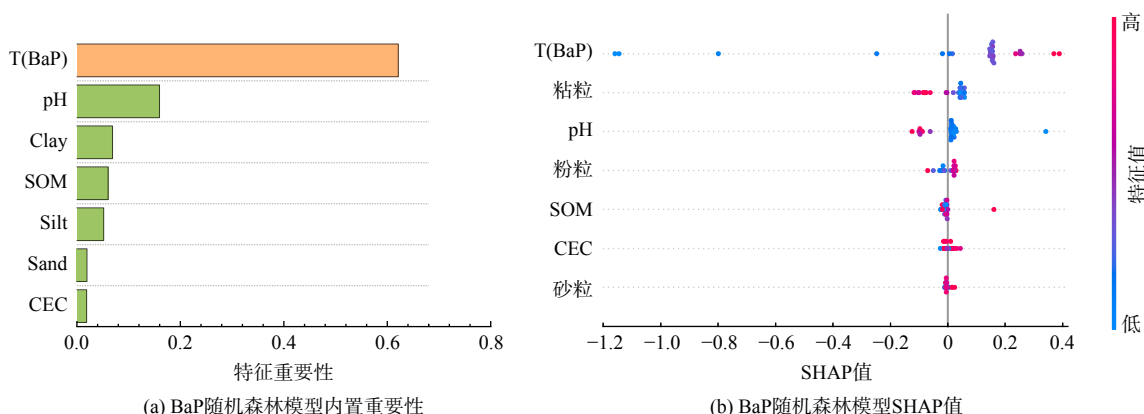


图 4 BaP 生物可给量随机森林特征重要性

Fig. 4 Random forest feature importance of BaP bioaccessibility content

型在训练集和测试集的 R^2 均值高于逐步回归方法构建的预测模型，说明随机森林回归法构建的预测模型在预测结果上更加可靠，且泛化能力好。因此用该方法构建的预测模型进行实际案例验证。

2.4 案例验证

选择了大连化工厂、白银冶炼厂、六盘水冶炼厂、福州冶炼厂、湖南冶炼厂和北京焦化厂 6 个污染地块进行随机森林预测模型的验证，案例地块土壤性质如表 5 所示。采用最优参数 (As 预测模型树的数量为 86, BaP 预测模型树的数量为 31)，预测案例地块的生物可给量，通过对比模型预测值和实测值，开展随机森林预测模型的案例验证。验证结果如图 5 所示，除六盘水冶炼厂，其他 5 个案例地块都在 95% 置信带内，造

表 5 案例地块土壤性质数据

Table 5 Soil properties data of contaminated plot

序号	案例地块	污染物	总量/(mg·kg ⁻¹)	pH	SOM/(g·kg ⁻¹)	CEC/(cmol·kg ⁻¹)	粘粒/%	粉粒/%	砂粒/%
1	大连化工厂		225.80	8.17	231.02	5.30	11.00	72.00	17.00
2	白银冶炼厂		169.36	6.23	52.58	5.20	5.62	74.08	20.30
3	六盘水冶炼厂	As	345.56	7.21	121.54	10.00	44.43	25.16	30.41
4	福州冶炼厂		1.28	9.29	1.72	4.02	3.41	60.00	36.59
5	湖南冶炼厂		280.00	7.66	16.38	9.02	37.00	57.80	5.18
6	北京焦化厂	BaP	46.60	7.71	23.20	8.10	1.95	17.06	80.99

成这一差异的原因可能为六盘水冶炼厂的生产时间较其他地块生产时间短，还可能与该地块土壤中铁质量分数较高有关。6 个案例地块都在 95% 预测带内，且拟合曲线 R^2 为 0.97，说明随机森林构建的预测模型预测效果较好，在实际地块的应用中可以采用随机森林回归模型对可给量进行预测。

3 结论

1) 分析发现土壤中 As 和 BaP 的生物可给量均与总量呈极显著正相关性 ($P < 0.01$)，与 pH 呈极显著负相关性，此外 As 的生物可给量与 CEC 呈极显著负相关性，BaP 生物可给量与粘粒质量分数呈极显著负相关性。

2) 分别采用逐步回归法和随机森林回归法构

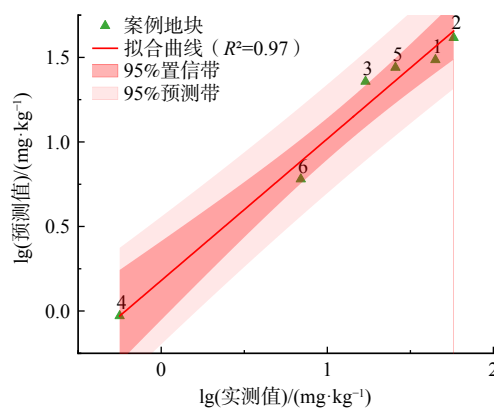


图 5 案例地块验证结果 (数字代表地块序号)

Fig. 5 Contaminated sites verification results (numbers represent serial numbers)

建 As 和 BaP 生物可给量预测模型, 比较 2 种模型训练集和测试集的 R^2 大小, 发现随机森林回归模型较逐步回归模型更可靠且泛化能力强。构建的随机森林预测模型的特征重要性与 Spearman 相关性分析结果一致。

3) 采用随机森林法对案例地块的生物可给量进行预测, 结合案例验证结果, As 和 BaP 随机森林回归预测模型对 6 个典型污染地块的预测效果较好 ($R^2=0.97$), 说明构建的预测模型适用于我国东北老工业基地、长三角、西南等区域 As 和 BaP 污染场地。

参考文献

- [1] LI C X, LI M, ZENG J Q, et al. Migration and distribution characteristics of soil heavy metal (loid) s at a lead smelting site[J]. *Journal of Environmental Sciences*, 2024, 135: 600-609.
- [2] 张施阳. 钢铁厂遗留场地土壤重金属和多环芳烃的污染特征及健康风险评估[J]. *环境污染与防治*, 2022, 44(10): 1336-1342.
- [3] CAO S Z, DUAN X L, ZHAO X G, et al. Health risks from the exposure of children to As, Se, Pb and other heavy metals near the largest coking plant in China[J]. *Science of the Total Environment*, 2014, 472: 1001-1009.
- [4] 贾晓洋, 姜林, 夏天翔, 等. 焦化厂土壤中 PAHs 的累积、垂向分布特征及来源分析[J]. *化工学报*, 2011, 62(12): 3525-3531.
- [5] CHRISTIAN G, AGNES O, ERIC W, et al. Bioaccessibility of PAH from Danish soils[J]. *Journal of Environmental Science and Health Part A*, 2007, 42(9): 1233-1239.
- [6] 中华人民共和国生态环境部. 建设用地土壤污染风险评估技术导则: HJ 25.3—2019[S]. 北京: 中国环境科学出版社, 2019.
- [7] Alexander M. Aging, bioavailability, and overestimation of risk from environmental pollutants[J]. *Environmental Science & Technology*, 2000, 34(20): 4259-4265.
- [8] EOM I C, RAST C, VEBER A M, et al. Ecotoxicity of a polycyclic aromatic hydrocarbon (PAH) contaminated soil[J]. *Ecotoxicology and Environmental Safety*, 2007, 67(2): 190-205.
- [9] 陈卓, 张丹, 吴志远, 等. 基于形态及生物可给性的汞污染场地概率风险[J]. *环境科学研究*, 2021, 34(11): 2748-2756.
- [10] 姜林, 钟茂生, 张丹, 等. 污染场地土壤多环芳烃 (PAHs) 生物可利用浓度的健康风险评估方法[J]. *生态环境学报*, 2011, 20(Z1): 1168-1175.
- [11] 中华人民共和国生态环境部. 建设用地土壤污染修复目标值制定指南 (试行) [EB/OL]. [2022-12-28]. <https://www.mee.gov.cn/xxgk/2018/xxgk/xxgk06/202212/W020221228543185584361.pdf>, 2022.
- [12] 李慧, 韩雅静, 朱晓辉, 等. 基于生物可给性的农用地土壤重金属复合污染非致癌健康风险评估[J]. *环境科学研究*, 2023, 36(4): 783-793.
- [13] 李小娟, 徐佳燕, 张剑宇, 等. 利用 in vitro 试验的城市土壤重金属健康风险研究及其对中国土壤污染调查的启示[J]. *天津大学学报 (自然科学与工程技术版)*, 2020, 53(10): 1001-1012.
- [14] 陈晓晨, 韩泽亮, 张剑宇, 等. 中国典型土壤中铅的生物可给性的影响因素分析与健康风险评估[J]. *生态环境学报*, 2021, 30(1): 165-172.
- [15] LU M J, LI G Y, YANG Y, et al. A review on in-vitro oral bioaccessibility of organic pollutants and its application in human exposure assessment[J]. *Science of the Total Environment*, 2021, 752: 142001.
- [16] YU L B, DUAN L C, NAIDU R, et al. Abiotic factors controlling bioavailability and bioaccessibility of polycyclic aromatic hydrocarbons in soil: Putting together a bigger picture[J]. *Science of the Total Environment*, 2018, 613-614: 1140-1153.
- [17] GIROUARD E, ZAGURY G J. Arsenic bioaccessibility in CCA-contaminated soils: Influence of soil properties, arsenic fractionation, and particle-size fraction[J]. *Science of the Total Environment*, 2009, 407(8): 2576-2585.
- [18] 范婧婧. 基于生物可给性的场地重金属及半挥发有机污染物健康风险评估研究[D]. 河北师范大学, 2020: 36-37.
- [19] 黄淑婷, 肖荣波, 黄飞, 等. 土壤不同粒级中 Cr(VI) 生物可给性及健康风险[J]. *生态毒理学报*, 2023, 18(1): 405-414.
- [20] 李继宁, 侯红, 魏源, 等. 株洲市农田土壤重金属生物可给性及其人体健康风险评估[J]. *环境科学研究*, 2013, 26(10): 1139-1146.
- [21] 武慧君, 方风满, 武家园, 等. 煤矿型城市小学校园表层土壤重金属生物可给性与健康风险研究[J]. *土壤通报*, 2017, 48(5): 1247-1255.
- [22] XIE K, XIE N, LIAO Z, et al. Bioaccessibility of arsenic, lead, and cadmium in contaminated mining/smelting soils: Assessment, modeling, and application for soil environment criteria derivation[J]. *Journal of Hazardous Materials*, 2023, 443 (Pt B): 130321.
- [23] 张加文, 田彪, 罗晶晶, 等. 土壤重金属生物可利用性影响因素及模型预测[J]. *环境科学*, 2022, 43(7): 3811-3824.
- [24] LUO X S, YU S, LI X D. The mobility, bioavailability, and human bioaccessibility of trace metals in urban soils of Hong Kong[J]. *Applied Geochemistry*, 2012, 27(5): 995-1004.
- [25] 高晓红, 李兴奇. 多元线性回归模型中无量纲化方法比较[J]. *统计与决策*, 2022, 38(6): 5-9.
- [26] BETHANY M, ARKO L, JAGANNATH A. Object-based random forest classification of Landsat ETM+and WorldView-2 satellite imagery for mapping lowland native grassland communities in Tasmania, Australia[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2018, 66: 46-55.
- [27] HUANG B Y, LU Q X, TANG Z X, et al. Machine learning methods to predict cadmium (Cd) concentration in rice grain and support soil management at a regional scale[J/OL]. *Fundamental Research* [2023-3-10]. <https://www.sciencedirect.com/science/article/pii/S2667325823000663>.
- [28] YANG H R, HUANG K, ZHANG K, et al. Predicting heavy metal adsorption on soil with machine learning and mapping global distribution of soil adsorption capacities[J]. *Environmental Science & Technology*, 2021, 55(20): 14316-14328.
- [29] 生态环境部. 全国建设用地土壤环境管理信息系统[EB/OL]. <http://114.251.10.109/page/shareuserlogin.html>.
- [30] 张逸飞, 曹佳. 土壤属性数据 pH 缺失的插补方法[J]. *计算机系统应用*, 2021, 30(1): 277-281.
- [31] MENG Y, YANG N H, QIAN Z L, et al. What makes an online review more helpful: An interpretation framework using XGBoost and SHAP values[J]. *Theoretical Applied Electronic Commerce*, 2021, 16: 466-490.
- [32] LI S W, LI J, LI H B, et al. Arsenic bioaccessibility in contaminated soils: Coupling in vitro assays with sequential and HNO³ extraction[J]. *Journal of Hazardous Materials*, 2015, 295: 145-152.
- [33] LU Y, YIN W, HUANG L, et al. Assessment of bioaccessibility and exposure risk of arsenic and lead in urban soils of Guangzhou City, China[J]. *Environmental Geochemistry and Health*, 2011, 33: 93-102.
- [34] 张玉, 熊杰, 唐翠梅, 等. 某污染场地土壤重金属生物可给性及其对修复目标的影响研究[J]. *中国氯碱*, 2019(6): 41-47.
- [35] ZHU X, LI M Y, CHEN X Q, et al. As, Cd, and Pb relative bioavailability in contaminated soils: Coupling mouse bioassay with UBM assay[J]. *Environment*

- [International](#), 2019, 130: 104875.
- [36] XIAO W W, LIN G B, HE X M, et al. Interactions among heavy metal bioaccessibility, soil properties and microbial community in phyto-remediated soils nearby an abandoned realgar mine [J]. [Chemosphere](#), 2022, 286: 131638.
- [37] LIAO Q, HE L X, TU G Y, et al. Simultaneous immobilization of Pb, Cd and As in soil by hybrid iron-, sulfate- and phosphate-based bio-nanocomposite: Effectiveness, longterm stability and bioavailability/bioaccessibility evaluation [J]. [Chemosphere](#), 2021, 266: 128960.
- [38] 冯康宏, 范缙, Lik Ung Stanley Hii, 等. 基于生物可给性的某冶炼厂土壤重金属健康风险评价 [J]. 中国环境科学, 2021, 41(1): 442-450.
- [39] 陈奕. 基于生物可给性分析工业场地土壤重金属污染的人体健康风险 [J]. 生态毒理学报, 2020, 15(5): 319-326.
- [40] 郭晓欣, 范婧婧, 周友亚, 等. 焦化场地典型多环芳烃类污染物精细化风险评估 [J]. 生态毒理学报, 2021, 16(1): 155-164.
- [41] 范婧婧, 周友亚, 王淑萍, 等. 基于 DIN 测试的场地土壤 PAHs 生物可给性及健康风险研究 [J]. 环境科学研究, 2020, 33(11): 2629-2638.
- [42] 钟茂生, 彭超, 姜林, 等. 老化土壤中 As 的人体可给性控制因素及健康风险 [J]. 环境科学研究, 2015, 28(2): 267-274.
- [43] ELBANA T A, SELIM H M, AKRAMI N, et al. Freundlich sorption parameters for cadmium, copper, nickel, lead, and zinc for different soils: Influence of kinetics [J]. [Geoderma](#), 2018, 324: 80-88.
- [44] CUI Y S, CHEN X C. Lead (Pb) and arsenic (As) bioaccessibility in various soils from South China [J]. [Environmental Monitoring and Assessment](#), 2011, 177(1/2/3/4): 481-492.
- [45] PORTET-KOLTALO F, GARDES T, DEBRET M, et al. Bioaccessibility of polycyclic aromatic compounds (PAHs, PCBs) and trace elements: Influencing factors and determination in a river sediment core [J]. [Journal of Hazardous Materials](#), 2020, 384: 121499.
- [46] CRAMPON M, BODILIS J, LE DERF F, et al. Alternative techniques to HPCD to evaluate the bioaccessible fraction of soil-associated PAHs and correlation to biodegradation efficiency [J]. [Journal of Hazardous Materials](#), 2016, 314: 220-229.
- [47] CELIS R, DE JONGE H, DE JONGE LW, et al. The role of mineral and organic components in phenanthrene and dibenzofuran sorption by soil [J]. [European Journal of Soil Science](#), 2006, 57: 308-19.
- [48] GU J J, YANG B, BRAUER M, et al. Enhancing the evaluation and interpretability of data-driven air quality models [J]. [Atmospheric Environment](#), 2021, 246: 118125.

(责任编辑: 金曙光)

Effect factors and model prediction of arsenic and benzo[a]pyrene bioaccessibility in the soil of contaminated sites

BAO Haining^{1,2}, XIONG Jie², ZHANG Chaoyan², GUO Xiaoxin², ZHAO Ying², ZHOU Youya^{2,*}

1. Chinese Research Academy of Environmental Sciences, Beijing 100012, China; 2. Technical Centre for Soil, Agriculture and Rural Ecology and Environment, Ministry of Ecology and Environment, Beijing 100012, China

*Corresponding author, E-mail: zhou_yy69@163.com

Abstract Compared to on-site sampling and testing bioaccessibility, constructing predictive models for bioaccessibility proves to be more cost-effective and time-efficient. To investigate the relationship between the bioaccessibility content of arsenic (As) and benzo[a]pyrene (BaP) and soil properties, this study conducted a statistical analysis of data on As and BaP bioaccessibility, bioaccessibility content, and soil properties from 12 research articles and 3 risk assessment reports. The relationship between bioaccessibility content and soil properties was analyzed. Subsequently, this study constructed predictive models for bioaccessibility content using stepwise regression analysis and random forest regression analysis. The results showed that the bioaccessibility content of As was positively correlated with the total amount, and negatively correlated with soil pH and CEC. The bioaccessibility content of BaP was positively correlated with the total amount, and negatively correlated with soil pH and clay content. Prediction models for the bioaccessibility content of As and BaP were constructed using stepwise regression analysis and random forest regression. Upon comparing the R^2 values for training and testing datasets of the two models, it was found that the random forest regression predictive model outperformed the stepwise regression predictive model. The feature importance of random forest prediction model was consistent with the correlation analysis results. Consequently, the random forest regression predictive model was used for site-specific validation, which yielded positive results. The results of cases validation showed that the random forest regression model for As and BaP had good prediction performance for six typical contaminated sites ($R^2=0.97$). The result of this study can provide technical support for the application of bioaccessibility in heavy metal and semi-volatile organic compounds contaminated sites.

Keywords contaminated site; arsenic; benzo[a]pyrene; bioaccessibility; stepwise regression; random forest