

卢合峰, 阎秀兰, 刘思言, 等. 典型城市工业集聚区土壤重金属污染精准刻画及健康风险评估[J]. 华中农业大学学报, 2023, 42(6): 185-195.
DOI: 10.13300/j.cnki.hnlkxb.2023.06.022

典型城市工业集聚区土壤重金属污染 精准刻画及健康风险评估

卢合峰¹, 阎秀兰², 刘思言^{2,3}, 苏艳超⁴, 杨潇²

1. 邢台市生态环境局邢东新区分局, 邢台 054001; 2. 中国科学院地理科学与资源研究所/陆地表层格局与模拟
院重点实验室, 北京 100101; 3. 中国科学院大学, 北京 100049; 4. 石家庄市平山环境监测中心, 石家庄 050400

摘要 为提高工业集聚区土壤污染刻画模型的精度, 以河北省某在产工业集聚区土壤重金属作为研究对象, 对比研究距离权重法(IDW)、普通克里金(OK)、支持向量机(SVM)和梯度提升决策树(GBDT)等不同插值方法在处理浓度非平稳、有偏数据的刻画精度问题。结果显示: 该场地土壤中主要污染物为As, GBDT在刻画土壤As污染时表现出最高的精度($R^2=0.9115$); GBDT可视化结果发现, As的浓度分布格局为“斑块聚集”, 并且表现出明显向深层迁移的趋势; 相关性分析结果表明, As浓度在场地土壤中的分异行为主要与土壤岩性和水文地质条件有关; 蒙特卡罗模拟风险评估结果显示, 场地土壤成人和儿童的总致癌风险指数均超过指导值, 并且儿童遭受的非致癌性和致癌性风险高于成人。

关键词 土壤污染; 空间分布; 机器学习; 梯度提升决策树; 蒙特卡罗模拟

中图分类号 X53; X820.4 **文献标识码** A **文章编号** 1000-2421(2023)06-0185-11

随着我国工业化进程的高速发展, 工业集聚区已成为我国支撑制造行业发展的主要载体, 但以产业集聚为特征的工业模式给区域环境保护带来巨大挑战^[1]。由于工业集聚区中的生产功能、工艺过程都不尽相同, 使其重金属污染特征呈现突出的复杂性和综合性, 并会通过不同迁移路径进入土壤, 威胁周围生态环境和人体健康^[2-3]。因此, 对在产工业集聚区土壤的污染特征进行系统准确的调查评估, 是全面认识城市工业集聚地的生产活动对周边环境和人类健康影响的重要组成部分。

精准刻画污染物空间分布特征是准确认识场地土壤污染状况的关键。由于工业集聚区场地具有高污染异质、多污染类型、多过程耦合的特点, 重金属在表生环境的迁移累积研究仍然存在一些困难^[4]。传统方法主要是采用地统计模型模拟污染分布, 通过反映场地的污染浓度变化, 探究重金属的污染特征和迁移过程^[5-7]。然而, 采用不同深度的密集土壤钻孔取样来描绘土壤重金属的空间分布费时且昂贵。同时为了不影响在产企业的正常生产过程以及

场地条件等限制, 实际工程中往往难以保证采样密度和空间分布的均匀性, 由此产生有偏的钻井数据限制了模型模拟结果的准确性和可靠性^[8]。Liu等^[8]利用随机森林与普通克里金(ordinary Kriging, OK)相结合的模型, 对某大型砷渣场地土壤中As的空间分布进行了精准刻画。该模型适用于场地范围较大且均匀的抽样分布, 但对于有偏钻井数据并不适用。因此, 我们试图弥补该研究方向的不足, 利用先进的机器学习算法耦合多源环境数据, 通过构建不同变量之间弱关系或非线性关系, 在不额外增加采样成本下提高模型模拟的准确性和可靠性^[9-10]。

本研究以河北省某在产城市工业集聚区土壤为研究对象, 利用有偏钻孔数据, 识别场地土壤的污染因子, 分别采用反距离权重法(inverse distance weighting, IDW)、普通克里金(OK)、支持向量机(support vector machine, SVM)和梯度提升决策树(gradient boosting decision tree, GBDT)等4种不同的空间插值方法构建非平稳耦合模型, 分析各模型刻画结果的精度, 并通过蒙特卡罗模拟方法对工业

收稿日期: 2023-10-25

基金项目: 国家自然科学基金项目(42207456; U21A2023)

卢合峰, E-mail: xthbtr@163.com

通信作者: 阎秀兰, E-mail: yanxl@igsnr.ac.cn

聚集区重金属污染土壤进行健康风险评估,旨在为场地污染土壤的精准刻画提供新思路以及为在产工业聚集区的污染诊断和科学防控提供高效的技术工具。

1 材料与方法

1.1 场地概况

研究区位于河北省无极县境内(38°13'53.30"N, 114°58'16.81"E),占地面积0.91 km²(图1A)。研究区位于磁河、沙河冲积-洪积扇区,水

文地质条件异质性较强。土壤沉积物自西向东逐渐增厚;土壤从上至下依次为杂填土层、砂土层、粉质黏土层、砂砾层(图1B)。该地区属暖温带半干旱大陆性季风气候,年平均气温8~19℃,年平均降雨量570 mm。

该工业聚集区于2006年开始建立并运营,目前现存企业21家,主要以原料化工、材料制造、机械设备、纺织印染企业为主。结合实际企业走访调查及生产工艺过程分析,主要潜在污染因子包括As、Pb、Cd、Ni、Cu和Hg。

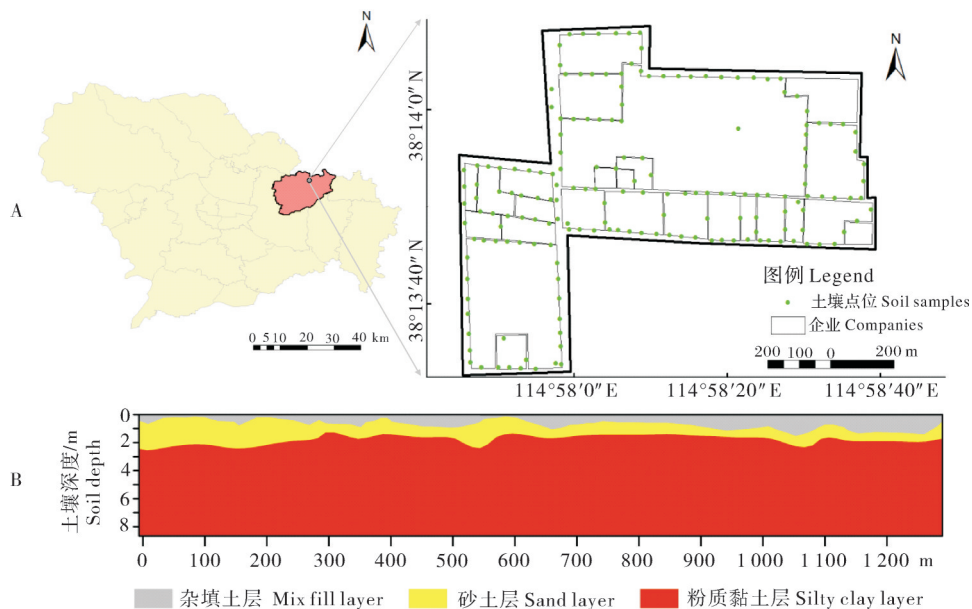


图1 研究区采样点位(A)和土壤岩性(B)

Fig.1 Locations of sample sites(A), soil layering profile(B) in the study area

1.2 样品采集及分析

为了明确工业活动对土壤重金属累积的影响,结合工业聚集区内企业分布情况以及HJ 25.1—2019《建设用地土壤污染状况调查技术导则》进行点位布设。2022年2月共布设172个土壤采样点位。结合土壤岩性和HJ/T 166—2004《土壤环境监测技术规范》要求,采集0~8 m土壤样品:杂填层采样间隔0.5 m,砂土层和粉质粘土层采样间隔2.0 m(图1B)。

对采集到的土壤样品进行风干、粉碎、筛分(孔径2 mm)和强酸消解,地下水样品采用0.45 μm滤膜过滤后,进行重金属含量分析^[11]。样品中的Cu、Ni、Pb和Cd参考HJ 803—2016《土壤和沉积物12种金属元素的测定王水提取-电感耦合等离子体质谱法》进行测定,As和Hg采用GB/T 22105.2—2008《土壤质量 总汞、总砷、总铅的测定 原子荧光法》进行测定。采用HJ 1068—2019《土壤 粒度的测定吸液管法

和比重计法》测定土壤粒度(黏粒、粉粒、砂粒、砾石)含量。采用HJ 962—2018《土壤pH值的测定 电位法》测定土壤pH值,采用HJ 613—2011《土壤 干物质和水分的测定 重量法》计算含水量。通过制定相同的空白值测定、回收实验和平行测定程序来进行质量控制和保证。

1.3 数据分析

根据HJ/T 166—2004《土壤环境监测技术规范》,采用土壤污染指数(pollution index, PI;公式中以 I_p 表示)对供试样品中的单一重金属污染现状进行评价;采用Nemerow综合污染指数(comprehensive pollution index, NPI;公式中以 I_{NP} 表示)对土壤中的多种重金属复合污染现状进行评价:

$$I_p = C_i/S_i \quad (1)$$

$$I_{NP} = \sqrt{P_{i,ave}^2 + P_{i,max}^2}/2 \quad (2)$$

式(1)~(2)中, C_i 为土壤样品*i*指标的实测结

果, mg/kg, S_i 为研究区土壤背景含量, mg/kg; $P_{i,ave}$ 为土壤所有单一污染指数的平均值, $P_{i,max}$ 为土壤所有单一污染指数的最大值。根据 PI 和 NPI 值, 土壤重金属污染程度被划分为 5 种污染等级: $PI/NPI < 1$ 为未污染; $1 \leq PI/NPI < 2$ 为轻微污染; $2 \leq PI/NPI < 3$ 为轻度污染; $3 \leq PI/NPI < 5$ 为中度污染; $PI/NPI \geq 5$ 为重度污染。

1.4 插值模型建立

反距离权重法 (IDW) 是一种加权平均方法, 以插值点与相邻采样点之间的距离为权重^[12]。IDW 假定每个测量点都有一种局部影响, 而这种影响会随着距离的增大而减小。在预测过程中, 每个样本值的权重与预测点值成正比, 这些权重值的总和为 1。具体计算过程:

$$d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} \quad (3)$$

$$w_i = \frac{1/d_i}{\sum_{i=1}^n 1/d_i} \quad (4)$$

$$Z_0 = \sum_{i=1}^n w_i Z(x_i, y_i) \quad (5)$$

式(3)~(5)中, (x_0, y_0) 为待估点的空间位置, (x_i, y_i) 为第 i 个样点的空间位置, d_i 为待估点到第 i 个样点的距离, m; w_i 为第 i 个样点的权重; Z_0 为第 i 个样点的预测值, mg/kg。

普通克里金 (OK) 模型是基于变异函数理论和结构分析, 对有限区域内的区域化变量进行无偏最优估计^[12]。OK 插值之前需要对每个变量的点对进行正态分布转换, 并且需要从数据中移除全局趋势和局部趋势^[10]。使用加权最小二乘理论数学模型拟合半变异理论模型。本研究使用对数变换对采集土壤样品的 As 含量数据进行正态分布转换。对数变换、趋势移除和 OK 插值使用 ArcGIS 10.6 中的地统计分析模块进行的。

支持向量机 (SVM) 是通过在 N 维空间中找到最佳超平面 $w^T x + b = 0$, 使特征空间上的训练数据集 T 的所有数据与该平面的距离最小。SVM 模型是通过将输入向量映射到一个新的高维特征空间 $\Phi(x)$, 使用核函数 $K(x_i, x_j)$ 替代目标函数和决策函数的内积, 构造新特征空间中的最优分离超平面。SVM 模型基于结构风险最小化理论, 使得学习器能够得到全局最优化, 使其减低了对数据规模和数据分布的要求^[9]。

$$f(x) = w^T x + b \quad (6)$$

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)\} \quad (7)$$

$$\gamma = \frac{|w^T x + b|}{\|w\|} \quad (8)$$

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (9)$$

其中, $f(x)$ 为模型输出值, w 为决策面的法向量, 决定了决策面的方向; b 决定了决策面的位置。 $x_i \in R^N$, x 为 N 维特征空间上的第 i 个特征向量, $y_i \in \{-1, +1\}$, y_i 为类标记。 γ 为样本点 x 到决策面的距离。

梯度提升决策树 (GBDT) 是一种集成机器学习算法, 通过最小化误差梯度来拟合提升决策树。GBDT 模型是将许多弱学习器 (单个决策树) 组合在一起, 得出一个强学习器^[13]。采用决策树作为弱学习器使 GBDT 具有较好的解释性, 能够自动发现特征间的高阶关系, 并且也不需要数据进行处理, 同时也适用于二元特征与连续特征同时存在的数据集^[14]。具体算法如下:

$$g_0(x) = \arg \min_c \sum_{i=1}^n L(y_i, c) \quad (10)$$

$$\gamma_{i,m} = - \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x)=f_{m-1}(x)} \quad (11)$$

$$c_{m,j} = \arg \min_c \sum_{x_i \in R_{m,j}} L(y_i, f_{m-1}(x_i) + c) \quad (12)$$

$$h_m(x_i) = \sum_{j=1}^{|R_m|} c_{m,j} I(x_i \in R_{m,j}) \quad (13)$$

$$G(x) = \sum_{i=1}^M \sum_{j=1}^J c_{m,j} I(x_i \in R_{m,j}) \quad (14)$$

式(10)~(14)中, $g_0(x)$ 为初始化弱学习器, $L(y, f(x))$ 假设损失函数。 c 的平均值可以设置为样本 y 的平均值。 $\gamma_{i,m}$ 为进行 M 次迭代拟合残差。 $R_{m,j}$ 为 m 的回归树节点面积。 $h_m(x_i)$ 为损失函数。 $G(x)$ 为梯度提升决策树函数。

机器学习模型预测中使用的环境变量有: 采样点位经纬度、高程、土壤剖面深度、土壤岩性、黏粒、粉粒、砂粒、砾石含量、土壤含水量、pH、污染物含量和功能区类型。其中, 土壤点位信息、土壤理化性质和重金属含量通过实测数据获得, 功能区类型根据现场场地调查结合企业工业类型划分。功能区类型和土壤岩性使用特征编码进行数据预处理。

本研究以均方根误差 (RMSE, 公式中以 E_{RMS} 表示)、平均绝对误差 (MAE, 公式中以 E_{MA} 表示)、平均相对误差 (MRE, 公式中以 E_{MR} 表示) 和相对分析误差 (RPD, 公式中以 R_{PD} 表示) 作为评价指标, 评价不同模型的精度^[10]。RMSE、MAE 和 MRE 值越小, RPD 值越大, 误差越小, 预测精度越高。线性回归决定系数 (R^2) 用于评估不同模型的实际值 (a_i) 和预测值 (b_i) 的拟合程度, R^2 值越大, 模型的拟合度越好。

$$E_{RMS} = \sqrt{\frac{1}{n} (\sum_{i=1}^n (a_i - b_i)^2)} \quad (15)$$

$$E_{MA} = \frac{1}{n} (\sum_{i=1}^n |a_i - b_i|) \quad (16)$$

$$E_{MR} = \frac{1}{n} \sum_{i=1}^n \left| \frac{a_i - b_i}{a_i} \right| \quad (17)$$

$$R^2 = \frac{1 - \sum_{i=1}^n (a_i - \bar{a}_i)^2}{\sum_{i=1}^n (a_i - \bar{a}_i)^2} \quad (18)$$

$$R_{PD} = \frac{1}{\sqrt{1 - R^2}} \quad (19)$$

表1 输入参数的含义及蒙特卡罗概率风险评估的取值

Table 1 Input parameters and values in health risk assessment with Monte Carlo simulator

参数 Parameter	含义 Description	单位 Unit	类型 Type	取值 Values	
				成人 Adults	儿童 Children
BW	平均体质量 Average body weight	kg	Lognormal	LN(67.48, 12.60)	LN(16.68, 1.48)
EF	暴露频率 Exposure frequency	d/a	Point	93.75	350
ED	暴露期 Exposure duration	a	Normal	N(25, 8.25)	N(6, 1.25)
IR _{oral}	每日经口摄入土壤量 Ingestion rate	mg/d	Triangular	TRI(4, 30, 52)	TRI(66, 103, 161)
InhR	每日空气呼吸量 Inhalation rate	m ³ /d	Lognormal	LN(16.30, 2.81)	LN(7.5, 1.5)
SA _s	暴露皮肤表面积 Exposed skin area	cm ²	Lognormal	LN(5 427, 579)	LN(1 592, 141)

致癌风险超过 1.00×10^{-6} 或非致癌风险超过 1.00, 该区域为健康风险超标的污染区域。致癌和非致癌风险计算公式如下:

$$H_{Q,i} = A_{DD,i} / RfD_i \quad (20)$$

$$H_1 = \sum H_{Q,i} \quad (21)$$

$$C_{R,i} = A_{DD,i} \times S_{F,i} \quad (22)$$

$$T_{CR} = \sum C_{R,i} \quad (23)$$

式(20)~(23)中, $H_{Q,i}$ 为元素 i 非致癌风险危险商指数; H_1 是总非致癌风险危险商指数, $C_{R,i}$ 为元素 i 致癌健康风险指数; T_{CR} 是总致癌风险指数; 式中其余参数的含义及取值参考 HJ25.3—2019《建设用地土壤污染风险评估技术导则》。

本研究运用蒙特卡罗模拟法的不确定性分析方法中的概率分析, 将污染物浓度、暴露受体参数、人体参数作为不确定性因素, 确定其概率分布类型, 运用蒙特卡罗方法随机模拟各敏感参数, 计算出土壤污染物对健康风险的概率分布。蒙特卡罗计算模拟进行 10 000 次迭代, 敏感性分析用于各输入参数对健康风险结果的贡献情况。

1.6 数据处理

使用 IBM SPSS Statistics 25.0 软件进行统计学分析。使用 Origin 2021 的“Correlation Plot”工具包计算 Pearson 相关系数。IDW 和 OK 插值使用 ArcGIS 10.6 中的地统计分析模块进行。本研究使用

1.5 场地健康风险评估

由于本研究区域周围有村庄以及住宅区, 环境敏感受体需同时考虑成人和儿童。土壤重金属考虑的暴露风险途径包括经口摄入、皮肤接触和吸入土壤颗粒物。本研究地块土壤暴露评估参照 HJ 25.3—2019《建设用地土壤污染风险评估技术导则》中第二类用地暴露评估模型计算土壤暴露量, 计算公式以及涉及的蒙特卡罗概率风险评估的取值参考文献 [15] 以及表 1。

MATLAB R2022a 软件结合高斯径向基函数(RBF)核函数拟合 SVM 模型。GBDT 模型是使用 Python 3.11 软件进行模型计算。蒙特卡罗模拟法中的人类概率风险使用 Crystal ball 软件进行评估。采用 Origin 2021 进行图形绘制, 使用 ArcGIS 10.6 绘制重金属含量空间分布图。

2 结果与分析

2.1 场地土壤污染特征

表 2 为场地土壤重金属含量描述性统计结果。由表 2 可知, Cu、Ni、Pb、Cd 和 As 在所有样品中均有检出, Hg 的检出率为 93.07%。与 GB 36600—2018《土壤环境质量 建设用地土壤污染风险管控标准(试行)》二类筛选值相比, As 的超标率最高, 达到 6.93%, 最大超标倍数为 17.33 倍; Pb 仅有 1 个样品出现超标, 超标倍数为 1.69 倍, 其他重金属均未出现超标现象。变异系数(CV)是用来反映指标在区域尺度内的稳定程度。Cu、Ni、Pb、Cd、As、Hg 等重金属的空间变异性分别为 176%、24%、207%、603%、401%、216%, 排序为: Cd > As > Hg > Pb > Cu > 35% > Ni, 除 Ni 外的重金属在场地中表现出较高的空间变异性(CV > 35%), 说明土壤中重金属受人类活动影响较大^[16]。

表2 场地土壤重金属含量描述性统计

mg/kg

重金属 Heavy metal	最小值 Min	最大值 Max	中位数 Median	平均值 Mean	标准差 SD	筛选值 Screening values	背景值 Background values
Cu	2.00	768.00	14.00	20.20	35.54	18 000	32.20
Ni	4.00	85.00	26.50	26.96	6.54	900	50.10
Pb	10.00	1 350.00	16.00	19.49	40.35	800	34.50
Cd	0.01	35.10	0.06	0.16	0.97	65	0.68
As	0.90	1 040.00	2.38	20.50	82.20	60	15.20
Hg	ND	0.80	0.02	0.028	0.061	38	0.10

注 Note:ND:未达到检出限 None detected.

采用单因子污染指数PI和Nemerow污染指数NPI进行土壤中污染物评价。参照土壤污染等级划分标准,土壤平均PI值依次为Hg(0.003 5)<Cd(0.008 1)<Cu(0.010)<Pb(0.049)<V(0.15)<Ni(0.18)<1<As(1.03),仅有As处于轻微污染,其余均为未污染。在所有样本中,有7.73%As处于污染水平(PI>2),其中As最高PI值可达52.00,表明本场地土壤中As污染程度最重。本场地土壤重金属NPI的范围为0.053~37.150,其中,5.23%的土壤为重度污染,1.70%的土壤为中度污染,1.70%的土壤为低度污染。图2表明,高As含量是导致本研究区NPI值异常的主要原因。

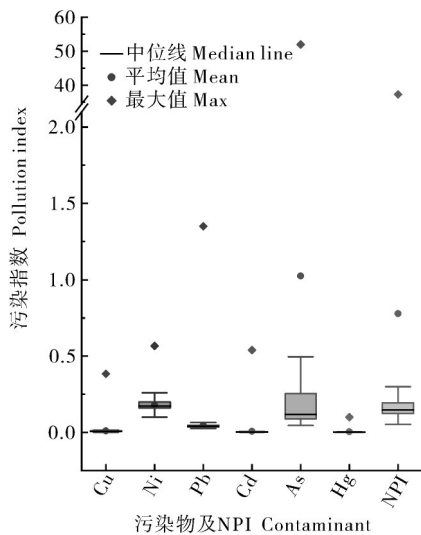


图2 土壤中污染物的污染指标

Fig.2 The pollution indexes of contaminants in soil

2.2 场地土壤重金属空间分布

为了克服土壤As浓度非平稳分布对平稳假设的影响,以及钻孔数据有偏造成的刻画结果精度低的难题,本研究在As浓度数据基础上,通过耦合水文地质参数、重金属含量等多源辅助环境数据来构建As

污染可视化模型,以期提高模型模拟场地土壤As空间分布的刻画精度。首先使用Pearson相关性分析方法,将土壤As浓度与多源辅助数据进行关联,厘清各参数对As在土壤中的迁移扩散和局部富集的重要性,从而为解耦空间非平稳关系以及有偏数据权重调整建立基础。图3表明,场地水文地质参数、土壤重金属含量均与As浓度表现出相关性。其中,As与重金属Cu、Ni、Pb、Cd和Hg以及含水量呈显著正相关。As-土壤剖面深度(-0.13)和pH(-0.17)呈显著负相关($P \leq 0.01$)。As含量还与土壤颗粒表现出显著相关性($P \leq 0.01$)。

使用4种不同的插值方法(IDW、OK、SVM和GBDT)对场地As污染进行空间刻画分析。根据RMSE、MAE和MRE的统计结果,GBDT模型的RMSE、MAE和MRE均为最低,其次是SVM、OK和IDW模型(图4),表明GBDT模型能够最为准确地反映土壤中As空间分布。4种模型的RPD值分别为3.36、2.10、1.30和1.18(表3)。根据反映预测能力的RPD值的分级标准,OK和IDW模型的RPD均小于1.4,这可能是由于传统地统计模型只能粗略估计场地土壤As污染的空间分布,而不适用于有偏数据的模拟。值得注意的是,GBDT和SVM模型的RPD>2.0,说明基于统计关系方法的机器学习模型在非平稳认知、数据纠偏等方面具备较高可靠性,表现出优异的模拟预测能力^[9]。其中,GBDT模拟结果具有最大的 R^2 值以及最小的偏差误差(表3),这表明较之其他3种模型方法,GBDT在刻画场地土壤污染时具有明显的优势。

为了更好地反映As污染的分布情况,采用GBRT对土壤As污染进行空间可视化(图5)。结果表明,As污染的空间分布格局表现为“斑块聚集”,说明土壤As分布直接受工业活动的影响^[17]。结合对

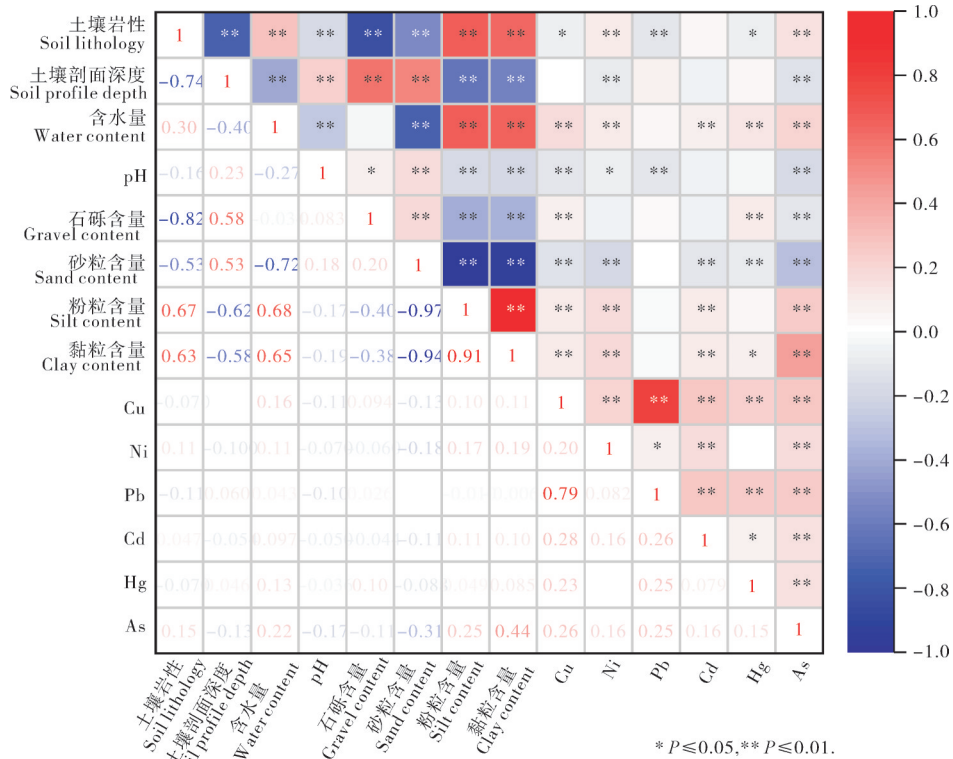


图3 重金属与土壤基本理化性质的Pearson相关性分析

Fig.3 Correlation hot map between heavy metals and soil parameters

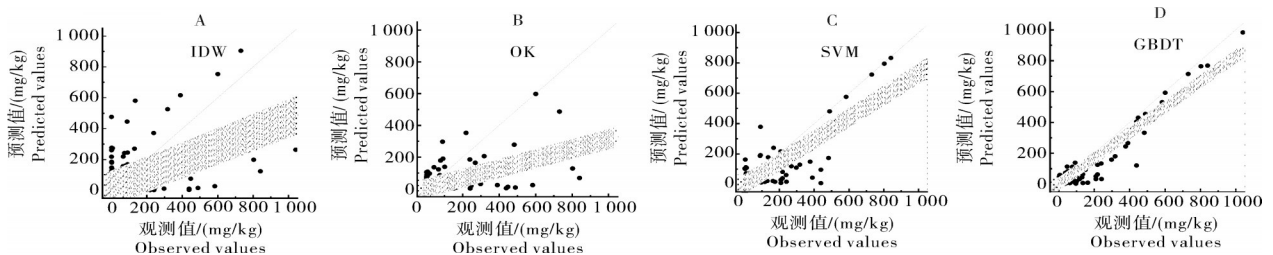


图4 不同模型的散点图拟合

Fig.4 Scatterplots of different models fitting

表3 不同模型的验证结果

Table 3 Scatterplots of different models validating results

项目 Item	IDW	OK	VSM	GBDT
R^2	0.284 7	0.412 2	0.773 7	0.911 5
RMSE	73.53	64.25	39.10	25.32
MAE	22.61	15.41	13.54	10.19
MRE	3.84	3.01	2.56	2.23
RPD	1.18	1.30	2.10	3.36

工业聚集区各企业现场调查走访以及资料收集分析可知,As污染区域主要集中于化工企业的原料堆存区、化肥和化工企业的生产加工区。在土壤垂直剖

面上,As在各深度均出现聚集现象,剖面分布特征表现为随着深度加深,As浓度先上升后下降。此外,剖面上不同深度As污染面积分别占总场地的5.56%、0.22%、3.88%、11.36%和26.06%,表明本场地的As浓度处于相对较高的水平,并且主要集中在6~8m区间。此区间土壤的垂向渗透系数为 1.91×10^{-6} cm/s,为弱透水层,具有较好的污染阻隔能力,因此,As污染集中累积在此区间。

2.3 场地土壤污染人体健康风险评估

本研究分别使用均值参数的确定性评估方法与蒙特卡罗模拟方法对研究区域进行健康风险评估。不确定性分析中通常认为风险概率分布90%~

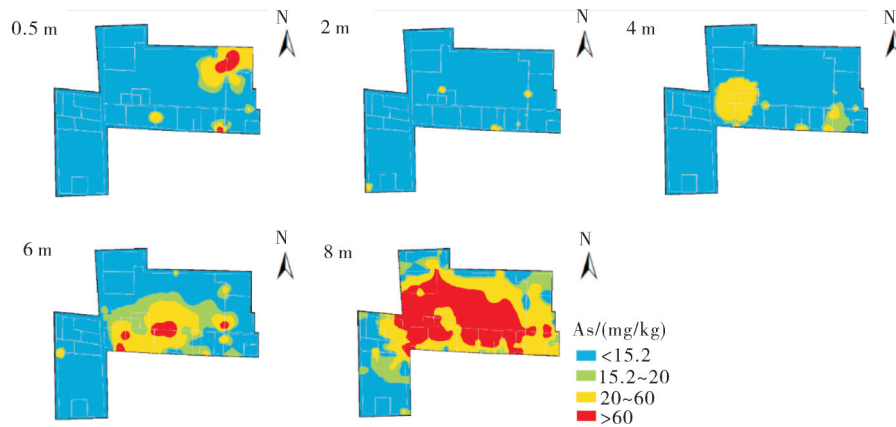


图5 使用GBRT模型刻画不同深度土壤中As的空间分布

Fig.5 Spatial distribution of As in soil at different depths using GBRT modeling

99.9%分位值为合理范围内的最大暴露,采用95%分位值作为场地最终风险值,具体评估结果见表4。蒙特卡罗模拟模型的风险概率分布图如图6A和6D所示。土壤中重金属对成人的非致癌危险商指数的平均值和95%分位值均低于风险管控值1.00,而致癌风险指数的平均值和95%分位值均大于风险管控值 1.00×10^{-6} ,表明土壤中的重金属几乎不会对成人健康造成非致癌风险,但会有较强的致癌风险。计算各重金属的致癌风险在总致癌风险的占比中发

现,使用场地污染物均值进行的确性风险评估大大低估了场地重金属对人体的健康风险^[15]。通过敏感度分析发现,影响致癌风险敏感度主要受污染物含量控制,其中As占成人总非致癌风险的97.32%(图6E)。土壤重金属对儿童的非致癌危险商在概率分布的百分位为88.30%(图6C)^[18]。使用风险概率分布95%分位值与风险管控值比较发现,本场地来自土壤的非致癌和致癌风险对于儿童均处于不可接受范围内。

表4 基于不同模型模拟的土壤污染物人体健康风险

Table 4 Human health risks of soil contaminants simulated in different models

项目 Item	成人 Adults			儿童 Children			
	确定性评估方法(均值参数) Deterministic assessment method (Mean parameter)	蒙特卡罗模拟方法(95%分位值) Monte Carlo simulation (95th percentile)	标准偏差 SD	确定性评估方法(均值参数) Deterministic assessment method (Mean parameter)	蒙特卡罗模拟方法(95%分位值) Monte Carlo simulation (95th percentile)	标准偏差 SD	
HQ	Cu	7.53×10^{-4}	2.70×10^{-3}	1.41×10^{-3}	4.07×10^{-3}	1.50×10^{-2}	7.54×10^{-3}
	Ni	7.37×10^{-3}	1.23×10^{-2}	2.54×10^{-3}	1.92×10^{-2}	3.06×10^{-2}	5.83×10^{-3}
	Pb	1.56×10^{-3}	5.78×10^{-3}	3.32×10^{-3}	3.44×10^{-2}	1.36×10^{-1}	7.90×10^{-2}
	Cd	9.27×10^{-3}	1.06×10^{-2}	6.20×10^{-2}	1.53×10^{-2}	5.77×10^{-2}	9.20×10^{-2}
	As	2.85×10^{-2}	1.14×10^{-1}	1.38×10^{-1}	4.37×10^{-1}	1.79	3.37
HI	Hg	2.86×10^{-4}	1.09×10^{-3}	6.55×10^{-4}	9.80×10^{-4}	3.76×10^{-3}	2.27×10^{-3}
	As	4.77×10^{-2}	1.37×10^{-1}	1.19×10^{-1}	5.11×10^{-1}	1.89	1.72
CR	Ni	3.34×10^{-10}	5.78×10^{-10}	1.25×10^{-10}	5.57×10^{-10}	9.22×10^{-10}	1.95×10^{-10}
	Pb	8.79×10^{-9}	3.23×10^{-8}	1.87×10^{-8}	7.71×10^{-8}	3.06×10^{-7}	1.78×10^{-7}
	Cd	2.19×10^{-7}	2.48×10^{-7}	1.46×10^{-6}	5.17×10^{-7}	1.04×10^{-6}	3.38×10^{-6}
	As	4.15×10^{-6}	1.66×10^{-5}	2.00×10^{-5}	1.53×10^{-5}	6.26×10^{-5}	1.18×10^{-4}
TCR	4.38×10^{-6}	1.69×10^{-5}	1.66×10^{-5}	1.59×10^{-5}	6.37×10^{-5}	5.99×10^{-5}	

注: Note: HQ: 非致癌风险危险商指数 Non carcinogenic risk hazard quotient index; HI: 总非致癌风险危险商指数 Total non carcinogenic risk hazard quotient index; CR: 致癌健康风险指数 Carcinogenic health risk index; TCR: 总致癌风险指数 Total carcinogenic risk index.

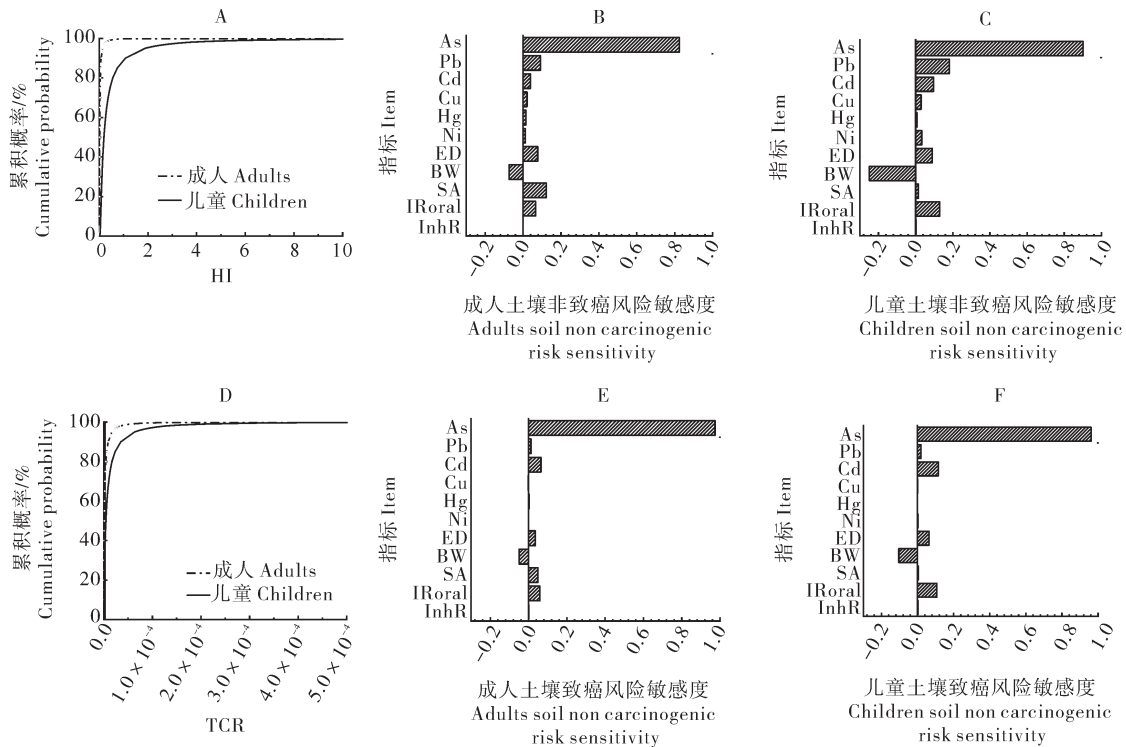


图6 土壤总非致癌和总致癌的风险概率分布以及各参数敏感度结果

Fig.6 Probability distribution for hazard index and total carcinogenic risk in soil and sensitivity results for various parameters

3 讨论

由于地下环境的复杂性,尽管部分研究改善了污染物浓度的空间非平稳性和钻井数据的偏差校正,但影响土壤污染深层迁移行为和因素的贡献仍不明确。本研究结果发现,As的深层迁移受土壤理化性质以及水文地质条件影响。Pearson相关性结果表明(图3),As在垂直方向上的空间分布可能是依靠土壤溶液进行^[19-20]。一方面,As可与土壤溶液中水分散性胶体结合,与胶体运移一起,从土壤表层随地表径流迁移,或沿着土壤孔隙和裂缝向深层土壤迁移^[21]。另一方面,受载体效应影响,土壤颗粒随着降雨和径流的移动会加速土壤中As的垂向迁移^[17]。其中,砂粒和黏粒2个因素与土壤As浓度之间显著相关,分别为-0.31和0.44,这可能是由于砂粒较容易被土壤水携带,导致As发生深层迁移^[22]。粉质黏土具有多孔结构,可提供大量吸附点位从而抑制As的迁移^[23]。一般来说,场地土壤pH值越低,As的迁移性越高。这是由于土壤溶液中的As主要以无机砷氧阴离子的形式存在^[24],高pH值下会促进As与黏土矿物或土壤有机物形成复合物,减弱As的迁移能力^[25]。然而,本研究场地土壤随着深度加深,pH值

逐渐降低。因此,深层环境可能更有利于As的迁移和扩散。

根据模拟结果可知,不同的插值方法得到的刻画精度差异显著。OK和IDW难以建立污染物含量与环境因素之间的空间相关性,忽略了污染物迁移扩散和水文地质条件带来的非平稳问题。因此,计算得到的精度很低^[14]。针对有偏钻孔数据直接采用传统的地统计插值方法得到的估计结果会存在污染物含量高估或低估的现象,无法准确反映污染水平。而基于统计关系的机器学习算法,由于考虑了场地土壤污染物的空间非平稳特征,通过引入场地水文地质条件环境因素(土壤pH、含水量、土壤岩性、土壤剖面深度、石砾含量、砂粒含量、粉粒含量和黏粒含量)以及工业生产活动因素(功能区分布),显示出对非平稳、有偏数据优异的适用性,从而显著提升了空间刻画的精度。与传统地统计模型不同,SVM是通过最小化预测结果与实际结果之间的差异,实现回归任务。因此,在处理小样本时具有较好的泛化性能,对在产工业聚集区的污染空间刻画有较好的适用性^[9]。但由于SVM中核函数选择是根据经验选取,无法结合实际数据情况构造最佳的SVM算法,

因此,在模型预测过程中存在一定的局限性^[26]。然而,GBRT模型具有良好的预测污染物含量和环境影响因素之间非线性关系的能力,可实现有偏数据的平稳态转化。因此,对本研究场地土壤的重金属污染刻画表现出最优的准确度和适应性^[27]。综上所述可知,基于GBRT的集成模型与其他算法相比表现出更好的拟合性能,可以作为同类型有偏污染数据模拟的推荐方法,从而在不额外增加采样工作的前提下,提高污染物的空间分布特征模拟和迁移过程预测的精度。

关于重金属污染土壤环境风险评估方面,目前国内外普遍使用的导则都是基于输入参数的确定性值,易受数据波动及其他不确定性参数的影响,造成场地健康风险评价结果存在极大的差异。因此,使用均值参数的确定性风险评估方法得到健康风险评价价值往往难以反映区域的整体情况。然而,蒙特卡罗模拟可通过使用随机参数值进行迭代计算,降低了由参数选取、数据波动带来的不确定性,并通过参数因子敏感度判断健康风险的主控因素,对人体健康风险进行更准确的评估^[28]。对比研究区域成人和儿童的健康风险发现,使用场地污染物均值进行的确定性风险评估大大低估了场地对儿童的健康风险,并且儿童遭受的非致癌性和致癌性风险比成人更大。本研究中,通过敏感度结果分析,除了重金属含量影响较大外,受体对象的体质量对健康风险不确定性的贡献率出现负值,说明较小体型的受体对象会面临较大的健康风险。儿童的平均体质量(BW)贡献率的绝对值显著高于成人。此外,对儿童健康风险而言,每日经口摄入量(IR_{oral})的贡献率高于成人。儿童特有的生理和行为模式,如高呼吸频率和吮吸手指等行为,也是产生健康风险的主要因素^[29]。这同样解释了尽管本场地Cd含量没有超标,但Cd在土壤CR值的95%分位值大于 1.00×10^{-6} ,其主要原因是Cd污染主要是由经口摄入途径进入人体,不确定输入参数IR_{oral}对致癌健康风险的贡献率较高,为13.01%。因此,本研究使用蒙特卡罗模拟法,对变量及结果的不确定性进行量化,不但能计算出风险值的范围和概率水平,而且能有效降低由参数选取、数据波动带来的不确定性,并通过参数因子敏感度判断健康风险的主控因素,对人体健康风险进行更精确的评估^[18]。

参考文献 References

- [1] SU Y, YU Y Q. Spatial agglomeration of new energy industries on the performance of regional pollution control through spatial econometric analysis[J/OL]. *Science of the total environment*, 2020, 704: 135261 [2023-10-25]. <https://doi.org/10.1016/j.scitotenv.2019.135261>.
- [2] PODGORSKI J, BERG M. Global threat of arsenic in groundwater[J]. *Science*, 2020, 368(6493): 845-850.
- [3] ZHU L, ZHANG X X, ZHANG J C, et al. Saltwater intrusion weakens Fe-(oxyhydr)oxide-mediated (im)mobilization of Ni and Zn in redox-fluctuating soil-groundwater system[J/OL]. *Water research*, 2022, 221: 118799 [2023-10-25]. <https://doi.org/10.1016/j.watres.2022.118799>.
- [4] HOU S Y, ZHAO X, LIU Y, et al. Spatial analysis connects excess water pollution discharge, industrial production, and consumption at the sectoral level[J/OL]. *NPJ clean water*, 2022, 5: 4 [2023-10-25]. <https://doi.org/10.1038/s41545-022-00152-7>.
- [5] RAN H Z, GUO Z H, YI L W, et al. Pollution characteristics and source identification of soil metal(loid)s at an abandoned arsenic-containing mine, China[J/OL]. *Journal of hazardous materials*, 2021, 413: 125382 [2023-10-25]. <https://doi.org/10.1016/j.jhazmat.2021.125382>.
- [6] ZENG J Q, LI C X, WANG J T, et al. Pollution simulation and remediation strategy of a zinc smelting site based on multi-source information[J/OL]. *Journal of hazardous materials*, 2022, 433: 128774 [2023-10-25]. <https://doi.org/10.1016/j.jhazmat.2022.128774>.
- [7] 范俊楠, 贺小敏, 杜维, 等. 基于标准化方法和统计分析法的土壤重金属环境基线值分析比较[J]. *华中农业大学学报*, 2021, 40(1): 160-167. FAN J L, HE X M, DU W, et al. Analyzing and comparing environmental baseline values of heavy metals in soil based on standardized method and statistical method [J]. *Journal of Huazhong Agricultural University*, 2021, 40 (1): 160-167 (in Chinese with English abstract).
- [8] 陶欢, 廖晓勇, 曹红英, 等. 场地土壤污染物含量三维刻画的研究进展[J]. *地理学报*, 2022, 77(3): 559-573. TAO H, LIAO X Y, CAO H Y, et al. Research progress of three-dimensional delineation of soil pollutants at contaminated sites [J]. *Acta geographica sinica*, 2022, 77(3): 559-573 (in Chinese with English abstract).
- [9] LIU G, ZHOU X, LI Q, et al. Spatial distribution prediction of soil As in a large-scale arsenic slag contaminated site based on an integrated model and multi-source environmental data[J/OL]. *Environmental pollution*, 2020, 267: 115631 [2023-10-25]. <https://doi.org/10.1016/j.envpol.2020.115631>.
- [10] GUO Z H, ZHANG Y X, XU R, et al. Contamination vertical distribution and key factors identification of metal(loid)s in

- site soil from an abandoned Pb/Zn smelter using machine learning[J/OL]. *Science of the total environment*, 2023, 856: 159264 [2023-10-25]. <https://doi.org/10.1016/j.scitotenv.2022.159264>.
- [11] ZHAI Y Z, JIANG Y, CAO X Y, et al. Valuation of ecosystem damage induced by soil-groundwater pollution in an arid climate area: framework, method and case study[J/OL]. *Environmental research*, 2022, 211: 113013 [2023-10-25]. <https://doi.org/10.1016/j.envres.2022.113013>.
- [12] QIAO P W, LI P Z, CHENG Y J, et al. Comparison of common spatial interpolation methods for analyzing pollutant spatial distributions at contaminated sites[J]. *Environmental geochemistry and health*, 2019, 41(6): 2709-2730.
- [13] ZHAO B, ZHU W X, HAO S F, et al. Prediction heavy metals accumulation risk in rice using machine learning and mapping pollution risk[J/OL]. *Journal of hazardous materials*, 2023, 448: 130879 [2023-10-25]. <https://doi.org/10.1016/j.jhazmat.2023.130879>.
- [14] PENG Y X, CHEN J, XIE E Z, et al. Three-dimensional spatial prediction of Zn in the soil of a former tire manufacturing plant using machine learning and readily attainable multisource auxiliary data[J/OL]. *Environmental pollution*, 2023, 318: 120931 [2023-10-25]. <https://doi.org/10.1016/j.envpol.2023.120931>.
- [15] ZHOU Y, JIANG D D, DING D, et al. Ecological-health risks assessment and source apportionment of heavy metals in agricultural soils around a super-sized lead-zinc smelter with a long production history, in China[J/OL]. *Environmental pollution*, 2022, 307: 119487 [2023-10-25]. <https://doi.org/10.1016/j.envpol.2022.119487>.
- [16] SUN Y, LIANG X J, XIAO C L. Assessing the influence of land use on groundwater pollution based on coefficient of variation weight method: a case study of Shuangliao City[J]. *Environmental science and pollution research international*, 2019, 26(34): 34964-34976.
- [17] ZHANG H, YIN A J, YANG X H, et al. Use of machine-learning and receptor models for prediction and source apportionment of heavy metals in coastal reclaimed soils[J/OL]. *Ecological indicators*, 2021, 122: 107233 [2023-10-25]. <https://doi.org/10.1016/j.ecolind.2020.107233>.
- [18] HUANG B, YUAN Z J, LI D Q, et al. Effects of soil particle size on the adsorption, distribution, and migration behaviors of heavy metal (loid)s in soil: a review[J]. *Environmental science processes & impacts*, 2020, 22(8): 1596-1615.
- [19] CHEN R H, TENG Y G, CHEN H Y, et al. A coupled optimization of groundwater remediation alternatives screening under health risk assessment: an application to a petroleum-contaminated site in a typical cold industrial region in Northeastern China[J/OL]. *Journal of hazardous materials*, 2021, 407: 124796 [2023-10-25]. <https://doi.org/10.1016/j.jhazmat.2021.124796>.
- [20] GUO T, ZHOU Y J, CHEN S C, et al. The influence of periphyton on the migration and transformation of arsenic in the paddy soil: rules and mechanisms[J/OL]. *Environmental pollution*, 2020, 263: 114624 [2023-10-25]. <https://doi.org/10.1016/j.envpol.2020.114624>.
- [21] ZONG Y T, XIAO Q, LU S G. Distribution, bioavailability, and leachability of heavy metals in soil particle size fractions of urban soils (northeastern China)[J]. *Environmental science and pollution research international*, 2016, 23(14): 14600-14607.
- [22] O'CONNOR D, PAN S Z, SHEN Z T, et al. Microplastics undergo accelerated vertical migration in sand soil due to small size and wet-dry cycles[J]. *Environmental pollution*, 2019, 249: 527-534.
- [23] WANG J X, ZHANG W J. Evaluating the adsorption of Shanghai silty clay to Cd(II), Pb(II), As(V), and Cr(VI): kinetic, equilibrium, and thermodynamic studies[J/OL]. *Environmental monitoring and assessment*, 2021, 193(3): 131 [2023-10-25]. <https://doi.org/10.1007/s10661-021-08904-7>.
- [24] XUE Q, RAN Y, TAN Y Z, et al. Arsenite and arsenate binding to ferrihydrite organo-mineral coprecipitate: implications for arsenic mobility and fate in natural environments[J]. *Chemosphere*, 2019, 224: 103-110.
- [25] YANG X, LIU S Y, LIANG T, et al. Ball-milled magnetite for efficient arsenic decontamination: insights into oxidation-adsorption mechanism[J/OL]. *Journal of hazardous materials*, 2022, 427: 128117 [2023-10-25]. <https://doi.org/10.1016/j.jhazmat.2022.128117>.
- [26] GAYE B, ZHANG D Z, WULAMU A. Improvement of support vector machine algorithm in big data background[J/OL]. *Mathematical problems in engineering*, 2021, 2021: 1-9 [2023-10-25]. <https://doi.org/10.1155/2021/5594899>.
- [27] ZHENG J T, WANG P, SHI H Y, et al. Quantitative source apportionment and driver identification of soil heavy metals using advanced machine learning techniques[J/OL]. *Science of the total environment*, 2023, 873: 162371 [2023-10-25]. <https://doi.org/10.1016/j.scitotenv.2023.162371>.
- [28] MUKHERJEE I, SINGH U K, SINGH R P, et al. Characterization of heavy metal pollution in an anthropogenically and geologically influenced semi-arid region of East India and assessment of ecological and human health risks[J/OL]. *Science of the total environment*, 2020, 705: 135801 [2023-10-25]. <https://doi.org/10.1016/j.scitotenv.2020.135801>.
- [29] CHEN H X, XU F N, CHEN Z Z, et al. Arsenic and cadmium removal from water by a calcium-modified and starch-stabilized ferromanganese binary oxide[J]. *Journal of environmental sciences*, 2020, 96: 186-193.

Accurate characterization and health risk assessment of heavy metal pollution in soil of typical urban industrial agglomeration areas

LU Hefeng¹, YAN Xiulan², LIU Siyan^{2,3}, SU Yanchao⁴, YANG Xiao²

1. *Xingdong New Area Branch, Xingtai Ecological and Environmental Bureau, Xingtai 054001, China;*

2. *Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences/Key Laboratory of Land Surface Patterns and Simulation, Beijing 100101, China;*

3. *University of Chinese Academy of Sciences, Beijing 100049, China;*

4. *Shijiazhuang Pingshan Environment Monitoring Center, Shijiazhuang 050400, China*

Abstract The heavy metals in the soil of an industrial agglomeration areas in Hebei Province were used to conduct a comparative study on the problem of characterization accuracy in handling non-stationary concentration and biased data with different interpolation methods including distance weighting (IDW), ordinary Kriging (OK), support vector machine (SVM), and gradient enhanced decision tree (GBDT) to improve the accuracy of models for characterizing the soil pollution in industrial agglomeration areas. The results showed that the main pollutant in the soil of this site was arsenic, and GBDT exhibited the highest accuracy in characterizing arsenic pollution in soil ($R^2=0.9115$). The results of GBDT visualization showed that the concentration distribution pattern of arsenic was "patchy aggregation" and had a good vertical migration capacity. The results of correlation analysis showed that the differentiation behavior of Arsenic concentration in the soil of this site was mainly related to the soil lithology and hydrogeological conditions. The results of Monte Carlo-based simulation showed that the total cancer risk index of both adults and children in the soil of this site exceeded the guidance value, and children suffered from higher non carcinogenic and carcinogenic risks than adults.

Keywords soil pollution; spatial distribution; machine learning; gradient boosting decision tree; Monte Carlo-based simulation

(责任编辑:陆文昌)