第37卷,第1期	Ē	自然	紫 资	源	遥	感
2025年02月	REMOTE S	ENSIN	IG FOR	NATU	RAL	RESOURCES

## doi: 10.6046/zrzyyg.2023229

引用格式:钟晴,麦麦提吐尔逊·艾则孜,米热古力·艾尼瓦尔,等.绿洲城市土壤砷含量高光谱估算[J].自然资源遥感, 2025,37(1): 188-194. (Zhong Q, Mamattursun E, Mireguli A, et al. Hyperspectral inversion of arsenic content in soil in an oasis city[J]. Remote Sensing for Natural Resources, 2025,37(1): 188-194.)

# 绿洲城市土壤砷含量高光谱估算

钟晴<sup>1</sup>,麦麦提吐尔逊·艾则孜<sup>1,2</sup>,米热古力·艾尼瓦尔<sup>1</sup>,郝海宇<sup>3</sup> (1.新疆师范大学地理科学与旅游学院,乌鲁木齐 830054; 2.新疆师范大学新疆干旱区湖泊环境与资源 实验室,乌鲁木齐 830054;3.新疆师范大学物理与电子工程学院,乌鲁木齐 830054)

摘要: 砷(As)是具有强致癌性的类金属元素,快速、准确地监测土壤中As 元素含量尤为重要。首先,以乌鲁木齐 市表层土壤为研究对象,采集 84 组土壤样品,并测定其 As 含量和原始光谱反射率,用 Pearson 相关分析对土壤原 始光谱及12种光谱变换下的光谱反射率与土壤As含量之间的关系进行检验,筛选出特征波段;然后,基于偏最小 二乘回归(partial least squares regression, PLSR)、随机森林回归(random forest regression, RFR)以及支持向量机回归 (support vector machine regression, SVMR),构建As含量高光谱反演模型;最后,选取决定系数R<sup>2</sup>、均方根误差(root mean square error, RMSE)和平均绝对误差(mean absolute error, MAE)来评估高光谱模型的反演预测能力。结果表 明:对原始光谱数据进行微分变换能够有效增强光谱特征,提高土壤光谱反射率与 As 含量之间的相关性。3 种模 型的反演预测能力由高到低依次为: RFR>SVMR>PLSR,其中,基于均方根二阶微分的 RFR 模型 R<sup>2</sup>为 0.821, RMSE 为 0.143 mg/kg, MAE 为 0.523 mg/kg, 模型拟合效果最好, 具有较高的稳定性和预测精度。研究可为构建绿洲城市 土壤 As 含量高光谱反演模型提供科学依据。

关键词:城市土壤;砷;高光谱反演;光谱变换;反演模型

中图法分类号: TP 79; X 833 文献标志码: A

文章编号: 2097-034X(2025)01-0188-07

#### 引言 0

城市土壤是生态环境的重要组成部分,城市土 壤环境健康关系到生态系统安全以及居民身体健 康<sup>[1]</sup>。砷(As)元素是一种类金属元素,在土壤环境 中难降解、可迁移转化,多以有毒化合物形式存在, 生物毒性强,具有强致癌性<sup>[2-4]</sup>。因此,快速、准确 地监测土壤中 As 元素含量尤为重要。传统的土壤 As 含量监测方法效率低,耗时耗力且监测范围有 限,所用试剂也会对土壤造成二次污染[5-6]。高光 谱技术具有光谱分辨率高、成本低、效率高、用时短 等特点,可以快速无损地对土壤中多种元素进行分 类和定量分析,进行大范围的原位监测,避免复杂的 采样步骤,对环境监测和进一步的决策支持具有重 要意义[7-8]。

部分学者利用高光谱遥感技术,构建了特定区 域土壤中As元素含量高光谱估算模型,并取得了一 定的成果。杨君怡等<sup>[9]</sup>利用偏最小二乘回归(par-

tial least squares regression, PLSR)方法建立了肇庆 市土壤 As 含量高光谱反演模型,发现土壤 As 含量 的最佳估测模型是基于反射率倒数对数的一阶微分 的 PLSR 模型;郑光辉等<sup>[10]</sup>发现土壤中有机质会掩 盖部分特征信息,有机质含量较低时,苏南某市的土 壤 As 含量与光谱数据所构建的高光谱估算效果更 好;李志远等[11]利用多元线性逐步回归、单光谱变 换指标 PLSR 和多光谱变换指标偏最小二乘回归 (M-PLSR)方法构建石家庄市土壤 As 含量估算模 型,发现采取 M-PLSR 方法土壤 As 含量模型估算 值与实测值最为接近:王泽强等[12]将随机森林 (random forest, RF)模型与"时间-空间-光谱"3个 组合特征进行结合,对山东省济南市长清区土壤 As 含量进行了空间预测,并得到了最佳的空间预测精 度,其决定系数  $R^2$ 为 0.90,均方根误差(root mean square error, RMSE)为 0.77 mg/kg, 四分位数间距 (ratio of performance to inter-quartile range, RPIQ)为 5.68; 史广等<sup>[13]</sup>基于多源环境数据和 RF、反距离权 重插值(inverse distance weighted, IDW)、逐步线性

基金项目:国家自然科学基金项目"绿洲地下水重金属污染风险防控理论与技术研究"(编号:U2003301)资助。

收稿日期: 2023-07-24;修订日期: 2023-11-26

第一作者:钟晴(1998-),女,硕士研究生,主要从事绿洲土壤环境安全研究。Email: 13235366308@163.com。

通信作者: 麦麦提吐尔逊・艾则孜(1981-), 男, 博士, 教授, 主要从事绿洲生态环境演变研究。Email: oasiseco@126.com。

回归模型,对中国西南部农田土壤中 As 的空间分布 进行模拟,发现 RF 模型的模拟结果比较符合研究 区的实际污染状况; 袁自然等<sup>[14]</sup>提出一种特征提 高型竞争性自适应重加权算法(competitive adaptive reweighted sampling,CARS)选取特征波段的粒子群 算法(particle swarm optimization,PSO)优化支持向 量机(support vector machine,SVM)的农田土壤 As 含量高光谱估测分析方法,有效提高了洪湖市农田 土壤 As 含量高光谱估算精度。此外,易兴松等<sup>[15]</sup> 利用 PLSR 模型对贵州省贵定县土壤 As 含量进行 高光谱估算,发现基于一阶微分的 400~2 400 nm 的 便携式地物光谱仪 ASD 获取的光谱数据结合 PLSR 模型的预测精度最高,其  $R^2$ 值为 0.69, *RMSE* 为 0.149 mg/kg,相对分析误差(relative percent deviation, RPD)为 1.36。

以上研究表明,土壤 As 含量高光谱估算模型的 预测精度会受到区域位置、土壤类型以及污染水平 等因素的影响,且As在土壤中难降解、可迁移转化, 过度累积易造成土壤污染及潜在健康、生态风险问 题。麦麦提吐尔逊·艾则孜等<sup>[16]</sup>经过研究发现干 旱区绿洲土壤对部分重金属的吸附能力很大,因此, 有必要研究绿洲城市土壤 As 含量高光谱估算方法, 以期对土壤 As 含量进行快速精确的实时动态监测。 基于此,以新疆乌鲁木齐市土壤 As 含量作为研究对 象,对土壤原始光谱数据进行微分变换和 Pearson 相关性分析智能选取特征波段,构建随机森林回归 (random forest regression, RFR), PLSR 和支持向量 机回归(support vector machine regression, SVMR)3 种模型,进行土壤 As 含量的高光谱反演预测,并对 比以上3种模型的高光谱反演预测结果,确定适合 研究区实际情况的土壤 As 含量的最佳预测模型,以 期为快速、精确地实现土壤 As 含量预测提供理论参 考与更具适用性的技术手段。

1 研究区概况及数据源

### 1.1 研究区概况

乌鲁木齐市人口众多,工业发达,地处亚欧大陆 腹地、天山山脉中段北麓、准噶尔盆地南缘,是新疆 首府和西北地区第二大城市,也是丝绸之路经济带 上的核心城市之一,在"一带一路"倡议中占据重要 地位。气候属温带大陆性干旱气候,多年平均气温 为 6.7 °C,多年平均降水量为 280 mm,多年平均蒸 发量为 2 730 mm,土壤类型主要为灰漠土。

#### 1.2 土样采集与测定

2021年4月在新疆乌鲁木齐市采集了84个表

层(0~20 cm)土壤样品(如图 1 所示)。在每个样 点,100 m×100 m内布设9个子样点,用非金属采样 工具采集约1 kg 土样,密封保存。所有土样室内自 然风干,剔除杂物后研磨,过20 目尼龙筛后分成2 份,一份用于 As 含量测定,一份用于光谱测量。土 样经过四酸法消解后,利用原子荧光光度计(AFS-933)测定了 As 元素的含量,检测依据为 GB/T 22105.2—2008。As 元素含量测定过程中,利用国家 土壤标准参比物质(GSS-12,北疆土壤)、平行样和 重复样进行质量控制,测试回收率为 93.66%,在允 许范围内。





#### 1.3 光谱数据测定与处理

利用便携式地物光谱仪(ASD FieldSpec 3),在 室外空旷且干扰因素少的区域测定土样光谱数据, 光谱波长范围为 350~2 500 nm。测定光谱数据时, 在 2 m×2 m 的黑色卡纸上放 40 cm×40 cm 的白板 进行定标,以获取绝对反射率。测定时光谱仪传感 器探头垂直于土壤表面。每个土壤样品的光谱数据 测定 15 次,传感器探头每隔 10 min 对准白板进行 优化。利用 ViewSpecPro 软件对测定的 15 条光谱 曲线进行均值化处理,将其算术平均值作为土样原 始光谱反射率。对均值处理后的所有光谱曲线统一 去除 350~399 nm,1 350~1 430 nm,1 781~1 970 nm 和2 401~2 500 nm 噪声较大的异常波段,共输出 1730个波段用于后续建模研究。采用 Savitzky-Golay法对土样原始光谱进行平滑处理,去除"毛刺" 噪声。不同土样的光谱曲线如图 2 所示,不同样本 的光谱曲线形态较为相似,平滑后的光谱曲线在 1 780 nm 和 1 970 nm 波长附近变化明显,数值略有 提高。



Fig.2 Original and Savitzky-Golay smoothing spectral reflectance curves

#### 研究方法 2

### 2.1 光谱变换与特征波段选取

光谱测定中存在随机误差,且As在土壤中含量 低,光谱响应信号微弱[17],采用原始数据光谱直接 反映特征波段比较困难,需根据As的光谱吸收特征 筛选特征波段。因此,对平滑处理后的原始光谱数 据分别进行一阶微分(FD)、二阶微分(SD)、倒数一 阶微分(RTFD)、倒数二阶微分(RTSD)、对数一阶 微分(LTFD)、对数二阶微分(LTSD)、均方根一阶微 分(RMSFD)、均方根二阶微分(RMSSD)、倒数的对 数一阶微分(ATFD)、倒数的对数二阶微分 (ATSD)、对数的倒数一阶微分(RLFD)、对数的倒 数二阶微分(RLSD)等光谱变换。对土壤样品 As 的实测值与原始光谱数据和 12 种光谱变换后的数 据进行 Pearson 相关性检验,筛选出相关系数较强 的光谱特征波段。平滑处理、光谱变换及相关性分 析均在 MATLAB(R2019b)软件中完成。

### 2.2 模型建立与检验

验证集

20

采用 RFR, PLSR 和 SVMR 这 3 种回归方法进 行反演建模。RFR 模型的核心算法是利用多棵决 策树对样本进行训练并预测的机器学习算法,集合 众多决策树,采用 Bootsrap 重抽样方法随机抽样构 建不同的分类模型并组成多分类模型系统,泛化能

力高<sup>[18]</sup>。PLSR 模型是 1993 年提出的双线性模型. 综合了主成分分析、典型相关分析和线性回归分析 的特点,可以解决多重线性回归中的多重共线性问 题,常用作高光谱数据建模的通用方法[19-20]。 SVMR 模型是非线性预测模型,主要基于内核的机 器学习方法,使用特定的传递核函数将光谱数据矩 阵映射到高维特征空间,并建立一个超平面作为决 策曲面,以间隔最大化分割原则对不同样本进行分 割,进而对土壤重金属含量进行反演预测<sup>[15]</sup>。模型 的构建在 Python 软件中完成。

把 Pearson 相关性分析筛选的特征波段作为自 变量,土壤 As 含量作为因变量,利用 RFR, PLSR 和 SVMR 模型对土壤 As 含量进行回归分析,并采用 R<sup>2</sup>, RMSE 和平均绝对误差 (mean absolute error, MAE)作为评价指标。RMSE 和 MAE 的值越小,模 型预测能力越好,模型鲁棒性越高; R<sup>2</sup>为负值,说明 模型的预测能力不好,R<sup>2</sup>为正值,且越趋近于1,说 明预测模型的拟合程度越好,拥有更好的稳定性。

3 结果与分析

1.47

# 3.1 土壤 As 含量统计分析

对 84 个土壤样品的 As 含量进行统计分析,结 果如表1所示。

表1 各数据集 As 含量描述性统计

	Tab.1         Descriptive statistics of As content in each dataset							
组别	样品数	范围/(mg・kg <sup>-1</sup> )	平均值/(mg・kg <sup>-1</sup> )	标准差/(mg・kg <sup>-1</sup> )	变异系数	土壤背景值/(mg・kg <sup>-1</sup> )		
全部	84	6.00~13.80	10.28	1.32	0.13			
校准集	64	6.84~13.80	10.30	1.25	0.12	11.20		

10.07

参考乌鲁木齐市土壤背景值[21],采集的土样 As 含量平均值为 10.28 mg/kg,低于土壤背景值, As 含量最大值为13.80 mg/kg,等于相应背景值的1.23 倍,表明部分样点土壤 As 含量富集较明显。为保证 校准集与验证集 As 含量范围一致,将 64 个土样作

 $6.00 \sim 11.80$ 

为校准集,用于建模;20个土样作为验证集,用于 检验模型。表1可见,校准集和验证集的 As 含量平 均值、标准差和变异系数基本一致,表明划分比较合 理,可以用于后续建模。

0.15

第1期

#### 3.2 相关性检验与特征波段筛选

将土壤 As 含量实测值与土壤原始光谱数据和 12 种光谱变换数据的波段进行 Pearson 相关性检 验,得到土壤 As 含量与光谱数据的相关系数(图 3)。其中,达极显著水平(P<0.01)的相关系数临界 值为±0.278,对应的自由度为 83。红色曲线代表原 始光谱数据与土壤 As 含量之间相关性,蓝色曲线为 12 种光谱变换数据与土壤 As 含量之间相关性。土 壤原始光谱数据与As含量的相关性差,没有对应的特征波段。由图3可以看出,变换后的土壤光谱数据与As含量的相关性均有不同程度的提高,显著相关的波段数量也明显增多。根据土壤光谱数据与As含量的相关系数的大小,筛选出各种光谱数据变换形式下绝对值大于0.278的波段作为特征波段,用于建立土壤As含量高光谱反演预测模型。



图 3 土壤 As 含量与光谱反射率及其变换的相关系数

Fig.3 Correlation coefficient between soil As content and spectral radiance and spectral transformation

## 3.3 不同反演模型对 As 含量反演的影响

基于 Python 软件,将 64 个样本作为校准集进 行建模,并将模型的随机种子数设为 91。由于 RFR 模型具有随机性,决策树的数量会干扰模型的预测 性能,在考虑模型性能、模型运行时间和样本数量等 因素后,将 RF 的决策树的数量设为 7。用原始光谱 数据和 12 种微分变换筛选出的特征波段作为模型 的自变量 *X*,As 含量作为因变量 *Y*,利用 RFR,PLSR 和 SVMR 这 3 种回归模型进行建模比对,验证特征 波段对不同模型预测能力的影响,优选出研究区土 壤 As 含量的最佳预测模型(表 2)。

由表 2 可知,在 RFR 模型中,除 RTFD 和 RLFD 这 2 种变换形式外,其余 10 种微分变换筛选出的特 征波段与 As 含量的模型反演预测能力均高于原始 光谱数据与 As 含量的模型反演预测能力,表明微分 变换后模型反演预测能力会显著提高。在 PLSR 模型中,除 RLSD 变换形式下的模型反演预测能力低于原始光谱数据建模外,其余 11 种微分变换筛选的特征波段参与构建的模型反演预测能力均高于原始光谱数据建模。在 SVMR 模型中,12 种微分变换形式下的模型反演预测能力均低于原始光谱数据与As 含量的模型反演预测能力,微分变换后,筛选出的相关性高的光谱波段数量也明显增加,但是建模效果并不明显,说明 SVMR 模型的反演预测能力更易受到波段数量的影响。

原始光谱数据参与构建的3种模型中,SVMR 的反演预测能力比 RFR 和 PLSR 模型的反演预测 能力要好。但经过微分变换和特征波段筛选的光谱 数据参与构建的3种模型中,RFR 和 PLSR 的反演 预测能力比SVMR模型要好。RFR模型中,基于

•	192	•

自然资源遥感

表 2 反演模型精度参数统计

Tab.2 Statistics of precision parameter of inversion model									
	RFR			PLSR			SVMR		
光谱变换	<b>D</b> <sup>2</sup>	RMSE/	MAE/	<b>D</b> <sup>2</sup>	RMSE/	MAE/	n <sup>2</sup>	RMSE/	MAE/
	ĸ	$(mg \cdot kg^{-1})$	$(mg \cdot kg^{-1})$	ĸ	$(mg \cdot kg^{-1})$	$(mg \cdot kg^{-1})$	ĸ	$(mg \cdot kg^{-1})$	$(mg \cdot kg^{-1})$
R <sup>①</sup>	0.604	0.280	0.802	0.555	0.240	0.798	0.697	0.107	0.894
FD	0.644	0.134	0.748	0.646	0.159	0.832	0.459	0.031	1.008
SD	0.678	0.230	0.652	0.653	0.170	0.807	0.231	0.029	1.051
RTFD	0.575	0.130	0.780	0.646	0.157	0.829	0.614	0.034	0.968
RTSD	0.626	0.183	0.739	0.632	0.168	0.828	0.277	0.031	1.040
LTFD	0.626	0.163	0.768	0.657	0.163	0.819	0.585	0.036	0.982
LTSD	0.669	0.208	0.692	0.623	0.165	0.843	0.177	0.029	1.059
RMSFD	0.641	0.121	0.769	0.646	0.163	0.828	0.541	0.036	0.991
RMSSD	0.821	0.143	0.523	0.588	0.181	0.852	0.252	0.031	1.049
ATFD	0.605	0.168	0.775	0.657	0.163	0.819	0.585	0.036	0.982
ATSD	0.679	0.219	0.720	0.623	0.165	0.843	0.177	0.029	1.059
RLFD	0.591	0.154	0.777	0.556	0.280	0.804	0.334	0.031	1.037
RLSD	0.710	0.219	0.680	0.498	0.208	0.886	0.209	0.031	1.054

①R 为未经过变换的原始光谱曲线。

RMSSD-RFR 模型的反演预测能力最好,其  $R^2$ 为 0.821, *RMSE*为 0.143 mg/kg, *MAE*为 0.523 mg/kg, 其他光谱变换方式建模的  $R^2$ 均超过 0.575。PLSR 模型反演预测的结果波动范围小,基于 LTFD-PLSR 和 ATFD-PLSR 模型的反演预测效果最好,  $R^2$ 均为 0.657,除 RLSD-PLSR 模型的  $R^2$ 为 0.498 外,其余光 谱变换方式建模的  $R^2$ 均超过 0.556。SVMR 模型所 用的光谱指标与 RFR 和 PLSR 模型相同, 原始光谱 数据建模的  $R^2$ 为 0.697, 12 种微分变换方式建模的  $R^2$ 范围为 0.177~0.614, *RMSE*均小于0.1 mg/kg, *MAE*的波动范围在 0.968~1.059 mg/kg 之间。

# 3.4 最优模型对比与讨论

实测值与模型预测值对比如图 4 所示。从图 4 可见,RFR 模型的拟合效果好,稳定性强,预测能力 良好; PLSR 模型的预测能力较稳定; SVMR 模型预 测结果的范围波动较大,且其原始光谱数据的建模 效果优于微分变换的光谱数据建模。3 种模型的反 演预测能力为 RMSSD-RFR > R-SVMR > LTFD-PLSR = ATFD-PLSR。可以看出,RMSSD-RFR 模 型拟合效果优于 PLSR 模型和 SVMR 模型,说明 RFR 建模对研究区 As 含量高光谱估算具有更高的 稳定性和可靠性。



Fig.4 Comparison of measured and predicted values of modeling

李志远等<sup>[11]</sup>把河北石家庄矿区的土壤作为研 究对象,对土壤中As含量进行高光谱估算,发现 M-PLSR建模效果最佳,其*R*<sup>2</sup>为0.659,*RMSE*为 0.312 mg/kg;李全坤等<sup>[22]</sup>的研究结果表明,浙江温 岭土壤以水稻土、黄壤和红壤为主,其最优预测模型 为PLS模型,*R*<sup>2</sup>为0.755,*RMSE*为0.720 mg/kg;杨 君怡等<sup>[9]</sup>研究发现,广东肇庆土壤以红壤为主,土 壤As含量最优预测模型为FD-PLSR模型,其*R*<sup>2</sup>为 0.790,*RMSE*为0.832 mg/kg;郭云开等<sup>[23]</sup>对湖南岳 阳耕地土壤中As含量高光谱反演,结果表明,遗传 算法结合 SVM 模型为最优预测模型,其*R*<sup>2</sup>为0.821, *RMSE*为0.084 mg/kg;对新疆吐鲁番葡萄园土壤的 相关研究结果表明<sup>[24]</sup>,基于 LTFD 的地理加权回归 模型为葡萄园土壤 As 含量最优预测模型,其 *R*<sup>2</sup>为 0.935,*RMSE* 为 0.402 mg/kg; 袁自然等<sup>[14]</sup>的研究表 明,湖北洪湖黄壤和棕壤中 As 含量最优预测模型为 PSO-SVM 模型,其 *R*<sup>2</sup>为 0.982,*RMSE* 为 0.522 mg/kg。 与以上研究区土壤 As 含量高光谱估算研究成果相 比,本研究构建的土壤 As 含量最优高光谱反演模型 为 RMSSD-RFR 模型,*R*<sup>2</sup>和 *RMSE* 分别为 0.821 和 0.143 mg/kg,属于中等水平。不同研究结果得到的 土壤 As 含量高光谱反演模型估算精度有一定差异, 这可能与区域土壤类型、土地利用方式、土壤中 As 元 素生物地球化学特征以及模型的适用性和鲁棒性等

#### 不同有关。

第1期

# 4 结论

本文利用不同光谱分析方法处理后的光谱数据,对比研究了 RFR,PLSR 和 SVMR 等机器学习法 对典型绿洲城市——乌鲁木齐市土壤 As 含量的预 测能力,得出以下结论:

1) 乌鲁木齐市土壤中 As 元素的含量平均值为 10.28 mg/kg,低于相应土壤背景值, As 含量最大值 为相应背景值的 1.23 倍,表明部分区域土壤 As 含 量富集较明显。

2) 光谱变换能够增强光谱特征, 提高土壤高光 谱数据与 As 含量之间的相关性。运用原始光谱数 据和微分变换后的光谱数据构建 RFR, PLSR 和 SVMR 模型, 结果表明模型的反演预测能力均有相 应的变化, 3 种模型均具有一定的预测能力。

3)3种模型的反演预测能力由高到低依次为: RFR>SVMR>PLSR,其中 RMSSD-RFR 的反演预测 能力最高,其 *R*<sup>2</sup>为 0.821,*RMSE* 为 0.143 mg/kg, *MAE* 为 0.523 mg/kg,说明 RFR 模型拟合效果好,具 有较高的稳定性和预测能力,可以为构建绿洲城市 土壤 As 含量高光谱估算模型提供支持。同时,由于 区域位置、土壤类型和污染水平存在差异,构建的最 佳模型是否适用其他土壤类型和地区需进一步深入 研究。

#### 参考文献(References):

- [1] 郭学飞,曹颖,焦润成,等.基于高光谱的北京铁矿区土壤重金 属镍元素含量反演研究[J].土壤通报,2021,52(4):960-967.
   Guo X F,Cao Y,Jiao R C, et al.An inversion of soil nickel contents with hyperspectral in iron mine area of Beijing [J].Chinese Journal of Soil Science,2021,52(4):960-967.
- [2] 温其谦,阎秀兰,申俊峰,等.半壁山金矿矿业活动区砷赋存的 矿物特征及其对农田土壤砷累积的影响[J].环境科学,2019, 40(11):5090-5097.

Wen Q Q, Yan X L, Shen J F, et al. Mineral characteristics of arsenic in the active area of the Banbishan gold mine and its effect on arsenic accumulation in farmland soil [J]. Environmental Science, 2019,40(11):5090–5097.

- [3] 邵金秋,温其谦,阎秀兰,等.天然含铁矿物对砷的吸附效果及 机制[J].环境科学,2019,40(9):4072-4080.
   Shao J Q, Wen Q Q, Yan X L, et al. Adsorption and mechanism of arsenic by natural iron-containing minerals [J]. Environmental Science, 2019,40(9):4072-4080.
- [4] 贺军亮,张淑媛,查 勇,等,高光谱遥感反演土壤重金属含量研究进展[J].遥感技术与应用,2015,30(3):407-412.
  He J L, Zhang S Y, Zha Y, et al. Review of retrieving soil heavy metal content by hyperspectral remote sensing [J].Remote Sensing Technology and Application,2015,30(3):407-412.
- [5] Tan K, Wang H M, Chen L H, et al. Estimation of the spatial distri-

bution of heavy metal in agricultural soils using airborne hyperspectral imaging and random forest [J].Journal of Hazardous Materials,2020,382:120987.

[6] 晏红波,韦晚秋,卢献健,等.基于高光谱特征的土壤含水量遥 感反演方法综述[J].自然资源遥感,2022,34(2):1-9.doi:10. 6046/zrzyg.2021126.

Yan H B, Wei W Q, Lu X J, et al. A review of remote sensing inversion methods for estimating soil water content based on hyperspectral characteristics [J]. Remote Sensing for Natural Resources, 2022,34(2):1-9.doi:10.6046/zrzyyg.2021126.

- [7] Sawut R, Kasim N, Abliz A, et al. Possibility of optimized indices for the assessment of heavy metal contents in soil around an open pit coal mine area [J]. International Journal of Applied Earth Observation and Geoinformation, 2018, 73; 14-25.
- [8] 王嘉芃,徐建国,沈家晓,等.德兴铜矿矿山重金属污染修复效 果高光谱遥感评价[J].自然资源遥感,2023,35(3):284-291. doi:10.6046/zrzyyg.2022196.

Wang J P, Xu J G, Shen J X, et al. Evaluating the remediation effect of heavy metal pollution in the Dexing copper mine based on hyperspectral remote sensing[J].Remote Sensing for Natural Resources, 2023,35(3):284-291.doi:10.6046/zrzyyg.2022196.

- [9] 杨君怡,钱乐祥,郑艳伟,等.基于高光谱的土壤砷和铜元素含量估算[J].科技通报,2018,34(11):112-119,124.
   Yang JY,Qian LX,Zheng YW, et al. Estimation of the content of arsenic and copper in soil based on hyperspectral data [J].Bulletin of Science and Technology,2018,34(11):112-119,124.
- [10] 郑光辉,周生路,吴绍华.土壤砷含量高光谱估算模型研究[J]. 光谱学与光谱分析,2011,31(1):173-176.
  Zheng G H,Zhou S L,Wu S H.Prediction of As in soil with reflectance spectroscopy [J].Spectroscopy and Spectral Analysis,2011, 31(1):173-176.
- [11] 李志远,邓 帆,贺军亮,等.土壤重金属砷的高光谱估算模型
  [J].光谱学与光谱分析,2021,41(9):2872-2878.
  Li Z Y, Deng F, He J L, et al. Hyperspectral estimation model of heavy metal arsenic in soil [J].Spectroscopy and Spectral Analysis,2021,41(9):2872-2878.
- [12] 王泽强,张冬有,徐夕博,等.基于遥感时-空-谱特征及随机森
   林模型的土壤重金属空间分布预测[J].环境科学,2024,45
   (3):1713-1723.

Wang Z Q,Zhang D Y,Xu X B,et al.Distribution prediction of soil heavy metals based on remote sensing temporal-spatial-spectral features and random forest model [J].Environmental Science, 2024,45(3):1713-1723.

[13] 史 广,刘 庚,赵 龙,等.基于多源环境数据和随机森林模型的农田土壤砷空间分布模拟[J].环境科学学报,2020,40(8): 2993-3000.

Shi G, Liu G, Zhao L, et al. Prediction of arsenic for farmland soil based on multi source environmental data and random forest model [J]. Acta Scientiae Circumstantiae, 2020, 40(8); 2993–3000.

[14] 袁自然,魏立飞,张杨熙,等.优化 CARS 结合 PSO-SVM 算法 农田土壤重金属砷含量高光谱反演分析[J].光谱学与光谱分 析,2020,40(2):567-573.

Yuan Z R, Wei L F, Zhang Y X, et al. Hyperspectral inversion and analysis of heavy metal arsenic content in farmland soil based on optimizing CARS combined with PSO-SVM algorithm [J]. Spectroscopy and Spectral Analysis, 2020, 40(2):567–573.

[15] 易兴松,兰安军,文锡梅,等.基于 ASD 和 GaiaSky-mini 的农田 土壤重金属污染监测[J].生态学杂志,2018,37(6):1781-1788. · 194 ·

Yi X S, Lan A J, Wen X M, et al. Monitoring of heavy metals in farmland soils based on ASD and GaiaSky-mini [J].Chinese Journal of Ecology, 2018, 37(6):1781-1788.

- [16] 麦麦提吐尔逊·艾则孜,阿吉古丽·马木提,艾尼瓦尔·买买提.新疆焉耆盆地辣椒地土壤重金属污染及生态风险预警[J]. 生态学报,2018,38(3):1075-1086.
  Mamattursun E, Ajigul M, Anwar M.Soil heavy metal pollution and ecological risk warning assessment of pepper field in Yanqi Basin, Xinjiang[J]. Acta Ecologica Sinica, 2018, 38(3):1075-1086.
- [17] 安柏耸,王雪梅,黄晓宇,等.基于连续小波变换的土壤重金属 镉含量的高光谱估测[J].地球与环境,2023,51(2):246-253.
  An B S, Wang X M, Huang X Y, et al. Hyperspectral estimation of heavy metal cadmium content in soil based on continuous wavelet transform[J].Earth and Environment,2023,51(2):246-253.
- [18] Liu W W, Li M J, Zhang M Y, et al. Hyperspectral inversion of mercury in reed leaves under different levels of soil mercury contamination [J]. Environmental Science and Pollution Research International, 2020, 27(18):22935-22945.
- [19] Leone A P, Viscarra-Rossel R A, Amenta P, et al. Prediction of soil properties with PLSR and vis-NIR spectroscopy: Application to Mediterranean soils from southern Italy [J]. Current Analytical Chemistry, 2012, 8(2):283-299.
- [20] Wei L F, Pu H C, Wang Z X, et al. Estimation of soil arsenic content with hyperspectral remote sensing[J].Sensors, 2020, 20(14): 4056.
- [21] 阿地拉·艾来提,麦麦提吐尔逊·艾则孜,靳万贵,等.新疆库

尔勒市道路积尘重金属污染及健康风险评价[J].中国地质, 2020,47(6):1915-1925.

Adila H, Mamattursun E, Jin W G, et al. Pollution and health risks assessment of heavy metals of road dust in Korla City, Xinjiang[J]. Geology in China, 2020, 47(6):1915–1925.

[22] 李全坤,赵万伏,文宇博,等.基于可见—近红外反射光谱的典型农田重金属污染风险分类研究[J].高校地质学报,2022,28(6):799-813.

Li Q K, Zhao W F, Wen Y B, et al. Classification of heavy metal contamination risk in typical agricultural soils by visible and near infrared reflectance spectroscopy [J].Geological Journal of China Universities, 2022, 28(6):799–813.

[23] 郭云开,张思爱,谢晓峰,等.基于 GA-SVM 的耕地土壤重金属 含量高光谱反演方法的研究[J].土壤通报,2021,52(4):968-974.

Guo Y K, Zhang S A, Xie X F, et al. The hyperspectral inversion method of heavy metal contents in cultivated soils based on GA-SVM [J].Chinese Journal of Soil Science, 2021, 52(4):968-974.

[24] 阿依努尔·麦提努日,麦麦提吐尔逊·艾则孜,李新国.基于地 理加权回归模型的葡萄园土壤砷含量高光谱反演[J].环境监 测管理与技术,2021,33(4):45-48.
Aynur M, Mamattursun E, Li X G. Hyperspectral inversion of arsenic content in vineyard soil based on geographically weighted regression model [J]. The Administration and Technique of Environmental Monitoring,2021,33(4):45-48.

# Hyperspectral inversion of arsenic content in soil in an oasis city

ZHONG Qing<sup>1</sup>, MAMATTURSUN Eziz<sup>1, 2</sup>, MIREGULI Ainiwaer<sup>1</sup>, HAO Haiyu<sup>3</sup>

(1. College of Geographical Science and Tourism, Xinjiang Normal University, Urumqi 830054, China; 2. Xinjiang Laboratory of Lake Environment and Resources, Xinjiang Normal University, Urumqi 830054, China; 3. College of Physics and

Electronic Engineering, Xinjiang Normal University, Urumqi 830054, China)

Abstract: Arsenic (As) is a metalloid element with high carcinogenicity, rendering it particularly important to detect As content in soils in a swift and accurate manner. The study focused on the topsoil in Urumqi City, where 84 soil samples were collected and tested for their As content and original spectral reflectance. This study examined the relationships of As content in the soils with the spectral reflectance under the original spectra and 12 spectral transformations using the Pearson correlation analysis, followed by screening characteristic bands. Hyperspectral models for the inversion of As content in soils were developed using partial least squares regression (PLSR), random forest regression (RFR), and support vector machine regression (SVMR). Finally, the prediction performance of the hyperspectral models was elevated based on the coefficients of determination ( $R^2$ ), root-mean-square errors (RMSEs), and mean absolute errors (MAEs). The results indicated that applying differential transformations to the original spectral data can effectively enhance the spectral features and improve the correlation between spectral reflectance and As content in soils. The prediction performance of the hyperspectral models decreased in the order of RFR, SVMR, and PLSR. The RFR model based on root-mean-square second order differentiation (RMSSD-RFR) exhibited the best fitting effects and the highest prediction stability, with  $R^2$  of 0.821, a RMSE of 0.143 mg/kg, and a MAE of 0.523 mg/kg. This study provides a scientific basis for developing hyperspectral models for the inversion of As content in soils in an oasis city.

Keywords: urban soil; As; hyperspectral inversion; spectral transformation; inversion model