

DOI: 10.11766/trxb202411180444

CSTR: 32215.14.trxb202411180444

杞应涛, 甘淑, 袁希平, 胡琳, 胡建开, 卢成卓. 利用直接校正算法与分数阶微分对矿区土壤 Pb 和 Zn 含量的定量反演与分析[J]. 土壤学报, 2026, 63 (1): 241–250.

QI Yingtao, GAN Shu, YUAN Xiping, HU Lin, HU Jiankai, LU Chengzhuo. Quantitative Inversion of Pb and Zn Content in Mining Area Soils Based on Direct Standardization Algorithm and Fractional Order Derivative[J]. Acta Pedologica Sinica, 2026, 63 (1): 241–250.

## 基于直接校正算法与分数阶微分定量反演矿区土壤 Pb 和 Zn 含量\*

杞应涛<sup>1</sup>, 甘淑<sup>1, 2†</sup>, 袁希平<sup>1, 2</sup>, 胡琳<sup>1, 2</sup>, 胡建开<sup>1</sup>, 卢成卓<sup>1</sup>

(1. 昆明理工大学国土资源工程学院, 昆明 650093; 2. 云南省高校高原山区空间信息测绘技术应用工程研究中心, 昆明 650093)

**摘 要:** 高光谱技术为土壤重金属含量的快速、精准监测提供了全新的解决方案。然而, 利用实验室光谱建立的模型在实际应用中泛化能力较弱; 此外, 直接使用遥感影像光谱数据反演土壤重金属含量时, 受成像时天气状况以及地面环境等因素的影响, 导致模型精度较低, 难以准确反映研究区重金属含量的分布情况。本研究以云南省会泽县矿山镇某尾矿区为研究对象, 获取 56 个表层土壤样本的高光谱反射率 (地面和影像) 以及 Pb、Zn 的含量。首先, 采用直接校正 (DS) 算法结合实验室光谱数据对高分 5 号影像数据进行光谱校正; 随后, 使用 Box-Cox 变换对 Pb 和 Zn 含量进行正态化处理; 接着, 通过分数阶微分 (FOD) 对校正后的光谱进行变换, 并利用 Boruta 算法筛选特征波段; 最后, 构建随机森林和 XGBoost 反演模型。研究表明, DS 算法可有效消除土壤粒径和含水量等干扰因素对影像光谱的影响; Box-Cox 变换解决了 Pb 和 Zn 含量的偏态分布问题; FOD 有效增强了细节光谱特征, Boruta 算法选出特征波段显著提升了反演精度; 此外, XGBoost 模型在处理复杂特征交互和非线性关系的回归问题时, 展现出更高的预测精度; 在该研究区 Pb 含量的最佳反演模型为 0.8 Order-Boruta-XGBoost, Zn 含量的最佳反演模型为 1.6 Order-Boruta-XGBoost, 两个最佳反演模型具有较好的鲁棒性。本研究为利用高光谱技术反演矿区土壤中 Pb 和 Zn 含量提供了可靠的参考方法。

**关键词:** 土壤高光谱; 重金属; 分数阶微分; 直接校正算法; XGBoost 模型

中图分类号: X87 文献标志码: A

## Quantitative Inversion of Pb and Zn Content in Mining Area Soils Based on Direct Standardization Algorithm and Fractional Order Derivative

QI Yingtao<sup>1</sup>, GAN Shu<sup>1, 2†</sup>, YUAN Xiping<sup>1, 2</sup>, HU Lin<sup>1, 2</sup>, HU Jiankai<sup>1</sup>, LU Chengzhuo<sup>1</sup>

(1. Faculty of Land and Resources Engineering, Kunming University of Science and Technology, Kunming 650093, China; 2. Application Engineering Research Center of Spatial Information Surveying and Mapping Technology in Plateau and Mountainous Areas set by Universities in Yunnan Province, Kunming 650093, China)

\* 国家自然科学基金项目 (62266026)、云南省教育厅科学研究基金项目 (2025J0079) 和昆明理工大学分析测试基金项目 (2024M20232201156) 资助 Supported by the National Natural Science Foundation of China (No. 62266026), the Scientific Research Foundation of the Education Department of Yunnan Province (No. 2025J0079), and the Analysis and Testing Foundation of Kunming University of Science and Technology (No. 2024M20232201156)

† 通讯作者 Corresponding author, E-mail: gs@kust.edu.cn

作者简介: 杞应涛 (1997—), 男, 云南楚雄人, 硕士研究生, 主要从事利用高光谱遥感技术定量反演研究。E-mail: 15758532473@163.com

收稿日期: 2024-11-18; 收到修改稿日期: 2025-07-20; 网络首发日期 (www.cnki.net): 2025-09-18

**Abstract:** 【 Objective 】 Hyperspectral technology provides a novel solution for the rapid and accurate monitoring of heavy metal content in soils. However, models developed using laboratory spectra often have limited generalizability in practical applications. Additionally, directly estimating soil heavy metal concentrations from remote sensing imagery is often hampered by factors such as weather conditions and surface environment at the time of image acquisition, which leads to reduced model accuracy and limits the ability to accurately reflect the spatial distribution of heavy metals in the study area. 【 Method 】 In this study, a tailings area in Kuanshan Town, Huize County, Yunnan Province, was selected as the research site. A total of 56 surface soil samples were collected, and both ground-based and image-based hyperspectral reflectance, as well as Pb and Zn concentrations, were obtained. First, the Direct Standardization (DS) algorithm, combined with laboratory spectra, was used to correct the GF-5 imagery. Subsequently, the Box-Cox transformation was applied to normalize the skewed distributions of Pb and Zn concentrations. Then, fractional order derivative (FOD) was performed on the corrected spectra, and the Boruta algorithm was used to identify informative spectral bands. Finally, Random Forest and XGBoost models were developed for the inversion of heavy metal concentrations. 【 Result 】 The results indicate that the DS algorithm effectively mitigated the influence of soil particle size and moisture content on image spectra. The Box-Cox transformation resolved the skewness distribution problem of Pb and Zn content. FOD effectively enhanced detailed spectral features, and the optimal feature band combinations selected by the Boruta algorithm significantly improved the inversion accuracy. Furthermore, the XGBoost demonstrated superior predictive performance in handling complex feature interactions and nonlinear regression problems. 【 Conclusion 】 The optimal inversion model for Pb content in the tailings area was a 0.8 Order-Boruta-XGBoost model, while for Zn content it was the 1.6 Order-Boruta-XGBoost model. Both models exhibited good robustness. This study provides a reliable reference method for using hyperspectral technology to invert Pb and Zn content in mining area soils.

**Key words:** Soil hyperspectrum; Heavy metals; Fractional order derivative; Direct standardization algorithm; XGBoost model

土壤作为地球上所有生态系统存在的基础,健康的土壤是实现人类社会可持续发展的关键因素之一<sup>[1-3]</sup>。然而矿产资源开发导致矿区土壤重金属污染严重,已成为亟待解决的环境难题<sup>[4]</sup>。如何快速准确获取矿区土壤中重金属含量空间分布信息,是制定治理措施和保障人民生命健康的前提。

传统的重金属含量检测方法,虽可揭示特定区域土壤重金属含量的空间分布信息,但由于操作过程复杂、成本高且缺乏时效性,因此难以实现对土壤重金属含量的实时、快速检测<sup>[5]</sup>。而高光谱遥感技术因光谱分辨率高、连续性强以及多波段同时成像能力等优势,能够准确反演土壤中重金属以及其他元素信息<sup>[6]</sup>。室内光谱反演重金属含量的研究作为野外光谱和遥感影像光谱反演土壤重金属含量的基础,也是目前大多数研究所聚焦的领域。野外光谱和遥感影像光谱易受温度、湿度、光照条件等环境因素干扰,数据稳定性与准确性难以保障,致使相应的研究成果在实际应用和推广中还处于探索阶段<sup>[7]</sup>。为了消除环境因素对光谱数据的影响,目前研究常用直接校正(Direct Standardization, DS)算法和分段直接校正(Piecewise Direct Standardization,

PDS)算法对野外或影像光谱数据进行校正,取得较好的建模效果<sup>[8]</sup>。由于高光谱数据冗余度高、自相关性较强,不同预处理和特征变量优选方法能够有效突出光谱特征和优化波谱响应机制<sup>[9]</sup>。其中常用的预处理方法包括包络线去除、倒数对数、标准正态变换、一阶微分和二阶微分等。有研究表明,相较于传统的整数阶微分,使用分数阶微分(Fractional Order Derivative, FOD)能够更加敏锐地捕捉波段的曲率和斜率变化,在建模精度提升方面表现优异<sup>[10]</sup>。丁松滔等<sup>[11]</sup>研究表明,利用 FOD 处理影像光谱数据,并结合偏最小二乘法反演矿区重金属 Pb、Zn、Ni 具有可行性。蒋宇恒等<sup>[12]</sup>构建 FOD 模型对农田 Zn、Ni 含量开展反演研究,结果表明相较于整数阶微分, FOD 建模效果更优。此外, Boruta 特征变量优选算法可有效精简最优特征组合,对简化模型结构、提高模型稳定性和泛化能力效果显著<sup>[13-14]</sup>。Box-Cox 变换作为一种有效的统计方法,能够显著提高重金属含量数据的正态性,为后续建模提供更为可靠的数据基础<sup>[15]</sup>。

目前,针对土壤重金属含量呈偏态分布区域以及利用 DS 算法消除高光谱影像数据受环境变量等

因素干扰的系统性研究鲜有涉及。本文以云南省会泽县矿山镇某尾矿区为研究对象,在室内采集了 56 个表层土壤样本的高光谱数据以及 Pb、Zn 含量数据,同时收集高分 5 号 (GF-5) 影像光谱数据。首先使用 DS 算法联合室内光谱数据校正 GF-5 影像光谱数据,以及利用 Box-Cox 变换使 Pb 和 Zn 含量符合正态性;随后对校正后的光谱数据进行 FOD 光谱变换,并使用 Boruta 算法筛选特征波段;最后,构建随机森林 (Random Forest, RF) 和极限梯度提升树 (eXtreme Gradient Boosting, XGBoost) 反演模型。本研究目的包括:(1) 基于室内光谱数据对星载高光谱数据进行校正,建立多源异质光谱数据的同化机制,消除大气传输与环境噪声引起的光谱畸变;(2) 利用 Box-Cox 算法解决矿区重金属含量偏态分布问题;(3) 通过 FOD 凸显更多细节光谱信息并结合 Boruta 算法筛选出最优特征波段组合;(4) 系统性评估 RF 和 XGBoost 模型反演矿区土壤重金属含量的性能。

## 1 材料与方法

### 1.1 研究区概况和试验数据收集

研究区位于云南省会泽县矿山镇某铅锌尾矿区 (26°37'39"~26°38'56"N, 103°42'17"~103°43'47"E), 属于典型温带高原季风气候,年平均气温和降雨量分别为 13℃和 807mm。该区域铅锌矿床是川滇黔铅锌成矿区内典型的大型富铅锌矿床之一<sup>[16]</sup>。该矿区在开采和冶炼过程中,由于废气、废水、废料和尾矿的堆放及淋溶,导致矿区周围土壤重金属含量不同程度地超标,对周边生态环境造成了严重影响<sup>[17]</sup>。2024 年 3 月,采用梅花形 5 点混合采样法 (范围为: 5 m×5 m) 进行采样。采集深度为 0~20 cm,去除石块、动植物残体等杂质后约取 1 kg 土样,共采集 56 个土壤样品,同时记录每个采样点的地理坐标及周围环境情况。研究区地理位置、采样点分布以及实地照片如图 1 所示。

将采集的土壤样品带回实验室,进行自然风干处理后,使用玛瑙研磨机将样品研磨过 100 目筛,以确保样品的均匀性和一致性。处理过后的样品分为两份,一份使用尼通 XL3t 950 手持矿石元素分析仪测定土壤中 Pb 和 Zn 的含量;另一份使用 ASD

Field Spec 3 地物波谱仪测定光谱数据,波段范围为 350 nm~2 500 nm,重采样间隔设置为 1 nm,得到的波段数为 2 151 个。此外,本研究使用的高分 5 号 (GF-5) 高光谱影像采集于 2024 年 3 月 9 日,通过“国家遥感数据和应用服务平台”(https://www.cpeos.org.cn/home/#/) 申请获取,其空间分辨率为 30 m,波段范围为 400~2 500 nm,共 330 个波段。获取的影像数据需经过完整的预处理流程:首先对影像数据进行辐射定标、大气校正和正射校正;随后进行光谱重采样,使其与室内测定光谱具有相同的光谱分辨率,确保数据的可比性和分析的准确性。

### 1.2 直接校正算法

DS 算法是一种对光谱直接校正的算法,可以通过实验室测定的室内光谱来校正野外光谱或高光谱影像光谱,从而消除环境因素对光谱的影响<sup>[13]</sup>。本研究室内光谱 ( $X_{lab}$ ) 和 GF-5 影像光谱 ( $X_{GF5}$ ) 的矩阵维度均为  $n \times p$ ,其中  $n$  为转换光谱的数量, $p$  为波段的数量。室内和影像光谱的 DS 转换关系式为:

$$X_{lab}=X_{GF5}B+E \quad (1)$$

式中, $B$  为由  $X_{lab}$  和  $X_{GF5}$  共同决定的未知参数转换矩阵,维度为  $P \times P$ ;  $E$  为背景残差矩阵,表示为:

$$E=\lambda d_s^T \quad (2)$$

式中, $d_s$  为由基线差异产生的  $P \times 1$  的矩阵; $\lambda$  为所有列向量值均为 1 的  $n \times 1$  的矩阵。将式 (2) 代入式 (1) 可得:

$$X_{lab}=X_{GF5}+\lambda d_s^T \quad (3)$$

为了计算未知的转换矩阵  $B$ ,首先需要引入一个  $n \times n$  阶的中心化矩阵  $C_n$ ,可表达为:

$$C_n=I_n-\left(\frac{1}{n}\right)\lambda\lambda^T \quad (4)$$

式中, $I_n$  为一个  $n \times n$  的单位矩阵。将式 (3) 两边同时乘以  $C_n$ ,因  $C_n\lambda d_s^T=0$ ,并且  $C_nX_{lab}$  和  $C_nX_{GF5}$  分别为  $X_{lab}$  和  $X_{GF5}$  的均值中心矩阵,分别用  $X_{lab}$  和  $X_{GF5}$  表示,则式 (3) 可以表示为:

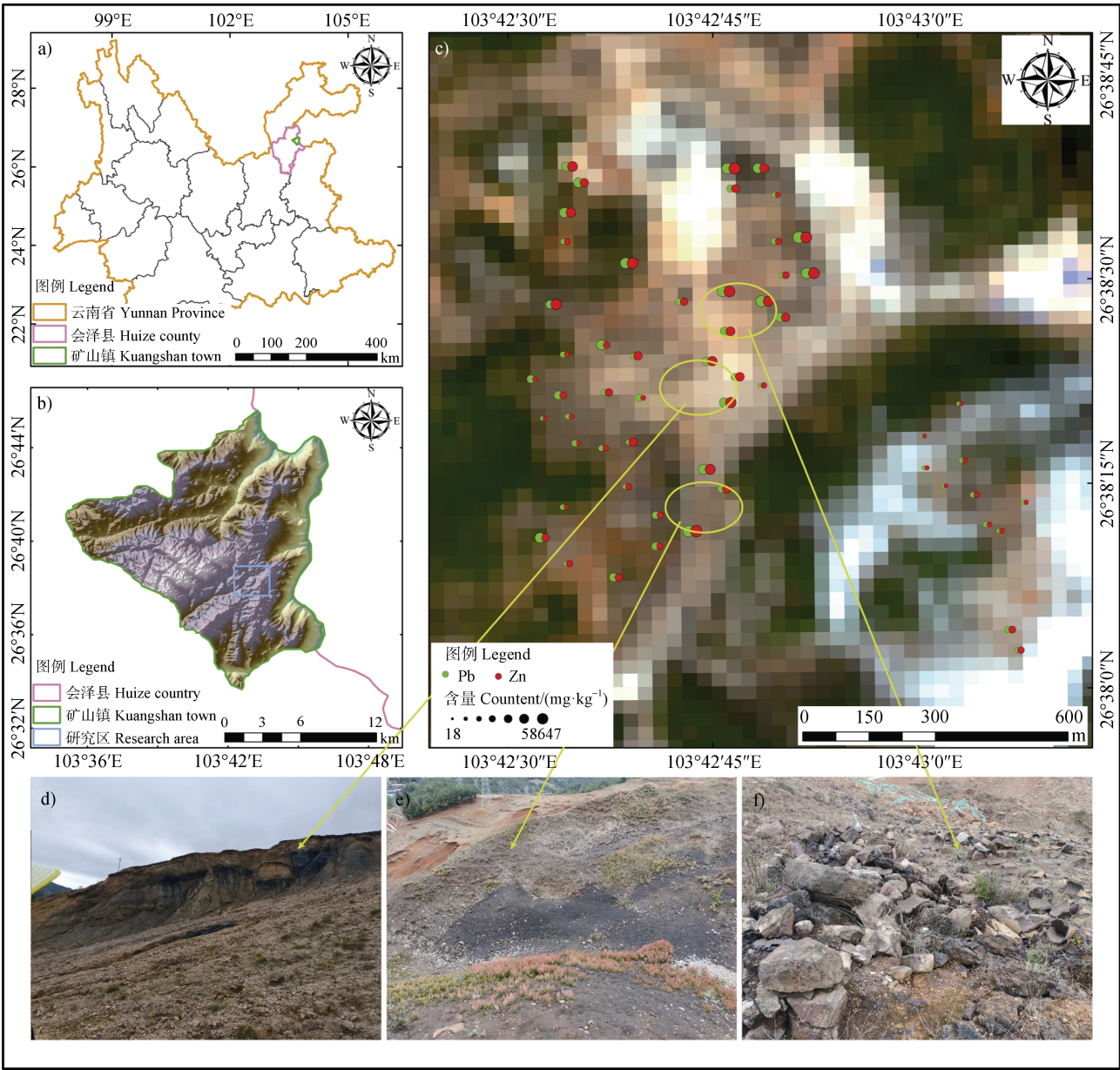


图 1 研究区位置和采样点分布示意图

Fig. 1 Schematic map of the study area location and distribution of sampling points

$$\bar{X}_{lab} = \bar{X}_{GF5}B \quad (5)$$

$$d_s = \bar{X}_{lab}^T - B^T \bar{X}_{GF5}^T \quad (7)$$

经过最小二乘变为：

$$B = \bar{X}_{GF5}^+ \bar{X}_{lab} \quad (6)$$

式中， $\bar{X}_{lab}^T$  和  $\bar{X}_{GF5}^T$  为维度为  $1 \times P$  的行向量，分别由  $X_{lab}$  和  $X_{GF5}$  每一列均值构成。

最终经过上述 DS 转换后的影像光谱可表示为：

$$X'_{lab} = X_{GF5}B + E \quad (8)$$

式中， $+$ 表示  $\bar{X}_{GF5}$  的广义逆矩阵。通过转换矩阵  $B$  可以计算出背景残差矩阵  $E$ ，将式 (3) 两边同时乘以  $(1/n) \lambda^T$  得：

### 1.3 分数阶微分

分数阶微分 (FOD) 将传统整数阶微分推广至



任意阶，在光谱分析时能够更加敏锐地捕捉到光谱反射率细节的变化。常用的分数阶微分方法主要包括：Riemann-Liouville（R-L）、Grunwald-Letnikov（G-L）和 Caputo<sup>[18]</sup>。其中，G-L 基于无穷级数进行定义，适合数值计算，能够处理复杂的信号和系统，因此是在光谱分析中进行 FOD 计算的常用方法。

1.4 建模方法

本文使用 Boruta 特征变量优选算法进行特征波段筛选<sup>[13]</sup>，构建了随机森林（Random Forest，RF）和极限梯度提升树（eXtreme Gradient Boosting，XGBoost）反演模型<sup>[10]</sup>，预测研究区内土壤 Pb 和 Zn 的含量。模型精度的评价指标选择决定系数（ $R^2$ ）、均方根误差（RMSE）和相对预测偏差（RPD）。其中，RPD 值将模型预测性能分为三类：A 类（ $RPD>2.0$ ），表示预测能力较强；B 类（ $1.4<RPD<2.0$ ），表示预测能力中等；C 类（ $RPD<1.4$ ），表示预测能力较弱<sup>[14]</sup>。

2 结 果

2.1 土壤中 Pb 和 Zn 原始含量及 Box-Cox 变换后含量的统计特征

研究区 Pb 和 Zn 含量的统计特征结果见表 1。对于原始数据，Pb 含量范围为 17.95~33442.41 mg.kg<sup>-1</sup>，标准差为 7086.17 mg.kg<sup>-1</sup>，偏度为 2.92；Zn 含量范围为 45.85~58647.12 mg.kg<sup>-1</sup>，标准差为 11628.44 mg.kg<sup>-1</sup>，偏度为 3.23。在土壤调查中，变异系数（Coefficient of Variance，CV）能够有效反映土壤中重金属含量的空间分布差异<sup>[19]</sup>；Pb 和 Zn 的 CV 均超过 100%，其含量在该研究区内呈现高度变异，说明其空间分布差异较大；Pb 和 Zn 的 Pearson 相关系数达到 0.82（ $P<0.01$ ）。经过 Box-Cox 转换后，Pb 和 Zn 含量的范围跨度显著减小，数据由高度变异转变为中度变异，偏度也接近于 0，说明转换后的数据更符合正态分布，同时 Pearson 相关系数提升 0.09，进一步凸显了二者之间的关联性。

表 1 Pb 和 Zn 含量的统计特征

Table 1 The statistical characteristics of Pb and Zn content

数据类型	元素	范围	标准差 Standard deviation /	偏度	变异系数 Coefficient	二者相关系数
Data type	Element	Range / ( mg.kg <sup>-1</sup> )	( mg.kg <sup>-1</sup> )	Skewness	of variance/%	Correlation coefficient
原始数据	Pb	17.95~33442.41	7086.17	2.92	176.8	0.82
Original data	Zn	45.85~58647.12	11628.44	3.23	186.8	
Box-Cox 数据	Pb	3.12~13.90	2.58	-0.11	29.8	0.91
Box-Cox data	Zn	3.59~9.15	1.43	-0.17	22.2	

2.2 光谱数据预处理和直接校正

光谱数据在采集过程中容易受仪器、人为操作以及外部环境等诸多因素的影响，导致测定的光谱反射率中混入大量随机噪声，因此需对室内光谱数据和影像光谱反射率进行预处理。首先采用 Savitzky-Golay（SG）滤波技术对光谱数据进行平滑处理，以降低噪声的影响；其次，在 350~429 nm、996~1 005 nm 和 2 401~2 500 nm 的波段范围内，受到传感器自身缺陷的影响，光谱反射率出现明显的振荡，导致其信噪比降低。此外，由于水汽的强烈吸收作用，在 1 344~1 424 nm 和 1 804~1 954 nm 的波段中，光谱反射率出现了明显的“阶跃”现象，

因此需要将这些波段从后续的处理流程中排除；室内和影像光谱数据经过上述处理后，保留的有效波段数为 1 733 个，图 2a 和图 2b 分别展示了经过预处理后的室内光谱和 GF-5 影像光谱反射率。预处理完成后，运用 DS 算法结合室内光谱数据对 GF-5 影像数据进行光谱校正，校正后的影像光谱数据如图 2c 所示，不仅提升了影像光谱反射率波形与室内光谱反射率之间的相似度，而且使得影像光谱曲线更为平滑，反射率也有显著提高。这充分表明，DS 能够有效抑制土壤粒径、含水量和温度等环境因素对影像光谱反射率的影响。将经过以上处理后的影像光谱反射率作为原始光谱反射率（Origin spectral

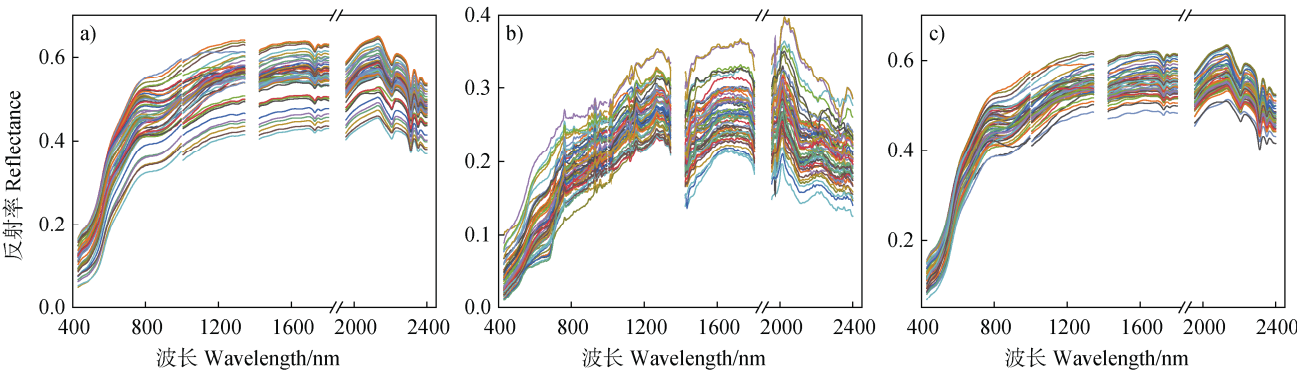


图 2 实验室光谱曲线 (a)、GF-5 影像光谱曲线 (b) 和校正后光谱曲线 (c)

Fig. 2 Laboratory spectrum curve (a), GF-5 image spectrum curve (b), and corrected spectrum curve (c)

reflectance, OR), 并将其应用于后续的建模分析。

2.3 光谱分数阶微分与特征波段选取

图 3 展示了使用 G-L 分数阶微分方法对 OR 进行光谱变换的结果。在 0~2 阶的范围内, 以 0.2 阶为步长进行 10 次光谱变换。结果表明, 随着微分阶数的增加, 光谱反射率逐渐降低并趋近于零, 波峰和波谷的特征逐渐鲜明, 但其强度差异逐渐减小, 这一变化规律符合 G-L 定义的数学原理: 当微分步长小于波峰和波谷的固有宽度时, 细微的光谱差异被逐阶放大, 光谱特征会被显著增强<sup>[20]</sup>。然而值得注意的是, 高阶微分 ( $\geq 1.0$  阶) 会放大波段间差异较大的噪声信号, 尤其是 1 400 nm 和 2 000 nm 附近出现显著的高频噪声干扰。相比之下, 低阶微分 (如 0.2 阶和 0.4 阶) 生成的光谱曲线更为平滑, 噪声较小。因此, 不同阶数的 G-L 分数阶微分在增强光

谱特征和抑制噪声方面表现出显著差异, 这为基于光谱特征的土壤分析提供了分数阶微分的优化选择依据。

对 OR 进行分数阶微分处理后, 运用 Boruta 算法筛选特征波段。如图 4 所示, 经 Boruta 算法筛选得到的 Pb 和 Zn 特征波段分布基本一致, 其中高阶筛选获得的特征波段数量较多, 而低阶筛选获得的相对较少。使用高阶微分 ( $\geq 1.0$  阶) 时, 两种重金属元素特征波段主要集中在 700~1 000 nm、1 400~1 750 nm 和 2 250~2 400 nm 范围内, 而低阶微分处理下, 特征波段少部分位于可见光区域, 更多集中于近红外和短波红外区域。

2.4 土壤重金属含量预测模型的构建

本次研究采用 RF 和 XGBoost 算法构建矿区土壤中 Pb 和 Zn 含量的估算模型, 所有模型均在 Python 的 scikit-learn 和 XGBoost 库中实现。模型以 Boruta 算法筛选的各阶次特征波段作为自变量, 转换后 Pb 和 Zn 的含量为因变量, 其中 70% 为训练集, 30% 为验证集。

图 5 展示了 RF 模型在不同微分阶数下对 Pb 和 Zn 含量估算的各项指标, 随着微分阶数的增加, Pb 和 Zn 的  $R^2$  呈现先下降后逐渐上升的趋势。对于 Pb, 模型精度在 1.8 阶时最高, 验证集的  $R^2$  为 0.618, RMSE 为 1.517, PRD 为 1.618, 其次是 1.4 阶和 2 阶的  $R^2$  分别为 0.51 和 0.575, RMSE 分别为 1.517 和 1.718, RPD 分别为 1.535 和 1.618; 1.4 阶、1.8 阶和 2 阶的 PRD 均介于 1.4 和 2.0 之间, 说明对 Pb 的预测能力中等, 而其余阶数的 RPD 小于 1.4, 则模型预测能力较差。对于 Zn, 模型精度在 2 阶时最

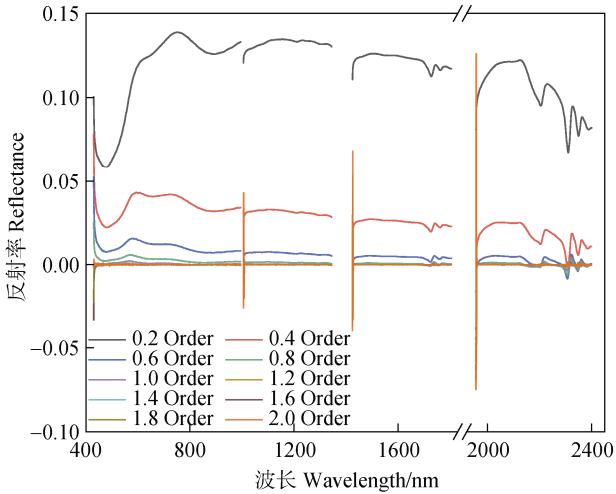


图 3 各阶次分数阶微分光谱曲线

Fig. 3 Fractional order derivative spectral curves of different orders

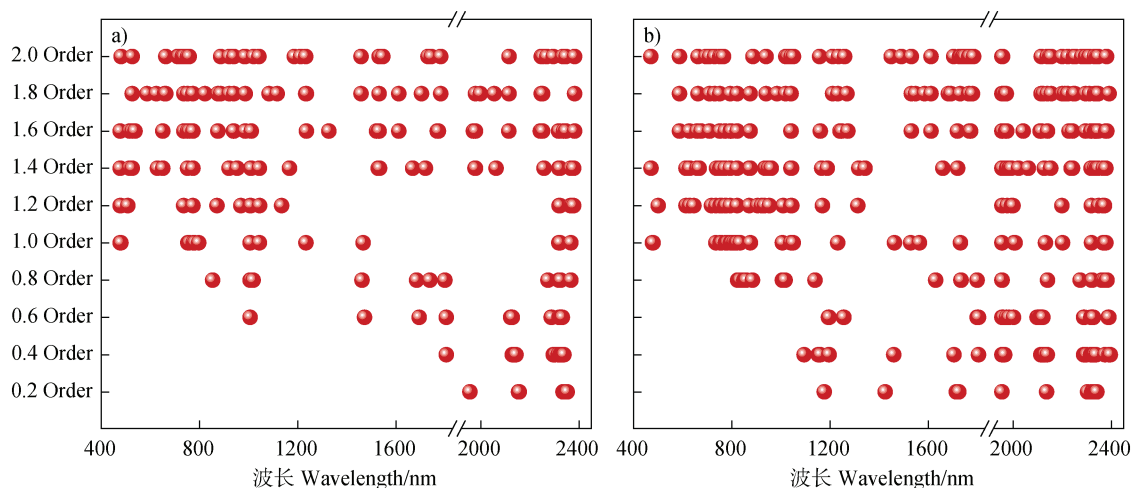


图 4 Boruta 算法选取的 Pb (a) 和 Zn (b) 特征波长分布

Fig. 4 Distribution of feature wavelengths selected by the Boruta Algorithm for Pb (a) and Zn (b)

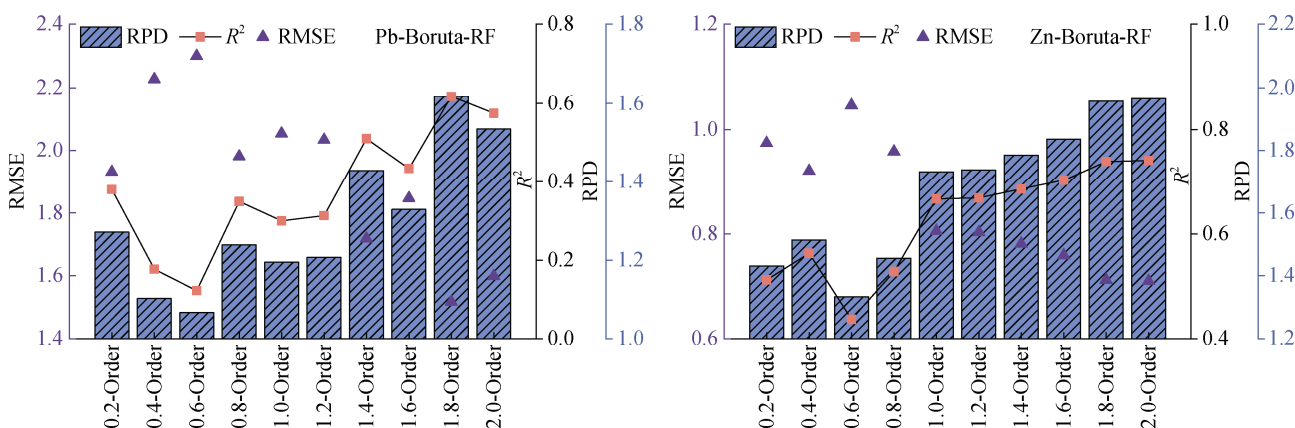


图 5 不同微分阶数下 RF 模型对 Pb 和 Zn 含量估算的各项指标

Fig. 5 Evaluation metrics for the estimation of Pb and Zn content using the RF model at different differential orders

高, 验证集的  $R^2$  为 0.741, RMSE 为 0.71, PRD 为 1.965, 预测能力中等; 在 0.6 阶时, 其 RPD 小于 1.4, 预测能力较差; 而其余阶数的 RPD 均介于 1.4 和 2 之间, 表明其对 Zn 含量的预测能力中等。

图 6 展示了 XGBoost 模型在不同微分阶数下对 Pb 和 Zn 估算的各项指标, 对于 Pb, 在 0.8 阶时模型精度最高, 其验证集的  $R^2$  为 0.852, RMSE 为 0.946, PRD 为 2.595, 其预测能力较好 ( $RPD > 2.0$ ); 0.4 阶、0.6 阶、1.2 阶、1.4 阶、1.6 阶、1.8 阶和 2 阶预测能力中等 ( $1.4 < RPD < 2.0$ ); 0.2 阶和 1.0 阶预测能力较差 ( $RPD < 1.4$ )。对于 Zn, 在 1.6 阶时模型精度最高, 其验证集的  $R^2$  为 0.877, RMSE 为 0.489, PRD 为 2.849; 0.8 阶、1 阶、1.4 阶、1.6 阶和 1.8 阶其预测能力较好 ( $RPD > 2.0$ ); 0.2 阶、0.4 阶、0.6

阶、1.2 阶和 2 阶预测能力中等 ( $1.4 < RPD < 2.0$ )。

为了更加直观地展示模型在预测 Pb 和 Zn 含量方面的性能, 绘制了 Pb 和 Zn 最佳反演模型验证集的实测值和预测值拟合图 (见图 7)。从图中可以看出, 预测值和实测值基本分布在 1:1 直线周围, 表明模型具有较高的预测精度。因此将 0.8 Order-Boruta-XGBoost 和 1.6 Order-Boruta-XGBoost 分别作为预测该矿区 Pb 和 Zn 含量的最佳反演模型, 可有效揭示研究区内铅锌含量的分布情况。

运用最佳反演模型对研究区内裸土区域 Pb 和 Zn 含量进行估算 (见图 8)。结果显示, 研究区西北部 Pb 和 Zn 含量明显高于其他区域, 是污染最为严重的区域; 从实地调查获取的信息 (见图 1) 为该结论提供了有力支撑, 西北部区域曾是历史遗留的

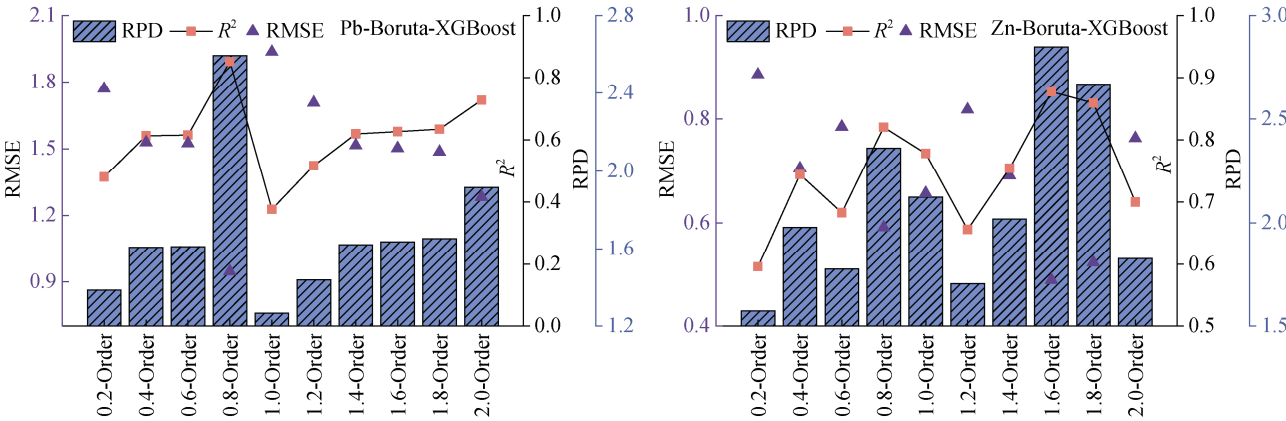


图 6 不同微分阶数下 XGBoost 模型对 Pb 和 Zn 含量估算的各项指标

Fig. 6 Evaluation metrics for the estimation of Pb and Zn content using the XGBoost model at different differential orders

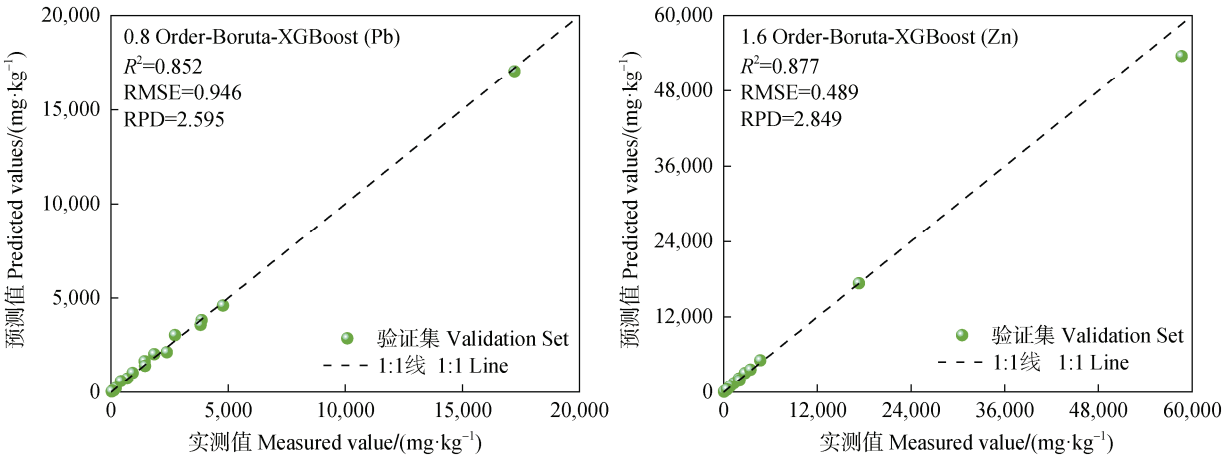


图 7 实测值和预测值的拟合图

Fig. 7 Fitting plot of the measured and predicted values

采矿和冶炼场地，堆积了大量废渣和废料，因此人类活动是导致该区域重金属污染严重的主要原因。将 Pb 和 Zn 的 Pearson ( $P<0.01$ ) 相关系数达到 0.82 (见表 1) 与反演结果的空间分布特征进行分析，两种重金属的高度相关性和空间分布一致性揭示了它们可能具有复合污染特征或同源性。以上均从不同角度验证了最佳反演模型的准确性和可靠性。

### 3 讨 论

本研究以云南会泽县矿山镇某铅锌尾矿区为例，采用 DS 算法结合室内光谱数据对 GF-5 影像光谱数据进行校正，消除多种环境因素对影像光谱的干扰。通过这一方法，旨在获得更加纯净的高光谱数据，从而更加准确地反演研究区内 Pb 和 Zn 含量

的空间分布。研究表明，使用 DS 算法能够消除土壤粒径、含水量和温度等环境因素对影像光谱反射率的影响，从而有效地提升了模型精度，这与 Zhang 等<sup>[13]</sup>的研究结果一致。

由于受人为活动的影响，土壤重金属含量通常呈偏态分布。在这种情况下直接利用原始光谱数据与重金属实测值构建反演模型，不仅会增加模型的复杂性和计算难度，还可能因为含量数据不符合正态分布而导致模型精度降低，无法准确反映研究区的重金属含量空间分布情况。使用 Box-Cox 转换能够使重金属含量服从正态分布，从而提高反演结果的准确性<sup>[15]</sup>。

在分数阶微分处理中，高阶微分能够突出信号中的高频变化，有助于揭示更多显著的特征波段，这些波段通常包含丰富的细节信息，有助于区分不



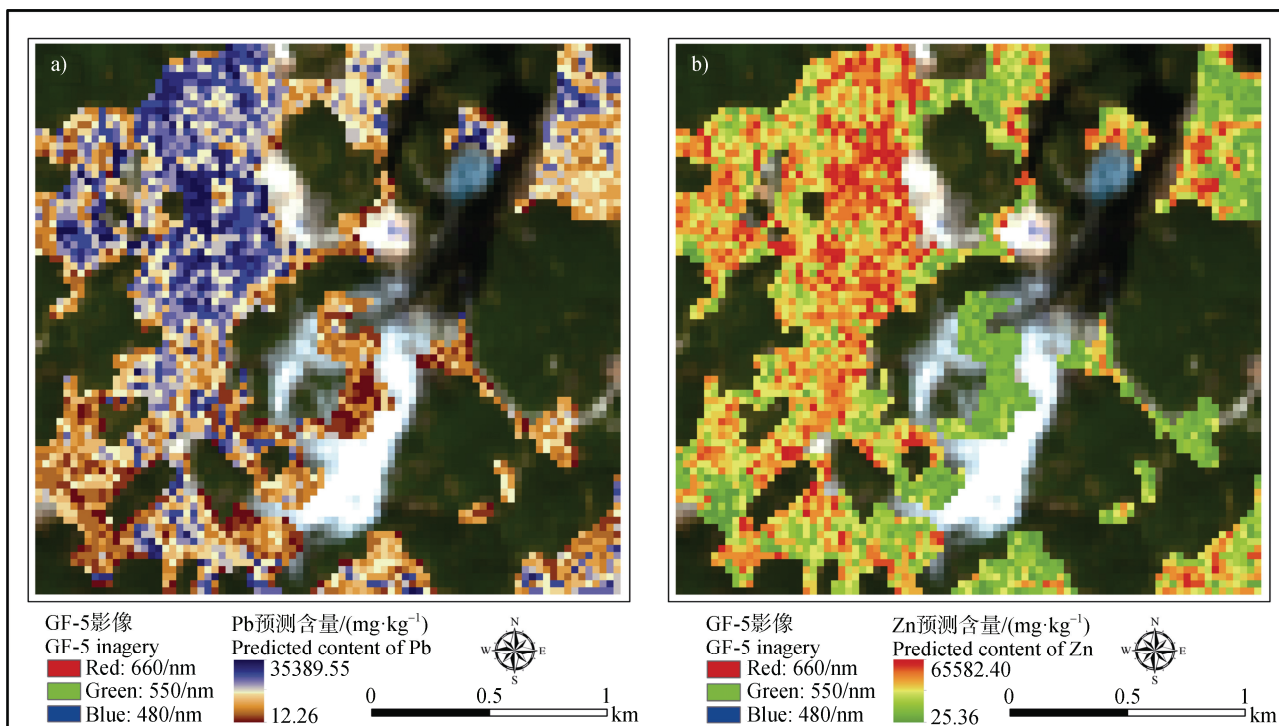


图 8 应用最佳反演模型估算的 Pb (a) 和 Zn (b) 含量分布图

Fig. 8 Distribution map of Pb (a) and Zn (b) content estimated by the optimal inversion model

同物质或元素的光谱特征。相比之下,低阶微分(0.2 阶~1 阶)则保留光谱的整体趋势和平滑变化,忽略了细微波动,导致许多波段的特征变化不明显,从而使用 Boruta 算法筛选出的特征波段数量较少。

RF 和 XGBoost 是两种被广泛应用的集成学习算法。RF 通过构建多棵决策树,并利用投票机制对数据进行预测,虽然在高维数据和处理多样化特征方面表现良好,但在面对复杂的非线性关系时,其效果通常不如梯度提升类算法。相比之下, XGBoost 是基于梯度提升树(GBDT)的改进算法,通过逐步拟合残差构建新树,并采用加权优化来增强模型性能;同时, XGBoost 还引入了正则化(L1/L2 正则化)、树的深度限制、列抽样等技术,能够有效防止过拟合,从而提高了模型的鲁棒性和精度。由于高光谱数据与重金属含量之间存在众多复杂的非线性关系, XGBoost 在此类任务中通常优于 RF,尤其是在处理复杂特征交互和非线性模式时表现更为出色。

## 4 结 论

(1) 通过 DS 算法建立室内测定光谱和 GF-5 影

像光谱之间的定量关系,从而有效消除土壤粒径、含水量和温度等环境因素对影像光谱数据的干扰。同时, Box-Cox 转换解决了重金属含量数据的偏态分布问题,使其更加接近正态分布。两种数据预处理方法共同为更加准确地反演矿区土壤中重金属含量奠定了坚实基础。(2) 使用分数阶微分构建的反演模型精度基本优于传统整数阶微分模型,表明不同阶数的分数阶微分能够增强土壤重金属光谱中的微弱信息,从而使 Boruta 算法能更好地捕捉到细节光谱特征,提升了模型精度。(3) 该研究区土壤 Pb 含量的最佳反演模型为 0.8 Order-Boruta-XGBoost, 而 Zn 含量的最佳反演模型为 1.6 Order-Boruta-XGBoost。这两种最佳反演模型在该矿区 Pb 和 Zn 含量反演中表现出较高的精度和较强鲁棒性,绘制的 Pb 和 Zn 含量空间分布图与实地调查结果相吻合。

## 参考文献 (References)

- [1] Wu Y F, Li X, Yu L, et al. Review of soil heavy metal pollution in China: Spatial distribution, primary sources, and remediation alternatives[J]. Resources, Conservation and Recycling, 2022, 181: 106261.
- [2] Li Y Z, Zhang J Z, Jia J Y, et al. Research progresses on farmland soil ecosystem multifunctionality[J]. Acta



- Pedologica Sinica, 2022, 59 (5): 1177—1189. [李奕赞, 张江周, 贾吉玉, 等. 农田土壤生态系统多功能性研究进展[J]. 土壤学报, 2022, 59 (5): 1177—1189.]
- [ 3 ] Chang C Y, Yang J, Wu J, et al. Deliberations on collaborative supervision of risk control and remediation for soil pollution of constructed land[J]. Acta Pedologica Sinica, 2025, 62 (5): 1—13. [常春英, 杨婕, 吴俭, 等. 建设用地土壤污染风险管控和修复联动监管的若干思考[J]. 土壤学报, 2025, 62 (5): 1—13.]
- [ 4 ] Wang Y J. Research progress and prospect on ecological disturbance monitoring in mining area[J]. Acta Geodaetica et Cartographica Sinica, 2017, 46 (10): 1705—1716. [汪云甲. 矿区生态扰动监测研究进展与展望[J]. 测绘学报, 2017, 46 (10): 1705—1716.]
- [ 5 ] Liu Z H, Lu Y, Peng Y P, et al. Estimation of soil heavy metal content using hyperspectral data[J]. Remote Sensing, 2019, 11 (12): 1464.
- [ 6 ] Jin M T, Yuan H, Liu B, et al. Review of the distribution and detection methods of heavy metals in the environment[J]. Analytical Methods, 2020, 12 (48): 5747—5766.
- [ 7 ] Zhang X, Ding S T, Cen Y, et al. Soil heavy metal Pb content estimation method by combining field spectra with laboratory spectra[J]. Geomatics and Information Science of Wuhan University, 2022, 47(9): 1479—1485. [张霞, 丁松滔, 岑奕, 等. 结合野外与实验室光谱的土壤 Pb 含量反演[J]. 武汉大学学报 (信息科学版), 2022, 47 (9): 1479—1485.]
- [ 8 ] Shi Y Y, Li J Y, Chu X L. Progress and applications of multivariate calibration model transfer methods[J]. Chinese Journal of Analytical Chemistry, 2019, 47 (4): 479—487. [史云颖, 李敬岩, 褚小立. 多元校正模型传递方法的进展与应用[J]. 分析化学, 2019, 47 (4): 479—487.]
- [ 9 ] Sun G Y, Fu H, Zhang A Z, et al. Singular spectrum analysis method for hyperspectral imagery feature extraction: A review and evaluation[J]. Acta Geodaetica et Cartographica Sinica, 2023, 52 (7): 1148—1163. [孙根云, 付航, 张爱竹, 等. 高光谱影像奇异谱分析特征提取方法: 综述与评价[J]. 测绘学报, 2023, 52 (7): 1148—1163.]
- [ 10 ] Chen L H, Lai J, Tan K, et al. Development of a soil heavy metal estimation method based on a spectral index: Combining fractional-order derivative pretreatment and the absorption mechanism[J]. Science of the Total Environment, 2022, 813: 151882.
- [ 11 ] Ding S T, Zhang X, Shang K, et al. Estimating soil heavy metal from hyperspectral remote sensing images base on fractional order derivative[J]. Journal of Remote Sensing, 2023, 27 (9): 2191—2205. [丁松滔, 张霞, 尚坤, 等. 基于分数阶微分的土壤重金属高光谱遥感图像反演[J]. 遥感学报, 2023, 27 (9): 2191—2205.]
- [ 12 ] Jiang Y H, Yan B, Zhuang Q Y, et al. Quantitative inversion model of soil heavy metals Zn and Ni based on fractional order derivative[J]. Spectroscopy and Spectral Analysis, 2024, 44 (10): 2850—2857. [蒋宇恒, 晏博, 庄清源, 等. 基于分数阶微分的土壤重金属锌和镍的定量反演模型研究[J]. 光谱学与光谱分析, 2024, 44 (10): 2850—2857.]
- [ 13 ] Zhang B, Guo B, Zou B, et al. Retrieving soil heavy metals concentrations based on GaoFen-5 hyperspectral satellite image at an opencast coal mine, Inner Mongolia, China[J]. Environmental Pollution, 2022, 300: 118981.
- [ 14 ] Mao J H, Zhao H Q, Jin Q, et al. Comparative study on the hyperspectral inversion methods for soil heavy metal contents in Hebei lead-zinc tailings reservoir areas[J]. Transactions of the Chinese Society of Agricultural Engineering, 2023, 39 (22): 144—156. [毛继华, 赵恒谦, 金倩, 等. 河北铅锌尾矿库区土壤重金属含量高光谱反演方法对比[J]. 农业工程学报, 2023, 39 (22): 144—156.]
- [ 15 ] Zhao H L, Gan S, Yuan X P, et al. Prediction of low Zn concentrations in soil from mountainous areas of central Yunnan Province using a combination of continuous wavelet transform and Boruta algorithm[J]. International Journal of Remote Sensing, 2023, 44 (15): 4753—4774.
- [ 16 ] Han R S, Wu P, Zhang Y, et al. New research progress in metallogenic theory for rich Zn-Pb- (Ag-Ge) deposits in the Sichuan-Yunnan-Guizhou Triangle (SYGT) area, southwestern Tethys[J]. Acta Geologica Sinica, 2022, 96 (2): 554—573. [韩润生, 吴鹏, 张艳, 等. 西南特提斯川滇黔成矿区富锗铅锌矿床成矿理论新进展[J]. 地质学报, 2022, 96 (2): 554—573.]
- [ 17 ] Chen W, Liu Q, Wang B, et al. Evaluation and source analysis of heavy metal pollution in cultivated soil of a lead-zinc mining area in Huize County, Yunnan Province, China[J]. Journal of Agro-Environment Science, 2024, 43 (5): 1036—1044. [陈文, 刘奇, 王豹, 等. 云南省会泽县某铅锌矿区耕地土壤重金属污染评价及来源解析[J]. 农业环境科学学报, 2024, 43 (5): 1036—1044.]
- [ 18 ] Benkhetto N, Brito da Cruz A M C, Torres D F M. A fractional Calculus on arbitrary time scales: Fractional differentiation and fractional integration[J]. Signal Processing, 2015, 107: 230—237.
- [ 19 ] Liu J L, Liu L, Ma X Y, et al. Spatial variability of soil salt in different soil layers at different scales[J]. Journal of Basic Science and Engineering, 2018, 26 (2): 305—312. [刘继龙, 刘璐, 马孝义, 等. 不同尺度不同土层土壤盐分的空间变异性研究[J]. 应用基础与工程科学学报, 2018, 26 (2): 305—312.]
- [ 20 ] Jalalinejad H, Tavakoli A, Zarmehi F. A simple and flexible modification of Grünwald-Letnikov fractional derivative in image processing[J]. Mathematical Sciences, 2018, 12 (3): 205—210.

(责任编辑: 檀满枝)