基于高光谱的尾矿有机质含量估测模型研究*

吴梓敬[1],谷海红[1,2,3,4,5],成少平[1],周征[6],任宇松[1],邵佳丽[1],艾艳君[1],田桂娥[1,2,3]

(1.华北理工大学 矿业工程学院, 河北 唐山市 063210;
2.河北省矿业开发与安全技术重点实验室, 河北 唐山市 063210;
3.河北省矿区生态修复产业技术研究院, 河北 唐山市 063210;
4.矿产资源绿色开发与生态修复协同创新中心, 河北 唐山市 063210;
5.唐山市资源与环境遥感重点实验室, 河北 唐山市 063210;
6.华北理工大学 机械工程学院, 河北 唐山市 063210)

摘要:矿产资源的过度开发加剧了矿区生态环境压力,尾矿作为含潜在重金属的固体废弃物,其堆存对土壤有机质(SOM)的破坏已成为生态修复的关键问题。基于高光谱遥感技术建立尾矿SOM监测体系,以河北唐山某矿区为研究对象,通过ASD地物光谱仪与GF-5卫星获取植被冠层高光谱数据,采集对应点位尾矿样本进行理化性质分析,结合Boruta算法筛选敏感波段,构建多元逐步线性回归(SMLR)、偏最小二乘回归(PLSR)、随机森林(RF)和极端梯度提升(XGBoost)等四类模型并进行精度验证。结果表明,ASD光谱数据采用XGBoost模型的预测效果最优($R^2=0.92$,$RMSE=0.37$),适用于精细化监测;GF-5卫星数据以XGBoost模型表现最佳($R^2=0.73$,$RMSE=0.53$),可满足大范围快速评估需求。研究证实,通过光谱数学变换与线性、非线性模型协同优化,可实现地面-空天多尺度SOM监测,较传统检测方法效率显著提升。研究结果可为矿区生态修复提供科学依据,促进矿区环境管理和矿业可持续发展。

关键词:尾矿有机质;高光谱技术;尾矿监测;估测模型
中图分类号:TD167 **文献标识码**:A **文章编号**:1005-2763(2026)01-0186-09

Study on Prediction Model of Tailings Organic Matter Content Based on Hyperspectral Technology

WU Zijing[1], *GU Haihong*[1,2,3,4,5], *CHENG Shaoping*[1], *ZHOU Zheng*[6], *REN Yusong*[1], *SHAO Jiali*[1], *AI Yanjun*[1], *TIAN Guie*[1,2,3]

(1.College of Mining Engineering, North China University of Science and Technology, Tangshan, Hebei 063210, China;
2.Hebei Provincial Key Laboratory of Mining Development and Safety Technology, Tangshan, Hebei 063210, China;
3.Hebei Industrial Technology Institute of Mine Ecological Remediation, Tangshan, Hebei 063210, China;
4.Collaborative Innovation Center for Green Development and Ecological Restoration of Mineral Resources, Tangshan, Hebei 063210, China;
5.Tangshan Key Laboratory of Resources and Environmental Remote Sensing, Tangshan, Hebei 063210, China;
6.College of Mechanical Engineering, North China University of Science and Technology, Tangshan, Hebei 063210, China)

Abstract: Over-exploitation of mineral resources aggravates the ecological environmental pressure in mining areas. As solid wastes containing potential heavy metals, tailings cause damage to soil organic matter (SOM) through stockpiling, which has become a key issue in ecological restoration. The SOM monitoring system of tailings was established based on hyperspectral remote sensing technology. Taking a mining area in Tangshan, Hebei Province as the research object, vegetation canopy hyperspectral data were obtained using an ASD ground spectrometer and GF-5 satellite, and tailings samples at the corresponding points were collected for physicochemical property analysis. After screening sensitive bands with the Boruta algorithm, four types of models such as multiple stepwise linear regression (SMLR), partial least squares regression (PLSR),

* **收稿日期**:2025-02-05
基金项目:国家自然科学基金项目(52204186);河北省自然科学基金项目(E2021209152);河北省中央引导地方科技发展资金项目(246Z7610G);唐山市科技研发平台培养计划项目(2020TS003b);唐山市科技计划项目(20130210b);河北省研究生创新资助项目(CXZZBS2023129)
作者简介:吴梓敬(2000—),男,福建福州人,硕士研究生,研究方向为环境生态修复。E-mail:zijing1264294689@163.com
通信作者:谷海红(1982—),女,河北唐山人,博士,副教授,研究方向为矿区生态修复与重建。E-mail:haihonggu1982@hotmail.com

random forest (RF) and extreme gradient boosting (XGBoost) were constructed and verified for accuracy. The results show that the XGBoost model has the best prediction effect on the ground ASD spectral data ($R^2=0.92$, $RMSE=0.37$), which is suitable for fine monitoring. The GF-5 satellite data perform best with the XGBoost model ($R^2=0.73$, $RMSE=0.53$), which meets the needs of large-scale rapid assessment. The study confirms that ground-space multi-scale SOM monitoring can be realized through the collaborative optimization of spectral mathematical transformations and linear and nonlinear models, with efficiency significantly enhanced compared with traditional detection methods. The research results can provide a scientific basis for ecological restoration of mining areas, promoting environmental management of mining area and the sustainable development of mining industry.

Key words: Tailings organic matter, Hyperspectral technology, Tailings monitoring, Prediction model

0 引言

矿产资源是国家经济发展和工业化的重要基础,对能源供应和基础设施建设具有不可替代的作用[1]。上世纪我国矿产资源开发以提升产能为主,环境保护关注度较低[2]。尾矿作为选矿生产排放的"废弃物",含细颗粒矿物、金属元素及有害物质,易通过雨水迁移和风蚀污染地下水及农田[3],其物理化学性质监测至关重要。尾矿中的有机质既可促进植物生长、改善土壤结构,又通过污染物相互作用影响环境特性[4]。传统监测手段难以及时获取动态信息,导致尾矿库的修复管理困难[5]。

遥感技术的发展为地表特征监测提供了新途径,该技术通过多光谱/高光谱成像系统,可非破坏性地获取土壤、植被等物理化学信息[6-7]。多光谱遥感通过波段组合反映土壤健康状况,高光谱遥感则利用精细光谱分辨率捕捉具体特征[8]。任建强等[9]通过作物冠层高光谱和 EO-1 Hyperion 高光谱卫星数据,对冬小麦地上干生物量进行反演。马磊等[10]应用铜锣山矿山公园往年的 GF-2 卫星遥感影像,监测植被覆盖度的变化,分析植被覆盖度恢复趋势及空间分布特征,从而推断矿山生态修复的效果。遥感技术还可预警土壤退化,为环境管理提供依据[11]。

高光谱技术在尾矿监测中具有独特优势,通过高分辨率光谱数据解析物质成分,可精准评估尾矿物理化学性质及污染特征[12]。现有研究多集中于农业土壤分析,张进等[13]通过无人机光谱构建土壤全氮预测模型。叶勤等[14]在实验室利用 ASD 地物光谱仪对横山县野外采集的土壤样品进行光谱测定,利用一阶导数光谱与土壤有机质(SOM)建立多元逐步线性回归(SMLR)模型,并对模型精度进行评价。但尾矿监测研究相对匮乏,其复杂成分对光谱反射产生显著影响,亟待特征提取技术的突破[15]。

高光谱技术兼具定点污染识别和大范围反演能力,通过卫星光谱数据可评估矿区环境风险[16]。虽在土壤盐分监测中已有成功案例,如王惠敏等[17]通过高分五号高光谱影像,针对新疆焉耆土壤中 5 种盐分含量,使用偏最小二乘、岭回归和支持向量机建立盐分估测模型,但尾矿有机质与重金属污染的高光谱分析仍处起步阶段。发展结合机器学习的高光谱模型,可提升监测精度,为尾矿治理提供科学支撑[19]。

本研究以河北唐山某矿区为研究对象,采用 GF-5 卫星和 ASD 地物光谱数据,结合光谱变换与指数分析,运用偏最小二乘回归(PLSR)、SMLR、RF 和极端梯度提升(XGBoost)算法构建尾矿有机质估算模型。通过对比不同尺度高光谱数据的建模效果,建立快速可靠的矿区有机质监测方法,为尾矿环境治理和资源利用提供技术依据,对实现矿山可持续管理和生态恢复具有重要实践意义。

1 材料与方法

1.1 研究区概况

本研究选用河北唐山某矿区作为研究对象,矿床类型为沉积变质铁矿床。该矿区采用露天开采方式,建有选矿厂、尾矿库、排土场等。由于矿区内部植被大量被破坏,生态环境质量显著下降。因此,自 2008 年起,该矿区在尾矿库、排土场等生态破坏严重区域持续实施植被恢复工程,结合区域气候、土壤条件等因素,筛选种植了刺槐、紫花苜蓿、黑麦草、狗尾草等先锋植物。

1.2 尾矿样品采集

本研究于 2019 年 7 月对研究区进行了为期 2 d 的采样工作。尾矿取样采用 S 形布点方法,同时使用手持 GPS 仪定位获取坐标信息,采集过程使用五点取样法,共计采样点 88 处。将采集的尾矿样品混合均匀后留存 1 kg,装入自封袋密封并带回实验室,于阴凉处自然风干后过 100 目筛,然后采用重铬酸钾-外加热法[20]测定尾矿的有机质(SOM)含量。

1.3 ASD 地物光谱仪的光谱采集

采用 ASD FieldSpec 4 地物光谱仪进行光谱数据采集。ASD 地物光谱仪涵盖可见光和红外波段,采样时间间隔为 0.2 s,重采样间隔为 1 nm,共有 2 151个输出波段,波长范围为 350～2500 nm。由于 GF-5 卫星无法直接获取被植被覆盖的尾矿光谱反射率,为了使试验结果具有对比性,本文选取植物冠层的三片叶片,ASD 地物光谱仪经预热 30 min 进行黑白板校正,每片子叶采集 5 条光谱曲线。采用 ASD 地物光谱仪自带的光谱处理软件 ViewSpecPro 进行光谱数据检查,剔除明显错误的光谱曲线,并求得算术平均值作为最终光谱曲线。

1.4 GF-5 卫星的光谱采集与预处理

GF-5 光谱波段范围达到 350～2 500 nm,具有 300 个光谱通道,光谱分辨率达到 5 nm。本研究所用的 GF-5 高光谱数据采集时间为 2019 年 7 月 21 日,标识号为 GF5_AHSI_E118.86_N40.19_20190721_006385_L10000051439,云量为 0。GF-5 高光谱数据经以下 6 步预处理。

(1) 坏波段剔除:采用 ENVI 5.3.1 软件去除受噪声及大气吸收干扰的波段;

(2) 辐射定标:通过辐射定标公式将 DN 值转换为辐射亮度;

(3) 坏线修复:利用 ENVI 的 Replace Bad Line 模块插值修复异常数据线;

(4) 条纹校正:应用全局去条纹法消除因传感器响应不均导致的条带噪声;

(5) 大气校正:采用 FLAASH 工具消除大气散射影响,还原地物真实反射率;

(6) 几何校正:以 Landsat-8 OLI 影像(2018 年 7 月)为基准,通过 Map-to-Map 方法选取河流交汇点等控制点,完成空间配准。

预处理后,提取采样点植被冠层端元光谱用于建模,确保数据可靠性与地物光谱真实性。

1.5 光谱变换

光谱数据常受到环境噪声和测量误差的影响,对光谱数据进行数学变换有助于提升数据的分析质量和准确性。通过数学变换,可以有效减少光谱数据的基线漂移和噪声,降低数据冗余,并突出光谱的特征信息,提高信号的清晰度[21]。本研究采用连续统去除(SG)、反射率一阶微分(FD)和反射率二阶微分(SD)对 ASD 和 GF-5 数据进行数学变换,以提取某些光谱的隐藏信息。

1.6 特征选择算法

高光谱数据包含大量波段,覆盖可见光至近红外的宽光谱范围。Boruta 算法是一种基于随机森林(Random Forest, RF)模型的特征选择方法,适用于处理这种高维度、高相关性的数据。该算法通过随机森林模型训练原始数据,并引入影子特征(随机打乱原始特征生成),比较原始特征与影子特征的重要性。若某一原始特征的重要性显著高于影子特征的最大重要性,则视该特征为重要特征,因此 Boruta 算法能自动处理冗余和不相关特征,适合于处理尾矿的复杂高光谱数据。

1.7 光谱指数

合适的光谱指数对模型预测结果有很大影响,由于植被覆盖会导致无法直接获取尾矿的光谱反射率,但尾矿条件的差异会造成植被特征差异,因此可通过作物冠层的光谱指数监测作物在不同生育期的生长特征。这意味着选用合适的光谱指数对提升模型精度至关重要,本研究采用的光谱指数[22-25]及其计算公式见表 1。

表 1 光谱指数及其计算公式

Table 1 Spectral index and its calculation formula

光谱指数	计算公式
归一化植被指数(Normalized Difference Vegetation Index,NDVI)	$(NIR-R)/(NIR+R)$
归一化盐分指数(Normalized Difference Salinity Index,NDSI)	$(R-NIR)/(R+NIR)$
增强植被指数(Enhanced Vegetation Index,EVI)	$G\times(NIR-R)/(NIR+C_1\times R-C_2\times B+L)$
差值植被指数(Difference Vegetation Index,DVI)	$NIR-R$
比值植被指数(Ratio Vegetation Index, RVI)	NIR/R
三角植被指数(Triangular Vegetation Index, TVI)	$60\times(NIR-G)-100\times(R-G)$
绿色归一化植被指数(Green Normalized Difference Vegetation Index,GNDVI)	$(NIR-G)\times(NIR+G)$
土壤调节植被指数(Soil Adjusted Vegetation Index,SAVI)	$(NIR-R)\times(1+L)/(NIR+R+L)$
改进土壤调节植被指数(Optimized Soil Adjusted Vegetation Index,OSAVI)	$1.16\times(NIR-R)/(NIR+R+0.16)$
可见光大气阻抗指数(Visible Atmospherically Resistant Index,VARI)	$(G-R)/(G+R-B)$
叶绿素植被指数(Green Chlorophyll Vegetation Index,GCVI)	$(NIR/G)-1$
归一化差值指数(Normalized difference Index,NDI)	$(G-R)/(G+R)$

注:NIR 表示近红外波段;R 表示红波段;G 表示绿波段;B 表示蓝波段;$C_1=6$,$C_2=7.5$;L 为土壤调节系数,取值范围为 0～1。

1.8 模型构建与精度验证

本研究针对尾矿库高光谱数据的复杂性,选取四种线性和非线性方法构建有机质预测模型。传统线性模型假设输入输出线性相关,但难以捕捉复杂非线性关系,因此采用 RF、XGBoost 等非线性方法进行对比分析。

RF 通过集成多棵决策树提升预测性能,采用袋外抽样和随机选择特征降低过拟合风险,并通过投票或平均机制聚合结果,其双重随机性设计在高维数据处理中优势显著。XGBoost 作为梯度提升算法的优化版本,通过序列化迭代构建决策树修正残差,创新性引入 L_1/L_2 正则化控制复杂度,结合并行计算与内存优化以提升效率,在工程实践中表现卓越。PLSR 通过降维解决高维数据多重共线性问题,将变量投影至低维潜在空间,并最大化变量间协方差提取有效成分,适用于变量数超样本量的场景。SMLR 采用逐步筛选机制平衡模型复杂度与模型解释性,但对噪声敏感,需结合交叉验证进行优化。

采用决定系数(R^2)和均方根误差($RMSE$)评估模型。R^2(取值范围 0 ~ 1)用于衡量模型解释方差的能力,R^2 值越接近 1,则表明预测性能越强;$RMSE$ 反映预测值与实测值的平均偏差程度,值越小,则模型精度越高。两指标互补,可全面评价模型性能。相关计算公式如下:

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (\hat{Y} - Y_i)^2}{\sum_{i=1}^{N} (Y_i - \bar{Y})^2} \tag{1}$$

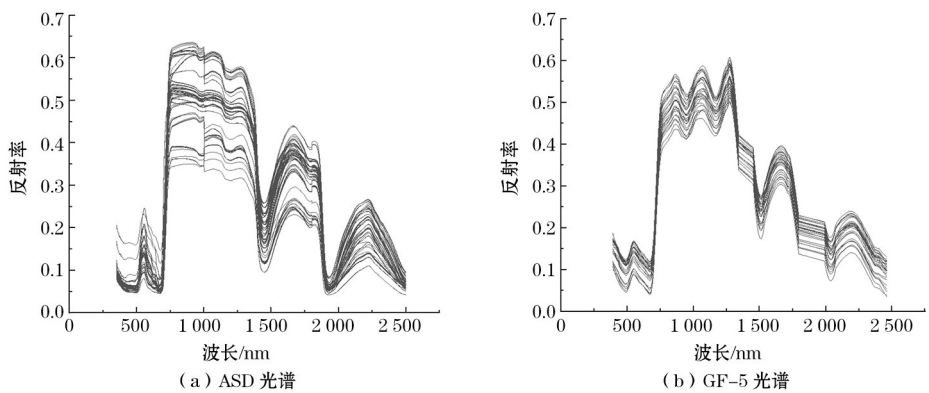
$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{Y} - Y_i)^2} \tag{2}$$

式中:Y_i 为样本集指标含量的实测值;$\hat{Y}$ 为样本集指标含量的预测值;$\bar{Y}$ 为样本集指标含量实测值的平均值。

2 模型结果与分析

2.1 数据预处理结果分析

应用 ASD 地物光谱仪和 GF-5 卫星采集了研究区的原始光谱反射曲线(如图 1 所示)。采集的 ASD 原始光谱间的特征差异性更明显,通过对比两组曲线,两种尺度下的光谱曲线具有多个相似特征。在 400~720 nm 的部分可见光范围内,平均反射率维持在 0.2 以下,光谱展现出“两谷一峰”的特征。在 670~760 nm 波段出现“红边”,反射率显著提升。在 800~1 300 nm 范围内,光谱曲线呈现波浪状起伏趋势。而当波长大于 1 300 nm 时,曲线迅速下滑,且在波长 1 400 nm 和 1 900 nm 附近形成了两个明显的吸收谷。



(a) ASD 光谱　　(b) GF-5 光谱

图 1 ASD 地物光谱仪和 GF-5 卫星采集的尾矿库原始光谱反射曲线

Fig.1 Original spectral reflection curves of tailings pond collected by ASD groud spectrometer and GF-5 satellite

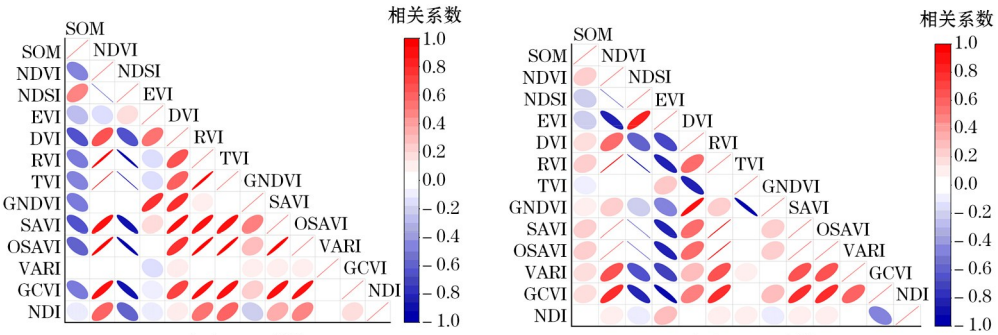
2.2 特征选择结果

本研究采用 Boruta 特征选择算法,实现对复杂光谱信息的精准筛选与解析。研究区尾矿经光谱变换后,ASD 的有机质特征光谱区间为 741 ~ 980 nm、1 041~1 174 nm、1 231~1 377 nm 及 1 589~1 833 nm,共 133 个波段。GF-5 的有机质特征光谱区间为 899.38~1 122.75 nm、1 596.23~1 780.18 nm 及 2 159.31~2 218.30 nm,共 38 个波段。ASD 波段数量基数较大,所筛选出的特征波段数量多于 GF-5。ASD 与 GF-5 的有机质光谱敏感波段具有相似性,敏感区域重合度为 74%。

2.3 相关性分析

将 SOM 与选定的光谱指数进行相关性分析,判定光谱指数建模的可行性,并去除相关性较弱的光谱指数,以保证后续预测模型的准确性,如图 2 所示。在 ASD 光谱中,SOM 与 SAVI、SOM 与 DVI

的相关性最强,相关系数为-0.65和-0.67,其次为SOM与RVI、SOM与OSAVI,相关系数为-0.58和-0.51,在剩余变量中,SOM与VARI、SOM与NDI的相关性较弱。在GF-5光谱中,光谱指数与SOM的相关性有所减弱,其中NDVI、OSAVI、SAVI、RVI等4个光谱指数与SOM的相关性最强,而NDI、TVI、GNDVI等3个光谱指数与SOM的相关性最弱,剔除不参与后续建模。



(a) ASD光谱　　(b) GF-5光谱

图2　尾矿SOM含量与光谱指数间的相关性分析

Fig.2　Correlation analysis between SOM content and spectral index of tailings

2.4　建模结果与模型性能指标

本研究通过R语言构建了4种不同的模型(RF、XGBoost、PLSR和SMLR)。为了确保结果的可靠性与有效性,采用系统化的参数调整策略,并通过留一交叉验证对模型的泛化误差进行评估,不同建模方法与光谱变换组合下的尾矿SOM含量预测模型精度见表2。

RF模型在SD、SG、FD和SI建模中均有良好的预测效果,在ASD光谱变换中除原始光谱(S)外,其余变换的R^2值均能达到0.7以上,其中SG-RF的R^2、*RMSE*分别为0.84和0.44,能够较为准确地进行预测;在GF-5光谱中,RF模型的R^2值有所降低,总体R^2值在0.6左右。XGBoost模型整体预测水平同样较为稳定,其中,在ASD光谱中,XGBoost-FD为最优预测模型,R^2、*RMSE*分别为0.92和0.37;在GF-5光谱中,相较于ASD光谱建模,XGBoost模型的精度有所下降,总体精度低于RF模型,但XGBoost-FD为最优GF-5预测模型,R^2、*RMSE*分别为0.73和0.53。相较于RF和XGBoost模型,PLSR模型的精度有所降低,在S、SG、FD、SI中,R^2值均在0.5左右,不适用于尾矿SOM的预测,但在SD变换中,PLSR模型的预测效果远高于其他变换,ASD和GF-5光谱SD变换的R^2值分别为0.79和0.64,可作为大范围预测的备选模型之一。在所选用的4种模型中,SMLR模型整体预测精度最差,在ASD和GF-5光谱中,除ASD光谱中的SD变换外,其余变换的R^2值均小于0.5,不具备预测SOM的能力,但与PLSR模型相似,SMLR模型在SD变换中的预测精度均高于其他变换,可能是线性回归模型对于预测SD变换的光谱有一定的优势。

表2　不同建模方法与光谱变换组合下的尾矿SOM含量估测模型精度

Table 2　Accuracy of prediction model of SOM content in tailings under different modeling methods and spectral transformation combinations

建模方法	ASD光谱			GF-5光谱		
	反射率及变换	R^2	*RMSE*	反射率及变换	R^2	*RMSE*
RF	S	0.68	0.76	S	0.51	0.80
	SG	0.84	0.44	SG	0.64	0.82
	FD	0.78	0.46	FD	0.65	0.54
	SD	0.72	0.69	SD	0.64	0.90
	SI	0.75	0.94	SI	0.51	1.10
XGBoost	S	0.69	0.81	S	0.48	0.79
	SG	0.78	0.57	SG	0.57	0.51
	FD	0.92	0.37	FD	0.73	0.53
	SD	0.69	0.77	SD	0.59	0.89
	SI	0.66	0.63	SI	0.43	0.94
PLSR	S	0.51	0.68	S	0.46	1.01
	SG	0.49	0.77	SG	0.43	1.06
	FD	0.51	0.68	FD	0.56	0.61
	SD	0.79	0.34	SD	0.64	0.50
	SI	0.41	0.82	SI	0.44	1.10
SMLR	S	0.25	1.69	S	0.38	0.78
	SG	0.21	2.07	SG	0.28	1.52
	FD	0.32	1.19	FD	0.44	1.84
	SD	0.64	1.16	SD	0.45	0.52
	SI	0.48	0.90	SI	0.22	1.60

注:S为原始光谱;SG为原始光谱连续统去除变换;FD为原始光谱一阶微分变换;SD为原始光谱二阶微分变换;SI为筛选的光谱指数。

图 3 为 ASD 与 GF-5 光谱最优预测值与实测值的验证散点图,通过拟合实测值与预测值的线性方程,能够更加直观地分析各个模型的预测精度。由图 3 可知,ASD 光谱中模型的预测精度整体大于 GF-5,可能与实测光谱的准确性更高有关,且不同反射率数学变换的最优模型大体也具有一致性。

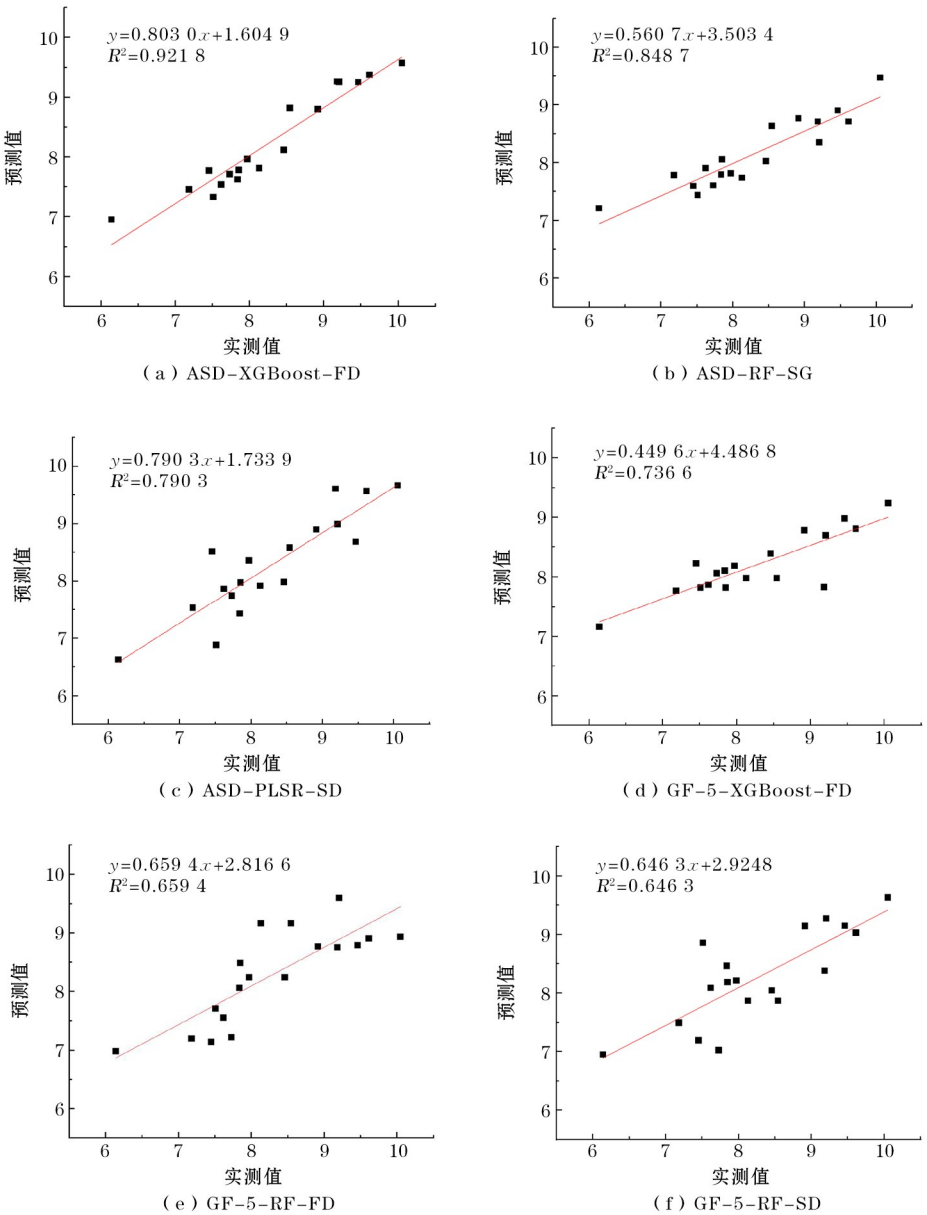


图 3 ASD 与 GF-5 光谱最优预测值与实测值的验证散点图

Fig.3 Scatter plots for verifying the optimal prediction values of ASD and GF-5 spectra with measured values

通过统计训练结果与性能评价指标发现,在 ASD 光谱变换中,ASD-XGBoost-FD 模型最为精确稳定,其次为 ASD-RF-SG 和 ASD-PLSR-SD 模型,这些模型均具有良好的 SOM 预测能力。在 GF-5 光谱变换中,最优预测模型是 GF-5-XGBoost-FD,其次是 GF-5-RF-FD 和 GF-5-RF-SD。在数据变换对应的模型中,ASD 光谱的整体预测精度排序为 RF>XGBoost>PLSR>SMLR,GF-5 光谱的整体预测精度排序为 XGBoost>RF>PLSR>SMLR。

3 讨论

尾矿是选矿作业的废弃产物,有用目标组分含量较低,但其成分复杂,传统的遥感数据分析方法难以有效提取和分析尾矿中的有用信息,且数据量大,尤其是高光谱数据中存在大量噪声和冗余信息,增加了研究的难度。由于缺乏有效的特征提取方法,

导致对尾矿库 SOM 含量等重要指标的预测不够准确,进而影响了环境治理和决策。而 Boruta 算法的全自动化过程和鲁棒性使其特别适用于处理尾矿高光谱这种高维度和复杂的数据集,可以更好地识别真正具有预测能力的光谱特征,过滤掉噪声和无关信息,从而提高了模型的准确性和稳定性[26]。因此,本研究选用 Boruta 算法进行特征光谱筛选,共筛选出 ASD 数据的特征光谱 133 个,GF-5 数据的特征光谱 38 个,去除了低敏感度数据,提升了模型运行效率。

在所选择的模型训练结果中,FD 和 SD 变换在光谱建模中表现出更优越的效果。因为这两种变换能够有效地减少基线漂移和背景噪声,从而显著增强光谱信号的分辨率和稳定性[27]。除了这些传统的变换方法外,SG 变换通过识别光谱中的连续趋势去除基线漂移和背景噪声,也具有不错的建模效果。虽然光谱指数与 SOM 含量的建模效果不如以上方法,但光谱指数仍可以作为有效的特征参数,反映尾矿 SOM 含量与光谱之间的响应关系。

贾增慧等[27]通过多种线性与非线性回归方法,利用高光谱数据对玉米产量进行估产,非线性回归建模效果均优于线性回归模型。LI 等[28]通过分析田间试验中的叶片高光谱反射率数据,对叶绿素含量进行无损预测,非线性模型如支持向量机(SVM)、RF 等均具有良好的预测性能。在矿区有机质估测结果中,非线性回归模型同样拥有较高的精度,非线性学习方法(如 XGBoost 和 RF)模型的预测效果明显优于传统的线性方法(如 PLSR 和 SMLR)。XGBoost 和 RF 模型通过构建大量决策树并结合其结果,能够更好地捕捉光谱数据中的复杂非线性关系[29]。这些模型的优势在于其强大的非线性拟合能力和鲁棒性,能够有效处理高维度、非线性和噪声较大的数据。

相较之下,PLSR 模型虽然在解释某些变量间相关性方面具有独特价值,但在本研究设定的条件下,其预测性能未能达到 RF 和 XGBoost 模型的水平。线性模型往往无法充分捕捉数据中的复杂模式和非线性特征,进而导致预测精度偏低[30]。SMLR 模型虽能有效应对多重共线性问题,但在模型精度和泛化能力上的表现亦未超越前两者。因此,在尾矿 SOM 的预测中,非线性模型(如 XGBoost 和 RF)能够提供更高的预测准确性和稳定性,表现出更强的预测能力。

4 结论

本研究以环境污染风险高的尾矿库为研究对象,根据 ASD 地物光谱仪、GF-5 卫星获取的高光谱数据,结合不同光谱变换与光谱指数,使用 PLSR、SMLR、RF 和 XGBoost 等 4 种方法建立尾矿 SOM 估测模型,得出如下的主要结论。

(1) 在研究区尾矿 SOM 的高光谱预测模型研究中,采用一阶、二阶微分变换等数据预处理方法,结合 XGBoost、RF 等非线性建模方法,可以显著提升模型的性能。这些预测方法对于 ASD 与 GF-5 的高光谱数据均适用。

(2) 在模型预测结果中,ASD 数据模型精度较高,最优模型 ASD-XGBoost-FD($R^2=0.92$,$RMSE=0.37$)能够进行较为精确的尾矿 SOM 估测,适用于精细化的监测与分析。相对而言,GF-5 数据模型精度较低,最优模型 GF-5-XGBoost-FD($R^2=0.73$,$RMSE=0.53$)只能进行大范围的粗略估测,适用于宏观的区域性监测。研究结果可为矿山环境修复及动态监测提供参考。

(3) 尾矿 SOM 含量受多种因素影响,如尾矿成分、水分条件、植被覆盖等,单一模型往往难以完全考虑所有影响因素的复杂交互作用,因此探索应用深度学习等新技术,以提升模型的自动化能力和精确度,并加强模型结果的解释性,使决策者能够理解模型的预测依据,进而提供实际可操作的尾矿管理建议。

参考文献(References):

[1] 王根锁,刘宏磊,武强,等.碳中和背景下废弃矿山环境正效应资源化开发利用[J].煤炭科学技术,2022,50(6):321-328.
WANG Gensuo, LIU Honglei, WU Qiang, et al. Resource development and utilization of positive environmental impacts of abandoned mines under carbon neutrality[J]. Coal Science and Technology,2022,50(6):321-328.

[2] 彭丽娜,施木俊,胡浩,等.固体废弃物在矿山生态修复中的应用研究[J].资源节约与环保,2023(6):45-48.
PENG Li'na, SHI Mujun, HU Hao, et al. Research on the application of solid waste in ecological restoration of mines [J]. Resources Economization & Environmental Protection, 2023(6):45-48.

[3] 胡海峰,廉旭刚,蔡音飞,等.山西黄土丘陵采煤沉陷区生态环境破坏与修复研究[J].煤炭科学技术,2020,48(4):70-79.
HU Haifeng, LIAN Xugang, CAI Yinfei. Study on ecological environment damage and restoration for coal mining-subsided area in loess hilly area of Shanxi Province [J]. Coal Science and Technology,2020,48(4):70-79.

[4] LIU S L, LI W P, QIAO W, et al. Zoning method for mining-induced environmental engineering geological patterns considering the degree of influence of mining activities on phreatic aquifer[J].Journal of Hydrology,2019,578:124020.

[5] 晏闻博,柳丹,彭丹莉,等.重金属矿山生态治理与环境修复技术进展[J].浙江农林大学学报,2015,32(3):467-477.
YAN Wenbo, LIU Dan, PENG Danli, et al. Technology advances of ecological restoration and environmental remediation of heavy metal mines[J]. Journal of Zhejiang A & F University,2015,32(3):467-477.

[6] 刘彦平,罗晴,程和发.高光谱遥感技术在土壤重金属含量测定领域的应用与发展[J].农业环境科学学报,2020,39(12):2699-2709.
LIU Yanping, LUO Qing, CHENG Hefa. Application and development of hyperspectral remote sensing technology to determine the heavy metal content in soil[J]. Journal of Agro-Environment Science,2020,39(12):2699-2709.

[7] 童庆禧,张兵,张立福.中国高光谱遥感的前沿进展[J].遥感学报,2016,20(5):689-707.
TONG Qingxi, ZHANG Bing, ZHANG Lifu. Current progress of hyperspectral remote sensing in China[J]. National Remote Sensing Bulletin,2016,20(5):689-707.

[8] 么嘉棋,常兔宇,王梦然,等.新一代水文水资源监测卫星SWOT数据特征、应用与展望[J].地球科学进展,2024,39(4):374-390.
YAO Jiaqi, CHANG Huanyu, WANG Mengran, et al. Characteristics, application, and prospects of a new generation hydrological and water resources monitoring satellite: SWOT[J]. Advances in Earth Science, 2024, 39(4):374-390.

[9] 任建强,吴尚蓉,刘斌,等.基于Hyperion高光谱影像的冬小麦地上干生物量反演[J].农业机械学报,2018,49(4):199-211.
REN Jianqiang, WU Shangrong, LIU Bin, et al. Retrieving winter wheat above-ground dry biomass based on Hyperion hyperspectral imagery[J]. Transactions of the Chinese Society for Agricultural Machinery,2018,49(4):199-211.

[10] 马磊,李成,司洪涛,等.矿山公园植被恢复动态监测及影响因素——以渝北铜锣山为例[J].中国水土保持科学(中英文),2023,21(6):111-120.
MA Lei, LI Cheng, SI Hongtao, et al. Dynamic monitoring and influencing factors of vegetation restoration on a mine park: a case study of Tongluo mountain[J]. Science of Soil and Water Conservation,2023,21(6):111-120.

[11] HU K, FENG X Y, ZHANG Q, et al. Review of satellite remote sensing of carbon dioxide inversion and assimilation [J].Remote Sensing,2024,16(18):3394.

[12] ZHAO S H, WANG Q, LI Y, et al. An overview of satellite remote sensing technology used in China's environmental protection[J]. Earth Science Informatics, 2017,10(2):137-148.

[13] 张进,吴春笃,张波,等.基于无人机多光谱的土壤全氮含量反演模型研究[J].江苏农业科学,2024,52(20):32-41.
ZHANG Jin, WU Chundu, ZHANG Bo, et al. Inversion model of soil total nitrogen content based on UAV multi-spectrum[J]. Jiangsu Agricultural Sciences,2024,52(20):32-41.

[14] 叶勤,姜雪芹,李西灿,等.基于高光谱数据的土壤有机质含量反演模型比较[J].农业机械学报,2017,48(3):164-172.
YE Qin, JIANG Xueqin, LI Xican, et al. Comparison on inversion model of soil organic matter content based on hyperspectral data[J]. Transactions of the Chinese Society for Agricultural Machinery,2017,48(3):164-172.

[15] 余龙,李军,贺霖,等.高光谱图像类别独立的域适应分类[J].遥感学报,2024,28(3):610-623.
YU Long, LI Jun, HE Lin, et al. Class-independent domain adaptation for hyperspectral image classification[J]. National Remote Sensing Bulletin,2024,28(3):610-623.

[16] 陈珠琳,李添雨,张耀方,等.基于高光谱数据的密云水库水生态空间地物精细分类[J].空间科学学报,2024,44(1):103-113.
CHEN Zhulin, LI Tianyu, ZHANG Yaofang, et al. Land cover classification from hyperspectral data in the water ecological space of Miyun Reservoir[J]. Chinese Journal of Space Science,2024,44(1):103-113.

[17] 王惠敏,于磊,徐凯磊,等.基于高分五号高光谱影像的干旱区盐渍化土壤盐分含量估算[J].光谱学与光谱分析,2023,43(7):2278-2286.
WANG Huimin, YU Lei, XU Kailei, et al. Estimation of salt content of saline soil in arid areas based on GF-5 hyperspectral image[J]. Spectroscopy and Spectral Analysis,2023,43(7):2278-2286.

[18] JIAO L L, SUN W W, YANG G, et al. A hierarchical classification framework of satellite multispectral/hyperspectral images for mapping coastal wetlands[J]. Remote Sensing,2019,11(19):2238.

[19] 鲍士旦.土壤农化分析[M].3版.北京:中国农业出版社,2000.
BAO Shidan. Soil agrochemical analysis[M]. 3rd edition. Beijing:China Agricultural Press,2000.

[20] LI W Y, XIANG Y Z, LIU X C, et al. Estimation of soil moisture content based on fractional differential and optimal spectral index[J].Agronomy,2024,14(1):14010184.

[21] LIU H H, GUO B, YANG X C, et al. High spatiotemporal resolution vegetation index time series can facilitate enhanced remote sensing monitoring of soil salinization[J]. Plant and Soil,2024,510(1):1-23.

[22] 姚昊昕,李清泉,赵亮,等.淮河流域暖季极端高温干旱复合事件的演变特征及其与气候和植被的关系[J].生态学报,2024,44(13):5596-5608.
YAO Haoxin, LI Qingquan, ZHAO Liang, et al. Evolutionary characteristics of extreme high temperature and drought composite events in the warm season of the Huaihe River Basin and their relationship with climate and vegetation[J]. Acta Ecologica Sinica, 2024, 44(13):5596-5608.

[23] 王欣怡,王昌昆,马海艺,等.基于双时相卫星遥感光谱指数估算土壤有机质含量[J].土壤,2023,55(5):1106-1113.
WANG Xinyi, WANG Changkun, MA Haiyi, et al. Estimation of soil organic matter content based on dual temporal satellite remote sensing spectral index[J]. Soils, 2023,55(5):1106-1113.

[24] QUAN Z J, CHENG Y X, TSUBO M, et al. Sensitivity and regulation factors of soil organic carbon content in steppe and desert-steppe grasslands of the Mongolian Plateau[J].Plant and Soil,2024,509(1):1-17.

[25] GUO H X, WU F, YANG K, et al. Sentinel-2 multispectral satellite remote sensing retrieval of soil cu content changes at different pH levels[J].Agronomy,2024, 14(10):2182.

[26] NIAN Y, SU X, YUE H, et al. Estimation of the rice aboveground biomass based on the first derivative spectrum and Boruta algorithm[J].Frontiers in Plant Science,2024, 15:1396183.

[27] 贾增慧,张继真,郝航,等.基于无人机高光谱影像的田块尺度玉米估产与生育时期优选[J].农业资源与环境学报,2024,41(6):1-14.
JIA Zenghui, ZHANG Jizhen, HAO Hang, et al. Field-scale maize yield estimation and growth stages optimization based on UAV hyperspectral image [J]. Journal of Agricultural Resources and Environment,2024,41(6):1-14.

[28] LI D, HU Q Q, RUAN S Q, et al. Utilizing hyperspectral reflectance and machine learning algorithms for non-destructive estimation of chlorophyll content in citrus leaves [J].Remote Sensing,2023,15(20):15204934.

[29] NAVROZIDIS I, PANTAZI X E, LAGOPODI A, et al. Application of machine learning for disease detection tasks in olive trees using hyperspectral data[J]. Remote Sensing, 2023,15(24):15245683.

[30] LIU Y, FENG H, FAN Y, et al. Improving potato above ground biomass estimation combining hyperspectral data and harmonic decomposition techniques [J]. Computers and Electronics in Agriculture,2024,218:108699.