

杨吉存, 郭兵, 韩保民. 基于 CR-FOD 变换 XGBoost 模型的高光谱土壤盐分反演及可解释性分析[J]. 土壤学报, 2026, YANG Jicun, GUO Bing, HAN Baomin. Hyperspectral Soil Salinity Inversion and Interpretability Analysis Based on CR-FOD Transform and XGBoost Model[J]. Acta Pedologica Sinica, 2026,

# 基于 CR-FOD 变换 XGBoost 模型的高光谱土壤盐分反演及可解释性分析\*

杨吉存, 郭兵<sup>†</sup>, 韩保民

(山东理工大学建筑工程与空间信息学院, 山东淄博 255000)

**摘要:** 本研究以黄河三角洲东营市为研究区, 基于高光谱数据开展土壤盐分估算。对原始光谱进行 Savitzky-Golay (S-G) 滤波和多元散射校正 (Multiplicative Scatter Correction, MSC) 预处理后, 进行倒数 (1/R)、倒数对数 (log(1/R)) 与连续统去除 (Continuum Removal, CR) 变换, 并结合分数阶微分 (Fractional Order Derivative, FOD) 处理, 构建 10 种典型二维光谱指数作为特征。在此基础上, 采用贝叶斯优化 (Bayesian Optimization, BO) 方法训练偏最小二乘回归 (Partial Least Squares Regression, PLSR)、卷积神经网络 (Convolutional Neural Network, CNN)、极限梯度提升 (eXtreme Gradient Boosting, XGBoost) 和支持向量机 (Support Vector Machine, SVM) 4 种模型。结果表明: 连续统去除组合二阶微分后的归一化差值指数 (Normalized Difference Index, NDI) 与土壤含盐量相关性最高 ( $|r|=0.91$ ); 基于该变换的 XGBoost 模型反演性能最优, 测试集决定系数 ( $R^2$ )、均方根误差 (Root Mean Square Error, RMSE) 与相对分析误差 (Residual Prediction Deviation, RPD) 分别达 0.94、0.85  $\text{g}\cdot\text{kg}^{-1}$  和 4.33; SHAP (SHapley Additive exPlanation) 可解释性分析进一步揭示, 广义指数 1 (Generalized Index 1, GDI 1) 为模型中最重要的特征。本研究为区域盐渍化高光谱监测提供了有效方法支持。

**关键词:** 光谱变换; 分数阶微分; 光谱指数; 盐渍化; XGBoost; 东营市

中图分类号: X144 文献标志码: A

## Hyperspectral Soil Salinity Inversion and Interpretability Analysis Based on CR-FOD Transform and XGBoost Model

YANG Jicun, GUO Bing<sup>†</sup>, HAN Baomin

(School of Civil Engineering and Geomatics, Shandong University of Technology, Zibo, Shandong 255000, China)

**Abstract:** 【Objective】 Under the global context of climate change and anthropogenic impacts, soil salinization has become increasingly severe. However, traditional salinization monitoring suffers from being time-consuming, labor-intensive, and costly. Hyperspectral-based salinization monitoring often relies on single mathematical transformations and one-dimensional spectral information, while also exhibiting poor model interpretability. Research utilizing combined spectral transformations to construct spectral indices for salinization estimation urgently requires in-depth exploration. Thus, this study aims to fully exploit spectral information, enhance data sensitivity, and establish a high-precision, interpretable salinization inversion model based on spectral indices.

\* 国家自然科学基金项目 (42471329) 和山东省高等学校“青创团队计划”团队项目 (2022KJ224) 资助 Supported by the National Natural Science Foundation of China (No.42471329) and the Scientific Innovation Project for Young Scientists in Shandong Provincial Universities (No.2022KJ224)

<sup>†</sup> 通讯作者 Corresponding author, E-mail: guobingjl@163.com

作者简介: 杨吉存(1999-), 男, 山东青岛人, 硕士研究生, 主要研究方向为高光谱反演与盐渍化。E-mail: 24407010856@stumail.sdut.edu.cn

收稿日期: 2025-08-02; 收到修改稿日期: 2026-03-04; 网络首发日期 (www.cnki.net):

**【Method】** Dongying City was selected as the study area, where hyperspectral datasets were collected through field surveys, and soil samples were analyzed in the laboratory for salinity determination. The samples were divided into training and testing sets in a 7:3 ratio based on salinity gradients. Spectral data were preprocessed using Savitzky-Golay (S-G) filtering and Multiplicative Scatter Correction (MSC). Four spectral transformations were applied: Reflectance ( $R$ ), Reciprocal ( $1/R$ ), Logarithm of Reciprocal ( $\log(1/R)$ ), and Continuum Removal (CR). The Fractional Order Derivative (FOD) transformation was subsequently performed on each form. Ten types of two-dimensional spectral indices were constructed from the combined transformed data at each derivative order. Optimal band combinations and differential orders were identified by assessing correlation coefficients with soil salt content (SSC). Using these spectral indices as features and measured salinity as the dependent variable, four machine learning models—Partial Least Squares Regression (PLSR), Convolutional Neural Network (CNN), eXtreme Gradient Boosting (XGBoost), and Support Vector Machine (SVM)—were constructed. The hyperparameters of all models were optimized using the Bayesian Optimization (BO) algorithm, which iteratively fitted a probabilistic surrogate model to guide the search for hyperparameters that minimize cross-validation error. Each model was trained and tuned via ten-fold cross-validation. Performance was evaluated using the Coefficient of Determination ( $R^2$ ), Root Mean Square Error (RMSE), and Residual Prediction Deviation (RPD). The best-performing model was further interpreted using SHapley Additive exPlanations (SHAP) to identify influential spectral features. **【Result】** Results demonstrated that: (1) FOD effectively enhances spectral sensitivity by highlighting gradient information during spectral curve variations; (2) Mathematical transformations combined with FOD significantly improve correlations between spectral data and SSC; (3) The 2-order NDI index after CR treatment achieved the highest absolute correlation coefficient ( $|r|=0.91$ ) with SSC; (4) The CR-FOD-XGBoost model delivered optimal accuracy (testing set:  $R^2=0.94$ ,  $RMSE=0.85 \text{ g}\cdot\text{kg}^{-1}$ ,  $RPD=4.33$ ); (5) In the optimal model, GDI1 contributed most significantly while DI clusters adjacent to zero contributed minimally. **【Conclusion】** Collectively, this study demonstrates that combining spectral transformations to construct indices with Bayesian-optimized XGBoost modeling effectively improves soil salinity inversion accuracy, providing scientific foundations for salinization control and ecological sustainability. Future research should focus on enhancing spectral sensitivity responsiveness to further improve model performance, thereby advancing theoretical frameworks for sustainable land-use and environmental conservation strategies.

**Key words:** Spectral transformation; Fractional order derivative (FOD); Spectral indices; Salinization; XGBoost; Dongying city

土壤盐渍化是指水分蒸发后盐分在表层积累的过程<sup>[1]</sup>。黄河三角洲作为我国滨海盐渍化的典型区域, 关乎国家粮食安全与生态屏障建设, 实现该地区的盐渍化监测尤为重要<sup>[2]</sup>。

光谱数据变换能够有效实现光谱数据增强, 广泛应用于数据分析等领域<sup>[3-5]</sup>。张贤龙等<sup>[6]</sup>对光谱进行一阶微分等 15 种数学变换, 发现变换后能有效提升模型精度。分数阶微分 (Fractional Order Derivative, FOD) 变换具有“全局性”和“记忆性”优势, 通过在整数阶微分间连续插值, 能捕捉光谱微分变换中的渐变特征, 进一步提升光谱敏感性<sup>[7-8]</sup>。有学者将常规数学变换与 FOD 变换结合, 指出数学变换可调整全局光谱形态、突出物质吸收或反射特征, 而 FOD 能精细化光谱反射率、捕捉非线性响应细节, 从而共同提高光谱敏感性<sup>[9]</sup>。常规数学变换及 FOD 变换虽能提升光谱敏感性, 但无法实现数据降维。竞争自适应重加权等算法在光谱特征提取中应用广泛, 但未能充分利用波段间相关信息, 存在微弱特征被湮没的局限。采用二维光谱指数, 可通过波段全组合优化, 利用波段互补或耦合抑制噪声, 同时增强特征敏感性。如马瑜蔓等<sup>[10]</sup>基于 FOD 变换后的数据构建光谱指数进行油菜产量预测, 取得了较高的精度。

偏最小二乘回归 (Partial Least Squares Regression, PLSR) 作为一种线性多变量回归模型,

在盐渍化监测中应用广泛,但对复杂非线性关系的捕捉能力有限,面对复杂光谱数据时需采用更稳健的模型以提升精度<sup>[11]</sup>。极限梯度提升(eXtreme Gradient Boosting, XGBoost)作为集成学习模型,在非线性强、特征冗余的场景中表现出高效性<sup>[12]</sup>。Ye等<sup>[13]</sup>借助XGBoost进行土壤砷含量预测,模型可靠性较高。支持向量机(Support Vector Machine, SVM)通过核函数映射处理高维非线性光谱数据,在土壤属性预测中兼具泛化性与稳定性,适用于小样本盐渍化监测任务。高曦文等<sup>[14]</sup>利用SVM对土壤盐渍化进行分类,准确率达85%,表现出较好的适应性。

常规数学变换虽能提升光谱敏感性,但难以捕捉非线性细节,且一维数据无法充分利用波段间的信息。PLSR在处理非线性数据时存在局限性,SVM面临核函数选择与解释性弱的限制,卷积神经网络(Convolutional Neural Network, CNN)与XGBoost则为黑箱模型,可解释性较差。为此,本研究以东营市为研究区,融合数学变换与FOD构建二维光谱指数,对比PLSR、CNN、XGBoost及SVM四种模型,并对最优模型进行SHAP可解释性分析,以实现土壤含盐量的高精度监测。

## 1 材料与方法

### 1.1 研究区概况

如图1所示,东营市位于山东省北部黄河三角洲核心地带(36°55'–38°10' N, 118°07'–119°10' E),海拔介于1~28 m,年均温12.5~12.8 °C,年均降水量550~600 mm。区域地势低平、地下水位高(受潮水补给)、排水不畅,蒸发作用使盐分通过毛细作用向表土迁移,呈现典型滨海盐碱化特征。

在本研究区,滨海盐土面积占比最大,约为45%;砂姜黑土的面积占比最小,为0.8%。

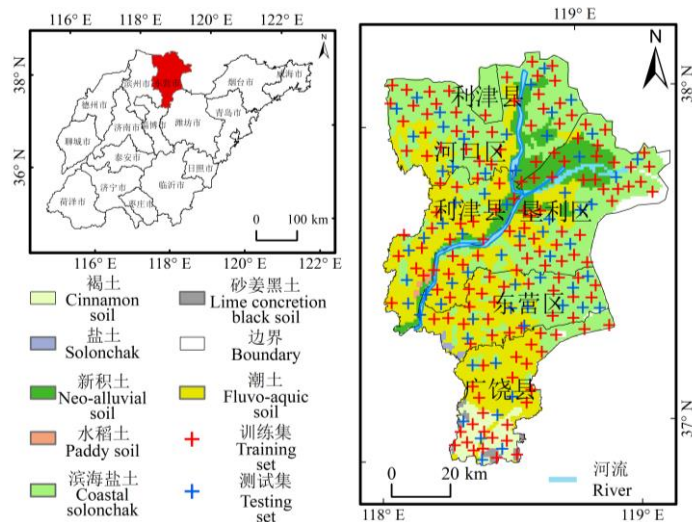


图1 研究区概况

Fig. 1 Study area overview

### 1.2 数据来源

本研究所需数据为土壤类型数据与实地测量数据。土壤类型数据来源于资源环境科学与数据中心(<https://www.resdc.cn/>),空间分辨率为1 km,含12个土纲、61个土类、227个亚类。实地测量数据包括盐分数据与高光谱数据:共布设213个采样单元,同时进行GPS定位,五点梅花法采集0~20 cm表层土壤,收集后经风干、过筛、四分法取样,制备1:5土水质量比浸提液,采用电极法测电导率,经标准曲线转换为全盐量,平行测定3次取均值。光

谱测量于晴朗无风 11:00~14:00 使用 SVC HR1024 (350~2500 nm) 同步进行, 测量前进行白板校准。采用双视场观测: 25°视场垂直 1.1 m 扫描 (地面直径 0.5 m) 与 4°视场距地 15 cm 精细测量 (地面直径 5 cm), 各采样 5 次取均值, 基于变异系数倒数自适应加权融合, 以兼顾宏观代表性及局部精度。最终以加权平均光谱构建土壤盐分反演模型。

### 1.3 数据预处理

由于数据采集过程中易受噪声影响, 采用 S-G 滤波对原始光谱数据进行平滑去噪处理<sup>[15]</sup>, 平滑窗口设置为 7, 多项式次数设置为 3。为减弱或消除散射效应引起的光谱基线漂移, 对去噪后的数据进行多元散射校正处理。本研究中的数据处理过程均在 Matlab R2022a 中完成。

为确保建模的稳定性, 减少单次划分带来的偶然性, 在模型训练过程中, 采用重复随机划分的方法, 选取 70% 的样本用于训练 ( $n=149$ ), 30% 的样本用于测试 ( $n=64$ ), 重复次数为 100 次。

### 1.4 数学变换

本研究中采用三种数学变换对高光谱反射率数据 ( $R$ ) 进行处理, 分别为倒数 ( $1/R$ )、倒数对数 ( $\log(1/R)$ ) 以及连续统去除 (CR), 以提高光谱数据的敏感性<sup>[6]</sup>。

### 1.5 分数阶微分

采用 Grünwald-Letnikov (G-L) 分数阶微分变换进行数据分析, 对于定义在区间  $[s, t]$  上的连续函数  $f(x)$ , 其在点  $x$  处的分数阶微分形式如式 (1) 所示<sup>[16]</sup>。

$$d_t^\alpha f(x) = \lim_{h \rightarrow 0} \frac{1}{h^\alpha} \sum_{m=0}^{\lfloor \frac{t-s}{h} \rfloor} (-1)^m \frac{\Gamma(\alpha+1)}{m! \Gamma(\alpha-m+1)} f(x-mh) \quad (1)$$

式中,  $h$  为步长;  $t$  和  $s$  分别为微分上限与下限;  $\Gamma$  为 Gamma 函数;  $\alpha$  为微分阶次, 当  $\alpha=0, 1, 2$  时, 分别对应原始光谱数据、一阶微分和二阶微分;  $m$  为点  $x$  向后离散跳跃的步数,  $m \in [0, \frac{t-s}{h}]$ 。在本研究中, 令  $h$  为 1,  $s$  和  $t$  分别为光谱起止波长位置, 令  $k = t-s$ ,  $m \in [0, k]$ , 则式 (1) 可表示为:

$$\frac{d^\alpha f(x)}{d\lambda^\alpha} \approx f(x) + (-\alpha)f(x-1) + \frac{(-\alpha)(-\alpha+1)}{2} f(x-2) + \dots + \frac{\Gamma(-\alpha+1)}{k!(-\alpha+k+1)} f(x-k) \quad (2)$$

在本研究中, 将波长  $\lambda$  作为自变量  $x$ , 则连续函数  $f(x)$  为波长反射率 ( $R$ 、 $\log(1/R)$ 、 $1/R$ 、CR)。阶数的设置参考了前人研究<sup>[16-17]</sup>并兼顾计算效率, 选取 0 至 2 阶的范围, 以 0.25 为间隔, 共计 9 个微分阶数 (0、0.25、0.5、0.75、1.0、1.25、1.5、1.75、2.0)。该方案能在有效捕捉光谱曲线形态渐变特征的同时, 避免因阶数过密 (如 0.1 间隔) 导致的计算成本高昂与特征冗余问题。

### 1.6 指数构建

与单波段光谱数据相比, 光谱指数能利用波段间的相互作用有效抑制噪声、增强敏感信号, 从而更好地量化地表参数, 在高光谱数据的处理中具有广泛的应用<sup>[18]</sup>。基于数学变换以及 FOD 变换后的高光谱数据, 构建 10 种典型二维光谱指数, 其相关计算公式如表 1 所示。

表 1 10 种典型光谱指数及计算公式

Table 1 Ten typical spectral indices and their calculation formulas

光谱指数 Spectral index	计算公式 Calculation formula	依据 Rationale
------------------------	--------------------------------	-----------------

最优光谱指数 <sup>[17]</sup> Optimal spectral index (OSI)	$\frac{(1+0.45)(2R_i^\alpha + 1)}{R_i^\alpha + 0.45}$	经验系数 0.45 对干扰反射率( $R_i$ )及敏感反射率( $R_j$ )优化, 提高监测精度与鲁棒性
土壤调节光谱指数 <sup>[17]</sup> Soil adjusted spectral index(SASI)	$\frac{(1+0.5)(R_i^\alpha - R_j^\alpha)}{R_i^\alpha + R_j^\alpha + 0.5}$	结合差值信号放大, 归一化噪声及经验系数优化的优势, 放大土壤盐分信号
乘积指数 <sup>[17]</sup> Product index(PI)	$R_i^\alpha R_j^\alpha$	强化多波段协同响应, 突出波段敏感
加和指数 <sup>[17]</sup> Summing index(SI)	$R_i^\alpha + R_j^\alpha$	反射率线性叠加, 强化敏感波段的协同效应或通过波段互补抑制局部干扰
氮平面域指数 <sup>[17]</sup> Nitrogen planar domain index(NPDI)	$(R_i^\alpha + R_j^\alpha)R_j^\alpha$	加和项能够综合波段的反射率特征, 平衡光照差异或增强盐分的协同效应, 乘积项能够放大辅助波段的贡献, 进一步强化反射率特征。
广义指数 1 <sup>[17]</sup> Generalized inde(GDI1)	$\frac{R_i^{\alpha 2} - R_j^{\alpha 2}}{R_i^{\alpha 2} + R_j^{\alpha 2}}$	非线性平方运算, 放大盐分敏感波段的反射差异, 归一化运算增强抗干扰能力, 消除光照影响
差值指数 <sup>[19]</sup> Difference index(DI)	$R_i^\alpha - R_j^\alpha$	计算两波段的反射率线性差值, 放大盐分引起的突变特征
比值指数 <sup>[19]</sup> Ratio index(RI)	$\frac{R_i^\alpha}{R_j^\alpha}$	通过反射率比值突出盐分敏感波段的比例关系
归一化指数 <sup>[19]</sup> Normalized index(NDI)	$\frac{R_i^\alpha - R_j^\alpha}{R_i^\alpha + R_j^\alpha}$	将两波段的反射率线性差值归一化, 以消除不同光照条件下的量级差异
广义指数 2 <sup>[20]</sup> Generalized inde(GDI2)	$\sqrt{R_i^{\alpha 2} + R_j^{\alpha 2}}$	欧几里得范数指数, 度量了两个波段反射率的综合强度或总幅度, 对地物的整体亮度或反射率敏感, 能融合两个波段所承载的能量信息。

注:  $R_i$  及  $R_j$  表示在波长  $i, j$  处的波长反射率 ( $i \neq j$ ),  $\alpha$  表示微分阶次。Note:  $R_i$  and  $R_j$  represent the spectral reflectance at wavelengths  $i$  and  $j$ , respectively ( $i \neq j$ ), and  $\alpha$  denotes the differential order.

### 1.7 模型构建及精度评定

PLSR 是一种针对高维、多重共线性数据的多元统计方法<sup>[21]</sup>。CNN 是一种能自动从数据中提取深层特征的深度学习方法。本研究构建的 CNN 包含 2 个卷积层 (分别含 16 个和 8 个长度为 3 的滤波器)、2 个最大池化层及 1 个含 64 个神经元的全连接层, 最终连接输出层<sup>[22]</sup>。为防止过拟合, 采用高斯噪声进行数据增强, 引入 L2 正则化与 Dropout 层限制模型复杂度, 并采用均方误差进行验证。XGBoost 是一种基于梯度提升的集成学习算法<sup>[23]</sup>, 其关键调优参数包括学习率、最大树深度、子采样比例, 以及 L1 和 L2 正则化项。SVM 是一种通过寻找最大间隔超平面的监督学习算法<sup>[24]</sup>, 其关键调优参数为惩罚参数  $C$  和核系数  $\gamma$ 。模型采用  $R^2$ 、RMSE 和 RPD 进行综合评估<sup>[25]</sup>。上述模型的构建均在 Python 3.13 中实现。

### 1.8 贝叶斯优化

贝叶斯优化通过构建概率代理模型来引导搜索过程, 利用已有评估结果不断更新后验分布, 并通过采集函数平衡“探索”与“利用”, 从而能以更少的迭代次数使参数逼近全局最优, 显著降低计算成本, 尤其适用于高维或耗时场景<sup>[26]</sup>。本研究采用十折交叉验证结合贝叶斯优化进行超参数调优。

### 1.9 SHAP 可解释性

SHAP 是一种基于合作博弈论中 Shapley 值的模型解释方法。其核心思想是将模型预测值分解为各特征的贡献值 (即 SHAP 值), 从而量化每个特征对单个预测结果的影响<sup>[27]</sup>。该方法通过加权平均法计算特征贡献, 提供兼具全局与局部视角的可加性解释, 在解析复杂机器学习模型时兼具公平性与较高的实用性。

## 2 结果

### 2.1 光谱变换特征

图 2 所示,原始光谱曲线(图 2a)呈现先增后降的趋势:可见光波段反射率上升明显,近红外波段增速放缓。在 1400 nm 和 1900 nm 附近存在明显吸收谷,在 2200 nm 处出现显著反射峰。这些特征分别对应土壤表层水与黏土矿物的 O-H 吸收、蒙脱石类矿物以及高岭石中的 Al-OH 振动<sup>[28]</sup>。经 S-G 滤波与多元散射校正后,光谱曲线更为光滑、集中(图 2b)。连续统去除进一步突出了 1400 nm、1900 nm 及 2200 nm 附近的吸收与反射特征,并在 400 nm、900 nm 和 2300 nm 附近也呈现出明显的光谱响应(图 2c)。而 1/R 与  $\log(1/R)$  后的光谱趋势与原始光谱相反(图 2d、图 2e)。

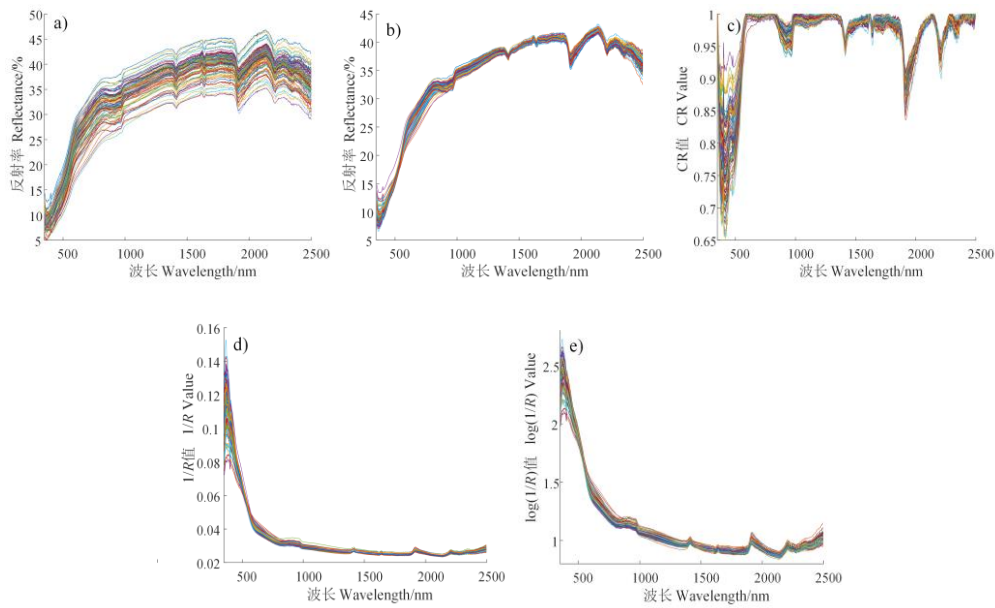


图 2 光谱曲线 (a) 原始光谱 (b) 预处理后光谱 (c) CR (d) 1/R (e)  $\log(1/R)$

Fig. 2 Spectral curves (a) Original spectra (b) Preprocessed spectra (c) CR (d) 1/R (e)  $\log(1/R)$

对不同变换后的光谱进行 FOD 变换,如图 3 所示,在不同的数学变换下,随着微分阶次的升高,曲线均逐渐趋向于  $y=0$  的直线,基线漂移现象逐渐减弱。以预处理后的光谱为例,0~0.75 阶曲线能基本保持原光谱曲线特征,随阶次升高,波峰波谷特征发生一定程度的变化。在 1900 nm 附近,当微分阶次在 0~1 阶时,其吸收谷特征逐渐凸显,1~2 阶时逐渐减弱;在 2100 nm 附近,当微分阶次在 1.25~2 阶时,其吸收峰特征变化最为明显。

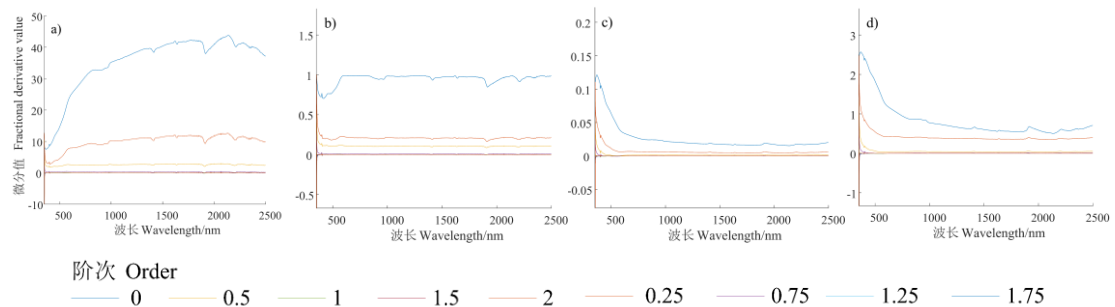


图 3 不同数学变换经 FOD 变换后的光谱曲线 (a) 预处理后光谱 (b) CR (c) 1/R (d)  $\log(1/R)$

Fig. 3 Spectral curves after fod transformation under different mathematical treatments (a) Preprocessed spectra (b) CR (c) 1/R (d)  $\log(1/R)$

图 4 所示,  $1/R$  及  $\log(1/R)$  变换后的微分光谱在 500~900 nm 波段呈正相关, 在 900~1400 nm 波段呈负相关, 正负相关性与原始光谱相反。相关性随着微分阶次的升高呈现正负交替的现象。CR 变换后的相关系数最大值最高, 达 0.50, 对应 2 阶微分和 673 nm。该波长与白色盐结晶的反射特征及黏质土壤在可见光波段对光的散射增强有关。此外, 900 nm、1400 nm 和 1900 nm 附近也表现出较高相关性。其中, 900 nm 位于水汽弱吸收带, 高盐分通过降低土壤持水能力影响该波段反射率; 1400 nm 和 1900 nm 则与黏土矿物相关, 盐渍化过程破坏土壤结构, 改变矿物表面性质, 从而影响其光谱响应。

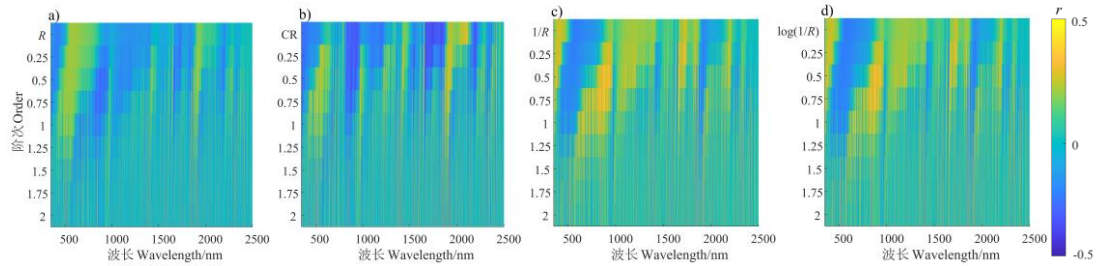
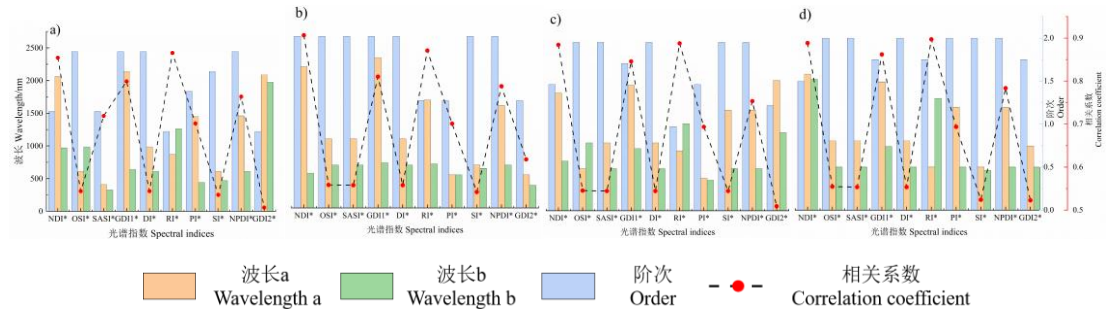


图 4 组合变换光谱反射率与土壤含盐量相关系数 (a)预处理后光谱 (b)CR (c)1/R (d)log(1/R)

Fig. 4 Correlation coefficients between combined transformation spectral reflectance and soil salt content (a) Preprocessed spectra (b) CR (c) 1/R (d) log(1/R)

## 2.2 最优光谱指数筛选

分别将不同变换的光谱反射率两两组合, 计算光谱指数, 并计算光谱指数与盐分的相关性。图 5 所示, 在全部变换组合中, 经 CR 处理后构建的 2 阶 NDI 与土壤含盐量的  $|r|$  最高(0.91), 其对应的波长组合为 2191 nm 与 535 nm, 其次为经  $1/R$  处理后的 1 阶 RI ( $|r|=0.90$ )。2191 nm 位于短波红外区域, 蒸发盐矿物(石膏( $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$ )、芒硝( $\text{Na}_2\text{SO}_4 \cdot 10\text{H}_2\text{O}$ )、泡碱( $\text{Na}_2\text{CO}_3 \cdot 10\text{H}_2\text{O}$ )等)中羟基与结晶水在该波段形成强吸收特征峰; 535 nm 处于植被绿峰附近, 对土壤亮度与颜色敏感。NDI 通过比值差分强化了二者的反差信号, 而 RI 通过比值放大了表层脱水光谱信号。



注: \*表示通过  $P < 0.05$  显著性检验 Note: \* indicates statistical significance at  $P < 0.05$ .

图 5 最优光谱指数波长及微分阶次组合 (a) 预处理后光谱 (b)CR (c)1/R (d)log(1/R)

Fig. 5 Optimal wavelength and differential order combinations of spectral indices (a) Preprocessed spectra (b) CR (c) 1/R (d) log(1/R)

## 2.3 模型构建结果

依据图 5, 将四种组合光谱变换下所筛选的最优光谱指数作为特征, 构建四种模型, 精度结果如表 2 所示。在 PLSR、CNN 以及 XGBoost 模型中, 经数学变换后的建模结果精度优于预处理后的数据, 表明组合光谱变换在一定程度上能够更好地表征土壤盐渍化程度。经 CR 处理的光谱数据在 XGBoost 模型中精度最高, 训练集  $R^2$ 、RMSE 及 RPD 分别为 0.95、0.71  $\text{g} \cdot \text{kg}^{-1}$  和 4.61, 测试集  $R^2$ 、RMSE 及 RPD 分别为 0.94、0.85  $\text{g} \cdot \text{kg}^{-1}$  和 4.33, 相较于最差模型(经预处理的 PLSR 模型), 最优模型的测试集精度  $R^2$ 、RPD 分别提升 0.1、1.82, RMSE 下

降  $1.11 \text{ g} \cdot \text{kg}^{-1}$ ，表明该最优模型在预测土壤盐渍化时具有一定的优势。

表2 模型精度

Table 2 Model precision

模型 Model	处理类型 Processing type	$R_c^2$	$R_p^2$	RMSE <sub>c</sub> /( $\text{g} \cdot \text{kg}^{-1}$ )	RMSE <sub>p</sub> /( $\text{g} \cdot \text{kg}^{-1}$ )	RPD <sub>c</sub>	RPD <sub>p</sub>
PLSR	预处理	0.85	0.84	1.96	2.01	2.63	2.51
	CR	0.90	0.89	1.24	1.35	3.23	3.04
	1/R	0.87	0.87	1.60	1.63	2.81	2.78
	log(1/R)	0.86	0.84	1.83	1.98	2.71	2.55
CNN	预处理	0.91	0.88	1.33	1.57	3.42	2.93
	CR	0.92	0.92	1.19	1.25	3.64	3.55
	1/R	0.94	0.93	0.93	1.14	4.33	3.92
	log(1/R)	0.93	0.89	1.09	1.44	3.79	3.14
XGBoost	预处理	0.93	0.90	1.08	1.28	3.85	3.24
	CR	0.95	0.94	0.71	0.85	4.61	4.33
	1/R	0.94	0.93	0.91	1.07	4.35	3.85
	log(1/R)	0.93	0.91	1.01	1.35	3.81	3.46
SVM	预处理	0.92	0.91	1.24	1.38	3.73	3.47
	CR	0.90	0.89	1.26	1.40	3.24	3.11
	1/R	0.89	0.88	1.35	1.55	3.08	2.97
	log(1/R)	0.93	0.91	1.10	1.37	3.85	3.42

注：图中角标  $c$  和  $p$  分别表示训练集和测试集，下同。Note: The subscript letters  $c$  and  $p$  in the figure denote the training set and testing set, respectively, the same as below.

进一步对最优模型（CR 结合 FOD 变换的 XGBoost 模型）进行 SHAP 可解释性分析，如图 6 所示，GDI1 对模型的贡献度最高，且大量样点呈现正向驱动，少量高值点呈现负向驱动，其次为 NDI 和 RI，特征分布呈现明显的颜色梯度，正向驱动作用较高，低值点分布于 SHAP 负半轴，在 SHAP 正半轴呈现的红色、紫色特征较为明显，具有一定的弱正相关性，DI 点群紧邻零值，对模型的贡献性最小。

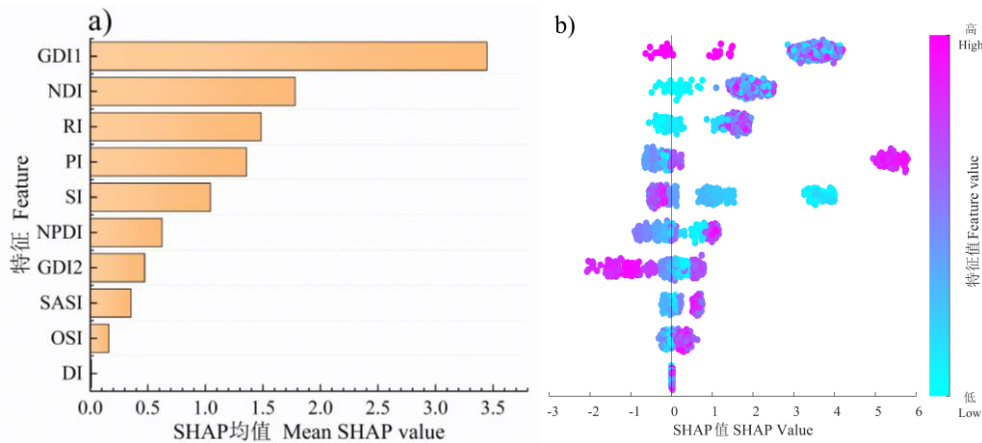


图6 SHAP 解释性分析 (a)SHAP 均值条形图 (b)SHAP 蜂群图 (以彩图发表)

Fig. 6 SHAP interpretability analysis (a) SHAP mean value bar chart (b) SHAP swarm plot.

### 3 讨论

高光谱数据虽信息丰富,但存在波长冗余与共线性问题,需进行特征筛选<sup>[29]</sup>。本研究通过对光谱进行常规数学变换(CR、 $1/R$ 、 $\log(1/R)$ )与分数阶微分(FOD)的组合处理,有效降低了数据冗余,提升了反演精度<sup>[30]</sup>。组合变换能够实现功能互补, $1/R$ 放大弱吸收特征, $\log(1/R)$ 提升低反射区信噪比,CR消除背景干扰、凸显与盐分相关的吸收带,FOD则通过全局微分特性精细挖掘隐含光谱变异,与有关学者的研究结论一致<sup>[31-32]</sup>。

一维光谱易受噪声与重叠峰干扰,二维光谱指数通过多波段关联可揭示波段间的协同变化。本研究构建的10种指数中,经CR变换的2阶NDI指数与含盐量的相关性最高,与唐子茹等<sup>[33]</sup>在干旱区研究中得出的最优指数(RI)存在差异,源于数据源(高光谱与多光谱)及盐渍环境(滨海与内陆)的不同,表明光谱指数需适配具体的土壤质地及环境特征。

基于CR-FOD变换的XGBoost模型预测精度优于PLSR、CNN与SVM。这是由于XGBoost模型能自适应学习非线性关系,突破PLSR的线性局限;正则化与二叉树结构可有效抑制高频噪声,并对光谱指数实现稳健分段建模,从而在小样本下获得更优的泛化性能。相比之下,CNN难以从已高度提纯的全局光谱指数中提取额外特征,且小样本易致模型学习能力不足<sup>[34]</sup>。该结论与巴亚岚等<sup>[35]</sup>在新疆农田的研究( $R^2$ 达0.72)具有一致性。

研究数据来自特定时段与区域,时空普适性不足;未涵盖季节与耕作管理下的水盐动态,长期监测能力待验证;滨海盐渍区最优方法向其他盐渍环境(如内陆干旱区、碱化土)迁移需跨区域对比评估。未来拟引入多时相数据、开发自适应特征选择算法,并在更广泛的生态区验证,以增强方法鲁棒性与推广价值。

### 4 结论

FOD结合数学变换可有效增强光谱对土壤盐分的敏感性;其中CR后2阶NDI与盐分相关性最高( $|r|=0.91$ ),表明变换方式能够显著影响指数敏感性。模型对比表明,基于CR-FOD的XGBoost模型表现最优,其梯度提升与正则化机制结合贝叶斯超参数优化,能有效学习多指数间的复杂关系,在强泛化能力下实现高预测精度,而PLSR相对较弱。SHAP可解释性分析表明,最优模型中GDII贡献最显著,DI贡献最低,各光谱指数对预测的贡献存在明显差异。本研究可为滨海盐渍区高光谱土壤盐分反演提供方法参考。

#### 参考文献 (References)

- [1] Fu P J, Bu Y K, Ma C J, et al. Spatio-temporal inversion of soil salinity in the Yellow River Delta region based on GEE[J]. *Bulletin of Surveying and Mapping*, 2024(11): 13-20, 26. [付萍杰, 补沅坤, 马池杰, 等. 基于GEE的黄河三角洲地区土壤含盐量时空反演[J]. *测绘通报*, 2024(11): 13-20, 26.]
- [2] Geng Q M, Tang X Y, Li X C, et al. Research progress and prospect of saline soil monitoring based on remote sensing technology[J]. *World Sci-Tech R & D*, 2025, 47(1): 60-81. [耿梦清, 唐夏雨, 李雪草, 等. 基于遥感技术的盐渍土监测研究进展与展望[J]. *世界科技研究与发展*, 2025, 47(1): 60-81.]
- [3] Chen K, Li C, Tang R. Estimation of the nitrogen concentration of rubber tree using fractional calculus augmented NIR spectra[J]. *Industrial Crops and Products*, 2017, 108: 831-839.
- [4] Wu Y H, Zhao H Q, Mao J H, et al. Study on hyperspectral inversion model of soil heavy metals in typical lead-zinc mining areas[J]. *Spectroscopy and Spectral Analysis*, 2024, 44(6): 1740-1750. [吴艳花, 赵恒谦, 毛继华, 等. 典型铅锌矿区土壤重金属含量高光谱反演模

- 型研究[J]. 光谱学与光谱分析, 2024, 44(6): 1740-1750.]
- [5] Li H, Yu H, Cao Y Y, et al. Hyperspectral prediction of soil organic matter content using CARS-CNN modelling[J]. Spectroscopy and Spectral Analysis, 2024, 44(8): 2303-2309. [李浩, 于滔, 曹永研, 等. 利用 CARS-CNN 模型的土壤有机质含量高光谱预测[J]. 光谱学与光谱分析, 2024, 44(8): 2303-2309.]
- [6] Zhang X L, Zhang F, Zhang H W, et al. Optimization of soil salt inversion model based on spectral transformation from hyperspectral index[J]. Transactions of the Chinese Society of Agricultural Engineering, 2018, 34(1): 110-117. [张贤龙, 张飞, 张海威, 等. 基于光谱变换的高光谱指数土壤盐分反演模型优选[J]. 农业工程学报, 2018, 34(1): 110-117.]
- [7] Wang H P, Chen P, Dai J W, et al. Recent advances of chemometric calibration methods in modern spectroscopy: Algorithms, strategy, and related issues[J]. Trends in Analytical Chemistry, 2022, 153: 116648.
- [8] Tian A H, Zhao J S, Zhang S J, et al. Hyperspectral estimation of saline soil electrical conductivity based on fractional derivative[J]. Chinese Journal of Eco-Agriculture, 2020, 28(4): 599-607. [田安红, 赵俊三, 张顺吉, 等. 基于分数阶微分的盐渍土电导率高光谱估算研究[J]. 中国生态农业学报, 2020, 28(4): 599-607.]
- [9] Yasenjiang Kahaer, Yang S T, Nigara Tashpolat, et al. Hyperspectral estimation of soil electrical conductivity based on fractional order differentially optimised spectral indices[J]. Acta Ecologica Sinica, 2019, 39(19): 7237-7248. [亚森江·喀哈尔, 杨胜天, 尼格拉·塔什甫拉提, 等. 基于分数阶微分优化光谱指数的土壤电导率高光谱估算[J]. 生态学报, 2019, 39(19): 7237-7248.]
- [10] Ma Y M, Duan B, Xu B C, et al. Rapeseed yield prediction based on fractional-order differentiation and UAV hyperspectral index optimization[J]. Transactions of the Chinese Society of Agricultural Engineering, 2025, 41(10): 166-175. [马瑜蔓, 段博, 徐宾灿, 等. 基于分数阶微分和无人机高光谱指数优选的油菜产量预测[J]. 农业工程学报, 2025, 41(10): 166-175.]
- [11] Liu L, Ji M, Buchroithner M. Transfer learning for soil spectroscopy based on convolutional neural networks and its application in soil clay content mapping using hyperspectral imagery[J]. Sensors, 2018, 18:3169.
- [12] Chen T, Guestrin C. Xgboost: A scalable tree boosting system[C]//Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016: 785-794.
- [13] Ye M, Zhu L, Li X, et al. Estimation of the soil arsenic concentration using a geographically weighted XGBoost model based on hyperspectral data[J]. Science of the Total Environment, 2023, 858: 159798.
- [14] Gao X W, Jia K L, Mao H X, et al. Study on hyperspectral quantitative classification of soil salinization based on wavelet transform and heterogeneous SVM[J]. Modern Electronics Technique, 2021, 44(3): 155-161. [高曦文, 贾科利, 毛鸿欣, 等. 基于小波变换及异质 SVM 方法的土壤盐渍化高光谱定量分类研究[J]. 现代电子技术, 2021, 44(3): 155-161.]
- [15] Guo B, Bai H R, Zhang B, et al. Inversion of soil zinc contents using hyperspectral remote sensing based on random forest and continuous wavelet transform in an opencast coal mine[J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(10): 138-147. [郭斌, 白昊睿, 张波, 等. 基于 RF 和连续小波变换的露天煤矿土壤锌含量高光谱遥感反演[J]. 农业工程学报, 2022, 38(10): 138-147.]
- [16] Zhang Y Q, Zhao Q C, Liu Q Y, et al. Inversion model of hyperspectral water quality

- parameters based on FOD and optimal spectral characteristics[J]. *Spectroscopy and Spectral Analysis*, 2025, 45(3): 842-851. [张雨晴, 赵起超, 刘其悦, 等. 基于 FOD 和最优光谱特征的高光谱水质参数反演模型[J]. *光谱学与光谱分析*, 2025, 45(3): 842-851.]
- [17] Huang H Y, Ding Q D, Zhang J H, et al. Ground-based hyperspectral inversion of salinization and alkalization of different soil layers in farmland in Yinbei area, Ningxia, China[J]. *Chinese Journal of Applied Ecology*, 2024, 35(11): 3073-3084. [黄华雨, 丁启东, 张俊华, 等. 基于地面高光谱的宁夏银北地区农田不同土层盐碱化信息反演[J]. *应用生态学报*, 2024, 35(11): 3073-3084.]
- [18] Jones A L, Ormondroyd A, Hayes F, et al. Reflections of stress: Ozone damage in broadleaf saplings can be identified from hyperspectral leaf reflectance[J]. *Environmental Pollution*, 2024, 360: 124642.
- [19] Sun Y N, Li X Y, Shi H B, et al. Optimizing the inversion of soil salt in salinized wasteland using hyperspectral data from remote sensing[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2022, 38(23): 101-111. [孙亚楠, 李仙岳, 史海滨, 等. 基于高光谱数据的盐荒地和耕地土壤盐分遥感反演优化[J]. *农业工程学报*, 2022, 38(23): 101-111.]
- [20] Shang T H. Quantitative inversion of soil organic matter in Yinchuan Plain area of Ningxia based on hyperspectral remote sensing[D]. Yinchuan: Ningxia University, 2022. [尚天浩. 基于高光谱遥感的银川平原土壤有机质定量反演[D]. 银川: 宁夏大学, 2022.]
- [21] Yang P, Hu J, Hu B, et al. Estimating soil organic matter content in desert areas using in situ hyperspectral data and feature variable selection algorithms in southern Xinjiang, China[J]. *Remote Sensing*, 2022, 14(20): 5221.
- [22] Lian Y Q, Su D H, Shi S X. Carbon peak prediction in fujian province based on combined STIRPAT and CNN-LSTM models[J]. *Environmental Science*, 2025, 46(1): 10-18. [连艳琼, 苏墩煌, 施生旭. 基于 STIRPAT 和 CNN-LSTM 组合模型的福建省碳达峰预测[J]. *环境科学*, 2025, 46(1): 10-18.]
- [23] Zhang B, Zhang Y, Jiang X. Feature selection for global tropospheric ozone prediction based on the BO-XGBoost-RFE algorithm[J]. *Scientific Reports*, 2022, 12(1): 9244.
- [24] Zhao M, Chang C Y, Wang Z R, et al. Driving force analysis and prediction simulation of soil salinization in coastal area of Yellow River Delta[J]. *Chinese Agricultural Science Bulletin*, 2024, 40(6): 75-83. [赵铭, 常春艳, 王卓然, 等. 黄三角濒海区土壤盐渍化的驱动力分析及预测模拟[J]. *中国农学通报*, 2024, 40(6): 75-83.]
- [25] Nie X J, Hong W W, Gill Ammara, et al. Hyperspectral estimation of coal-derived carbon mass fraction in mine soil based on the CWT-CARS-CNN integrated method[J]. *Journal of Henan Polytechnic University(Natural Science)*, 2024, 43(3): 91-100. [聂小军, 洪雯雯, Gill Ammara, 等. 基于 CWT-CARS-CNN 综合方法的矿区土壤煤源碳质量分数高光谱估测[J]. *河南理工大学学报(自然科学版)*, 2024, 43(3): 91-100.]
- [26] Zhao H Y, Zhou F, Yang W J, et al. Data-driven effectiveness evaluation modeling and simulation of anti-missile equipment system[J]. *Journal of National University of Defense Technology*, 2025, 47(3): 81-89.
- [27] Yang G, Qiao X, Zuo Q, et al. Root-Zone salinity in irrigated arid farmland: Revealing driving mechanisms of dynamic changes in China's Manas River basin over 20 Years[J]. *Remote Sensing*, 2024, 16(22): 4294.
- [28] Peng J, Zhou Q, Zhang Y Z, et al. Effect of soil organic matter on spectral characteristics of soil[J]. *Acta Pedologica Sinica*, 2013, 50(3): 517-524. [彭杰, 周清, 张杨珠, 等. 有机质对

- 土壤光谱特性的影响研究[J]. 土壤学报, 2013, 50(3): 517-524. ]
- [29] Zhang Z, Ding J, Wang J, et al. Prediction of soil organic matter in northwestern China using fractional-order derivative spectroscopy and modified normalized difference indices[J]. *Catena*, 2020, 185: 104257.
- [30] Liang Z Y, Chen S H, Li N, et al. Estimation of available phosphorus content in soil with different particle sizes based on hyperspectral remote sensing[J]. *Soil and Fertilizer Sciences in China*, 2024(12): 230-239. [梁智永, 陈署晃, 李娜, 等. 不同粒径土壤有效磷光谱估测[J]. *中国土壤与肥料*, 2024(12): 230-239.]
- [31] Guo Y P, Wang X M, Zhao F, et al. Hyperspectral inversion of the RF model for soil salinity in oasis tillage layer based on optimal mathematics and wavelet transform[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2025, 41(3): 83-93. [郭艳萍, 王雪梅, 赵枫, 等. 基于最优数学和小波变换的绿洲耕层土壤盐分 RF 高光谱反演[J]. *农业工程学报*, 2025, 41(3): 83-93.]
- [32] Jin H, Peng J, Bi R, et al. Comparing laboratory and satellite hyperspectral predictions of soil organic carbon in farmland[J]. *Agronomy*, 2024, 14:175.
- [33] Tang Z R, Wu T, Tan S L, et al. Soil salinization monitoring model based on remote sensing derivative processing and optimal spectral index[J]. *Hubei Agricultural Sciences*, 2024, 63(8): 223-230. [唐子茹, 吴彤, 谭世林, 等. 基于遥感导数处理和最优光谱指数的土壤盐渍化监测模型[J]. *湖北农业科学*, 2024, 63(8): 223-230.]
- [34] Yoon H I, Lee H, Yang J S, et al. Predicting models for plant metabolites based on PLSR, AdaBoost, XGBoost, and LightGBM algorithms using hyperspectral imaging of brassica juncea[J]. *Agriculture*, 2023, 13:1477.
- [35] Ba Y L, Zhang Z T, Xie P L, et al. Inverting soil salinity of farmland in Xinjiang by integrating Sentinel-1/2 and environmental variables[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2024, 40(16): 171-179. [巴亚岚, 张智韬, 谢坪良, 等. 集成 Sentinel-1/2 和环境变量的新疆农田土壤含盐量反演[J]. *农业工程学报*, 2024, 40(16): 171-179.]

(责任编辑: 檀满枝)