

# 基于可解释机器学习的河套平原盐碱农田 土壤水分和有机质含量估算

丁启东<sup>1</sup>, 黄华雨<sup>1</sup>, 张俊华<sup>1</sup>, 杨凌彬<sup>1</sup>, 贾科利<sup>2</sup>, 郝凤霞<sup>3</sup>

(1. 宁夏大学 生态环境学院, 宁夏 银川 750021; 2. 宁夏大学 地理科学与  
规划学院, 宁夏 银川 750021; 3. 宁夏大学 化学化工学院, 宁夏 银川 750021)

**摘要:** [目的] 针对传统方法在盐碱化农田土壤水分(SMC)和有机质含量(SOMC)监测中存在效率低下的问题,探索高光谱数据结合可解释机器学习的估算方法,以期为河套平原盐碱化土壤信息快速获取和土壤质量评价提供理论依据。[方法] 以地面高光谱反射率及实测 SMC 和 SOMC 为数据源,对光谱数据采用分数阶微分(FOD)变换并构建光谱指数,基于偏最小二乘回归(PLSR)、支持向量机(SVM)和随机森林(RF)算法建模,并引入夏普利加性解释(SHAP)方法解析变量对模型预测结果的相对贡献,提升模型的解释性。[结果] ①经 1.25 阶微分变换后构建的光谱指数与 SMC 和 SOMC 间相关性最强,其中,广义差异指数(GDI)和最优光谱指数(OSI)与 SMC 和 SOMC 间相关系数最大,分别为 0.505 4 和 0.682 5。②RF 模型对 SMC 和 SOMC 的估算精度远高于 PLSR 和 SVM;SMC 和 SOMC-RF 模型验证集( $R^2$ )、均方根误差(RMSE)和相对分析误差(RPD)分别为 0.734, 3.28, 2.07 及 0.870, 1.53, 2.43。③SHAP 分析发现,氮平面域指数(NPDI)和比值指数(RI)分别在 SMC 和 SOMC 的建模估算中贡献度最大,且 NPDI, OSI 和差值指数(DI)对 SMC 的建模贡献度累计达到 68.58%;RI, GDI 和 NPDI 对 SOMC 的建模贡献度累计达到 61.86%。[结论] FOD 联合光谱指数在高光谱数据的有效利用中具有明显优势,RF 模型在土壤属性估算中展现了较高的精度和鲁棒性,SHAP 分析有效揭示了不同变量对目标变量的贡献度。NPDI, RI, OSI 和 DI 等光谱指数在盐碱化农田 SMC 和 SOMC 的建模估算中贡献显著。

**关键词:** 高光谱; 土壤盐碱化; 机器学习; 夏普利加性解释; 数字土壤制图

文献标识码: B

文章编号: 1000-288X(2025)04-0184-14

中图分类号: S153.6

**文献参数:** 丁启东, 黄华雨, 张俊华, 等. 基于可解释机器学习的河套平原盐碱农田土壤水分和有机质含量估算[J]. 水土保持通报, 2025, 45(4): 184-197. Ding Qidong, Huang Huayu, Zhang Junhua, et al. Estimation of soil moisture and organic matter content in saline-alkaline farmland of Hetao Plain based on interpretable machine learning [J]. Bulletin of Soil and Water Conservation, 2025, 45(4): 184-197. DOI: 10.13961/j.cnki.stbctb.2025.04.033; CSTR: 32312.14.stbctb.2025.04.033.

## Estimation of soil moisture and organic matter content in saline-alkaline farmland of Hetao Plain based on interpretable machine learning

Ding Qidong<sup>1</sup>, Huang Huayu<sup>1</sup>, Zhang Junhua<sup>1</sup>, Yang Lingbin<sup>1</sup>, Jia Keli<sup>2</sup>, Hao Fengxia<sup>3</sup>

(1. College of Ecology and Environmental Science, Ningxia University, Yinchuan, Ningxia 750021, China; 2. College of Geographical Sciences and Planning, Ningxia University, Yinchuan, Ningxia 750021, China; 3. College of Chemistry & Chemical Engineering, Ningxia University, Yinchuan, Ningxia 750021, China)

**Abstract:** [Objective] To address the inefficiencies of traditional approaches for monitoring the soil moisture content (SMC) and soil organic matter content (SOMC) in saline-alkaline farmlands, an estimation method that integrates hyperspectral data with interpretable machine learning was investigated. The goal was to establish a

收稿日期: 2025-01-08

修回日期: 2025-05-18

采用日期: 2025-05-19

资助项目: 国家重点研发计划项目“河套平原盐碱农田水肥盐精准诊断与地力长效提升关键技术及产品研发”(2021YFD1900602); 国家自然科学基金项目“时空谱数据融合的宁蒙引黄灌区土壤盐碱化时空演变与风险评价”(42467036); 宁夏科技创新领军人才项目(2022GKLRXLX02)

第一作者: 丁启东(1995—), 男(汉族), 甘肃省天水市人, 硕士研究生, 研究方向为遥感监测、精准农业与土壤地力提升。Email: 2535643120@qq.com。

通信作者: 张俊华(1977—), 女(汉族), 宁夏回族自治区中卫市人, 博士, 研究员, 主要从事精准农业与土地质量提升研究。Email: zhangjunhua728@163.com。

郝凤霞(1978—), 女(汉族), 宁夏回族自治区中宁人, 硕士, 副教授, 主要从事土壤环境方面的研究工作。Email: hao\_fx@nxu.edu.cn。

theoretical foundation for the rapid acquisition of soil information and quality assessments of the Hetao Plain, China. [Methods] Ground-based hyperspectral reflectance data and field-measured SMC and SOMC were used as the primary data sources. Spectral data were processed using a fractional-order differential (FOD) transformation, and various spectral indices were constructed. The models were developed using partial least squares regression (PLSR), support vector machines (SVM) and random forest (RF). To enhance interpretability, the Shapley additive explanations (SHAP) method was employed to evaluate the relative contribution of each variable to model predictions. [Results] ① The spectral indices derived from the 1.25-order differential transformation showed the highest correlation with the SMC and SOMC. In particular, the generalized difference index (GDI) and optimal spectral index (OSI) exhibited the strongest correlations, with coefficients of 0.505 4 and 0.682 5, respectively. ② The RF model significantly outperformed the PLSR and SVM models in estimating both the SMC and SOMC. For the validation datasets, the RF models achieved  $R^2$  values of 0.734 and 0.870, root mean square errors of 3.28 and 1.53, and recognition-primed decisions of 2.07 and 2.43 for SMC and SOMC, respectively. ③ The SHAP analysis indicated that the normalized plane domain index (NPDI) and ratio index (RI) were the most influential variables for estimating the SMC and SOMC, respectively. The combined contributions of the NPDI, OSI and difference index (DI) to SMC modeling reached 68.58%, whereas RI, GDI and NPDI collectively contributed 61.86% to SOMC modeling. [Conclusion] The integration of FOD and spectral indices enhanced the utility of hyperspectral data. The RF model demonstrated superior accuracy and robustness in estimating soil properties, whereas the SHAP analysis effectively elucidated the contribution of individual variables. Spectral indices (such as NPDI, RI, OSI and DI) played significant roles in modeling SMC and SOMC in saline-alkaline farmland.

**Keywords: hyperspectral; soil salinization and alkalinization; machine learning; shapley additive explanations (SHAP); digital soil mapping**

土壤盐碱化是干旱与半干旱地区普遍存在且亟待解决的生态与农业问题,该现象不仅破坏土壤结构,显著抑制养分循环和作物生长,还降低农田产出效率<sup>[1]</sup>。土壤水分(soil moisture content, SMC)和有机质含量(soil organic matter content, SOMC)是土壤健康与农田生产力的核心指标,对土壤结构、微生物活性有着深远影响<sup>[2]</sup>。传统的土壤监测方法,如地面采样和实验室分析,虽然能够提供精准的土壤属性数据,但面临着高成本、低效率等问题,更无法满足大范围、实时监测的需求<sup>[3]</sup>。相比之下,高光谱遥感因其光谱信息丰富和获取效率高,正逐渐成为实现土壤属性快速、无损监测的重要手段<sup>[4]</sup>。利用高光谱数据估算 SMC 和 SOMC,不仅能克服传统监测方法的局限性,还能为盐碱化农田土壤信息的快速获取提供技术支撑。

在基于高光谱数据对土壤属性的估算中,数据预处理与特征选择对建模效果有重要影响<sup>[5]</sup>。有研究表明高光谱数据经一阶微分变换后对耕地 SOMC 的估算效果最优<sup>[6]</sup>。然而,整数阶微分变换会忽略与目标变量相关的细微光谱特征。此外,高阶微分变换还存在过度平滑等问题,可能导致一些关键光谱信息的丢失,进而影响模型的精度。针对以上问题,本研究设置细微的微分间隔,通过选取合适的阶数

来避免过度平滑,并保留与目标变量有效的光谱信息。除反射率变换方法外,特征提取对模型精度亦有显著影响。尽管相关性分析<sup>[5]</sup>、灰度关联法以及竞争性自适应重加权采样等方法广泛用于特征选择,但选取的单一波段往往难以全面表征土壤的复杂性。相比之下,构建光谱指数可有效融合多波段信息,增强与土壤属性的相关性<sup>[7]</sup>。尚天浩等<sup>[8]</sup>通过 FOD 变换构建光谱指数,实现了银川平原 SOMC 的高精度估算( $R^2=0.87$ )。在高光谱估算 SMC 的研究中,FOD 结合光谱指数同样展现出较高精度<sup>[9]</sup>。但上述研究对不同光谱指数在 SMC 与 SOMC 建模中的贡献度尚缺乏系统量化评估,仍需进一步研究。

机器学习模型凭借其非线性建模能力和强鲁棒性,广泛应用于高光谱数据分析<sup>[10]</sup>。偏最小二乘回归(partial least squares regression, PLSR)、支持向量机(support vector machine, SVM)和随机森林(random forest, RF)等算法已被成功用于土壤属性估算<sup>[11]</sup>。李天乐等<sup>[12]</sup>在利用高光谱数据估算 SOMC 时发现,SVM 模型表现优于 PLSR,估算精度达到  $R^2=0.85$ ,体现出 SVM 较强的非线性拟合能力。在宁夏银北地区盐碱化 SMC 的估算研究中,SVM 同样展现出最佳的建模效果<sup>[13]</sup>。同时,RF 算法在 SOMC 估算中表现最优<sup>[14]</sup>,  $R^2$  达到 0.93。尽管机器学习模型

在精度上表现优越,其“黑箱”特性使得模型内部决策机制仍不透明,限制了其在科学研究中的解释力<sup>[15]</sup>。为提升模型可解释性,夏普利加性解释(shapley additive explanations, SHAP)作为一种解释型算法,可对模型预测结果进行可视化解析,定量评估各变量对预测结果的贡献<sup>[16]</sup>。叶森等<sup>[17]</sup>在北京市延庆区和房山区永久农田 SOMC 估算中,通过 SHAP 识别并剔除贡献度较低的变量,重构模型后精度显著提升。在辽西地区 SMC 的估算中,研究者将 RF 与 SHAP 结合,系统阐明了不同变量对模型预测的贡献<sup>[18]</sup>。基于上述成果,本研究引入 SHAP 方法,旨在系统量化并识别对 SMC 与 SOMC 建模贡献显著的关键变量,从而为盐碱化农田 SMC 与 SOMC 的高精度估算提供理论依据与变量优选支持。

河套平原是中国西北重要的农业产区,土壤盐碱化已成为制约其农田生产力和生态可持续发展的关键因素。本研究结合地面高光谱数据、机器学习算法与 SHAP 方法,构建 SMC 与 SOMC 的估算模型,并量化不同光谱指数在建模中的贡献。以期为

河套平原及类似地区盐碱化土壤的肥力监测与地力提升提供方法支持和数据依据。

## 1 材料与方法

### 1.1 研究区概况

河套平原位于内蒙古和宁夏回族自治区,属于旱半干旱区,是黄河沿岸的冲积平原,面积约  $2.50 \times 10^4 \text{ km}^2$ 。地势平坦,依靠黄河水自流灌溉,盛产小麦、玉米、高粱等,是中国重要的农业区和商品粮基地。然而,灌溉水渗漏抬升地下水位,导致土壤盐碱化加剧,肥力下降。本研究采样点涵盖 9 个典型示范区(图 1)。

(1) 吴忠市红寺堡区。属中温带干旱气候区,该区多年平均蒸发量为 2 387 mm,是降水量的 9 倍。农田灌溉多采用大水漫灌方式,加之排水不畅,导致地下水位上升,次生盐碱化现象明显。

(2) 银川市兴庆区。属西北内陆温带干旱气候区,农田排水系统较好,土壤盐碱化程度较轻,但地处黄河沿岸,导致地下水位偏高。

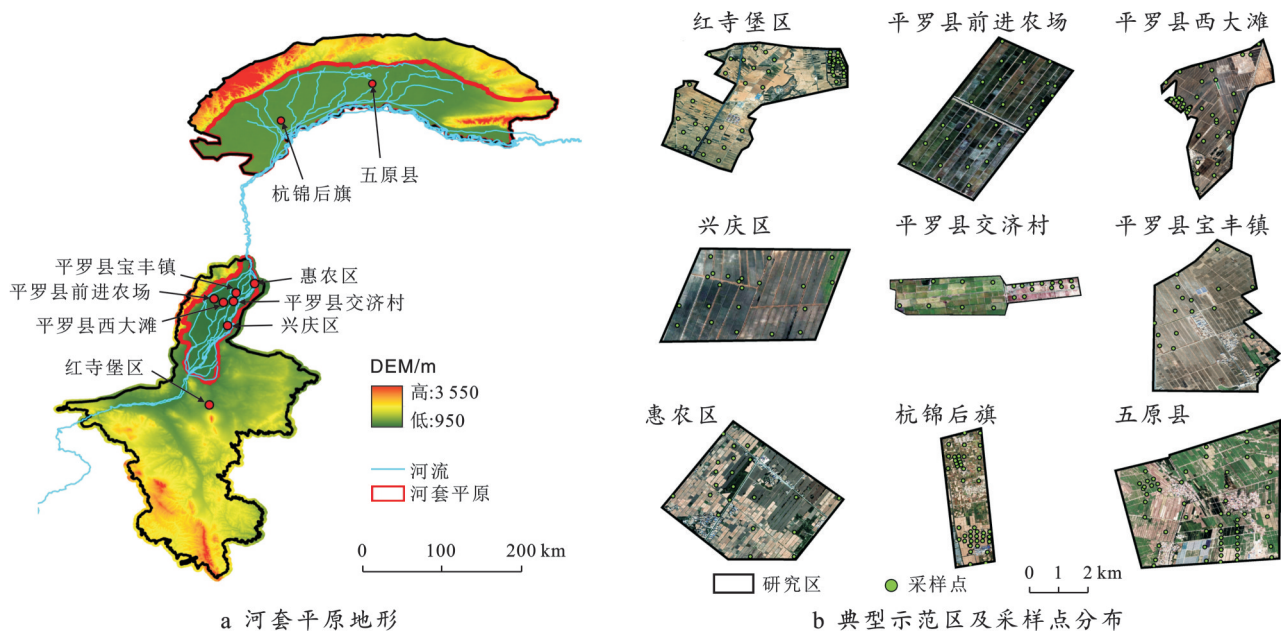


图 1 研究区地形及采样点情况

Fig.1 Condition of sampling plots and topography of study area

(3) 平罗县。属温带干旱荒漠气候区,降水分布不均,蒸发强度大。盐碱地主要分布在前进农场、西大滩、宝丰镇和交济村。地势整体平缓低洼,灌溉条件良好但排水不畅,农田排水沟深 1.5~1.8 m,接近地下水位。

(4) 惠农区。属西北内陆温带干旱气候区,该地区地处灌区下游,是上游排水排盐的聚集区,同时,受黄河水顶托作用,导致排水不畅,多为重度盐碱地。

(5) 内蒙古杭锦后旗和五原县。属中温带大陆

性季风气候,海拔高度 1 000~1 800 m,地势由南向北倾斜,蒸降比接近 10:1,土壤类型主要有淡栗钙土、灌淤土、潮土和盐化土等盐碱化程度高,pH 值平均 8.55。

### 1.2 高光谱数据采集与处理

样品采集均在春灌之前进行,采样前利用 ArcGIS 10.8 布设采样网格。具体采样时间为 2022 年 3 月中旬至 4 月上旬,在宁夏的红寺堡区、兴庆区、前进农场、宝丰镇、交济村及西大滩等 6 个示范区完

成采集任务;2023年3月下旬至4月上旬,则分别在宁夏惠农区以及内蒙古五原县和杭锦后旗的3个示范区开展了样品采集工作。

高光谱数据由ASD FieldSpec 4地物光谱仪(350~2 500 nm)获取,测量安排在晴朗、无风的10:00—14:00进行。每个样点先进行白板校准,随后采集5条光谱曲线,取均值作为代表光谱。数据经ViewSpec Pro处理并导出,为减少噪声干扰,剔除350~499 nm及2 401~2 500 nm波段。

### 1.3 土壤样品采集与指标测定

为确保土壤样品具有代表性,每个样点以高光谱测定点为中心,在直径1 m范围内采用五点采样法采集表层土(0—20 cm),将采集的土样充分混合后,取约500 g均匀样品,装入洁净自封袋中,密封保存以备后续试验分析。9个示范区采集有效样品共计383份。将土壤样品带回实验室,首先取部分样品采用烘干法测定SMC,剩余样品风干、过筛后,采用外加热重铬酸钾法测定SOMC。

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

式中: $x$ 为自变量; $\alpha$ 为阶数, $\alpha=0.0$ 表示原始光谱, $\alpha=1.0$ 表示一阶微分, $\alpha=2.0$ 则表示二阶微分。

### 1.5 光谱指数

光谱指数是通过对不同波段反射率进行数学组合运算,用于提取物体的光谱特征信息。其原理基于不同物质对电磁波的反射、吸收和散射特性差异,结合特定波段选择,能够有效地增强目标物质的光谱特征。不同类型的光谱指数通过选择适当的波段组合,利用反射率差异,能够有效地表征土壤特性<sup>[20]</sup>。

本研究基于FOD变换后的光谱反射率构建差值指数(difference index, DI);乘积指数(product

### 1.4 光谱变换

分数阶微分(FOD)通过灵活调整微分阶数,有效增强高光谱数据中的关键信息特征。相比传统的整数阶微分方法,FOD在提升光谱特征敏感性的同时,能够更有效地抑制噪声,并增强光谱曲线的平滑性与分辨能力<sup>[19]</sup>。目前常用的FOD计算方法主要包括Caputo, Riemann-Liouville和Grünwald-Letnikov(G-L)3种形式。本研究基于G-L定义采用公式(1)在MATLAB R2023 b软件中编程实现FOD变换,具体步骤如下:

首先对400~2 400 nm间的原始高光谱数据采用Savitzky-Golay(SG)9点平滑进行预处理,为降低光谱波段间的共线性,以10 nm为间隔对数据进行重采样,最终获得201个波段,然后对重采样后的光谱数据进行0~2阶(间隔0.25)的FOD变换,最后,基于不同阶数的FOD光谱构建多种光谱指数,并分别计算其与SMC及SOMC之间的相关性,从中筛选出最优波段组合与光谱指数。

index, PI);比值指数(ratio index, RI);加和指数(sum index, SI);广义差异指数(generalized difference index, GDI);最优光谱指数(optimal spectral index, OSI);土壤调节光谱指数(soil-adjusted spectral index, SASI)和氮平面域指数(nitrogen planar domain index, NPDI)。这些光谱指数的构建充分考虑了不同土壤属性的光谱响应特性,旨在提高SMC和SOMC的估算精度。光谱指数构建公式见表1。

表1 光谱指数及公式

Table 1 Spectral index and formula

光谱指数	计算公式	光谱指数	计算公式
差值指数(DI)	$R_{\lambda_1} - R_{\lambda_2}$	广义差异指数(GDI)	$\frac{R_{\lambda_1}^2 - R_{\lambda_2}^2}{R_{\lambda_1}^2 + R_{\lambda_2}^2}$
乘积指数(PI)	$R_{\lambda_1} * R_{\lambda_2}$	最优光谱指数(OSI)	$\frac{(1+0.45)(2R_{\lambda_2} + 1)}{(R_{\lambda_1} + 0.45)}$
比值指数(RI)	$R_{\lambda_1}/R_{\lambda_2}$	土壤调节光谱指数(SASI)	$\frac{(1+0.5)(R_{\lambda_1} - R_{\lambda_2})}{(R_{\lambda_1} + R_{\lambda_2} + 0.5)}$
加和指数(SI)	$R_{\lambda_1} + R_{\lambda_2}$	氮平面域指数(NPDI)	$(R_{\lambda_1} + R_{\lambda_2}) * R_{\lambda_2}$

注: $\lambda_1$ 和 $\lambda_2$ 为400~2 400 nm间两个不相等的任意波段。

### 1.6 建模方法与评价

#### 1.6.1 建模方法

偏最小二乘回归(PLSR)是一种典型的线性降维

回归方法,适用于特征维度高、变量间存在多重共线性的情况。该方法通过将自变量矩阵 $X$ 和因变量矩阵 $Y$ 同时投影到一组新的潜在变量上,构建回归模型,其

核心思想是最大化  $X$  与  $Y$  的协方差<sup>[7]</sup>。其建模过程可表示为:

$$X = T \cdot P^T + E, \quad Y = T \cdot Q^T + F \quad (2)$$

式中:  $T$  为得分矩阵,  $P$  和  $Q$  分别为载荷矩阵,  $E$  和  $F$  为残差项。最终预测通过回归系数矩阵实现:

$$\hat{Y} = X \cdot B \quad (3)$$

潜变量个数通过交叉验证选择,以最小化验证集上的均方根误差(RMSE)。

支持向量机(SVM)是一种基于统计学习理论的非线性回归方法,具有良好的泛化能力。采用径向基函数(radial basis function, RBF)作为核函数<sup>[7]</sup>,其形式为:

$$K(X_i, Y_j) = \exp(-\gamma \|X_i - Y_j\|^2) \quad (4)$$

式中:  $\gamma$  为核参数,控制样本之间的影响范围。SVM 回归的目标是构建一个对训练数据拟合良好且对未知样本具有强泛化能力的函数  $f(x)$ ,其优化目标为最小化以下目标函数:

$$\min_{\omega, b, \epsilon, \epsilon^*} \left[ \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\epsilon^i + \epsilon_i^*) \right] \quad (5)$$

约束条件为:

$$\begin{cases} y_i - [\omega \cdot \varphi(x_i) + b] \leq \epsilon + \epsilon_i \\ [\omega \cdot \varphi(x_i) + b] - y_i \leq \epsilon + \epsilon_i \\ \epsilon_i, \epsilon_i^* \geq 0 \end{cases} \quad (6)$$

式中:  $C$  为惩罚系数;  $\varphi(x_i)$  为非线性映射函数;  $\epsilon$  为不敏感损失函数的宽度。模型参数  $C$  和  $\gamma$  通过网格搜索(grid search)结合交叉验证优化,以 RMSE 最小为准则选择最优组合。

随机森林(RF)是一种基于决策树的集成学习算法,采用 Bootstrap 自助采样法生成多个训练子集,并在每棵树的分裂过程中随机选择部分特征参与建模,从而提高模型的多样性与泛化能力。最终预测结果为所有树输出的平均值<sup>[21]</sup>:

$$\hat{y} = \frac{1}{T} \sum_{i=1}^T h_i(x) \quad (7)$$

式中:  $h_i(x)$  为每棵回归树的预测结果;  $T$  为总树数。每棵树的生成过程中,特征选择采用随机子集策略(通常为  $\sqrt{p}$ ,其中  $p$  为总特征数),以降低模型的方差并增强鲁棒性。

本研究将整体数据采用 5 折交叉验证法划分为 70% 用于训练集,30% 用于验证集。模型建立及数据划分通过 Python 环境在 Pycharm 中编程实现。

### 1.6.2 精度评价

模型精度通过建模决定系数( $R_c^2$ )、验证决定系数( $R_v^2$ )、均方根误差(RMSE)和相对分析误差(RPD)综

合评估。其中,  $R^2$  越接近 1, RMSE 越小,表明模型性能越优。RPD 用于判断预测能力,划分如下:  $RPD \leq 1.4$  表示预测效果较差,  $1.4 < RPD \leq 2.0$  为一般水平,  $2.0 < RPD \leq 2.5$  表示预测较好,  $RPD > 2.5$  说明模型预测效果优秀<sup>[22]</sup>。

### 1.7 夏普利加性解释

夏普利加性解释(SHAP)是基于博弈论的模型解释方法,旨在为复杂的机器学习模型提供全局和局部的可解释性。SHAP 通过计算每个特征在给定预测中的贡献来帮助理解模型的决策过程。该方法的核心思想来源于博弈论中的夏普利值(Shapley value),其本质上是对每个特征“边际贡献”的度量<sup>[23]</sup>。以上过程基于 Python 环境,通过 PyCharm 软件调用 shap 库中的 TreeExplainer 类对最佳模型进行解释,夏普利值计算公式如下:

$$\phi_i(f) = \sum_{S' \subseteq S \setminus \{x_i\}} \frac{|S'|! [n - |S'| - (1)!]}{n!} \{f(S' \cup \{x_i\}) - f(S')\} \quad (8)$$

式中:  $\phi_i(f)$  为特征  $\{x_i\}$  的夏普利值;  $f(S')$  为给定特征集合  $S'$  时模型的预测值;  $|S'|$  为子集  $S'$  中特征的数量;  $n$  为特征的总数量。

## 2 结果与分析

### 2.1 土壤水分和有机质含量描述性统计特征

研究区 SMC 和 SOMC 描述性统计结果见表 2。SMC 为 2.11%~37.60%,平均值为 17.81%,变异系数为 35.83%,表明 SMC 在空间上变异程度较高。SOMC 为 2.24~23.68 g/kg,平均值为 9.39 g/kg,其中 SOMC 低于 20 g/kg 样品占比达到 99.22%,表明整体 SOMC 水平偏低,变异系数为 43.45%,表明 SOMC 的空间变异性较高。

为直观呈现不同数据集中样本分布的统计特征,绘制了训练集与验证集对应的小提琴图(图 2)。由图 2 可知,SMC 和 SOMC 在总样本集、训练集和验证集中的分布较为一致,且均值和标准差变化不大,表明数据划分科学合理,具备较高的代表性。

在 SMC 和 SOMC 的分布上,各子集的均值和标准差相对稳定,变异系数变化较小,进一步表明数据集划分没有引入显著的偏差。SMC 样本总集、训练集和验证集均值分别为 17.81%,17.82% 和 17.79%(图 2a),而 SOMC 对应数据集均值分别为 9.39 g/kg,9.40 g/kg,9.36 g/kg(图 2b)。表明训练集和验证集能够有效代表整个数据集的分布特征,确保了模型训练和评估的科学性与稳定性。

表 2 土壤水分和有机质含量描述性统计

Table 2 Descriptive statistics of soil moisture and organic matter content

指标	最大值	最小值	平均值	标准差	变异系数/%
土壤水分含量/%	37.60	2.11	17.81	6.25	35.83
土壤有机质含量/(g·kg <sup>-1</sup> )	23.68	2.24	9.39	4.08	43.45

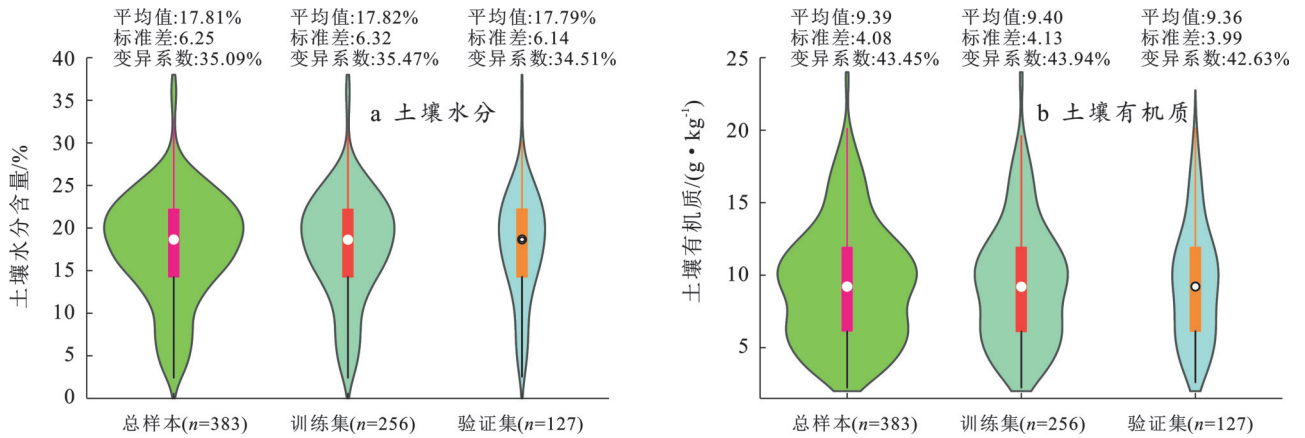


图 2 土壤水分和有机质含量建模数据集划分统计总结

Fig. 2 Statistical summary of modelling dataset division for soil moisture and organic matter content

2.2 不同等级土壤水分和有机质含量光谱曲线变化特征

依据 5% 为间隔和全国第二次土壤普查 SOMC 分级标准对 SMC 和 SOMC 进行等级划分,并将各等级光谱数据进行平均化(图 3)。由图 3a 可知,随着 SMC 从 <5% 增加到 35%~40%,整体光谱反射率逐渐降低,尤其在近红外(800~1 300 nm)和短波红外(1 300~2 400 nm)波段,这主要归因于水分吸收效应的增强,导致在 1 400, 1 900, 2 200 nm 波段形成显著

吸收峰,吸收深度随水分含量的增加而加深。同时,低水分含量的土壤光谱表现出更高的反射率和明显的纹理特征。同样地,图 3b 表明随 SOMC 的增加(<6 g/kg 到 20~30 g/kg)整体光谱反射率逐渐下降,尤其在近红外(900~1 300 nm)和短波红外(2 000 nm 附近)波段表现明显,反映出有机质中羰基(C=O)或羟基(OH)基团的吸收特征。高 SOMC (>20 g/kg)的土壤光谱吸收特征更为显著,而低 SOMC (<6 g/kg)则表现出较高的反射率。

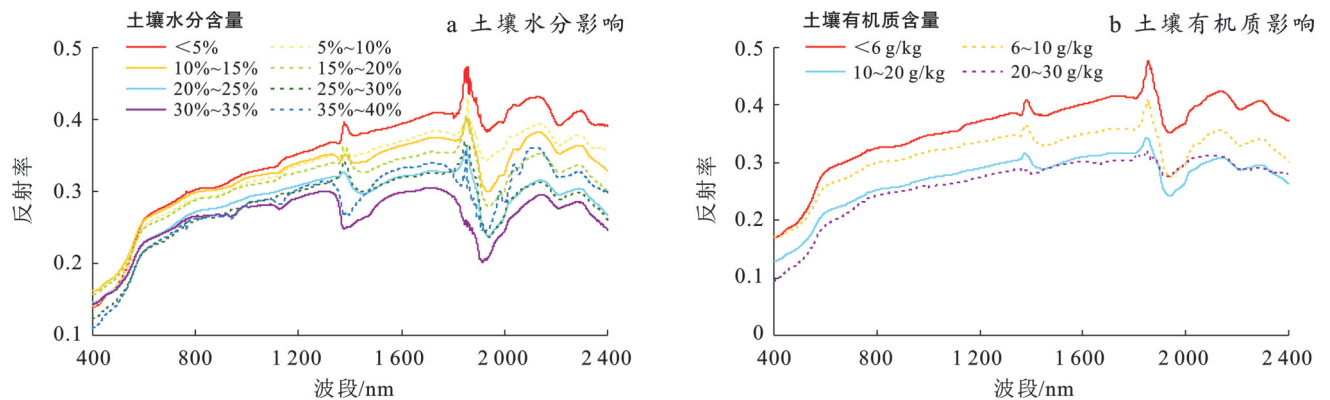


图 3 不同土壤水分和有机质含量原始光谱曲线变化特征

Fig.3 Variation characteristics of raw spectral curves with varying soil moisture and organic matter contents

2.3 分数阶微分变换后的光谱曲线变化特征

图 4 中显示了高光谱反射率经不同微分阶数变换过程光谱曲线的变化特征。从 0 阶即原始光谱可以看出,反射率曲线整体平滑,主要反映目标物体的固有光谱特征,曲线细节变化较少。随着阶数从 0.25 逐渐增加至 0.75 阶,光谱曲线的细节特征逐渐显现,

特别是在波峰与波谷附近的变化更为明显,但整体曲线仍保持一定的平滑性,此阶段经微分变换有效增强了光谱特征的梯度信息,适用于提取低频特征。从 1.0 阶开始,波峰和波谷的振荡显著增强,波形变化更加清晰,有助于区分相邻波段间的差异,但同时可能放大光谱中的噪声。进一步提高阶数至 1.25 阶

至 1.75 阶后,高频振荡特征愈发明显,曲线的平滑性减弱,信号中的细微变化被进一步放大,但噪声也随

之增加。最终在 2.0 阶高阶整数微分时,光谱曲线表现出较大幅度的高频振荡,信噪比显著下降。

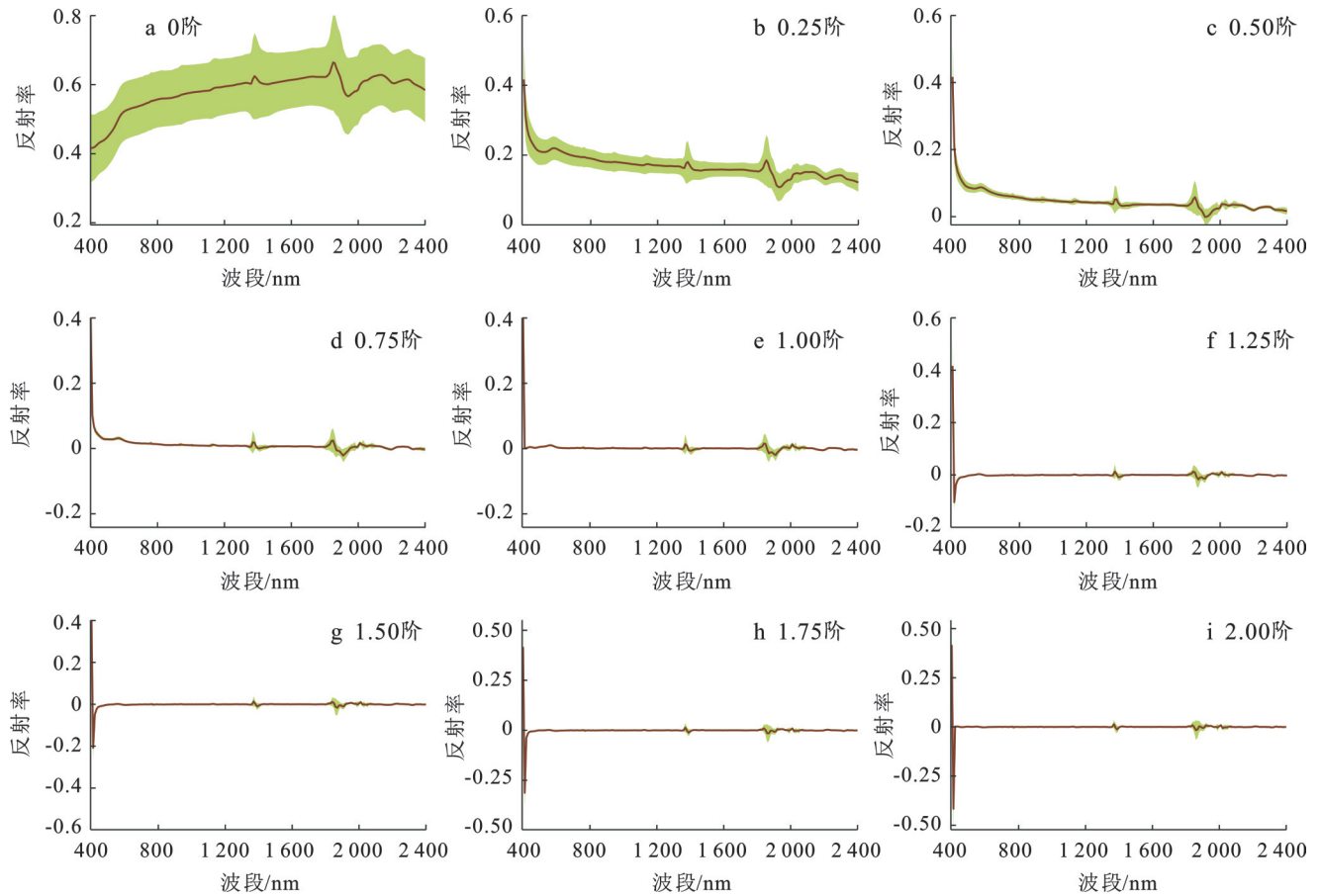


图 4 基于分数阶微分变换的光谱曲线变化特征

Fig.4 Spectral curve variation characteristics based on fractional differential transformation

#### 2.4 土壤水分和有机质含量对应光谱指数及其二维相关性

基于 1.25 阶微分变换后的光谱反射率与 SMC 和 SOMC 间构建的光谱指数相关性最强,两种属性与各光谱指数以及对应最优波段组合结果见表 3。可以看出,SMC 与 GDI 表现出最高的相关性,最大相关系数绝对值 (maximum absolute correction coefficient, MACC) 为 0.505 4,最优波段组合为 (1 190, 1 550 nm)。其次,OSI 和 DI 的相关性也较高,分别为 0.503 9 和 0.501 4。而 RI 的相关性最低,MACC 仅为 0.457 3,说明其在水分含量监测中的效果相对较弱。对于 SOMC 的相关性分析,OSI 在所有光谱指数中表现最优,MACC 达到 0.682 5,其最优波段组合为 (1 990, 1 610 nm)。此外,SASI 和 SI 的相关性也较高,分别为 0.671 2 和 0.586 3。而 PI 的相关性较低,MACC 值为 0.592 2。从波段组合来看,短波红外区域的波段(如 1 190, 1 550 nm)与 SMC 和 SOMC 的相关性中表现突出。此外,近红外和可见光波段(如 810 nm)也在 SOMC 的敏感性分析中起到重要作用。

图 5 为 SMC 与各光谱指数间的二维相关矩阵热图。由图 5 可知,各光谱指数与 SMC 的相关性具有显著的波段依赖性,其中在中红外波段(1 400, 1 900 nm 附近)表现出较强的正负相关性,而在短波长区域(400~1 000 nm)相关性较弱,多呈现绿色或浅蓝色,表明短波长波段对 SMC 的敏感性较低。具体而言,DI, SI 和 GDI 在 1 400, 1 900 nm 等关键水分吸收带附近具有显著相关性,反映了其对水分特征的良好响应;RI 和 SASI 在中红外范围内表现出较高的正负相关性,而 PI 和 OSI 的相关性分布则较为复杂,表现为高频振荡。此外,NPDI 在短波长(500~800 nm)和中红外波段(1 400, 1 900 nm)均表现出较高相关性,体现出其对 SMC 特征的有效捕获能力。

图 6 为 SOMC 与各光谱指数间的二维相关矩阵热图。各光谱指数与 SOMC 在中红外波段(1 600~2 400 nm)和近红外波段(800~1 300 nm)表现出较高的相关性。DI 和 GDI 在近红外波段与 SOMC 表现出较强的正相关,而在中红外波段,负相关显著增强。RI 和 OSI 则在近红外与中红外波段均表现出正负交

替的分布,表明其构建机制对波段间比值关系的敏感性较强。PI和SASI在短波长(400~700 nm)与中红外波段的相关性较低,多呈现绿色或浅蓝色区域。此外,SI和NPDI在中红外波段相关性显著增强,表

明它们对 SOMC 吸收特征的捕捉能力较强,其中 NPDI 的相关性在多个波段交替出现高相关性区间。总体来看,8种光谱指数在不同波段对 SOMC 的响应特性差异显著。

表 3 土壤水分和有机质含量光谱最优变换阶次及光谱指数二维相关性  
Table 3 Optimal spectral transformation order and two-dimensional correlation of spectral indices for soil moisture and organic matter content

光谱指数	土壤水分			土壤有机质含量		
	最优阶次	MACC	光谱组合/nm	最优阶次	MACC	光谱组合/nm
DI	1.25	0.501 4**	(2 210,1 640)	1.25	0.670 9**	(1 190,810)
PI	1.25	0.471 4*	(1 340,540)	1.25	0.592 2**	(1 530,810)
RI	1.25	0.457 3*	(1 840,840)	1.25	0.544 1**	(2 160,850)
SI	1.25	0.496 1*	(1 120,1 460)	1.25	0.586 3**	(1 320,1 130)
GDI	1.25	0.505 4**	(1 190,1 550)	1.25	0.535 9**	(1 530,810)
OSI	1.25	0.503 9**	(1 340,590)	1.25	0.682 5**	(1 990,1 610)
NPDI	1.25	0.463 4*	(1 340,2 200)	1.25	0.570 1**	(2 300,490)
SASI	1.25	0.501 7**	(2 210,1 640)	1.25	0.671 2**	(1 190,810)

注:\*\*表示在  $p < 0.01$  水平相关显著; \*表示在  $p < 0.05$  水平相关显著。各光谱参数名称见表 1。下同。

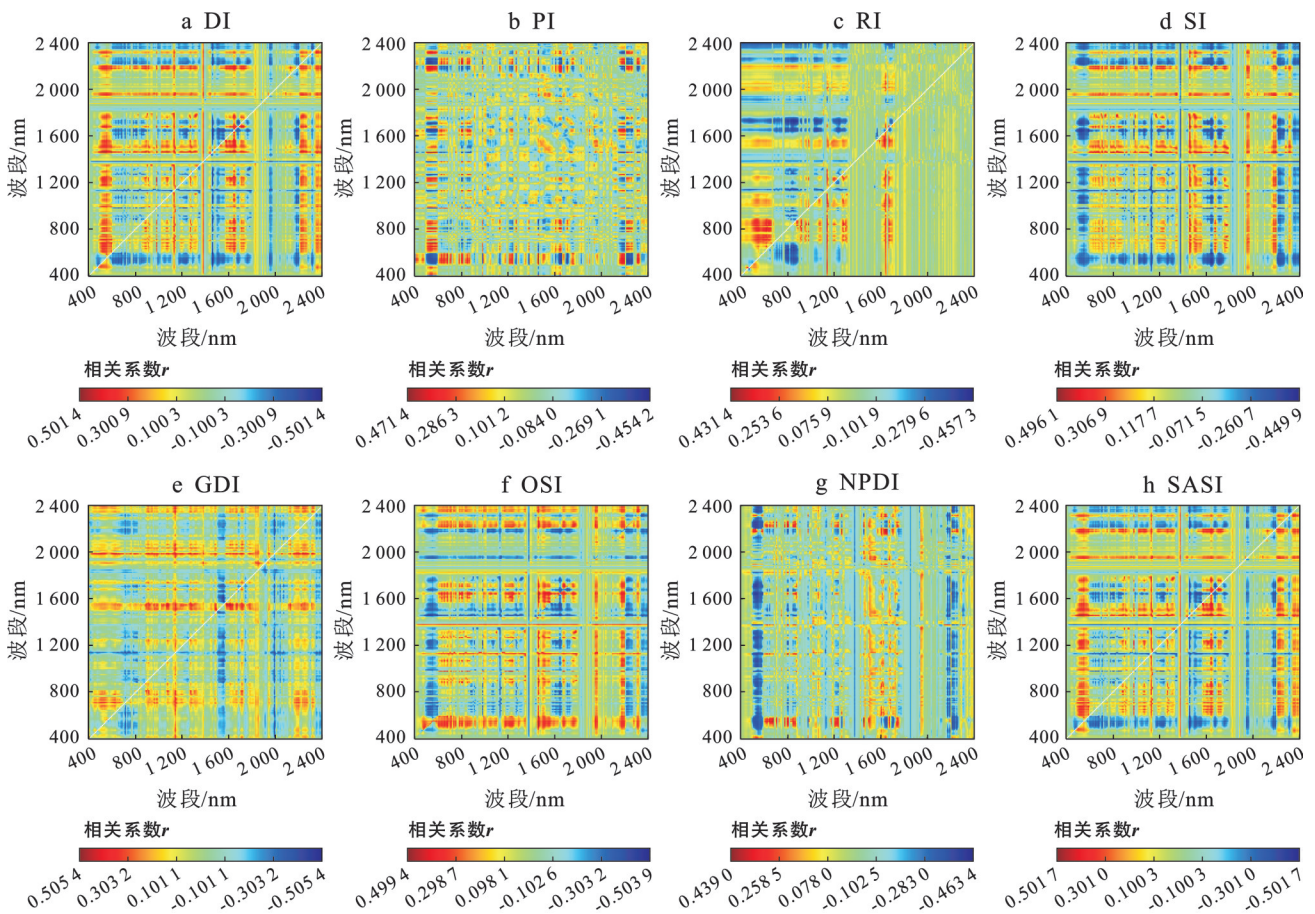


图 5 土壤水分含量与光谱指数间的二维相关性矩阵

Fig.5 Two-dimensional correlation matrix between soil moisture and spectral index

2.5 模型的建立与验证

表 4 为基于 1.25 阶微分变换联合光谱指数的 SMC 和 SOMC 建模结果。相比于 PLSR 和 SVM, RF 模型表现出显著的优势。

在 SMC 建模中,RF 的验证集拟合优度  $R^2$  达到 0.734, RMSE 最小为 3.28, 相对分析误差(RPD)为 2.07, 达到有效预测水平( $RPD > 2.0$ ), 而 PLSR 和 SVM 的  $R^2$  分别为 0.248 和 0.212, RPD 均低于 2.0, 预

测能力有限。

在 SOMC 建模中,RF 模型的验证集  $R^2$  高达 0.870, RMSE 为 1.53, RPD 达到 2.43, 同样优于 PLSR 和 SVM, 后两者的  $R^2$  分别为 0.322 和 0.369,

RPD 均为 1.12, 未达到有效预测水平。综合来看,RF 模型能够更好地处理光谱指数与 SMC 和 SOMC 间的非线性关系,其鲁棒性和对复杂特征的适应性使其在 SMC 和 SOMC 建模中表现最优。

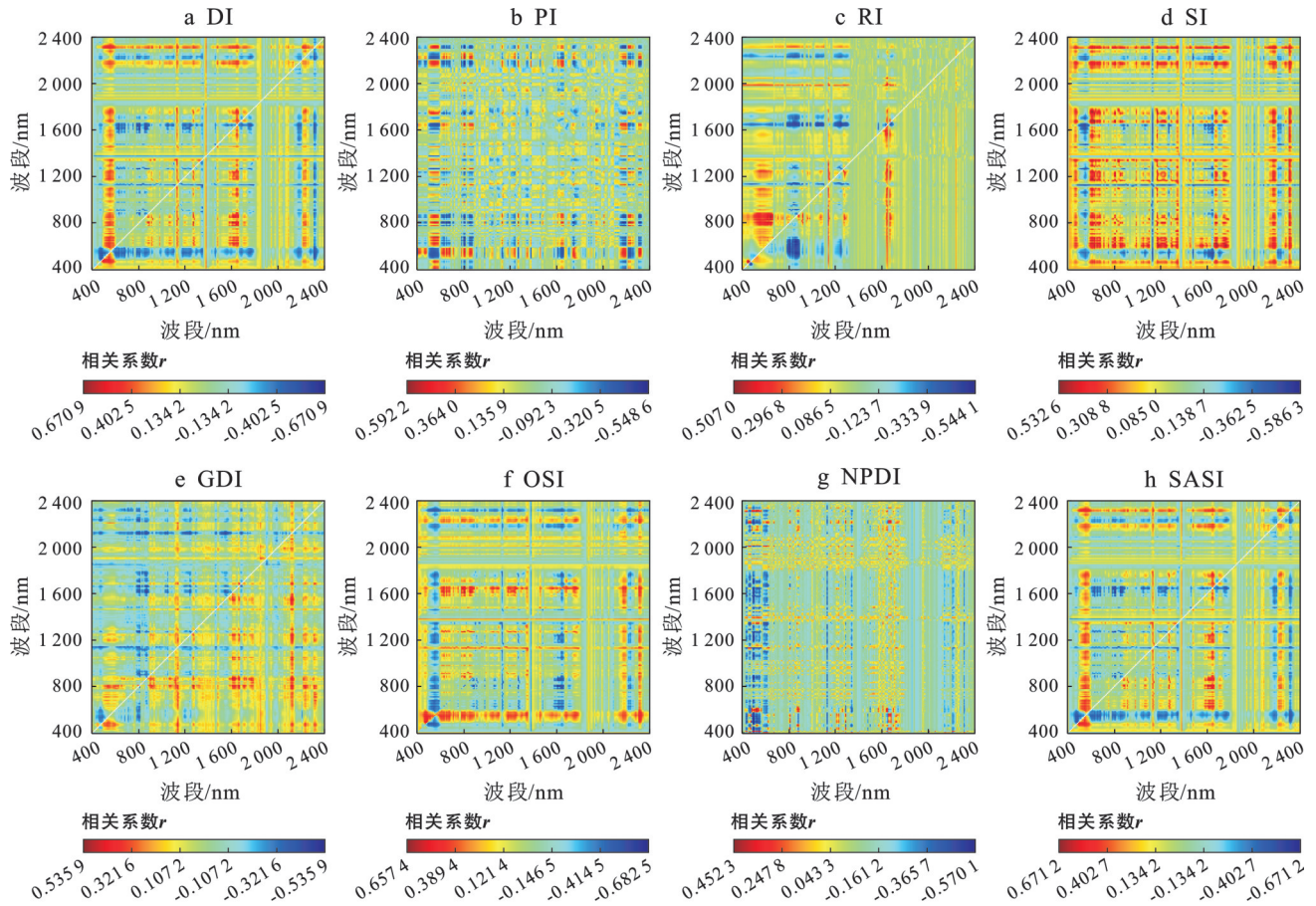


图 6 土壤有机质含量与光谱指数间的二维相关性矩阵

Fig.6 Two-dimensional correlation matrix between soil organic matter content and spectral index

表 4 模型精度与验证结果

Table 4 Model accuracy and validation results

方法	土壤水分				土壤有机质含量			
	$R_c^2$	$R_p^2$	RMSE	RPD	$R_c^2$	$R_p^2$	RMSE	RPD
PLSR	0.270	0.248	5.18	1.31	0.341	0.322	3.33	1.12
SVM	0.284	0.212	5.67	1.20	0.428	0.369	3.31	1.12
RF	0.798	0.734	3.28	2.07	0.896	0.870	1.53	2.43

## 2.6 模型可视化解释

基于 SHAP 值的全局解释结果揭示了 SMC 和 SOMC 预测模型中不同变量的贡献度(图 7)。在 SMC 模型中,NPDI 以最高 SHAP 值对模型输出贡献最大,贡献度占比达 34.29%。其次,OSI 和 DI 贡献度分别占 20% 和 14.29%,而 RI 和 GDI 等变量则对模型提供了补充信息。在 SOMC 模型中,RI 贡献度最高,占比 23.20%,表明其敏感地捕捉了有机质深色特性及其光谱响应机制,GDI 和 NPDI 分别以

20.62% 和 18.04% 的贡献度紧随其后。整体来看,SMC 和 SOMC 估算主导变量的差异反映了两种土壤属性光谱响应机理的不同,其中 SMC 的主导变量主要与吸收和散射路径变化相关,而 SOMC 的主导变量则与其深色特性及特定波段吸收能力密切相关。该结果从全局视角明确了变量重要性和光谱机理关系,不仅验证了模型构建的科学性,也为土壤属性精准预测和模型优化提供了理论依据与技术支持。

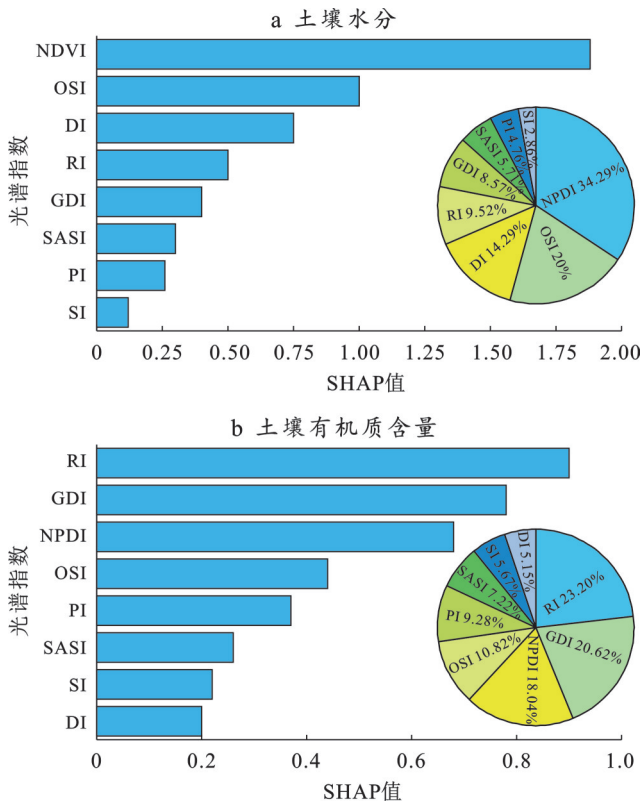


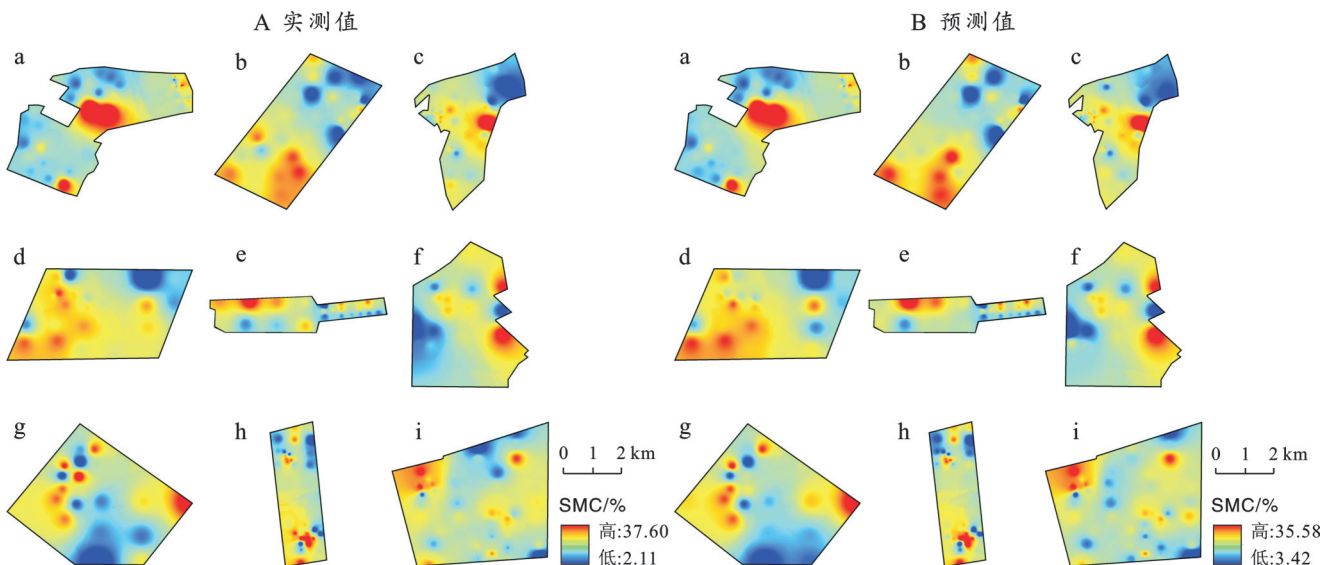
图7 基于SHAP解释的土壤水分和有机质含量RF模型全局分析

Fig.7 Global analysis of RF models for soil moisture and organic matter content based on SHAP interpretation

### 2.7 数字土壤制图

基于SMC实测值和1.25阶-RF模型预测值,利用IDW进行空间插值(图8)。从SMC实测值(8A)与预测值(8B)的空间分布对比可以看出,RF模型整体上能够较好地反映不同区域SMC的分布趋势,尤其在红寺堡、前进农场和西大滩等区域,高、低水分区域的分布位置较为一致。然而,部分区域如兴庆和交济,实测值与预测值在局部存在一定偏差,表现为高值低估和低值高估现象,但该现象仅在小范围区域存在。总体来看,RF模型在捕捉高、低水分极端值方面表现良好,且模型具有较高的预测性能和稳定性。

图9为基于实测SOMC(9A)和1.25阶-RF模型预测值(9B)的空间插值分布结果。整体来看,RF模型整体上较好地反映了各区域SOMC的空间分布趋势,特别是在红寺堡、前进农场和西大滩等区域,高值区与低值区的分布位置基本吻合。然而,在兴庆、交济和杭锦后旗等区域,实测值与预测值在局部高值或低值的空间位置上存在一定差异,部分区域的预测值在高值区略低于实测值。而惠农和交济在低值区的分布位置表现出较好的一致性,但模型对部分中值区域的空间过渡特征描述略显平滑。



注:图中a-i代表9个典型示范区,依次为红寺堡区,前进农场,西大滩,兴庆区,交济村,宝丰镇,惠农区,杭锦后旗,五原县。下同。

图8 基于RF模型的土壤水分含量空间分布

Fig.8 Spatial distribution of soil moisture based on RF model

## 3 讨论

### 3.1 分数阶微分和光谱指数与目标变量的适用性差异

高光谱遥感在土壤属性估算中的应用已得到广

泛证实,而FOD作为光谱变换的新方法,在突出特征光谱信息方面显示出优良潜力。整数阶微分在增强光谱特征的同时,引入高频噪声可能性较大<sup>[7]</sup>。FOD可以灵活调整阶数,兼顾光谱数据的高频和低频信息,从而实现光谱特征的平滑提取和噪声抑制。但

也有研究表明经一阶微分变换对 SOMC 建立的 SVM 模型反演精度最高<sup>[24]</sup>。本研究通过比较不同阶数的变换效果发现 1.25 阶微分变换对增强 SMC 和 SOMC 相关光谱特征更有效,这与在江汉平原 SOMC 估算中得出的结论高度一致<sup>[25]</sup>。这表明在不同研究地区,SOMC 在光谱中的体现在本质上具有相似性。而 Liu Ya 等<sup>[26]</sup>表明 0.8 阶微分变换对 SOMC 估算效果最佳。Ge Xiangyu 等<sup>[27]</sup>则发现 0.8 阶变换驱动模型对土壤盐分展示最佳估算效果。更有研究发现 1.5 阶微分变换对 SMC 的预测效果最佳<sup>[28]</sup>。因此,不同地区不同土壤属性甚至相同类型地区相同土壤属性对应的最佳微分变换仍存在差异。原因在

于不同地区的土壤类型、矿物质组成等因素不同,进而影响到光谱信号的反射特征,因此最佳微分结果存在差异。此外,气候条件对土壤的物理化学特性和光谱响应也会产生影响,从而进一步影响了最佳微分阶数的选择<sup>[29]</sup>。光谱指数通过特定波段的组合反映目标参数的物理或化学特性。本研究发现 GDI 和 OSI 分别与 SMC 和 SOMC 间相关性最强。SMC 主要通过影响近红外区域的吸收特性而表现出显著的光谱响应,而 SOMC 则通过其对光谱信号的散射作用在可见光和近红外区域形成独有的特征波段<sup>[30]</sup>。GDI 通过反射率差异突出水分变化特征,而 OSI 通过加权组合波段反射率强化了 SOMC 的光谱响应<sup>[31]</sup>。

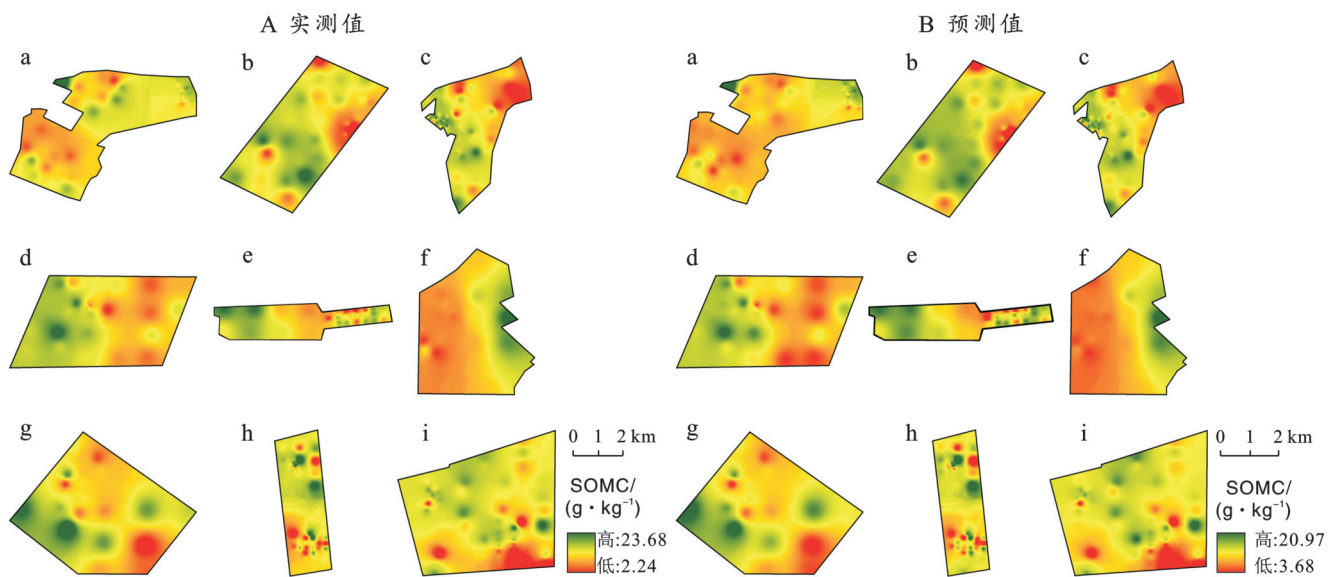


图9 基于RF模型的土壤有机质含量(SOMC)空间分布

Fig.9 Spatial distribution of soil organic matter content (SOMC) based on RF model

### 3.2 不同建模方法对土壤属性估算性能的差异分析

本研究发现RF算法在SMC和SOMC估算中的表现显著优于PLSR和SVM,这一结果既与模型自身的特性有关,也与土壤属性的光谱响应机制密切相关。PLSR作为一种线性回归方法,其建模能力依赖于目标变量与光谱特征之间的线性关系。而在复杂的土壤系统中,目标属性与光谱信号之间的关系通常是非线性的。如SOMC通过其对光谱散射的影响表现出复杂的多峰特征,而PLSR模型由于其线性假设的局限性,难以有效捕捉这一特征。刘英等<sup>[32]</sup>在矿区开展的SMC预测研究中,建立的PLSR模型 $R^2$ 达到0.640,该结果远高于本研究PLSR模型精度。原因主要源于研究区域的土壤特性和光谱响应之间的差异。首先,矿区的土壤可能相对均质,光谱特征较为一致,因此PLSR模型在矿区的SMC预测中取得较好结果。而在盐碱化农田,土壤盐碱性的空间异质性导致光谱响应更加复杂,且土壤的物理化学

性质可能导致光谱特征的高度变异。此外,矿区和农田的土壤样本数量、空间分布和采样密度也可能不同,进而影响模型的泛化能力和预测精度。SVM尽管通过核函数的引入一定程度上能够表征非线性关系,但其建模过程对核函数的选择和参数优化高度敏感,且在高维特征空间中容易出现过拟合或计算效率降低的问题<sup>[33]</sup>,导致其在本研究中的表现一般。相比之下,RF模型通过随机抽样和决策树的集成结构,能够在不依赖特定数据分布假设的情况下,捕捉光谱变量与目标属性之间的非线性交互关系,从而在高维光谱数据建模中表现出明显的优势<sup>[34]</sup>。

### 3.3 SHAP值分析揭示变量的贡献与变量优选建议

本研究基于SHAP值分析揭示了不同光谱指数在SMC和SOMC预测中的贡献存在的差异性。具体而言,NPDI,OSI和DI在SMC的估算中贡献度明显高于其他指数,而在SOMC预测中,RI,NPDI和GDI的贡献度则更加突出。这一结果不仅揭示了不

同土壤属性对光谱指数的依赖模式的多样性,还反映出土壤属性与光谱特征之间潜在的非线性关系与高度复杂的交互作用。进一步的分析表明,尽管SMC和SOMC分别与GDI和OSI之间的相关性最强,但SHAP值的分布却揭示了这些变量对具体样本的实际贡献并不完全一致。意味着单纯的相关性强并不等同于变量在模型中的贡献度高。GDI在SMC较高区域可能受限于其他表面特征,而在SOMC预测中,光谱特征可能受到更多的非线性因素干扰,如SOMC分解程度、土壤类型及其微生物活动变化等。因此,在特定样本或区域中,相关性较强的变量可能因其信号表达的复杂性或受其他因素的调节而对模型贡献度较低。

基于以上结论,在实际应用中,应将相关性分析作为特征预筛选工具,而将SHAP或其他模型内部重要性度量方法作为核心建模变量筛选手段。此外,针对光谱变量间高度共线的特性,建议结合特征降维与嵌入式选择方法,提升模型稳定性与泛化能力。这一发现强调了变量选择时不应仅依赖于相关性的强弱来判断变量的贡献度,而要结合变量与目标属性之间潜在的机制性联系以及其在不同环境背景下的表现。因此,未来研究应进一步探讨土壤属性估算中的复杂性,特别是在土壤类型和地理环境异质性影响下,如何更加精准地选取与目标属性最相关的特征,以提升模型的泛化能力和预测精度。

## 4 结论

本文采用FOD(步长0.25)对原始高光谱数据进行变换,并基于不同阶次微分变换后的光谱数据构建8种光谱指数,依据各光谱指数与SMC和SOMC间的相关性来筛选最佳变量,基于PLSR,SVM和RF算法构建SMC和SOMC估算模型,得到以下结论。

(1) 河套平原平均SMC为17.81%,但SOMC平均值为9.39 g/kg,整体水平较低。光谱数据经1.25阶微分变换构建的光谱指数与SMC和SOMC间相关性最强,其中,SMC和SOMC分别与GDI和OSI间相关性最强,MACC分别为0.505 4和0.682 5。

(2) RF模型对SMC和SOMC预测效果最优,SVM次之,PLSR效果最差,SMC和SOMC-RF模型 $R^2$ 分别为0.734和0.870,RF模型较其他2种方法更适于干旱地区盐碱农田SMC和SOMC的估算。

(3) SHAP结果表明,NPDI,OSI和DI是SMC建模估算的最佳光谱指数,而SOMC最佳估算光谱指数分别为RI,GDI和NPDI。与SMC和SOMC相关性最强的光谱指数对2种属性的建模估算贡献均

不是最高。NPDI,OSI,RI和DI是盐碱化SMC和SOMC建模估算的最佳光谱指数变量。

## 参考文献(References)

- [1] 陆宝金,田生昌,左忠,等.盐渍化土地可持续利用研究综述及展望[J].宁夏大学学报(自然科学版),2023,44(1):79-88.  
Lu Baojin, Tian Shengchang, Zuo Zhong, et al. Review and prospect on sustainable utilization of salinized land [J]. Journal of Ningxia University (Natural Science Edition), 2023,44(1):79-88.
- [2] 丁启东,王怡婧,张俊华,等.利用CARS算法联合协变量估算盐碱农田土壤水分和有机质含量[J].应用生态学报,2024,35(5):1321-1330.  
Ding Qidong, Wang Yijing, Zhang Junhua, et al. Estimation of soil moisture and organic matter content in saline alkali farmland by using CARS algorithm combined with covariates [J]. Chinese Journal of Applied Ecology, 2024,35(5):1321-1330.
- [3] Geng Jing, Tan Qiuyuan, Lv Junwei, et al. Assessing spatial variations in soil organic carbon and C:N ratio in northeast China's black soil region: Insights from Landsat-9 satellite and crop growth information [J]. Soil and Tillage Research, 2024,235:105897.
- [4] Zhao Mingsong, Gao Yingfeng, Lu Yuanyuan, et al. Hyperspectral modeling of soil organic matter based on characteristic wavelength in East China [J]. Sustainability, 2022,14(14):8455.
- [5] 丁启东,王怡婧,张俊华,等.基于高光谱信息的宁夏土壤有机质含量估算[J].生态学杂志,2024,43(8):2531-2538.  
Ding Qidong, Wang Yijing, Zhang Junhua, et al. Estimation of soil organic matter content in Ningxia based on hyperspectral information [J]. Chinese Journal of Ecology, 2024,43(8):2531-2538.
- [6] 李嘉琦,冯宇华,陈署晔,等.基于高光谱的土壤有机质及全氮估测[J].新疆农业科学,2024,61(10):2491-2499.  
Li Jiaqi, Feng Yuhua, Chen Shuhuang, et al. Estimation of soil organic matter and total nitrogen based on hyperspectral technology [J]. Xinjiang Agricultural Sciences, 2024,61(10):2491-2499.
- [7] 丁启东,王怡婧,张俊华,等.基于高光谱信息的宁夏引黄灌区中低产田土壤水分和有机质含量估算[J].应用生态学报,2023,34(11):3011-3020.  
Ding Qidong, Wang Yijing, Zhang Junhua, et al. Estimation of soil water and organic matter content in medium and low yield fields of Ningxia Yellow River irrigation area based on hyperspectral information [J]. Chinese Journal of Applied Ecology, 2023,34(11):3011-3020.
- [8] 尚天浩,陈睿华,张俊华,等.基于分数阶微分联合光谱指数估算银川平原土壤有机质含量[J].应用生态学报,

- 2023,34(3):717-725.
- Shang Tianhao, Chen Ruihua, Zhang Junhua, et al. Estimation of soil organic matter content in Yinchuan Plain based on fractional derivative combined with spectral indices [J]. *Chinese Journal of Applied Ecology*, 2023, 34(3):717-725.
- [9] 王怡婧,陈睿华,张俊华,等.基于分数阶微分技术的土壤水盐信息高光谱反演[J].*应用生态学报*,2023,34(5):1384-1394.
- Wang Yijing, Chen Ruihua, Zhang Junhua, et al. Hyperspectral inversion of soil water and salt information based on fractional order derivative technology [J]. *Chinese Journal of Applied Ecology*, 2023,34(5):1384-1394.
- [10] Ben Abbes A, Jarray N, Farah I R. Advances in remote sensing based soil moisture retrieval: Applications, techniques, scales and challenges for combining machine learning and physical models [J]. *Artificial Intelligence Review*, 2024,57(9):224.
- [11] 仇皓雷,王海燕.机器学习在土壤性质预测研究中的应用进展[J].*生态学杂志*,2025,44(1):283-294.
- Qiu Haolei, Wang Haiyan. Application of machine learning to the prediction of soil properties: A review [J]. *Chinese Journal of Ecology*, 2025,44(1):283-294.
- [12] 李天乐,赵泉华,贾淑涵,等.土壤有机质含量反演方法:土壤高光谱数据微分-瞬时频率变换[J].*地球信息科学学报*,2024,26(7):1733-1744.
- Li Tianle, Zhao Quanhua, Jia Shuhan, et al. Soil organic matter content inversion methods: Differential-instantaneous frequency transform of soil hyperspectral data [J]. *Journal of Geo-Information Science*, 2024, 26(7):1733-1744.
- [13] 尚天浩,贾萍萍,孙媛,等.宁夏银北地区盐碱化土壤水分光谱特征及模型拟合精度分析[J].*水土保持通报*,2020,40(4):183-189.
- Shang Tianhao, Jia Pingping, Sun Yuan, et al. Spectral characteristics of soil moisture in salinized soil and model fitting accuracy in northern Yinchua City, Ningxia Hui Autonomous Region [J]. *Bulletin of Soil and Water Conservation*, 2020,40(4):183-189.
- [14] 李武耀,买买提·沙吾提,买买木提·巴拉提.基于分数阶微分的土壤有机质含量高光谱反演研究[J].*激光与光电子学进展*,2023,60(7):404-411.
- Li Wuyao, Mamat Sawut, Maihemuti Balati. Fractional differential-based hyperspectral inversion of soil organic matter content [J]. *Laser & Optoelectronics Progress*, 2023,60(7):404-411.
- [15] Rohmer J, Belbeze S, Guyonnet D. Insights into the prediction uncertainty of machine-learning-based digital soil mapping through a local attribution approach [J]. *Soil*, 2024,10(2):679-697.
- [16] Pan Mengyao, Xia Bisheng, Huang Wenbo, et al. PM<sub>2.5</sub> concentration prediction model based on random forest and SHAP [J]. *International Journal of Pattern Recognition and Artificial Intelligence*, 2024, 38(5): 2452012.
- [17] 叶森,朱琳,刘旭东,等.基于连续小波变换、SHAP和XGBoost的土壤有机质含量高光谱反演[J].*环境科学*,2024,45(4):2280-2291.
- Ye Miao, Zhu Lin, Liu Xudong, et al. Hyperspectral inversion of soil organic matter content based on continuous wavelet transform, SHAP, and XGBoost [J]. *Environmental Science*, 2024,45(4):2280-2291.
- [18] 付平凡,杨晓静,苏志诚,等.基于集成学习的土壤含水量预测研究:以辽西地区为例[J].*土壤*,2023,55(3):671-681.
- Fu Pingfan, Yang Xiaojing, Su Zhicheng, et al. Prediction of soil moisture content based on ensemble learning: A case study of western Liaoning Province [J]. *Soils*, 2023,55(3):671-681.
- [19] Lao Congcong, Chen Junying, Zhang Zhitao, et al. Predicting the contents of soil salt and major water-soluble ions with fractional-order derivative spectral indices and variable selection [J]. *Computers and Electronics in Agriculture*, 2021,182:106031.
- [20] Dharumarajan S, Lalitha M, Kalaiselvi B, et al. Remote sensing of soils: Spectral signatures and spectral indices [M]. *Remote Sensing of Soils*, 2024.
- [21] 施光耀,杨思琪,张劲松,等.基于机器学习算法的高海拔地区臭氧影响因素重要性分析[J].*宁夏大学学报(自然科学版)*,2024,45(2):196-202.
- Shi Guangyao, Yang Siqi, Zhang Jinsong, et al. Importance analysis of ozone influencing factors in high-altitude regions based on machine learning algorithms [J]. *Journal of Ningxia University (Natural Science Edition)*, 2024,45(2):196-202.
- [22] 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.
- Huang Huayu, Ding Qidong, Zhang Junhua, 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.
- [23] Letif M, Bahar R, Mezouar N. The use of machine learning models and SHAP interaction values to predict the soil swelling index [J]. *Periodica Polytechnica Civil Engineering*, 2025,69(1):239-250.
- [24] 侯化刚,王丹阳,马斯琦,等.黄河三角洲不同盐渍度土壤有机质含量的高光谱预测研究[J].*中国农业科学*,2023,56(10):1905-1919.
- Hou Huagang, Wang Danyang, Ma Siqi, et al. Hyperspectral prediction of organic matter in soils of different

- salinity levels in the Yellow River delta [J]. *Scientia Agricultura Sinica*, 2023, 56(10):1905-1919.
- [25] Hong Yongsheng, Liu Yaolin, Chen Yiyun, et al. Application of fractional-order derivative in the quantitative estimation of soil organic matter content through visible and near-infrared spectroscopy [J]. *Geoderma*, 2019, 337:758-769.
- [26] Liu Ya, Zhang Yuewen, Lu Huizhong, et al. Application of fractional-order differential and ensemble learning to predict soil organic matter from hyperspectra [J]. *Journal of Soils and Sediments*, 2024, 24(1):361-372.
- [27] Ge Xiangyu, Ding Jianli, Teng Dexiong, et al. Exploring the capability of Gaofen-5 hyperspectral data for assessing soil salinity risks [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2022, 112:102969.
- [28] Li Wangyang, Xiang Youzhen, Liu Xiaochi, et al. Estimation of soil moisture content based on fractional differential and optimal spectral index [J]. *Agronomy*, 2024, 14(1):184.
- [29] Yuan Jing, Gao Jichao, Yu Bo, et al. Estimation of soil organic matter content based on spectral indices constructed by improved Hapke model [J]. *Geoderma*, 2024, 443:116823.
- [30] Wu Dongqing, Lei Lei, Xie Meiting, et al. High-performance metamaterial light absorption from visible to near-infrared assisted by anti-reflection coating [J]. *Photonics*, 2023, 10(9):998.
- [31] Zhang Xinle, Dong Chang, Liu Huanjun, et al. Methodology for regional soil organic matter prediction with spectroscopy: Optimal sample grouping, input variables, and prediction model [J]. *Remote Sensing*, 2024, 16(3):565.
- [32] 刘英,范凯旋,裴为豪,等.基于PLSR和LSSVM模型的土壤水分高光谱反演[J].*矿业安全与环保*, 2024, 51(5):147-153.
- Liu Ying, Fan Kaixuan, Pei Weihao, et al. Hyperspectral inversion of soil moisture based on PLSR and LSSVM models [J]. *Mining Safety & Environmental Protection*, 2024, 51(5):147-153.
- [33] 陈睿华,尚天浩,张俊华,等.不同光谱类型对银川平原土壤含盐量反演精度的影响与校正[J].*应用生态学报*, 2022, 33(4):922-930.
- Chen Ruihua, Shang Tianhao, Zhang Junhua, et al. Effects of different spectra types on the accuracy and correction of soil salt content inversion in Yinchuan Plain, China [J]. *Chinese Journal of Applied Ecology*, 2022, 33(4):922-930.
- [34] Jia Pingping, He Wei, Hu Yi, et al. Inversion of coastal cultivated soil salt content based on multi-source spectra and environmental variables [J]. *Soil and Tillage Research*, 2024, 241:106124.
- ~~~~~
- (上接第9页)
- [20] Zhang Qingwei, Wang Zhanli, Wu Bing, et al. Identifying sediment transport capacity of raindrop-impacted overland flow within transport-limited system of interrill erosion processes on steep loess hillslopes of China [J]. *Soil and Tillage Research*, 2018, 184:109-117.
- [21] 雷志栋.土壤水动力学[M].北京:清华大学出版社, 1988:77-82.
- Lei Zhidong. *Soil hydrodynamics* [M]. Beijing: Qinghua University Press, 1988.
- [22] Mangangka I R. The decline of soil infiltration capacity due to high elevation groundwater [J]. *Civil Engineering Dimension*, 2008, 10(1):35-39.
- [23] La Follette P, Ogden F L, Jan A. Layered green and ampt infiltration with redistribution [J]. *Water Resources Research*, 2023, 59(7): e2022WR033742.
- [24] Chu Xuefeng, Mariño M A. Determination of ponding condition and infiltration into layered soils under unsteady rainfall [J]. *Journal of Hydrology*, 2005, 313(3/4):195-207.
- [25] Fox G A, Wilson G V. The role of subsurface flow in hillslope and stream bank erosion: A review [J]. *Soil Science Society of America Journal*, 2010, 74(3):717-733.
- [26] 武秀荣.片沙覆盖黄土区沙土二元结构坡面侵蚀过程研究[D].陕西 杨凌:西北农林科技大学, 2015.
- Wu Xiurong. *Studying erosion process of aeolian sand over loess surface in hilly region of the Loess Plateau covered by aeolian sand* [D]. Yangling, Shaanxi: Northwest A & F University, 2015.
- [27] 曹晓娟,谢林好,张风宝,等.沙层特性对沙盖黄土坡面产流产沙变化贡献的定量分析[J].*地理学报*, 2019, 74(5):962-974.
- Cao Xiaojuan, Xie Linyu, Zhang Fengbao, et al. Quantifying the contributions of sand layer characteristic to variations of runoff and sediment yields from sand-covered loess slopes during simulated rainfall [J]. *Acta Geographica Sinica*, 2019, 74(5):962-974.
- [28] Hotta W, Suzuki S, Hori M. On contraction of three-dimensional multiple shear mechanism model for evaluation of large scale liquefaction using high performance computing [J]. *Geosciences*, 2019, 9(1):38.
- [29] 汤珊珊,高海东,李占斌,等.坡面覆沙后侵蚀泥沙颗粒分选特性[J].*农业工程学报*, 2017, 33(2):125-130.
- Tang Shanshan, Gao Haidong, Li Zhanbin, et al. Characteristics of particle separation of erosion sediment in slop surface covered with sand [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2017, 33(2):125-130.