

刘东, 王世航, 赵明松, 刘峰, 徐胜祥. 基于高光谱和机器学习技术的矿区原状剖面土壤铜组分预测[J]. 土壤学报, 2026,  
LIU Dong, WANG Shihang, ZHAO Mingsong, LIU Feng, XU Shengxiang. Prediction of Cu Fractions in Intact Soil Profiles of Mining  
Areas Using Hyperspectral Imagining and Machine Learning [J]. Acta Pedologica Sinica, 2026,

# 基于高光谱和机器学习技术的矿区原状剖面土壤铜组分预测\*

刘东<sup>1,2</sup>, 王世航<sup>1</sup>, 赵明松<sup>1</sup>, 刘峰<sup>2,3</sup>, 徐胜祥<sup>2,3†</sup>

(1. 安徽理工大学空间信息与测绘工程学院, 安徽淮南 232001; 2. 中国科学院南京土壤研究所, 南京 211135;  
3. 中国科学院大学, 北京 100049)

**摘要:** 高光谱遥感技术在监测和评价土壤重金属污染方面具有较大的研究价值。为探究高光谱成像 (HSI:400~1010 nm) 光谱预测原状土壤剖面 5 种土壤铜形态属性的潜力: 总铜、弱酸可提取态铜 (F1)、可还原态铜 (F2)、可氧化态铜 (F3) 和残渣态铜 (F4)。本研究以江西省乐安河流域农田土壤为研究对象, 共采集了 22 个深度约 100 cm 的原状土壤剖面样品, 分别测定土壤剖面样品的光谱数据及其 Cu 含量, 建立土壤 Cu 含量预测模型, 比较偏最小二乘回归法 (Partial Least Squares Regression, PLSR)、随机森林 (Random Forest, RF)、Cubist 混合线性回归决策树 (Cubist Regression Tree, Cubist)、高斯过程回归 (Gaussian Process Regression, GPR) 和支持向量机 (Support Vector Machine, SVM) 方法与不同光谱预处理方法对土壤 Cu 含量预测精度的影响。结果显示, RF、Cubist、GPR 和 SVM 这 4 种机器学习算法在  $R^2$  上普遍优于线性 PLSR 模型, 表现出更高的预测精度; 采用吸光度转换结合一阶导数 (Abs+FD) 组合方法进行预处理后, 基于 SVM 建立的模型在独立验证集中对 5 种土壤铜属性的预测效果较好 (F1:  $R^2_p=0.78$ ,  $RMSE_p=0.56$  mg·kg<sup>-1</sup>; F2:  $R^2_p=0.78$ ,  $RMSE_p=0.40$  mg·kg<sup>-1</sup>; F3:  $R^2_p=0.67$ ,  $RMSE_p=1.33$  mg·kg<sup>-1</sup>; F4:  $R^2_p=0.70$ ,  $RMSE_p=2.91$  mg·kg<sup>-1</sup>; 总铜:  $R^2_p=0.67$ ,  $RMSE_p=3.64$  mg·kg<sup>-1</sup>)。这表明高光谱结合机器学习技术可实现对土壤剖面重金属多种形态含量的精准、快速预测, 对深理解解重金属在土体中的迁移转化规律及开展区域土壤污染风险评估具有重要意义。

**关键词:** 土壤重金属; 高光谱; 化学形态; 机器学习

中图分类号: TP79 文献标志码: A

## Prediction of Cu Fractions in Intact Soil Profiles of Mining Areas Using Hyperspectral Imagining and Machine Learning

LIU Dong<sup>1,2</sup>, WANG Shihang<sup>1</sup>, ZHAO Mingsong<sup>1</sup>, LIU Feng<sup>2,3</sup>, XU Shengxiang<sup>2,3†</sup>

(1. School of Geomatics, Anhui University of Science and Technology, Huainan, Anhui 232001, China; 2. Institute of Soil Science, Chinese Academy of Sciences, Nanjing 211135, China; 3. University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** 【Objective】Hyperspectral remote sensing technology holds considerable research value for monitoring and assessing heavy metal contamination in soils. However, it is unclear how this technology can be used to detect different

\* 国家自然科学基金项目 (42271369) 和中国科学院南京土壤研究所“十四五”规划自主部署项目 (ISSASIP2202) 资助 Supported by the National Natural Science Foundation of China (No.42271369), and "14th Five-Year Plan" Autonomously Deployed Project of Institute of Soil Science, Chinese Academy of Sciences (No.ISSASIP2202)

† 通讯作者 Corresponding author, E-mail: sxxu@issas.ac.cn

作者简介: 刘东 (1997-), 男, 江苏徐州人, 硕士研究生, 主要从事数字土壤制图研究。E-mail:dongll1203@163.com  
收稿日期: 2025-10-08; 收到修改稿日期: 2026-03-16; 网络首发日期 (www.cnki.net):

heavy metal fractions in soil. **【Method】** This study collected 22 intact soil profile samples with depths of approximately 100 cm from farmland soils in the Le'an River Basin, Jiangxi Province, China. The samples were used to investigate the potential of hyperspectral imaging (HSI, 400-1010 nm) for predicting five copper (Cu) fractions in intact soil profiles, including total Cu, weak acid-extractable Cu (F1), reducible Cu (F2), oxidizable Cu (F3), and residual Cu (F4). After the spectral data and Cu contents of the soil profile samples were measured, prediction models for soil Cu contents were established. Several modeling methods were applied to investigate the effect of different spectral preprocessing techniques on prediction accuracy, including partial least squares regression (PLSR), random forest (RF), Cubist regression tree (Cubist), Gaussian process regression (GPR), and Support vector machine (SVM). **【Result】** The results show that the four machine learning algorithms, namely RF, Cubist, GPR, and SVM, generally outperformed the linear PLSR model in terms of  $R^2$ , demonstrating higher predictive accuracy. After preprocessing with the combined absorbance transformation and first derivative method (Abs+FD), the SVM-based model achieved relatively good predictive performance for the five soil Cu fractions in the independent validation set (F1:  $R_p^2 = 0.78$ ,  $RMSE_p = 0.56 \text{ mg}\cdot\text{kg}^{-1}$ ; F2:  $R_p^2 = 0.78$ ,  $RMSE_p = 0.40 \text{ mg}\cdot\text{kg}^{-1}$ ; F3:  $R_p^2 = 0.67$ ,  $RMSE_p = 1.33 \text{ mg}\cdot\text{kg}^{-1}$ ; F4:  $R_p^2 = 0.70$ ,  $RMSE_p = 2.91 \text{ mg}\cdot\text{kg}^{-1}$ ; Total Cu:  $R_p^2 = 0.67$ ,  $RMSE_p = 3.64 \text{ mg}\cdot\text{kg}^{-1}$ ). **【Conclusion】** These findings indicate that HIS combined with machine learning can effectively predict multiple heavy metal fractions in soil profiles, which is of great significance for improving our understanding of the migration and transformation of heavy metals in soil and for conducting regional soil pollution risk assessments.

**Key words:** Soil heavy metals; Hyperspectral; Chemical speciation; Machine learning

土壤是生态系统的重要组成部分，对生态环境质量和人类健康具有重要意义<sup>[1]</sup>。土壤中的重金属污染物主要来自土壤自然风化和人为活动，重金属污染会导致土壤环境质量的恶化和土壤生态功能的丧失<sup>[2]</sup>。长期以来，土壤环境质量标准主要基于重金属的总量。然而，大量研究和实践表明，重金属的环境效应、迁移能力和生物毒性并不主要取决于其总量，而是取决于其在土壤中存在的化学形态和空间分布<sup>[3-4]</sup>。重金属进入土壤后，会与土壤中各组分发生复杂的物理、化学反应，形成具有不同稳定性和活性的形态<sup>[5]</sup>。因此，研究土壤重金属的化学组分是从“表象”深入“机理”的关键，具有极其重要的理论和实践意义。

传统的土壤重金属含量测定方法主要依赖于化学分析，如电感耦合等离子体质谱（ICP-MS）、原子吸收分光光度法（AAS）、原子荧光光谱法（AFS）等，这些方法具有较高的准确性，但耗时长、成本高、效率低，且不适合分析大量土壤样本<sup>[6]</sup>。而高光谱成像（HSI）获取信息速度快、周期短、光谱分辨率高，其可见光和近红外区域（VNIR: 400~2500 nm）可提供高质量的空间和光谱信息，目前在预测土壤属性方面有了较为广泛的应用，例如预测土壤碳、氮、有机质<sup>[7-9]</sup>和土壤重金属含量<sup>[6,10-13]</sup>。其中，重金属高光谱反演的机理主要基于黏土矿物、氧化铁、有机质等重金属吸附物与不同重金属之间的内在相关性，根据重金属吸附物的光谱特性，实现土壤重金属含量的间接反演<sup>[14]</sup>。此外，大部分高光谱反演土壤重金属含量的研究主要针对 0~20 cm 表层土壤的重金属总量预测，且土壤样品普遍采取了风干、研磨、过筛等前处理。例如，Zheng 等<sup>[13]</sup>基于平潭岛 72 个表层土壤样点，建立了土壤砷浓度的高光谱反演模型。杞应涛等<sup>[12]</sup>采集了 56 个表层土壤样品，经风干和研磨等前处理后进行高光谱测定，对铅锌矿区 Pb、Zn 含量进行了预测。相较之下，针对原状土壤剖面重金属含量反演的研究尚不多见，目前仅有王滢艺等<sup>[6]</sup>对矿区原状土壤剖面总铜的高光谱反演进行了探索。

反演土壤重金属含量的预测模型根据其算法原理，大致分为线性模型与非线性模型两大类。常见的线性建模方法主要包括多元线性回归（MLR）和偏最小二乘回归（PLSR）等<sup>[13]</sup>。Yang 等<sup>[15]</sup>通过逐步多元回归（SMR）和 PLSR 建立高光谱预测模型，发现 PLSR 法建立的重金属含量预测模型具有更高的预测精度和更好的拟合效果， $R^2$  在 0.7 以上。周瑶等<sup>[10]</sup>在湘南某多金属矿区采集了 43 个土壤样品，采用多种光谱变换方法，建立了预测矿区土壤 As 含量的 PLSR 模型，结果显示优化光谱指数为自变量构建的 PLSR 模型性能更优（ $R^2=0.91$ ,  $RMSE=46.8 \text{ mg}\cdot\text{kg}^{-1}$ ）。随着人工智能的快速发展

展,机器学习模型已经表现出显著的预测能力,在土壤学领域获得了广泛应用。为估算土壤中重金属含量并解析高维光谱数据与土壤重金属含量之间的复杂非线性关系,诸多学者采用了非线性建模方法,并将光谱数据与机器学习算法相结合,显著提高了预测模型的精度。常见的机器学习算法包括支持向量机(Support Vector Machine, SVM)<sup>[6]</sup>、人工神经网络(Artificial Neural Network, ANN)<sup>[7]</sup>、高斯过程回归(Gaussian Process Regression, GPR)<sup>[16]</sup>、随机森林(Random Forest, RF)<sup>[17]</sup>、Cubist混合线性回归决策树(Cubist Regression Tree, Cubist)<sup>[18]</sup>等。Zhong等<sup>[19]</sup>在乌鲁木齐采集了88个土壤样品,获得了土壤镍含量和 Related 的高光谱数据,发现基于倒数一阶微分的RF模型反演土壤Ni含量的效果优于PLSR、SVM模型,最高精度可达 $R^2=0.86$ , RMSE=1.32 mg·kg<sup>-1</sup>。陶超等<sup>[20]</sup>在铅锌矿区采集了83组土壤样品,通过对比PLSR与SVM两种模型对土壤Pb和Zn含量反演效果,结果发现SVM在小样本数据和高维特征问题方面具有明显优势( $R^2>0.68$ )。Shang等<sup>[21]</sup>采用SVM和GPR等模型建立土壤高光谱数据与土壤Cu含量的关系,分析发现GPR是预测土壤Cu含量的最佳模型( $R^2=0.83$ , RMSE=364.61 mg·kg<sup>-1</sup>)。上述案例表明,非线性模型的预测性能通常优于线性模型,且不同的机器学习方法的预测性能也存在差异。

针对上述研究不足,本文以江西省乐安河沿岸农田为研究对象,利用高光谱成像(HSI)和机器学习算法对原状土壤剖面中总铜及其化学组分进行预测,通过PLSR和4种机器学习模型(RF、GPR、Cubist、SVM)预测土壤Cu含量,并比较不同光谱预处理方法对模型精度的影响,筛选预测该区域土壤剖面总铜及其化学组分的最佳模型,为其他利用高光谱数据预测土壤重金属组分的研究提供合理的建模思路和方法。

## 1 材料与方 法

### 1.1 研究区概况

研究区位于江西省乐安河流域,包括德兴市与乐平市,处在28°38′-29°18′N,116°53′-118°06′E之间(图1)。该区域属亚热带湿润季风气候区,气候温暖湿润,主要的土壤类型为水稻土、黄壤、红壤。德兴铜矿位于江西省德兴市,处于江西省东北部,是亚洲露天开采规模最大的铜矿,主要矿石矿物有黄铁矿和黄铜矿。德兴铜矿周边水系主要有三个流域,分别为大坞河、浮溪河和乐安河<sup>[4]</sup>。大坞河流经德兴铜矿区域腹地,受矿业活动影响最为显著,开采产生的废水通过大坞河排入乐安河,会对周边灌溉农田造成重金属污染。主要的重金属污染物包括As、Cd、Cr、Cu、Hg、Pb等元素<sup>[6]</sup>。

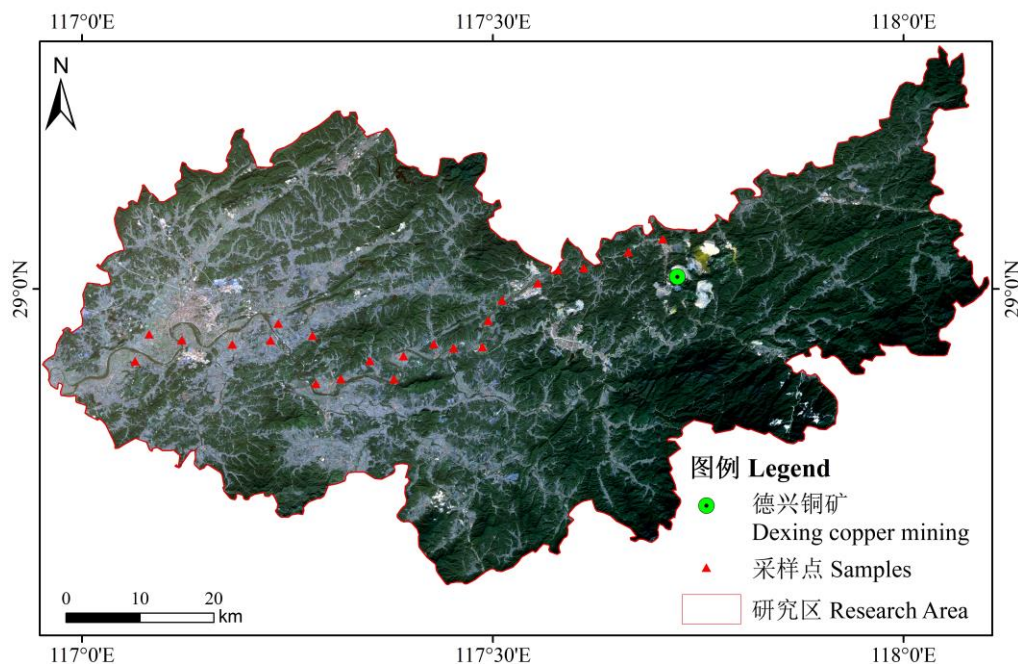


Fig. 1 Location of the study area and distribution of sampling sites

## 1.2 土壤剖面采集及预处理数据处理

根据污染物在乐安河流域的扩散方向,于 2022 年 12 月在河流沿岸农田布设 22 个采样点,共采集 22 个土柱样品。原状土柱样品用汽油动力钻(内径约 8.4 cm)采集,总共采集 22 个深度为 95~105 cm 的土柱样品,并使用手持 GPS 仪精确记录采样位置。样品采集后,使用 PVC 管和保鲜膜保存。

在实验室中用不锈钢丝将每个原状土芯垂直切成 2 个半径为 4.2 cm、高为 95~105 cm 的半土柱,挑选其中较为完整的半土柱,以 5 cm 为间距从表层向下划分成 19~21 个土层,共计 442 个土层。使用光谱范围为 396~1019 nm 的推扫式仪器在暗室中扫描整个土壤半剖面的 HSI 图像。该仪器主要由以下部件组成:ImSpector V10E 高光谱成像仪、高性能的电荷耦合器件相机(CCD)、由两个 150 W 卤钨灯组成的照明系统,以及定制的移动样品台。高光谱相机可获取 396~1019 nm 范围内的 1040 个波段,光谱分辨率约为 0.6 nm,每线成像包含 1024 个像素,单个像素尺寸为 40  $\mu\text{m}$ ×40  $\mu\text{m}$ <sup>[7]</sup>。从每个图像中依次选择 5 cm 间隔的矩形区域,对应送去实验室分析的样品。计算每个子样品的平均光谱,最终总共获得了 442 条反射光谱数据,将用于模型开发。

## 1.3 土壤样品分析

根据 BCR 提取法,将土壤中重金属形态分为弱酸可提取态(F1)、可还原态(F2)、可氧化态(F3)和残渣态(F4) 4 种形态<sup>[3,22-23]</sup>。根据光谱测量前划分出的 5 cm 间距,每个土壤剖面采集约 20 个土壤分析样品,总共得到 442 个土壤样品。土壤样品在室内经自然风干后用木棒碾碎,之后剔除其中的砂砾及植物残体等杂质,研磨后过 0.149 mm(100 目)筛。过筛后的土壤样品采用 0.11 mol·L<sup>-1</sup> 的 CH<sub>3</sub>COOH 溶液,经振荡、离心后分离上清液(用于测定 F1);向上一步的残渣中加入 0.5 mol·L<sup>-1</sup> 的 NH<sub>2</sub>OH-HCl 溶液,经振荡、离心后分离上清液(用于测定 F2);而后向上一步残渣中加入 8.8 mol·L<sup>-1</sup> 的 H<sub>2</sub>O<sub>2</sub> 溶液后添加 1 mol·L<sup>-1</sup> 的 NH<sub>4</sub>OAc,经振荡、离心后分离上清液(用于测定 F3)。最后使用 HCl-HF-HClO<sub>4</sub> 混合溶液对以上三步所剩余残渣进行消解,溶出存在于残渣中的重金属(用于测定 F4)。最终,使用 AAS 法分别测定土壤中 Cu 各组分含量。

## 1.4 土壤光谱数据预处理

土壤样品的结构特性和光谱仪器的工作状态等因素会影响反射光谱的测量,产生光谱随机噪声、基线漂移等,进一步影响模型的预测精度。光谱预处理主要用于消除光谱数据中无关信息和噪声,

使其更好地符合数据分析的目标。因此, 首先将反射光谱数据的波段范围缩减至 400~1010 nm, 以排除边缘背景噪声干扰。其次, 为了进一步提高信噪比、突出光谱的特征信息, 共采用 18 种光谱预处理方法及其组合对原始光谱反射率数据进行变换, 具体包括: 原始光谱反射率 (Raw)、吸光度转换 (Abs)、Gap-Segment 导数 (GS)、Savitzky-Golay 平滑 (SG)、一阶导数 (FD)、二阶导数 (SD)、多元散射校正 (MSC)、标准正态变量 (SNV)、去趋势化 (DT)、连续统去除 (CR)、Abs+SG、Abs+FD、Abs+SD、Abs+GS、GS+SNV、GS+MSC、SG+SNV、SG+MSC。

### 1.5 反演模型选择

利用主成分分析对采集到的 442 条光谱数据进行光谱均匀性检验, 通过计算 Hotelling T<sup>2</sup>椭圆进行判定, 结果显示所有样本均落在椭圆内, 未发现异常离散点。随后, 使用均值±3 倍标准差的方法<sup>[28]</sup>对土壤样品中总铜含量异常值进行剔除, 剔除 9 份异常样本。接着, 利用 Kennard-Stone (KS) 算法将 433 条光谱数据的 70% 作为建模集 (共 303 个样本), 30% 作为独立验证集 (共 130 个样本)。该算法是一种广泛应用于光谱分析和机器学习的经典样本选择方法, 可确保在特征空间中具有代表性和均匀分布的子样本<sup>[8]</sup>。基于 VNIR 和 Cu 含量实测值, 本研究采用 PLSR、Cubist、GPR、RF、SVM 在建模集上分别构建土壤剖面总铜及各组分的 VNIR 预测模型。

PLSR 是一种线性回归建模方法, 该方法是由主成分分析、相关分析和线性回归分析等多种思想发展而来的多元数据分析方法, 能够很好地处理自变量之间的多重共线性、样本数少于变量数以及计算复杂度高等问题, 是一种常用的线性建模方法<sup>[24]</sup>。

RF 是一种基于决策树的集成学习算法, 它通过构建多个决策树来进行预测和分类。随机森林算法是通过训练多个决策树来生成模型, 然后综合利用多个决策树进行分类。随机森林算法训练速度快, 无需特征选择即可处理高维数据, 具有很强的数据集适应性, 在土壤重金属高光谱反演中表现良好<sup>[13]</sup>。

Cubist 是一种基于树结构的局部线性模型, 其核心机制在于递归地划分预测变量空间<sup>[18]</sup>。它将数据集划分为具有相似特征子集, 并为每个子集生成线性回归规则, 选择最合适的预测变量进行回归。该模型还结合了基于实例的调整, 以提高预测准确性<sup>[25]</sup>。

GPR 是一种基于高斯随机函数的机器学习方法, 适合处理复杂的高维、小样本及非线性问题<sup>[26]</sup>。该模型可以对预测输出做出概率解释, 具有易实现、推断灵活、超参数自适应优化等优点。

SVM 是一种基于核方法的非线性机器学习模型, 它使用核函数来处理复杂的非线性问题, 具有良好的性能。使用核函数, SVM 将数据集映射到较原始空间更高维的特征空间, 使样本在新的特征空间中线性分离。SVM 可以提取关键样本信息, 减少冗余样本<sup>[16]</sup>。此外, SVM 在解决小样本、非线性和高维模式识别问题上也表现出许多特有的优势。

5 种模型的参数优化和构建在 R (4.4.2) 软件中进行, 利用 R 的 caret 包对每种建模方法的参数进行网格搜索, 根据十折交叉验证的结果确定均方根误差 (Root Mean Square Error, RMSE) 最小的参数组合为模型最优参数。每种建模方法都针对 18 种光谱预处理的光谱数据进行建模参数的网格搜索。

### 1.6 模型精度评估

采用决定系数 ( $R^2$ )、均方根误差 (RMSE)、相对分析误差 (RPD) 三个评价参数对交叉验证 (以下标 cv 表示) 和独立验证 (以下标 p 表示) 的结果进行模型精度评价。其中  $R^2$  越接近于 1, 模型解释能力越强; RMSE 越小, 模型预测精度越高。RPD 是数据的标准差 (SD) 与模型预测的均方根误差 (RMSE) 的比值, 它衡量的是模型预测的精度相对于数据自身波动性的表现。当  $RPD < 1.4$  时, 表明模型的预测能力不可靠; 当  $1.4 \leq RPD < 2.0$  时, 表明模型的预测能力相对可靠, 但可能存在改进空间; 当  $RPD \geq 2.0$  时, 表明模型的预测精度很高, 足以可靠地区分不同样本之间的差异<sup>[8]</sup>。

### 1.7 数据统计分析

本次研究所涉及软件包括 R (4.4.2)、ArcGIS10.8、SPSS、Origin2024。其中 PLSR、RF、GPR、Cubist、SVM 等 5 种模型构建及参数优化均在 R 中调试完成；数据统计分析与相关分析在 SPSS 软件中完成；绘图在 ArcGIS 和 Origin 软件中完成。

## 2 结果

### 2.1 土壤 Cu 的描述性统计

研究区土壤剖面中 Cu 及其组分的描述性统计特征如表 1 所示。土壤剖面全集样本总铜变化范围为 8.94~73.46 mg·kg<sup>-1</sup>，F1 的变化范围为 0.03~9.82 mg·kg<sup>-1</sup>，F2 的变化范围为 0.01~24.48 mg·kg<sup>-1</sup>，F3 的变化范围为 0.10~15.16 mg·kg<sup>-1</sup>，F4 的变化范围为 7.91~48.07 mg·kg<sup>-1</sup>。其中，F1 含量最低，但变异系数较大；F2 变异系数最大，分布不均匀，呈现强烈的空间异质性。F4 含量最高且稳定，是土壤中铜的主要赋存形态。由于采集的土壤剖面样品从表层至底层（约 100 cm 深）覆盖范围大，加之土壤本身具有较强的空间变异性，这会导致样本具有显著的异质性。从表 1 中可以看出，整体数据集和训练集及验证集具有相似的描述性特征。这表明 Kennard-Stone 算法的分割方案是合理可行的。

5 种土壤铜属性（总铜、F1、F2、F3 和 F4）之间均呈现显著正相关关系（表 2）。其中，F1 ( $r=0.61$ )、F2 ( $r=0.86$ )、F3 ( $r=0.62$ ) 和 F4 ( $r=0.92$ ) 与总铜均表现出较强的相关性。这些铜属性之间的密切关联，有助于进一步解释高光谱成像（HSI）在土壤铜预测中的应用机理。

表 1 土壤剖面 Cu 含量的统计特征

Table 1 Statistical characteristics of Cu concentration in soil profile

数据集 Data set	变量 Variable	范围 Range/ (mg·kg <sup>-1</sup> )	均值 Mean/ (mg·kg <sup>-1</sup> )	中值 Median/ (mg·kg <sup>-1</sup> )	偏度 Skewness	标准差 Standard Deviation/ (mg·kg <sup>-1</sup> )	变异系数 Coefficient of variation/%
全集 Whole set (n=433)	F1 <sup>①</sup>	0.03~9.82	1.02	0.63	2.83	1.26	123.53
	F2 <sup>②</sup>	0.01~24.48	1.33	0.79	5.74	2.08	156.39
	F3 <sup>③</sup>	0.10~15.16	2.42	1.93	2.15	3.33	137.60
	F4 <sup>④</sup>	7.91~48.07	22.28	21.04	0.72	6.71	30.12
	总铜	8.94~73.46	27.06	25.84	1.19	8.73	32.26
建模集 Calibration set (n=303)	F1 <sup>①</sup>	0.03~9.82	0.99	0.60	3.09	1.28	129.29
	F2 <sup>②</sup>	0.12~24.48	1.45	0.71	5.06	2.41	166.21
	F3 <sup>③</sup>	0.10~15.16	2.38	1.71	2.14	2.34	98.32
	F4 <sup>④</sup>	7.91~48.07	22.39	20.96	0.69	7.26	32.43
	总铜	8.94~73.46	27.21	25.48	1.17	9.60	35.28
验证集 Validation set (n=130)	F1 <sup>①</sup>	0.03~6.14	1.09	0.70	2.20	1.22	111.93
	F2 <sup>②</sup>	0.01~7.19	1.05	0.87	3.61	0.85	80.95
	F3 <sup>③</sup>	0.12~14.92	2.52	2.19	2.20	2.32	92.06
	F4 <sup>④</sup>	8.88~39.87	22.04	21.71	0.66	5.24	23.77
	总铜	13.24~48.59	26.70	26.09	0.70	6.25	23.41

①弱酸溶解态铜 Weak acid soluble copper (F1)，②可还原性铜 Reducible copper (F2)，③有机结合态铜 Organic-bound copper (F3)，④残渣态铜 Residual copper (F4)。下同。 The same below.

表 2 土壤总铜与各组分之间的相关系数

Table 2 Correlation coefficients between total Cu and its chemical fractions in soils

铜组分 Cu Fraction	F1	F2	F3	F4	总铜
F1	1.00				
F2	0.45**	1.00			
F3	0.88**	0.48**	1.00		
F4	0.31**	0.72**	0.3**	1.00	
总铜 Total Cu	0.61**	0.86**	0.62**	0.92**	1.00

\*\*表示  $P < 0.01$ 。 \*\* represents  $P < 0.01$ .

## 2.2 不同光谱预处理对土壤剖面 Cu 组分预测模型的影响

为了比较不同光谱预处理方法对土壤剖面 Cu 的预测模型精度的影响,本研究采用 18 种光谱预处理方法,并基于 PLSR、RF、Cubist、GPR 和 SVM 等 5 种模型,利用 VNIR 对铜含量进行预测分析。各模型在最佳预处理方法下的评估结果如表 3 所示。经过预处理后,各模型的精度得到了提高。PLSR 模型中 SNV 在预测 F1 (验证集:  $R^2_p=0.68$ ,  $RMSE_p=0.70 \text{ mg}\cdot\text{kg}^{-1}$ ) 与 F3 (验证集:  $R^2_p=0.64$ ,  $RMSE_p=1.40 \text{ mg}\cdot\text{kg}^{-1}$ ) 时具有较高的预测精度。而在 RF 模型中,GS 对 F1 (验证集:  $R^2_p=0.73$ ,  $RMSE_p=0.64 \text{ mg}\cdot\text{kg}^{-1}$ ) 与 F3 (验证集:  $R^2_p=0.73$ ,  $RMSE_p=1.21 \text{ mg}\cdot\text{kg}^{-1}$ ) 的预测效果较好。GPR 模型采用 Abs+GS 预处理后,对 F1、F3 和 F4 均表现出较强的预测能力。Cubist 模型结合 GS+MSC 在 F2、F4 及总铜的预测表现突出。而 SVM 模型在 Abs+FD 预处理下,整体上对总铜以及 4 种组分均能取得稳定的预测结果。这表明预处理方法的有效性高度依赖于模型类型与预测目标。

在独立验证中,经 GS+MSC 预处理后的 Cubist 模型与 Abs+FD 预处理后的 SVM 模型,在 5 种 Cu 属性的预测中,其 RPD 的值均大于 1.4, RMSE 较小且  $R^2$  接近于 1,显然这两种预处理后建立的模型具有相对较好的预测能力。

表 3 不同预处理方法下土壤剖面 Cu 含量的预测模型

Table 3 Predictive models of Cu concentration in soil profiles under different pretreatment methods

铜组分 Cu Fraction	模型 Model	最佳预处理方法 Best method	建模集 Calibration set			验证集 Validation set		
			$R^2_{cv}$	$RMSE_{cv}$	$RPD_{cv}$	$R^2_p$	$RMSE_p$	$RPD_p$
弱酸溶解态铜 Weak acid soluble copper (F1)	PLSR	SNV	0.54	0.87	1.47	0.68	0.70	1.75
	Cubist	GS+SNV	0.62	0.79	1.62	0.80	0.54	2.23
	GPR	Abs+GS	0.60	0.80	1.58	0.71	0.68	1.80
	RF	GS	0.55	0.86	1.48	0.73	0.64	1.90
	SVM	Abs+GS	0.70	0.70	1.82	0.80	0.54	2.24
可还原性铜 Reducible copper (F2)	PLSR	SNV	0.66	1.41	1.71	0.51	0.65	1.31
	Cubist	GS+MSC	0.84	1.01	2.39	0.85	0.36	2.38
	GPR	FD	0.75	1.20	2.01	0.78	0.40	2.12
	RF	Abs+FD	0.79	1.10	2.19	0.77	0.42	2.02
有机结合态铜 Organic-bound copper (F3)	SVM	FD	0.84	0.99	2.43	0.81	0.38	2.27
	PLSR	SNV	0.59	1.51	1.55	0.64	1.40	1.66
	Cubist	Abs+SD	0.55	1.58	1.48	0.74	1.18	1.96
	GPR	Abs+GS	0.64	1.40	1.67	0.70	1.27	1.83
	RF	GS	0.64	1.41	1.66	0.73	1.21	1.92

	SVR	Abs+FD	0.73	1.22	1.91	0.67	1.33	1.74
残渣态铜	PLSR	GS+SNV	0.55	4.92	1.47	0.25	4.70	1.12
Residual	Cubist	GS+MSC	0.72	3.81	1.91	0.68	3.01	1.74
copper (F4)	GPR	Abs+GS	0.66	4.26	1.71	0.61	3.33	1.57
	RF	GS+MSC	0.67	4.21	1.72	0.62	3.45	1.52
	SVM	Abs+FD	0.66	4.24	1.71	0.70	2.91	1.80
总铜	PLSR	GS+SNV	0.57	6.41	1.50	0.28	5.83	1.07
Total copper	Cubist	GS+MSC	0.69	5.32	1.80	0.65	3.78	1.65
	GPR	Abs+FD	0.72	5.15	1.86	0.64	3.77	1.66
	RF	GS+MSC	0.69	5.43	1.77	0.63	4.02	1.55
	SVM	Abs+FD	0.76	4.74	2.03	0.67	3.64	1.71

注: Raw: 原始光谱反射率; Abs: 吸光度转换; GS: Gap-Segment 导数; SG: Savitzky-Golay 平滑; FD: 一阶导数; SD: 二阶导数; MSC: 多元散射校正; SNV: 标准正态变量; DT: 去趋势化; CR: 连续统去除。组合方法(如 GS+SNV)表示按顺序应用相应方法。Note:

Raw-Original spectral reflectance; Abs-Absorbance transformation; GS-Gap-Segment derivative; SG-Savitzky-Golay smoothing; FD-First derivative; SD-Second derivative; MSC-Multiplicative scatter correction; SNV-Standard normal variate; DT-Detrending; CR-Continuum removal. Combinations (e.g., GS+SNV) indicate that the corresponding methods were applied sequentially.

### 2.3 不同机器学习模型预测土壤剖面 Cu 组分的精度比较

为了确定最佳的土壤 Cu 预测模型,基于 3 种精度评价参数比较了 PLSR 模型与 4 种机器学习模型。在验证集中,PLSR 对于 5 种土壤 Cu 形态属性预测的 RPD 值明显低于其他模型。与传统的线性模型 PLSR 相比,4 种非线性模型(Cubist、GPR、RF、SVM)预测土壤剖面总铜及其组分的精度有明显提升。为更好地比较 5 种模型的预测性能,Cubist、GPR、RF、PLSR 均采用预测效果最好的预处理方法,SVM 采用 Abs+FD 预处理方法。从 5 种模型的独立验证集预测结果中可以发现(图 2-图 6),4 种机器学习模型的拟合精度均明显高于 PLSR 模型。而在 4 种机器学习模型中,Cubist 与 SVM 模型在 5 种模型中均能够取得较好的预测效果,其预测值和实测值较均匀且集中分布在 1:1 线附近。在独立验证中,SVM 在 5 种土壤 Cu 属性的预测中较为出色(F1:  $R^2_p=0.78$ ,  $RMSE_p=0.56 \text{ mg}\cdot\text{kg}^{-1}$ ; F2:  $R^2_p=0.78$ ,  $RMSE_p=0.40 \text{ mg}\cdot\text{kg}^{-1}$ ; F3:  $R^2_p=0.67$ ,  $RMSE_p=1.33 \text{ mg}\cdot\text{kg}^{-1}$ ; F4:  $R^2_p=0.70$ ,  $RMSE_p=2.91 \text{ mg}\cdot\text{kg}^{-1}$ ; 总铜:  $R^2_p=0.67$ ,  $RMSE_p=3.64 \text{ mg}\cdot\text{kg}^{-1}$ )。通过对验证集中铜属性预测结果的分析发现,对于总铜的预测,各模型的拟合精度排序为: SVM>Cubist>GPR>RF>PLSR。然而,在针对不同土壤组分进行预测时,SVM 在 F1 和 F4 组分上为最优模型,而在 F2 和 F3 组分上则未能表现出最佳性能。

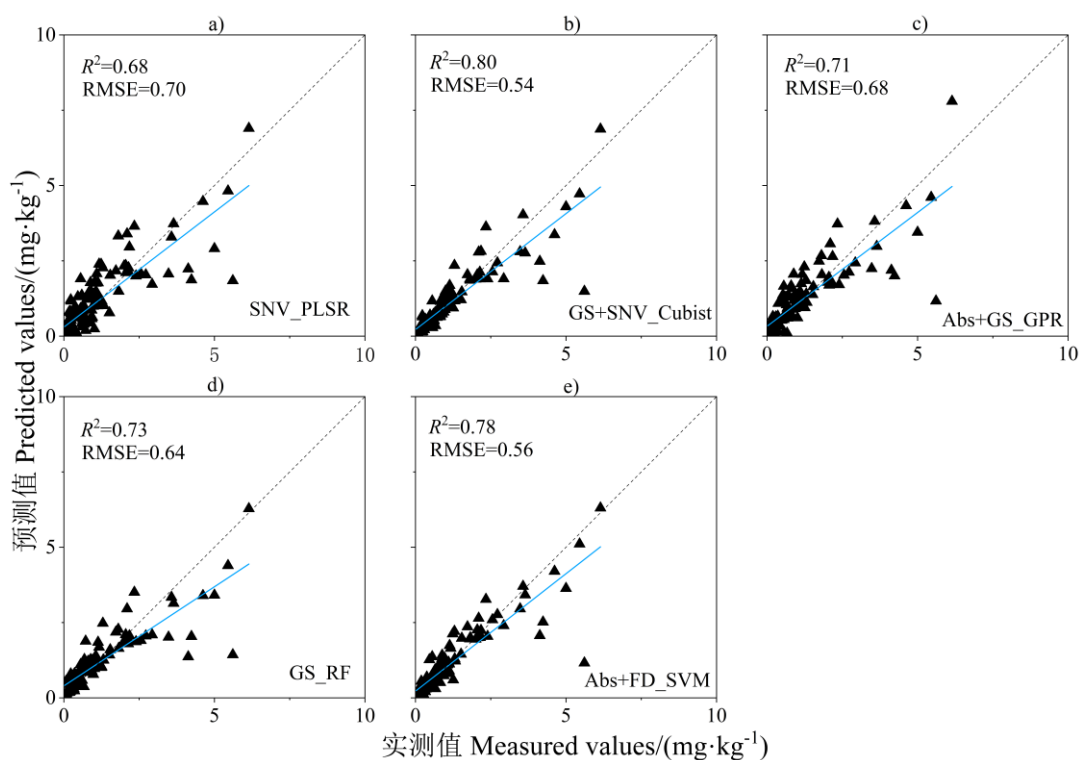


图 2 使用验证集中的不同机器学习技术预测值与实测值弱酸溶解态铜 (F1) 的散点图

Fig. 2 Scatter plots of predicted versus measured values of weak acid soluble copper (F1) using various machine learning techniques on the validation set

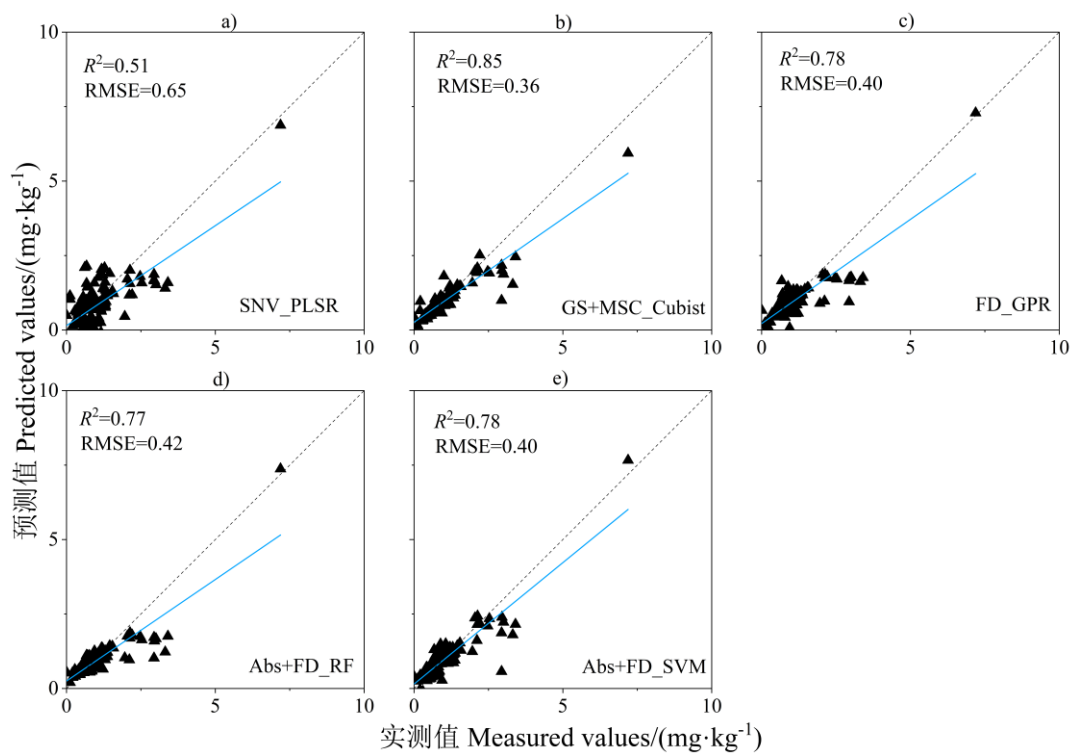


图 3 使用验证集中的不同机器学习技术预测值与实测值的可还原性铜 (F2) 的散点图

Fig. 3 Scatter plots of predicted versus measured values of reducible copper (F2) using various machine learning techniques on the validation set

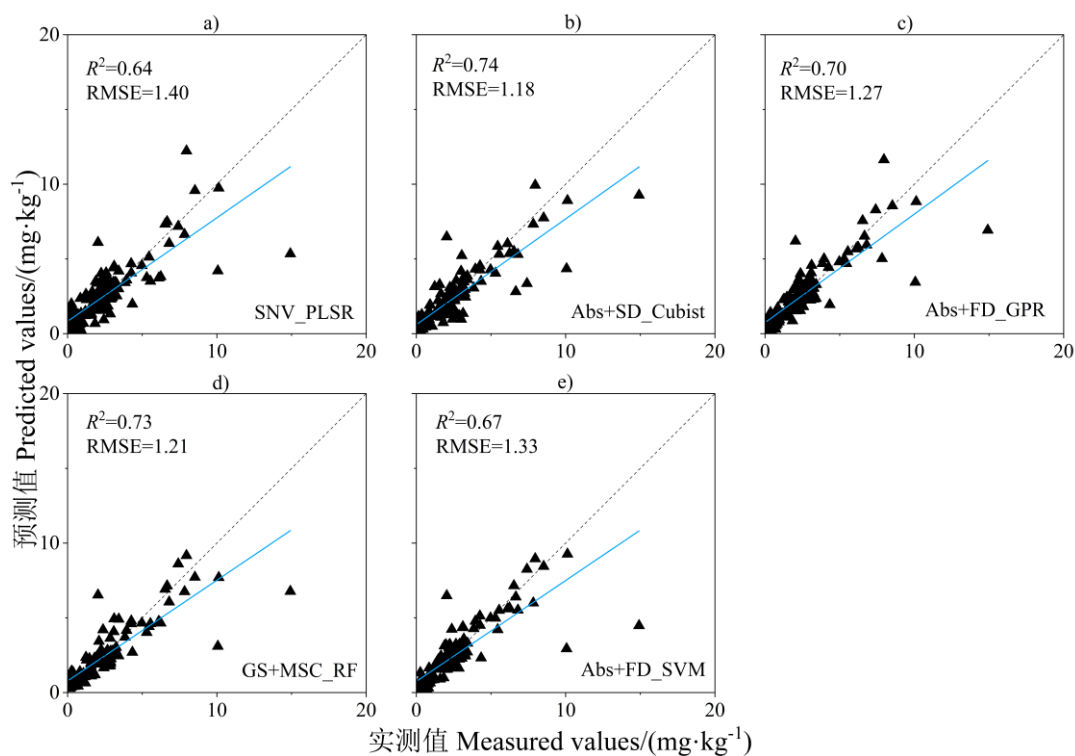


图 4 使用验证集中的不同机器学习技术预测值与实测值的有机结合态铜 (F3) 的散点图

Fig. 4 Scatter plots of predicted versus measured values of organically bound copper (F3) using various machine learning techniques on the validation set

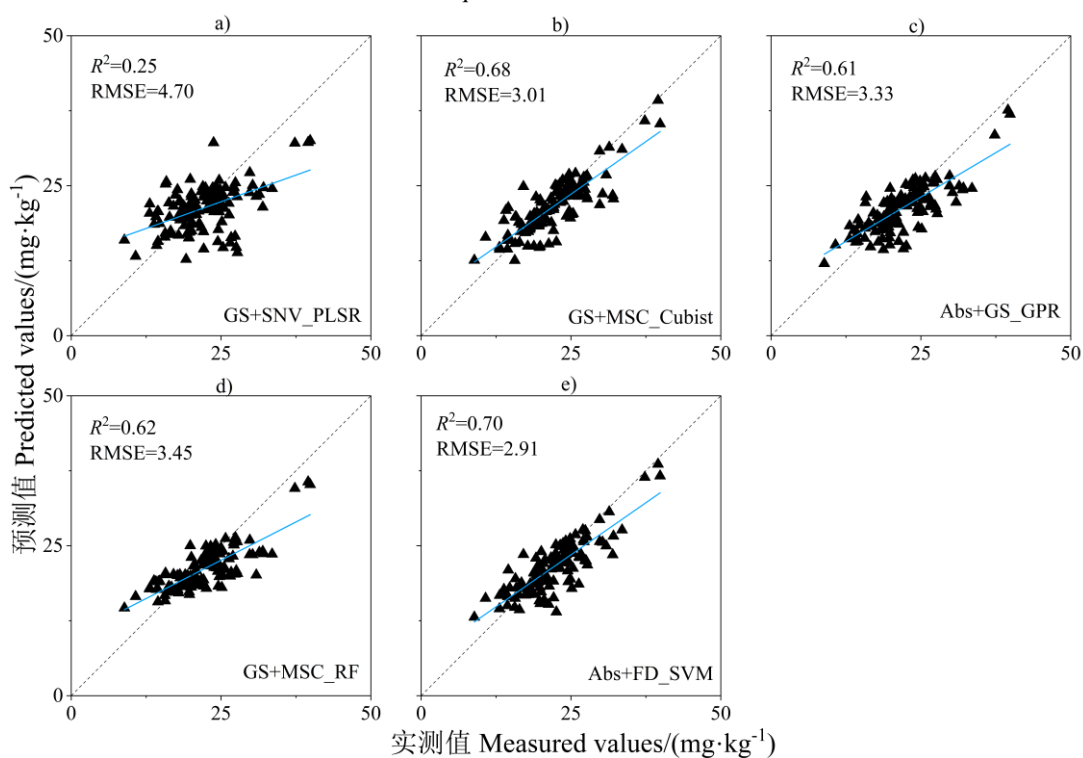


图 5 使用验证集中的不同机器学习技术预测值与实测值的残渣态铜 (F4) 的散点图

Fig. 5 Scatter plots of predicted versus measured values of residual copper (F4) using various machine learning techniques on the validation set

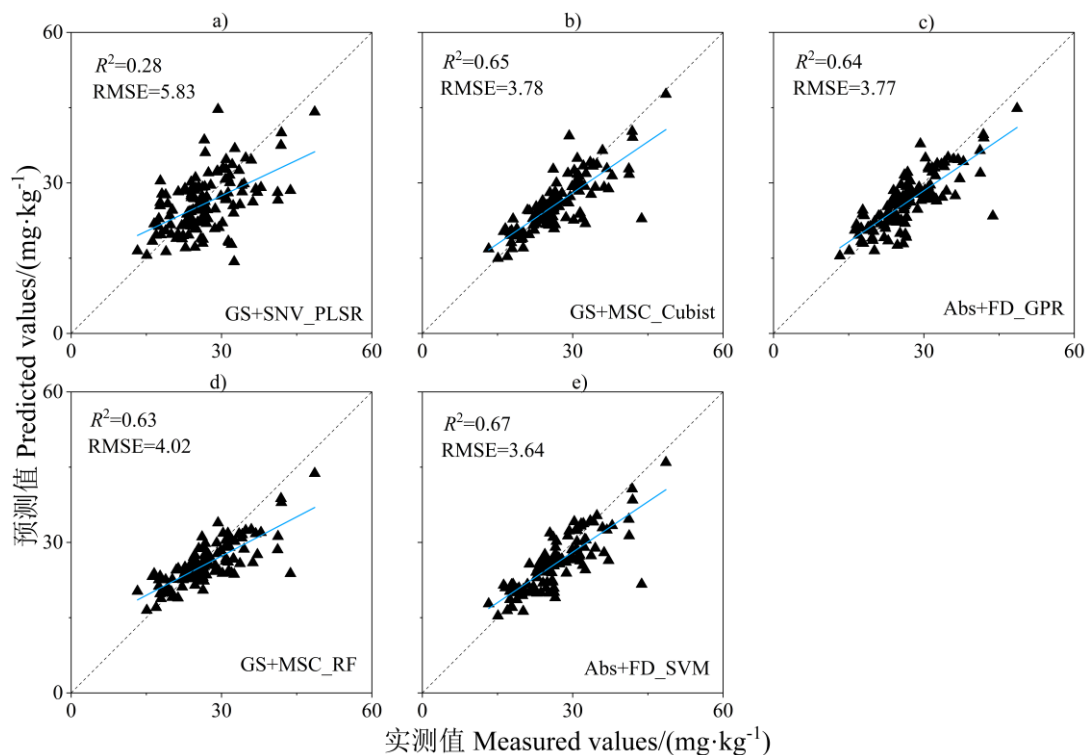


图 6 使用验证集中的不同机器学习技术预测值与实测值的土壤总铜的散点图

Fig. 6 Scatter plots of predicted versus measured values of total copper in soil using various machine learning techniques on the validation set

## 2.4 影响光谱变量的相对重要性

为了探讨机器学习模型预测原状土壤剖面总铜及其各组分的内在机理,计算 SVM 预测模型中每个波长的相对重要性(图 7)。从光谱相对重要性上来看, SVM 模型识别的所有土壤 Cu 属性的光谱指示性波段主要出现在 400~470 nm 和 940~1010 nm 光谱区间。其中重要性较为突出的波长为 408、409、410、411、412、413、450、461、960 nm。此外, 530~570 nm 及 800~900 nm 的部分光谱区域对预测 F1 与 F3 也比较重要。

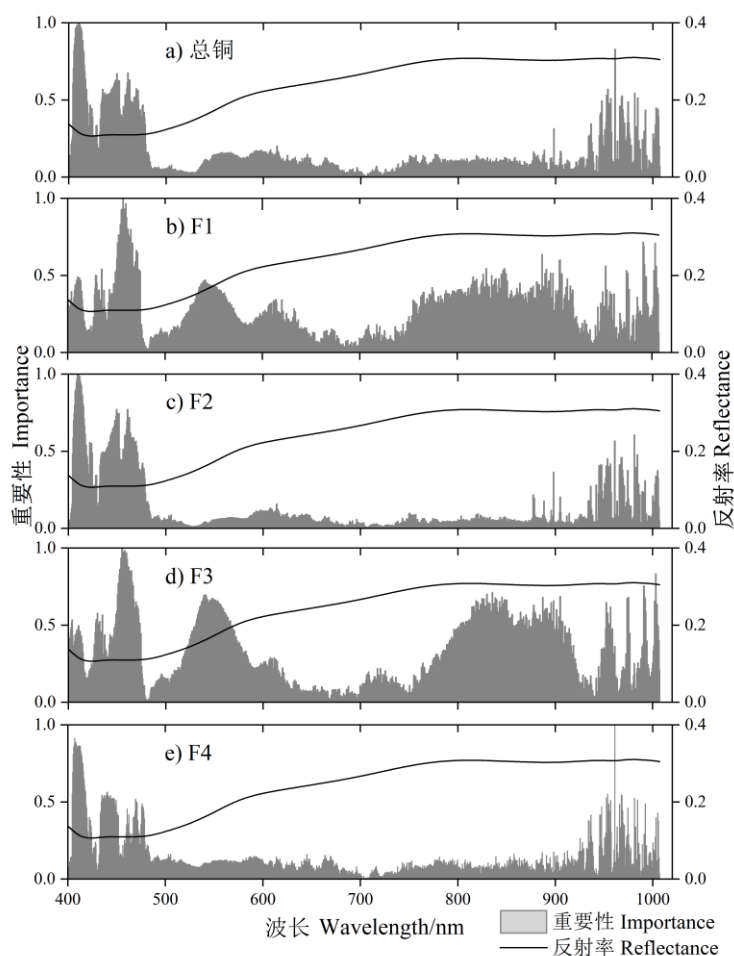


图 7 SVM 模型预测土壤 Cu 属性的光谱变量相对重要性

Fig. 7 Relative importance of spectral variables for predicting soil Cu attributes by the SVM model

### 3 讨论

本研究基于土壤光谱反射特征与重金属含量的相关性,以乐安河沿岸农田为对象,通过采集原状土壤样本,测定其反射光谱,并进行实验室理化分析,采用 PLSR 与机器学习方法,对土壤剖面总铜及其化学形态含量进行预测反演。

采用合适的预处理方法,可以更有效地突出土壤的特征光谱波段,增强光谱曲线中的反射峰和吸收谷等关键信息,提升高光谱数据模拟精度。本研究建立了多种土壤 Cu 高光谱遥感定量反演模型。研究发现基于 Abs+FD 预处理的 SVM 模型与基于 GS+SNV 预处理的 Cubist 可以较为准确地预测土壤剖面总铜及各化学组分的浓度(表 4)。但仅表明在本研究中这两种预处理效果相对较好,其并不具有普适性。Hou 等<sup>[27]</sup>采用 SG 和 MSC 等多种光谱预处理建立 PLSR 模型估算多种土壤重金属含量,结果表明 SG 平滑结合 MSC 和倒数对数变换对反射光谱进行预处理后可获得最高的模型精度 ( $R^2=0.88$ ,  $RMSE=1.29$ )。张杨等<sup>[28]</sup>在喀斯特矿区对比了 9 种高光谱数据预处理方法,建立了 PLSR 与 RF 的高光谱反演模型,结果显示多元散射校正结合一阶微分 (MSC+FD) 的预处理方法是土壤 As 和 Sb 的最优光谱预处理方法 ( $R^2$  均大于 0.9)。这表明并没有某一种单一的光谱预处理方法或其组合总能获得最佳预测能力,同一种建模方法中不同预处理方法建立的模型预测精度也存在着很大

差距。这主要由于土壤原始光谱中混杂的噪声类型多样、建模算法对数据结构的响应机制不同,以及预处理在抑制噪声的同时可能削弱与重金属赋存状态相关的光谱信息。因此,不存在普遍适用的最优预处理方案,其效果需结合具体研究对象与建模方法进行系统评估。

为确定最佳的土壤 Cu 的预测模型,比较 Abs+FD 预处理下的 SVM 模型与最佳预处理方法下的 PLSR、Cubist、GPR、RF 模型。与 PLSR 模型相比,4 种机器学习算法 Cubist、RF、GPR 和 SVM 对土壤剖面 Cu 含量的预测能力均有显著提升,其中 SVM 与 Cubist 模型的表现优于 RF、GPR 模型。此外,在验证集独立验证阶段,发现 SVM 模型对 5 种土壤铜属性均具有更好的预测能力(独立验证的  $R^2$  范围为 0.67~0.78)。Cubist 模型在 5 种土壤铜属性的独立验证中也表现出出色的预测能力。PLSR 模型对 5 种 Cu 属性的预测精度最低。作为一种线性回归方法,PLSR 在建模过程中难以充分捕捉此类复杂的非线性相互作用机制,因而在预测不同化学形态的 Cu 含量时表现不佳。前人研究结果表明,SVM 模型在预测土壤属性方面的表现较为出色<sup>[9,29-30]</sup>,这与本文的研究结果一致。但 SVM 的预测性能并不总是优于其他机器学习算法。Ma 等<sup>[31]</sup>在复垦矿山土壤中重金属反演研究过程中,比较了 PLSR、SVM、RF 和 ELM 对重金属 Cd、As、Pb 的预测性能,结果发现 RF 算法在土壤重金属反演中具有最高的精度。Niu 等<sup>[9]</sup>通过采用光谱指数、CARS 算法结合 PLSR、SVM 和 RF 等机器学习方法,建立了评估总氮含量的高光谱估计模型,结果表明结合 CARS 算法的 SVM 模型能够快速估算砂姜黑土中全氮含量( $R^2=0.96$ , RMSE=0.06)。周伟等<sup>[32]</sup>利用 PLSR、SVM、RF 模型对三江源区 SOM 含量进行建模估算,发现采用 RF 模型可有效反演土壤有机质含量( $R^2=0.78$ )。Xu 等<sup>[30]</sup>比较了 PLSR、RF 和 SVM 等 3 种模型在土壤剖面上对不同土壤 Fe 属性的预测效果,结果表明 SVM 对 6 种土壤 Fe 属性均表现出优于其他测试模型的性能。综上所述,非线性的机器学习模型在大多数情况下对土壤属性的预测精度显著优于线性模型。然而,对于预测不同土壤属性而言,并不存在一种普遍适用的最优机器学习模型,其预测性能往往因具体研究对象和数据特征而异。

研究表明,土壤中重金属的含量与黏土矿物、铁锰氧化物和有机质等组分的吸附作用存在显著关联<sup>[14]</sup>。这些成分不仅影响土壤光谱的形态和反射率,而且表现出特殊的光谱特征,从而为利用高光谱数据反演土壤重金属污染状况提供了坚实的理论依据。根据计算 SVM,预测模型中每个波长的相对重要性(图 3),SVM 模型识别的所有土壤 Cu 参数的光谱指示性波段主要出现在 400~470 和 940~1010 nm 光谱区间,这与 Meng 等<sup>[33]</sup>、王滢艺等<sup>[6]</sup>研究结果一致。其中,400~470 nm 处于铁锰氧化物的强吸收带<sup>[34]</sup>。而本研究中 940~1010 nm 也是土壤 Cu 属性的光谱指示性波段,这可能与磁铁矿中的 OH<sup>-</sup>离子形成的络合物及近红外区域的分子基团与铜离子形成额外的配位化合物有关<sup>[33]</sup>。此外,Cu<sup>2+</sup>在 530 nm 附近有一个反射峰值,也是反演土壤 Cu 的敏感波段<sup>[35]</sup>,这可能对应于 F1 与 F3 在 530~570 nm 区间较高的光谱相对重要性。显然,实验中计算的光谱指示性波段符合相关规律,大部分位于 Cu 的光谱特征波段,是反演土壤 Cu 的敏感波段。

本研究仍存在一定的局限性。首先,在超参数优化方面,本研究采用了网格搜索方法,该方法虽直观但计算成本较高,并不适合高维参数空间中的高效搜索。未来研究可引入贝叶斯优化等自动化超参数优化方法以提高模型调优效率与模型性能。与网格搜索相比,贝叶斯优化能够在人为给定的超参数搜索范围内找到全局最优的超参数组合,而且优化速度更快<sup>[36]</sup>。其次,本研究在建模时直接使用了全部光谱波段,未进行系统的特征变量筛选,这可能会引入光谱冗余信息或噪声。后续工作可进一步结合特征选择方法(如 CARS 和 Boruta 算法<sup>[12]</sup>等),以增强模型的解释性与泛化能力。其中,CARS 算法是一种广泛应用于光谱分析的特征选择方法<sup>[37]</sup>。该算法基于自适应重加权采样技术,以回归系数等指标衡量各波长变量的重要性,通过迭代竞争机制逐步剔除冗余信息,最终筛选出最优的特征子集。此外,后续研究将结合 HSI 和机器学习技术,绘制原状土壤剖面中铜组分含量的垂直分布图,以揭示其迁移与富集规律,深化对土壤重金属污染程度与机理的理解。

## 4 结论

(1) 乐安河沿岸农田原状土壤剖面中, 总铜及其不同化学组分含量在垂向上呈现分异特征, 各形态铜的变化规律并不一致, 且其含量范围和变异程度存在差别。总铜与各化学组分之间均呈显著正相关, 不同铜组分之间具有一定内在联系, 其含量变化能够在相应光谱特征中得到响应。(2) 光谱预处理能够增强光谱变量与土壤铜含量之间的相关性, 并提高铜组分预测模型的性能。不同光谱预处理方法对模型性能的影响存在差异, 且这种差异与建模对象及所采用的建模方法密切相关。(3) 土壤剖面总铜及其不同化学组分含量与高光谱信息之间的关系较为复杂, 非线性建模方法能够较好地表征其变化特征。不同机器学习模型对土壤铜组分的响应规律表征存在差异, 其中 SVM 模型在总铜及各形态铜含量预测中均表现出较好的预测效果。

### 参考文献 (References)

- [1] Chu D. Effects of heavy metals on soil microbial community[J]. IOP Conference Series: Earth and Environmental Science, 2018, 113(1): 012009.
- [2] Wang J X, Fu H Y, Xu D M, et al. The remediation mechanisms and effects of chemical amendments for heavy metals in contaminated soils: A review of literature[J]. Polish Journal of Environmental Studies, 2022, 31(5): 4511-4522.
- [3] Cao Q Y, Huang Z H. Review on speciation analysis of heavy metals in polluted soils and its influencing factors[J]. Ecological Science, 2017, 36(6): 222-232. [曹勤英, 黄志宏. 污染土壤重金属形态分析及其影响因素研究进展[J]. 生态科学, 2017, 36(6): 222-232.]
- [4] Gao Z R, Chang Y H, Zhao Y Y. Chemical speciation characteristics and significance of heavy metals in the soil of the river basin adjacent to the Dexing copper mining area, Jiangxi Province[J]. Acta Petrologica et Mineralogica, 2025, 44(1): 167-180. [高知睿, 常玉虎, 赵元艺. 江西德兴铜矿附近流域土壤中重金属化学形态特征及意义[J]. 岩石矿物学杂志, 2025, 44(1): 167-180.]
- [5] He J, Li C X, Tan X Y, et al. Driving factors for distribution and transformation of heavy metals speciation in a zinc smelting site[J]. Journal of Hazardous Materials, 2024, 471: 134413.
- [6] Wang Y Y, Xu S X, Zhao Y C, et al. Prediction of Cu concentrations in intact soil profiles based on VNIR and machine learning algorithms[J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(7): 336-344. [王滢艺, 徐胜祥, 赵永存, 等. 基于 VNIR 和机器学习算法的原状土剖面 Cu 含量预测[J]. 农业工程学报, 2022, 38(7): 336-344.]
- [7] Xu S X, Wang M Y, Shi X Z. Hyperspectral imaging for high-resolution mapping of soil carbon fractions in intact paddy soil profiles with multivariate techniques and variable selection[J]. Geoderma, 2020, 370: 114358.
- [8] Zhao M S, Xie Y, Lu L M, et al. Modeling for soil organic matter content based on hyperspectral feature indices[J]. Acta Pedologica Sinica, 2021, 58(1): 42-54. [赵明松, 谢毅, 陆龙妹, 等. 基于高光谱特征指数的土壤有机质含量建模[J]. 土壤学报, 2021, 58(1): 42-54.]
- [9] Niu Z, Shi L, Qiao H B, et al. Construction of a hyperspectral estimation model for total nitrogen content in Shajiang black soil[J]. Journal of Plant Nutrition and Soil Science, 2023, 186(2): 196-208.
- [10] Zhou Y, Cheng Y S, Wang D P, et al. Hyperspectral inversion of soil arsenic content in polymetallic mining areas based on optimized spectral index combined with PLSR[J]. The Chinese Journal of Nonferrous Metals, 2024, 34(2): 653-667. [周瑶, 成永生, 王丹平, 等. 优化光谱指数结合 PLSR 的多金属矿区土壤 As 含量高光谱反演[J]. 中国有色金属学报, 2024, 34(2): 653-667.]
- [11] Hong Y S, Chen Y Y, Shen R L, et al. Diagnosis of cadmium contamination in urban and suburban soils using visible-to-near-infrared spectroscopy[J]. Environmental Pollution, 2021, 291: 118128.

- [12] Qi Y T, Gan S, Yuan X P, et al. 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. [杞应涛, 甘淑, 袁希平, 等. 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[J]. *土壤学报*, 2026,63(1):241-250.]
- [13] Zheng M D, Luan H J, Liu G S, et al. Ground-based hyperspectral retrieval of soil arsenic concentration in Pingtan Island, China[J]. *Remote Sensing*, 2023, 15(17): 4349.
- [14] Zhou W, Yang H, Xie L J, et al. Hyperspectral inversion of soil heavy metals in Three-River Source Region based on random forest model[J]. *Catena*, 2021, 202: 105222.
- [15] Yang H F, Xu H, Zhong X N. Prediction of soil heavy metal concentrations in copper tailings area using hyperspectral reflectance[J]. *Environmental Earth Sciences*, 2022, 81(6): 183.
- [16] Wang Y Y, Zhao Y C, Xu S X. Application of VNIR and machine learning technologies to predict heavy metals in soil and pollution indices in mining areas[J]. *Journal of Soils and Sediments*, 2022, 22(10): 2777-2791.
- [17] Zhao Z P, Sun Y M, Jia W W, et al. Prediction of vanadium contamination distribution pattern through remote sensing image fusion and machine learning[J]. *Remote Sensing*, 2025, 17(7): 1164.
- [18] Oppong Sarkodie V Y, Vašát R, Pouladi N, et al. Predicting soil organic carbon stocks in different layers of forest soils in the Czech Republic[J]. *Geoderma Regional*, 2023, 34: e00658.
- [19] Zhong Q, Eziz M, Sawut R, et al. Application of a hyperspectral remote sensing model for the inversion of nickel content in urban soil[J]. *Sustainability*, 2023, 15(18): 13948.
- [20] Tao C, Wang Y J, Zou B, et al. Assessment and analysis of migrations of heavy metal lead and zinc in soil with hyperspectral inversion model[J]. *Spectroscopy and Spectral Analysis*, 2018, 38(6): 1850-1855. [陶超, 王亚晋, 邹滨, 等. 土壤重金属铅、锌高光谱反演模型可迁移能力分析研究[J]. *光谱学与光谱分析*, 2018, 38(6): 1850-1855.]
- [21] Shang K, Xiao C C, Gan F P, et al. Estimation of soil copper content in mining area using ZY1-02D satellite hyperspectral data[J]. *Journal of Applied Remote Sensing*, 2021, 15(4): 042607.
- [22] Zhong X L, Zhou S L, Huang M L, et al. Chemical form distribution characteristic of soil heavy metals and its influencing factors[J]. *Ecology and Environment*, 2009, 18(4): 1266-1273. [钟晓兰, 周生路, 黄明丽, 等. 土壤重金属的形态分布特征及其影响因素[J]. *生态环境学报*, 2009, 18(4): 1266-1273.]
- [23] Huang Q, Tong F, Wang B, et al. Mechanisms of Mn-modified biochar to mediate inorganic arsenic speciation in the soil-solution system[J]. *Acta Pedologica Sinica*, 2025, 62(3): 752-765. [黄琴, 童非, 王博, 等. 锰改性生物质炭调控土壤-溶液体系中砷赋存形态的机制研究[J]. *土壤学报*, 2025, 62(3): 752-765.]
- [24] Liu J L, Wang F, Han J Q, et al. Study on spectral characteristics and quantitative estimation of soil salinity based on fractional order derivative[J]. *Remote Sensing Technology and Application*, 2025, 40(2): 344-358. [刘佳琳, 王飞, 韩剑桥, 等. 基于分数阶微分的土壤含盐量光谱特征及定量估算研究[J]. *遥感技术与应用*, 2025, 40(2): 344-358.]
- [25] Win A, Minasny B, Ringrose-Voase A, et al. Improving the accuracy of digital soil mapping using remote sensing and topography covariates in the Central Dry Zone of Myanmar[J]. *Geoderma Regional*, 2025, 42: e01001.
- [26] Bi L, Xiong W L. Small sample expansion method and application based on data distribution[J]. *Control Engineering of China*, 2019, 26(7): 1431-1436. [毕略, 熊伟丽. 基于数据分布的小样本扩充方法及应用[J]. *控制工程*, 2019, 26(7): 1431-1436.]
- [27] Hou L, Li X J, Li F. Hyperspectral-based inversion of heavy metal content in the soil of coal mining areas[J]. *Journal of Environmental Quality*, 2019, 48(1): 57-63.
- [28] Zhang Y, Wu L N, Ouyang K C. Hyperspectral inversion study of heavy metals As and Sb in soil from karst mining area[J]. *Laser & Infrared*, 2025, 55(7): 1081-1092. [张杨, 吴琳娜, 欧阳坤长. 喀斯特矿区土壤重金属 As、Sb 高光谱反演研究[J]. *激光与红外*, 2025, 55(7): 1081-1092.]
- [29] 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.]
- [30] Xu S X, Zhao Y C, Wang M Y, et al. A comparison of machine learning algorithms for mapping soil iron parameters indicative of pedogenic processes by hyperspectral imaging of intact soil profiles[J]. *European Journal of Soil Science*, 2022, 73(1): e13204.
- [31] Ma W B, Tan K, Du P J. Predicting soil heavy metal based on random forest model[C]//2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS): IEEE, 2016: 4331-4334.
- [32] Zhou W, Xie L J, Yang H, et al. Hyperspectral inversion of soil organic matter content in the three-rivers source region[J]. *Chinese Journal of Soil Science*, 2021, 52(3): 564-574. [周伟, 谢利娟, 杨晗, 等. 基于高光谱的三江源区土壤有机质含量反演[J]. 土壤通报, 2021, 52(3): 564-574.]
- [33] Meng C, Hong M, Hu Y C, et al. Using optimized spectral indices and machine learning algorithms to assess soil copper concentration in mining areas[J]. *Sustainability*, 2024, 16(10): 4153.
- [34] Xie X L, Sun B, Hao H T. Relationship between visible-near infrared reflectance spectroscopy and heavy metal of soil concentration[J]. *Acta Pedologica Sinica*, 2007, 44(6): 982-993. [解宪丽, 孙波, 郝红涛. 土壤可见光-近红外反射光谱与重金属含量之间的相关性[J]. 土壤学报, 2007, 44(6): 982-993.]
- [35] Huang C P, Liu B, Zhang X, et al. Study on band selection and optimal spectral resolution for prediction of Cu contamination in soils[J]. *Remote Sensing Technology and Application*, 2010, 25(3): 353-357, 341. [黄长平, 刘波, 张霞, 等. 土壤重金属 Cu 含量遥感反演的波段选择与最佳光谱分辨率研究[J]. 遥感技术与应用, 2010, 25(3): 353-357, 341.]
- [36] Li H X, Song D L, Kong J N, et al. Evaluation of hyperparameter optimization techniques for traditional machine learning models[J]. *Computer Science*, 2024, 51(8): 242-255. [李海霞, 宋丹蕾, 孔佳宁, 等. 传统机器学习模型的超参数优化技术评估[J]. 计算机科学, 2024, 51(8): 242-255.]
- [37] Zhang X, Sun Y X, Shang K, et al. Soil Cd content retrieval from hyperspectral remote sensing data based on organic matter characteristic spectral bands[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2024, 55(1): 186-195. [张霞, 孙友鑫, 尚坤, 等. 基于有机质特征谱段的土壤 Cd 含量高光谱遥感反演[J]. 农业机械学报, 2024, 55(1): 186-195.]

(责任编辑: 檀满枝)