

DOI: 10.11766/trxb202502140060

CSTR: 32215.14.trxb202502140060

史泽峰, 徐明星, 赵成义, 焦彩霞, 陈桐, 曾荣, 卢婧, 郑光辉. 基于无人机成像光谱数据和机器学习算法的土壤盐分反演与制图研究[J]. 土壤学报, 2026, 63 (3): 704–717.

SHI Zefeng, XU Mingxing, ZHAO Chengyi, JIAO Caixia, CHEN Tong, ZENG Rong, LU Jing, ZHENG Guanghui. Research on Soil Salinity Inversion and Mapping Based on UAV Imaging Spectroscopy Data and Machine Learning Algorithms[J]. Acta Pedologica Sinica, 2026, 63 (3): 704–717.

基于无人机成像光谱数据和机器学习算法的土壤盐分反演与制图研究*

史泽峰¹, 徐明星², 赵成义¹, 焦彩霞¹, 陈桐¹, 曾荣¹, 卢婧³, 郑光辉^{1†}

(1. 南京信息工程大学地理科学学院, 南京 210044; 2. 浙江省地质院国土空间生态修复所, 杭州 310007; 3. 江苏省不动产登记中心, 南京 210017)

摘要: 土壤盐渍化严重制约农业可持续发展, 精准监测土壤盐分对农业管理和生态保护至关重要。本研究利用无人机成像光谱技术结合机器学习算法, 探索滨海地区土壤盐分含量(SSC)的反演与空间制图方法。通过竞争自适应重加权采样(CARS)算法筛选特征波段并计算光谱指数, 采用递归特征消除法(RFE)筛选光谱指数, 利用偏最小二乘回归(PLSR)、支持向量回归(SVR)、随机森林回归(RFR)构建 6 种不同光谱变换的全波段预测模型和利用 SVR、RFR、极端梯度提升(XGBoost)和反向传播神经网络(BPNN)构建光谱指数预测模型, 并通过精度评估选择最佳模型进行 SSC 空间制图。结果表明: 研究区实测 SSC 范围为 1.23~8.96 g·kg⁻¹, 均值为 3.12 g·kg⁻¹; 全波段模型中, 经 SG 平滑处理的原始光谱的 RFR 模型精度最高; 光谱指数模型中, 基于特征选择的 XGBoost 模型表现最优; 反演结果揭示了研究区土壤盐分中低含量广泛分布、高值零散分布的空间特征, XGBoost 适合全面预测整体分布, 而经 SG 平滑处理的原始光谱的 RFR 更适合低盐分情况的分布。本研究创新性地结合了全波段优化光谱指数与传统光谱指数来构建 SSC 预测模型, 为无人机成像光谱技术在田间尺度 SSC 快速监测提供了范例。

关键词: 盐渍化; 成像光谱; 机器学习; 土壤盐分; 光谱指数

中图分类号: S127/TP79 **文献标志码:** A

Research on Soil Salinity Inversion and Mapping Based on UAV Imaging Spectroscopy Data and Machine Learning Algorithms

SHI Zefeng¹, XU Mingxing², ZHAO Chengyi¹, JIAO Caixia¹, CHEN Tong¹, ZENG Rong¹, LU Jing³, ZHENG

* 国家重点研发计划项目(2023YFE0208100)、国家自然科学基金项目(42371060, 42130405)、自然资源部 2024 年度部省合作项目(2024ZRBSHZ115)资助 Supported by the National Key Research and Development Program of China(No. 2023YFE0208100), the National Natural Science Foundation of China (Nos. 42371060, 42130405) and Ministry of Natural Resources 2024 Provincial Cooperation Project (No. 2024ZRBSHZ115)

† 通讯作者 Corresponding author, E-mail: 001524@nuist.edu.cn

作者简介: 史泽峰(2001—), 男, 浙江绍兴人, 硕士研究生, 主要从事土壤遥感。E-mail: 202312100007@nuist.edu.cn

收稿日期: 2025-02-14; 收到修改稿日期: 2025-08-08; 网络首发日期(www.cnki.net): 2026-01-05

Guanghui^{1†}

(1. School of Geographical Science, Nanjing University of Information Science and Technology, Nanjing 210044, China; 2. Department of Territorial Space Ecological Restoration, Zhejiang Institute of Geosciences, Hangzhou 310007, China; 3. Jiangsu Real Estate Registration Center, Nanjing 210017, China)

Abstract: [Objective] Soil salinization seriously restricts the sustainable development of agriculture, and the accurate monitoring of soil salinity is crucial for agricultural management and ecological protection. This study combined unmanned aerial vehicle (UAV) imaging spectroscopy with machine learning algorithms to explore the inversion and spatial mapping of soil salt content (SSC) in coastal areas. [Method] Feature bands were selected using the Competitive Adaptive Reweighted Sampling (CARS) algorithm, and spectral indices were calculated. Spectral indices were selected using the Recursive Feature Elimination (RFE) method. Utilizing PLSR, SVR, and RFR, this study developed prediction models for all spectral bands based on six different spectral transformations, and spectral index prediction models were built using SVR, RFR, XGBoost, and BPNN. The best model was chosen for SSC spatial mapping through accuracy evaluation. [Result] The results showed that the measured soil salt content (SSC) in the study area ranged from 1.23 to 8.96 g kg⁻¹, with a mean of 3.12 g kg⁻¹. Among the full-spectrum models, the random forest regression (RFR) model based on raw spectra processed with Savitzky-Golay (SG) smoothing demonstrated the highest accuracy. For the spectral index models, the extreme gradient boosting (XGBoost) model with feature selection performed the best. The inversion results revealed that low-to-moderate soil salinity was widely distributed across the study area, with high salinity values scattered sporadically. While XGBoost was well-suited for predicting the overall spatial distribution of soil salinity, the RFR model based on SG-smoothed raw spectra was more effective for mapping areas with low salinity. [Conclusion] This study innovatively combined full-spectrum optimized spectral indices with traditional ones to build a SSC prediction model, offering a new technical path for rapid SSC monitoring in coastal regions using UAV imaging spectroscopy.

Key words: Salinization; Imaging spectroscopy; Machine learning; Soil salinity; Spectral index

土壤盐渍化是指在自然和人为因素的交互作用下,土壤中可溶性盐类含量不断增加,导致土壤肥力下降和农作物减产的一种土壤退化现象,极端情况下土壤表层会形成明显的盐结晶^[1]。盐渍化已经成为生态环境面临的重大挑战之一,对区域生态环境保护构成威胁^[2-4]。在我国,盐碱地面积近1亿hm²,主要分布在西北、东北、华北和滨海地区,大约占世界盐碱地总面积的十分之一^[5]。2024年中央一号文件^[6]聚焦推进乡村全面振兴,明确提出要加强盐碱地的综合治理,包括盐碱地的监测、评估以及土壤盐分的估算。因此,为实现农业可持续发展,对土壤盐分进行估算、监测及制图刻不容缓。

传统的土壤盐渍化监测是人工野外采样-室内测定,该方法费时、费力且成本高,难以精准反映盐渍化的空间变异特征。近年来,随着遥感技术的发展,基于遥感技术的土壤盐渍化监测和预测研究逐渐增多^[7]。许多学者利用卫星遥感影像建立基于光谱指数的土壤盐渍化预测模型并进行土壤盐渍化空间制图。例如Hassan等利用Landsat 8 OLI卫星图像计算的盐度指数和波段组合来估算孟加拉国巴

格哈特的土壤盐分并制图^[8];田义超等^[9]利用国产资源一号02D(ZY1-02D)卫星多光谱数据提取相关特征参数,评价了五种不同机器学习算法的性能,并预测了土壤盐分含量和空间分布。尽管卫星遥感技术可成功预测土壤盐分,但存在空间分辨率和时间分辨率低的不足,且土壤样本采集时间与卫星过境时间的差异可能影响其应用的有效性。

随着无人机技术的兴起,无人机遥感已成为区域土壤盐渍化监测的重要手段。相较于卫星遥感,无人机遥感具有高空间分辨率、实时、灵活和波段连续性强的优势^[10]。研究表明,无人机遥感在监测农田盐分方面表现出显著潜力^[7, 11]。例如,Wang等^[12]利用深度学习和无人机多光谱数据估算向日葵田的土壤盐分水平。李振^[13]探究了无人机高光谱的土壤盐分光谱响应规律,研究表明二维特征光谱能提高模型精度。

前人研究发现,不同的光谱变换处理方法、建模方法等都会影响预测模型的精度^[14-15]。经过变换处理的光谱模型可以提高土壤盐分预测准确度^[16]。不同研究表明,光谱数据的最佳预处理方法因研究

条件而异。李振^[13]比较了不同的光谱预处理方法,发现最优的去噪预处理方法能够提高预测模型精度。黄帅等^[17]发现对数二阶微分光谱构建的模型最优,关红等^[18]发现倒数一阶微分变换构建的模型最优,张贤龙等^[19]发现一阶导数和倒数一阶导数构建的模型较好。此外,赵文举等^[20]发现不同作物类型或不同土壤深度下的最优预测模型存在差异。高光谱遥感影像具有高维性和信息冗余等特征,可通过特征质量分析、特征选择、多特征优化等方法进行降维,从而进一步提高模型预测精度^[21]。

目前大多数研究采用无人机多光谱数据反演土壤盐分,且研究区多集中于干旱半干旱区。而基于无人机高光谱影像反演滨海地区土壤盐分的研究相对较少,缺少长江沉积物母质形成的滨海土壤盐分遥感反演特征研究。本文以滨海地区多年耕作农田为研究区域,通过 6 种光谱变换方法构建了全波段土壤盐分含量(SSC)预测模型。同时,创新地结合全波段优化光谱指数和传统光谱指数,构建了光谱指数 SSC 预测模型。在此基础上,本研究在田间尺度上实现 SSC 的反演和空间制图,并对比分析了制图结果,以探讨 SSC 最佳预测模型与制图方法。

1 材料与方 法

1.1 研究区域

慈溪市位于浙江省宁波市境内,坐落在杭州湾南岸。该市地处北亚热带南缘,属于亚热带季风气候,年均气温为 16℃,年平均降水量为 1 300mm。研究区位于慈溪市北部国家现代农业产业园内,农田面积约为 6.53 hm²。研究区土壤为潮盐土,呈中性或微碱性,成土母质为滨海沉积物,颗粒匀细、质地均一,含有可溶性盐分,其主要作物为水稻和小麦。

1.2 研究数据

1.2.1 无人机成像光谱数据采集 高光谱成像光谱数据采集于 2022 年 1 月 18 日 12:00~13:00 之间,天气晴朗无风。飞行平台为 DJI M300 无人机,无人机飞行高度为 120 m,对应图像空间分辨率为 3.12 cm,航向和旁向重叠率分别为 70%和 80%;成像光谱仪为 Cubert S185,光谱覆盖范围为 450~950 nm,共 125 个波段,光谱分辨率为 8 nm。飞行

前进行采集白帧和黑帧数据用于后期反射率计算和校正。因无人机高光谱数据不包含高精度地理坐标信息,在数据采集前于每个采样点位置中心放置一张 A4 纸作为定位标志。

1.2.2 土壤取样和水溶性盐分测定 土壤样本于 2022 年 1 月 18 日采集,当时研究区域处于裸土期,受植被等因素影响较小。无人机数据采集完成后,在每个采样位置以 A4 纸为中心的 1 m 范围内,采用五点采样法采集表层 0~5 cm 的土壤样品,样本混合装入布袋中并编号,样本总数为 40 个。

采集的土壤样本经风干、研磨和过 10 目筛等预处理。采用 5:1 水土比制备浸提液,通过质量法测定水溶性盐分含量^[22]。具体如图 1 所示。

1.2.3 光谱数据的提取与预处理 在采样点中心位置设置一个内半径为 15 cm、外半径为 80 cm 的圆环,以圆环内像元平均光谱数据为该采样点的原始光谱数据(R)^[23]。采用 Savitzky-Golay(SG)平滑法^[24]去除锯齿噪音同时保持光谱曲线的整体趋势^[25]。随后,对经 SG 平滑处理的光谱进行标准正态变量(SNV)、多重散射校正(MSC)、倒数的对数(Log(1/R))、倒数对数的一阶导数(FDR(Log(1/R)))和一阶导数(FDR)变换。

1.3 高光谱特征波段筛选

竞争自适应重加权采样(CARS)算法是一种基于迭代统计分析的光谱变量筛选方法,其核心思想源于达尔文“适者生存”理论。该算法通过蒙特卡洛采样技术构建模型,并利用偏最小二乘回归(PLSR)回归系数绝对值的百分比来衡量目标变量重要性。在经过 N 次迭代后,会逐步筛选出采样变量^[26]。最终,对比基于所筛选变量建立的 PLSR 模型的交互验证均方差(RMSECV),确定误差最小的变量集合为特征波段。

1.4 光谱指数构建

本研究选择 18 个传统光谱指数和 3 个全波段优化光谱指数(通过遍历所有可能的双波段组合,计算差值型、比值型、归一化型等形式的光谱指数,并联合相关性分析筛选最敏感波段组合)构建光谱指数模型,具体光谱指数如表 1 所示。

1.5 机器学习方法和精度评估

1.5.1 机器学习方法 (1)RFR 随机森林回归(Random Forest Regression)算法是由 Breiman^[34]于

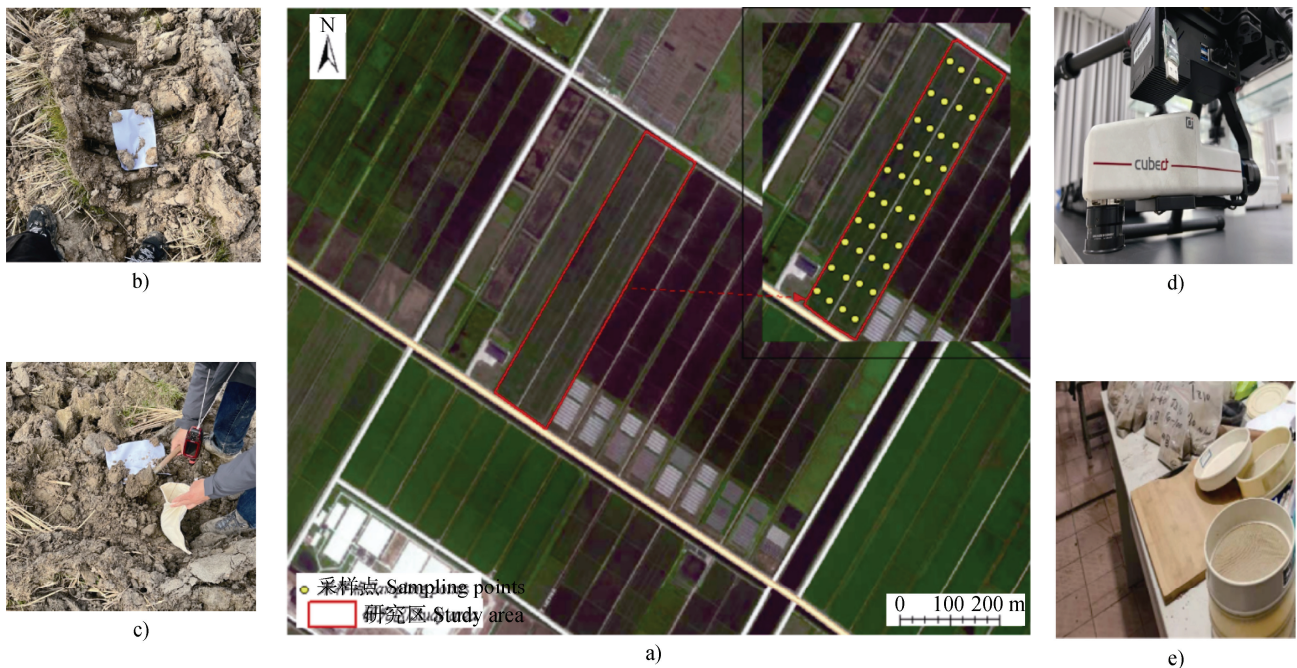


图 1 研究区采样点空间分布和实验流程 (a: 样点分布图, b: 标志点布设, c: 土壤采样, d: Cubert S185 成像光谱仪, e: 样本研磨筛选)

Fig. 1 Spatial distribution of sampling sites and experimental procedures in the study area (a) sampling location map; (b) marker placement; (c) soil sampling; (d) Cubert S185 imaging spectrometer; (e) sample grinding and sieving

2001 年提出的机器学习算法, 其适用于回归问题。其核心思想是通过构建多棵回归决策树, 并集成各树的预测结果, 实现对变量的回归预测。本研究在全波段模型中使用默认的决策树数量即 100 个来进行预测。

(2) SVR 支持向量回归 (Support Vector Regression) 算法由 Cortes^[35]于 1995 年提出的回归算法。其主要原理是通过在高维空间中找到一个最优的回归函数来适应数据。本文设定核函数类型为“rbf”, 采用网格搜索法 (Grid search) 进行参数寻优。

(3) BPNN 反向传播神经网络 (Back Propagation Neural Network) 的核心思想是通过多层神经网络来学习数据中的复杂模式, 其基本原理包括前向传播、误差计算、反向传播和参数更新。本文构建 BPNN 时并按照训练结果误差较小的标准, 通过多次试验, 确定其学习效率为 0.001, 激活函数为“relu”以及加入了正则化参数减少其过拟合问题。

(4) XGBoost 极端梯度提升 (eXtreme Gradient Boosting) 算法是一种基于梯度提升决策树的优化和

增强型机器学习算法。该算法由 Chen^[36]于 2016 年提出, 通过引入二阶泰勒展开式来优化目标函数, 并加入正则化, 从而能够更加精确地拟合数据。在训练过程中, 多个弱学习器 (通常是决策树) 依照前一轮的残差依次进行调整和训练。本文通过多次试验调整模型参数, 并确定了适合本研究数据的参数, 以达到误差最小化的效果。

1.5.2 模型优化与精确度评估 过拟合现象通常出现在模型性能较差的情况下, 其特征是模型在训练集上表现良好但在测试集上表现较差。为此, 可以采用数据增强和特征选择等方法: 通过向数据中添加噪声来扩充样本量, 从而增强模型泛化能力; 同时利用递归特征消除 (RFE) 法去除冗余特征, 以增强模型的稳定性和预测精度。

本研究采用 7:3 的比例划分训练集与测试集, 并结合 K 折交叉验证, 使用决定系数 (R^2) 和均方根误差 (RMSE) 作为评估指标, 以检验模型的稳定性与预测能力^[37]。决定系数 (R^2) 表示预测值与真实值之间的拟合程度, 值越接近于 1, 说明模型的精度越高。均方根误差 (RMSE) 主要用于模型验证, 衡量预测值与真实值之间的差异, 数值越小

代表模型精度越好^[38]。计算公式如下所示：

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{oi} - y_{pi})^2}{\sum_{i=1}^n (y_{oi} - \overline{y_{oi}})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{oi} - y_{oi})^2} \quad (2)$$

式中， y_{oi} 为真实值， y_{pi} 为预测值， $\overline{y_{oi}}$ 为真实值的平均值， n 为总样本个数。

2 结 果

2.1 土壤盐分统计特征

研究区的土壤盐分含量最小值和最大值分别为 $1.23 \text{ g}\cdot\text{kg}^{-1}$ 和 $8.96 \text{ g}\cdot\text{kg}^{-1}$ ，平均值为 $3.12 \text{ g}\cdot\text{kg}^{-1}$ ，符合中等盐渍化的标准^[39]；变异系数为 0.51，表明样本值存在较大差异（表 2）。此外，峰度和偏度分别为 4.43 和 1.91，说明数据呈现出高值的右偏态分布特征。

2.2 SSC 与土壤光谱之间的相关性

2.2.1 SSC 和全波段反射率之间的相关性 图 2

表 1 光谱指数计算公式

Table 1 Equation for spectral index calculation

光谱指数 Spectral index	公式 Formula	参考文献 Reference
S1 (盐分指数)	Blue/Red	[27]
S2 (盐分指数)	(Blue-Red) / (Red+Blue)	[27]
S3 (盐分指数)	(Green×Red) / Blue	[27]
S4 (盐分指数)	(Blue×Red) ^{0.5}	[27]
S5 (盐分指数)	(Blue×Red) / Green	[27]
S6 (盐分指数)	(Red×NIR) / Green	[27]
SI (盐分指数)	(Blue+Red) ^{0.5}	[28]
SI1 (盐分指数)	(Green×Red) ^{0.5}	[29]
SI2 (盐分指数)	(Green ² +Red ² +NIR ²) ^{0.5}	[30]
SI3 (盐分指数)	(Green ² +Red ²) ^{0.5}	[30]
SI5 (盐分指数)	(NIR-Red) / (Green-Red)	[30]
SI1reg (盐分指数)	(Green×RedEdge) ^{0.5}	[31]
SI3reg (盐分指数)	(Green ² +RedEdge ²) ^{0.5}	[28]
NDSI (归一化盐分指数)	(Red-NIR) / (Red+NIR)	[3]
NDSIreg (调节归一化差异盐分指数)	(RedEdge-NIR) / (RedEdge+NIR)	[31]
SI*S2 (组合盐分指数)	Blue×(Blue-Red) / Red×(Blue+Red)	
SI-T (盐分指数)	(Blue/NIR) × 100	[3]
INT1 (盐分指数)	(Green+Red) / 2	[32]
RI (比值光谱指数)	Band _i / Band _j	[33]
DI (差值光谱指数)	Band _i - Band _j	[33]
NDI (归一化光谱指数)	(Band _i - Band _j) / (Band _i + Band _j)	[33]

表 2 土壤盐分含量统计分析

Table 2 Statistics and analysis of soil salt content (SSC)

样本数 Observations	最小值 Minimum/ ($\text{g}\cdot\text{kg}^{-1}$)	最大值 Maximum/ ($\text{g}\cdot\text{kg}^{-1}$)	平均值 Mean/($\text{g}\cdot\text{kg}^{-1}$)	标准差 SD	变异系数 CV	峰度 Kurtosis	偏度 Skewness
40	1.23	8.96	3.12	1.61	0.51	4.43	1.91

展示了土壤光谱和土壤盐分含量之间的相关系数趋势。图中的原始光谱经过 SG 平滑处理，其他光谱变换均基于 SG 平滑后的原始光谱进行。结果表明，

原始光谱和 $\text{Log}(1/R)$ 变换，与其他光谱变换相比，与土壤盐分的相关性较弱（图 2a，图 2e）；相比之下，SNV、MSC、FDR 和 $\text{FDR}(\text{Log}(1/R))$ 变换

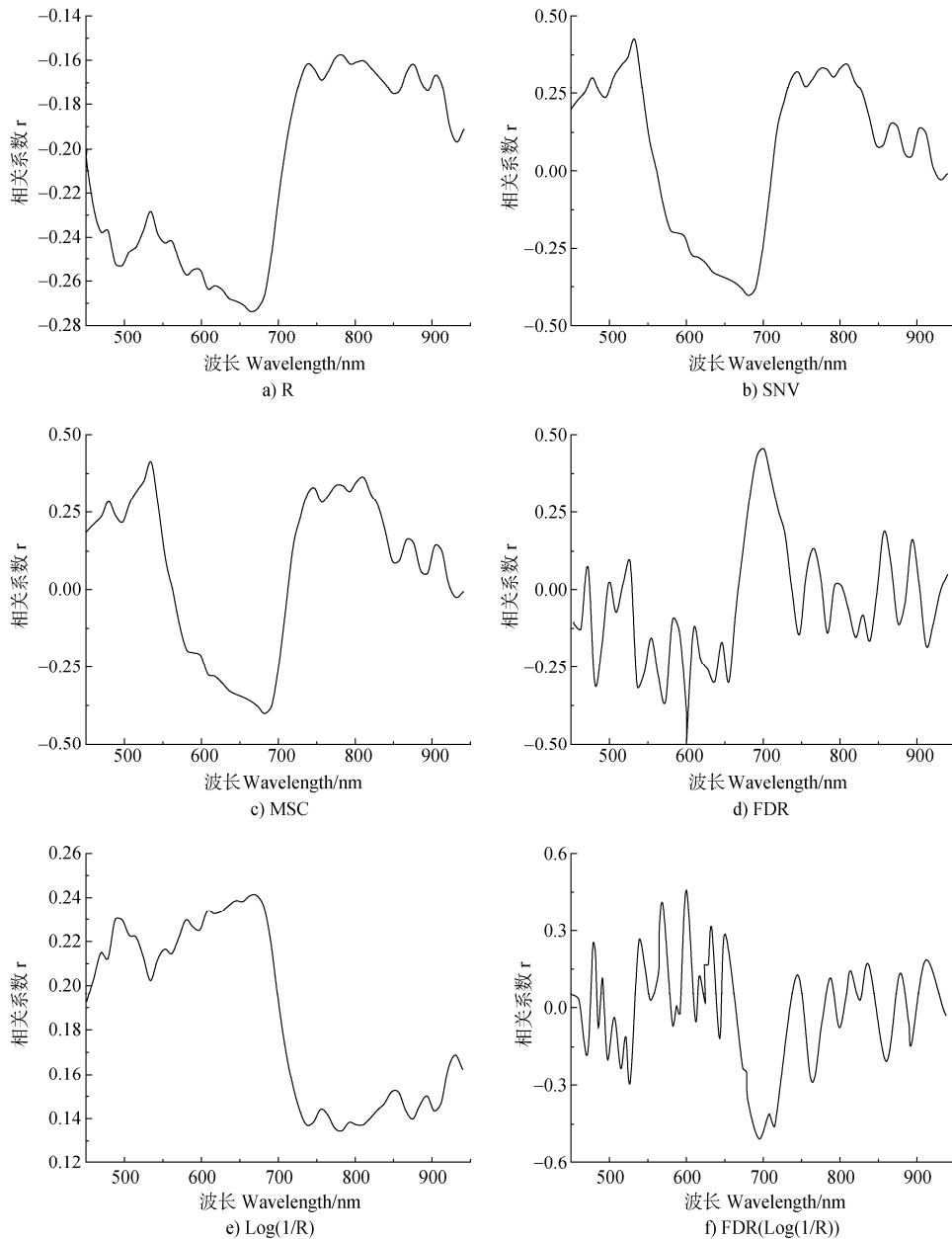


图 2 土壤光谱与 SSC 的相关性趋势

Fig. 2 Correlation trend between soil spectra and soil salt content (SSC)

后光谱与 SSC 的相关系数均有所提高;SNV 与 MSC 变换后的光谱与 SSC 的最大相关系数分别达到 0.42 和 0.4 (图 2b, 图 2c)。FDR 与 FDR (Log (1/R)) 变换后的光谱与 SSC 的最大相关系数达到 0.5 左右 (图 2d, 图 2f)。除 Log (1/R)、FDR 变换之外, 其他变换在大约 670~800 nm 的波长范围内 (即红光和红边间) 相关系数呈波动上升趋势, 并在该范围内的特定波长位置达到极大值。

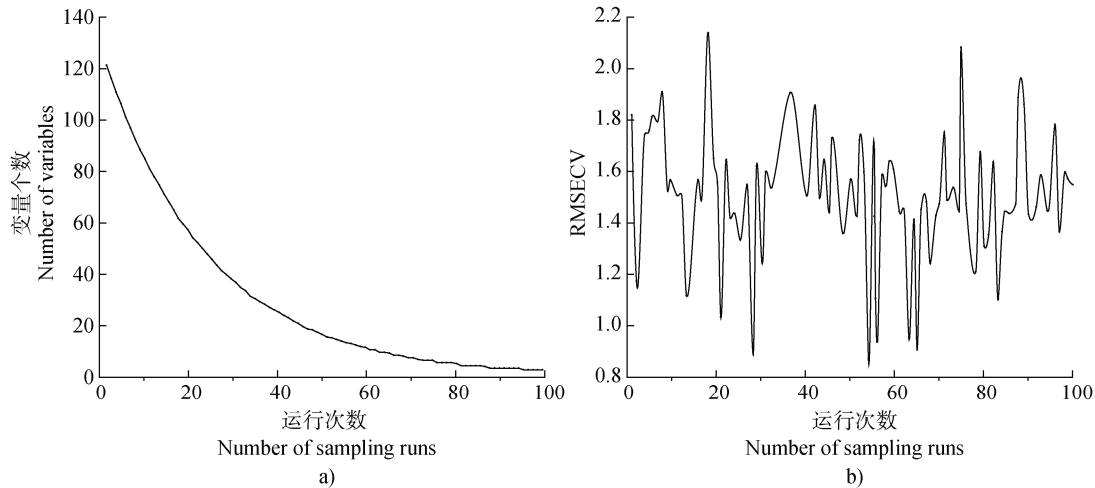


图 3 原始光谱 CARS 的降维结果

Fig. 3 Results of the original spectral by CARS downscaling

表 3 CARS 特征波段筛选结果

Table 3 CARS characteristic wavelength selection outcome

指标 Indicator	最佳运行		
	次数 Optimal number of runs	特征波段数 Number of feature bands	特征波段 Featured band/nm
SSC-SG-R	54	14	478、490、510、 530、570、646、 658、670、694、 702、730、 738、894、930

根据表 3、图 2 和无人机 S185 波长范围, 本研究选择 0.49、0.57、0.67、0.702、0.93 μm 作为蓝光、绿光、红光、红边和红外波段。利用 R 语言中 Cor 函数对 SSC 和传统光谱指数进行相关性分析, 结果表明传统光谱指数间相关性较高 (图 4a)。因此, 在构建光谱指数模型时, 需通过特征选择来降低传

2.2.2 SSC 和光谱指数之间的相关性 在本研究中, 鉴于全波段模型中原始波段的 RFR 模型表现最优, 采用 CARS 算法对经 SG 平滑处理的原始光谱反射率 (R) 进行特征波长筛选。随着 CARS 算法运行次数增加, 变量个数在前 40 次采样中显著减少 (图 3), 表明筛选出的变量逐渐减少。当运行次数为 54 次时, RMSECV 达到最小值, 表明此时光谱中与土壤盐分含量无关的信息最少。特征波长筛选见表 3。

统光谱指数间多重共线性。由表 4 可知, 实测点土壤盐分含量与传统光谱指数的相关系数中, 盐分指数与土壤盐分相关系数的绝对平均值为 0.21, 其中相关性较优的为 S3、S4、S5 等 8 个指数。

图 4b、图 4c、图 4d 展示了全波段优化光谱指数的相关性热力图 (等势图)。通过遍历 125 个波段的组合计算, 筛选出相关系数最大的敏感波段组合。结果表明, RI (比值光谱指数) 的最敏感波段组合为 690、726 nm, 相关系数为 0.55; DI (差值光谱指数) 的最敏感波段组合为 606、602 nm, 相关系数为 0.5; NDI (归一化光谱指数) 的最敏感波段组合为 690、726 nm, 相关系数为 0.54。

2.3 SSC 建模

2.3.1 基于全波段的 SSC 建模 本研究利用六种光谱形式分别构建 PLSR、SVR 和 RFR 模型, 结果见表 5。在 PLSR 建模结果中, 原始光谱的建模结果最佳, 其在建模集和预测集的 R^2 分别为 0.28 和 0.39。其中, 基于 FDR 变换光谱的 PLSR 建模结果表现为过拟合。在 SVR 建模结果中, 基于 FDR、

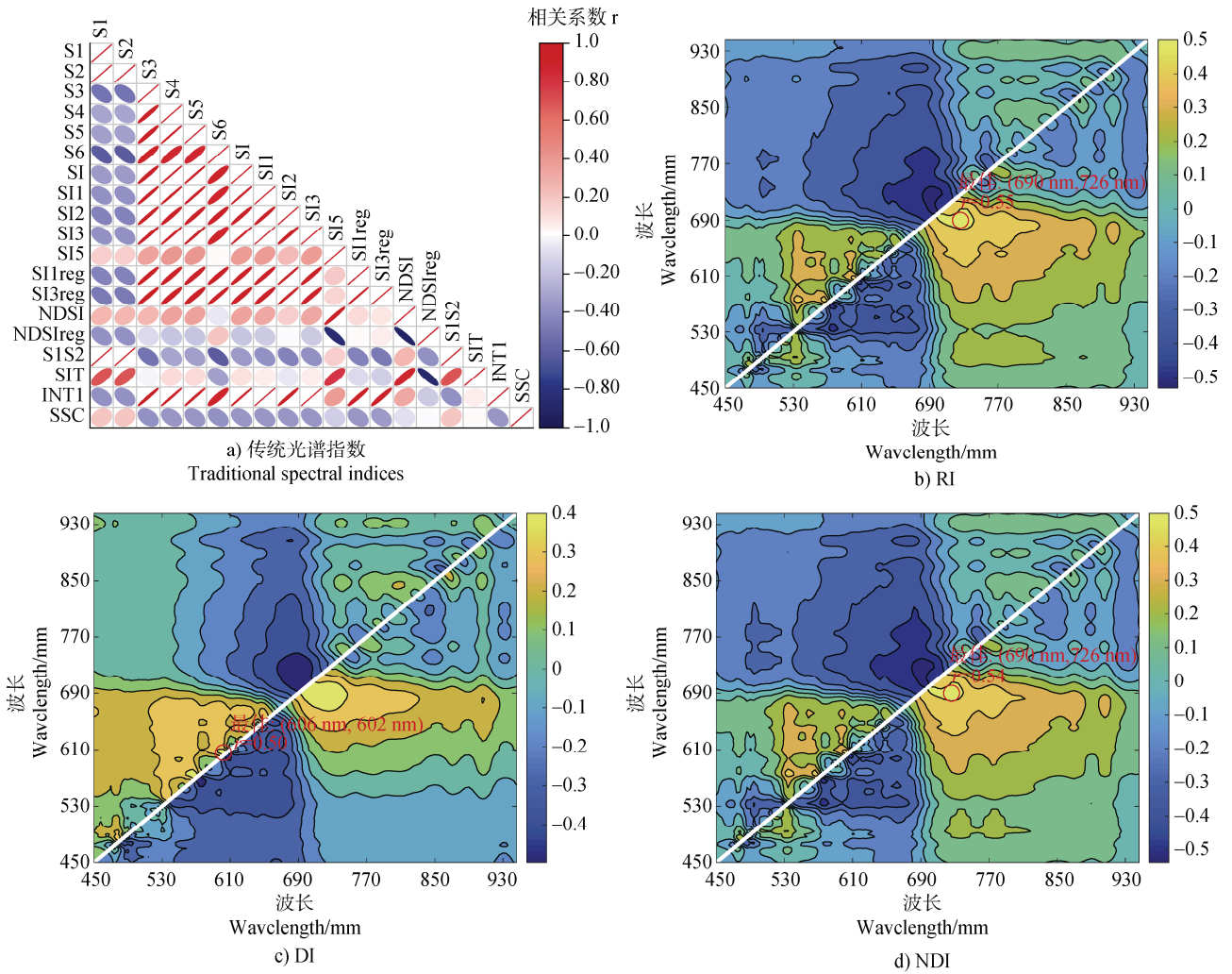


图 4 传统光谱指数以及全波段优化光谱指数和土壤盐分含量的相关性热力图

Fig. 4 Correlation heatmap of traditional spectral indices and full-band optimized spectral indices with soil salt content

表 4 土壤盐分含量与传统光谱指数的相关系数

Table 4 Correlation coefficient of soil salt content and traditional spectral indices

指数类型	变量	r	指数类型	变量	r	指数类型	变量	r
Index type	Variable	r	Index type	Variable	r	Index type	Variable	r
盐分指数 Salinity index	S1	0.12	盐分指数 Salinity index	S2	0.13	盐分指数 Salinity index	S3	-0.26
	S4	-0.27		S5	-0.28		S6	-0.22
	SI	-0.25		SI1	-0.26		SI2	-0.24
	SI3	-0.26		SI5	0.25		SI1reg	-0.23
	SI3reg	-0.22		NDSI	-0.21		NDSIreg	0.04
	S1×S2	0.11		SI-T	-0.09		INT1	-0.26

Log (1/R) 和 FDR (Log (1/R)) 变换光谱的 SVR 建模结果优于其他三种。相比之下，基于 SNV 和

MSC 变换光谱的建模结果过拟合，不适合预测。在 RFR 建模结果中，原始光谱的预测结果最佳，其次

表 5 全波段的建模结果

Table 5 Results of the full-spectrum modeling

光谱 Spectrum	模型 Model	R_c^2	RMSE _c	R_p^2	RMSE _p
R	PLSR	0.28	1.30	0.39	1.34
	SVR	0.34	1.41	0.46	0.89
	RFR	0.73	0.82	0.81	0.71
SNV	PLSR	0.13	1.41	0.37	1.38
	SVR	0.61	0.89	0.25	1.64
	RFR	0.61	0.92	0.76	0.86
MSC	PLSR	0.12	1.67	0.27	0.84
	SVR	0.58	1.01	0.23	1.47
	RFR	0.55	1.01	0.65	1.04
FDR	PLSR	0.41	1.35	0.15	1.03
	SVR	0.48	1.31	0.35	0.68
	RFR	0.62	0.94	0.76	0.84
Log (1/R)	PLSR	0.13	1.34	0.27	1.63
	SVR	0.51	1.28	0.69	0.42
	RFR	0.62	0.93	0.73	0.92
FDR (Log (1/R))	PLSR	0.33	1.44	0.19	0.87
	SVR	0.51	1.27	0.45	0.59
	RFR	0.78	0.65	0.30	1.35

是基于 SNV、FDR 和 Log (1/R) 变换光谱。然而基于 FDR (Log (1/R)) 变换光谱的 RFR 建模结果为过拟合, 预测能力差。综合来看, 在这三种预测方法中, RFR 模型的稳定性和建模结果明显优于其

他两种方法, 其中最优模型为基于原始光谱的 RFR 模型。

2.3.2 基于光谱指数的 SSC 建模 为防止出现过拟合现象, 本研究引入数据增强和特征选择。其中, 特征选择采用递归特征消除 (RFE) 方法, 其通过反复拟合模型并评估特征的重要性^[40], 逐步剔除最不重要的特征。在本研究中, 选择中位数作为阈值, 以确定特征的重要性排序, 直到达到指定的特征数量或理想的模型性能水平。本研究初始光谱指数共有 21 个, 经 RFE 特征选择后, 最终保留 10 个关键光谱指数, 特征选择结果如图 5 所示, 其中 RI、S3、SI5 等重要性贡献率较高。由表 6 可知, 基于特征选择的建模结果中, XGBoost 模型在训练集和测试集的 R^2 均大于等于 0.8, 展现出最高的精度, 优于其他三种模型的结果。

2.4 基于最优模型的 SSC 反演

根据表 5 和表 6 中的建模结果, 选择最优 SSC 预测模型, 并绘制散点图 (图 6)。从图 6 中可以观察到两个预测模型的预测值和真实值之间存在不同程度的偏差。当 SSC 含量高于 $8 \text{ g}\cdot\text{kg}^{-1}$ 时, 基于特征选择的 XGBoost 模型比基于原始光谱 RFR 模型更接近实测值。总体而言, 基于特征选择的 XGBoost 模型预测值和真实值更接近 1:1 线, 表明其预测精度较高。

本研究将 RFR 和 XGBoost 这两个最优模型应用于成像光谱数据, 通过逐像素反演生成 SSC 空间分布图 (图 7)。结果显示, 研究区域内大部分呈现中低盐分反演值, 高值区域分布较为分散。表 7 统

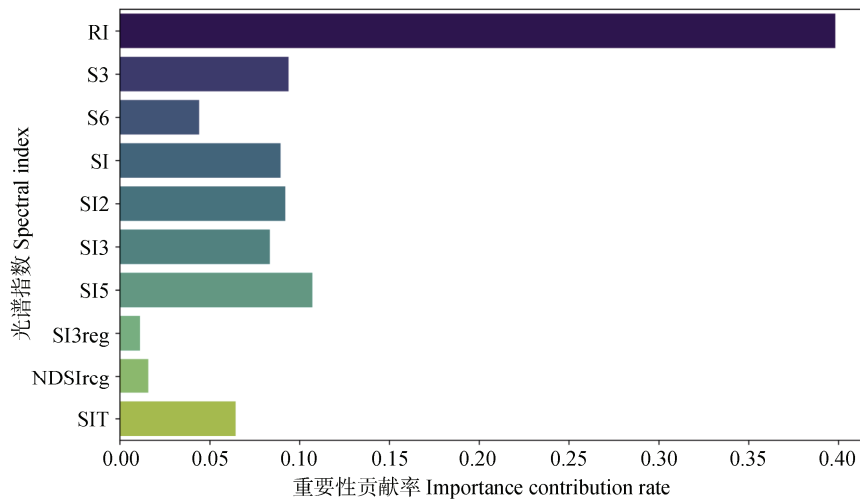


图 5 RFE 特征选择结果

Fig. 5 RFE feature selection results

表 6 基于特征选择的不同模型精度

Table 6 Accuracy of different models based on feature selection

模型 Model	R_c^2	RMSE _c	R_p^2	RMSE _p
SVR	0.89	0.51	0.75	0.76
RFR	0.90	0.49	0.73	0.80
XGBoost	0.96	0.28	0.85	0.60
BPNN	0.87	0.53	0.71	0.79

计结果表明, 基于原始光谱的 RFR 模型的平均值与实测均值高度吻合, 在预测低盐分水平时表现较好。然而, 其最大值预测显著低于实测极值, 且标准差也较实际偏低, 反映出对高盐分极值捕捉能力和数

据波动性方面存在一定局限性。相比之下, XGBoost 模型的最大值预测更接近实测极值, 尽管其标准差偏低, 但分布形态与实测数据更具相似性, 表明 XGBoost 模型在高盐分响应能力和整体分布特征捕捉能力方面表现良好。因此, 若研究重点在于高盐分区域识别和整体分布表征, XGBoost 模型更适合; 而针对低盐分区域分析和平均值精度优化, 基于原始光谱的 RFR 模型更具优势。此外, IDW 和 Kriging 作为空间插值方法, 其插值结果呈现独特的高值集聚特征 (见图 7), 这可能与采样点高值在空间上的集中分布有关。整体而言, IDW 和 Kriging 的插值结果以中低值为主, 与机器学习模型反演的分布图相似。

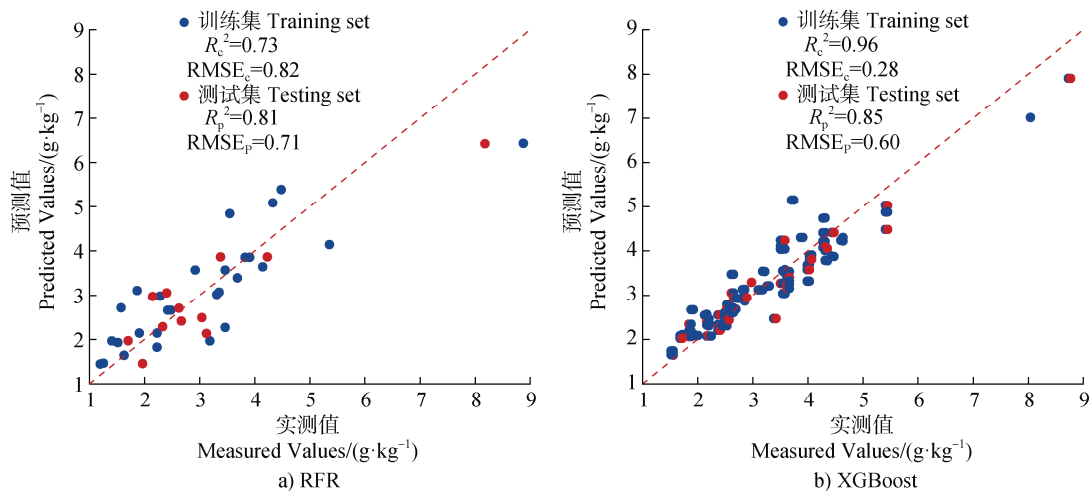


图 6 最优模型预测散点图

Fig. 6 Scatter plot of predictions from the optimal model

3 讨论

本研究通过从无人机成像光谱数据中筛选特定波段, 并结合传统光谱指数和二维光谱指数, 实现了对土壤盐分的有效反演。在利用全波段数据反演土壤盐分的过程中, 发现经过 MSC 和 SNV 变换后的光谱与土壤盐分的相关性趋势呈现出一致性, 而 $\text{Log}(1/R)$ 变换后的光谱和原始光谱的相关性趋势相反。此外 FDR 与 $\text{Log}(1/R)$ 变换后的光谱相关性趋势总体相反, 但在 450~490 nm 区间的趋势一致。尽管这些光谱变换能够增强部分吸收或反射特性, 但在后续的反演模型精度评价中发现, 原始光谱的 RFR 模型精度最高。这表明, 虽然一阶

导数或其他光谱变换方法可以提升与土壤盐分之间的相关性, 但可能会破坏原始光谱的特征和维度, 导致信息丢失, 并影响回归模型的性能。因此, 这些光谱变换方法的普适性仍需根据具体情况进行验证和调整。在全波段模型中, 理论上支持向量回归 (SVR) 和随机森林回归 (RFR) 模型应具有相似的预测精度, 但实际结果显示两者存在较大差异。这一差异主要源于样本量有限, 特征维度高以及 SVR 模型对噪声更敏感, 而 RFR 模型能自动处理高维特征数据, 对小样本数据更具有鲁棒性^[41]。传统光谱指数与土壤盐分含量的相关性分析, 以及光谱指数间相关性 (图 4a, 表 4) 表明, 各光谱指数间存在较强的相关性, 这说明若将全部光谱指数作为模型输入变量, 可能会出现过拟合现象, 凸显了进行特征选择的必要性。

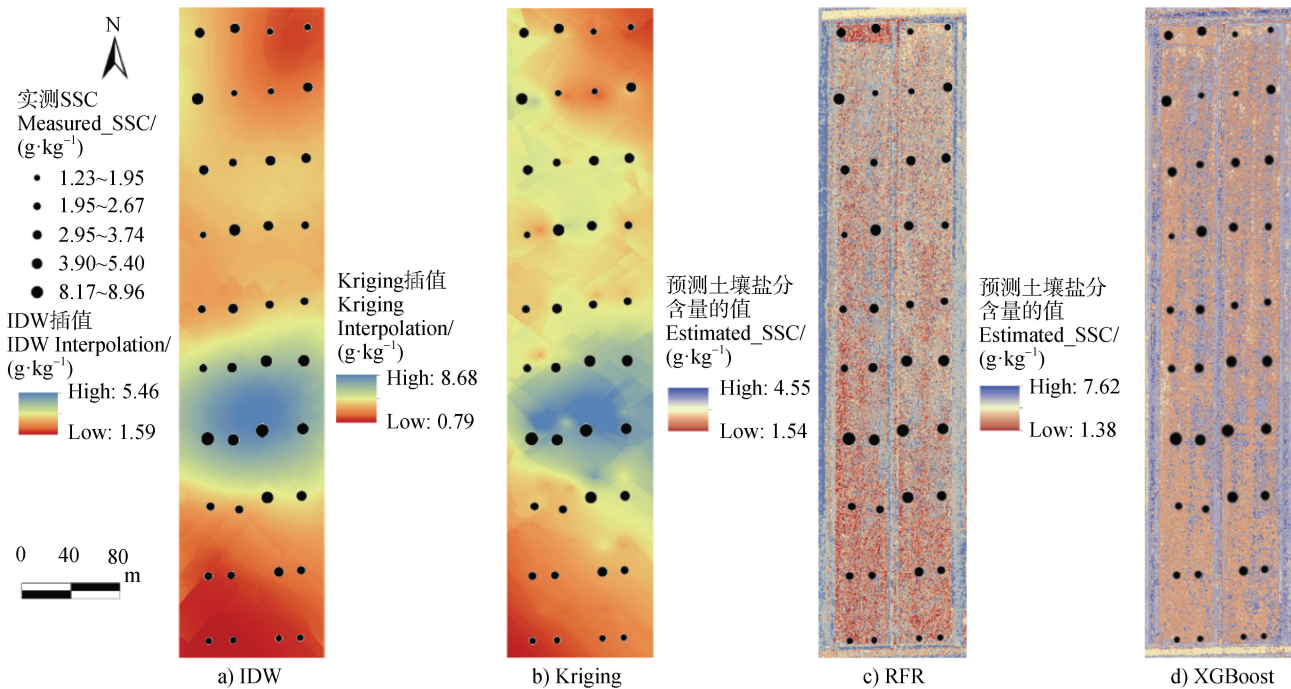


图 7 基于不同方法的土壤盐分含量空间分布图

Fig. 7 Spatial distribution map of SSC based on different methods

表 7 基于不同方法的土壤盐分含量统计特征

Table 7 Statistical features of soil salt content based on different methods

方法类型 Method category	最小值 Minimum/ ($\text{g}\cdot\text{kg}^{-1}$)	最大值 Maximum/ ($\text{g}\cdot\text{kg}^{-1}$)	平均值 Mean/ ($\text{g}\cdot\text{kg}^{-1}$)	标准差 SD
RFR	1.54	4.55	3.13	0.91
XGBoost	1.38	7.62	4.11	0.89
IDW	1.59	5.46	3.08	0.93
Kriging	0.79	8.68	3.04	1.04
实测值 Measured value	1.23	8.96	3.12	1.61

机器学习算法近年来已被广泛应用于土壤盐分研究,并取得显著成果。本研究以无人机高光谱图像为数据源,选用了四种机器学习算法进行土壤盐分反演。结果显示,经过数据增强和特征选择后的SVR模型和RFR模型在精度上表现类似,表明在特征维度较小且数据量相同时,这两种算法对土壤盐分的反演能力趋于一致,但具体应用场景仍需具体分析。同时基于特征选择的XGBoost模型在反演土壤盐分时表现出最高的精度,其在训练集和验证集的 R^2 分别为0.96、0.85, RMSE分别为0.28和0.60。

这一精度高于王怡婧等^[42]使用MSC变换下敏感波段组合的随机森林回归模型(R^2 分别为0.937、0.427, RMSE为3.74);并与程俊恺等^[43]基于一阶微分转换光谱结合随机森林回归方法的模型精度(R^2 分别为0.896、0.8854, RMSE为0.4131)相当。本研究表明, XGBoost机器学习算法在此研究区土壤盐分反演中,相较于其他回归模型具有更好的拟合效果,同时特征选择的引入有效地提升了模型性能。然而该模型对不同时空尺度的土壤盐分反演能力及其普适性仍需进一步探索和验证。

机器学习算法的选择、参数的优化和特征选择等均会对模型精度产生影响,从而导致预测结果的不确定性。相较于多光谱遥感数据,本研究采用的无人机高光谱遥感技术提供了更丰富的波段信息,这在提高土壤盐分含量反演精度的同时也引入了数据冗余,增加了数据处理和建模的复杂度。尽管本研究的样本数量有限,但研究区域为一块 6.53 hm² 的裸土期多年耕作农田,面积较小且采样点布设均匀,属于小样本、小尺度的土壤盐分反演。已有研究表明^[44-45]小样本数据在一定条件下能够实现较高的反演精度。此外,实测土壤盐分含量中存在两个高值点,这些高值点可能是局部盐分富集区(如灌溉排水不良或施肥不均区域),其存在反映了农田盐分的真实空间异质性。为应对样本数量有限的挑战,本研究在光谱指数模型中采用了数据增强和正则化参数,并结合特征选择方法。这些方法有效扩充了样本数量并降低了过拟合的风险,提升了模型的稳定性和预测精度。特征选择在小样本数据建模中起着至关重要的作用。通过递归特征消除(RFE)方法,本研究有效筛选出敏感光谱指数,降低了特征间的多重共线性,同时保留了关键信息,使模型更加高效和准确。以上方法的应用为小样本条件下的土壤盐分反演提供了有效的解决方案。

在未来的研究中,可以通过增加实测土壤样本数量,并结合敏感波段组合和空间特征,进一步丰富特征选择方法,以增强土壤盐分反演模型的稳定性和普适性,为今后实现空天地一体化监测提供科学依据和技术支持^[46]。此外,土壤盐分的监测与预测应从更大范围、更精细尺度以及更全面的角度出发,充分考虑地理大数据的特性以及应用要求,优化模型设计,以提升模型的实用性和推广性^[47]。

4 结 论

本研究以慈溪滨海地区的土壤盐分含量为研究对象,利用无人机成像光谱数据,建立基于全波段数据和光谱指数的预测模型,并通过最优模型对研究区土壤盐分含量进行了空间制图和分析,得到以下结论:全波段模型中经 SG 平滑的原始光谱 RFR 模型表现出最高的预测精度;在光谱指数模型中,基于特征选择的 XGBoost 模型表现出最优的预测性能,适用于本研究区的土壤盐分反演。反演结果图

表明,该地区土壤盐分空间分布呈现出中低含量广泛分布、高值零散分布的特征;XGBoost 模型较为适用于全面、准确地预测土壤盐分的整体分布,而基于原始光谱的 RFR 模型更适用于较低盐分及其平均值的预测。因此,本研究创新地结合全波段优化光谱指数和传统光谱指数构建的预测模型,可以为无人机成像光谱技术在田间尺度 SSC 快速监测提供了范例。

参考文献 (References)

- [1] Abrol I P, Yadav J S P, Massoud F I. Salt-affected soils and their management[M]. Food & Agriculture Organization, 1988.
- [2] Ivushkin K, Bartholomeus H, Bregt A K, et al. Global mapping of soil salinity change[J]. Remote Sensing of Environment, 2019, 231: 111260.
- [3] Allbed A, Kumar L, Aldakheel Y Y. Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-spatial resolution imageries: Applications in a date palm dominated region[J]. Geoderma, 2014, 230: 1—8.
- [4] Metternicht G I, Zinck J A. Remote sensing of soil salinity: Potentials and constraints[J]. Remote Sensing of Environment, 2003, 85 (1): 1—20.
- [5] Map of Global Saline Soil Distribution Released[J]. China Agricultural Comprehensive Development, 2021 (10): 64. [世界盐渍土壤分布图发布[J]. 中国农业综合开发, 2021 (10): 64.]
- [6] Xinhua Review. Central Committee of the Communist Party of China, and State Council of the People's Republic of China. Opinions on Learning and Applying the Experience of the "Thousand Villages Demonstration and Ten Thousand Villages Renovation" Project to Effectively Promote Comprehensive Rural Revitalization[J]. 2024. [新华时评. 中共中央国务院关于学习运用“千村示范,万村整治”工程经验有力有效推进乡村全面振兴的意见[J]. 2024.]
- [7] Ivushkin K, Bartholomeus H, Bregt A K, et al. UAV based soil salinity assessment of cropland[J]. Geoderma, 2019, 338: 502—512.
- [8] Hassan R, Ahmed Z, Islam M T, et al. Soil salinity detection using salinity indices from Landsat 8 satellite image at Rampal, Bangladesh[J]. Remote Sensing in Earth Systems Sciences, 2021, 4: 1—12.
- [9] Tian Y C, Zheng D L, Zhang Q, et al. Inversion of coastal soil salinity in Qinzhou Bay based on Chinese satellite ZY-1 02D and machine learning algorithms[J]. China Environmental Science, 2024, 44 (1): 371—385. [田义超, 郑丹琳, 张强, 等. 基于国产资源一号 02D 卫星和机器学习算法的钦州湾滨海土壤盐分反演[J]. 中国环境科学, 2024, 44 (1): 371—385.]

- [10] Yan D D, Li J T, Yao X Y, et al. Integrating UAV data for assessing the ecological response of *Spartina alterniflora* towards inundation and salinity gradients in coastal wetland[J]. *Science of the Total Environment*, 2022, 814: 152631.
- [11] Wang D Y, Chen H Y, Wang G F, et al. Study on the inversion of severe saline soil salinity in the Yellow River Delta using UAV multispectral data[J]. *Scientia Agricultura Sinica*, 2019, 52 (10): 1698—1709. [王丹阳, 陈红艳, 王桂峰, 等. 无人机多光谱反演黄河口重度盐渍土盐分的研究[J]. *中国农业科学*, 2019, 52 (10): 1698—1709.]
- [12] Wang Y X, Qu Z Y, Yang W, et al. Inversion of soil salinity in the irrigated region along the southern bank of the Yellow River using UAV multispectral remote sensing[J]. *Agronomy*, 2024, 14 (3): 523.
- [13] Li Z. Estimation of farmland soil salinity based on UAV hyperspectral data[D]. Jinan : Shandong Normal University, 2023. [李振. 基于无人机高光谱的农田土壤盐分估测研究[D]. 济南: 山东师范大学, 2023.]
- [14] Rinnan Å, Van Den Berg F, Engelsen S B. Review of the most common pre-processing techniques for near-infrared spectra[J]. *Trends in Analytical Chemistry*, 2009, 28 (10): 1201—1222.
- [15] Lucà F, Conforti M, Castrignanò A, et al. Effect of calibration set size on prediction at local scale of soil carbon by Vis-NIR spectroscopy[J]. *Geoderma*, 2017, 288: 175—183.
- [16] Zhu C M, Ding J L, Zhang Z P, et al. Exploring the potential of UAV hyperspectral image for estimating soil salinity: Effects of optimal band combination algorithm and random forest[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2022, 279: 121416.
- [17] Huang S, Ding J L, Li X, et al. Hyperspectral characteristics analysis and modeling of soil salinization[J]. *Chinese Journal of Soil Science*, 2016, 47 (5): 1042—1048. [黄帅, 丁建丽, 李相, 等. 土壤盐渍化高光谱特征分析与建模[J]. *土壤通报*, 2016, 47 (5): 1042—1048.]
- [18] Guan H, Jia K L, Zhang Z N, et al. Spectral characteristics analysis and modeling of salinized soil[J]. *Remote Sensing for Land and Resources*, 2015, 27 (2): 100—104. [关红, 贾科利, 张至楠, 等. 盐渍化土壤光谱特征分析与建模[J]. *国土资源遥感*, 2015, 27 (2): 100—104.]
- [19] Zhang X L, Zhang F, Zhang H W, et al. Optimization of hyperspectral index models for soil salinity retrieval based on spectral transformation[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.]
- [20] Zhao W J, Ma F F, Ma H, et al. Soil salinity inversion model based on UAV multispectral images[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2022, 38 (24): 93—101. [赵文举, 马芳芳, 马宏, 等. 基于无人机多光谱影像的土壤盐分反演模型[J]. *农业工程学报*, 2022, 38 (24): 93—101.]
- [21] Su H J. Dimensionality reduction of hyperspectral remote sensing images: progress, challenges, and prospects[J]. *Journal of Remote Sensing*, 2022, 26 (8): 1504—1529. [苏红军. 高光谱遥感影像降维: 进展, 挑战与展望[J]. *遥感学报*, 2022, 26 (8): 1504—1529.]
- [22] Soil Testing. Part 16: Determination of soil water-soluble salt content: NY/T 1121.16-2006[S], 2006. [土壤检测. 第 16 部分: 土壤水溶性盐总量的测定: NY/T 1121.16-2006[S], 2006.]
- [23] Zheng G H, Chen T, Wang Y, et al. Rapid monitoring of the spatial distribution of soil organic matter using unmanned aerial vehicle imaging spectroscopy[J]. *Annals of GIS*, 2024, 30 (3): 367—381.
- [24] Savitzky A, Golay M J E. Smoothing and differentiation of data by simplified least squares procedures[J]. *Analytical Chemistry*, 1964, 36 (8): 1627—1639.
- [25] Long Q, Li H T, Zhang Y Y, et al. Upconversion nanoparticle-based fluorescence resonance energy transfer assay for organophosphorus pesticides[J]. *Biosensors and Bioelectronics*, 2015, 68: 168—174.
- [26] Zhou J P, Xu Y P, Gu X H, et al. High-precision mapping of soil organic matter based on UAV imagery using machine learning algorithms[J]. *Drones*, 2023, 7 (5): 290.
- [27] Abbas A, Khan S. Using remote sensing techniques for appraisal of irrigated soil salinity[C]//International Congress on Modelling and Simulation (MODSIM). Modelling and Simulation Society of Australia and New Zealand, 2007: 2632—2638.
- [28] Nicolas H, Walter C. Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data[J]. *Geoderma*, 2006, 134 (1/2): 217—230.
- [29] Ding J L, Yu D L. Monitoring and evaluating spatial variability of soil salinity in dry and wet seasons in the Werigan-Kuqa Oasis, China, using remote sensing and electromagnetic induction instruments[J]. *Geoderma*, 2014, 235: 316—322.
- [30] Bannari A, Guedon A M, El - Harti A, et al. Characterization of slightly and moderately saline and sodic soils in irrigated agricultural land using simulated data of advanced land imaging (EO - 1) sensor[J]. *Communications in Soil Science and Plant Analysis*, 2008, 39 (19/20): 2795—2811.
- [31] Zhang S M, Zhao G X, Lang K, et al. Integrated satellite,

- unmanned aerial vehicle (UAV) and ground inversion of the SPAD of winter wheat in the reviving stage[J]. *Sensors*, 2019, 19 (7): 1485.
- [32] Wang J Z, Ding J L, Yu D L, et al. Capability of Sentinel-2 MSI data for monitoring and mapping of soil salinity in dry and wet seasons in the Ebinur Lake region, Xinjiang, China[J]. *Geoderma*, 2019, 353: 172—187.
- [33] Cao X Y, Ding J L, Ge X Y, et al. Estimation of soil electrical conductivity based on spectral indices and machine learning algorithms[J]. *Acta Pedologica Sinica*, 2020, 57 (4): 867—877. [曹肖奕, 丁建丽, 葛翔宇, 等. 基于光谱指数与机器学习算法的土壤电导率估算研究[J]. *土壤学报*, 2020, 57 (4): 867—877.]
- [34] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45: 5—32.
- [35] Cortes C. Support-vector networks[J]. *Machine Learning*, 1995, 20 (3): 273—297.
- [36] Chen T Q, Guestrin C. Xgboost: A scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. San Francisco: Association for Computing Machinery, 2016: 785—794.
- [37] Nash J E, Sutcliffe J V. River flow forecasting through conceptual models part I—A discussion of principles[J]. *Journal of Hydrology*, 1970, 10 (3): 282—290.
- [38] Powell S L, Cohen W B, Healey S P, et al. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data : A comparison of empirical modeling approaches[J]. *Remote Sensing of Environment*, 2010, 114 (5): 1053—1068.
- [39] National Marine Standards Technical Committee (SAC/TC 283). Technical specifications for monitoring and evaluation of salinization of marine soils: HY/T 0320-2021[S]. Beijing: China Standards Press, 2021. [全国海洋标准化技术委员会 (SAC/TC 283). 滨海土壤盐渍化监测与评价技术规程: HY/T 0320-2021[S].北京: 中国标准出版社, 2021.]
- [40] Zhang X T, Huang W, Fu P H, et al. Research on digital soil mapping based on feature selection algorithm[J]. *Acta Pedologica Sinica*, 2024, 61 (3): 635—647. [张晓婷, 黄魏, 傅佩红, 等. 基于特征筛选算法的数字土壤制图研究[J]. *土壤学报*, 2024, 61 (3): 635—647.]
- [41] Hastie T, Tibshirani R, Friedman J H, et al. The elements of statistical learning: Data mining, inference, and prediction[M]. New York, USA: Springer, 2009.
- [42] Wang Y J, Ding Q D, Zhang J H, et al. Hyperspectral UAV remote sensing and machine learning-based inversion of soil moisture-salinity information[J]. *Chinese Journal of Applied Ecology*, 2023, 34 (11): 3045—3052. [王怡婧, 丁启东, 张俊华, 等. 基于无人机高光谱遥感和机器学习的土壤水盐信息反演[J]. *应用生态学报*, 2023, 34 (11): 3045—3052.]
- [43] Cheng J K, Feng X L, Chen L B, et al. Model optimization for cultivated land soil salinity inversion using UAV hyperspectral data[J]. *Chinese Journal of Applied Ecology*, 2024, 35 (11): 3085—3094. [程俊恺, 冯秀丽, 陈立波, 等. 基于无人机高光谱数据的耕地土壤盐分反演模型优选[J]. *应用生态学报*, 2024, 35 (11): 3085—3094.]
- [44] Wu X, Wang C J, Fan L Q, et al. Applicability analysis of salinization assessment indices based on multispectral remote sensing for soil salinity prediction in Yinbei Irrigation District, Ningxia[J]. *Remote Sensing for Land and Resources*, 2021, 33 (2): 124—133. [吴霞, 王长军, 樊丽琴, 等. 基于多光谱遥感的盐渍化评价指数对宁夏银北灌区土壤盐度预测的适用性分析[J]. *国土资源遥感*, 2021, 33 (2): 124—133.]
- [45] Kumar P, Tiwari P, Biswas A, et al. Spatio-temporal assessment of soil salinization utilizing remote sensing derivatives, and prediction modeling: Implications for sustainable development[J]. *Geoscience Frontiers*, 2024, 15 (6): 101881.
- [46] Zhang Z T, Chen Q D, Huang X Y, et al. Soil salinity monitoring model using airborne and satellite remote sensing upscaling based on a weighted algorithm[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53 (9): 226—238, 251. [张智韬, 陈钦达, 黄小鱼, 等. 基于加权算法的空-天遥感升尺度土壤含盐量监测模型[J]. *农业机械学报*, 2022, 53 (9): 226—238, 251.]
- [47] Wang J Z, Ding J L, Ge X Y, et al. Monitoring soil salinization on the basis of remote sensing and proximal soil sensing: Progress and perspective[J]. *Journal of Remote Sensing*, 2024, 28 (9): 2187—2208. [王敬哲, 丁建丽, 葛翔宇, 等. 基于星地传感技术的土壤盐渍化监测进展与展望[J]. *遥感学报*, 2024, 28 (9): 2187—2208.]

(责任编辑: 陈德明)