

基于 Kubelka – Munk 理论的稻田土壤含水率 高光谱反演研究

魏 薇¹ 赵刘文涛¹ 何 林¹ 李思汉¹ 单国富¹ 杜 俊^{1,2}

(1. 华中农业大学工学院, 武汉 430070; 2. 农业农村部长江中下游农业装备重点实验室, 武汉 430070)

摘要: 稻田土壤含水率是影响水稻生长的关键指标,利用高光谱技术实现其快速检测具有重要应用价值。针对高光谱在稻田土壤含水率宽幅变化条件下反演精度受限的问题,基于 Kubelka – Munk (KM) 理论建立辐射传输模型,利用实测稻田土壤高光谱及含水率逐波长求解吸收系数和散射系数,并通过 2 种途径提高含水率反演精度。以吸收散射比作为输入特征,采用最小二乘 (Ordinary least squares, OLS) 与多层感知机 (Multi-layer perceptron, MLP) 结合构建含水率半经验模型 (KM – OLS – MLP); 利用吸收系数和散射系数重建模拟光谱,扩充训练集样本实现数据增强 (KM-data augmentation, KM – DA), 并构建含水率预测模型。结果表明,稻田土壤质量含水率在 26.33% ~ 61.09% 区间时, KM – OLS – MLP 模型精度最高,测试集决定系数 R^2 为 0.924 1, 均方根误差 RMSE 为 2.41%; 模拟光谱在测试集上的 RMSE 和结构相似性指数 (Structural similarity index, SSIM) 分别为 0.008 1 和 0.996 4。数据增强后, 偏最小二乘回归 (Partial least squares regression, PLSR)、支持向量回归 (Support vector regression, SVR) 以及 BP 神经网络 (Back propagation neural network, BPNN) 模型精度均得到提高。BPNN 模型精度提高最显著, 模拟光谱与实测光谱比例为 2:1 时, 测试集 R^2 从 0.790 6 提高至 0.898 7, RMSE 从 4.00% 降低至 2.78%。该研究可为含水率宽幅变化条件下的稻田土壤含水率高光谱反演提供参考。

关键词: 稻田土壤含水率; Kubelka – Munk 理论; 高光谱反演; 吸收系数; 散射系数; 数据增强

中图分类号: S152.7 **文献标识码:** A **文章编号:** 1000-1298(2026)19-0365-10

OSID:



Hyperspectral Inversion of Soil Moisture Content in Paddy Fields Based on Kubelka – Munk Theory

Wei Wei¹ Zhao Liuwentao¹ He Lin¹ Li Sihan¹ Shan Guofu¹ Du Jun^{1,2}

(1. College of Engineering, Huazhong Agricultural University, Wuhan 430070, China

2. Key Laboratory of Agricultural Equipment in Mid-lower Yangtze River,
Ministry of Agriculture and Rural Affairs, Wuhan 430070, China)

Abstract: Soil moisture content in paddy fields is a critical indicator affecting rice growth, and its rapid detection using hyperspectral technology holds significant application value. To address the limitations of hyperspectral inversion accuracy under wide-ranging variations in paddy soil moisture content, a radiative transfer model was established based on the Kubelka – Munk (KM) theory. Using the measured hyperspectral data and the corresponding soil moisture content values of paddy soil, the absorption and scattering coefficients were estimated at each wavelength. Two approaches were employed to enhance the inversion accuracy of soil moisture content. Firstly, a semi-empirical model (KM – OLS – MLP) was developed by using the absorption-to-scattering ratio as an input feature, combining ordinary least squares (OLS) and multi-layer perceptron (MLP). Secondly, simulated spectra were reconstructed using the absorption and scattering coefficients to augment the training dataset (KM-data augmentation, KM – DA), and a soil moisture prediction model was constructed. The results showed that when the mass

收稿日期: 2026 – 04 – 08 修回日期: 2026 – 05 – 14

基金项目: 国家自然科学基金项目 (32101625, 32271994)、湖北省技术创新计划项目 (2024BB047)、中央高校基本科研业务费专项资金项目 (2662024JC002) 和湖北省农业关键核心技术攻关项目 (HBNYHXGG2023 – 2)

作者简介: 魏薇 (1981 –), 女, 教授, 博士, 主要从事智慧农业技术与装备研究, E-mail: weiwei1981@mail.hzau.edu.cn

通信作者: 杜俊 (1987 –), 男, 副教授, 博士, 主要从事现代农业装备设计与测控研究, E-mail: dujun@mail.hzau.edu.cn

moisture content of paddy soil ranged from 26.33% to 61.09%, the KM-OLS-MLP model achieved the highest accuracy, with a coefficient of determination (R^2) of 0.9241 and a root mean square error (RMSE) of 2.41% on the test set. The simulated spectra exhibited an RMSE of 0.0081 and a structural similarity index (SSIM) of 0.9964 on the test set. After data augmentation, the accuracy of the partial least squares regression (PLSR), support vector regression (SVR), and back propagation neural network (BPNN) models was improved, with BPNN model showing the most significant enhancement. When the ratio of simulated to measured spectra was 2:1, the test set R^2 was increased from 0.7906 to 0.8987, and the RMSE was decreased from 4.00% to 2.78%. The research result can provide a valuable reference for hyperspectral inversion of soil moisture content in paddy fields under wide-varying moisture conditions.

Key words: soil moisture in paddy fields; Kubelka-Munk theory; hyperspectral inversion; absorption coefficient; scattering coefficient; data augmentation

0 引言

土壤含水率(Soil moisture content, SMC)是调控作物生长、养分利用及农田水效率的核心指标,直接影响农业产量与可持续性^[1]。与旱地不同,稻田 SMC 因淹水或干湿交替而呈现高值且宽幅变化的特征,该特征显著影响水稻根系发育与产量形成,调控土壤养分转化等生态过程^[2]。因此,实现稻田 SMC 精准测量对优化水分管理、保障水稻高产高效与农业节水具有重要意义。

传统 SMC 检测方法有烘干法、时域反射法等,虽精度较高,但存在操作复杂、耗时费力、易受环境干扰等局限,难以满足精准农业需求。近年来,光谱技术凭借快速、无损、无需复杂前处理等优势,为土壤成分分析提供了新途径^[3]。该技术能够建立光谱数据与 SMC 之间的定量关系,从而实现对待测样品水分含量的高效、非破坏性检测^[4]。当前大多数使用实测农田 SMC 进行建模的研究中^[5-7],质量含水率不超过 35%,且变化范围不超过 15 个百分点。部分研究通过制备不同含水率的样本用于建模,虽然变化范围略大,但土壤样本经过了风干、研磨等实验室处理^[8],模型实用性有限。而稻田在淹水状态下的土壤质量含水率可达 50%~70%^[9-10],在干湿交替的情形下含水率变化范围超过 30 个百分点。土壤光谱反射率随含水率增加而降低,但下降速率逐渐减缓;当含水率超过田间持水量后,反射率转为随含水率增加而升高^[11]。该现象使得模型难以充分捕捉光谱特征与稻田 SMC 之间的内在关联,目前针对稻田土壤含水率的光谱反演研究较为匮乏。

当前,在土壤含水率高光谱反演研究中,通常采用机器学习算法或神经网络构建预测模型^[12-14],并通过预处理、特征提取^[15]等方式提高模型精度。这些方法均属于纯数据驱动的方法,仅依赖土壤光谱

数据与回归目标值之间的映射关系,缺乏对光与土壤介质相互作用物理过程的描述,导致模型可解释性差、泛化能力受限。辐射传输模型通过模拟光在土壤介质中的散射、吸收与传输过程,能够揭示光谱响应背后的机理。其中,Kubelka-Munk(KM)理论是描述光在不透明介质中传播与吸收、散射关系的二向通量辐射传输模型,已被广泛用于造纸与纺织品的反射检测、粉末及粗糙表面的红外检测等领域^[16]。然而,目前将 KM 理论应用于土壤光谱分析的研究尚处于起步阶段。文献[17-18]验证了 KM 理论应用于土壤光谱分析的可行性,但所用样品均经过了实验室处理,且暂未将模拟光谱应用于土壤成分含量反演。

针对上述问题,本文将基于 KM 理论建立土壤二向通量辐射传输模型,利用实测稻田土壤高光谱及含水率,逐波长求解吸收与散射系数,重建模拟光谱,并通过 2 种途径提高 SMC 高光谱反演精度。探究 KM 理论应用于稻田土壤光谱分析及 SMC 反演的适用性,以期为含水率宽幅变化条件下的稻田土壤含水率高光谱反演提供技术支撑。

1 研究区域与实验数据

1.1 研究区域

本实验采集的土壤样本来自上海市崇明区横沙岛(31°20'20.59"N,121°50'15.49"E),如图 1 所示。横沙岛位于长江入海口最东端,地势北低南高,岛面平坦,平均海拔 2.8 m,总面积 54.4 km²。属亚热带季风气候,年均气温 15.4℃,年降水量 1 140 mm,盛产水稻、棉花、油菜等农作物。

1.2 样本及数据获取

于 2025 年 8 月 9—20 日在上海市横沙岛采集稻田土壤样本,处于当地水稻生长的分蘖期,分别在灌水前后随机选取不同田块进行采样,以确保涵盖从低于到高于田间持水量的水分条件,乃至土壤淹

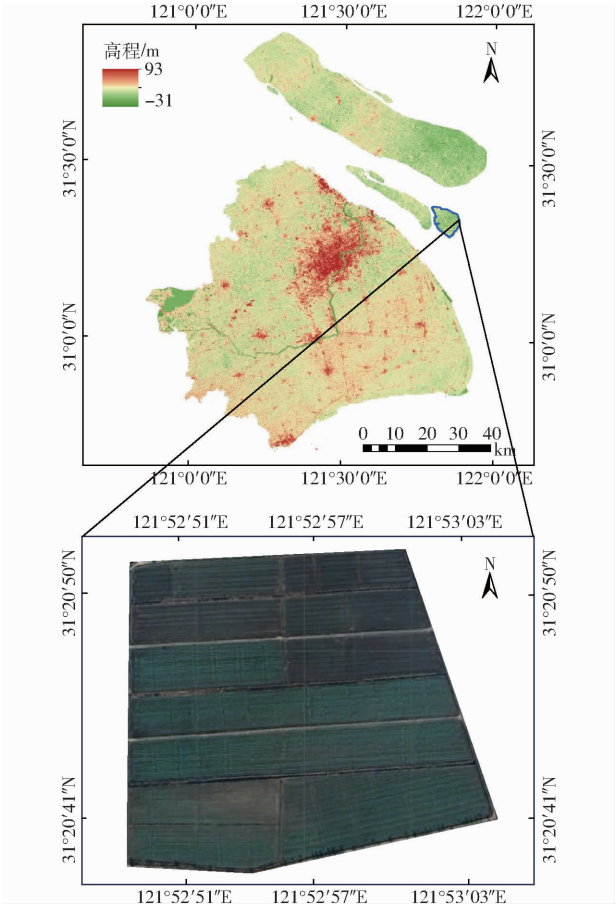


图 1 研究区域
Fig. 1 Study area

水状态。每个田块内的采样点间隔 5 m × 5 m, 采集深度 0 ~ 20 cm 的原状土壤共 194 份。立即将每个样本取出约 35 g 装入洗净干燥的铝盒中, 土壤表面刮平后使用 ASD FieldSpec 4 地物光谱仪及其接触式探头测量土壤样本的高光谱^[19], 每 10 min 进行一次白板校正, 每个样品在 3 个不同位置进行光谱测量, 取 3 个不同位置测量的反射率均值作为一个土壤样本最终的反射率光谱数据。光谱波长范围为 350 ~ 2500 nm, 步长为 1 nm。为减少仪器自身引起的噪声, 删除了光谱两端的噪声波段, 最终使用波长 450 ~ 2 400 nm 的光谱数据, 原始光谱图如图 2 所示。使用烘干法测定土壤质量含水率, 测得稻田土壤样本质量含水率变化范围为 26. 33% ~ 61. 09%。

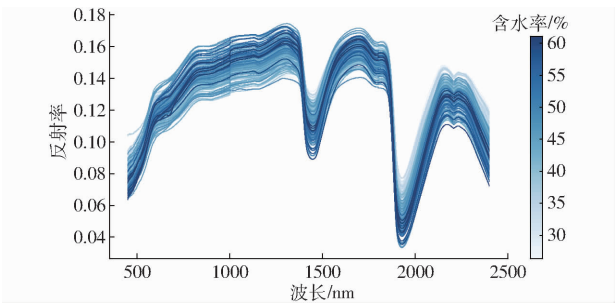


图 2 原始光谱图
Fig. 2 Original spectrum

2 研究方法

基于 KM 理论构建土壤二向通量辐射传输模型, 结合稻田土壤高光谱, 计算吸收与散射系数, 通过 2 种途径提升 SMC 反演精度。第 1 种基于 KM 理论构建 SMC 半经验模型, 第 2 种基于 KM 理论重建模拟光谱, 进行数据增强后建立 SMC 反演模型, 整体技术路线如图 3 所示。

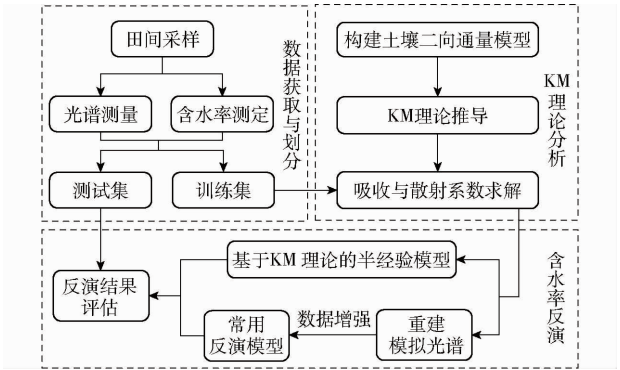


图 3 技术路线
Fig. 3 Technical route

2.1 Kubelka – Munk 理论

KM 理论用于描述光在具有散射与吸收特性的微粒体系中的光学行为^[20], 因此能够描述光线在土壤内部发生吸收效应和散射效应的过程。吸收效应消耗光, 而散射效应改变光的传播方向, 未被吸收的光经过散射最终在土壤表面形成了漫反射。采用一个二向通量辐射传输模型表示此过程^[18, 21], 如图 4 所示, 竖直向下定义为正方向, 设土壤厚度为 H , 光线在土壤中穿透的距离为 z 。入射通量为 I , 其吸收效应和散射效应分别为 K_I 、 S_I 。反射通量为 J , 其吸收效应和散射效应分别为 K_J 、 S_J 。

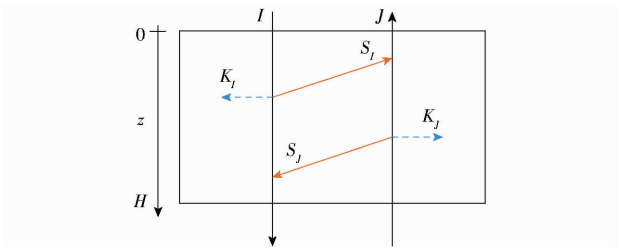


图 4 土壤二向通量 KM 模型
Fig. 4 KM model of soil two-flux

根据该模型, 得到微分方程组为

$$\begin{cases} dI = -(K_I + S_I) dz + S_J dz \\ dJ = (K_J + S_J) dz - S_I dz \end{cases} \quad (1)$$

式中 dz ——光线穿透的厚度微元
 dI ——入射通量关于厚度的微分
 dJ ——反射通量关于厚度的微分

土壤漫反射率 R 即为表面处的反射通量与入射通量之比, 表示为

$$R = \frac{J(z=0)}{I(z=0)} \tag{2}$$

当光源稳定时, $I(z=0)$ 为定值, 随着土壤厚度逐渐增加, 能够透过的光线通量几乎为 0, 反射通量趋于稳定, 则 $J(z=0)$ 也趋于稳定, 结合式(1)和式(2)得到土壤漫反射率计算式为

$$R = 1 + \frac{K}{S} - \sqrt{\left(\frac{K}{S}\right)^2 + 2\frac{K}{S}} \tag{3}$$

式中 K ——吸收系数
 S ——散射系数

同样也能够转换为反射率表示吸收散射比, 即 Kubelka - Munk 公式, 表示为

$$\frac{K}{S} = \frac{(1-R)^2}{2R} \tag{4}$$

已有研究证实, 土壤的反射光谱信息主要源自表层土壤, 仅有极少部分光线可穿透深度超过 1 mm 的土层^[22]。在本研究中测量土壤样品光谱时, 装入铝盒中的土壤厚度约为 2.5 cm, 满足透过的光线通量几乎为零的条件。由于土壤表面微观结构凹凸不平, 且测量过程中接触式探头与土壤表面紧密贴合, 致使大部分空气介质被排出, 此时菲涅尔反射率数值极小^[23], 整体反射率主要由吸收效应和散射效应来决定, 因此可用式(3)表示当前情形下的土壤反射率。

在运用 KM 理论时, 通常将不同物质的吸收与散射系数按含量加权线性叠加来表示整体的吸收与散射系数^[17]。在研究 SMC 相关问题中, 可将样品视为土壤固体和水分 2 种物质成分, 样本整体的吸收和散射系数表示为

$$\begin{cases} K = k_1(1-\theta) + k_2\theta \\ S = s_1(1-\theta) + s_2\theta \end{cases} \tag{5}$$

式中 k_1 ——土壤固体吸收系数
 s_1 ——土壤固体散射系数
 k_2 ——水分吸收系数
 s_2 ——水分散射系数
 θ ——土壤含水率

2.2 系数求解方法

结合式(3)和式(5), 在已知稻田土壤反射率光谱和含水率的条件下, 基于训练集数据求解系数 k_1 、 k_2 、 s_1 、 s_2 , 以分析土壤固体和水分逐波长的吸收与散射效应。本研究采用遗传算法 (Genetic algorithm, GA) 计算 k_1 、 k_2 、 s_1 、 s_2 的初值, 种群规模为 100, 迭代次数为 500。采用自适应矩估计 (Adaptive moment estimation, Adam) 对遗传算法计算的初值进一步优化, 学习率为 5×10^{-4} , 损失函数 L 设计为

$$L = L_{\text{RMSE}} + \lambda_{\text{shape}} L_{\text{shape}} + \lambda_{\text{amp}} L_{\text{amp}} + \lambda_{\text{phy}} L_{\text{phy}} \tag{6}$$

式中 L_{RMSE} ——均方根误差项
 L_{shape} ——形状平滑项
 L_{amp} ——方差匹配项
 L_{phy} ——物理合理项
 λ_{shape} 、 λ_{amp} 、 λ_{phy} ——权重

L_{RMSE} 作为主要优化项, 用于直接衡量模拟值与真实值之间的整体偏差。 L_{shape} 项旨在保持光谱形状的平滑性, 其构成包括光谱角匹配、一阶导数差异、二阶导数差异以及平滑约束。 L_{amp} 使模拟光谱在样本间的幅值方差逼近实测光谱的方差分布。 L_{phy} 项用于确保模拟结果的物理合理性, 具体约束包括吸收系数与散射系数为非负值, 以及相关物理表达式中根号内取值的非负性。

2.3 光谱预处理方法

原始光谱受到环境、仪器等因素的影响, 导致光谱中存在噪声、基线漂移等干扰, 因此需要进行光谱预处理。 Savitzky - Golay 滤波^[4]通过局部多项式拟合, 抑制噪声干扰, 保留光谱轮廓与吸收特征的细节信息; 一阶导数能够降低基线背景的影响、突出光谱曲线的细微拐点, 从而凸显与土壤成分相关的特征波段。

2.4 特征提取方法

光谱存在高维度、共线性的特点, 特征提取能够消除冗余信息并提升模型效率与泛化能力。 本研究采用的特征提取方法包括连续投影算法 (Successive projections algorithm, SPA)、竞争性自适应重加权采样 (Competitive adaptive reweighted sampling, CARS) 和变量重要性投影 (Variable importance in projection, VIP)^[6]。

2.5 基于 KM 理论的半经验模型

反射率光谱中吸收效应与散射效应相互耦合, 通过式(4)将反射率映射为具有明确物理意义的光学参数, 即吸收散射比, 从而凸显水分主导的吸收特征。 采用最小二乘 (Ordinary least squares, OLS) 构建吸收散射比与 SMC 之间的线性关系, 表示为

$$M_{\text{Linear}} = \sum_{i=1}^n \partial_i \left(\frac{K}{S} \right)_i + \beta \tag{7}$$

式中 M_{Linear} ——线性模型输出
 n ——波长数量
 ∂_i ——自变量系数
 β ——截距

在线性模型基础上增加一个残差补偿模型, 以刻画并补偿未被线性部分拟合的非线性关系^[24]。 残差补偿模型的预测目标为线性模型输出与实测 SMC 之间的差值, 采用多层感知机 (Multi-layer perceptron, MLP) 作为基础架构。

残差补偿模型的输入特征包括吸收散射比 $\left(\frac{K}{S}\right)_i$ 、线性模型输出 M_{Linear} 及其在各波长处的贡献度 $\partial_i\left(\frac{K}{S}\right)_i$ 。通过纳入线性模型的预测结果以及内部信息,能够更有效地修正其系统性偏差,拟合非线性关系。MLP 训练轮数为 1 000,学习率为 1.75×10^{-4} ,激活函数为 ReLU,并加入早停机机制防止模型过拟合。基于 KM 理论的半经验模型(KM – OLS – MLP)结构如图 5 所示,可表示为

$$M = \sum_{i=1}^n \partial_i \left(\frac{K}{S}\right)_i + \beta + C_{\text{Residual}} \tag{8}$$

式中 M ——半经验模型输出
 C_{Residual} ——残差补偿模型输出

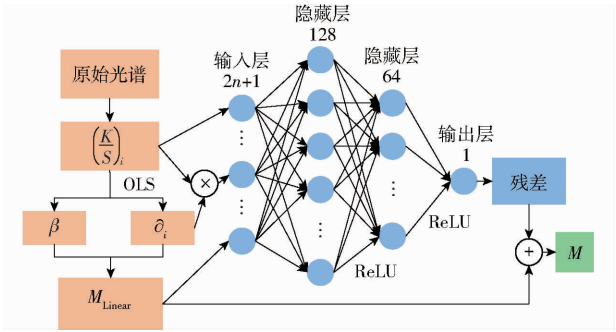


图 5 半经验模型结构图

Fig. 5 Semi-empirical model structure

2.6 基于 KM 理论的数据增强

本研究采用的训练集为中小规模数据集,故引入数据增强 (Data augmentation, DA) 以提高模型的反演精度^[25]。以训练集的 SMC 数据作为标签,吸收系数和散射系数作为约束,依据式(3)和式(5),重建满足 KM 理论方程且物理一致性高的模拟光谱,将其加入训练集以扩充样本数量,实现数据增强 (KM – DA)。在测试集内采用同样的方式重建模拟光谱,以评估重建质量以及验证系数求解结果的准确性。

分别生成与原始训练集样本数量比例为 1:1、2:1、3:1 及 4:1 的模拟光谱。由于求解所得的系数仅与波长有关,同一含水率下重建的模拟光谱相同。因此在模拟光谱中叠加标准差为 0.000 1 的随机高斯噪声,该噪声强度相对于光谱反射率数值极小,可在不影响模拟光谱质量的前提下增加样本多样性。随后,将模拟光谱与实测光谱合并,构成增强训练集。使用反射率光谱建立 SMC 反演模型,对比增强前后的模型表现。采用的回归模型包括偏最小二乘回归 (Partial least squares regression, PLSR)、支持向量回归 (Support vector regression, SVR) 以及 BP 神经网络 (Back propagation neural network,

BPNN)^[12],模型内部的超参数均采用网格搜索的方式调优。

2.7 样本划分与评估指标

从所有样本中随机选取 155 份作为训练集样本,剩余 39 份作为测试集样本^[26],含水率统计结果如表 1 所示。训练集与测试集样本的最小值、最大值、平均值、标准差以及变异系数基本一致。进一步采用 Levene 检验^[27]验证方差齐性 ($D = 1.946\ 9, p = 0.164\ 5 > 0.05$),表明训练集与测试集 SMC 离散程度无显著差异。采用 Kolmogorov – Smirnov 检验验证整体分布一致性 ($D = 0.151\ 4, p = 0.422\ 4 > 0.05$),表明训练集与测试集 SMC 服从同一整体分布,数据集划分合理且具有代表性,可用于后续模型训练与评估。

表 1 稻田土壤含水率描述性统计
Tab. 1 Descriptive statistics of soil moisture

in rice fields					%
数据集	最小值	最大值	平均值	标准差	变异系数
全集	26.33	61.09	41.15	7.48	18.18
训练集	26.33	60.74	40.81	7.09	17.36
测试集	27.92	61.09	42.50	8.74	20.56

本研究模型评估包含 2 类:SMC 反演模型精度评估和模拟光谱的重建质量评估。SMC 反演模型的精度评估指标采用决定系数 (Coefficient of determination, R^2) 与均方根误差 (Root mean square error, RMSE)。模拟光谱的重建质量评估指标采用 R^2 、RMSE、峰值信噪比 (Peak signal-to-noise ratio, PSNR)、结构相似性指数 (Structural similarity index, SSIM) 以及光谱角制图法 (Spectral angle mapper, SAM)。各指标的计算公式参考文献[15,28]。在评估中, R^2 与 SSIM 值越接近 1、PSNR 值越高、RMSE 与 SAM 值越低,表明评估结果越优。

3 结果与分析

3.1 系数求解结果

将原始反射率转换为吸收散射比,如图 6 所示,进一步将土壤视为固体与水分,求解出土壤固体与水分的吸收系数与散射系数曲线如图 7a、7b 所示。由各曲线对比可知,吸收系数普遍高于散射系数,表明可见-近红外光在土壤中的吸收作用强于散射作用。具体而言,土壤固体的散射曲线变化较为平缓;其吸收曲线在波长 700 nm 与 2 300 nm 附近呈现明显峰值,在波长 1 500 nm 及 1 900 nm 附近则出现明显谷值。水分的散射曲线在波长 1 400 ~ 1 500 nm 与 1 900 ~ 2 100 nm 范围内出现两处显著谷值,表明水分在上述波段的散射效应较弱。与此同时,水分

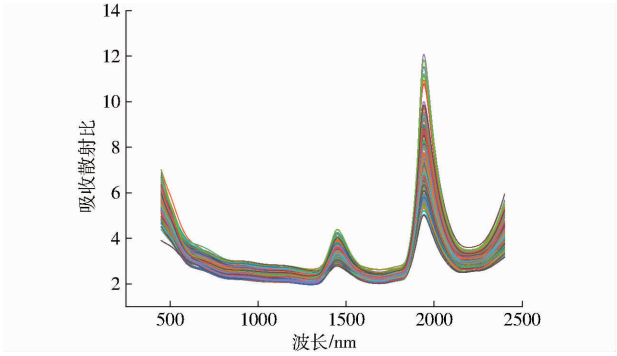


图6 吸收散射比

Fig. 6 Absorption-to-scattering ratio

的吸收曲线在波长 1 900 ~ 2 100 nm 范围内呈现突出的峰值。

为更直观地分析水分吸收与散射效应的综合影响,本研究进一步绘制了水分吸收散射比曲线,如图 7c 所示。该曲线在波长 1 420 ~ 1 470 nm 和 1 850 ~ 2 050 nm 两段区间内呈现出明显峰值,且与整体吸收散射比曲线趋势相似,表明在这些波段内水分的吸收效应与散射效应之间差距显著,对土壤整体反射率的影响最为突出,因而可被视为水分的敏感波段。水分敏感波段广泛被认为位于波长 1 450 nm

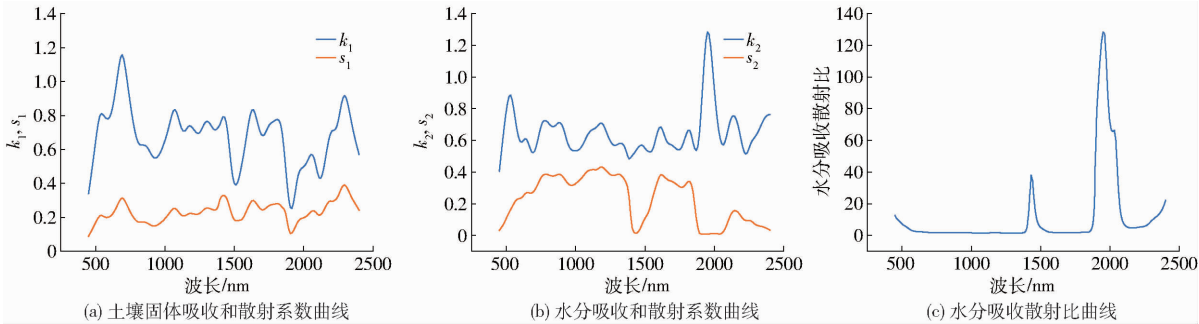


图7 系数求解结果

Fig. 7 Coefficient solution results

与 1 950 nm 附近^[29],与本研究得出的敏感波段相一致,验证了 KM 理论应用于稻田土壤高光谱的有效性。

3.2 半经验模型反演结果

使用由 KM 理论计算的吸收散射比作为模型的输入特征,由于输入数据存在多重共线性,为规避 OLS 参数估计不稳定进而导致模型出现显著过拟合现象,初步在由吸收散射比得出的 2 个水分敏感波段内,采用等间隔的方式选取 60 个波长点作为建模的输入波长。基于 OLS 线性模型的可加性,采用回归系数与吸收散射比乘积的绝对值表征波长对预测结果的贡献程度。不同含水率样本的吸收散射比在尺度上存在差异,为更稳健地评估各波长在预测中的重要性,将每个样本中各波长点的贡献度进行归一化(即相对于该样本的总贡献度计算其占比),随后对训练集所有样本的归一化贡献占比取平均值,得到各波长的平均贡献占比。剔除平均贡献占比低于 0.002 的波长点,并综合考虑模型在训练集上的拟合表现,最终在 2 个水分敏感波段内选择了 38 个波长点。图 8 展示了这些波长点在训练集中的平均贡献占比,其中橙色标记表示贡献占比最高的 10 个波长点。

基于 KM 理论的半经验模型反演结果如表 2 所示,该模型在 KM 理论得到的水分敏感波段中根据贡献占比选择波长的过程相当于特征提取,因此表

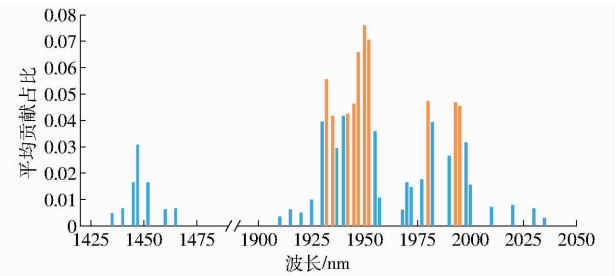


图8 平均贡献占比(训练集)

Fig. 8 Average contribution share (training set)

中列出了传统特征提取方法与回归模型组合,并采用反射率光谱进行建模的结果。图 9 展示了不同特征提取方法所选择的波长,结果表明,SPA、VIP 和 CARS 方法所提取的部分波长不属于水分敏感波段,且特征波长在位置上存在显著差异。传统回归模型中,CARS 特征提取结合 SVR 模型效果最优,测试集 R^2 为 0.869 5, RMSE 为 3.16%。半经验模型仅用 OLS 拟合线性关系时,测试集 R^2 为 0.892 7, RMSE 为 2.86%,加入 MLP 补充非线性关系后,模型效果最优,测试集 R^2 为 0.924 1, RMSE 为 2.41%。为了直观地反映模型对高含水率样本的预测性能,计算了测试集内高含水率样本(实测含水率高于 50%)的预测 RMSE,并按不同模型绘制了折线图,如图 10 所示。传统回归模型对高含水率样本的预测 RMSE 均较高,而 KM-OLS 对高含水率样本的预测 RMSE 显著低于传统回归模型,表明 KM 理

论有效提高了对高含水率样本的预测性能,加入 MLP 后预测性能进一步提高。

表 2 反演精度
Tab.2 Inversion accuracy

方法	波长数	训练集		测试集	
		R^2	RMSE/%	R^2	RMSE/%
PLSR – SPA	40	0.880 8	2.45	0.853 0	3.35
PLSR – VIP	120	0.812 2	3.07	0.816 1	3.75
PLSR – CARS	559	0.874 7	2.51	0.855 1	3.33
SVR – SPA	40	0.869 7	2.56	0.838 8	3.51
SVR – VIP	120	0.857 9	2.67	0.839 8	3.50
SVR – CARS	559	0.914 9	2.07	0.869 5	3.16
BPNN – SPA	40	0.930 6	1.87	0.830 2	3.60
BPNN – VIP	120	0.863 2	2.62	0.811 1	3.80
BPNN – CARS	559	0.822 5	2.99	0.805 3	3.86
KM – OLS	38	0.896 3	2.28	0.892 7	2.86
KM – OLS – MLP	38	0.947 4	1.62	0.924 1	2.41

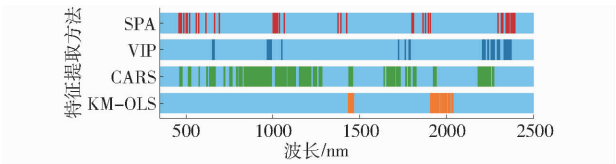


图 9 不同特征提取方法所选波长

Fig. 9 Wavelengths selected by different feature extraction methods

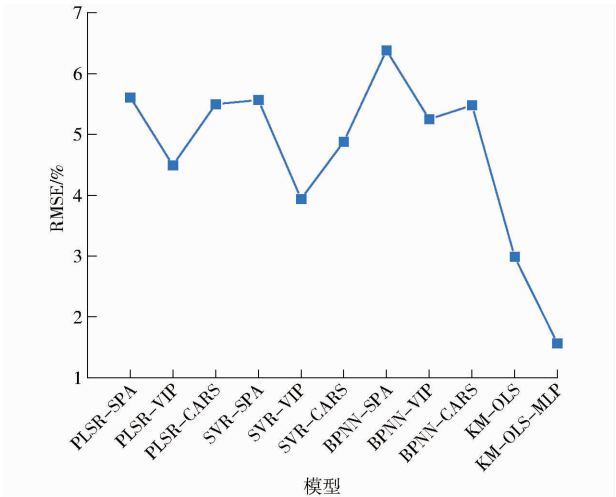


图 10 不同模型对高含水率样本的预测 RMSE

Fig. 10 Prediction RMSE of different models for high water content samples

3.3 数据增强后的反演结果

3.3.1 模拟光谱重建结果

图 11 展示了测试集中重建的模拟光谱与实测光谱对比图,分别呈现了经遗传算法初步求解系数后重建的模拟光谱,以及进一步采用 Adam 优化器结合损失函数优化系数后重建的结果。对比显示,所构建的损失函数有效地约束了参数求解过程,从而提高了模拟光谱的重建质量。最终模拟光谱评估

结果如表 3 所示,测试集中 R^2 为 0.912 2,表明模拟光谱整体拟合效果较好。具体而言, RMSE 为 0.008 1、PSNR 为 27.540 7,表明模拟光谱在幅值上拟合精度较高, SSIM 为 0.996 4、SAM 为 0.031 6,表明结构和形状上与实测光谱相似性较高。采用 t – SNE 方法对测试集光谱数据进行降维与可视化,在三维空间中分别构建实测光谱与模拟光谱的 95% 置信椭圆,用以表征样本分布情况,结果如图 12 所示,2 类椭圆重合度高,表明模拟光谱的分布特征与实测光谱具有良好的一致性。上述结果表明, KM 理论能够重建出符合物理机理的模拟光谱,同时验证了所求吸收系数和散射系数的准确性。

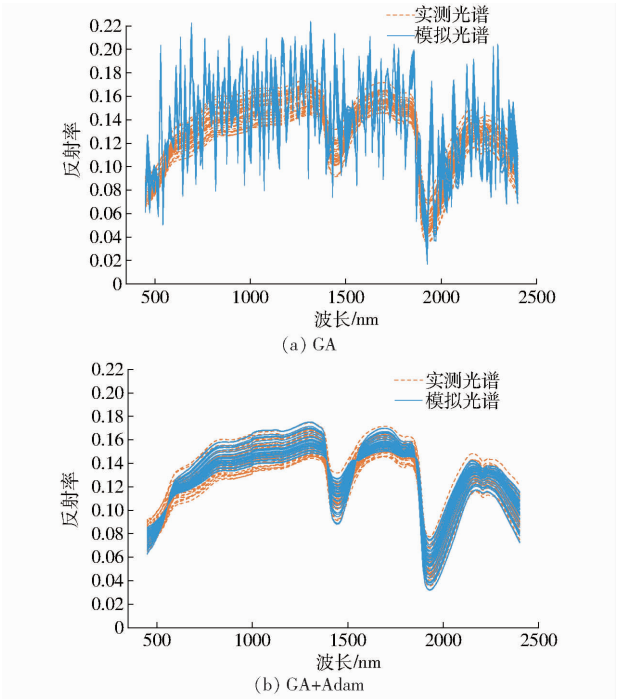


图 11 模拟光谱与实测光谱对比

Fig. 11 Comparison of simulated spectra and measured spectra

表 3 模拟光谱评估结果

Tab.3 Simulated spectral evaluation results

数据集	R^2	RMSE	PSNR	SSIM	SAM
训练集	0.927 4	0.007 2	28.757 2	0.997 1	0.029 9
测试集	0.912 2	0.008 1	27.540 7	0.996 4	0.031 6

3.3.2 模型反演结果

分别将按比例 1:1、2:1、3:1 和 4:1 重建的模拟光谱加入训练集进行数据增强 (KM – DA), 有效扩充了训练样本的数量和多样性。使用全波段反射率建立含水率反演模型, 并与仅用真实样本训练的模型反演精度进行对比, 结果如表 4 所示。在训练样本数量增强 1 至 4 倍的条件下, 3 种 SMC 反演模型的建模性能均优于仅用真实样本训练时的表现。同时, 训练集与测试集 R^2 之间的差距缩小, 表明模型

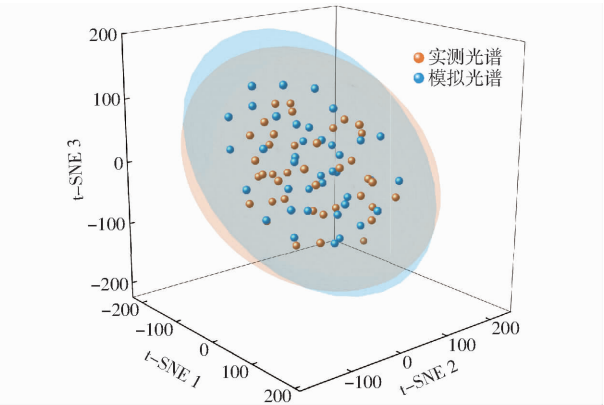


图 12 t-SNE 可视化三维图

Fig. 12 t-SNE three-dimensional visualization

的泛化能力得到提升。PLSR 模型在 3:1 时效果最优,测试集 R^2 从 0.830 2 提高至 0.881 0, RMSE 从 3.60% 降至 3.02%;SVR 模型在 3:1 时效果最优,测试集 R^2 从 0.838 5 提高至 0.893 4, RMSE 从 3.51% 降至 2.85%;BPNN 模型在 2:1 时效果最优,测试集 R^2 从 0.790 6 提高至 0.898 7, RMSE 从 4.00% 降至 2.78%。不同建模方法对应的最佳增强样本数量不同,随着增强样本数量的增多,SMC 反演模型的精度提高趋于饱和。结果表明,由 KM 理论重建的模拟光谱不仅具备良好物理可解释性,还能够作为数据增强的样本提高稻田 SMC 反演模型的精度。

表 4 数据增强前后的反演精度对比

Tab. 4 Comparison of modeling accuracy before and after data augmentation					
方法	增强 样本数	训练集		测试集	
		R^2	RMSE/%	R^2	RMSE/%
PLSR	0	0.963 2	1.36	0.830 2	3.60
	155	0.935 2	1.80	0.872 9	3.12
	310	0.931 4	1.86	0.869 1	3.16
	465	0.922 3	1.97	0.881 0	3.02
	620	0.913 1	2.09	0.872 0	3.13
SVR	0	0.966 5	1.30	0.838 5	3.51
	155	0.965 7	1.31	0.858 1	3.29
	310	0.934 5	1.81	0.871 1	3.14
	465	0.926 0	1.93	0.893 4	2.85
	620	0.908 2	2.15	0.891 6	2.88
BPNN	0	0.910 4	2.12	0.790 6	4.00
	155	0.902 2	2.22	0.856 3	3.31
	310	0.937 9	1.77	0.898 7	2.78
	465	0.871 8	2.54	0.883 7	2.98
	620	0.898 6	2.26	0.871 5	3.13

4 讨论

高光谱技术已成为 SMC 这一关键农田指标的重要检测方法,然而针对稻田此类含水率变化范围较大的土壤类型,相关研究仍较为匮乏,且常用的方

法均为纯数据驱动,缺乏物理理论的支撑,可解释性不足。而 KM 理论作为辐射传输理论,通过描述光在介质中的传输过程,能够揭示光谱响应背后的机理,可用于土壤光谱分析。

本研究基于稻田土壤高光谱和 KM 理论,构建土壤二向通量辐射传输模型,并将土壤视为固体和水分两部分,经过公式推导与系数求解,根据吸收散射比曲线识别出了波长 1 450 nm 和 1 900 nm 附近的水分敏感波段。水的近红外吸收主要来源于 O—H 振动的泛音和组合带,从分子光谱学角度看,这对应于非简谐振动势阱中的高阶能级跃迁,因而会出现泛音跃迁以及不同振动模式之间的组合跃迁。KM 理论则将土壤反射率视为吸收效应与散射效应耦合后的有效光学响应,因此水分引起的 O—H 振动吸收变化主要体现在吸收系数和吸收散射比中。具体而言,水分在波长 1 450 nm 附近对应 O—H 伸缩振动第一泛音^[30],在波长 1 900 nm 附近对应 O—H 伸缩与弯曲振动的组合带,液态水中由于氢键网络的存在,不同氢键环境会使这些吸收带发生展宽、位移并相互叠加^[31],从而在波长 1 450 nm 和 1 900 nm 附近形成明显的水分吸收特征。因此,水分引起的 O—H 振动吸收变化,使得吸收系数和吸收散射比在波长 1 450 nm 和 1 900 nm 附近出现局部增强,而颗粒尺度、孔隙结构和界面粗糙度等散射效应则被相对削弱。该现象反映了水分子 O—H 振动吸收在土壤多相散射介质中的有效表征,表明 KM 系数能够从物理机理层面提取并凸显水分的本征光学响应。

基于 KM 理论的半经验模型将反射率映射为具有物理意义的吸收散射比作为输入特征,而水分主要通过吸收效应来影响土壤反射率,因此这一转换使得输入特征更接近光传输过程中介质的吸收贡献。采用 OLS 拟合吸收散射比与含水率的线性关系,模型效果优于传统方法,同时模型结构简洁、可解释性高,证明了吸收散射比反演 SMC 的可行性。采用 MLP 建立残差补偿模型与 OLS 组合后,模型效果进一步提升,尤其是对高含水率样本的反演精度,凸显了组合建模的优势。其原因在于,土壤整体反射率随含水率增加通常呈下降趋势,但当含水率超过田间持水量时,反射率随含水率增加反而上升^[11],呈现出明显的非线性变化特征。该非线性关系在含水率宽幅变化的稻田土壤中尤为突出。残差补偿模型通过引入线性模型的预测结果及其内部信息,能够更有效地捕捉该非线性模式,从而进一步提高了稻田 SMC 的反演精度。

基于 KM 理论的数据增强方法中,在吸收系数

和散射系数的约束下,重建的模拟光谱在数值、形状以及分布上与实测光谱高度吻合,物理一致性高。使用不同数量的模拟光谱加入训练集实现数据增强,并构建 SMC 反演模型,相对于仅用真实样本训练模型的精度均有提高。在实际采样过程中,难以保证 SMC 的均匀分布,该问题在含水率宽幅变化的稻田土壤中尤为突出,导致部分含水率区间(尤其是高含水率)的样本稀缺,进而对模型的训练效果产生不利影响。引入模拟光谱进行数据增强后,不仅增加了训练样本的总数量,而且有效补充了稀缺样本,从而提升了 SMC 反演模型精度和泛化能力。

目前应用于土壤含水率反演的辐射传输模型还包括多层辐射传输模型(MARMIT)和 Hapke 模型等。MARMIT 模型^[32]将湿润土壤视为干土覆盖上薄水膜,能够较好地刻画表层水分对土壤反射率的影响,以及模拟湿润土壤的反射率,但使用该模型反演含水率不仅要获取湿润光谱,还需要将土壤经过实验室处理后得到干燥光谱作为参考基准。Hapke 模型^[33]能够综合考虑单次、多次散射以及观测几何效应,但反演含水率时需要多角度的湿润光谱、对应的干燥光谱,以及观测几何参数。以上 2 种模型,对输入数据的要求较高,模型适用条件严格。相比之下,本研究基于 KM 理论,仅需湿润土壤高光谱与含水率即可建立模型,进而模拟稻田土壤光谱、提取水分

敏感波段,在含水率反演精度较高且遵循物理机制的同时,具有输入数据易获取、实用性较强的优势。

然而,基于 KM 理论的稻田土壤含水率高光谱反演研究仍然存在以下局限,光谱预处理方式较单一,仅对深 0 ~ 20 cm 的稻田土壤 SMC 进行建模,未涉及更深层的土壤。此外,仅在特定实验区域内反演精度较高,未来应该在不同区域和不同类型的土壤上进一步验证其普适性,以及将反演指标扩展至有机质、总氮等其他土壤关键指标。

5 结论

(1) 基于 KM 理论,将反映物理效应的吸收与散射系数作为半经验模型的输入特征和数据增强中的物理约束,通过 2 种途径提高了稻田土壤 SMC 高光谱反演精度。

(2) 基于 KM 理论的半经验模型(KM – OLS – MLP)通过 OLS 建立 SMC 与吸收散射比的线性关系,进一步采用 MLP 补偿非线性部分。测试集 R^2 为 0.924 1, RMSE 为 2.41%。

(3) 基于所求得的吸收与散射系数重建物理一致性高的模拟光谱,将其加入训练集实现数据增强(KM – DA),提高了稻田 SMC 反演模型的精度。BPNN 模型精度提高最显著,模拟光谱与实测光谱比例为 2:1 时,测试集 R^2 从 0.790 6 提高至 0.898 7, RMSE 从 4.00% 降低至 2.78%。

参 考 文 献

[1] 靳亚红,吴鑫森,甄文超,等. 基于采样点光谱信息窗口尺度优化的土壤含水率无人机多光谱遥感反演[J]. 农业机械学报,2024,55(1):316 – 327.
Jin Yahong, Wu Xinmiao, Zhen Wenchao, et al. UAV multispectral remote sensing inversion of soil moisture content based on window size optimization of spectral information at sampling points[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(1): 316 – 327. (in Chinese)

[2] 和玉璞,万家玮,王荣勇,等. 灌排调控下稻田水分侧渗过程规律分析[J]. 农业机械学报,2025,56(2):463 – 473.
He Yupu, Wan Jiawei, Wang Rongyong, et al. Analysis of lateral seepage in paddy field with controlled irrigation and drainage regulation[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(2): 463 – 473. (in Chinese)

[3] 熊静玲,朱西存,高华光,等. 基于 MSC 与 SVM 的夯土齐长城土壤含水率高光谱估测[J]. 土壤学报,2018, 55(6): 1336 – 1344.
Xiong Jingling, Zhu Xicun, Gao Huaguang, et al. Hyperspectral estimation of soil moisture content in rammed soil of Qi – Dynasty Great Wall based on MSC and SVM[J]. Acta Pedologica Sinica, 2018, 55(6): 1336 – 1344. (in Chinese)

[4] 刘浩,杨锡震,张蓓,等. 基于分数阶微分光谱指数的冬小麦根域土壤含水率估算模型[J]. 农业工程学报,2023,39(13): 131 – 140.
Liu Hao, Yang Xizhen, Zhang Bei, et al. Estimation model of soil moisture content in root domain of winter wheat using a fractional-order differential spectral index[J]. Transactions of the CSAE, 2023, 39(13): 131 – 140. (in Chinese)

[5] 唐子竣,张威,向友珍,等. 基于高光谱和机器学习模型的冬小麦土壤含水率监测研究[J]. 农业机械学报,2023,54(12): 350 – 358.
Tang Zijun, Zhang Wei, Xiang Youzhen, et al. Monitoring of soil moisture content of winter wheat based on hyperspectral and machine learning models[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(12): 350 – 358. (in Chinese)

[6] 杨文举,李奋华,康德奎,等. 基于敏感变量筛选的多光谱土壤含水率反演模型研究[J]. 节水灌溉,2025(6):45 – 54.
Yang Wenju, Li Fenhua, Kang Dekui, et al. Study on multispectral soil moisture inversion model based on sensitive variable

- screening[J]. Water Saving Irrigation, 2025(6):45–54. (in Chinese)
- [7] 李志军,陈国夫,支佳伟,等. 融合无人机光谱信息与纹理特征的大豆土壤含水率估测模型研究[J]. 农业机械学报, 2024,55(9):347–357.
- Li Zhijun, Chen Guofu, Zhi Jiawei, et al. Estimation model of soybean soil moisture content based on UAV spectral information and texture features[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024,55(9):347–357. (in Chinese)
- [8] 杨曦光,于颖. 基于试验反射光谱数据的土壤含水率遥感反演[J]. 农业工程学报, 2017,33(22):195–199.
- Yang Xiguang, Yu Ying. Remote sensing inversion of soil moisture based on laboratory spectral reflectance data[J]. Transactions of the CSAE, 2017, 33(22):195–199. (in Chinese)
- [9] Zhong J, Song Y, Yang M, et al. Strong N_2O uptake capacity of paddy soil under different water conditions[J]. Agricultural Water Management, 2023, 278:108146.
- [10] Perry H, Carrijo D, Linnquist B. Single midseason drainage events decrease global warming potential without sacrificing grain yield in flooded rice systems[J]. Field Crops Research, 2022, 276:108312.
- [11] 司海青,姚艳敏,王德营,等. 含水率对土壤有机质含量高光谱估算的影响[J]. 农业工程学报, 2015,31(9):114–120.
- Si Haiqing, Yao Yanmin, Wang Deying, et al. Hyperspectral prediction of soil organic matter contents under different soil moisture contents[J]. Transactions of the CSAE, 2015, 31(9):114–120. (in Chinese)
- [12] 李鑫星,梁步稳,白雪冰,等. 光谱技术在土壤水分含量检测中的研究进展[J]. 光谱学与光谱分析, 2020,40(12):3705–3710.
- Li Xinxing, Liang Buwen, Bai Xuebing, et al. Research progress of spectroscopy in the detection of soil moisture content[J]. Spectroscopy and Spectral Analysis, 2020, 40(12):3705–3710. (in Chinese)
- [13] Shin S K, Lee S J, Park J H. Prediction of soil properties using VIS-NIR spectroscopy combined with machine learning: a review [J]. Sensors, 2025, 25(16):5045.
- [14] 王勇,侯晨悦,杨锡震,等. 基于多种高光谱指标反演冻结土壤含水率的研究[J]. 节水灌溉, 2023(7):10–19.
- Wang Yong, Hou Chenyue, Yang Xizhen, et al. Inverting frozen soil moisture content based on various hyperspectral indexes [J]. Water Saving Irrigation, 2023(7):10–19. (in Chinese)
- [15] 刘凯,王玉峰,彭志青,等. 基于机器学习的高光谱土壤养分特征波段提取方法[J]. 光学学报, 2025, 45(6):222–231.
- Liu Kai, Wang Yufeng, Peng Zhiqing, et al. Machine learning for hyperspectral characteristic band extraction in soil nutrient analysis[J]. Acta Optica Sinica, 2025, 45(6):222–231. (in Chinese)
- [16] Milosevic M, Berets S L. A review of FT-IR diffuse reflection sampling considerations[J]. Applied Spectroscopy Reviews, 2002, 37(4):347–364.
- [17] 袁静,王鑫,颜昌翔. 不同含水量黑土土壤光谱反射率半经验模型构建[J]. 光谱学与光谱分析, 2019, 39(11):3514–3518.
- Yuan Jing, Wang Xin, Yan Changxiang. A semi-empirical model for reflectance spectral of black soil with different moisture contents[J]. Spectroscopy and Spectral Analysis, 2019, 39(11):3514–3518. (in Chinese)
- [18] Wu F, Tan K, Wang X, et al. A semi-analytical radiative transfer model for explaining soil spectral features[J]. International Journal of Applied Earth Observation and Geoinformation, 2023, 118:103250.
- [19] Lazaar A, Mouazen A M, Hammouti K E L, et al. The application of proximal visible and near-infrared spectroscopy to estimate soil organic matter on the Triffa Plain of Morocco[J]. International Soil and Water Conservation Research, 2020, 8(2):195–204.
- [20] Alcaraz de la Osa R, Iparraguirre I, Ortiz D, et al. The extended Kubelka–Munk theory and its application to spectroscopy[J]. ChemTexts, 2019, 6(1):2.
- [21] Mohseni M, Bastani S, Allonas X, et al. Effect of scattering nanoparticles on the curing behavior and conversion gradient of UV-curable turbid systems: two-flux Kubelka–Munk approach[J]. Progress in Organic Coatings, 2018, 115: 65–73.
- [22] Norouzi S, Sadeghi M, Liaghat A, et al. Information depth of NIR/SWIR soil reflectance spectroscopy[J]. Remote Sensing of Environment, 2021, 256:112315.
- [23] 温颜沙,邓建明,毛志华,等. 水面菲涅尔反射对 ~761 nm 叶绿素荧光估算的影响[J]. 遥感学报, 2018, 22(3):424–431.
- Wen Yansha, Deng Jianming, Mao Zhihua, et al. Effect of the fresnel reflection of a water surface on chlorophyll fluorescence line height at approximately 761 nm[J]. Journal of Remote Sensing, 2018, 22(3):424–431. (in Chinese)
- [24] 李诗文,胡梦云,李鲁宁,等. 基于飞秒等离子体光栅诱导击穿光谱与 ARHNet 的土壤铬元素定量分析(特邀)[J]. 中国激光, 2025, 52(19):191–199.
- Li Shiwen, Hu Mengyun, Li Luning, et al. Femtosecond plasma grating-induced breakdown spectroscopy combined with ARHNet for quantitative analysis of chromium in soil (invited) [J]. Chinese Journal of Lasers, 2025, 52(19):191–199. (in Chinese)
- [25] 唐磊,茅晔辉,蔡婧,等. 光谱数据增强方法及其应用进展[J]. 分析测试学报, 2025, 44(6):1227–1236.
- Tang Lei, Mao Yehui, Cai Jing, et al. Spectral data augmentation methods and their application progress [J]. Journal of Instrumental Analysis, 2025, 44(6):1227–1236. (in Chinese)

[19]倪美玉,杜子涛,曹卫刚.基于优化小波神经网络的作物温室环境温度预测模型[J].中国农机化学报,2024,45(12):73-80. DOI:10.13733/j.jcam.issn.2095-5553.2024.12.012.
Ni Meiyu, Du Zitao, Cao Weigang. Crop greenhouse environmental temperature prediction model based on optimized wavelet neural network[J]. Journal of China Agricultural Mechanization, 2024, 45(12): 73-80. (in Chinese)

[20]Jung D H, Kim H S, Jhin C, et al. Time-serial analysis of deep neural network models for prediction of climatic conditions inside a greenhouse[J]. Computers and Electronics in Agriculture, 2020, 173: 105402.

[21]张冠山,李天华,侯加林.考虑动态吸收率的玻璃温室覆盖层温度预测模型[J].农业工程学报,2020,36(5):201-211.
Zhang Guanshan, Li Tianhua, Hou Jialin. Temperature prediction model for glass greenhouse cladding considering dynamic absorptivity[J]. Transactions of the CSAE, 2020, 36(5): 201-211. (in Chinese)

[22]郝璇,任守纲,舒欣,等.基于 TimesTransformer 的农业设施环境预测与调控研究[J].南京农业大学学报,2026,49(3):653-664.
Hao Xuan, Ren Shougang, Shu Xin, et al. Research on environmental prediction and regulation of agricultural facilities based on TimesTransformer[J]. Journal of Nanjing Agricultural University, 2026,49(3):653-664. (in Chinese)

[23]Cai X, Liu T. State of health prediction for lithium-ion batteries using transformer-LSTM fusion model[J]. Applied Sciences, 2025, 15(7): 3747.



(上接第 374 页)

[26]于丰华,曹慧妮,金忠煜,等.基于 NSWOA-ELM 算法的水稻冠层氮素含量反演方法[J].农业机械学报,2025,56(7):532-540.
Yu Fenghua, Cao Huini, Jin Zhongyu, et al. Inversion method of nitrogen content in rice canopy based on NSWOA-ELM algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(7): 532-540. (in Chinese)

[27]Liu X, Ma T, Zhang H, et al. The accumulation of fungal not bacterial residue carbon is management-dependent under conventional and organic practices in apple-orchard soil[J]. Geoderma, 2024, 448: 116968.

[28]张鹏,翁建文,康晴,等.基于卷积神经网络的太阳光谱辐照度超分辨率重建方法[J].光子学报,2025,54(3):221-230.
Zhang Peng, Weng Jianwen, Kang Qing, et al. Super-resolution solar spectral irradiance reconstruction method based on convolutional neural network[J]. Acta Photonica Sinica, 2025, 54(3): 221-230. (in Chinese)

[29]Koch M, Schodlok M C, Guggenberger G, et al. Effects of water tension and surface roughness on soil hyperspectral reflectance [J]. Geoderma, 2021, 385: 114888.

[30]Wilson R H, Nadeau K P, Jaworski F B, et al. Review of short-wave infrared spectroscopy and imaging methods for biological tissue characterization[J]. Journal of Biomedical Optics, 2015, 20(3): 030901.

[31]Tsenkova R, Muncán J, Pollner B, et al. Essentials of aquaphotomics and its chemometrics approaches[J]. Frontiers in Chemistry, 2018, 6: 363.

[32]Bablet A, Vu P V H, Jacquemoud S, et al. MARMIT: a multilayer radiative transfer model of soil reflectance to estimate surface soil moisture content in the solar domain (400 ~ 2 500 nm) [J]. Remote Sensing of Environment, 2018, 217: 1-17.

[33]Ding A, Ma H, Zhao P, et al. Improved Hapke model to characterize soil moisture content variation[J]. Environmental Sciences Proceedings, 2024, 29(1): 76.