doi:10.6041/j.issn.1000-1298.2025.09.047

基于高光谱红边偏度、峰度和机器学习的植被叶绿素含量估算模型研究

姚付启[1]　曾凡超[1,2]　孙金伟[3]　饶志龙[1]　王子涵[1]

(1.鲁东大学水利土木学院, 烟台 264025; 2.中国科学院东北地理与农业生态研究所, 长春 130102;
3.鲁东大学资源与环境工程学院, 烟台 264025)

摘要: 叶绿素是植物光合作用的关键组分,能够反映植物的生理状况与健康水平。准确估算植被叶绿素含量对于监测植物生长状态具有重要意义。尽管传统红边参数在叶绿素含量估算中得到了广泛应用,但其存在一定的局限性。为弥补这一不足,引入红边偏度 S_{RE} 和红边峰度 K_{RE} 作为新的红边参数,探索其与叶绿素含量之间的关系,并以冬小麦和夏玉米为研究对象,采用多元线性回归(Multiple linear regression, MLR)、极端梯度提升树(Extreme gradient boosting, XGBoost)和反向传播神经网络(Backpropagation neural network, BPNN)3 种建模方法,比较以 S_{RE} 和 K_{RE} 作为输入参数的叶绿素含量估算模型与传统红边参数模型的表现。研究结果表明:S_{RE} 和 K_{RE} 与叶绿素含量之间存在显著负相关,随着叶绿素浓度的增加,S_{RE} 和 K_{RE} 均呈现下降趋势。采用 3 种建模方法的不同作物叶绿素含量估算模型中,以 S_{RE} 和 K_{RE} 为输入参数的模型预测精度明显优于传统红边参数模型,具有更高的 R^2,更低的 RMSE,残差的分布也更加集中,模型的稳定性和精度较高。在 3 种建模方法中,BPNN 模型表现最佳,其 R^2 显著高于其他两种方法。本研究证明了 S_{RE} 和 K_{RE} 相比传统红边参数更能全面地反映叶绿素含量的变化特征,且当结合 BPNN 模型进行建模时,能够有效提升植被叶绿素含量的估算精度,为植被叶绿素含量的高精度估算提供了一种新的技术路径,对作物生长状态监测和精准农业应用具有实际意义。

关键词: 叶绿素含量; 估算模型; 红边偏度; 红边峰度; 机器学习

中图分类号: S127　　**文献标识码**: A　　**文章编号**: 1000-1298(2025)09-0566-10　　**OSID**:

Estimating Vegetation Chlorophyll Content Based on Hyperspectral Red Edge Skewness, Kurtosis, and Machine Learning

YAO Fuqi[1]　ZENG Fanchao[1,2]　SUN Jinwei[3]　RAO Zhilong[1]　WANG Zihan[1]

(1. *School of Hydraulic and Civil Engineering, Ludong University, Yantai* 264025, *China*
2. *Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun* 130102, *China*
3. *School of Resources and Environmental Engineering, Ludong University, Yantai* 264025, *China*)

Abstract: Chlorophyll is a key component in plant photosynthesis and an important indicator of physiological status and plant health. Accurate estimation of chlorophyll content is essential for monitoring plant growth conditions. Although traditional rededge parameters were widely used in chlorophyll estimation, they presented certain limitations. To address these shortcomings, rededge skewness (S_{RE}) and rededge kurtosis (K_{RE}) were introduced as novel rededge parameters to explore their relationship with chlorophyll content. Taking winter wheat and summer maize as study subjects, three modeling approaches—multiple linear regression (MLR), extreme gradient boosting (XGBoost), and backpropagation neural network (BPNN)—were employed to compare the performance of chlorophyll estimation models by using the proposed parameters versus traditional rededge parameters. The results demonstrated that S_{RE} and K_{RE} were significantly negatively correlated with chlorophyll content, showing a decreasing trend as chlorophyll levels were increased. Across all three modeling methods and crop types, models incorporating S_{RE} and K_{RE} achieved higher prediction accuracy than those based on traditional rededge parameters, characterized by higher R^2, lower RMSE, and more concentrated residual distributions, indicating enhanced model stability and precision. Among the three modeling approaches,

收稿日期: 2025-04-03　修回日期: 2025-04-30

基金项目: 国家自然科学基金项目(51809284、51309016)和山东省自然科学基金项目(ZR2020ME254、ZR2020QD061)

作者简介: 姚付启(1981—),男,教授,博士,主要从事农业节水与农业信息技术研究,E-mail: fuqiyao163@163.com

the BPNN model consistently exhibited the best performance, with significantly higher R^2 values compared with that of MLR and XGBoost. The research result confirmed that S_{RE} and K_{RE} provided a more comprehensive representation of chlorophyll variation than traditional rededge parameters. When combined with BPNN modeling, they can substantially improve the accuracy of chlorophyll content estimation. These findings can offer a novel technical pathway for high-precision chlorophyll monitoring and had practical implications for crop condition assessment and precision agriculture.

Key words: chlorophyll content; estimation model; red edge skewness; red edge kurtosis; machine learning

0　引言

植被叶绿素是衡量植被生长状态、光合作用效率和生态健康的重要指标,对生态监测与环境评估具有重要意义[1]。准确获取叶绿素含量不仅能揭示植被的生长动态和健康水平,还能反映植被对环境变化的响应,为推动生态保护和可持续发展提供科学依据[2]。然而,传统的叶绿素测量方法往往需要复杂的采样和实验操作,具有破坏性且耗时,难以满足大范围、实时监测的需求,尤其在动态生态系统中应用受限[3]。尽管便携式叶绿素仪能够无损测定叶片的相对叶绿素含量,但受限于逐点测量的局限性,效率低下且难以满足大范围、高精度的监测需求[4]。

近年来,随着遥感技术的快速发展,高光谱遥感凭借其高分辨率、高效率和无损性的优势,成为植被生态参数估算的重要工具[5]。其中,基于红边波段的叶绿素估算方法因其对植被光谱变化的高度敏感性,被广泛应用于农业、林业及生态学研究领域[6]。传统红边参数以及其衍生参数已成为估算叶绿素含量的重要手段[7-9]。

尽管传统红边参数具有一定的估算能力,但仍存在明显的局限性,特别是在高浓度叶绿素条件下。叶绿素在红光区表现出强烈的吸收能力,其吸收过程遵循 Beer - Lambert 定律[10],但当叶绿素浓度较高时,几乎所有入射红光均被吸收,导致反射率在该波段迅速达到极限,即形成饱和现象。与此同时,植被内部复杂的多次散射效应使得部分光线的传播不完全依赖于叶绿素吸收,而受到叶片结构和排列方式的影响[11]。这种结构散射与吸收的联合作用导致红边区域反射曲线的斜率和幅度对浓度变化的响应从线性逐渐转为非线性并趋于平稳,进而限制了传统红边参数在叶绿素高浓度区间的敏感性和估算精度。从参数表达形式来看,传统红边参数多聚焦于红边波段的一阶导数反射率差值(D_{RE})、一阶导数反射率极值位置(λ_{RE}),或红边范围下的积分面积(SD_{RE})等,难以捕捉红边曲线整体形态的变化。尤其在不同叶绿素浓度对应的红边曲线具有相同的参数值时,会导致传统红边参数对叶绿素的判别能力下降。

为解决上述问题,本文引入统计学中的红边偏度(S_{RE})和红边峰度(K_{RE})作为新的红边参数,更全面地揭示红边波段光谱曲线的分布特征。具体而言,S_{RE}与 K_{RE}均基于红边范围内一阶导数光谱曲线计算所得,分别代表该导数曲线红边波段范围内的偏度和峰度。S_{RE}是描述光谱曲线分布偏斜程度的三阶统计量,用以衡量曲线相对于其平均导数值的偏移方向和不对称性[12]。随着叶绿素浓度的增加,红边区域反射曲线上升段位置会发生偏移,并出现陡峭变化,使得一阶导数曲线在红边范围内呈现出明显的非对称性。S_{RE}可敏感捕捉这种不对称性的变化,反映了由叶绿素吸收增强与冠层结构散射共同引起的红边跃变特征,是对红边形态变化的有效量化指标。K_{RE}则是衡量红边范围内光谱一阶导数分布尖锐程度的四阶统计量,即红边区域反射率变化速度的集中或分散状态[13]。K_{RE}不仅受到植被冠层结构(如叶片排列密度、角度变化)所引起的多重散射效应影响,同时也受叶绿素浓度变化的制约。当叶绿素浓度升高时,红边曲线逐渐趋于平滑,其导数曲线的尖锐程度下降,从而导致峰度降低。因此,K_{RE}能够揭示高浓度叶绿素条件下红边反射形态由尖锐向平缓演化的趋势,反映光谱信息在结构与生理变化下的响应机制。

为系统评估上述新参数在植被叶绿素含量估算中的有效性与适用性,本文以华北地区常见农作物(夏玉米和冬小麦)为研究对象,构建多种叶绿素含量估算模型。选择多元线性回归(MLR)[14]、极端梯度提升(XGBoost)[15]和反向传播神经网络(BPNN)[16]3 种建模方法进行比较,探索不同建模方法在不同输入参数组合下的精度表现。通过对比这 3 种建模方式在不同输入参数组合下的模型性能,验证新引入的 S_{RE}与 K_{RE}是否能显著提升估算精度,并探讨其在高光谱红边参数体系中的科学价值与实际应用潜力,旨在从光谱物理特性与统计特征结合的视角,为高光谱遥感在植被叶绿素监测中的应用提供新的思路和理论支持,推动叶绿素含量高效、无损估算方法发展。

1 研究方法

1.1 理论基础

1.1.1 传统红边参数

在植被光谱研究中,红边参数为用来描述红边区域光谱特征的几个关键指标。常见的红边参数包括 λ_{RE}、D_{RE}和 SD_{RE}。λ_{RE}定义为植被一阶导数光谱位于波段 680 ~ 760 nm 内的最大值位置波长[17];D_{RE}定义为植被一阶导数光谱位于波段 680 ~ 760 nm 内的反射率最大值[18];SD_{RE}定义为植被一阶导数光谱位于波段 680 ~ 760 nm 内所包围的面积,为红边波长范围内一阶导数光谱反射率总和[19]。

1.1.2 红边偏度、峰度

偏度是描述数据分布的不对称程度的统计量,用于衡量分布相对于平均值的偏斜方向和程度[20]。正偏(右偏):分布右尾较长,数据更多集中在平均值左侧。负偏(左偏):分布左尾较长,数据更多集中在平均值右侧。零偏:分布对称,例如正态分布。不同偏度如图 1 所示。

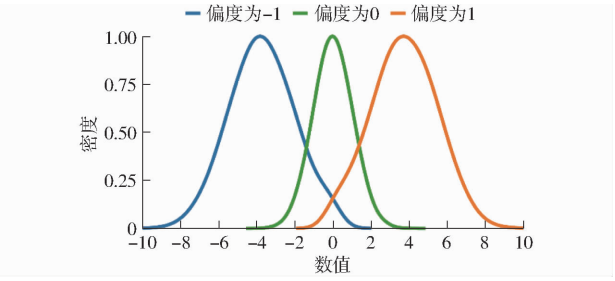


图 1 不同偏度示意图

Fig. 1 Schematic of different skewness values

S_{RE}为红边范围内一阶导数光谱反射率偏度,计算公式为

$$S_{RE}=\frac{\frac{1}{n}\sum_{i=1}^{n}(R'_i-\overline{R}')^3}{\left[\frac{1}{n}\sum_{i=1}^{n}(R'_i-\overline{R}')^2\right]^{3/2}} \tag{1}$$

式中 n——波段数量

R'_i——第 i 个波段位置一阶导数反射率

$\overline{R}'$——该范围内平均一阶导数反射率

峰度是描述数据分布形状“尖峰”或“平坦”程度的统计量[21]。高峰(尖峰):分布集中在均值附近,尾部较重;低峰(平峰):分布更平坦,尾部较轻;正态分布峰度:峰度为 3。不同峰度如图 2 所示。

K_{RE}为红边范围内一阶导数光谱反射率峰度,其计算公式为

$$K_{RE}=\frac{\frac{1}{n}\sum_{i=1}^{n}(R'_i-\overline{R}')^4}{\left[\frac{1}{n}\sum_{i=1}^{n}(R'_i-\overline{R}')^2\right]^2} \tag{2}$$

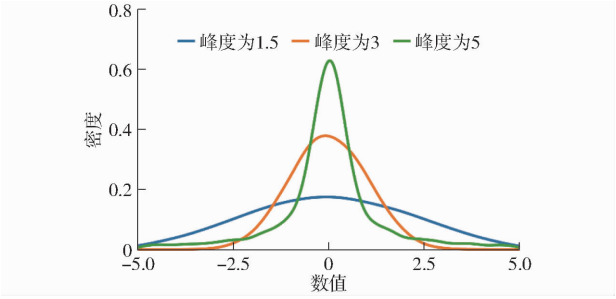


图 2 不同峰度示意图

Fig. 2 Schematic of different kurtosis values

1.2 实验地概况

实验于 2022 年 6 月至 2024 年 10 月在鲁东大学农业水资源高效利用实验地进行。实验地位于山东半岛北部濒临渤海,属于低山丘岭地形,平均海拔 47.8 m,为暖温带大陆性季风气候,气候湿润凉爽,冬日无严寒、夏日无酷暑,四季分明,并且主要盛行东北和西北风,年平均风速 2.7 m/s 左右。常年平均气温为 11.8 ~ 13.0℃,历史最高、最低气温为 36、-14℃,多年平均降水量为 651.9 ~ 722.2 mm,降水集中在 7—8 月,年均光照时数为 2 611.6 ~ 2 698.4 h,无霜期为 210 ~ 282 d。实验地土壤类型为壤土,平均干容重为 1.1 ~ 1.4 g/cm^3,田间持水量为 22% ~ 25%,凋萎系数为 13% ~ 15%,有机质质量分数为 3% ~ 5%,pH 值为 6 ~ 7。

1.3 数据采集

1.3.1 叶绿素相对含量测定

采用 SPAD-502PLUS 型便携式叶绿素仪对作物叶绿素相对含量(SPAD)进行测量。每株作物选择 5 个叶片测量,每个叶片测量 10 次,取平均值。每隔 10 d 测量一次,保证覆盖到每个关键生育期,精确到 0.1。所有测量在 11:00—13:00 期间完成,确保数据一致性。优先在无雨、光照稳定、风速低于 2.5 m/s 的晴朗天气进行,以减小环境对数据的影响,并与冠层光谱数据采集时段保持一致性。

1.3.2 冠层光谱测定

采用 ASD 型便携式地物光谱仪(ASD FieldSpecHandHeld2)获取作物冠层光谱数据。光谱仪测量波长为 325 ~ 1 075 nm,波长精确度为 1 nm,光谱分辨率 3.0 nm,视场角 25°。为确保数据具有良好的可比性与代表性,测量严格选择在北京标准时间 11:00—13:00 的太阳高度角较高、光照稳定时段进行,天气要求晴朗、无云或少云,风速低于 2.5 m/s,避免阴影、风力、云层等干扰因素。

测量前,使用原厂标配的高反射率标准白板(反射率 0.99 以上,型号 Spectralon,误差 0.01 以内)进行光谱仪校准。具体校准流程为:将光谱仪

探头垂直指向白板，读取并平均 10 次参考光谱，确保标准曲线在波段 400 ~ 1 000 nm 内的反射率稳定并接近 100%。若 10 次读取的标准差超过 0.01，则重新进行校准。为减小环境变化（如光照强度变化）对仪器稳定性的影响，每连续测量30 min 后重新定标一次，确保光谱响应一致性与数据可比性。

光谱测量过程中，将探头垂直向下对准作物冠层中心，距离冠层顶部保持 10 ~ 15 cm，使冠层填满仪器视场角，模拟垂直遥感观测方式。每个采样点重复采集 10 条光谱曲线，剔除异常值，计算剩余光谱平均值作为最终冠层反射率。所有冠层测量点均与 SPAD 测定点一一对应，确保数据配对准确无误。

1.4　高光谱数据处理

1.4.1　原始光谱

光谱采集完成后，利用地物光谱仪配套软件 ViewSpec 导出原始数据，并采用 Savitzky－Golay 滤波器进行平滑处理。该滤波器使用二次多项式和 11 点滑动窗口，有效平滑曲线同时保留光谱特征。为定量评估其降噪效果，对处理前后波段 350 ~ 910 nm 内的标准差进行分析，结果显示标准差由 0.008 03 降至 0.006 78，下降幅度达 15.57%，表明噪声显著降低，光谱曲线更加平滑稳定。处理前后高光谱曲线对比如图 3 所示。

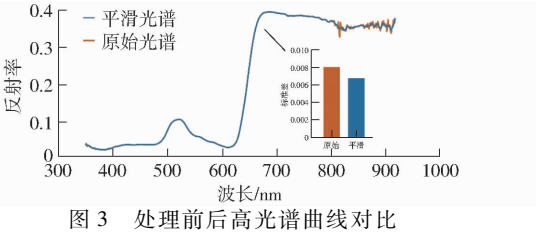


图 3　处理前后高光谱曲线对比

Fig. 3　Comparison of hyperspectral curves before and after treatment

1.4.2　一阶导数光谱

对原始光谱进行一阶导数变换，从而降低部分环境背景和仪器自身误差带来的影响[22]。计算方法为

$$R'_i = \frac{R_{i+1} - R_{i-1}}{2\Delta i} \tag{3}$$

式中　R_i——波长 i 处原始光谱反射率

Δi——波长 $i+1$ 到 i 的间隔，取 1 nm，其选择基于光谱仪分辨率（1 nm）与噪声抑制需求

1.5　模型和评估指标

1.5.1　多元线性回归

MLR 是一种传统的统计方法，用于建立因变量与两个或多个自变量之间的关系，假设其之间存在线性关联。由于其计算简便且具有较强的解释能力，MLR 广泛应用于作物指数高光谱反演研究[23]。在本研究中，植被叶绿素含量为因变量，（λ_{RE}、D_{RE}和 SD_{RE}）或（S_{RE}和 K_{RE}）为自变量。

MLR 模型可以表示为

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon \tag{4}$$

式中　y_i——SPAD

β_0——截距

β_1、β_2、…、β_n——输入参数 x_1、x_2、…、x_n 的回归系数

ε——误差项

在 R 语言中，MLR 模型是通过 stats 包中的 lm 函数进行开发的。模型性能通过 lm 函数拟合并评估，评估指标采用均方根误差（Root mean square error，RMSE）和决定系数 R^2。所有结果都进行了系统记录，对 MLR 模型性能提供了全面的概述。

1.5.2　极端梯度提升树

XGBoost 算法是一种基于 Boosting 思想的集成学习算法[24]。该模型具有并行和分布式计算能力，是目前速度最快、继承效果最好的决策树算法。XGBoost 模型有 3 个指标，分别为增益、频度和覆盖度[25]，以此来判断提升树的重要性。

XGBoost 模型又是梯度提升算法，假设模型是由 K 个 CART 组成，则表达式为

$$\Phi(x_i) = \sum_{k=1}^{K} f_k(x_i) \quad (f_k \in F) \tag{5}$$

其中　$F = \{f | f(x) = w_q(x)\}$

式中　$f_k(x_i)$——第 k 棵树的第 i 个节点的得分

K——样本总个数

F——所有可能的 CART 集合

$w_q(x)$——回归树叶子节点权重

XGBoost 与大多数机器学习模型相似，可以由控制模型精度的损失函数和平衡模型复杂度的正则化项来表示，有助于平滑模型的学习权重，避免过拟合现象的发生。一般公式设定为

$$L(\Phi) = l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{6}$$

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \| \omega \|^2 \tag{7}$$

式中　l——损失函数，可计算预测值 $\hat{y}_i$ 和目标值 y_i 之间的差异

Ω——模型惩罚项

γ——树中叶子复杂度

T——决策树中叶子个数

λ——平衡惩罚项参数

ω——叶片总得分[24]

XGBoost 模型使用 R 语言中的 xgboost 包(版本 x64 4.0.4)实现。超参数优化采用网格搜索结合 10 折交叉验证进行,调参范围包括:学习率{0.01, 0.05, 0.1},最大深度{3, 6, 9},子样本比例与特征采样比例{0.6, 0.7, 0.8},以及迭代次数{100, 200}。评估指标采用 RMSE 和 R^2,每组参数组合均在交叉验证下重复训练并评估,选取平均误差最低的参数组合作为最优模型。最终确定的参数为:学习率为 0.1,最大深度为 6,子样本比例与特征采样比例均为 0.7,迭代次数为 100。

1.5.3 反向传播神经网络

BPNN 是一种具有自适应学习能力和强大拟合能力的机器学习算法,通过反向传播算法优化网络权重,能够快速处理大量数据,提供高精度的预测,为实现各种预测目标提供了强有力的支持。BPNN 被广泛应用于多个领域,尤其在遥感影像分析和环境参数建模中表现出色[26]。

本研究基于 R 语言构建 BPNN 模型,包括输入层、隐含层和输出层共 3 层结构。输入层的神经元个数为 2(S_{RE}和 K_{RE})或 3(λ_{RE}、D_{RE}和 SD_{RE});隐含层神经元个数的选择直接影响模型的性能[27],为优化隐含层结构并防止过拟合,通过网格搜索结合 10 折交叉验证,确定最优隐含层神经元数,调参范围为{3,5,7,9},以平均交叉验证 RMSE 为选择依据,最终确定隐含层神经元数为 5。输出层仅包含 1 个神经元,用于输出 SPAD。激活函数采用线性激活函数,以避免输出范围限制,从而保留更多原始信息;网络权重通过正则化限制过拟合风险,训练次数最大为 500 次。

1.5.4 评估指标

为验证各模型可靠性与稳健性,随机选取 1/3 的样本集作为验证集开展验证工作。针对 MLR、XGBoost 及 BPNN 模型在 SPAD 模拟中的准确性评估,选用 R^2、RMSE 和残差 3 项关键指标进行量化分析。

2 结果与分析

2.1 SPAD 描述性统计分析

本研究分别收集了连续两年的夏玉米和冬小麦数据,并随机划分为训练数据集(占总数据的 2/3)和验证数据集(占总数据的 1/3),各数据集统计特性如表 1 所示。由表 1 可知,夏玉米的整体 SPAD 显著高于冬小麦,而夏玉米的 SPAD 范围小于冬小麦,表明夏玉米植株间的 SPAD 差异较小,冬小麦的差异较大。从变异系数统计结果来看,冬小麦训练数据集变异系数为 4.35%,验证数据集变异系数为 4.52%,均高于夏玉米数据集的变异系数。夏玉米训练数据集变异系数为 2.84%,验证数据集变异系数为 3.38%,表现出更低的变异性。上述结果表明,相较于冬小麦,夏玉米的 SPAD 分布更集中,数据波动较小,稳定性更高,而冬小麦的 SPAD 分布较为分散,个体间差异显著。

表 1 SPAD 测量值统计分析

Tab.1 Statistical analysis of SPAD measurements

作物种类	训练数据集				验证数据集			
	样本量	范围	平均值±标准差	变异系数/%	样本量	范围	平均值±标准差	变异系数/%
冬小麦	100	2.9~54.7	29.95±1.302	4.35	50	5.5~49.4	30.53±1.38	4.52
夏玉米	60	31.3~66.7	48.06±1.367	2.84	30	34.8~63.0	49.04±1.658	3.38

2.2 不同 SPAD 对应的红边偏度和峰度变化

图 4a 展示了不同 SPAD 下植被红边范围内光谱曲线变化特征,反映了红边光谱曲线形态随 SPAD 增加而发生的显著变化。具体而言,随着 SPAD 的增加,红边位置逐渐向右移动,表现出明显的“红边右移”现象。这表明叶绿素含量增加显著影响了红边光谱分布特征。同时,红边位置右移导致光谱曲线分布不对称性逐渐减弱,即红边偏度呈现出逐步减小趋势(图 4b)。此外,随着 SPAD 的增加,红边光谱曲线的峰形逐渐趋于平缓,红边峰度也随之逐步降低(图 4b)。这一趋势表明,SPAD 高,红边范围内的光谱曲线越集中且不尖锐,进一步反映了叶片结构和光谱特征的动态变化。这一现象直观地展现了红边偏度和峰度对 SPAD 变化的敏感响应,为深入理解红边光谱参数与叶绿素含量之间的关系提供了重要支持。

2.3 叶绿素与传统红边参数和红边偏度、峰度的相关性对比分析

相关性热图(图 5)展示了冬小麦和夏玉米 SPAD 与传统红边参数、S_{RE}和 K_{RE}的相关性。冬小麦和夏玉米的 SPAD 与 S_{RE}和 K_{RE}均呈显著负相关,且相关系数较高,均介于 -0.8 ~ -0.7 之间。SPAD 与 λ_{RE}呈显著正相关,相关系数为 0.7~0.8。冬小麦 SPAD 与 D_{RE}和 SD_{RE}也存在显著相关性,但

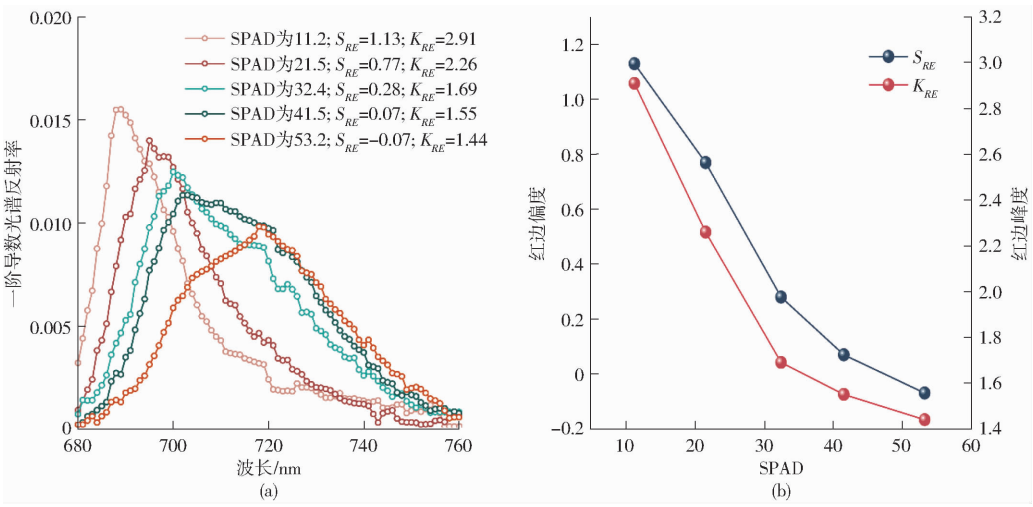


图 4　不同 SPAD 下红边光谱曲线的变化及红边偏度与峰度随 SPAD 的变化关系

Fig. 4　Changes of red edge spectral curves and relationship between red edge skewness and kurtosis with SPAD under different SPAD

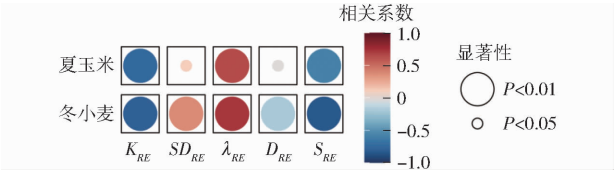


图 5　冬小麦和夏玉米的 SPAD 与红边参数的相关性热图

Fig. 5　Correlation heat map of SPAD values and red edge parameters of winter wheat and summer maize

相关系数较低，绝对值为 0.3 左右。夏玉米 SPAD 与 D_{RE} 和 SD_{RE} 的相关性不显著，相关系数较小，绝对值仅为 0.2 左右。综合来看，无论是夏玉米还是冬小麦，相较于传统红边参数，S_{RE} 和 K_{RE} 与 SPAD 的相关性更高，说明 S_{RE} 和 K_{RE} 能够更好地反映植物叶绿素浓度变化特征，为 SPAD 估算提供了更敏感和可靠的参数依据。

2.4　基于传统红边参数和红边偏度、峰度的 SPAD 估算模型研究

2.4.1　冬小麦 SPAD 估算模型

图 6 展示了基于传统红边参数（图 6a ~ 6c）与 S_{RE} 和 K_{RE}（图 6d ~ 6f）为输入变量，采用 MLR、

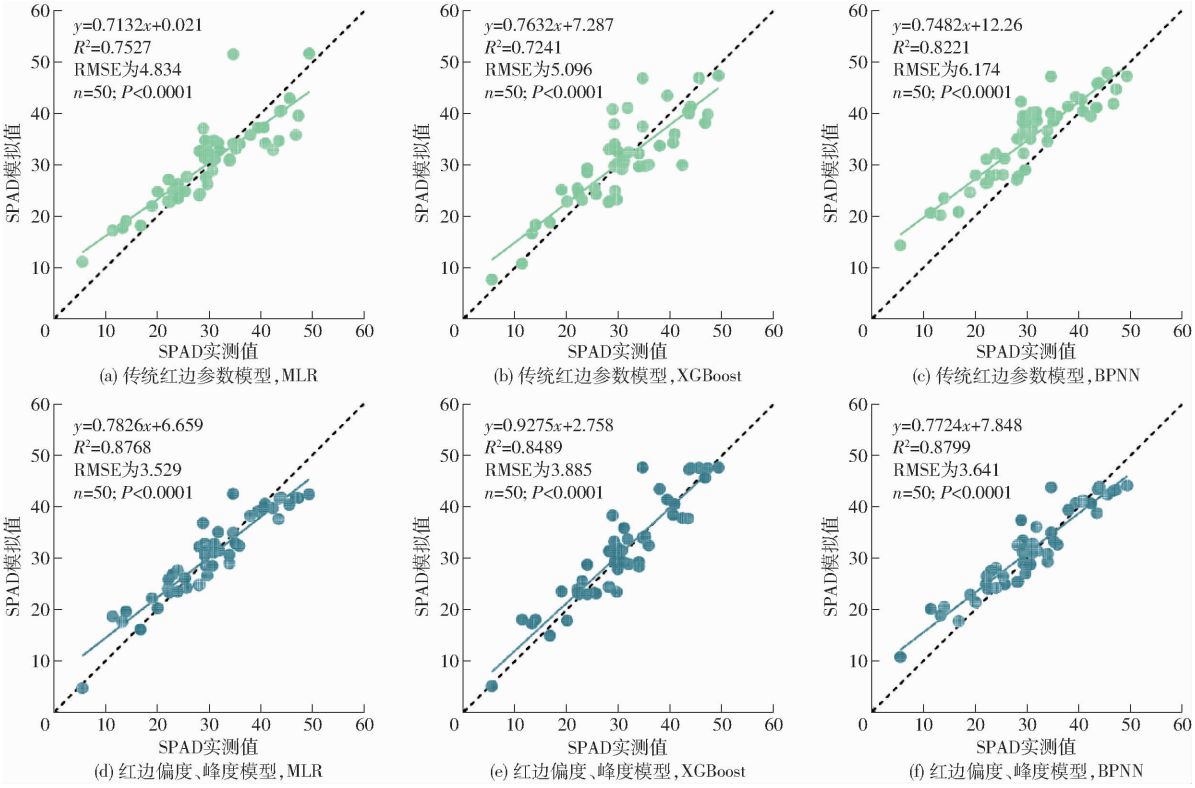


图 6　冬小麦 SPAD 估算模型的模拟值与实测值对比散点图

Fig. 6　Scatter plots comparing simulated and measured values of winter wheat SPAD value estimation model

XGBoost 和 BPNN 构建的冬小麦 SPAD 估算模型的散点图。结果表明,所有模型均在极显著水平上具有统计学意义($P<0.000\ 1$)。从输入参数角度分析,采用 S_{RE} 和 K_{RE} 为输入参数能够显著提升模型性能,相较于传统红边参数模型,MLR 模型 R^2 提高 16.49%,RMSE 降低 26.99%,XGBoost 模型 R^2 提高 17.24%,RMSE 降低 23.76%,BPNN 模型 R^2 提高 7.03%,RMSE 降低 41.03%,说明新型红边参数在提高冬小麦 SPAD 估算精度方面具有显著优势。从建模方法来看,BPNN 模型在不同输入条件下均表现出优越性能,拟合精度优于 MLR 与 XGBoost,其中,以 S_{RE} 和 K_{RE} 为输入的 BPNN 模型估算效果最佳($R^2=0.879\ 9$,RMSE 为 3.641),表明新型红边参数与 BPNN 方法结合,可显著提升冬小麦 SPAD 估算的准确性与模型稳定性。

2.4.2 夏玉米 SPAD 估算模型

图 7 呈现了以传统红边参数(图 7a ~ 7c)与 S_{RE} 和 K_{RE}(图 7d ~ 7f)为输入变量,分别采用 MLR、XGBoost 和 BPNN 方法构建的夏玉米 SPAD 估算模型的散点图。统计结果显示,所有模型均达到显著水平($P<0.05$),其中以 S_{RE} 和 K_{RE} 为输入变量的 BPNN 模型具有最高的显著性水平($P<0.000\ 1$),而以传统红边参数为输入的 XGBoost 模型显著性相对较低($P<0.05$)。从输入参数角度来看,基于 S_{RE} 和 K_{RE} 构建的模型在估算精度上表现更优,与传统红边参数模型相比,MLR 模型 R^2 提升 73.29%,RMSE 降低 18.15%;XGBoost 模型 R^2 提升 110.71%,RMSE 降低 24.67%;BPNN 模型 R^2 提升 104.31%,RMSE 降低 24.02%。这些结果表明,S_{RE} 和 K_{RE} 作为新型红边参数,在提升夏玉米 SPAD 估算准确性方面具有显著优势。在建模方法比较中,BPNN 模型整体性能最优,不同输入变量类型均展现出良好的拟合能力。其中,基于 S_{RE} 和 K_{RE} 构建的 BPNN 模型估算效果最佳,R^2 为 0.691 6,RMSE 为 5.664,表明新型红边参数与 BPNN 模型结合在估算夏玉米 SPAD 中具有较大潜力。

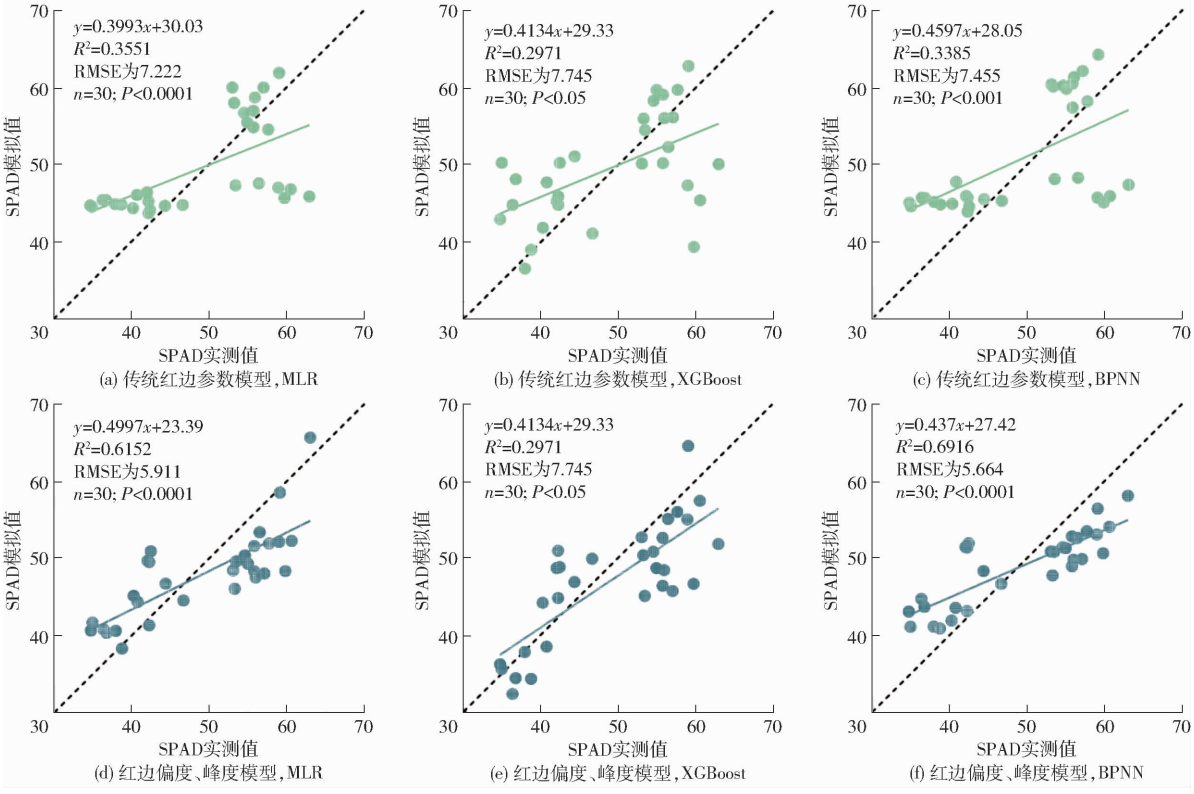


图 7 夏玉米 SPAD 估算模型的模拟值与实测值对比散点图

Fig. 7 Scatter plots comparing simulated and measured values of summer maize SPAD estimation model

2.4.3 模型对比

图 8 为冬小麦和夏玉米 SPAD 估算模型在不同输入参数下的残差分布。Re 表示以传统红边参数为输入参数;SK 表示以 S_{RE} 和 K_{RE} 为输入参数。总体来看,以 S_{RE} 和 K_{RE} 作为输入参数显著提升了模型性能。以传统红边参数为输入的 3 个模型中,残差分布较大,且存在明显的离散点,尤其是 MLR 模型。与传统红边参数相比,以 S_{RE} 和 K_{RE} 为输入参数使模型残差分布更为集中,误差范围明显减小,尤其是非线性模型(XGBoost 和 BPNN)表现出更强的鲁棒性和建模能力。

对于冬小麦数据,以 S_{RE} 和 K_{RE} 为输入的非线性模型表现优异,特别是 BPNN,其残差分布最为集中,误差范围最小,表现出较高的建模精度和稳定

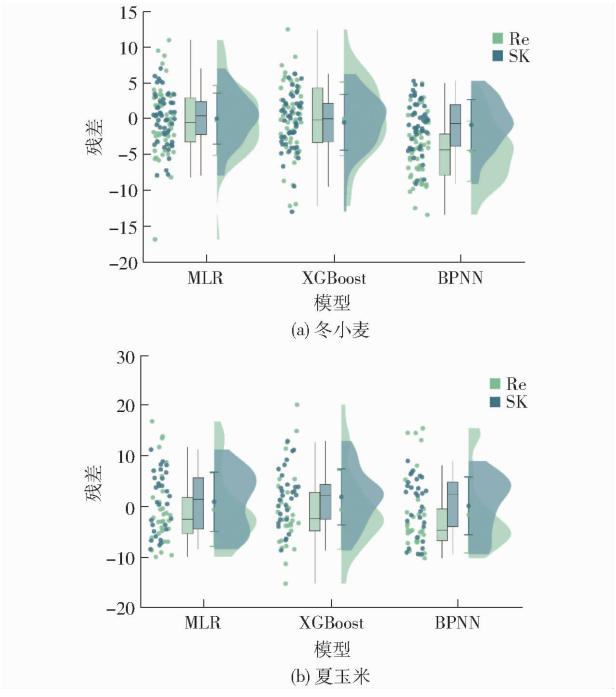


图 8　SPAD 估算模型模拟残差图

Fig. 8　Simulated residual plot of SPAD value estimation model

性。相比之下，MLR 模型以传统红边参数为输入时残差波动较大，存在显著离散点，即使引入 S_{RE} 和 K_{RE} 后有所改善，但仍难以充分捕捉冬小麦 SPAD 的非线性关系。

在夏玉米数据中，S_{RE} 和 K_{RE} 对模型性能的提升更为显著。以 S_{RE} 和 K_{RE} 为输入时，XGBoost 模型的误差范围显著减小，残差分布更集中，预测精度明显提高。BPNN 模型在夏玉米中的表现尤为突出，误差中值接近于零，分布范围最小，充分体现了其建模非线性关系的优势。MLR 模型的表现则受限于其线性特性，即使引入 S_{RE} 和 K_{RE} 后性能有所提升，但与非线性模型仍有较大差距。

综合来看，S_{RE} 和 K_{RE} 对冬小麦和夏玉米 SPAD 估算模型均表现出良好的普适性，显著提升了模型对作物光谱特征的适应能力。BPNN 模型在捕捉复杂光谱特征方面表现更优，为实现作物 SPAD 的精准估算提供了可靠的建模方法和技术路径。

3　讨论

3.1　红边偏度与峰度在叶绿素估算中的优势及机制

红边区域位于植物叶片对红光强吸收与近红外强反射之间的过渡区，是叶绿素含量变化最敏感的波段之一[28]。本研究所提出的 S_{RE} 和 K_{RE} 是对该区间内一阶导数光谱曲线的三阶与四阶统计特征，用以分别刻画该导数曲线的不对称性与尖锐程度，从而量化红边形态随叶绿素浓度变化所产生的细微变化[29]。当叶绿素浓度较低时，植物对红光吸收作用较弱，红边上升段变化迅速且陡峭[30]，其导数曲线呈现出不对称且尖锐的峰值形态，对应的红边偏度与峰度较高（图4）。随着叶绿素浓度的增加，红光吸收增强，近红外反射同步提高，红边位置随之发生“红移”，但由于光谱反射逐渐趋于饱和，导致原始曲线的反射率上升段变缓，其一阶导数曲线峰形也逐渐变得对称和平坦，最终表现为红边偏度与峰度值的逐步下降。即，S_{RE} 和 K_{RE} 在统计意义上揭示了红边范围内的一阶导数曲线由“非对称且尖锐”向“对称且平缓”演化的过程，这一变化正体现了高叶绿素浓度条件下植被光谱响应的饱和趋势。因此，相较于传统红边参数在高浓度叶绿素估算中面临的灵敏度下降问题，S_{RE} 和 K_{RE} 能够从红边曲线整体形态的变化中提取更加稳定和细致的光谱信息，提升模型在高浓度区间的判别能力和估算精度。

为了更全面地探讨 S_{RE} 和 K_{RE} 相较于传统红边参数在叶绿素变化过程中的响应特性，对不同红边参数在 SPAD 变化过程中的敏感性与饱和趋势进行了定量分析（图 9）。在 SPAD 低于 40 时，5 种红边参数对叶绿素变化均表现出较高的响应斜率，为 0.7 ~ 0.8，表明在低叶绿素浓度条件下，它们均具备良好的敏感性。然而，当 SPAD 超过 40 后，传统红边参数的响应斜率迅速下降，趋近于零，呈现出明显的饱和趋势，难以有效反映高叶绿素浓度的变化。相比之下，S_{RE} 和 K_{RE} 在高浓度叶绿素区间仍保持较高的响应斜率，为 0.6 ~ 0.7，显示出其在高浓度条件下依然具备较强的敏感性。因此，S_{RE} 和 K_{RE} 在一定程度上突破了传统红边参数在叶绿素高浓度区域易饱和的局限性，为提高叶绿素高值段的监测精度提供了更具潜力的光谱特征指标。

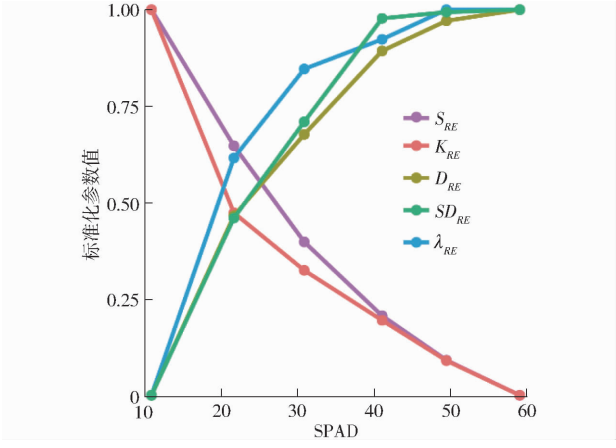


图 9　新红边参数和传统红边参数随 SPAD 升高变化曲线

Fig. 9　Variation curves of new and traditional red edge parameters to increase of SPAD

在建模方面，以 S_{RE} 和 K_{RE} 为输入参数的 BPNN

模型对叶绿素模拟展现出了较强的能力。BPNN 在高光谱建模中的优势主要体现在其强大的非线性映射能力[31]。高光谱遥感数据,尤其是红边区域的光谱信息,往往包含复杂且非线性的变化特征,这些特征与叶绿素浓度之间的关系受到多种生理、结构及环境因素的耦合作用,难以用简单的线性函数或决策树规则准确刻画[32]。MLR 模型在假设变量之间呈线性关系的前提下,难以有效提取这些光谱非线性信息;而 XGBoost 虽具备一定的非线性建模能力,其基于加法模型的结构更适用于离散或分段特征,对于红边光谱形态的连续变化响应存在识别能力不足的问题[33]。相比之下,BPNN 通过多层网络结构和非线性激活函数的组合[34],可对高光谱数据进行多层次特征抽象与表达,在无需人为设定映射函数的情况下自动捕捉输入与输出变量之间复杂的非线性关系[35],从而更精准地提取反映植被叶绿素变化的光谱响应特征,表现出更优的泛化能力与建模鲁棒性。

3.2 未来研究方向

本研究提出了基于 S_{RE} 和 K_{RE} 的统计特征,结合 BPNN 对夏玉米和小麦 SPAD 进行估算的新方法,显著提高了叶绿素含量估算精度,验证了 S_{RE} 和 K_{RE} 在红边曲线形态分析中的潜在价值。尽管 BPNN 在本研究中展现出较优的建模能力,但其性能在不同作物类型及环境条件下仍面临一定的不确定性与挑战。BPNN 模型在学习过程中依赖于训练样本的分布特征,若样本覆盖范围不足,或训练数据未能充分代表植被在不同光照、土壤、氮素水平等环境梯度下的光谱响应差异,模型在外部样本上的泛化能力可能受到限制,甚至出现过拟合现象。特别是在作物品种或叶型结构存在差异的情况下,红边曲线形态变化规律也可能发生偏移,从而影响 S_{RE} 和 K_{RE} 稳定性。此外,不同遥感观测条件(如传感器类型、观测角度、天空辐射状况)也可能对红边形态造成扰动,进一步加剧模型对样本环境的敏感性。因此,应高度重视模型对样本的依赖性和对外部条件变化的敏感性。一方面可通过多地多期采样、引入多样化的训练样本增强模型的泛化能力;另一方面,也可结合迁移学习、正则化技术或集成学习策略,降低模型对特定训练样本的过拟合风险。

未来的研究可以从以下几方面开展:①扩展数据来源,采集更多样本,涵盖不同作物种类、土壤类型及气候条件,以提高模型的鲁棒性和适用性。②结合其他环境因子与 S_{RE} 和 K_{RE} 构建多变量模型,以提升估算模型的精度与普适性。③优化模型算法,结合深度学习和物理模型,提高估算精度和计算效率。④探索 S_{RE} 和 K_{RE} 的其他潜在应用场景,如生物量估算、作物健康监测等,扩展其在农业生态研究中的应用范围。

4 结束语

采用传统红边参数以及 S_{RE} 和 K_{RE} 作为输入参数,结合 MLR、XGBoost 和 BPNN 建模方法,构建了冬小麦和夏玉米的叶绿素含量估算模型,系统探讨了 S_{RE} 和 K_{RE} 在叶绿素含量估算中的可靠性。研究结果表明,相较于传统红边参数模型,S_{RE} 和 K_{RE} 作为输入参数能够显著提升植被叶绿素含量的估算精度。在 3 种建模方法中,BPNN 模型在夏玉米和冬小麦叶绿素含量估算中的 R^2 更高,RMSE 更低,同时模型模拟残差分布更为集中,表现出较高的稳定性和预测精度。S_{RE} 和 K_{RE} 相比传统红边参数,能够更全面地反映叶绿素变化特征。当以 S_{RE} 和 K_{RE} 为输入参数,并结合 BPNN 模型进行建模时,可有效提升植被叶绿素含量估算精度。研究结果为植被叶绿素含量的高精度估算提供了一种创新方法,对植物生长状态监测和精准农业具有应用价值。

参 考 文 献

[1] KANCHEVA R, BORISOVA D, GEORGIEV G. Chlorophyll assessment and stress detection from vegetation optical properties [J]. Ecological Engineering and Environment Protection, 2014, 1: 34 - 43.

[2] HABOUDANE D, MILLER J R, TREMBLAY N, et al. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture[J]. Remote Sensing of Environment, 2002, 81(2 - 3): 416 - 426.

[3] QIU R, WEI S, ZHANG M, et al. Sensors for measuring plant phenotyping: a review[J]. International Journal of Agricultural and Biological Engineering, 2018, 11(2): 1 - 17.

[4] DONG T, SHANG J, CHEN J M, et al. Assessment of portable chlorophyll meters for measuring crop leaf chlorophyll concentration[J]. Remote Sensing, 2019, 11(22): 2706.

[5] YU H, KONG B, WANG Q, et al. Hyperspectral remote sensing applications in soil: a review[M]// PANDEY P C, SRIVASTAVA P K, BALZTER H, et al. Hyperspectral remote sensing. Amsterdam: Elsevier, 2020: 269 - 291.

[6] MAIN R, CHO M A, MATHIEU R, et al. An investigation into robust spectral indices for leaf chlorophyll estimation[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2011, 66(6): 751 - 761.

[7] ZHENG T, LIU N, WU L, et al. Estimation of chlorophyll content in potato leaves based on spectral red edge position[J].

IFAC－PapersOnLine, 2018, 51(17): 602－606.
[8] 钱彬祥,黄文江,叶回春,等.红边位置改进算法的冬小麦叶绿素含量反演[J].农业工程学报,2020,36(23):162－170.
QIAN Binxiang, HUANG Wenjiang, YE Huichun, et al. Inversion of winter wheat chlorophyll contents based on improved algorithms for red edge position[J]. Transactions of the CSAE, 2020, 36(23): 162－170. (in Chinese)
[9] 冯海宽,陶惠林,赵钰,等.利用无人机高光谱估算冬小麦叶绿素含量[J].光谱学与光谱分析,2022,42(11):3575－3580.
FENG Haikuan, TAO Huilin, ZHAO Yu, et al. Estimation of chlorophyll content in winter wheat based on UAV hyperspectral [J]. Spectroscopy and Spectral Analysis, 2022, 42(11): 3575－3580. (in Chinese)
[10] 王双喜,束美艳,顾晓鹤,等.利用 Beer－Lambert 消光定律遥感反演玉米叶面积指数[J].中国农业科技导报,2018,20(12):67－73.
WANG Shuangxi, SHU Meiyan, GU Xiaohe, et al. Retrieval of corn leaf area index using Beer－Lambert extinction law[J]. Journal of Agricultural Science and Technology, 2018, 20(12): 67－73. (in Chinese)
[11] USTIN S L, JACQUEMOUD S. How the optical properties of leaves modify the absorption and scattering of energy and enhance leaf functionality[M]//CAVENDER-BARES J, GAMON J A, TOWNSEND P A. Remote sensing of plant biodiversity. Cham: Springer, 2020: 349－384.
[12] ARELLANO-VALLE R B, AZZALINI A. On the unification of families of skew-normal distributions[J]. Scandinavian Journal of Statistics, 2006, 33(3): 561－574.
[13] ANNENKOV S Y, SHRIRA V I. Evaluation of skewness and kurtosis of wind waves parameterized by JONSWAP spectra[J]. Journal of Physical Oceanography, 2014, 44(6): 1582－1594.
[14] 蒋傲,丁玲,宋恩泽,等.基于无人机多光谱多时相小白菜 SPAD 监测研究[J].中国农村水利水电,2024(6):193－202,208.
JIANG Ao, DING Ling, SONG Enze, et al. UAV-based multispectral multi-temporal chard SPAD monitoring study[J]. China Rural Water and Hydropower, 2024(6): 193－202,208. (in Chinese)
[15] ZENG F, SUN J, ZHANG H, et al. Modeling soil respiration in summer maize cropland based on hyperspectral imagery and machine learning[J]. Frontiers in Environmental Science, 2025, 12: 1505987.
[16] 张驰,郭媛,黎明.人工神经网络模型发展及应用综述[J].计算机工程与应用,2021,57(11):57－69.
ZHANG Chi, GUO Yuan, LI Ming. Review of development and application of artificial neural network models[J]. Computer Engineering and Applications, 2021, 57(11): 57－69. (in Chinese)
[17] LI M, CHU R, SHA X, et al. Hyperspectral characteristics and scale effects of leaf and canopy of summer maize under continuous water stresses[J]. Agriculture, 2021, 11(12):1180.
[18] 王仁红,宋晓宇,李振海,等.基于高光谱的冬小麦氮素营养指数估测[J].农业工程学报,2014,30(19):191－198.
WANG Renhong, SONG Xiaoyu, LI Zhenhai, et al. Estimation of winter wheat nitrogen nutrition index using hyperspectral remote sensing[J]. Transactions of the CSAE, 2014, 30(19): 191－187. (in Chinese)
[19] BROGE N H, MORTENSEN J V. Deriving green crop area index and canopy chlorophyll density of winter wheat from spectral reflectance data[J]. Remote Sensing of Environment, 2002, 81(1): 45－57.
[20] JOANES D N, GILL C A. Comparing measures of sample skewness and kurtosis[J]. Journal of the Royal Statistical Society: Series D (The Statistician), 1998, 47(1): 183－189.
[21] ĐORIĆ D, NIKOLIĆ-ĐORIĆ E, JEVREMOVIĆ V, et al. On measuring skewness and kurtosis[J]. Quality and Quantity, 2009, 43(3): 481－493.
[22] SHEN L, GAO M, YAN J, et al. Hyperspectral estimation of soil organic matter content using different spectral preprocessing techniques and PLSR method[J]. Remote Sensing, 2020, 12(7): 1206.
[23] FEI S, XU D, CHEN Z, et al. MLR-based feature splitting regression for estimating plant traits using high-dimensional hyperspectral reflectance data[J]. Field Crops Research, 2023, 293: 108855.
[24] CHEN T, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 785－794.
[25] JIA Y, JIN S, SAVI P, et al. GNSS－R soil moisture retrieval based on an XGBoost machine learning aided method: performance and validation[J]. Remote Sensing, 2019, 11(14): 1655.
[26] ZHANG L, WANG F, SUN T, et al. A constrained optimization method based on BP neural network[J]. Neural Computing and Applications, 2018, 29(2): 413－421.
[27] 刘秀英,余俊茹,王世华.光谱特征变量和 BP 神经网络构建油用牡丹种子含水率估算模型[J].农业工程学报,2020,36(22):308－315.
LIU Xiuying, YU Junru, WANG Shihua. Estimation of moisture content in peony seed oil using spectral characteristic variables and BP neural network[J]. Transactions of the CSAE, 2020, 36(22): 308－315. (in Chinese)
[28] ZHANG N, SU X, ZHANG X, et al. Monitoring daily variation of leaf layer photosynthesis in rice using UAV-based multi-spectral imagery and a light response curve model[J]. Agricultural and Forest Meteorology, 2020, 291: 108098.

(下转第 595 页)

Applications, 2025, 61(7): 96 - 104. (in Chinese)
[26] 时雷,杨程凯,雷镜楷,等. 基于改进 YOLO v8s 的小麦小穗赤霉病检测研究[J]. 农业机械学报, 2024, 55(7): 280 - 289.
SHI Lei, YANG Chengkai, LEI Jingkai, et al. Wheat spikelet detection of fusarium head blight based on improved YOLO v8s [J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(7): 280 - 289. (in Chinese)
[27] TAN S, KUANG Z, JIN B. AppleYOLO: apple yield estimation method using improved YOLO v8 based on deep OC - SORT [J]. Expert Systems with Applications, 2025, 272: 126764.
[28] TAO H. A triple interference removal network based on temporal and spatial attention interaction for forest smoke recognition in videos[J]. Computers and Electronics in Agriculture, 2024, 218: 108756.
[29] LI C, ZHOU A, YAO A. Omni-dimensional dynamic convolution[J]. arXiv Preprint, arXiv: 2209.07947, 2022.
[30] ZHAO L, FU L, JIA X, et al. YOLO - BOS: An emerging approach for vehicle detection with a novel BRSA mechanism[J]. Sensors, 2024, 24(24): 8126.
[31] 杨断利,齐俊林,陈辉,等. 基于改进 YOLO v8n 模型的散养蛋鸡个体行为识别方法与差异分析[J]. 农业机械学报, 2024, 55(11): 112 - 123.
YANG Duanli, QI Junlin, CHEN Hui, et al. Individual behavioral identification and differential analysis of free-range laying hens based on improved YOLO v8n model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(11): 112 - 123. (in Chinese)
[32] YANG J, LI C, DAI X, et al. Focal modulation networks[J]. Advances in Neural Information Processing Systems, 2022, 35: 4203 - 4217.
[33] LIU Y, HU Y, CAO G, et al. SENet: super-resolution enhancement network for crowd counting[J]. Pattern Recognition, 2025, 162: 111420.
[34] FU H, WANG C, CHEN Z. Adaptive selection of spectral-spatial features for hyperspectral image classification using a modified-CBAM-based network[J]. Neurocomputing, 2025, 615: 128877.
[35] WU J, SUN M, JIANG C, et al. Context-based local-global fusion network for 3D point cloud classification and segmentation [J]. Expert Systems With Applications, 2024, 251: 124023.
[36] LI G, ZHAO T. Efficient image analysis with triple attention vision transformer[J]. Pattern Recognition, 2024, 150: 110357.
[37] 王芯蕊,王慧琴,王可,等. 基于改进 YOLO v8 的无人机视角下高速公路异常目标检测方法[J]. 计算机工程与应用, 2025, 61(11): 105 - 118.
WANG Xinrui, WANG Huiqin, WANG Ke, et al. Anomalous target detection method for highway from UAV viewpoint based on improved YOLO v8[J]. Computer Engineering and Applications, 2025, 61(11): 105 - 118. (in Chinese)

(上接第 575 页)

[29] YANG H, MING B, NIE C, et al. Maize canopy and leaf chlorophyll content assessment from leaf spectral reflectance: estimation and uncertainty analysis across growth stages and vertical distribution[J]. Remote Sensing, 2022, 14(9): 2115.
[30] ANJUM S, XIE X, WANG L, et al. Morphological, physiological and biochemical responses of plants to drought stress[J]. African Journal of Agricultural Research, 2011, 6(9): 2026 - 2032.
[31] 孙杰,杨静,丁书杰,等. 基于多注意力机制与编译图神经网络的高光谱图像分类[J]. 农业机械学报, 2024, 55(3): 183 - 192, 212.
SUN Jie, YANG Jing, DING Shujie, et al. Hyperspectral image classification based on multi-attention mechanism and compiled graph neural networks[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(3): 183 - 192, 212. (in Chinese)
[32] 马文勇,王训明. 基于高光谱分析的草地叶绿素含量估算研究进展[J]. 地理科学进展, 2016, 35(1): 25 - 34.
MA Wenyong, WANG Xunming. Progress on grassland chlorophyll content estimation by hyperspectral analysis[J]. Progress in Geography, 2016, 35(1): 25 - 34. (in Chinese)
[33] HAN L, YANG G, YANG X, et al. An explainable XGBoost model improved by SMOTE - ENN technique for maize lodging detection based on multi-source unmanned aerial vehicle images[J]. Computers and Electronics in Agriculture, 2022, 194: 106804.
[34] 王丽萍,张新明,李继伟,等. BP 网络激活函数选择及在径流预报模型中的应用[J]. 水力发电学报, 2014, 33(1): 29 - 36.
WANG Liping, ZHANG Xinming, LI Jiwei, et al. BP neural network activation function's selection and application to runoff forecasting model[J]. Journal of Hydroelectric Engineering, 2014, 33(1): 29 - 36. (in Chinese)
[35] 徐睿,梁循,齐金山,等. 极限学习机前沿进展与趋势[J]. 计算机学报, 2019, 42(7): 1640 - 1670.
XU Rui, LIANG Xun, QI Jinshan, et al. Advances and trends in extreme learning machine[J]. Chinese Journal of Computers, 2019, 42(7): 1640 - 1670. (in Chinese)