

基于冠层尺度高光谱分数阶微分的 薄壳山核桃叶片氮素含量估测

徐佳佳, 于磊, 傅根深, 燕李鹏, 黄庆丰, 唐雪海

1. 安徽农业大学 林学与园林学院, 合肥 230036;
2. 安徽农业大学 安徽省林木资源培育重点实验室, 合肥 230036

摘要: 植物叶片氮素含量是反映植物营养状况及长势的重要指标。高光谱技术能够无损、高效地估算植物生理生化参数, 为监测植物生长发育过程中的养分和健康状况提供可靠依据。为挖掘高光谱技术在薄壳山核桃养分估测中的应用潜力, 本研究以薄壳山核桃(长林和建德系列)为研究对象, 室外随机采集53株薄壳山核桃350—2500 nm的冠层高光谱后, 首先运用分数阶微分FOD(Fractional Order Derivative)进行冠层高光谱预处理; 然后联合两种两波段光谱指数, 探究薄壳山核桃叶片氮素含量LNC(Leaf Nitrogen Content)与冠层光谱的响应关系; 最后利用变量组合集群分析算法VCPA(Variable Combination Population Analysis)筛选建模变量, 分别构建冠层FOD单波段及FOD联合两波段光谱指数的极端梯度提升算法XGBoost(eXtreme Gradient Boosting)估算, 得到基于本试验条件下薄壳山核桃LNC适宜的估算模型。结果表明: 与原始光谱相比, FOD预处理后的冠层光谱与薄壳山核桃LNC的相关性增强, 提高了0.152; 在提高光谱特征与目标成分的相关性方面, FOD分别结合归一化光谱指数NDSI(Normalized Difference Spectral Index)和差值光谱指数DSI(Difference Spectral Index)的表现均比单波段效果更佳, 分别提高了0.250和0.277; VCPA变量选择方法最终筛选的光谱变量组合子集中同时包含强弱信息变量, 对提升估算模型精度具有重要的作用; 最优薄壳山核桃叶片氮素估算模型是1.5阶微分结合DSI模型, 模型预测集 $R^2_p=0.75$, RMSEP=1.32 g/kg。综上, 冠层高光谱可快速无损估算薄壳山核桃LNC, 分数阶微分结合两波段光谱指数可以显著提高光谱特征与目标变量的响应关系。

关键词: 薄壳山核桃, 冠层尺度, 高光谱遥感, 氮素, 分数阶微分, 光谱指数, 变量组合集群分析, 机器学习
中图分类号: O657.3/P2

引用格式: 徐佳佳, 于磊, 傅根深, 燕李鹏, 黄庆丰, 唐雪海. 2025. 基于冠层尺度高光谱分数阶微分的薄壳山核桃叶片氮素含量估测. 遥感学报, 29(5): 1164–1178

Xu J J, Yu L, Fu G S, Yan L P, Huang Q F and Tang X H. 2025. Estimation of leaf nitrogen content of *Carya illinoensis* using Fractional-Order Derivative of canopy-scale hyperspectral data. National Remote Sensing Bulletin, 29(5): 1164–1178 [DOI:10.11834/jrs.20243454]

1 引言

薄壳山核桃(*Carya illinoensis*)原产于北美东部地区, 是著名的落叶干果和木本油料树种(Zhang等, 2022)。坚果种仁色美味香, 富含人体所需的各种脂肪酸、氨基酸等营养物质, 其油脂亚麻酸含量高于橄榄油和茶油, 是优良的果材兼用树种, 具有良好的经济和生态效益(de Araújo等, 2021)。氮素是影响植物光合作用、光合产物积累和最终产量的重要生理生化指标(Berger等,

2020)。除了能够影响植物光合作用潜力和初级生产力, 氮素在反映植物营养缺乏等方面也起到指示作用(Simkin等, 2022)。获取并分析薄壳山核桃叶片氮素含量, 有助于获悉植株营养状况及长势变化, 进而精准指导施肥作业, 对后续提高果实品质和产量、监测植物群体或个体生长及养分状况等都具有极其重要意义(Yang等, 2021; Liu等, 2022)。

传统的氮素含量测定方法是现场进行破坏性

收稿日期: 2023-12-03; 预印本: 2024-06-03

基金项目: 国家自然科学基金(编号: 32171783)

第一作者简介: 徐佳佳, 研究方向为林业定量遥感。E-mail: xjj21720383@stu.ahau.edu.cn

通信作者简介: 唐雪海, 研究方向为林业定量遥感、森林资源动态监测、林业3S技术应用。E-mail: tangxuehai@ahau.edu.cn

取样,再带回实验室化学分析,难以做到大面积快速监测(Gao等,2019)。近年来,基于高光谱技术无损高效地估算有关植物生理生化指标得到快速发展,这种方法具有简洁性、敏感性和可信度等优势,适宜于大范围监测应用,为植物生长发育过程中的养分和健康状况评估及可持续管理提供了可靠的数据获取方法(Shu等,2022)。植物高光谱在地面观测层面主要有叶片和冠层两个尺度(Mirzaei等,2019)。其中,叶片尺度光谱特性主要由叶片内部结构和生化组分的散射、吸收特征所控制,可以用来估算叶片营养及健康状态指标的含量(Zhao等,2022)。Jiang等(2022)利用红树林叶片高光谱反射率,综合比较了估算精度、灵敏度、抗噪声性能、空间可视化质量等因素,构建出叶片相对叶绿素含量敏感的高光谱指数,得出叶片高光谱红边区域的波段组合能够有效地捕捉叶片叶绿素含量变化的结论。冠层尺度则包含了整株植物的光谱贡献,重点反映植物总光合有效辐射和初级生产力的能力,在产量量化方面具有重要作用(Robles-Zazueta等,2022)。但是,获取的原始光谱数据往往具有极高的空间复杂度。光谱曲线携带的信息中不仅能反映各种组成物质的成分和含量,还客观记录了观测时的温湿度、表面纹理、组织结构参数等非目标成分,再加上大量背景噪声的引入,使得谱峰重叠、吸收强度降低,进而影响模型的估算精度和鲁棒性(Yang等,2022)。因此,对光谱数据进行变换并增强不同波段范围的光谱特征是非常有必要的。目前,通过微分变换提高光谱数据与目标变量之间响应关系是一种常用的手段,如使用一阶和二阶为代表的整数阶微分变换消除光谱曲线中的背景干扰和基线漂移(Peng等,2018)。此外,阶数更高的整数阶也被尝试引入光谱预处理,但是随着阶数的增加,光谱中的高频噪声会进一步放大,信噪比降低,导致原始光谱中 useful 信息丢失或难以提取(Fu等,2019)。随着引入分数阶微分 FOD (Fractional Order Differentiation) 进行信号处理,高光谱最优的微分变换结果并不都在整数阶,也可能在整数阶之间的分数阶(Hasan等,2023)。

分数阶微分不仅细化了光谱间距,在小范围内也放大了微弱的光谱特征,一定程度上反映了光谱信息的变化,能够在整数阶微分之间找到更精细的插值反射光谱(Tian等,2018)。Fu等

(2019)探讨了基于零阶与二阶之间 Grunwald-Letnikov 分数阶微分的光谱预处理效果,通过研究不同分数阶下相关系数的变化趋势,发现分数阶可以显著提高相关系数,深入挖掘了光谱的潜在信息。Hu等(2021)基于近红外光谱分数阶微分研究了橡胶树叶片氮含量的无损估算方法,对比多个阶数建立橡胶树叶片氮含量估算模型,发现 0.6 阶和 1.6 阶比整数阶具有更好的模型估算性能。Cheng等(2021)运用分数阶微分和波段组合构建基于机器学习的地上植被有机碳含量估测模型,结果发现,相对于一阶和二阶微分,分数阶微分能捕捉到更细微的光谱特征,挖掘出海岸带湿地植被冠层光谱在估算地上植被有机碳含量中的应用潜力。

变量选择以特定的波长或波长区间为基础,一方面对模型进行简化,另一方面可得到稳健性好、解释性强的估算模型(Kamruzzaman等,2022)。近年来,变量选择在算法实现方面取得显著进展,发展出不需要引入数学模型仅从变量自身的预测性能出发的过滤法,如相关系数法和方差分析法(Li等,2022);也有需要引入数学模型并随机挑选单个变量或变量组合,再采用交叉验证的方法比较模型评价指标的包装法,如连续投影算法、竞争自适应加权算法(Cheng等,2021;Zheng等,2019);还有嵌入法,其与包装法区别在于不需要通过迭代的方式反复建模来筛选变量,如随机森林的变量重要性排序(Li等,2021b)。由于原始冠层高光谱数据存在较多重复和无用的冗余变量信息,为减少模型输入变量和缩短分析计算时间,需从全波段提取出对建模最有效的波长。

鉴于此,本研究拟以薄壳山核桃为研究对象,基于果实成熟期的薄壳山核桃冠层高光谱数据,运用分数阶微分进行光谱预处理,结合光谱指数定性和定量分析光谱响应关系,再利用变量组合集群分析算法筛选建模变量,最后通过机器学习算法构建最优的薄壳山核桃叶片氮素估算模型,以为大面积薄壳山核桃林遥感监测和施肥量化提供理论依据。

2 研究区与数据

2.1 研究区概况及样本选择

本研究样本数据来自中国安徽省合肥市肥东县白龙镇的薄壳山核桃试验示范基地(117°22'20"E—

117°23′10″E, 32°11′10″N—32°11′30″N)。该区域隶属北亚热带季风气候,光照充足,气候温和,降水适中,年平均降水量约为879.9 mm,生长季节平均气温15.5℃。该区域符合薄壳山核桃引种栽培适宜区条件,实际栽培的薄壳山核桃开花结实已取得成效,并表现出较强的抗逆性。本研究选择基地内具有诸多优良品质(早实、丰产、抗病)和显著经济效益的建德(J5、J35)及长林(C10、C21)系列共4个典型品种,共调查53株。由于基地每年都会在3月和10月施肥一次,因此数据采集时间定为2022年9月,正值薄壳山核桃果实成熟期,调查期前5个月内除必要的除草外,无其他管理措施(如浇水、施肥、打药等)。

2.2 数据获取

2.2.1 光谱数据采集

使用全波段地物光谱仪(ASD FieldSpec4 Wide-Res)捕获紫外光、可见光和近红外光谱(UV-Visible, 350—780 nm; Near Infrared, 780—1000 nm; 光谱分辨率为3 nm)及短波红外光谱(Short Wavelength Infrared-1, 1000—1800 nm; Short Wavelength Infrared-2, 1800—2500 nm; 光谱分辨率为30 nm),光谱仪采集光谱波段范围为350—2500 nm。由于供试薄壳山核桃样本垂直高度普遍高于5 m,而光谱仪自身的铠装光纤长度仅有1.5 m且裸光纤探头视场角为25°,因此将光谱仪铠装光纤通过光纤适配器连接光纤延长线,再使用定制支架将探头举至距离薄壳山核桃东南西北4个方向2 m处进行45°倾斜测量,此时距离地表形成一个直径超过1.8 m的圆锥体形探测空间,这个探测视野基本覆盖单株薄壳山核桃冠体顶层和中层的大部分,图1为野外作业现场图。每个方向光谱重复测10次,每株共40条光谱,在软件中去除异常光谱后用算术平均值作为薄壳山核桃最终的原始冠层光谱特征。

ASD光谱仪可观测350—2500 nm的光谱,输出的重采样波长间隔为1 nm,光谱波段总数达2151个。由于各领域划分光谱范围的标准不同,在具体的波段界线划定上又比较模糊,为避免本文引述相关波段概念出现混乱,借鉴Pu (2017)的划定方案,本研究采用的波段和波长划定详见表1。



图1 野外作业
Fig. 1 Field operations

表1 波段和波长划定情况表
Table 1 Waveband and wavelength delineation table

波段	波长/nm
紫外光	350—400
可见光	400—780
紫光	400—420
蓝光	420—440
靛蓝光	440—490
绿光	490—570
黄光	570—585
橙光	585—620
红光	620—780
红边	670—780
近红外	780—1000
短波红外1	1000—1800
短波红外2	1800—2500

2.2.2 氮素含量测定

单株薄壳山核桃叶片采摘需满足两个原则,即所选叶片落在探测视域和树体破坏最小化。因此,本试验采摘单株薄壳山核桃冠层中上部且分散在倾斜测量探测视域的东南西北4个方位,每个方位9片,总计36片叶子。叶片生长良好、无机械损伤、无病虫害。叶片采摘后立即装入信封袋,

带回实验室放入烘箱在 60 °C 下烘干至恒重后, 研磨过 60 目筛, 筛出的薄壳山核桃叶片粉末使用自动定氮仪利用凯氏定氮法测定叶片氮素含量 LNC (Leaf Nitrogen Content), 单位统一换算为 g/kg (杨贵军 等, 2019)。

2.3 数据处理及分析

2.3.1 分数阶微分

分数阶微分是整数阶微分学的拓展, 其计算方法与整数阶相似, 但阶数被任意扩展为分数。分数阶导数的定义包括 Grunwald-Letnikov 导数、Riemann-Liouville 导数和 Caputo 导数。采用在光谱数据处理和信息提取方面广泛应用的 Grunwald-Letnikov 导数进行光谱数据预处理 (Li 等, 2021a)。

首先观察函数的 n 阶导数 $f(t)$ 公式:

$$\frac{d^n}{dt^n} f(t) = \lim_{h \rightarrow 0} \frac{1}{h^n} \sum_{j=1}^n (-1)^j \binom{n}{j} f(t - jh) \quad (1)$$

式中, t 为微分上界, n 为阶数, h 为微分的步长, j 为求和的索引。

二项式 (1) 展开式可以写成:

$$(1 - z)^n = \sum_{j=0}^n (-1)^j \binom{n}{j} z^j = \sum_{j=0}^n \frac{(-1)^j n!}{j!(n-j)!} z^j \quad (2)$$

式中, z 为用于二项式展开的代数变量, n 为二项式的整数阶次, j 为遍历二项式的系数。

这样易将 n 阶导数公式直接拓展到非整数阶 α 的情形, 和整数阶不同的是, 二项式表达式不再是有限项的和, 而变成了无穷级数的形式, 即:

$$(1 - z)^\alpha = \sum_{j=0}^{\infty} (-1)^j \binom{\alpha}{j} z^j = \sum_{j=0}^{\infty} w_j z^j \quad (3)$$

式中, z 为用于二项式展开的代数变量, α 为二项式的非整数阶次, w_j 为 z^j 项的系数, 决定了每项在多项式中的权重。

拓展的二项式表达式则变成:

$$w_j = (-1)^j \binom{\alpha}{j} = \frac{(-1)^j \tau(\alpha + 1)}{\tau(j + 1) \tau(\alpha - j + 1)} \quad (4)$$

假设 $t \leq t_0$ 时, 函数 $f(t)$ 的值为零, 则无限项的和可以转换成有限项, 如此, 便可引入 Grünwald-Letnikov 分数阶导数公式:

$${}^{GL}D_t^\alpha f(t) = \lim_{h \rightarrow 0} \frac{1}{h^\alpha} \sum_{j=0}^{((t-t_0)/h)} (-1)^j \binom{\alpha}{j} f(t - jh) \quad (5)$$

式中, α 为微分的阶数, t_0 为微分的下界; t 为微分的上界, h 为微分的步长, 本文 $h=1$, τ 为 Gamma

函数。若 $\alpha = 0$ 时, ${}^{GL}D_t^\alpha f(t) = f(t)$ 。若 $\alpha = 1$ 或 2 时, 分别表示原函数的一阶微分和二阶微分。

本部分使用 MATLAB R2022a (MathWorks Inc., USA) 软件中 Xue 等编写的 FOTF 工具箱。本工具箱主要对微分的阶数 α 和微分的上下界 t 和 t_0 三项参数进行调优 (https://ww2.mathworks.cn/matlabcentral/file-exchange/60874-fotf-toolbox?s_tid=srchtitle_FOTF_1 [2023-12-03])。

2.3.2 光谱指数

使用少数几个波段反射率值的简单组合, 可以有效增强高光谱特征对地物理化参数的线性度量, 减弱由于背景因素差异造成的误差和不确定性影响。两波段组合形式的光谱指数将一维波长索引范围的光谱特征空间拓展到二维索引尺度, 充分联合光谱之间的相互关系 (Montero 等, 2023)。经测试, 差值光谱指数 DSI (Difference Spectral Index) 和归一化光谱指数 NDSI (Normalized Difference Spectral Index), 能够综合分析光谱数据与目标变量 (LNC) 的响应关系。因此本研究使用这两种两波段反射率值组合方式, 见式 (6) 和式 (7), 重采样间隔设置 5 nm 最合适。

$$NDSI_\alpha(R_i, R_j) = \frac{R_\alpha(i) - R_\alpha(j)}{R_\alpha(i) + R_\alpha(j)} \quad (6)$$

$$DSI_\alpha(R_i, R_j) = R_\alpha(i) - R_\alpha(j) \quad (7)$$

式中, $NDSI_\alpha(R_i, R_j)$ 和 $DSI_\alpha(R_i, R_j)$ 代表 α 阶下两波段组合构成的归一化光谱指数和差值光谱指数; $R_\alpha(i)$ 和 $R_\alpha(j)$ 分别代表 α 阶预处理后第 i 和第 j 个波段对应的光谱反射率, 索引范围 350—2500 nm。

2.3.3 特征变量筛选——变量组合集群分析

变量组合集群分析 VCPA (Variable Combination Population Analysis) 是一种综合考虑变量间随机组合时相互影响的特征筛选方法, 在解决高维特征难题时起到连续收缩变量空间的作用。VCPA 设计时将非线性迭代偏最小二乘 NIPALS (Nonlinear Iterative Partial Least Squares) 包裹进算法, 融合了指数量递减函数 EDF (Exponentially Decreasing Function)、二进制矩阵采样 BMS (Binary Matrix Sampling) 和模型集群分析 MPA (Model Population Analysis) 3 种方法不断迭代搜索出最优变量子集组合 (Yun 等, 2014)。

2.3.4 基于机器学习构建LNC模型

极端梯度提升 XGBoost (eXtreme Gradient Boosting) 算法是一种对梯度提升决策树 GBDT (Gradient Boosting Decision Tree) 改进的提升集成算法, 在当前全世界的数据挖掘和机器学习竞赛中非常流行。XGBoost通过加入正则项控制模型的复杂度, 提高模型的泛化能力, 防止过拟合。XGBoost能够支持并行运算, 可以极大快速地完成模型训练。此外, 由于在损失函数中添加了正则化项, 因此训练过程中的目标函数由两部分组成: 第1部分是梯度提升算法的损失项, 第2部分是正则化项 (Gao 等, 2022)。本研究主要基于决策树最大深度 (maxdepth) 进行参数调优。目标函数如下式所示:

$$L(\emptyset) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (8)$$

$$\sum_{k=1}^K \Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (9)$$

式中, n 为样本的数量; $l(y_i, \hat{y}_i)$ 为用来表征目标值 y_i 与预测值 $\hat{y}_i$ 之间的损失函数; γ 为手动设置的参数; ω 为决策树中所有叶节点的值构成的向量; T 为叶节点的个数, 即 $\sum_{k=1}^K \Omega(f_k)$ 是合计全部 t 棵树的复杂度。

本文按照 7 : 3 的比例随机划分容量为 53 的薄壳山核桃叶片氮素含量样本数据集, 利用 XGBoost 机器学习算法构建薄壳山核桃 LNC 最优估算模型。由于样本数量非常有限, 本文使用五折交叉验证方法固定随机种子数划分出验证集, 用于模型最优参数的选择和防止过拟合。

2.3.5 模型验证和评估

本文模型验证评价指标采用决定系数 R^2 (Coefficient of Determination) 和均方根误差 RMSE (Root Mean Square Error)。 R^2 越大, RMSE 越小, 说明模型预测效果越好。

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (10)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (11)$$

式中, x 和 y 分别为观测值和预测值, $\bar{x}$ 为观测值的平均值, n 为样本数。

3 结果与分析

3.1 薄壳山核桃叶片氮素含量及冠层高光谱特性分析

3.1.1 薄壳山核桃叶片氮素含量描述性分析

薄壳山核桃 LNC 的统计结果见表 2, 可见整体样本叶片氮素含量为 17.10—46.97 g/kg, 平均值为 26.43 g/kg。

表 2 薄壳山核桃 LNC 描述性统计结果

Table 2 Statistical results of LNC values in *Carya illinoensis* leaves

类别	最大值/ (g/kg)	最小值/ (g/kg)	均值/ (g/kg)	标准差/ (g/kg)	变异 系数/%
值	46.97	17.10	26.43	2.25	8.51

LNC 的 4 个等分区间内薄壳山核桃冠层高光谱平均反射率响应关系表明 (图 2), 处在不同氮素水平的冠层高光谱反射率整体变化趋势呈现一定规律。具体表现为: 可见光谱区 (500—780 nm) LNC 值处在 18.00—20.70 g/kg 和 20.70—25.75 g/kg 的光谱反射率很接近; 而 NIR 谱区 780—1000 nm 随着 LNC 值增加, 反射率逐步降低; 在 SWIR 谱区 1400—1800 nm 和 2000—2400 nm, 随着 LNC 值增加, 反射率逐步升高。这说明氮素调控薄壳山核桃长势变化可能存在过渡阶段, 在 LNC 减少或增加到超出过渡阶段临界值时, 植株长势改变方向有清晰的指向。

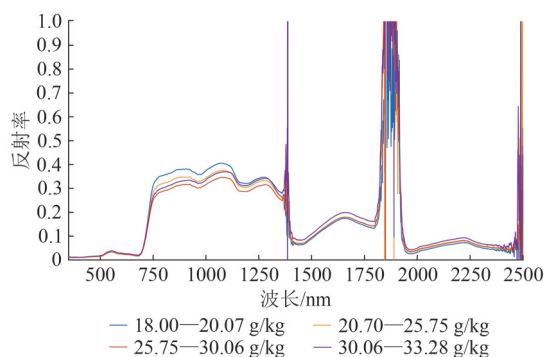


图 2 不同 LNC 梯度的平均冠层高光谱反射率响应关系
Fig. 2 Response relationship of average canopy hyperspectral reflectance with different LNC gradients

3.1.2 薄壳山核桃冠层光谱特性分析

薄壳山核桃冠层原始光谱特性结果见图 3 (a)。

可见冠层原始光谱反射率较低, 原因是冠层光谱特性除了由植物的内部结构和生化组分决定, 冠层结构参数 (叶面积指数、冠层消光系数和平均叶倾角分布均是影响太阳辐射在植物进行重新分配的重

要参数) 也是主要的影响因素 (屈永华 等, 2012), 同时还受大气、植被下垫面、太阳高度角、观测角度和方位等多因素影响 (Luo 等, 2022)。

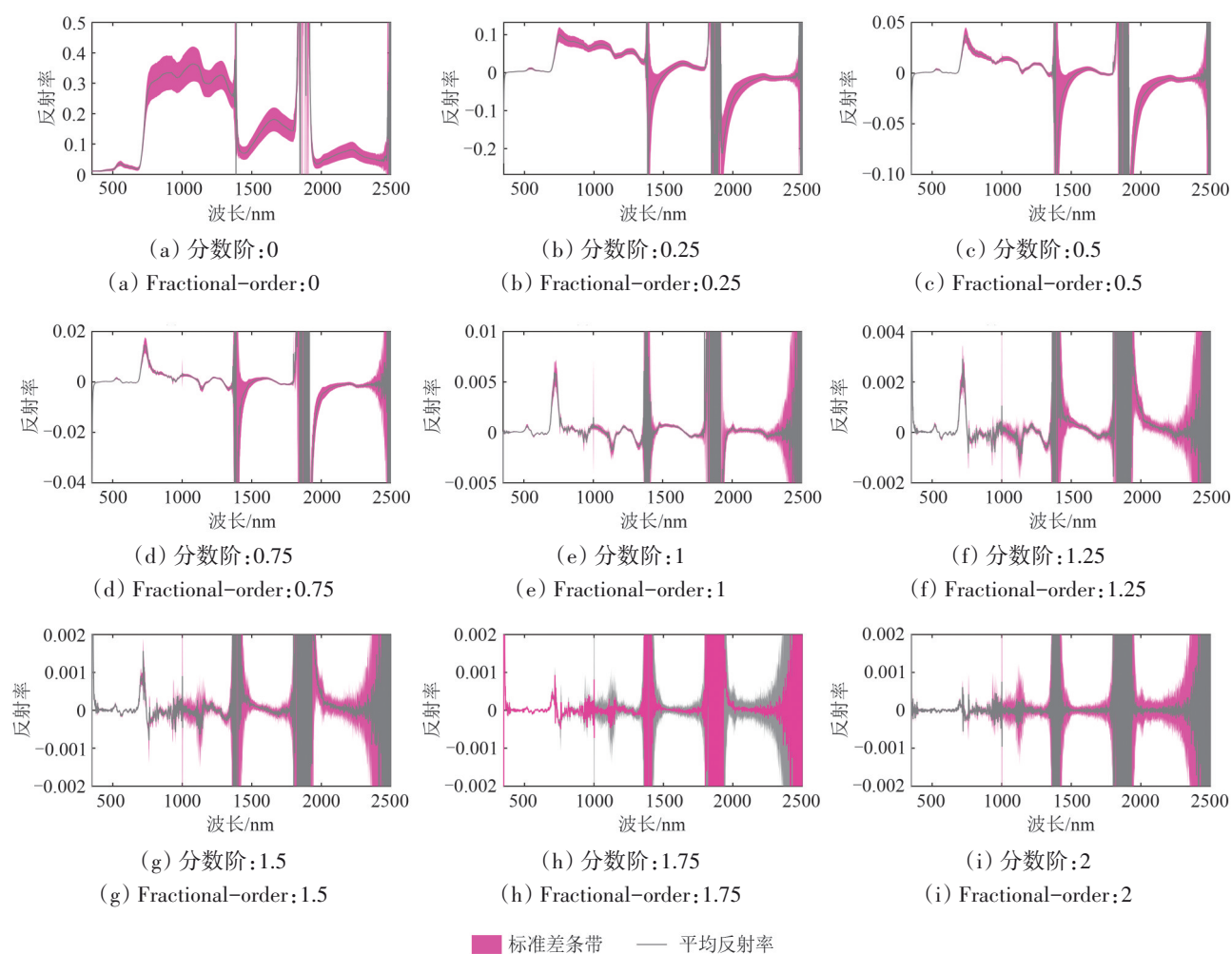


图3 冠层FOD光谱曲线对比

Fig. 3 Comparison of canopy FOD spectral curves

冠层FOD光谱 (0—2阶, 以0.25为增量) 结果表明从0阶到0.75阶, 整体光谱强度减弱了, 0阶光谱中的反射峰转化为若干个反射峰和吸收谷, 在780 nm附近出现光谱峰值; 在1450—1800 nm和1950—2350 nm反射率急速增加; 在0.75阶时1450 nm和1950 nm处反射率增加的斜率达到最大值; 从1阶到1.5阶, 1450—1800 nm和1950—2350 nm反射率呈下降趋势; 从1.5阶到2阶, 光谱曲线形态变化较小, 反射率为 $[-0.01, 0.01]$ (图3)。总体而言, 低阶数光谱能维持与原始 (0阶) 光谱相似的特征, 但是, 随着阶数增加, 光谱反射率减小, 反射强度逐渐稳定, 光谱曲线

形态特征丧失明显, 相应地部分光谱曲线出现明显波动。

3.2 薄壳山核桃冠层FOD光谱与LNC响应关系

3.2.1 冠层FOD单波段光谱与LNC响应关系

LNC与冠层单波段FOD光谱反射率的相关性分析结果见图4。可见: 受冠层结构和背景辐射等影响, 0阶冠层光谱与LNC的相关性较低, 二者绝对相关系数在1832 nm处达到最大 (0.384); 在500—1350 nm, 0阶冠层光谱与LNC呈负相关, 在1400—2400 nm呈正相关; 此外, 分数阶微分预处理后敏感波段有明显变化, 从0.25阶增加到1.25阶

时,冠层FOD光谱与LNC在500—750 nm呈显著正相关,频带数量先增加后减少。由表3可见二者相关系数随着阶数的增加而增加,在1.25阶的1033 nm波长处达到最高(0.536),之后逐渐下降。总体来说,随着分数阶递增,冠层FOD单波段光谱与LNC的敏感波段频带数量先增加后减少,二者相关系数表现出相同特征,且都高于原始(0阶)光谱。同时发现,分数阶微分光谱与LNC的相关性优于一阶微分和二阶微分光谱,这表明,与整数阶微分相比,分数阶微分更能有效地提高单波段光谱与目标变量之间的线性相关。

3.2.2 冠层FOD两波段光谱指数与LNC响应关系

LNC与FOD两波段光谱指数相关分析结果见图5和图6。可见随着阶数增加,FOD结合NDSI与LNC相关性总体呈现“下降—上升—下降—上升”趋势。结合NDSI处理后各微分阶数表现,选取处理效果显著的进行相关性展示。图5(a)中原始冠层光谱结合NDSI相关系数绝对值最大值从单波段处理的0.384提升至0.607;图5(b)中绿光(490—570 nm)和SWIR1(1400—1800 nm)联合

区域相关性有很大下降;图5(c)中NIR(780—1000 nm)和SWIR1(1500—1700 nm)联合区域出现高相关性,且呈现正负相关交替现象;图5(d)中对应联合区域的相关性更为破碎,高相关性区域比1阶处理更狭窄,在可见光区域(500—680 nm)表现较为明显。相比而言,Vis-NIR波段(550—1000 nm)和SWIR1波段(1000—1500 nm)与LNC的线性关联较强,其中550—1500 nm波长组合在0—1阶均有较高的相关性,并且越偏向可见光谱区线性相关越高。

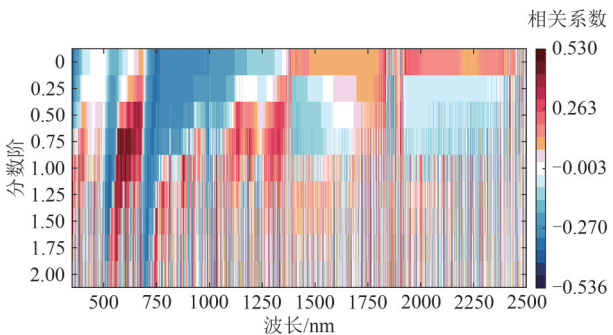
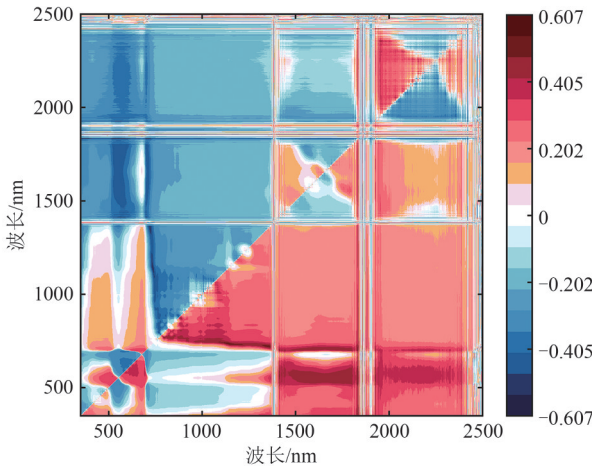


图4 LNC与冠层单波段FOD光谱反射率的相关性分析
Fig. 4 Correlation analysis of LNC and the canopy single-band FOD spectral reflectance

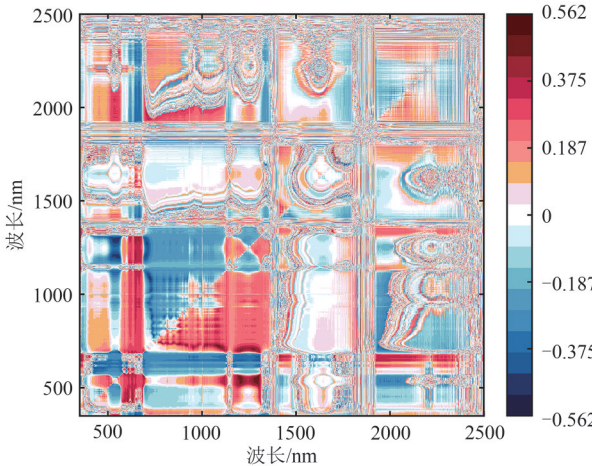
表3 LNC与冠层单波段FOD光谱反射率之间最大相关系数

Table 3 The maximum correlation coefficients between LNC and canopy single-band FOD spectral reflectance										
目标变量	类型	分数阶								
		0	0.25	0.5	0.75	1	1.25	1.5	1.75	2
LNC	最大相关系数	0.384	-0.437	0.400	0.469	-0.527	-0.536*	-0.522	-0.493	0.499
	对应波长/nm	1832	2421	647	584	1033	1033	1033	1033	1034

注: *代表单波段FOD光谱反射率与LNC之间最大相关系数。



(a) 分数阶:0
(a) Fractional-order:0



(b) 分数阶:0.5
(b) Fractional-order:0.5

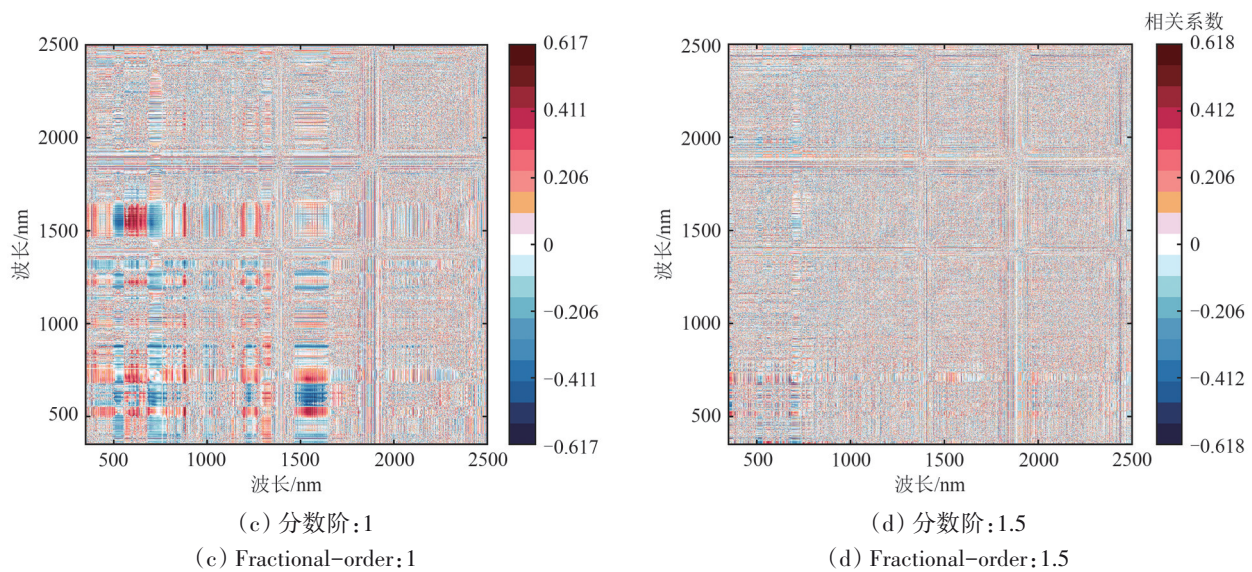


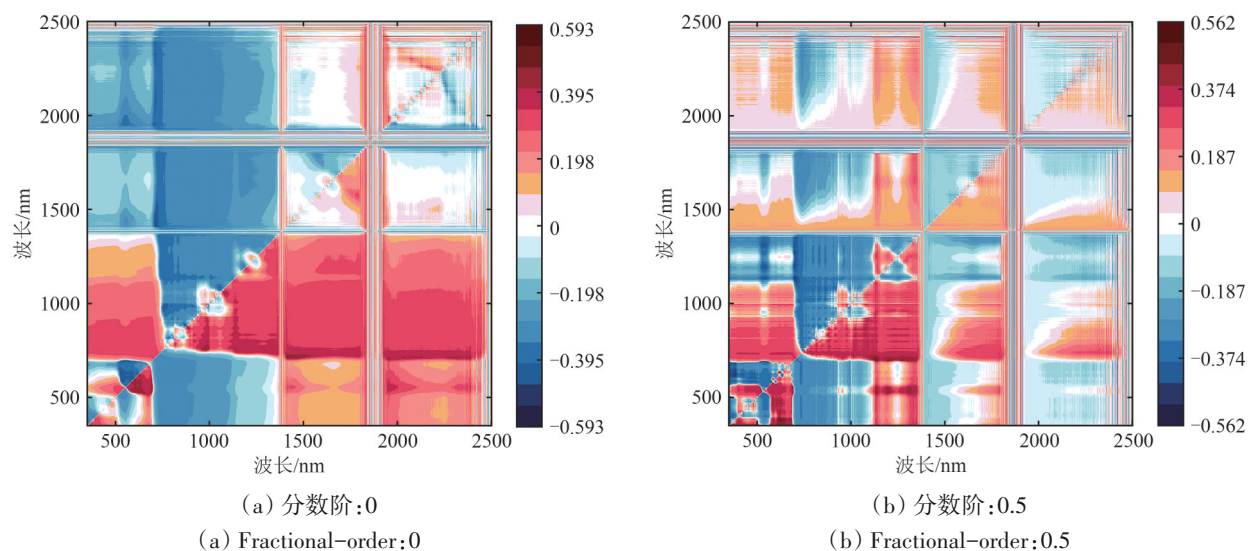
图5 LNC与FOD两波段光谱指数(NDSI)相关分析(横纵坐标分别代表350—2500 nm内的波长 $R(i)$ 和 $R(j)$,

右边的红白蓝颜色条带对应 Pearson 相关系数大小,颜色越深,则表明敏感程度越高,反之越低,下同)

Fig. 5 LNC and FOD two-band spectral index (NDSI) correlation analysis (The horizontal and vertical coordinates represent the wavelengths $R(i)$ and $R(j)$ in the range of 350—2500 nm, respectively, and the red-white-blue color bands on the right correspond to the size of the Pearson correlation coefficients, with the darker colors indicating higher sensitivity and vice versa. The same as below)

FOD结合DSI的相关性分析结果见图6。随着阶数增加,FOD结合DSI与LNC的相关性总体呈现“下降—上升—下降”趋势,与FOD结合NDSI相关性变化趋势略有不同。结合DSI处理后各微分阶数表现,选取处理效果显著的进行相关性展示。图6(a)中原始冠层光谱结合DSI的相关系数绝对值最大值从单波段处理的0.384提升至0.593,略低于NDSI处理的0.607;图6(b)中NIR(780—1000 nm)和SWIR1(1500—1700 nm)联合区域

出现正负相关性交替地带;图6(c)中可见光区域出现强相关条带,且总体上相关性呈现“十字交叉”的条纹分布;图6(d)中1.5阶处理下条纹分布比1.0阶分布更为密集,然而可见光区域仍然显现一条强相关条带。Vis-NIR波段(550—1000 nm)和SWIR1波段(1000—1500 nm)与LNC的线性关联也较强,这与FOD结合NDSI处理表现一致。



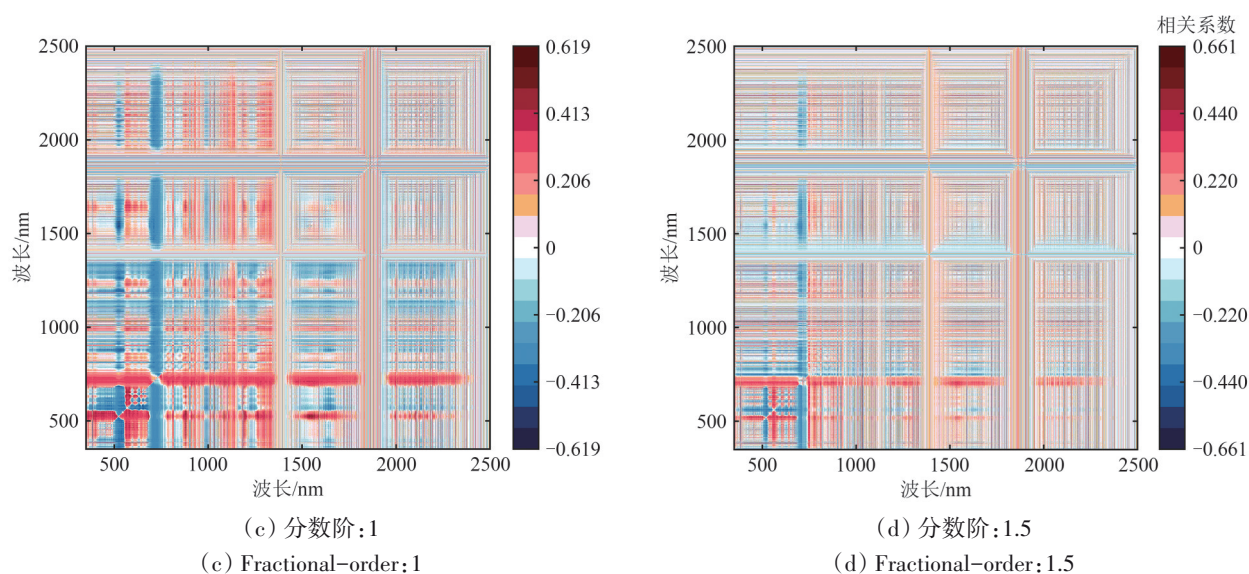


图6 LNC与FOD两波段光谱指数(DSI)相关分析

Fig. 6 LNC and FOD two-band spectral index (DSI) correlation analysis

综上,光谱反射率可以反映植被对光吸收的能力,吸收越多,光谱反射率越低,植被光能利用率就越高,而植被对光谱的吸收能力与叶片生化特性等有密切关系。如在可见光与近红外光波长范围内,叶绿素含量是影响植被光谱反射率的主导因素,氮素作为植物体内重要的营养元素,是叶绿素的主要组成成分,因此,氮素含量直接影响植物光系统活性,进而表现出氮素含量与可见光范围的光谱反射率呈负相关的特征,与近红外范围光谱反射率呈现正相关特征,这与油茶、苹果的叶片光谱采用微分处理的细节信息等研究结果一致(杨迈等,2024)。另一方面,氮素含量的变化会引起植物冠层和叶片形态结构的变化,从而引起光谱特征变化,导致光谱反射率与氮素高度相关。值得注意的是,FOD结合两波段光谱指数(NDSI和DSI)与LNC之间的最大绝对相关系数均高于原始(0阶)冠层两波段光谱指数。甚至在分数阶微分(1.5阶)处理下的冠层两波段光谱指数与LNC的相关性优于整数阶微分(1阶和2阶)光谱指数,这进一步说明了分数阶微分有提高光谱特征与目标变量(LNC)之间的线性相关的特点。

3.3 基于VCPA算法的光谱变量筛选

LNC的FOD冠层光谱及FOD与两波段光谱指数组合,再经过VCPA算法筛选出了最优变量组合子集,具体信息包括每个处理类型对应保留变量

个数和具体的人选波长位置(表4和表5)。由表4可知,经过分数阶微分预处理后保留变量数有所增加,这表明使用预处理方法可以提升VCPA挖掘潜在信息的能力;同时,从入选波长位置来看,LNC优选波段主要位于可见光谱区(Red)、NIR和SWIR2波段。但是,组成最优变量子集的波长并非都位于线性相关程度较高的敏感区间,如1033 nm、1535 nm、2140 nm等,进一步说明对光谱变量的选择不能简单地以最大相关系数所在波长为单一输入变量。由表5可知,FOD结合两波段光谱指数保留变量数也呈现出与表4相同的趋势,该结果证实,预处理结合光谱指数可以进一步提升VCPA筛选变量的潜力。与此同时,通过比较FOD结合NDSI与FOD结合DSI变量数发现,前者的变量数要略少于后者,这一现象可能与两种处理方式下对应的相关性分布差异有关,进一步表明特征筛选受不同处理方式的影响较大。LNC的两波段光谱指数在入选波长组合位置分布上呈现多数处在可见光谱区(Green、Red)、NIR和SWIR2联合区域,且随着分数阶增加,更多的NIR和SWIR2联合区域被选为特征变量。

3.4 基于薄壳山核桃冠层FOD光谱的LNC模型估算

LNC的FOD冠层光谱及FOD联合两种两波段光谱指数形式组合对应的27种机器学习模型(XGBoost)评价结果见表6。模型指标内容包括训

练集和预测集的 R^2 、RMSE。LNC 模型在训练集的表现 R_t^2 为 0.48—0.89, RMSET 为 1.12—2.37 g/kg, 在预测集的表现 R_p^2 为 0.21—0.75, RMSEP 为 1.32—3.07 g/kg, 可以看出, 不同组合类型得到模型的估算性能存在明显差异。

表 4 LNC 的 FOD 光谱反射率的最优变量组合子集结果
Table 4 The optimal variable combination subset results of FOD spectral reflectance of LNC

分数阶	保留变量数	入选波长位置/nm
0 阶	4	351、430、904、1515
0.25 阶	6	351、401、717、1103、1141、1535
0.5 阶	5	350、715、851、973、1332
0.75 阶	5	717、817、830、915、1138
1 阶	7	389、705、782、815、879、892、1332
1.25 阶	7	436、706、748、753、880、1332、2140
1.5 阶	5	748、758、888、1445、2294
1.75 阶	5	697、758、908、1034、2140
2 阶	6	350、717、757、924、1034、2140

表 5 LNC 的 FOD 联合两波段光谱指数的最优变量组合子集结果
Table 5 The optimal variable combination subset result of FOD joint two-band spectral indices of LNC

分数阶	保留变量数	入选波长位置/nm
0 阶-NDSI	4	(386,384)、(602,601)、(681,680)、(790,766)
0.25 阶-NDSI	5	(671,426)、(1474,695)、(1773,379)、(2055,550)、(2140,537)
0.5 阶-NDSI	5	(605,428)、(669,571)、(1157,693)、(1336,570)、(1591,1316)
0.75 阶-NDSI	4	(586,438)、(609,406)、(947,370)、(1482,779)
1 阶-NDSI	6	(585,505)、(790,389)、(879,557)、(974,891)、(1673,736)、(2071,703)
1.25 阶-NDSI	4	(754,545)、(898,513)、(917,441)、(992,557)
1.5 阶-NDSI	5	(531,458)、(615,508)、(804,643)、(876,379)、(2255,1347)
1.75 阶-NDSI	7	(653,540)、(731,556)、(811,406)、(920,681)、(961,822)、(995,824)、(2306,1674)
2 阶-NDSI	6	(580,520)、(797,493)、(826,650)、(910,849)、(987,857)、(2306,1475)
0 阶-DSI	5	(362,353)、(567,548)、(613,607)、(981,907)、(1010,789)
0.25 阶-DSI	5	(410,378)、(555,520)、(830,756)、(1128,1110)、(2186,2181)
0.5 阶-DSI	9	(403,370)、(547,511)、(879,830)、(979,735)、(1033,1014)、(1164,1154)、(1348,1343)、(2193,2192)、(2269,2250)
0.75 阶-DSI	7	(818,350)、(978,530)、(1033,1013)、(1343,1328)、(1795,1138)、(2139,2125)、(2186,2146)
1 阶-DSI	8	(487,401)、(521,469)、(996,777)、(1033,1013)、(1332,1330)、(1623,1589)、(2139,2125)、(2272,2152)
1.25 阶-DSI	7	(576,521)、(745,350)、(996,401)、(1245,1015)、(1563,1015)、(2107,2080)、(2264,2193)
1.5 阶-DSI	5	(702,381)、(853,747)、(1738,1033)、(2140,2082)、(2266,2239)
1.75 阶-DSI	8	(521,350)、(1257,1015)、(1332,1092)、(1515,1332)、(1650,1034)、(2140,2083)、(2223,2126)、(2266,2145)
2 阶-DSI	9	(1092,1034)、(1332,1092)、(1544,1009)、1650、1332)、(1719,1192)、(2126,2108)、(2140,2083)、(2223,2145)、(2316,2274)

综合分析经过冠层 FOD 光谱与 FOD 联合两波段光谱指数 (NDSI 和 DSI) 处理后的提升效果, FOD (0—2 阶) 处理中 R_t^2 为 0.40—0.67, R_p^2 为 0.21—0.59; FOD 结合 NDSI 处理的 R_t^2 为 0.64—0.89, R_p^2 为 0.41—0.73; FOD 结合 DSI 处理的 R_t^2 为 0.57—0.89, R_p^2 为 0.50—0.75, 表明随着光谱波段维数的

增加, LNC 模型估算精度上限会随之提高。冠层 FOD (1.25 阶) 光谱与 FOD 联合两波段光谱指数 (1.5 阶-NDSI 和 1.5 阶-DSI) 处理后的模型估算精度相比较一阶和二阶微分更高, 整体估算性能也更好。总体上来看, 分数阶与光谱指数 (NDSI 和 DSI) 组合的 R^2 平均值分别为 0.67 与 0.69, 可以得

出分数阶结合 DSI 对 LNC 整体估算性能的提升更加优越。

表 6 基于 XGBoost 的 LNC 估算模型评价结果
Table 6 Evaluation results of LNC estimation models based on XGBoost

组合类型	训练集		预测集	
	R^2_T	RMSET/(g/kg)	R^2_P	RMSEP/(g/kg)
0 阶	0.48	2.21	0.31	3.07
0.25 阶	0.40	2.37	0.35	2.98
0.5 阶	0.53	2.09	0.45	2.75
0.75 阶	0.58	1.99	0.22	3.28
1 阶	0.67	1.85	0.46	2.49
1.25 阶	0.63	2.13	0.59	1.68
1.5 阶	0.43	2.31	0.39	2.90
1.75 阶	0.53	2.10	0.39	2.91
2 阶	0.59	1.96	0.21	3.31
0 阶-NDSI	0.66	2.05	0.47	1.91
0.25 阶-NDSI	0.78	1.63	0.60	1.68
0.5 阶-NDSI	0.89	1.12	0.71	1.41
0.75 阶-NDSI	0.64	1.91	0.55	2.28
1 阶-NDSI	0.78	1.62	0.67	1.51
1.25 阶-NDSI	0.73	1.65	0.64	2.04
1.5 阶-NDSI	0.84	1.38	0.73	1.36
1.75 阶-NDSI	0.77	1.52	0.51	2.38
2 阶-NDSI	0.75	1.61	0.41	2.59
0 阶-DSI	0.57	2.00	0.50	2.62
0.25 阶-DSI	0.74	1.80	0.52	1.83
0.5 阶-DSI	0.76	1.51	0.60	2.36
0.75 阶-DSI	0.85	1.37	0.65	1.55
1 阶-DSI	0.77	1.47	0.44	2.78
1.25 阶-DSI	0.86	1.30	0.73	1.37
1.5 阶-DSI	0.89	1.12	0.75	1.32
1.75 阶-DSI	0.76	1.63	0.67	1.87
2 阶-DSI	0.81	1.53	0.63	1.60

为了进一步评价 LNC 最优模型的估测值与实测值间的变化情况，根据样本实测值和估测值制作 FOD 冠层光谱与 FOD 联合两波段光谱指数 (NDSI 和 DSI) 最优模型的散点图 (图 7—图 9)。由图 7 可知，1.25 阶光谱的 LNC 估算模型在训练集中 $R^2_T=0.63$ ，预测集中 $R^2_P=0.59$ ，表明模型能够解释训练样本中 63% 的变异程度，但模型在预测集中的表现略低。与此同时，观察 95% 置信区间的倾斜方向发现，估算 LNC 时会出现低值高估和高值

低估的现象。由图 8 可知，1.5 阶结合 NDSI 的 LNC 估算模型在训练集中 $R^2_T=0.84$ ，预测集中 $R^2_P=0.73$ ，模型表现优于 1.25 阶光谱的 LNC 估算模型，95% 置信区间的倾斜方向也出现了低值高估和高值低估的现象。由图 9 可知，1.5 阶结合 DSI 的 LNC 估算模型在训练集中的评估指标 $R^2_T=0.89$ ，RMSET 为 1.12 g/kg，在预测集中的 $R^2_P=0.75$ ，RMSEP 为 1.32 g/kg，分别解释薄壳山核桃 LNC 训练样本中 89% 和预测样本 75% 的变异程度，意味着该模型能够较好地通过冠层高光谱数据的特征信息准确地反映薄壳山核桃 LNC 的变化。模型的估测值与实测值散点均匀地分布在标准 1 : 1 线附近。

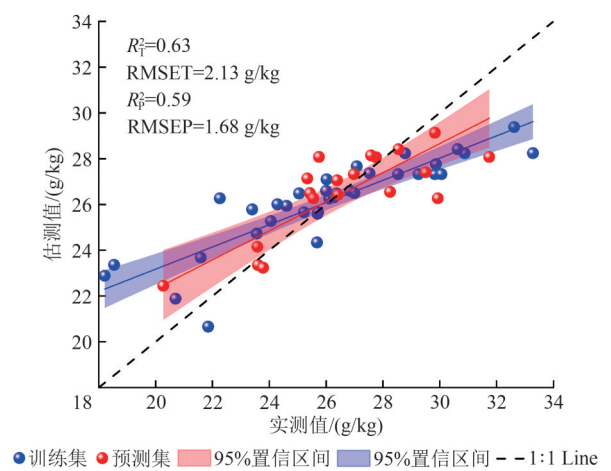


图 7 基于 FOD 光谱反射率的 LNC 实测值与估测值的 XGBoost 模型散点图
Fig. 7 XGBoost model scatterplot of the measured and estimated LNC values based on FOD spectral reflectance

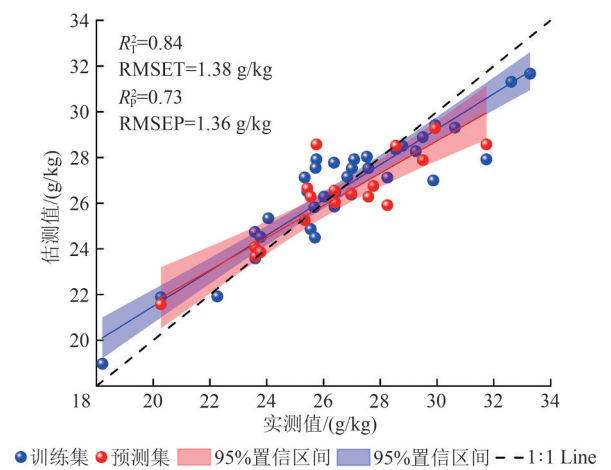


图 8 基于 FOD 两波段光谱指数(NDSI)的 LNC 实测值与估测值的 XGBoost 模型散点图
Fig. 8 XGBoost model scatterplot of the measured and estimated LNC values based on FOD two-band spectral index (NDSI)

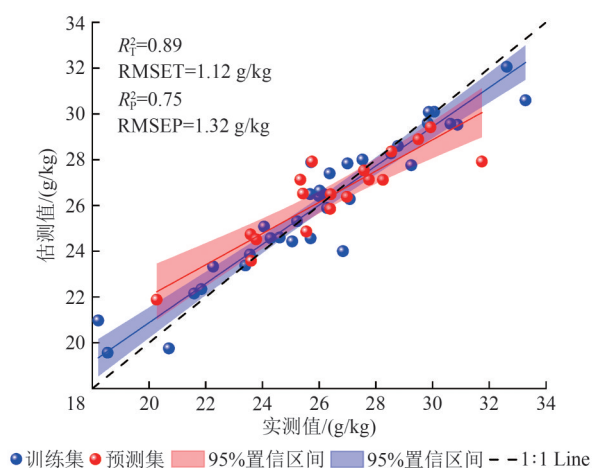


图9 基于FOD两波段光谱指数(DSI)的LNC实测值与估测值的XGBoost模型散点图

Fig. 9 XGBoost model scatterplot of the measured and estimated LNC values based on FOD two-band spectral index (DSI)

4 讨论

与传统单一的一阶导、二阶导等高光谱预处理相比, 本文综合运用多种分数阶微分联合光谱指数定量分析光谱特征与目标变量的响应关系, 进一步结合变量组合集群分析算法构建最优薄壳山核桃叶片氮素估算模型。

发现经过分数阶微分变换后, 光谱反射率能提供比原始光谱精度更高的分辨率和更清晰的光谱轮廓, 并增强了反射率与植物属性间的相关性(刘浩等, 2023)。对比原始光谱和FOD光谱与薄壳山核桃LNC的相关性分析, 不同分数阶处理下薄壳山核桃冠层光谱与目标变量的相关性均显著提高, 说明分数阶微分算法能够更好的表征光谱的细节信息。Hong等(2019)的结果也证实, FOD处理过的特定波长光谱与石油有机物的相关性比与原始光谱的相关性显著提升。

由于不同营养物质的叠加作用以及冠层结构参数的影响, 单个波段反射率的作用有限, 使用两个或多个波段的光谱反射率的比值经过线性或非线性组合构造的光谱指数, 将光谱的重要信息压缩到一个光谱指数通道, 能够有效减少背景效应, 增强光谱特征以提高对目标变量的敏感程度(刘爽等, 2021)。对比分数阶微分与单波段结合, 两波段光谱指数(NDSI和DSI)的结合与薄壳山核桃LNC的相关性提升效果更优。FOD与不同代数运算形式的光谱指数组合有效地增强光谱变换特征对薄壳山核桃LNC的线性度量, 很重要的原

因也就是不同方法之间形成优势互补, 减少干扰因子对所构建的光谱指数产生影响, 这个结果与Chen等(2022)的研究相吻合。

随着微分阶数的提高, 不论是单波段光谱还是两波段光谱指数(NDSI和DSI)与薄壳山核桃LNC的相关性与模型精度整体均呈现出先升后降的趋势。究其原因, 分数阶微分是整数阶微分变换的拓展与延伸, 可以提取整数阶微分无法表征的渐变信息。但是, 随着微分阶数的提高, 背景噪声被逐渐削弱, 高频噪声会被逐渐放大, 也减少了反射率数据中的潜在敏感信息, 致使光谱信息信噪比降低, 进而影响相关性与模型精度(向友珍等, 2023)。同时发现, 分数阶结合单波段反射率在1.25阶微分下与LNC相关性及其构建的预测模型精度均高于1阶与2阶整数微分。两波段光谱指数(NDSI和DSI)与分数阶1.5阶结合相关性及其构建的预测模型精度也均高于1阶与2阶整数微分, 这与Hu等(2021)的研究结果一致, 为进一步探索高光谱反射率分数阶微分变换提供参考。

20世纪60—70年代, 美国农业部(USDA)研究人员在实验室条件下通过捣碎多种植物干叶片, 获得400—2400 nm内42个生物化学成分相应的光谱吸收特征, 这奠定了光谱遥感估测叶片生化组分的基础。研究结果表明, 可见光谱区中Green波段对应的生化组分主要是叶绿素, Red主要是淀粉和蛋白质, NIR主要是氮素、淀粉、蛋白质、纤维素和木质素, SWIR2波段主要是氮素、淀粉、蛋白质和纤维素(刘良云, 2021)。因此, 氮在Green、Red范围的光谱特征实质上是氮素与叶绿素之间存在强相关性进行表达的, 在NIR和SWIR2范围的光谱特征则是通过淀粉、蛋白质、纤维素和木质素等含氮有机物所表达。另一方面氮素含量的变化会引起植物冠层和叶片形态结构变化进而引起光谱特征变化, 如NIR受叶片叶肉细胞排列和冠层结构的影响。我们的研究结果也证实了这一观点, LNC优选的单波段主要位于可见光谱区的Red、NIR和SWIR2, LNC的两波段组合在入选波长位置多数呈现在可见光谱区的Green和Red、NIR、SWIR2的联合区域, 这一结论与王仁红等(2014)认为小麦生育中后期冠层氮素反应敏感且受组织结构因素影响较小的观点相呼应。机器学习往往能很好的解释非线性关系, 但是,

模型的训练需要大量样本和适合的超参数。本文采用集成学习算法进行估算,有效地规避了小样本建模估算的问题,结合更多的理论和数据可能会带来更令人期待的估算结果,实现模型估测精确性和普适性的有效统一。

本文仅选择果实成熟期的薄壳山核桃建立了LNC估测模型,随着生长地域、品种、物候期的变化,冠层光谱与氮素含量的响应关系也会发生变化,该模型是否适用于薄壳山核桃不同品种、不同生育期的LNC的预测还需进一步验证。此外,本文利用地面高光谱特征探讨了薄壳山核桃冠层光谱尺度LNC估测中的可行性,随着样本数据和研究区域的增加,使用多光谱与无人机高光谱等多源遥感手段监测大面积核桃林地的养分也将成为可能。

5 结 论

本研究基于果实成熟期的薄壳山核桃冠层高光谱数据,综合运用分数阶微分进行光谱预处理,联合光谱指数定性和定量分析光谱响应关系,进一步结合变量组合集群分析算法筛选建模变量,最后通过机器学习算法构建最优的薄壳山核桃LNC估算模型。主要结论如下:

(1) 分数阶微分可以减少基线效应,消除背景噪音,通过细化光谱间距放大微弱的光谱吸收特征,继而提高光谱与目标变量的响应关系。与原始光谱相比,FOD处理后的光谱与LNC的相关性均大幅度的提高。FOD结合两波段光谱指数比单波段在提高光谱特征与目标成分的相关性效果更佳。

(2) 基于冠层高光谱分数阶微分算法,结合单波段和两波段组合的光谱指数(NDSI和DSI)及变量筛选(VCPA)构建了27个薄壳山核桃LNC的XGBoost模型。其中,分数阶结合单波段光谱或两波段光谱指数相比整数阶模型效果都更好,且分数阶微分结合两波段光谱指数模型精度优于单波段光谱。最优的氮素模型是1.5阶微分变换结合DSI模型,模型预测集 $R_p^2=0.75$,RMSEP=1.32 g/kg。

参考文献(References)

Berger K, Verrelst J, Féret J B, Wang Z H, Woche M, Strathmann M, Danner M, Mauser W and Hank T. 2020. Crop nitrogen monitoring: recent progress and principal developments in the context of

imaging spectroscopy missions. Remote Sensing of Environment, 242: 111758 [DOI: 10.1016/j.rse.2020.111758]

Chen L H, Lai J, Tan K, Wang X, Chen Y and Ding J W. 2022. Development of a soil heavy metal estimation method based on a spectral index: combining fractional-order derivative pretreatment and the absorption mechanism. Science of the Total Environment, 813: 151882 [DOI: 10.1016/j.scitotenv.2021.151882]

Cheng H, Wang J, Du Y K, Zhai T L, Fang Y and Li Z H. 2021. Exploring the potential of canopy reflectance spectra for estimating organic carbon content of aboveground vegetation in coastal wetlands. International Journal of Remote Sensing, 42(10): 3850-3872 [DOI: 10.1080/01431161.2021.1883201]

de Araújo A R, Sampaio G R, da Silva L R, Portal V L, Markoski M M, de Quadros A S, Rogero M M, da Silva Torres E A F and Marcadenti A. 2021. Effects of extra virgin olive oil and pecans on plasma fatty acids in patients with stable coronary artery disease. Nutrition, 91-92: 111411 [DOI: 10.1016/j.nut.2021.111411]

Fu C B, Gan S, Yuan X P, Xiong H G and Tian A H. 2019. Impact of fractional calculus on correlation coefficient between available potassium and spectrum data in ground hyperspectral and Landsat 8 image. Mathematics, 7(6): 488 [DOI: 10.3390/math7060488]

Gao P, Xu W, Yan T Y, Zhang C, Lv X and He Y. 2019. Application of near-infrared hyperspectral imaging with machine learning methods to identify geographical origins of dry narrow-leaved oleaster (*Elaeagnus angustifolia*) fruits. Foods, 8(12): 620 [DOI: 10.3390/foods8120620]

Gao W L, Zhou L, Liu S Q, Guan Y, Gao H and Hu J J. 2022. Machine learning algorithms for rapid estimation of holocellulose content of poplar clones based on Raman spectroscopy. Carbohydrate Polymers, 292: 119635 [DOI: 10.1016/j.carbpol.2022.119635]

Hasan U, Jia K, Wang L, Wang C Y, Shen Z Q, Yu W J, Sun Y S, Jiang H, Zhang Z C, Guo J F, Wang J Z and Li D. 2023. Retrieval of leaf chlorophyll contents (LCCs) in litchi based on fractional order derivatives and VCPA-GA-ML algorithms. Plants, 12(3): 501 [DOI: 10.3390/plants12030501]

Hong Y S, Chen S C, Liu Y L, Zhang Y, Yu L, Chen Y Y, Liu Y F, Cheng H and Liu Y. 2019. Combination of fractional order derivative and memory-based learning algorithm to improve the estimation accuracy of soil organic matter by visible and near-infrared spectroscopy. CATENA, 174: 104-116 [DOI: 10.1016/j.catena.2018.10.051]

Hu W F, Tang R N, Li C, Zhou T, Chen J and Chen K. 2021. Fractional order modeling and recognition of nitrogen content level of rubber tree foliage. Journal of Near Infrared Spectroscopy, 29(1): 42-52 [DOI: 10.1177/0967033520966693]

Jiang X P, Zhen J N, Miao J, Zhao D M, Shen Z, Jiang J C, Gao C J, Wu G F and Wang J J. 2022. Newly-developed three-band hyperspectral vegetation index for estimating leaf relative chlorophyll content of mangrove under different severities of pest and disease. Ecological Indicators, 140: 108978 [DOI: 10.1016/j.ecolind.2022.108978]

Kamruzzaman M, Kalita D, Ahmed T, Elmasry G and Makino Y. 2022. Effect of variable selection algorithms on model performance for

- predicting moisture content in biological materials using spectral data. *Analytica Chimica Acta*, 1202: 339390 [DOI: 10.1016/j.aca.2021.339390]
- Li C C, Wang Y L, Ma C Y, Ding F, Li Y C, Chen W N, Li J B and Xiao Z. 2021a. Hyperspectral estimation of winter wheat leaf area index based on continuous wavelet transform and fractional order differentiation. *Sensors*, 21(24): 8497 [DOI: 10.3390/s21248497]
- Li M G, Xu Y Y, Men J, Yan C H, Tang H S, Zhang T L and Li H. 2021b. Hybrid variable selection strategy coupled with random forest (RF) for quantitative analysis of methanol in methanol-gasoline via Raman spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 251: 119430 [DOI: 10.1016/j.saa.2021.119430]
- Li X K, Li Z Y, Yang Z Y, Qiu D, Li J M and Li B Q. 2022. A hybrid variable selection and modeling strategy for the determination of target compounds in different spectral datasets. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 275: 121123 [DOI: 10.1016/j.saa.2022.121123]
- Liu H, Yang X Z, Zhang B, Huang J L, Zhao X, Wu Y X, Xiang Y Z, Geng H S, Chen H R and Chen J Y. 2023. Estimation model of soil moisture content in root domain of winter wheat using a fractional-order differential spectral index. *Transactions of the Chinese Society of Agricultural Engineering*, 39(13): 131-140 (刘浩, 杨锡震, 张蓓, 黄嘉亮, 赵笑, 吴雨箫, 向友珍, 耿宏锁, 陈皓锐, 陈俊英. 2023. 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. *农业工程学报*, 39(13): 131-140) [DOI: 10.11975/j.issn.1002-6819.202212068]
- Liu L Y. 2021. *Hyperspectral Remote Sensing of Vegetation*. Wuhan: Hubei Science and Technology Press (刘良云. 2021. 植被高光谱遥感. 武汉: 湖北科学技术出版社)
- Liu S, Yu H Y, Zhang J H, Zhou H G, Kong L J, Zhang L, Dang J M and Sui Y Y. 2021. Study on inversion model of chlorophyll content in soybean leaf based on optimal spectral indices. *Spectroscopy and Spectral Analysis*, 41(6): 1912-1919 (刘爽, 于海业, 张郡赫, 周海根, 孔丽娟, 张蕾, 党敬民, 隋媛媛. 2021. 基于最优光谱指数的大豆叶片叶绿素含量反演模型研究. *光谱学与光谱分析*, 41(6): 1912-1919) [DOI: 10.3964/j.issn.1000-0593(2021)06-1912-08]
- Liu W J, Li Y J, Tomasetto F, Yan W Q, Tan Z F, Liu J and Jiang J M. 2022. Non-destructive measurements of *Toona sinensis* chlorophyll and nitrogen content under drought stress using near infrared spectroscopy. *Frontiers in Plant Science*, 12: 809828 [DOI: 10.3389/fpls.2021.809828]
- Luo S J, Jiang X Q, Yang K L, Li Y J and Fang S H. 2022. Multispectral remote sensing for accurate acquisition of rice phenotypes: impacts of radiometric calibration and unmanned aerial vehicle flying altitudes. *Frontiers in Plant Science*, 13: 958106 [DOI: 10.3389/fpls.2022.958106]
- Mirzaei M, Marofi S, Abbasi M, Solgi E, Karimi R and Verrelst J. 2019. Scenario-based discrimination of common grapevine varieties using in-field hyperspectral data in the western of Iran. *International Journal of Applied Earth Observation and Geoinformation*, 80: 26-37 [DOI: 10.1016/j.jag.2019.04.002]
- Montero D, Aybar C, Mahecha M D, Martinuzzi F, Söchting M and Wieneke S. 2023. A standardized catalogue of spectral indices to advance the use of remote sensing in earth system research. *Scientific Data*, 10(1): 197 [DOI: 10.1038/s41597-023-02096-0]
- Peng Y, Fan M, Wang Q H, Lan W J and Long Y T. 2018. Best hyperspectral indices for assessing leaf chlorophyll content in a degraded temperate vegetation. *Ecology and Evolution*, 8(14): 7068-7078 [DOI: 10.1002/ece3.4229]
- Pu R L. 2017. *Hyperspectral Remote Sensing: Fundamentals and Practices*. Boca Raton: CRC Press
- Qu Y H, Wang J D, Dong J and Jiang F B. 2012. Design and experiment of crop structural parameters automatic measurement system. *Transactions of the Chinese Society of Agricultural Engineering*, 28(2): 160-165 (屈永华, 王锦地, 董健, 姜富斌. 2012. 农作物冠层结构参数自动测量系统设计与试验. *农业工程学报*, 28(2): 160-165) [DOI: 10.3969/j.issn.1002-6819.2012.02.028]
- Robles-Zazueta C A, Pinto F, Molero G, Foulkes M J, Reynolds M P and Murchie E H. 2022. Prediction of photosynthetic, biophysical, and biochemical traits in wheat canopies to reduce the phenotyping bottleneck. *Frontiers in Plant Science*, 13: 828451 [DOI: 10.3389/fpls.2022.828451]
- Shu M Y, Zhou L, Chen H C, Wang X Q, Meng L and Ma Y T. 2022. Estimation of amino acid contents in maize leaves based on hyperspectral imaging. *Frontiers in Plant Science*, 13: 885794 [DOI: 10.3389/fpls.2022.885794]
- Simkin A J, Kapoor L, Doss C G P, Hofmann T A, Lawson T and Ramamoorthy S. 2022. The role of photosynthesis related pigments in light harvesting, photoprotection and enhancement of photosynthetic yield in planta. *Photosynthesis Research*, 152(1): 23-42 [DOI: 10.1007/s11120-021-00892-6]
- Tian A H, Zhao J S, Xiong H G, Gan S and Fu C B. 2018. Application of fractional differential calculation in pretreatment of saline soil hyperspectral reflectance data. *Journal of Sensors*, 2018: 8017614 [DOI: 10.1155/2018/8017614]
- Wang R H, Song X Y, Li Z H, Yang G J, Guo W S, Tan C W and Chen L P. 2014. Estimation of winter wheat nitrogen nutrition index using hyperspectral remote sensing. *Transactions of the Chinese Society of Agricultural Engineering*, 30(19): 191-198 (王仁红, 宋晓宇, 李振海, 杨贵军, 郭文善, 谭昌伟, 陈立平. 2014. 基于高光谱的冬小麦氮素营养指数估测. *农业工程学报*, 30(19): 191-198) [DOI: 10.3969/j.issn.1002-6819.2014.19.023]
- Xiang Y Z, Wang X, An J Q, Tang Z J, Li W Y and Shi H Z. 2023. Estimation of leaf area index of soybean based on fractional order differentiation and optimal spectral index. *Transactions of the Chinese Society for Agricultural Machinery*, 54(9): 329-342 (向友珍, 王辛, 安嘉琪, 唐子竣, 李汪洋, 史鸿棹. 2023. 基于分数阶微分 and 最优光谱指数的大豆叶面积指数估算. *农业机械学报*, 54(9): 329-342) [DOI: 10.6041/j.issn.1000-1298.2023.09.033]
- Yang C B, Feng M C, Song L F, Jing B H, Xie Y K, Wang C, Yang W D, Xiao L J, Zhang M J and Song X Y. 2022. Study on hyperspectral monitoring model of soil total nitrogen content based on fractional-order derivative. *Computers and Electronics in Agriculture*, 201: 107307 [DOI: 10.1016/j.compag.2022.107307]

- Yang G J, Zhao C J and Li Z H. 2019. Quantitative Remote Sensing of Crop Nitrogen Nutrition and Its Application. Beijing: Science Press (杨贵军, 赵春江, 李振海. 2019. 作物氮素定量遥感与应用. 北京: 科学出版社)
- Yang M, Li X L, Zheng Y, Ou C R and Sun S X. 2024. Estimation modelling of nitrogen content of Wetland plants in Dian-chi Lake Basin using hyperspectral technology. Journal of Southwest Forestry University, 44(5): 80-89 (杨迈, 李晓琳, 郑毅, 欧朝蓉, 孙仕仙. 2024. 基于高光谱技术的滇池流域湿地植物氮含量估算模型构建. 西南林业大学学报, 44(5): 80-89) [DOI: 10.11929/j.swfu.202303048]
- Yang Z F, Tian J C, Feng K P, Gong X and Liu J B. 2021. Application of a hyperspectral imaging system to quantify leaf-scale chlorophyll, nitrogen and chlorophyll fluorescence parameters in grapevine. Plant Physiology and Biochemistry, 166: 723-737 [DOI: 10.1016/j.plaphy.2021.06.015]
- Yun Y H, Wang W T, Tan M L, Liang Y Z, Li H D, Cao D S, Lu H M and Xu Q S. 2014. A strategy that iteratively retains informative variables for selecting optimal variable subset in multivariate calibration. Analytica Chimica Acta, 807: 36-43 [DOI: 10.1016/j.aca.2013.11.032]
- Zhang X D, Chang J, Ren H D, Wu Y P, Huang M, Wu S, Yang S P, Yao X H and Wang K L. 2022. Mineral nutrient dynamics in pecans (*Carya illinoensis*) 'Mahan' grown in southern China. Frontiers in Plant Science, 13: 1003728 [DOI: 10.3389/fpls.2022.1003728]
- Zhao K, Li Y H, Yan H L, Hu Q and Han D X. 2022. Regulation of light spectra on cell division of the unicellular green alga *Haematococcus pluvialis*: insights from physiological and lipidomic analysis. Cells, 11(12): 1956 [DOI: 10.3390/cells11121956]
- Zheng K Y, Feng T, Zhang W, Huang X W, Li Z H, Zhang D, Yao Y and Zou X B. 2019. Variable selection by double competitive adaptive reweighted sampling for calibration transfer of near infrared spectra. Chemometrics and Intelligent Laboratory Systems, 191: 109-117 [DOI: 10.1016/j.chemolab.2019.07.001]

Estimation of leaf nitrogen content of *Carya illinoensis* using Fractional-Order Derivative of canopy-scale hyperspectral data

XU Jiajia, YU Lei, FU Genshen, YAN Lipeng, HUANG Qingfeng, TANG Xuehai

1. School of Forestry and Landscape Architecture, Anhui Agricultural University, Hefei 230036, China;
2. Anhui Provincial Key Laboratory of Forest Resources and Silviculture, Anhui Agricultural University, Hefei 230036, China

Abstract: Leaf nitrogen content can indicate the nutritional status and growth of plants. Hyperspectral technology can be used to estimate plant physiological and biochemical indicators in a non-destructive and efficient manner, providing reliable data for assessing the nutrient and health status of plants during growth and development. In order to explore the potential of hyperspectral technology in estimating the nutrient content of *Carya illinoensis*, this study took *Carya illinoensis* (Changlin and Jiande series) as the research object. Canopy hyperspectral data of 53 *Carya illinoensis* were randomly collected from 350—2500 nm in the field. Fractional Order Derivative (FOD) was used to preprocess the canopy spectrum, then two types of two-band spectral indices were combined to investigate the response relationship between Leaf Nitrogen Content (LNC) and canopy spectrum of *Carya illinoensis*. Finally, the modeling variables were screened by Variable Combination Population Analysis (VCPA). eXtreme Gradient Boosting (XGBoost) estimation models of canopy FOD single-band and FOD combined two-band spectral indices were constructed respectively, so as to obtain a suitable LNC estimation model under the test conditions of this experiment. Results showed that compared with the raw spectrum, the correlation between the canopy spectrum after FOD pretreatment and the LNC of *Carya illinoensis* was enhanced by 0.152. In terms of improving the correlation between spectral features and target components, FOD combined with Normalized Difference Spectral Index (NDSI) and Difference Spectral Index (DSI), had better performance than that of single band, with an improvement of 0.250 and 0.277, respectively. The VCPA variable selection method ultimately screened a subset of spectral variable combinations containing both strong and weak information variables, playing an important role in improving the estimation model accuracy. The optimal *Carya illinoensis* LNC model is the 1.5th-order derivative transformation combined with DSI, with $R^2=0.75$ and RMSEP=1.32 g/kg. In summary, canopy hyperspectral technology can rapidly and nondestructively estimate LNC of *Carya illinoensis*, and FOD combined with two-band spectral indices can significantly improve the response relationship between spectral features and target variables.

Key words: *Carya illinoensis*, canopy scale, hyperspectral remote sensing, nitrogen, fractional order derivative, spectral index, variable combination population analysis, machine learning

Supported by National Natural Science Foundation of China (No. 32171783)