

基于不同预处理的油茶叶片 SPAD 值 高光谱遥感估算研究

李 明,胡岑龙*,陶光林

(黔东南州林业科学研究所,贵州 凯里 556000)

摘要:叶绿素作为油茶生长过程中的重要生化参数,其含量直接影响油茶的光合能力,利用高光谱遥感技术估测油茶叶片叶绿素含量,对监测油茶树体的长势和健康状况具有重要意义。该文以油茶为研究对象,采集 450 组油茶叶片光谱数据和 SPAD 值。采用 5 种光谱分析方法对原始光谱进行预处理;通过连续投影算法(SPA)对原始光谱和预处理后的光谱进行敏感特征波段筛选,分别筛选出 4,10,4,12,11,9 个敏感波段;最后,应用偏最小二乘回归(PLSR)基于上述处理提取的敏感波段分别建立油茶叶片 SPAD 值估算模型。结果表明,5 种处理方法中,采用 MSC 预处理提取特征波段建立的估测模型优于其他处理方法,建模集 R^2 , RMSE 和 RE 分别为 0.855, 2.361 和 3.194%;预测集 R^2 , RMSE 和 RE 分别为 0.818, 1.980 和 2.533%,预测效果最好,建立的模型能更好实现对油茶叶片 SPAD 值的估算。

关键词:油茶;SPAD 值;光谱预处理;连续投影算法;估算

中图分类号:Q433.4;Q945.11;S794.4

文献标志码:A

doi:10.3969/j.issn.1001-7380.2022.02.001

叶绿素作为植物进行光合作用的重要生化参数,其含量变化与植物光合速率、氮素水平以及植株健康状况等密切相关,在植物生长、营养状况及光合作用等方面具有指示作用^[1-2]。传统的叶绿素测量方法主要是通过野外采样和室内分析获得,该方法不但费时、费力,且对采集对象具有一定的破坏性和不可恢复性,难以实现大面积叶绿素含量的实时获取^[3]。当前已广泛使用 SPAD 值反映叶绿素含量,SPAD 值表示叶片单位面积叶绿素的相对含量,是一个无量纲的比值。与传统方法相比,SPAD 测定具有快速、无损和简便的特点,但仅能进行某个点或小范围的测量,不利于大范围叶绿素含量的监测^[4]。而随着遥感技术的不断发展,高光谱凭借其高空间分辨率、多波段和光谱信息丰富的特点和优势,使实时、快速监测植被叶绿素含量成为可能,为采用遥感无损大尺度监测植被叶绿素含量提供了有效手段^[5]。20 世纪 90 年代,Blackburn^[6]通过

光谱指数估计叶片色素浓度得出叶绿素与高光谱波段之间存在一定的关系。随后,许多学者利用高光谱原始数据或由原始数据构建的植被指数、基于经验统计模型估算叶绿素含量开展了大量的研究,但大多集中在农作物小麦、玉米、棉花等方面,而在森林植被方面,杜华强等^[7]基于偏最小二乘回归模型建立了马尾松叶绿素含量与光谱反射率及 9 个特征参数之间的预测模型,其精度远大于传统线性回归模型。林辉等^[8]以杉木成熟林为研究对象,通过分析叶绿素总量与原始光谱、原始光谱一阶微分之间的相关性,构建了杉木叶绿素含量的高光谱模型。随着研究的不断深入,模型构建方面开始使用机器学习模型,冯海宽等^[9]基于特征光谱参数利用随机森林模型较好估算了苹果叶片叶绿素含量。尼格拉·吐尔逊等^[10]基于光谱一阶导数的地理加权最小二乘支持向量回归模型较好实现了对红枣树叶片 SPAD 值的预测。

收稿日期:2022-01-21;**修回日期:**2022-02-22

基金项目:贵州省林业局青年人才基金项目“基于植物高光谱特征的油茶主要病虫害监测及诊断模型研究”(黔林科合 J 字[2020]06 号);贵州省林业局青年人才基金项目“黔东南州乡土油茶优良无性系选育及繁殖技术研究”(黔林科合 J 字[2020]12 号);黔东南州科学技术局基础研究项目“黔东南油茶炭疽病调查及绿色防控技术初步研究”(黔东南科合 J 字[2021]72 号)

作者简介:李 明(1986—),男,贵州镇远人,助理工程师,硕士。从事林业 3S 技术研究。

* **通信作者:**胡岑龙(1988—),男,贵州岑巩人,工程师。从事林木良种选育工作。

油茶(*Camellia oleifera* Abel.)为我国4大木本油料树种之一。本文分别对原始光谱进行一阶导数、S-G平滑、多元散射校正(MSC)、标准正态变量(SNV)及向量归一化(VN)处理,结合偏最小二乘回归模型进行油茶叶片SPAD值估算,探索简便、快捷的油茶叶片光谱预处理方法和油茶叶片SPAD值反演模型,为提高油茶叶绿素含量遥感估算精度,以便精准监测其生长及健康状况提供一定的理论依据和技术支持。

1 材料与方 法

研究区位于贵州省黎平县东风林场(东经109°11'15",北纬26°20'23"),研究对象为2021年7月采集的油茶叶片,为保证试验结果的代表性,在油茶试验林内随机选取林龄为11 a生长健康的油茶树50株作为样本,每个样本按照上、中、下3层随机各采集健康、完整、无损的叶片3片,共采集9片作为1个样本的样叶,将采集的叶片装入密封袋并编号,带回实验室进行光谱数据的测定。

1.1 数据采集

使用FX2000光谱仪(光谱仪波长范围为350—1100 nm,狭缝宽度25 μm,光谱分辨率0.73 nm,以钨石英卤素灯为光源)在暗室内进行油茶叶片光谱数据采集,为降低环境对测量结果的干扰,每次光谱测量前用白板进行定标,并用干燥纸巾将叶片表面擦拭干净,然后铺平放置于操作平台,探头在距叶片2 cm处垂直向下测量,每个样本的样叶重复测量5次,取其平均值作为该样本的光谱反射率。根据Chan等^[11]人的研究,叶绿素敏感波段主要分布在可见光—近红外波段。因此,本研究选择400—1000 nm波段进行油茶叶片特征波段的选择和叶绿素含量的估算研究。

为保证数据的一致性和精度,每个样本光谱测量完后随即采用HM-YD叶绿素仪进行SPAD值的测量,测量时避开主脉,每个样本的样叶随机选择5个不同部位进行测量,最终取平均值作为该样本的SPAD值。

1.2 光谱数据预处理

本文使用Unscrambler 10.4软件对原始光谱进行一阶导数(主要是消除不同程度的背景干扰,实现光谱基线校正、提高不同吸收特征的对比度和光谱特征值)、S-G平滑(消除数据噪声在提取地物光谱信息时产生的干扰)、多元散射校正(MSC,主要

是消除颗粒大小对光谱产生的光谱差异,增加光谱与数据间的相关性)、标准正态变量(SNV,主要用于消除表面散射及光程度变化所带来的光谱误差)和向量归一化(VN,主要通过光谱数据减去光谱的吸光度平均值,来消除光程变化对光谱产生的影响)5种预处理。

1.3 模型精度评价

为了提高估算模型的普适性,需对构建的估测模型进行精度评价,将样本实测值带入估测模型进行实际样本值与估测模型值之间的精度比较。本研究采用以下3种评价指标对构建的模型效果进行评价:

(1)决定系数(R^2)。 R^2 用于表示实测值与预测值间的拟合程度,数值范围为0—1。 R^2 越大,模型的稳定性和准确性越高。

(2)均方根误差(RMSE)。用RMSE来确定模型的预测能力。RMSE越小,模型的精度越高,预测能力越好。

(3)相对误差(RE)。 RE 表示预测值与实测值的偏差。 RE 越小,模型的估算能力越强,则预测值与实测值拟合效果更好。

$$R^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

$$RE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100\%$$

式中, y_i 为实测值, \hat{y}_i 为预测值, \bar{y}_i 为实测值平均值, n 为样本数。

2 结果与分析

2.1 光谱预处理

在原始光谱数据的采集中,不仅包含有用信息,同时还受到杂散光、仪器噪声、样品背景、基线漂移等因素的干扰,这些因素都影响了特征波长的选取,从而影响光谱的定量和定性分析结果。因此,为了使提取的光谱信息更加准确反映样本曲线的变化,需要对原始光谱进行预处理,尽量消除或减少光照强度、环境因素和噪声干扰对光谱信息的影响。本试验采用一阶导数、S-G平滑、MSC、SNV和VN方法对原始数据进行预处理,原始光谱与处理后的光谱对比见图1。

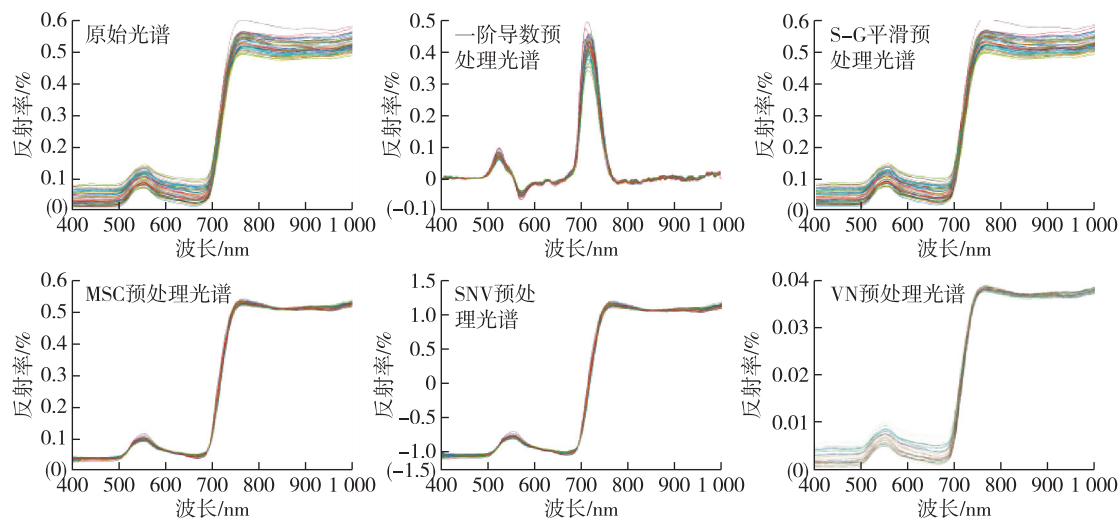


图1 原始光谱与预处理后的光谱

2.2 SPA 算法特征变量筛选

在对样本进行光谱信息采集时,采集的每条光谱曲线包含有数百甚至数千个数据点,同时,光谱信息中含有大量的冗余、共线和重叠信息,大量的数据明显影响建模的精度。本文利用连续投影算法^[12](SPA)筛选油茶光谱数据 SPAD 值的敏感特征波段。作为一种前向特征向量选择方法,它以较高的精度在光谱分析领域得到了广泛的应用。当用于光谱敏感特征波段的筛选时,可以通过设置特

征波段的数量来降低建模的复杂性,研究设置特征波段筛选的数目范围为 1—15,在均方根误差达最小时筛选出共线性最小、相关性最高的油茶叶片光谱对 SPAD 值敏感的特征波段。图 2 为通过 SPA 算法对原始光谱及预处理光谱筛选的特征波段,横轴表示波长,纵轴表示光谱反射率。由图 2 可以看出,在 500—1 000 nm 波段为主要的敏感特征波段分布范围,以 670—780 nm 波段范围的特征波段最多。

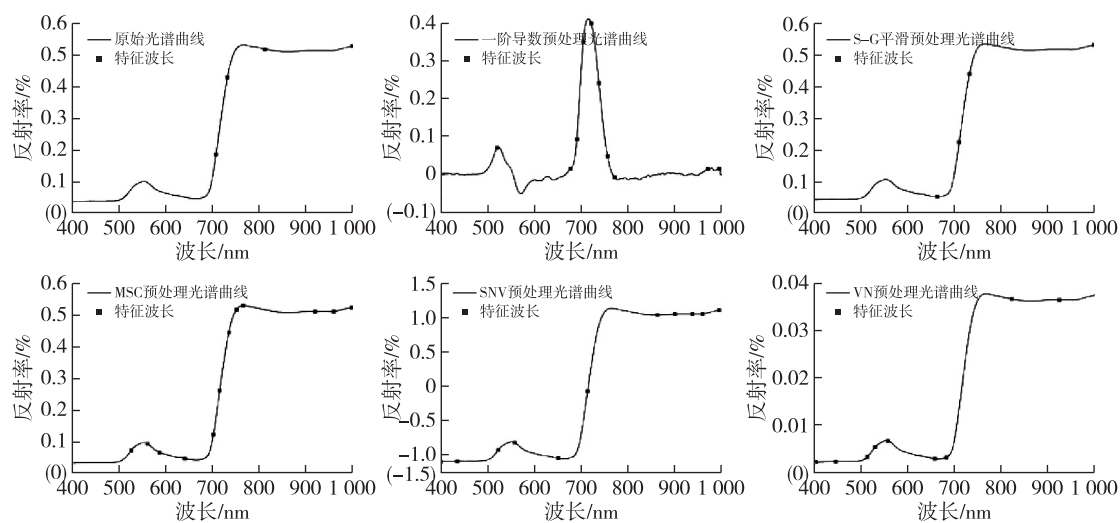


图2 基于 SPA 算法的特征波长选择结果

2.3 模型构建

在 R 软件下采用 pls 包进行偏最小二乘回归 (PLSR) 模型的构建,以连续投影算法从原始光谱和预处理光谱中提取的特征波长作为回归模型的自变量,以油茶叶片的 SPAD 值作为因变量,采用

SPXY 算法^[13]按 7:3 的比例进行样本划分,将油茶叶片 SPAD 样本划分为 35 个建模集和 15 个预测集,分别进行 PLSR 模型的构建和预测,结果如表 1 所示。

表 1 基于 PLSR 的油茶叶片 SPAD 值估算模型结果

处理 方式	波段数	建模集			预测集		
		R^2	RMSE	RE	R^2	RMSE	RE
原始光谱	4	0.597	3.980	5.861	0.794	2.142	2.802
一阶导数	10	0.830	2.558	4.625	0.780	2.299	2.807
S-G 平滑	4	0.623	3.853	5.924	0.835	1.921	2.440
MSC	12	0.855	2.361	3.194	0.818	1.980	2.533
SNV	11	0.854	2.397	4.559	0.769	2.173	2.755
VN	9	0.821	2.665	3.425	0.775	2.208	2.622

由表 1 建模集可知,光谱经过预处理后,建模精度较原始光谱数据均有不同程度的提高,对 5 种预处理方法(一阶导数、S-G 平滑、MSC、SNV 和 VN)建立的 PLSR 模型的精度对比发现,经过 MSC 预处理后的建模集决定系数 R^2 最高达 0.855。S-G 平滑建模集决定系数 R^2 最低为 0.623。SNV 预处理与 MSC

预处理的建模精度相差较小,从整个建模效果对比可知 $MSC > SNV > \text{一阶导数} > VN > S-G \text{ 平滑} > \text{原始光谱}$ 。

从表 1 的预测集可知,原始光谱预测精度 R^2 为 0.794,预测精度大于预处理的一阶导数、SNV 和 VN 模型的预测精度。5 种预处理方法中 S-G 平滑处理的预测精度最高,决定系数 R^2 达 0.835;其次是 MSC 预处理,决定系数 R^2 达 0.818。通过对预测集结果的残差分析可知,S-G 平滑处理预测结果残差的正态检验不通过,虽然决定系数 R^2 最高,但模型预测的结果不准确。通过模型综合分析可知,MSC 预处理的建模集和预测集的精度具有较好的一致性,建立的模型预测效果最好,能更好实现对油茶叶片 SPAD 值的估算。图 3 为不同预处理光谱 SPAD 值 PLSR 模型实测值和预测值散点图。

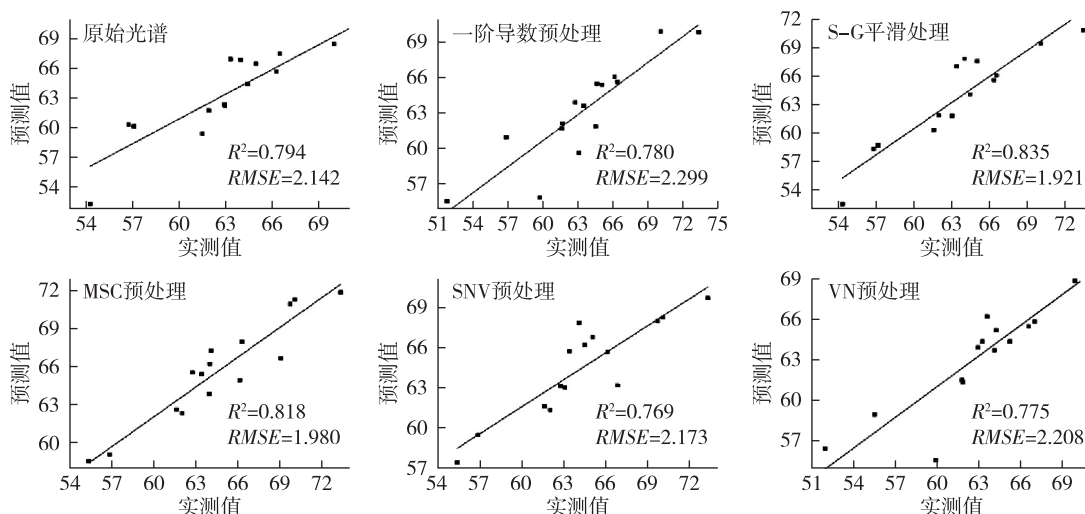


图 3 不同预处理光谱 SPAD 值 PLSR 模型实测值和预测值散点图

3 结论与讨论

本文以油茶叶片 SPAD 值为研究对象,以简化和提高油茶叶片 SPAD 值模型估测精度为研究目的,对原始光谱曲线分别进行了一阶导数、S-G 平滑、MSC、SNV 和 VN 预处理,然后通过 SPA 算法对油茶叶片光谱数据进行 SPAD 值敏感特征波段筛选,最后采用 SPXY 算法进行模型样本的划分,并建立了估算油茶叶片 SPAD 值的偏最小二乘回归模型,结果显示,光谱经不同变换和 SPA 算法处理后,有效降低了光谱信息的冗余,剔除了大量的重叠信息,还保证了筛选的特征波段之间共线性最小,减少了模型自变量的个数,降低了模型构建的复杂程

度,模型精度也得到了提高。

与直接采用原始光谱数据建模相比,光谱经不同变换后减少了各种干扰信息的影响,有价值波段信息得到增强,采用 MSC 预处理建立的估算模型精度最高,其建模集 R^2 , RMSE 和 RE 分别为 0.855, 2.361 和 3.194%;预测集 R^2 , RMSE 和 RE 分别为 0.818, 1.980 和 2.533%,预测效果最好,建立的模型可较好实现对油茶叶片 SPAD 值的估算。

叶绿素是植物叶片中吸收光能的主要物质,其浓度的多少对植株的生长情况产生重要影响^[14]。且植物叶片中其他营养元素的合成也与叶绿素内部结构有一定的联系,通过对叶绿素含量的监测,可有效估算植物营养状况及生理状态^[15]。而植物

叶片的营养状况与自身的光谱特征密切相关,植物叶片的光谱反射率随着叶绿素含量的不同而呈现出不同的光谱响应曲线^[16]。而高光谱遥感信息丰富,原始光谱易受环境、背景及噪声的影响,不可避免会携带一些与测试样本不相关的信息,如果直接将原始光谱用于建模,反演模型的精度将受到一定程度的制约。基于此,本文采用一阶导数、S-G 平滑、MSC、SNV 和 VN 共 5 种常用的光谱预处理方法,对预处理后的光谱信息进行特征波段选取,并尝试性地对油茶叶片 SPAD 值进行光谱建模。MSC 预处理建立的 SPAD 值模型预测能力最好, R^2 为 0.818,SNV 预处理预测效果最差, R^2 为 0.769。由于光谱曲线经 MSC 处理后部分有用信息得到增强,建立的模型质量较好。一阶导数、SNV 和 VN 预处理的模型预测能力,比直接采用原始光谱建模的预测能力较差,这说明不同的光谱预处理方法能直接影响模型预测效果^[17]。

由于高光谱数据包含的信息量大,对高光谱数据进行分析的方法也多种多样,如何从数百甚至数千个光谱信息中提取针对油茶叶绿素的敏感光谱参数,一直都是高光谱数据反演的重点和难点。本文只是进行了几种简单的光谱预处理,通过连续投影算法实现油茶叶片 SPAD 值预测模型构建,仍存在不足,未进行多种预处理方法的组合及多个模型对比,建立的模型以及筛选的最佳预处理方法还需进一步探索,模型的稳健性及普适性有待进一步完善。

参考文献:

- [1] CUI B, ZHAO Q J, HUANG W J, et al. Leaf chlorophyll content retrieval of wheat by simulated RapidEye, Sentinel-2 and EnMAP data[J]. Journal of Integrative Agriculture, 2019, 18(6): 1230-1245.
- [2] 李哲,张飞,陈丽华,等.光谱指数的植物叶片叶绿素含量估算模型[J].光谱学与光谱分析,2018,38(5):1533-1539.
- [3] 蔡庆空.基于小波变换的冬小麦叶面积指数与叶绿素含量遥感反演[D].北京:中国矿业大学(北京),2015.
- [4] LING Q, HUANG W, JARVIS P. Use of a SPAD-502 meter to measure leaf chlorophyll concentration in *Arabidopsis thaliana*[J]. Photosynthesis Research, 2011, 107(2): 209-214.
- [5] 曹英丽,邹焕成,郑伟,等.水稻叶片高光谱数据降维与叶绿素含量反演方法研究[J].沈阳农业大学学报,2019,50(1): 101-107.
- [6] BLACKBURN G A. Spectral indices for estimating photosynthetic pigment concentrations: A test using senescent tree leaves[J]. International Journal of Remote Sensing, 1998, 19(4): 657-675.
- [7] 杜华强,葛宏立,范文义,等.马尾松针叶光谱特征与其叶绿素含量间关系研究[J].光谱学与光谱分析,2009,29(11): 3033-3037.
- [8] 林辉,刘璇,臧卓,等.杉木叶绿素含量高光谱遥感模型研究[J].中南林业科技大学学报,2010,30(11):1-5.
- [9] 冯海宽,杨福芹,杨贵军,等.基于特征光谱参数的苹果叶片叶绿素含量估算[J].农业工程学报,2018,34(6):182-188.
- [10] 尼格拉·吐尔逊,苏磊·乃比,高健,等.GWLS-SVR 模型的红枣树叶片叶绿素含量估算[J].光谱学与光谱分析,2021,41(6):1730-1736.
- [11] CHAN C W, PAELINCKX D. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery[J]. Remote Sensing of Environment, 2008, 112(6):2999-3011.
- [12] GALVO R K H, ARAUJO MCU, FRAGOSO W D, et al. A variable elimination method to improve the parsimony of MLR models using the successive projections algorithm[J]. Chemometrics and Intelligent Laboratory Systems, 2008, 92(1):83-91.
- [13] 何勇.光谱及成像技术在农业中的应用[M].北京:科学出版社,2016.
- [14] YANG F, LI J L, GAN X Y, et al. Assessing nutritional status of *Festuca arundinacea* by monitoring photosynthetic pigments from hyperspectral data[J]. Computers & Electronics in Agriculture, 2010, 70(1): 52-59.
- [15] 周科琴,朱艳,杨杰,等.基于冠层高光谱参数的水稻叶片碳氮比监测[J].农业工程学报,2009,25(3):135-141.
- [16] 姚付启,张振华,杨润亚,等.基于红边参数的植被叶绿素含量高光谱估算模型[J].农业工程学报,2009,25(13):123-129.
- [17] 吴静珠,李慧,王克栋,等.光谱预处理在农产品近红外模型优化中的应用研究[J].农机化研究,2011,33(3):178-181.