

基于冠层高光谱数据的油茶炭疽病病情指数估测

李 明,陶光林*,廖华刚

(黔东南州林业科学研究所,贵州 凯里 556000)

摘要:利用一阶导数、S-G平滑及多元散射校正等光谱预处理方法对油茶炭疽病危害下原始光谱进行预处理后,采用SPXY样本划分法将65个样本按7:3将样本划分为45个校正集和20个预测集,再对不同预处理的光谱数据建立油茶炭疽病病情指数偏最小二乘回归模型。结果显示,在多种预处理方法中,S-G平滑预处理效果最好。通过连续投影算法(SPA)以及竞争性自适应重加权采样算法(CARS),从S-G平滑预处理的光谱中提取特征波长,进而构建基于偏最小二乘回归(PLSR)的油茶炭疽病病情指数的估测模型。试验结果发现,基于SPA所提特征波长建立的SPA-PLSR模型预测集 R_p^2 和预测均方根偏差(RMSEP)分别为0.700和0.072,基于CARS所提特征波长建立的CARS-PLSR模型预测集 R_p^2 和RMSEP分别为0.834和0.053;CARS-PLSR预测模型总体上要优于SPA-PLSR模型,可实现油茶炭疽病病情指数的估测。

关键词:油茶;炭疽病;连续投影算法;竞争性自适应重加权算法

中图分类号:Q433.4;S763.13;S794.4 **文献标志码:**A **doi:**10.3969/j.issn.1001-7380.2023.04.005

Estimation of anthracnose disease index in *Camellia oleifera* based on canopy hyperspectral data

Li Ming, Tao Guanglin*, Liao Huagang

(Qiandongnan Institute of Forestry, Kaili 556000, China)

Abstract: The original spectra of *camellia oleifera* affected by anthracnose disease were firstly preprocessed by first-order derivative, S-G smoothing, and multiplicative scattering correction. The SPXY (sample set partitioning based on joint x-y distance) method was then employed to divide 65 samples into 45 calibration sets and 20 prediction sets at a ratio of 7:3. Subsequently, partial least squares regression (PLSR) models for estimating the anthracnose disease index were established based on different preprocessed spectral data. The result showed that among various preprocessing methods, the S-G smoothing worked best. Feature wavelengths were extracted from the S-G smoothed spectra using successive projections algorithm (SPA) and competitive adaptive reweighting sampling algorithm (CARS). The PLSR models based on SPA-selected feature wavelengths (SPA-PLSR) and CARS-selected ones (CARS-PLSR) were constructed to predict the anthracnose disease index. Experimental results demonstrated that the SPA-PLSR model achieved a prediction set R_p^2 of 0.700 and RMSEP of 0.072, while the CARS-PLSR model got R_p^2 of 0.834 and RMSEP of 0.053. Overall, the CARS-PLSR model outperformed the SPA-PLSR model, and provided a feasible approach for estimating anthracnose disease index in *Camellia oleifera*.

Key words: *Camellia oleifera*; Anthracnose disease; Successive projections algorithm; Competency adaptive reweighting sampling

油茶炭疽病由油茶炭疽菌(*Colletotrichum gloeosporioides*)引起,是油茶(*Camellia oleifera*)生产

中的重要病害之一,尤其在长江以南的油茶种植区较为常见^[1]。它具有传播速度快、极易导致油

收稿日期:2023-05-10;修回日期:2023-06-14

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)

作者简介:李明(1986-),男,贵州镇远人,工程师,硕士。从事林业3S技术研究。

*通信作者:陶光林(1971-),男,高级工程师。从事野生植物引种栽培技术研究。

茶大规模感染的特点,感染后一般造成 10%—30% 的减产,严重感染情况下,减产超 50%,重度感染甚至造成绝收^[2]。当叶片受到病害危害时,叶片的外观会出现颜色和形状的改变,而叶片内部细胞结构、叶绿素、水分含量及其他生化成分也会跟着变化,叶片的反射光谱也随之变化,这就使得利用遥感技术估测病害的发生成为可能。当前病虫害的预测预报,主要还是依靠植保专家的田间调查、取样等方式进行,虽然调查结果具有较高的真实性和可靠性,但耗时、费力,且结果存在代表性差、滞后性及主观性强等缺点,难以实现大范围的病虫害监测和预报。因此,将遥感技术应用于油茶炭疽病的预测预报及防治具有重要意义。

近年来,随着高光谱遥感技术的成熟和普及,其在病虫害监测领域取得了长足的进步,并得到了广泛的应用。该技术宏观性强、监测精准、识别速度快、不具有破坏性^[3],已成为农业领域中一种理想的监测手段和重要发展方向。植株感病后,随着病害程度的加重,叶片的生理参数也将出现显著的变化,进而影响到反射光谱特征。可以通过分析植株受病害危害后的光谱信息提取敏感光谱,建立病情指数的估测模型,以实现对病害的准确测报。邓小玲等^[4]通过无人机高光谱遥感技术采用连续投影算法提取柑橘患病植株的特征波长与全波段波长建立模型,结果显示采用特征波长建模比使用全波段的精度更高;冯子恒等^[5]将 CARS-SPA(竞争性自适应重加权采样算法—连续投影算法)相结合,对通过一阶导数预处理的小麦白粉病高光谱数据进行特征波段进行提取,并建立了高斯过程回归模型。该模型具有较好的估算能力,能够有效提升小麦白粉病遥感监测精度;侯冰茹等^[3]通过应用 RF(随机蛙跳算法),SPA 和 CARS 算法从马铃薯晚疫病的光谱数据中提取特征波长,建立偏最小二乘回归预测模型对马铃薯晚疫病进行预测。结果表明,在采用的 3 种方法中,CARS 算法提取的特征波长建立的马铃薯晚疫病预测模型,效果最好,预测集 R_p^2 达 0.958 1。

本研究通过获取不同炭疽病感病程度的油茶冠层光谱数据,采用不同的光谱预处理方法对获取的光谱数据进行预处理,然后分别运用 SPA 和 CARS 进行特征波段提取,将提取的特征波段作为自变量、病情指数作为因变量建立 PLSR(偏最小二乘回归)估测模型,从而为实现高光谱遥感估测油

茶病害提供参考方法。

1 材料与方法

1.1 试验样地及病情指数获取

试验于 2022 年 5—8 月在黎平县东风林场的油茶基地进行。油茶树林龄为 11 a,株行距 2 m×3 m,平均树高 2.1 m,郁闭度 0.8。在研究区内选取 5 块 20 m×30 m 的样地,并以人工踏查的方法在样地内进行采样和病情指数调查,共采集样株 70 株,得到 70 组数据,将存在异常的 5 组数据进行剔除,最终获得了 65 组有效数据。调查结果显示,油茶炭疽病病情指数(DI)分布在 0—80.1% 之间,病害等级参照《油茶主要有害生物综合防治技术规程》进行划分^[6],具体病害程度分级及代表数值见表 1。

表 1 油茶炭疽病病害分级划分

级别	病情严重程度	代表数值
0	无明显病斑	0
I	病斑面积/叶片总面积<1/4	1
II	1/4≤病斑面积/叶片总面积<1/2	2
III	1/2≤病斑面积/叶片总面积<3/4	3
IV	病斑面积/叶片总面积≥3/4	4

通过对样地内受炭疽病感染的油茶冠层叶片进行统计,按下式计算出单株油茶的 DI^[7]

$$DI = \frac{\sum (x \times f)}{\sum f \times n} \times 100\%$$

式中, x 为各梯度的级值, f 为各梯度的叶片数, n 为最高梯度值 4。

1.2 光谱采集

采用 FX2000 非成像光谱仪(波长范围为 350—1 100 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)对油茶冠层进行高光谱数据采集。为减少外界因素的影响,确保获取数据的准确性,光谱采集时将光谱仪的探头与油茶冠层的垂直距离限制在 0.5 m 以内。每次光谱测量前用白板进行定标,并选取 DI 相同或相近的油茶冠层进行光谱测定,以降低环境对测量结果的干扰,保证数据的一致性和精度。每个样本连续采集 10 条光谱曲线,以其平均值作为样本的冠层反射率。

1.3 数据处理

由于油茶冠层结构的复杂性、仪器性能及其他外部环境的影响,原始光谱往往带有一定的噪声,对模型的稳定性具有不同程度的影响,需要对采集

的光谱数据进行预处理,将光谱噪声尽可能地从采集的光谱曲线中剔除,最大限度突出光谱的有用信息。因此研究采用 The Unscrambler © X 10.4 光谱处理软件进行光谱预处理。高光谱数据预处理的方法较多,本研究采用常见的光谱预处理方法主要包括一阶导数变换(FD)、多元散射校正(MSC)、标准正态变换(SNV)、S-G平滑和去趋势处理(DT)。在 MATLAB 2020a 下编写程序对光谱信息进行特征波长提取、样本划分和 PLSR 模型建立及评价。

1.4 特征波长选择算法

1.4.1 连续投影算法 研究通过使用连续投影算法^[8](SPA)进行炭疽病危害下油茶光谱特征波长提取。主要分为以下几个步骤:第1步,从光谱数据中筛选出共线性最低的变量子集。第2步,利用这些筛选出的变量构建回归模型,并从模型结果中挑选出均方根误差最小的变量子集,使用筛选出的各子集变量建立回归模型,根据模型结果,选出均方根误差最小的子集。第3步,对第2步筛选出的变量建立一个逐步回归模型,以确保模型的预测准确性,从而获取一个变量数量较少的集合,其中包含的波长变量就是最终确定的有效波长。

1.4.2 竞争性自适应重加权算法 CARS^[9]是一种基于蒙特卡罗采样和 PLSR 系数的特征波长选择方法,算法根据达尔文“适者生存”的原则,利用自适应重加权采样技术结合指数衰减函数,从而有效地筛选出具有较高可靠性的波长变量,从而达到更好的预测效果。通过十折交叉验证法,去除权重值较小的变量,从而确定 PLSR 模型中 RMSECV(交互验证均方根误差)值最低的子集,这些子集将构成 CARS 算法提取的最佳特征波长组合。

1.5 模型的建立与评价

在 MATLAB 2020a 软件下编程实现 PLSR 模型,以 SPA 和 CARS 提取的特征波长为自变量,以油茶炭疽病病情指数为因变量,采用 SPXY 样本划分法将 65 个样本按 7:3 的比例进行样本划分,划分为 45 个校正集和 20 个预测集,分别进行 PLSR 模型的构建和预测。

本研究使用以下指标对模型进行精度评价,建模集决定系数(Determination coefficients of cablibration, R_c^2)、验证集决定系数(Predicting determination of cablibration, R_p^2)、建模均方根误差(root mean square error of calibration, RMSEC)和预测均方根误差(root mean square error of prediction, RM-

SEP)。 R^2 越大,模型精度越高。RMSEC 和 RMSEP 与模型精度成反比,即其值越小,模型精度越高。

2 结果与分析

2.1 炭疽病危害下油茶冠层光谱曲线

由于受环境和仪器噪声的影响,350—400 nm 和 1050—1 100 nm 波段范围内光谱反射率波动变化较大,因此选择 400—1 050 nm 的光谱数据进行分析研究。图 1 为油茶炭疽病危害下不同病情指数的油茶冠层光谱曲线。由图 1 可以看出,在 550 nm 附近,由于叶片叶绿素对该范围的光吸收较少,形成了 1 个与叶绿素有关的小反射峰,称为“绿峰”。在红色波段 670 nm 附近,由于叶绿素的强吸收,形成吸收谷,称为“红谷”。在近红外波段,由于叶片多孔薄壁细胞结构对近红外光的强烈反射,在光谱曲线形成了最高峰。当油茶叶片受炭疽病危害后,叶片细胞遭到破坏,叶片结构特征和内部的生理生化参数发生变化,使得叶片内叶绿素水平下降。此时,光谱曲线出现明显的改变,特别是 550 nm 处的绿峰和近红外波段,光谱反射率随着病害程度的增加而减小,在病害感染后期,在光谱曲线上这一特征就更加明显。因此,根据光谱间的差异,进行高光谱油茶炭疽病病情指数的估测是可行的。

2.2 不同预处理的 PLSR 建模分析

为确定最适的预处理方法,研究通过对原始和不同预处理光谱数据建立的 PLSR 模型进行比较分析,确定最佳预处理方法。试验中得到不同预处理方法下的油茶炭疽病病情指数 PLSR 模型结果,如表 2 所示。由表 2 可以看出,相对于原始光谱数据,通过标准正态变换和多元散射校正处理后,模型预测集 R_p^2 比校正集 R_c^2 精度更低了,这是因为标准正态变换是对 1 条光谱进行处理,而多元散射校正则是对一组样本光谱曲线进行处理,处理后的光谱数据对建模的结果具有一定的制约^[10]。通过 FD、S-G 平滑及 DT 处理后,模型的精度都有不同程度的提高。几种预处理方法中,S-G 平滑预处理的数据建模对油茶炭疽病病情指数的预测效果最好。因此,将 S-G 平滑预处理后的数据作为下一步特征波段选择的光谱数据集。

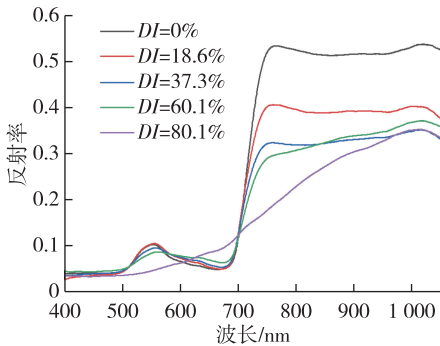


图 1 不同病情指数的油茶冠层光谱曲线

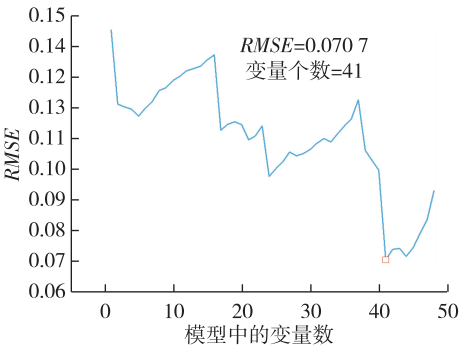


图 2 模型均方根误差随变量变化趋势

表 2 不同预处理方法的油茶炭疽病病情指数

PLSR 模型预测结果					
预处理方法	主成分数	校正集		预测集	
		R^2_c	RMSEC	R^2_p	RMSEP
FD	7	0.700	0.115	0.706	0.083
YS	4	0.677	0.116	0.703	0.076
SNV	7	0.616	0.126	0.601	0.076
S-G 平滑	4	0.727	0.103	0.812	0.056
MSC	5	0.740	0.103	0.524	0.082
DT	4	0.707	0.112	0.763	0.073

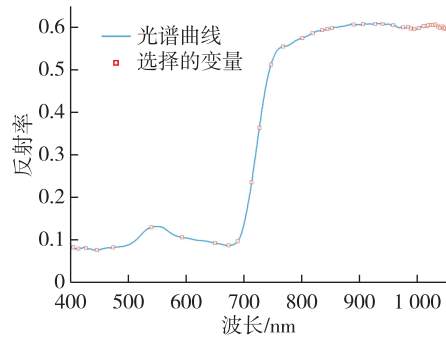


图 3 SPA 算法特征波长分布情况

2.3 特征波段筛选

2.3.1 SPA 特征波段筛选 采用连续投影算法从 S-G 平滑处理的光谱数据中提取特征波长,在参考前人区间设置的基础上,研究将选择变量数的最优区间设置为 $[2, 50]$ ^[11],在均方根误差达最小时筛选出共线性最小、与油茶炭疽病相关性最高的敏感特征波段,基于 SPA 算法对油茶炭疽病危害下病情指数敏感波段筛选结果如图 2 和图 3 所示。由图 2 可知,均方根误差 RMSE 随变量数的增加而发生显著变化,当变量数达 41 时(图 2 中空心方框的位置),RMSE 值最低,仅有 0.070 7。图 3 为 SPA 筛选特征波长结果,41 个特征波段(空心方框所示)在光谱曲线上的分布状态,其中 800—1 050 nm 波段范围的特征波段最多,占据了筛选结果的绝大部分。

2.3.2 CARS 特征波段筛选 在使用 CARS 算法进行油茶炭疽病病情指数特征波段筛选时,为了获得更准确的特征波段,将蒙特卡罗采样次数调整为 100 次。CARS 算法运行步骤及特征波段筛选结果如图 4 和图 5 所示。图 4 分别显示了 CARS 算法在 1 次运行中,随着蒙特卡罗采样次数的增加,变量数量、交叉验证 RMSECV 值以及各变量的回归系数的变化情况。从图 4(a)可以看到,由于指数衰减函数的影响,在前 10 次蒙特卡罗采样中,变量数急剧下

降,随后变量数减少的速度逐渐减慢。结果表明,该算法在关键变量的选择上具有“粗选”和“精选”2 个阶段。由图 4(b)可以看出, RMSECV 值表示算法提取的特征波长构建的 PLSR 模型的预测效果,当蒙特卡罗采样次不断增加时, RMSECV 值随着采样次数的增加不断变化。当采样次数达到 31 次时, PLSR 模型的 RMSECV 值最小,随着采样次数的继续增加, RMSECV 值又不断上升,当采样次达 65 次时, RMSECV 值变化趋于稳定,表明与 DI 预测相关的一些重要光谱信息被排除,从而降低了 PLSR 模型的准确度。因此,将第 31 次蒙特卡罗采样后获得的变量确定为预测油茶炭疽病病情指数(DI)的关键变量,共计 30 个特征波段。从图 5 可以看出,运用 CARS 算法筛选出的 30 个特征波段(空心方框)在光谱曲线上呈现出明显的分布趋势,其中,在 480—500, 670—690, 760—770 和 1 038—1 040 nm 的波段最常见和集中。

2.4 基于不同特征波段筛选方法的 PLSR 模型

基于 SPA 筛选的 41 个特征波段和 CARS 算法筛选出的 30 特征波段分别建立 SPA-PLSR 和 CARS-PLSR 模型对油茶炭疽病病情指数进行预测,模型结果见表 3,由表 3 可知, SPA-PLSR 模型的校

表 3 基于 SPA 和 CARS 算法的 PLSR 模型结果						
建模方法	波段数	主成分数	校正集		验证集	
			R_c^2	RMSEC	R_p^2	RMSEP
SPA-PLSR	41	5	0.758	0.097	0.700	0.072
CARS-PLSR	30	7	0.753	0.098	0.834	0.053

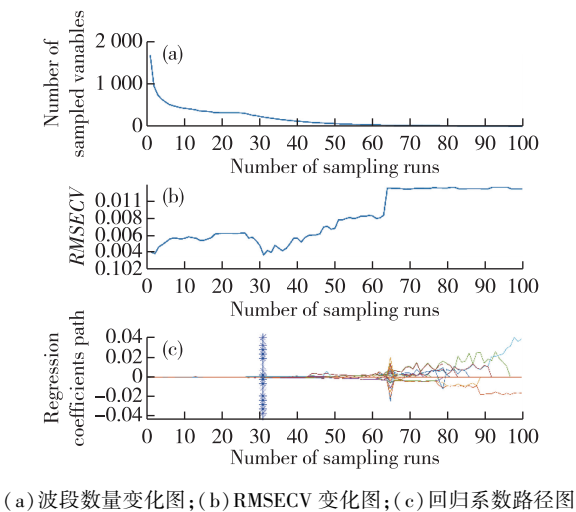


图 4 CARS 特征波段选择

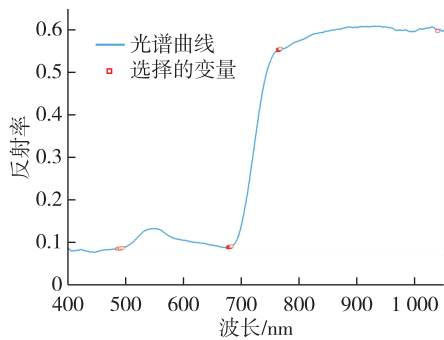


图 5 CARS 算法 SPA 算法特征波长分布情况

正集 R_c^2 与预测集 R_p^2 分别为 0.758 和 0.700, 相应的 $RMSEC$ 和 $RMSEP$ 为 0.097 和 0.072, 而 CARS-PLSR 模型的校正集 R_c^2 与预测集 R_p^2 为 0.753 和 0.834, 相应的 $RMSEC$ 和 $RMSEP$ 为 0.098 和 0.053。通过对 2 个模型的 R_p^2 和 $RMSEP$ 比较, 2 个模型都能较好的对油茶炭疽病病情指数进行预测, 但 SPA-PLSR 模型的预测精度低于 CARS-PLSR 模型, 这主要是由于它采用的特征选择过程没有受到任何监督, 只是尽可能地扩大了自变量的范围, 而没有建立一个完整的预测模型, 从而导致其变量解释能力受到了限制。通过采用 CARS 算法, 只使用了 30 个特征波段(占总波段的 1.78%), 这样就可

以有效地减少建模的变量数和维度, 从而降低了模型的复杂性, 大大缩短了建模时间。由此可知, 研究中 CARS-PLSR 模型在油茶炭疽病病情指数的预测上表现出了显著的优势, 它能够更准确地筛选出与油茶炭疽病病情指数有关的特征波段。

图 6 和图 7 为 SPA-PLSR 模型和 CARS-PLSR 模型油茶炭疽病病情指数实测值和预测值散点图。可以看出, CARS-PLSR 模型拟合的效果较好, 能较好地实现对油茶炭疽病病情指数的估测。

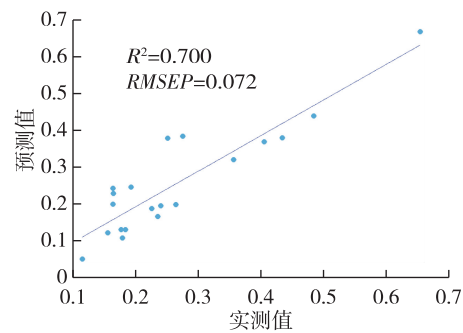


图 6 SPA-PLS 模型预测结果

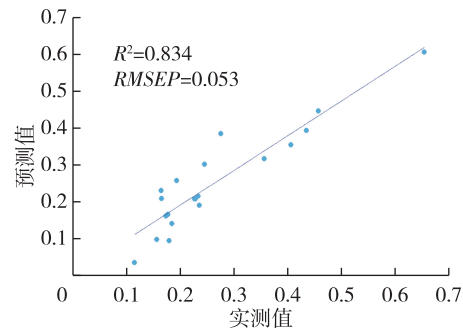


图 7 CARS—PLS 模型变量预测结果

3 结论

本文以油茶冠层炭疽病病情指数为研究对象, 利用高光谱遥感技术, 初步建立了冠层光谱特征波段与病情指数之间的 PLSR 估测模型, 简化和提高了油茶炭疽病病情指数估测精度。试验结果显示, 几种对原始光谱数据预处理方法中, S-G 平滑预处理的数据建模对油茶炭疽病病情指数的预测效果最好, 而基于 CARS 算法的 CARS-PLSR 预测模型则总体上要优于 SPA-PLSR 模型。分别采用 SPA 算法和 CARS 算法对 S-G 平滑预处理的光谱数据进行特征波段提取, SPA 算法提取了 41 个特征波段,

(下转第 52 页)