

# 基于荧光光谱的鲜茶叶片叶绿素含量定量分析

刘燕德\*, 林晓东, 高海根, 高雪, 王舜

华东交通大学光机电技术及应用研究所, 江西 南昌 330013

**摘要** 茶叶叶片叶绿素含量的准确监测对茶树的营养状况和生长态势具有重要意义, 为此基于叶绿素荧光光谱技术提出一种快速无损检测叶片叶绿素含量的方法。利用叶绿素荧光采集装置对茶叶叶片进行光谱采集, 并测量叶绿素相对含量。采用 S-G(Savitzky-Golay)平滑法对光谱进行预处理, 可以消除大量的噪声信号; 对所提方法与传统方法进行比较。实验结果表明, 采用所提方法能够有效消除无关变量, 对模型的优化可以得到较好的效果; 简化变量后所建立的偏最小二乘模型在预测集上的相关系数为 0.96, 方均根误差为 0.87, 在建模集上的相关系数为 0.96, 方均根误差为 0.95; 荧光光谱结合化学计量学方法可以为茶叶叶片叶绿素含量的定量分析提供一种快速简便的分析方法。

**关键词** 光谱学; 荧光光谱; 连续投影算法; 后向区间偏最小二乘法; 叶绿素

中图分类号 O657.3

文献标志码 A

doi: 10.3788/LOP202158.0830001

## Quantitative Analysis of Chlorophyll Content in Tea Leaves by Fluorescence Spectroscopy

Liu Yande\*, Lin Xiaodong, Gao Haigen, Gao Xue, Wang Sun

*Institute of Optics Mechanics Electronics Technology and Application, East China Jiaotong University, Nanchang, Jiangxi 330013, China*

**Abstract** Accurate monitoring of the chlorophyll content of tea leaves is of great significance to the nutritional status and growth of tea trees. Thus, a method for rapid and nondestructive detection of chlorophyll content of leaves is proposed on the basis chlorophyll fluorescence spectroscopy technology. The chlorophyll fluorescence collection device is used to collect the spectrum of tea leaves and measure the relative chlorophyll content. The Savitzky-Golay (S-G) smoothing method is used to preprocess the spectrum, which can eliminate a large number of noise signals. The proposed method is compared with the traditional method. Experimental results show that the proposed method can effectively eliminate irrelevant variables, and the optimization of the model can achieve better results. The partial least square model established after simplifying the variables has a correlation coefficient of 0.96 on the prediction set, and a root mean square error of 0.87. The correlation coefficient on the model set is 0.96, and the root mean square error is 0.95. The fluorescence spectroscopy and chemometric methods can provide a quick and easy analysis method for the quantitative analysis of tea leaf chlorophyll content.

**Key words** spectroscopy; fluorescence spectroscopy; successive projections algorithm; backward interval partial least squares; chlorophyll

**OCIS codes** 300.2530; 300.6170; 200.3050

收稿日期: 2020-07-30; 修回日期: 2020-08-31; 录用日期: 2020-09-10

基金项目: 国家自然科学基金(31760344)、江西省教育厅科学技术研究项目(GJJ190306)

\* E-mail: jxliuyd@163.com

## 1 引言

叶绿素含量是直接评价作物生理状态和长势的关键参数之一<sup>[1]</sup>,而且对农业的水肥管理和利用也具有重要意义,因此对其进行监测是判断植株健康程度和作物产量的有效方法之一。目前,测定叶绿素含量的方法主要有活体叶绿素仪法<sup>[2]</sup>、高效液相色谱法<sup>[3]</sup>、光声光谱法<sup>[4]</sup>和紫外分光光度法<sup>[5]</sup>,上述方法虽能准确测定叶绿素含量,但费时费力,而且需要破坏采样,难以满足实时、快速、无损和大面积监测的需求。光谱技术具有省时、便宜、样本量少和不破样品结构等特点,可以在不破坏植物结构的情况下实现对叶绿素含量的监测<sup>[6-7]</sup>。

目前,应用光谱技术测量植物叶绿素含量的报道很多。杨婧等<sup>[8]</sup>采用了一元线性回归和神经网络建立叶绿素含量估算模型,使用该模型对叶绿素含量进行测定。刘宁等<sup>[9]</sup>使用了便携式地物光谱仪(ASD)采集马铃薯在不同生长时期的光谱反射率变化特征,并同步测定叶绿素含量,通过筛选得到的光谱特征来建立偏最小二乘(PLS)模型,得到在筛选马铃薯叶绿素特征波长方面,随机蛙跳(RF)算法优于蒙特卡洛无信息变量消除(MC-UVE)法和竞争自适应重加权采样(CARS)法,得到的特征波长能够较全面地反映与马铃薯叶绿素相关的物质信息。石吉勇等<sup>[10]</sup>采用了联合区间偏最小二乘(SiPLS)法和净分析物(NAS)法提取近红外光谱的特征信息,建立了黄瓜叶片叶绿素光谱模型。杨艳阳等<sup>[11]</sup>使用了发光二极管(LED)的荧光光谱信息建立鲜茶叶的相对叶绿素含量模型,结果表明叶绿素含量与 685 nm 波长处的荧光强度之间有显著的相关性,但是 LED 的光源不稳定,而且在室外使用过程中太阳光对其造成的影响无法抵消。以上报道所使用的光谱设备均存在着价格昂贵、体积较大和不宜现场采集等缺点,所以迫切需要一种便捷且经济的光谱设备来实现田间作业。荧光光谱技术具有灵敏度高、选择性强和方法简单的优点,而且应用范围广泛,特别是在物质的种类检测中有着重要的应用价值。

鉴于此,本文提出基于荧光光谱的鲜茶叶片叶绿素含量定量分析模型,采用连续投影算法以及后向区间偏最小二乘法筛选特征波长以减少建模变量,以期提高建模的速度和精度。研究结果将对茶叶叶片的叶绿素含量的测量提供一定的参考价值。

## 2 材料与方法

### 2.1 实验材料

实验所使用的样品选自福建省无公害零农药有机茶基地,在每棵茶树的 4 个方向上随机选择 10 个不同高度和水平的冠层叶片,无损且规则形状的叶片如图 1 所示。将收集好的叶片送回实验室,共 120 片。采摘后拭去叶片表面的灰尘,接着使用去离子水清洗并装入密封袋中立即放置在 2 °C 的温度下全程保存。采用光谱-理化值共生距离(SPHY)法将样本集按照 2 : 1 的比例划分为校正集和预测集,即校正集有 80 片,预测集有 40 片。



图 1 无损且规则形状的叶片

Fig. 1 Lossless, regular-shaped blade

### 2.2 茶叶叶片荧光光谱的采集

图 2 为荧光光谱采集装置,该系统由激光光源、光纤、样品池、光谱仪和计算机组成。实验中使用的光谱仪是科研级别的光谱仪 QE65000,分辨率为 1.5 nm,积分时间为 7 ms ~ 15 min,工作范围为 200 ~ 1100 nm。光谱仪的侧面配有 USB 接口,与计算机相连,并由计算机直接供电。光谱仪中,线阵 CCD (Charge Coupled Device) 探测器的分辨率为 1044 pixel × 64 pixel。光谱仪配有 SMA905 接口,与同样接口的可见-近红外(VIS-NIR)光纤相连,光纤的纤芯直径为 1000 μm,数值孔径为 0.22,发散全角为 25.4°,激光光源的波长为 405 nm。采集过程中,积分时间设为 20 ms,激光功率设为 40 mW。光源与植物叶片间的夹角为 45°,光纤垂直对准植物叶片,在叶片表面的垂直高度为 3.0 cm。数据由海洋光学配套的软件 SpectraSuite 来采集。

### 2.3 茶叶叶片叶绿素含量的测定

手持式叶绿素无损测量仪(SPAD 502 Plus)可以在人眼不可见之前就对植物健康的细微变化或趋势进行定量分析。通过测量 650 nm 和 940 nm 波

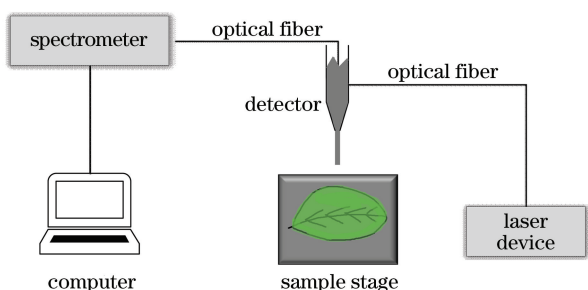


图 2 叶绿素荧光光谱的采集原理图  
Fig. 2 Capture schematic of chlorophyll fluorescence spectrum

长处的光密度差来确定叶片当前叶绿素的相对含量,叶绿素含量的测量范围为 $-9.9\sim 199.9$ ,测量尺寸为 $2\text{ mm}\times 3\text{ mm}$ ,光源为两个 LED,仪器的重复性在 $\pm 0.3$  SPAD 单位以内(SPAD 值介于 $-9.9\sim 199.9$ )。当测量叶绿素含量时,直接将叶片放置在 SPAD-520 PLUS 的感应部位上轻轻用力夹紧,屏幕上显示的数字即为该点的 SPAD 值。对相对叶绿素含量的平均值、最大值、最小值和标准差等数据进行统计,结果表 1 所示,其中 SD 为标准差,CV 为变异系数。

表 1 叶绿素含量的统计数据  
Table 1 Statistics on chlorophyll content

Sample	Maximum	Minimum	Mean	SD	CV
Tea leaf	$87.0\pm 0.3$	$61.6\pm 0.3$	78.93	5.09	0.06

## 2.4 光谱数据的预处理

数据的预处理是指尽量减少实验室仪器或环境变化对荧光信号的影响,在不丢失相关信息的情况下简化后续操作。这种预处理操作可确保不同样品光谱之间的可比性,并消除了杂散光和光谱中无关化学信息的影响。如果未对数据进行预处理操作,某些无关变量可能会在建模中占据主导,进而对模型产生偏差。S-G (Savitzky-Golay) 平滑去噪法是目前最常用的平滑法之一,可用于消除光谱采集过程中的噪声信号,为此提取光谱数据中的有效信息。S-G 平滑法主要是通过多项式对光谱数据进行平滑处理<sup>[12-13]</sup>。平滑点数在平滑处理过程中起着至关重要的作用,点数过多易丢失有效的光谱信号,点数过少噪声消除不干净。叶绿素在可见光波段有两个较强烈的吸收峰,分别位于 $430\sim 450\text{ nm}$ 的蓝紫光区域和 $640\sim 660\text{ nm}$ 的红光区域<sup>[14]</sup>,如图 3 所示,其中 $R_C$ 为标定集相关系数, $R_P$ 为预测集相关系数, RMSEC 为标定集的方均根误差, RMSEP 为预测集的方均根误差。从图 3 可以看到, $300\sim 500\text{ nm}$

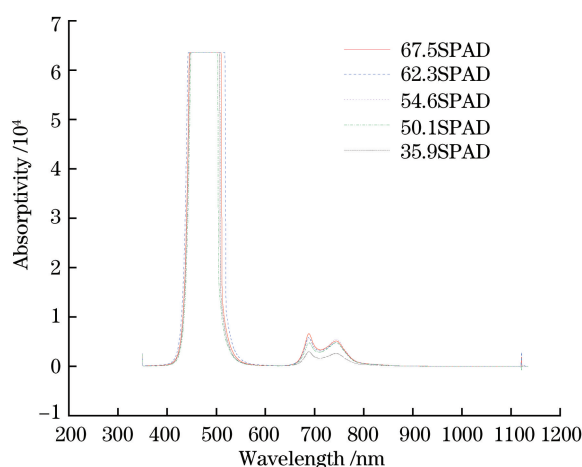


图 3 原始的全光谱图像

Fig. 3 Original full spectrum image

波段的噪声干扰大,且 $900\text{ nm}$ 之后的光谱有用信息较少,故取原始光谱的 $600\sim 900\text{ nm}$ 波段用于后续的建模分析,并采用 S-G 方法对所选区域进行平滑校正并建立 PLS 模型,处理前后的 PLS 值如表 2 所示。从表 2 可以看到,经过 5 点平滑后的模型效果最佳。

表 2 叶绿素相对含量处理前后的 PLS 值  
Table 2 PLS values before and after chlorophyll relative content treatment

Evaluation index	Before	After	
		3 points	5 points
$R_C$	0.8339	0.9129	0.9162
RMSEC	2.2360	2.2776	0.2747
$R_P$	0.8820	0.8675	0.8989
RMSEP	1.8811	1.9199	1.4831

## 2.5 偏最小二乘法

偏最小二乘回归(PLSR)<sup>[15]</sup>是一种有效的多元变量统计回归技术,其对光谱的分析非常有用,广泛用于建立具有大量且高度共线性的变量线性模型,可以使变量与少量样本相关。PLSR 不仅对共线变量不敏感,而且对大量的变量具有容忍度。校正模型一般采用全荧光光谱。实验将变量的选择作为减少 PLS 模型计算复杂度的方法,从而提高模型的可靠性和稳定性。

PLSR 广泛应用于光谱分析,可以将光谱矩阵  $\mathbf{X}$  与浓度矩阵  $\mathbf{Y}$  相关联。PLSR 模型可表示为

$$\mathbf{Y} = \mathbf{bX} + \mathbf{e}, \quad (1)$$

式中: $\mathbf{b}$  表示回归系数; $\mathbf{e}$  表示模型的残差矩阵。

## 2.6 连续投影算法

连续投影算法(SPA)<sup>[16]</sup>是一种前向循环的

变量选择算法。选定一个初始频点,接着在每一次迭代过程中加入新的频点,直到达到指定的频点数,该方法的本质是解决信息重叠,选择

最小冗余信息量的频点组合,从而解决共线性问题。荧光光谱的预处理及模型对比流程如图 4 所示。

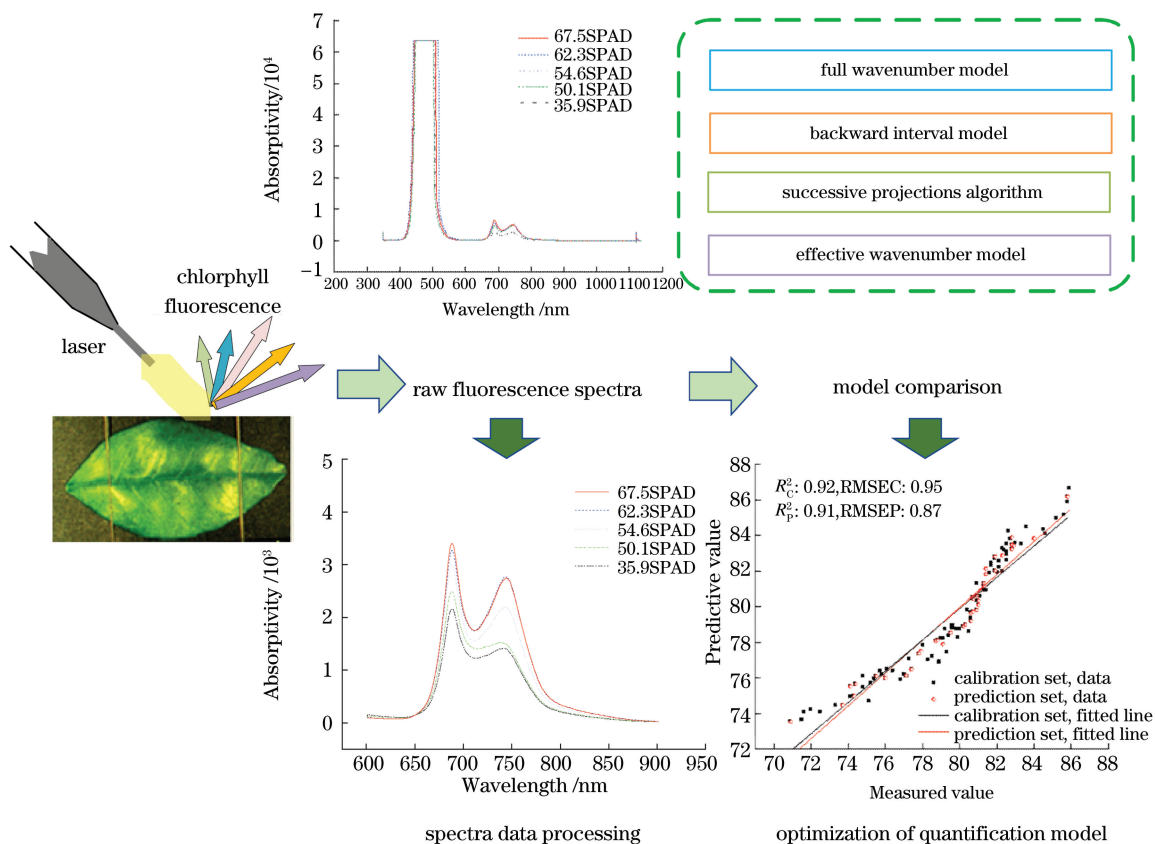


图 4 荧光光谱的预处理及模型对比流程

Fig. 4 Fluorescence spectrum pretreatment and model comparison process

由样本数  $q$  与波长点  $l$  组成一个光谱矩阵  $\mathbf{X}_{q \times l}$ , 将第  $l_0$  列作为基准向量, 并记波长数  $n$  ( $n < l-1$ ) 为需要提取的数目。从中选择一个波长, 计算出每个样本在该波长点处的响应, 构成的向量在未选中的波长上进行投影, 如此循环  $m$  次, 直至选中投影最大的向量为止, 将该向量对应的波长与其余波长进行线性组合, 使之与上一次选中的向量线性关系最小, 这就是连续投影算法, 具体步骤如下。

1) 令  $n=1$ , 则在矩阵中随意选一列, 记为  $\mathbf{X}_v$ , 其中  $v \in \{1, \dots, l\}$ 。

2) 将未选中的波长组成一个集合, 记为  $S$ , 计算出  $\mathbf{X}_v$  在  $S$  中的投影向量  $\mathbf{R}_{\mathbf{X}_v}$ 。

3) 给出最大投影向量的标号, 即  $Z(n) = \operatorname{argmax} \|\mathbf{R}_{\mathbf{X}_v}\|$ , 将其作为下一个循环的投影向量, 即  $\mathbf{X}_v = \mathbf{R}_{\mathbf{X}_v}$ 。

4)  $n=n+1$ , 如果  $n < l$ , 返回到步骤 2) 继续计

算, 最终提取向量  $\mathbf{U}_l = [0 \ 1 \ 2 \ \dots \ l_{n-1}]$ 。

完成上述步骤后, 可以获得  $n \times t$  对波长组合, 由  $\mathbf{X}_v$  和  $n$  共同决定的波长组合来建立模型<sup>[17-19]</sup>, 得到 RMSEP 值最小的模型波长组合为最佳组合。

## 2.7 后向区间偏最小二乘法

后向区间偏最小二乘 (BiPLS) 算法是由 Norgaard 等<sup>[20]</sup>开发的, BiPLS 模型是将数据集分成给定数量的等长区间间隔, 新的 PLS 模型是在每个区间间隔被忽略的情况下计算的。选择 20 个间隔, 每次去除一个区间间隔使其被排除在外所剩下区间的模型效果最优。循环此过程, 直到只存在一个区间间隔。

## 2.8 数据统计分析

通过校正集的相关系数 ( $R_c$ )、决定系数 ( $R_c^2$ ) 和方均根误差 (RMSEC) 来评价叶绿素含量与光谱数据之间的校正方程。通过验证集的相关系数

( $R_p$ )、决定系数( $R_p^2$ )和方均根误差(RMSEP)来评价校正模型的预测精度<sup>[21]</sup>。这些统计指标的表达式分别为

$$R_C = \sqrt{1 - \frac{\sum_{i=1}^{k_C} (y_{i,C} - \hat{y}_{i,C})^2}{\sum_{i=1}^{k_C} (y_{i,C} - y_{\text{mean},C})^2}}, \quad (2)$$

$$R_C^2 = 1 - \frac{\sum_{i=1}^{k_C} (y_{i,C} - \hat{y}_{i,C})^2}{\sum_{i=1}^{k_C} (y_{i,C} - y_{\text{mean},C})^2}, \quad (3)$$

$$M_{\text{RMSE},C} = \sqrt{\frac{1}{k_C} \sum_{i=1}^{k_C} (y_{i,C} - \hat{y}_{i,C})^2}, \quad (4)$$

$$R_P = \sqrt{1 - \frac{\sum_{i=1}^{k_P} (y_{i,P} - \hat{y}_{i,P})^2}{\sum_{i=1}^{k_P} (y_{i,P} - y_{\text{mean},P})^2}}, \quad (5)$$

$$R_P^2 = 1 - \frac{\sum_{i=1}^{k_P} (y_{i,P} - \hat{y}_{i,P})^2}{\sum_{i=1}^{k_P} (y_{i,P} - y_{\text{mean},P})^2}, \quad (6)$$

$$M_{\text{RMSE},P} = \sqrt{\frac{1}{k_P} \sum_{i=1}^{k_P} (y_{i,P} - \hat{y}_{i,P})^2}, \quad (7)$$

式中: $k_C$ 和 $k_P$ 分别表示校正集和验证集中的样本数; $y_{i,C}$ 和 $y_{i,P}$ 分别表示校正集和验证集中第 $i$ 个样本的测量值; $\hat{y}_{i,C}$ 和 $\hat{y}_{i,P}$ 分别表示校正集和验证集中第 $i$ 个样本的预测值; $y_{\text{mean},C}$ 和 $y_{\text{mean},P}$ 分别表示校正集和验证集中所有测量值的平均值。

### 3 结果与讨论

#### 3.1 茶叶叶片荧光光谱分析和测量数据描述

在 600~900 nm 区域内,茶叶样品的荧光光谱如图 5 所示。从图 5 可以看到,在不同相对叶绿素含量的情况下,除了荧光波段强度有明显差异外,样品间共同光谱的特征基本相似。叶绿素在可见光波段有两个较强烈的吸收峰,分别位于 430~450 nm 的蓝紫光区域和 640~850 nm 的红光区域。在 430~600 nm 的波段中,可能是由儿茶素、茶黄素、花青素和叶绿素的共同作用所存在的光谱,红光波段(620~780 nm)和近红光短波段(780~900 nm)的叶绿素荧光光谱主要来源于光合作用的

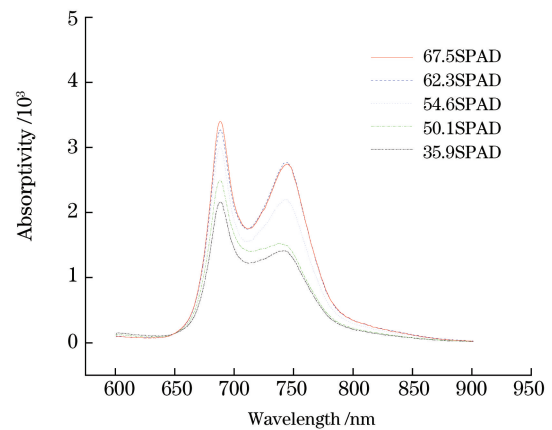


图 5 平滑处理后的鲜茶叶荧光光谱

Fig. 5 Fluorescence spectra of fresh tea leaves after smoothing treatment

叶绿素 a。

为了保证校正荧光光谱数据集模型的有效性和鲁棒性,需要合理的目标参数参考范围,因为过于狭窄的参考范围会限制模型的预测能力,所以采用基于光谱变量和真值信息的光谱-理化值共生距离(SPXY)算法<sup>[22-24]</sup>将样品分成标定集和预测集。校正集和预测集的描述性统计数据如表 3 所示。从表 3 可以看到,校正集与预测集的叶绿素含量范围和标准差均无显著性差异,并且样本集覆盖了足够大的范围,且校正集的范围大于预测集,这些特征有利于建立稳健的回归模型。因此,在校正集和预测集中,样本集的划分和参数的分布都是合适的。

表 3 校正集和预测集的描述性统计数据

Table 3 Descriptive statistics for calibration set and prediction set

Subset	Number of samples	Minimum	Maximum	Mean	SD
Calibration set	80	61.6±0.3	87.0±0.3	78.56	5.80
Prediction set	40	70.9±0.3	85.8±0.3	79.66	3.16

#### 3.2 PLS 定量模型的建立

为了定量叶片的相对叶绿素含量,光谱数据在通过 5 点平滑后,使用 600~900 nm 波段的荧光光谱建立定量模型。使用建立的 PLS 定量模型给出 X 和 Y 之间的定量关系,接着对预测集进行验证,评价定量模型的预测能力。表 4 为模型的评价指标结果。结果表明,荧光光谱结合化学计量学可以作

表 4 校正集和预测集的描述性统计数据

Table 4 Descriptive statistics of calibration set and prediction set

Method	Number of variable points	Calibration set		Prediction set	
		$R_c$	RMSE	$R_p$	RMSE
PLS	1044	0.9063	2.4359	0.9386	1.4871
UVE-PLS	160	0.9174	2.2208	0.9224	1.4073
SPA-PLS	10	0.9176	2.2896	0.9450	1.3069
BiPLS	280	0.9048	2.4545	0.9392	1.4861
BiPLS-SPA	14	0.9612	0.9483	0.9591	0.8863

为一种有效的叶绿素含量检测新方法。然而在 PLS 定量模型的建立过程中,采用过多的不相关或者线性变量会导致模型的效果降低,所以对模型进行优化。

### 3.3 模型的优化和比较

为了降低多变量单目标的计算复杂度,用于数据缩减和改善模型可预测性的变量选择算法已引起科研人员的关注。降低计算复杂度的关键在于最小化不相关变量的影响,并对敏感光谱区域和变量进行挑选。因此,对不同的变量选择算法进行对比,从而挑选出最优的光谱矩阵。

#### 3.3.1 无信息变量消除算法的波长筛选

无信息变量消除(UVE)法<sup>[25]</sup>的基本思想是将变量的回归系数平均值与其标准差的比值作为变量选择的衡量依据,表达式为

$$x_{cv} = \frac{\text{mean}(\mathbf{b}_j)}{\text{Std}(\mathbf{b}_j)}, \quad (8)$$

式中: $x_{cv}$ 为回归系数的平均值与标准差之比; $j$ 为光谱矩阵中的第 $j$ 列; $\mathbf{b}$ 为光谱矩阵。

最终的判断方法是先将一定数量的随机变量矩阵加入光谱矩阵中,然后通过交叉验证来建立 PLS 模型以得到回归系数矩阵,最后将随机变量矩阵的最大 CV 值( $x_{cv, \max}$ )作为阈值,当变量对应的  $x_{cv}$  值小于  $x_{cv, \max}$  值时,该变量被剔除。

图 6 为 UVE 法筛选的变量稳定性结果,其中两条水平虚线对应的稳定性值为筛选波长变量的阈值( $\pm 26.0486$ ),对上下阈值之间的光谱变量进行剔除,对上下阈值区间以外的光谱变量进行保留。图 7 为 UVE 变量筛选结果。从图 7 可以看到,经过 UVE 法筛选后,可以得到 160 个变量;经过 UVE 法筛选后的模型精度优于原始数据。

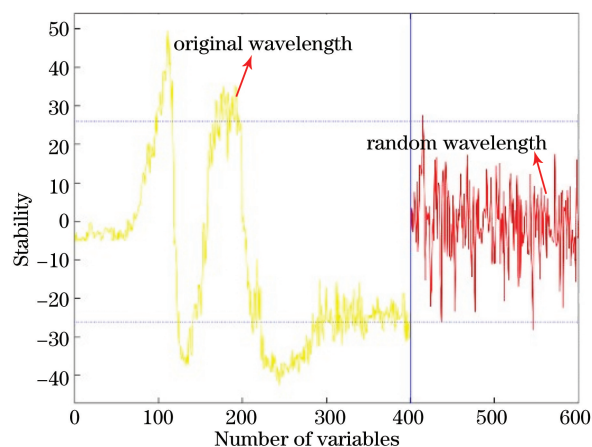


图 6 UVE 法筛选的变量稳定性结果

Fig. 6 Variable stability results of UVE method screening

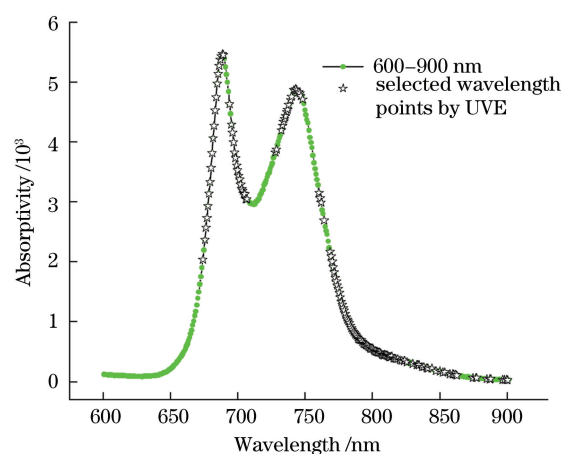


图 7 UVE 法筛选变量的结果

Fig. 7 Results of variable selection using UVE method

#### 3.3.2 连续投影算法的波长筛选

为了缩短信息处理的时间,使用最优特征变量而不是所有的光谱变量。信息光谱变量的选择还可以简化校正模型,并可以产生更稳健且准确的模型,采用 SPA 算法选择有效敏感波长点。实验中,在 600~900 nm 波段中对波长进行挑选,SPA 算法的基本思想是利用向量的投影分析找出包含最低限度的冗余信息的变量组,并且使变量之间达到最小的共线性,SPA 算法的变量筛选结果如图 8 所示。从图 8(a)可以看到,当 SPA 算法所选变量数为 10 时, RMSEP 值最低,并且随着所选变量数的进一步增加, RMSEP 值几乎不再降低,因此将这 10 个变量作为 SPA 算法选定的变量。从图 8(b)可以看到,选中的敏感波长点分别为 600.64, 682.44, 687.00, 697.63, 717.34, 730.95, 740.01, 743.03, 746.80, 767.88 nm, 所挑选的波长点均为叶绿素荧光的敏感位置,并且在模型的对比中, SPA-PLS 模型的预

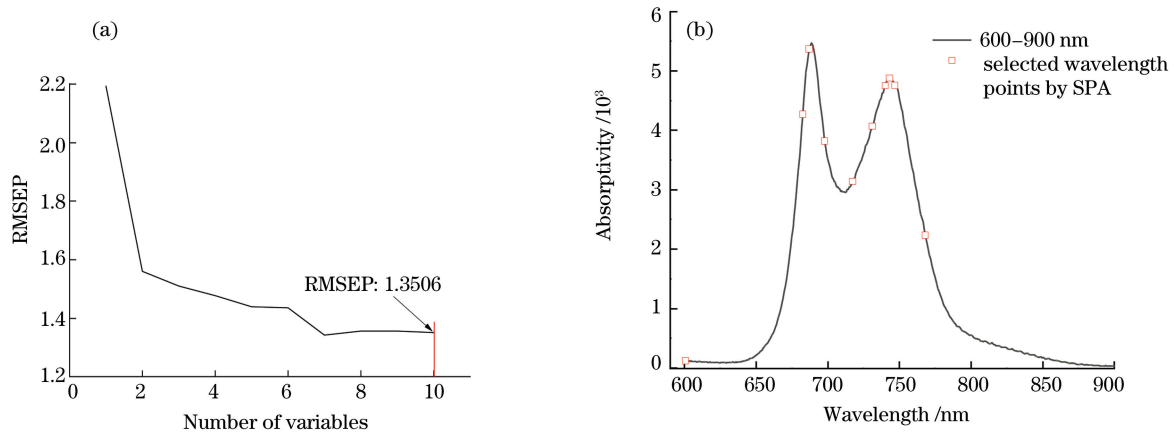


图 8 SPA 算法的变量筛选结果。(a) 选定变量的数目与 RMSEP 的关系曲线; (b) 选定变量在处理谱中的位置  
 Fig. 8 Results of variable selection using SPA algorithm. (a) Relationship between number of selected variables and RMSEP; (b) locations of selected variables in processed spectra

测能力都优于全谱 PLS 模型, 预测集的  $R_p$  值为 0.9450, RMSEP 值为 1.3069。结果表明, 为了减少计算量, 提高预测精度, 采用 SPA 算法挑选敏感波长是可行的。

### 3.3.3 后向区间偏最小二乘法结合 SPA 算法的波长筛选

SPA 算法是从 1044 个光谱变量中筛选出 10 个, 这大大减少了变量数, 因为 SPA 算法是在全谱范围内挑选敏感波长点, 所以存在模型过拟合的可能性, 故在 SPA 算法前加入后向区间偏最小二乘法以挑选出相关度高的波段, 之后再对波长进行挑选。

在区间 PLS 模型中, 数据集被分割成 20 个区间, PLS 模型在每次计算时都去掉一个区间。表 5 为 20 个子区间的模型评价指标相关数据, 每个子区间均包含 20 个变量。第一个被去除的区间间隔是 6, 当区间 6 被去除后剩余的 19 个区间的模型效果最佳, RMSECV 值最小。继续进行此过程, 直到只

剩下一个间隔区间(7), 得到每次去除区间后的模型参数如表 5 所示。从表 5 可以看到, 开始时, RMSECV 值随着子区间数的减少而减少, 但后来随子区间数的减少逐渐增加。当 RMSECV 值最小时, 所建立的 PLS 模型最佳, 此时 RMSECV 值为 2.3112, 校正集相关系数( $R_c$ )为 0.9048, 入选的子区间为(1,2,3,5,7,8,10,11,12,13,14,15,16,17), 共 280 个波长点。对预测集样本进行预测, RMSECV 值为 1.4861, 预测集相关系数( $R_p$ )为 0.9392。

BiPLS 算法从 1044 个光谱变量中筛选出 280 个, 这大大减少了变量数, 但是 PLS 的模型效果没有比全谱和 SPA 算法处理过后的模型好。可能的原因是 BiPLS 算法优选的 14 个区间范围较大, 仍存在一定的冗余信息且相邻的变量间仍存在较强的相关性, 所以对 BiPLS 算法优选后的 280 个变量进行 SPA 算法处理。

表 5 逐步选择最优间隔的结果

Table 5 Results of selecting optimal intervals step by step

Number intervals in model	Selected interval	Number of variables	RMSECV	Number intervals in model	Selected interval	Number of variables	RMSECV
20	6	400	2.6825	10	11	200	2.3902
19	9	380	2.4252	9	5	180	2.3643
18	4	360	2.3616	8	14	160	2.3413
17	18	340	2.3325	7	3	140	2.3402
16	19	320	2.3269	6	1	120	2.4077
15	20	300	2.3222	5	2	100	2.4248
14	17	280	2.3212	4	10	80	2.3955
13	16	260	2.3221	3	13	60	2.7288
12	15	240	2.3333	2	12	40	2.7245
11	8	220	2.3684	1	7	20	2.4189

在 BiPLS 算法得到的 280 个变量的基础上,采用 SPA 算法得到 14 个波长点,即 600.64, 661.12, 661.88, 666.45, 674.83, 683.20, 691.56, 693.84, 694.60, 702.94, 703.70, 705.22, 710.52, 779.89 nm。最终简化的 PLS 模型的预测结果如图 9 所示,其预测集相关系数  $R_p$  为 0.96, RMSE 为 0.87, 建模集的相关系数  $R_c$  为 0.96, RMSEC 为 0.95。结果表明,该方法对光谱的离散优化问题具有很大的吸引力,此外还可以将 BiPLS-SPA 程序应用于去除不相关的光谱信息。

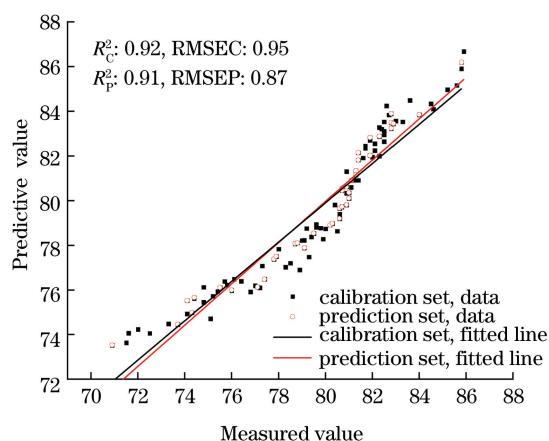


图 9 简化后的 PLS 模型的预测结果

Fig. 9 Prediction results of simplified PLS model

## 4 结 论

通过茶叶叶片的荧光光谱和 SPAD 数据验证了荧光光谱结合化学计量学方法对叶绿素相对含量的可预测性,对 UVE、SPA、BiPLS 和 BiPLS-SPA 算法进行比较,其中经 BiPLS-SPA 算法优化后所建立的模型的预测效果最好。实验结果表明,后向区间片最小二乘法与连续投影算法的结合是消除不相关光谱变量的有效变量选择工具之一,荧光光谱与有效的变量选择方法相结合在叶绿素相对含量的检测中具有很大的潜力,可能成为未来作物生长状况实时监测的一种有效技术。由于本次实验使用的是单点采集模式,可能对整片叶片的叶绿素相对含量产生误差,导致了校正集与预测集的方差较大,但是同时也说明了 PLS 模型的可收缩性较高。当然使用更多的数据和更科学的建模方法建立叶绿素含量估测模型将是未来的研究工作重点。

## 参 考 文 献

- [1] Liang L, Yang M H, Zhang L P, et al. Chlorophyll content inversion with hyperspectral technology for wheat canopy based on support vector regression algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering, 2012, 28(20): 162-171, 294.
- [2] Su Z S, Zhang X Z. Comparison of several methods for measuring plant chlorophyll content [J]. Plant Physiology Communications, 1989, 25(5): 77-78.
- [3] Shi J Y, Li W T, Guo Z M, et al. Nondestructive diagnostics of nitrogen and potassium deficiencies based on chlorophyll distribution features of cucumber leaves [J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(8): 264-269.
- [4] Chen B C, Fang J W, Du H. Photoacoustic spectroscopy for chlorophyll in plant leaves[J]. Plant Physiology Communications, 1993, 29(4): 276-278.
- [5] Pan L L, Xu X J, Tan J J, et al. Spectrophotometric determination of chlorophyll concentrations in maize leaves[J]. Chinese Journal of Analytical Chemistry, 2007, 35(3): 413-415.
- [6] Trypidis D, García-González D L, Lobo-Prieto A, et al. Real time monitoring of the combined effect of chlorophyll content and light filtering packaging on virgin olive oil photo-stability using mesh cell-FTIR spectroscopy [J]. Food Chemistry, 2019, 295: 94-100.
- [7] Liu J B, Han J C, Chen X, et al. Nondestructive detection of rape leaf chlorophyll level based on Vis-NIR spectroscopy [J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 222: 117202.
- [8] Yang J, Liao G P, Liu F, et al. Prediction of chlorophyll content of rape leaves with hyperspectral imaging technology [J]. Journal of Agricultural



- Science and Technology, 2020, 22(5): 86-96.
- 杨婧, 廖桂平, 刘凡, 等. 基于高光谱成像技术的油菜叶片叶绿素含量预测[J]. 中国农业科技导报, 2020, 22(5): 86-96.
- [9] Liu N, Xing Z Z, Qiao L, et al. Discussion on spectral variables selection of potato chlorophyll using model population analysis [J]. Spectroscopy and Spectral Analysis, 2020, 40(7): 2259-2266.
- 刘宁, 邢子正, 乔浪, 等. 基于模型集群的马铃薯叶绿素检测光谱变量筛选讨论[J]. 光谱学与光谱分析, 2020, 40(7): 2259-2266.
- [10] Shi J Y, Zou X B, Zhao J W, et al. NIR spectra in non-invasive measurement of cucumber leaf chlorophylls content[J]. Transactions of the Chinese Society for Agricultural Machinery, 2011, 42(5): 178-182, 141.
- 石吉勇, 邹小波, 赵杰文, 等. 黄瓜叶片叶绿素含量近红外光谱无损检测[J]. 农业机械学报, 2011, 42(5): 178-182, 141.
- [11] Yang Y Y, Chen B, Cai G M, et al. Study on nondestructive determination of chlorophyll content in tea canopy leaves by fluorescence spectroscopy[J]. Jiangsu Agricultural Sciences, 2012, 40(5): 270-272.
- 杨艳阳, 陈斌, 蔡贵民, 等. 茶鲜叶的荧光光谱特征与叶绿素含量之间的关系研究[J]. 江苏农业科学, 2012, 40(5): 270-272.
- [12] Sobron P, Wang A, Mayer D P, et al. Dalangtan saline playa in a hyperarid region of Tibet plateau: III. Correlated multiscale surface mineralogy and geochemistry survey [J]. Astrobiology, 2018, 18(10): 1277-1304.
- [13] Liu Y D, Zhang Y, Xu H, et al. Detection of sugar content of pomegranates from different producing areas based on near-infrared spectroscopy[J]. Laser & Optoelectronics Progress, 2020, 57(1): 013002.
- 刘燕德, 张雨, 徐海, 等. 基于近红外光谱检测不同产地石榴的糖度[J]. 激光与光电子学进展, 2020, 57(1): 013002.
- [14] Kalaji H M, Oukarroum A, Alexandrov V, et al. Identification of nutrient deficiency in maize and tomato plants by *in vivo* chlorophyll a fluorescence measurements[J]. Plant Physiol Bioch, 2014, 81: 16-25.
- [15] Ramadan Z, Hopke P K, Johnson M J, et al. Application of PLS and back-propagation neural networks for the estimation of soil properties [J]. Chemometrics and Intelligent Laboratory Systems, 2005, 75(1): 23-30.
- [16] Araújo M C U, Saldanha T C B, Galvão R K H, et al. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis[J]. Chemometrics and Intelligent Laboratory Systems, 2001, 57(2): 65-73.
- [17] Cheng Z, Zhang L Q, Liu H Y, et al. Successive projections algorithm and its application to selecting the wheat near-infrared spectral variables [J]. Spectroscopy and Spectral Analysis, 2010, 30(4): 949-952.
- 成忠, 张立庆, 刘赫扬, 等. 连续投影算法及其在小麦近红外光谱波长选择中的应用[J]. 光谱学与光谱分析, 2010, 30(4): 949-952.
- [18] Ye D D, Sun L J, Zou B R, et al. Non-destructive prediction of protein content in wheat using NIRS[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2018, 189: 463-472.
- [19] Wang J J, Shi T Z, Liu H Z, et al. Successive projections algorithm-based three-band vegetation index for foliar phosphorus estimation[J]. Ecological Indicators, 2016, 67: 12-20.
- [20] Norgaard L, Saudland A, Wagner J, et al. Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy [J]. Applied Spectroscopy, 2000, 54(3): 413-419.
- [21] Hao Y, Wu W H, Shang Q Y, et al. Analysis model of oleic and linoleic acids in camellia oil via near-infrared spectroscopy[J]. Acta Optica Sinica, 2019, 39(9): 0930004.
- 郝勇, 吴文辉, 商庆园, 等. 山茶油中油酸和亚油酸近红外光谱分析模型[J]. 光学学报, 2019, 39(9): 0930004.
- [22] Galvão R K H, Araújo M C U, José G E, et al. A method for calibration and validation subset partitioning[J]. Talanta, 2005, 67(4): 736-740.
- [23] Yang Z F, Xiao H, Zhang L, et al. Fast determination of oxide content in cement raw meal using NIR spectroscopy with the SPXY algorithm [J]. Analytical Methods, 2019, 11(31): 3936-3942.
- [24] Mao B H, Sun H, Liu H J, et al. Prediction of winter wheat chlorophyll content based on Gram-Schmidt and SPXY algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(S1): 160-165.
- 毛博慧, 孙红, 刘豪杰, 等. 基于正交变换与 SPXY 样本划分的冬小麦叶绿素诊断[J]. 农业机械学报,

2017, 48(S1): 160-165.

- [25] Liu Y D, Xu H, Sun X D, et al. On-line detection of soluble solids content of apples from different origins by visible and near-infrared spectroscopy[J]. Chinese

Optics, 2020, 13(3): 482-491.

刘燕德, 徐海, 孙旭东, 等. 不同产地苹果糖度可见近红外光谱在线检测[J]. 中国光学, 2020, 13(3): 482-491.