

## 基于高光谱技术的茶鲜叶茶多酚含量的估算模型

张民<sup>1a</sup>, 李银花<sup>1b,3</sup>, 袁晴春<sup>1a</sup>, 李娟<sup>1b,3</sup>, 戴思慧<sup>1b</sup>, 刘仲华<sup>1b,3</sup>, 李明<sup>1a,2\*</sup>

(1.湖南农业大学 a.工学院; b.园艺园林学院, 湖南 长沙 410128; 2.湖南省现代农业装备工程技术研究中心, 湖南 长沙 410128; 3.国家植物功能成分利用工程技术研究中心, 湖南 长沙 410128)

**摘要:**选取长势、色泽差异较小的9个茶树品种,使用ASD FieldSpec HandHeld 2光谱分析仪采集茶树冠层的光谱数据,通过对原始数据进行预处理去除噪音干扰,采用主成分分析方法得到茶树冠层特征波段520、765、821、940 nm,提取特征波段的反射率值,运用多元线性回归、一元线性回归、最小二乘法建立了光谱反射率与茶鲜叶茶多酚含量关系的估算模型。结果表明:最小二乘法模型的决定系数达到0.99;另选23个样品对模型进行验证,真实值与预测值的相关系数为0.97,相对误差为2.99%。

**关键词:**茶鲜叶;茶多酚;高光谱;最小二乘法模型

中图分类号: S571.1; O434.19

文献标志码: A

文章编号: 1007-1032(2015)04-0450-05

## Polyphenols inspecting method of living tea leaves based on hyperspectral reflection

Zhang Min<sup>1a</sup>, Li Yin Hua<sup>1b,3</sup>, Yuan Qingchun<sup>1a</sup>, Li Juan<sup>1b,3</sup>, Dai Sihui<sup>1b</sup>, Liu Zhonghua<sup>1b,3</sup>, Li Ming<sup>1a,2\*</sup>

(1.a.College of Engineering; b.College of Horticulture and Landscape, Hunan Agricultural University, Changsha 410128, China; 2.Hunan Provincial Engineering Technology Research Center for Modern Agricultural Equipment, Changsha 410128, China; 3.National Research Center of Engineering Technology for Functional Ingredients from Botanicals, Changsha 410128, China)

**Abstract:** Spectral data was captured for selected different growing and color of nine kind of tea by using ASD Field Spec Hand Held 2 spectrum analyzer. The raw data was pre-processing to decrease noise and improve acquisition. Multiple linear regression analysis, SG and PAC were employed to analysis the data. The optimization model was established to predict the best time of harvesting. The principal component characteristic bands of 520, 765, 821, 940 nm was obtained by smoothing, the first derivative, the second derivative and the pretreatment process. The determination coefficient is close to 0.99 for this model using least square method. By analyzing other 23 kinds of tea to verity the model, the correlation coefficient of real and predictive value is 0.97 as well as the average relative error of 2.99%.

**Keywords:** living tea leaves; polyphenols; hyperspectral; least square method model

为及时掌握茶叶的生长发育状况,通常需根据经验进行外观诊断,可靠性不高;或采用化学方法进行成分分析,但对植株有损伤,且过程较复杂,费时费力。高光谱遥感具有较高的光谱分辨率,通常达到 $10^{-2}$ 数量级,光谱信息能充分反映茶叶有效成分(包括多酚类、植物碱、氨基酸和蛋白质等含

氢基团成分)的特征信息,在多层次定量分析与应用中显现出良好前景<sup>[1-2]</sup>。刘建雄等<sup>[3]</sup>应用近红外光谱技术对茶园病虫害发生进行预测,结果预测值与实际值的决定系数 $R^2$ 为0.97,平均相对误差3.93%;赵杰文等<sup>[4]</sup>利用高光谱技术研究了茶树叶片中叶绿素含量及其分布;陈全胜等<sup>[5]</sup>基于高光谱图像技术

进行绿茶质量等级的评判研究,建立的等级判别模型预测识别率为 94%;胡永光等<sup>[6]</sup>利用可见-近红外非高光谱成像预测茶鲜叶全氮含量,建立了茶叶含氮量的估计模型。Dutta 等<sup>[7]</sup>利用卫星遥感和近红外光谱相结合的方法来鉴定 TV1 和 S3A3 两种茶叶的质量;Chaseok RYU 等<sup>[8]</sup>利用高光谱成像技术建立了茶叶产量估计模型。

笔者采用手持便携式光谱分析仪采集茶树冠层光谱数据,通过平滑、微分处理对光谱除噪后,选用主成分分析方法获得特征波段波长和反射率,运用高效液相色谱法测定茶鲜叶的茶多酚含量,建立了茶鲜叶茶多酚含量与高光谱特征波段反射率的估算模型,并进行了验证。现将结果报道如下。

## 1 茶树冠层高光谱数据的采集及处理

### 1.1 数据的采集

2013 年 4 月中旬至 10 月上旬,于湖南省长沙县高桥乡湖南省茶叶研究所(东经 113°19',北纬 28°29'),选择晴好天气,使用便携式近红外光谱仪 ASD Field Spec Hand Held 2(美国),采集 10:00—14:00 的茶树冠层高光谱数据。选择视场角为 25°,光谱仪器与茶树冠层的距离为 50 cm,检测区域的直径为 22 cm。抽取长势、色泽差异较小的 9 个茶树品种早白尖 5 号、金观音、福鼎大白茶、金萱、南江 1 号、碧香早、湘妃翠、寒绿、安徽 1 号,每

个样本区域测 4 个样点,每个样点采集 10 个光谱数据(通过剔除异常误差之后进行均化处理,设定仪器 2 s 扫描 1 次,为避免干扰每间隔 10~15 min 重新优化 1 次),取平均值作为样点的真实测量值。

每个样点采集 4 个茶鲜叶样品,按照文献<sup>[9]</sup>高效液相色谱法测定茶多酚含量。9 个茶树品种的春、夏、秋茶鲜叶的茶多酚含量列于表 1。

表1 茶树品种茶多酚的含量

品种	茶多酚含量/%		
	春茶	夏茶	秋茶
早白尖5号	22.823	11.823	17.969
金观音	26.866	17.650	19.095
南江1号	18.122	9.496	14.922
金萱	21.719	12.436	19.361
福鼎大白茶	23.698	13.243	25.811
碧香早	24.791	13.193	23.293
湘妃翠	18.093	10.283	21.874
寒绿	22.427	11.103	20.951
安徽1号	23.235	16.194	23.730

### 1.2 茶树冠层高光谱数据的处理

#### 1.2.1 茶树冠层的反射光谱特征

剔除茶树冠层光谱数据粗大误差的光谱曲线,对每个样点的光谱曲线进行均值处理,运用 View Spec Pro Version 6.0 软件得到光谱曲线图,如图 1 所示。

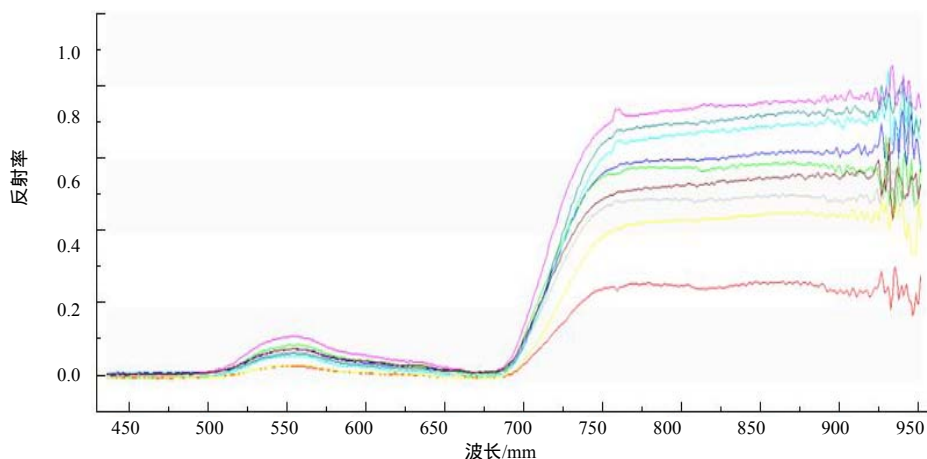


图 1 茶树冠层的高光谱图

Fig.1 Spectrum data of the sample tea

从图 1 可以看出,在光谱区间范围内,不同茶树品种冠层的光谱曲线的增长与递减是一致的。在波长 700 nm 处反射率开始有较大幅度的增大;在波长 765~960 nm 处各品种的反射率趋于稳定。

由于外界条件(如光线强度、空气中氧气浓度、仪器系统误差等)的影响,光谱曲线首尾两端出现了比较杂乱的曲线,去除全光谱数据中 325~425 nm 和 960~1 175 nm 的首尾噪声,以获取信噪比高的

波段,截取 425~960 nm 的平均光谱曲线进行研究。

在 425~900 nm 光谱范围内,9 个茶树品种冠层的平均样本曲线虽然走势基本相近,但在 520~670 nm 和 690~960 nm 波长范围内仍存在相对较大的光谱反射率差异。

### 1.2.2 光谱数据的预处理

为了减小噪声,提高信噪比,选取 4 种典型的平滑方法(“SG 卷积法”“多项式平滑”“相邻平均法”“傅里叶变换滤波”)进行比较分析。

将平滑处理后的数据进行线性拟合,得到不同平滑处理方法的模型精度(表 2)。

由表 2 可知,平滑处理后的模型预测均方差均优于未处理的,表明平滑处理可以去除由于测量过程中受环境因素影响而产生的大量细小而尖锐的杂峰,提高模型的精度。对比决定系数和预测均方

表 2 不同平滑处理方法的模型精度

数据平滑方法	$R^2$	$F$	预测均方差	平均值
SG 卷积方	0.91	1 763	5.190	13.180
相邻平均法	0.95	3 291	2.918	14.420
多项式平滑	0.91	1 741	5.650	14.780
傅里叶变换滤波	0.89	1 724	5.701	14.780
未处理	0.80	1 314	6.450	13.870

差值,相邻平均法的决定系数最高,预测均方差最小,因而采取相邻平均法对光谱数据进行平滑处理来去除噪音。

通过对茶树冠层光谱数据进行一阶和二阶导数处理,如图 2 所示,得到相关性的波段为 425~960 nm,一阶导数比二阶导数处理后的光谱线性变化率大,应用一阶导数进行光谱数据处理且设定差分宽度为 65,  $R$  为 0.98,预测均方差为 0.611,预测精度最高,可获得较好的预测效果。

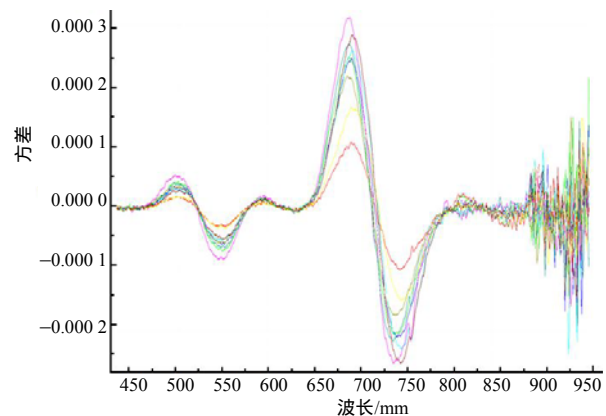
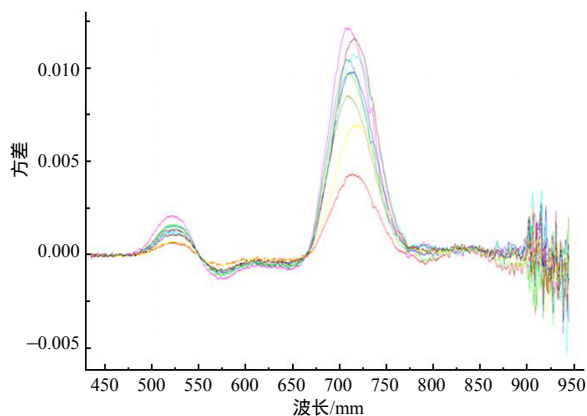


图 2 一阶和二阶导数处理后的特征波段

Fig.2 After the 1stDer and 2ndDer of the characteristics wave band

### 1.2.3 特征波段的选取

在去除噪音较大的 325~425 nm 与 960~1 175 nm 波段后仍有大量的波点数据。将每个样本的光谱数据中 425~960 nm 波长的波点通过 SPSS 软件进行主成分分析计算,第一主成分的方差贡献率为 82.68%,第二主成分的方差贡献率为 9.35%,第三主成分的方差贡献率为 6.32%,第四主成分的方差贡献率为 1.65%。根据主成分载荷阵,可以得到第一主成分波段 821 nm 的载荷为 0.99,第二主成分波段 765 nm 的载荷为 0.90,第三主成分中波段 940 nm 载荷率最高,第四主成分中波段 520 nm 载荷达到 0.88,故取 4 个主成分的波段反射率数据代表整个光谱数据的重要信息,据此可以提取出特征波段

和反射率,经过对原始光谱平滑处理与一阶导数预处理,由主成分分析得到特征波点,结合主成分分析中矩阵差,得到特征波段波长为 821、765、940、520 nm,特征波的反射率值分别用  $x_1$ 、 $x_2$ 、 $x_3$ 、 $x_4$  表示。

## 2 茶多酚预测模型的建立

### 2.1 高光谱特征波段的选取

由茶树冠层高光谱曲线的特征波段可以看出,不同品种之间的茶多酚含量与反射率存在差异,且同一品种的不同波段反射率也有明显不同。选用春茶茶多酚含量的数据作为建模基准,如表 3 所示。

表 3 茶多酚含量和不同特征波段反射率

**Table 3 Different samples of TP content and characteristics wave band reflectance**

品种	茶多酚含量%	$x_1$	$x_2$	$x_3$	$x_4$
早白尖 5 号	22.823	0.706	0.693	0.755	0.085
金观音	26.866	0.571	0.523	0.590	0.056
南江 1 号	18.122	0.741	0.740	0.791	0.108
金萱	21.719	0.730	0.727	0.762	0.092
福鼎大白茶	23.698	0.754	0.557	0.634	0.064
碧香早	24.791	0.694	0.662	0.745	0.060
湘妃翠	18.093	0.754	0.727	0.774	0.114
寒绿	22.427	0.725	0.721	0.752	0.088
安徽 1 号	23.235	0.758	0.680	0.695	0.072

表 3 数据表明，近红外光谱反射率  $x_1$ 、 $x_2$ 、 $x_3$ 、 $x_4$  和茶多酚含量存在相关性，多酚含量较高的品种吸收光能力较强，反射率较低。

### 2.2 茶多酚含量预测模型的建立

运用最小二乘法、多元线性回归、多项式法分

别对 4 个特征波段进行处理，得到茶多酚估算模型 (表 4)。其中， $y$  为茶多酚含量， $x$  为高光谱特征波段反射率。

表 4 不同方法建立的估算模型

**Table 4 The model equation based on the different mean**

模型	拟合方程	$R$	预测均方差	显著性
多项式法	$y = -0.0012x^4 + 0.1111x^3 - 3.8036x^2 + 57.5x - 322.81$	0.89	13.61	0.01
多元线性回归法	$y = 39.08 - 14.09x_1 + 15.89x_2 - 12.57x_3 - 129.01x_4$	0.97	2.41	0.01
最小二乘法	$y = 40.558 - 12.185x_1 + 14.57x_2 - 11.857x_3 - 129.111x_4$	0.99	1.96	0.00

从表 4 可以看出，最小二乘法建立的模型相关性最好，相关系数  $R$  达到 0.99，预测均方差为 1.96。最优模型为  $y = 40.558 - 12.185x_1 + 14.570x_2 - 11.857x_3 - 129.111x_4$ 。

结果表明，预测值与实测值的相关系数  $R$  为 0.97，平均相对误差值为 2.99%，说明对茶多酚的预测值与真实值误差小，可以应用于茶鲜叶茶多酚含量估算。

### 3 模型的验证

选取 23 个茶树品种对所建立的模型进行检验，采集茶树冠层的光谱数据，并将其特征波段的反射率值代入定量模型得出预测值，与实测值进行比较，如图 3 所示。

### 4 结论

通过采集茶树冠层的高光谱数据，对原始数据进行预处理，去除噪音干扰，采用主成分分析方法得到特征波段，提取出特征波段的反射率值。采用高效液相色谱测定样品的茶多酚含量，运用多元线性回归、一元线性回归、最小二乘法建立光谱反射率与茶鲜叶茶多酚含量的回归方程，其中最小二乘法模型精度最高， $R$  为 0.99，预测均方差为 1.96。选取 23 个茶树品种对模型进行验证，预测值与实测值的  $R$  值为 0.97，相对平均误差值为 2.99%，表明对检验样品中的茶多酚含量预测值与实测值误差较小，可实现利用高光谱技术对茶鲜叶茶多酚含量进行在线估算。

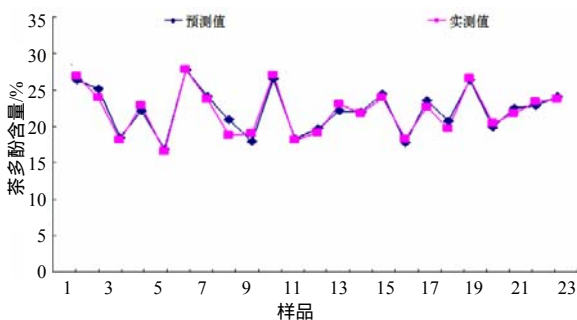


图 3 茶多酚含量实测值与预测值

Fig.3 The true data and predictive value scatter plot

## 参考文献:

- [1] 于海燕, 应义斌. 农产品品质近红外光谱分析结果影响因素研究综述[J]. 农业工程学报, 2005, 21(11): 168-171.
- [2] Sun D W. Hyperspectral Imaging for Food Quality Analysis and Control[M]. San Diego: Academic Press / Elsevier, 2009.
- [3] 刘建雄, 李明, 陶栋材, 等. 基于高光谱技术的茶尺蠖危害程度研究[J]. 湖南农业大学学报: 自然科学版, 2013, 39(3): 315-318.
- [4] 赵杰文, 王开亮, 欧阳琴, 等. 高光谱技术分析茶树叶片中叶绿素含量及分布[J]. 光谱学与光谱分析, 2011(2): 512-515.
- [5] 胡永光, 李萍萍, 母建华, 等. 基于可见-近红外光谱技术预测茶鲜叶全氮含量[J]. 光谱学与光谱分析, 2008, 28(12): 2821-2825.
- [6] 陈全胜, 赵杰文, 蔡健荣, 等. 利用高光谱图像技术评判茶叶的质量等级[J]. 光学学报, 2008, 28(4): 669-674.
- [7] Dutta R, Stein A, Bhagat R M. Integrating satellite images and spectroscopy to measuring green and black tea quality[J]. Food Chemistry, 2013, 127: 866-874.
- [8] Chanseok R Y U, Masahito Suguri, Miki Umeda. Estimation of the quantity and quality of green tea using hyperspectral sensing[J]. Journal of the Japanese Society of Agricultural Machinery, 2010, 72(1): 46-53.
- [9] 李健. 不同季节绿茶中茶多酚、咖啡碱和氨基酸含量分析[J]. 南京晓庄学院学报, 2012 (6): 57-59.

责任编辑: 罗慧敏

英文编辑: 吴志立

## 简 讯

## 由湖南农业大学牵头组建的“湖南畜禽安全生产协同创新中心” 通过专家评估考察

由湖南农业大学动物科学技术学院牵头, 中国科学院亚热带农业生态研究所和湖南省畜牧兽医研究所作为核心协同单位, 湖南新五丰股份有限公司、唐人神集团股份有限公司、佳和农牧股份有限公司、湖南临武舜华鸭业发展有限责任公司、湖南湘佳牧业股份有限公司等作为主要参与单位共同组建的“湖南畜禽安全生产协同创新中心”已顺利通过湖南省教育厅的形式审查、外围专家评审、专家会议评审和现场考察。

该中心立足于保障畜禽安全生产重大需求, 以服务区域经济社会和畜禽产业健康发展为目标, 以相对独立的非法人实体运行, 实行管理委员会领导下的主任负责制, 由印遇龙院士担任中心首席科学家, 文利新教授担任中心主任。设“多目标畜禽品种培育”、“多样化的区域饲料资源安全高效利用”、“多样化的区域型畜禽生产模式研究”3个创新平台和1个“公共服务与战略研究”平台。中心现有成员80人, 其中中国工程院院士1人, 百人计划、教育部新世纪人才、百篇优秀博士论文获得者等高层次人才16人。人员专业背景涵盖了畜牧学、兽医学、生物学等多个学科。1个创新团队入选湖南省高校创新团队。

该中心自组建以来, 着力探索人才培养机制改革, 重点构建了围绕产业服务的复合应用型卓越人才培养体系和培养方式, 将创新创业教育贯穿全过程, 与企业联合创办了22个创新创业人才班和3个卓越人才班。创建了国家生猪产业技术创新战略联盟, 与企业共建博士后工作站3个、博士生工作站6个、研究院和工程技术中心8个, 承担了64项国家、行业、地方和企业的重大重点科技项目, 取得了一批重要创新成果。协同攻关开发了首个国家二类新中兽药——博落回散(商品名: 美佑壮), 获2014年湖南省科技进步一等奖; 以湖南地方品种桃源黑猪为母本, 培育了国家级新品种——湘村黑猪; 开拓了新兴产业——功能性农牧业, 成功研发了国内外第一个低胆固醇功能性猪肉。