

基于 siPLS 法建立藤茶中二氢杨梅素 近红外定量分析方法

陈科力 姚 琪

(湖北中医药大学 教育部中药资源和中药复方重点实验室 武汉 430065)

摘 要 目的:建立藤茶中二氢杨梅素近红外定量模型,用于其快速定量分析。方法:采集 45 批藤茶样品的近红外光谱,以高效液相色谱法测定样品中二氢杨梅素的含量,并作为参照,以组合区间偏最小二乘法(siPLS 法)对特征谱段进行优化,建立二氢杨梅素的定量模型并对模型进行评价。结果:光谱预处理选用矢量归一化法(VN),确定 9000~8500 cm^{-1} 、8000~7000 cm^{-1} 、6000~5500 cm^{-1} 为建模谱段。内部交叉验证中,校正均方差(RMSECV)为 2.38%,决定系数(R^2)为 92.16%;外部验证中,模型对训练集预测均方差(RMSEE)为 0.993%,对验证集预测均方差(RMSEP)为 2.87%,其决定系数为 90.67%,相对预测偏差(RPD)为 3.27。当样品中二氢杨梅素的含量范围为 9.77%~46.13%时,平均预测回收率为 97.57%。结论:所建模型预测能力好,可用于藤茶中二氢杨梅素的快速定量分析。

关键词 近红外光谱法;藤茶;二氢杨梅素;siPLS;高效液相色谱法;定量分析

中图分类号 O657.33;R284.1 文献标识码 A 文章编号 1672-4321(2018)01-0041-05

Quantitative Model for Dihydromyricetin in *Ampelopsis grossedentata* by Near Infrared Spectroscopy Based on siPLS

Chen Keli, Yao Qi

(Key Laboratory of National Education Ministry on Traditional Chinese Medicine Resource and Compound Prescription, Hubei University of Chinese Medicine, Wuhan 430065, China)

Abstract Objective: To establish a NIR quantitative analysis model of Dihydromyricetin and a rapid method for the determination of Dihydromyricetin in *Ampelopsis grossedentata*. Methods: Using 45 batches of *Ampelopsis grossedentata*, 45 NIR spectral data were collected and preliminarily treated, the characteristic spectrum was optimized by the method of siPLS, the contents of dihydromyricetin were determined as the reference category by the method of HPLC, a NIR quantitative model of dihydromyricetin was established to determinate the dihydromyricetin in *Ampelopsis grossedentata* in a rapid way. Results: The results showed that the root mean square error of cross-validation (RMSECV) was 2.38%, the coefficient of internal cross-validation of dihydromyricetin (R^2) was 92.16%, the root mean square error of estimation (RMSEE) was 0.993%, the root mean square error prediction (RMSEP) was 2.87%, the coefficient of external validation of dihydromyricetin (R^2) was 90.67%, the relative prediction deviation (RPD) was 3.27, and the average predicted recoveries were 97.57% when the contents of dihydromyricetin ranged from 9.77% to 46.13%, and the spectral data got preprocessing by the method of VN in the calibration regions of 9000-8500 cm^{-1} , 8000-7000 cm^{-1} , and 6000-5500 cm^{-1} . Conclusion: The structure of dihydromyricetin model is proved to be rational. The model provides an efficient approach for the rapid determination of dihydromyricetin in *Ampelopsis grossedentata*.

Keywords near infrared spectroscopy; *Ampelopsis grossedentata*; dihydromyricetin; siPLS; HPLC; quantitative analysis

藤茶为葡萄科蛇葡萄属显齿蛇葡萄 (*Ampelopsis grossedentata*) 的干燥藤茎和叶,主产于湖北、湖南、

福建等地^[1],主治感冒风热,咽喉肿痛,黄疸型肝炎等。藤茶主要的有效成分为黄酮类化合物,其中二

收稿日期 2017-09-10

作者简介 陈科力(1947-),男,教授,博士生导师,研究方向:中药资源及其品质研究, E-mail:kelichen@126.com

基金项目 重大新药创制国家科技重大专项“中药新药安全性检测技术与标准研究”(2014ZX09304307001)

氢杨梅素的含量较高^[2,3]。研究表明,二氢杨梅素具有较好的降压、降脂、降糖和保肝作用,目前藤茶成为一些地区较常用的保健茶。但不同产地、不同采收期、不同批次藤茶中二氢杨梅素的含量差异很大,在本文的调查中,其含量范围为 9%~46%,在以往的测定中,曾发现更低含量。因此,建立一种快速分析方法有利于随时监控市售藤茶的质量。

藤茶的定量分析方法通常采用高效液相色谱法(HPLC)^[4],分析时间较长。近红外光谱(NIRS)技术是近年来发展迅速的光谱技术之一,其主要反映 C-H、O-H、N-H 等含氢基团的倍频与合频吸收,具有分析速度快、样品无损、无污染、操作简单等特点^[5]。在定性以及定量分析方面已得到广泛应用^[6,7]。拟采用 NIRS 技术与化学计量学方法相结合,采用组合区间偏最小二乘法(synergy interval partial least squares, siPLS)建立定量模型,用于藤茶中二氢杨梅素的快速定量分析。

1 仪器与材料

1.1 仪器

MPA 型傅里叶变换近红外光谱仪(德国 Bruker 公司),配有固体积分球漫反射附件、OPUS7.5 软件;MATLAB2014a 软件(美国 MathWorks 公司);UltiMate3000 高效液相色谱仪(美国 Dionex 公司),配有 UltiMate3000 四元梯度洗脱泵,UV170U 检测器;电子分析天平(瑞士 Precisa 公司)。

1.2 药材与试剂

45 批藤茶样品收集于湖北恩施、湖南张家界、福建漳州、福建泰宁。经湖北中医药大学药教研室陈科力教授鉴别,均为葡萄科显齿蛇葡萄(*Ampelopsis grossedentata*)的藤茎。二氢杨梅素对照品(成都曼思特生物科技有限公司,批号 MUST-17031501);甲醇为色谱纯;水为超纯水。

2 方法与结果

2.1 含量测定

参照文献[8,9]采用 HPLC 法测定样品中的二氢杨梅素含量。

2.1.1 对照品溶液的制备

精密称取二氢杨梅素对照品,置于 10 mL 容量瓶中,加甲醇定容,制成 1.002 g/L 的二氢杨梅素对照品溶液。

2.1.2 供试品溶液的制备

取干燥后的本品粉末(过 60 目筛)0.02 g,精密称定,浸于 10 mL 甲醇溶液中,超声提取 45 min,放冷,过滤,转至 10 mL 容量瓶中,定容摇匀,用 0.45 μm 微孔滤膜滤过,即得供试品溶液。

2.1.3 检测条件

色谱柱:Welchrom-C₁₈(250 mm×4.6 mm×5 mm);流动相:甲醇-0.05%磷酸溶液(35:65);流速 0.8 mL/min;进样量:20 μL;柱温:30℃;检测波长:290 nm。

2.1.4 标准曲线

分别精密量取对照品溶液 0、0.4、0.8、1.2、1.6、1.8 mL 置于 2 mL 容量瓶中,分别加入甲醇至刻度,摇匀,用 0.45 μm 微孔滤膜滤过,取续滤液,所得溶液按上述色谱条件进样测定。绘制峰面积(Y)对对照品浓度(X)曲线,得到二氢杨梅素含量的回归方程为 $Y = 1030X - 2.2249$ ($r = 0.9991$, $n = 5$),线性关系良好。

2.1.5 样品含量测定

按 2.1.3 色谱条件进样测定,记录 20 min 内色谱图,计算得供试品中二氢杨梅素含量在 9.77%~46.13% 范围内。

2.2 近红外光谱采集

分别将 45 批藤茶样品打粉,过 60 目筛,干燥,混合均匀置于石英样品杯中,在分辨率为 16 cm⁻¹、扫描时间为 64 次、扫描范围为 12500~4000 cm⁻¹ 条件下,每个样品重复扫描 3 次,求平均光谱。结果表明 45 批藤茶样品的近红外光谱图具有相似峰形(图 1),特征峰主要集中在 9000~4000 cm⁻¹。因此,建模谱段初步选定为 9000~4000 cm⁻¹。

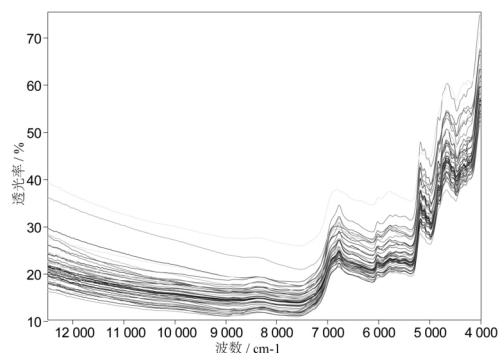


图 1 45 批藤茶样品的近红外光谱图

Fig. 1 Near infrared spectra of 45 batches of *Ampelopsis grossedentata*

2.3 二氢杨梅素含量的近红外光谱 siPLS 定量模型

2.3.1 样品集的划分

在 MATLAB2014a 中,采用 K-S 算法^[10],按照训

训练集和验证集 2 : 1 划分样本建立定量校正模型。各样品集样品浓度测定值的变化范围、平均值和样品量如表 1。

表 1 样品集信息
Tab.1 Information of sample set

样品集	C max / %	C min / %	\bar{C} / %	样品量
训练集	46.13	9.77	30.17	30
验证集	44.32	10.39	27.97	15

2.3.2 模型的建立与验证方法

采用 siPLS 算法将样品的光谱信息和含量信息进行匹配, 建立 siPLS 定量分析模型, 用于藤茶中二氢杨梅素的快速定量分析。建模过程中, 训练集通过内部交叉验证来优化模型以及评价模型性能, 以模型的交叉验证均方差 (RMSECV)、决定系数 (R^2) 作为评价指标; 采用外部验证对所建模型的预测能力进行评价, 以模型对训练集预测均方差 (RMSEE)、对验证集预测均方差 (RMSEP) 及其决定系数 (R^2) 作为评价指标。其中, 在内部交叉验证中, RMSECV 越小, R^2 越大, 所建模型越合理; 在外部验证中, RMSEE 不得大于 RMSECV, RMSEP 越小, R^2 越大, 结果与实测值越接近, 模型精度越高, 预测能力越强。此外, 所建模型对预测结果的相对预测偏差 (RPD) 作为评价模型预测能力的关键指标。RPD 为 1.75 ~ 2.25 时, 具有一定的建模意义; RPD 为 2.25 ~ 3 时, 建模比较成功; RPD 为 3 ~ 4 时, 模型具有较强实用性^[11]。建模过程中, 模型的主因子数 (Rank) 不宜过大, 否则会造成模型过拟合。因此, Rank 值在 1 ~ 10 范围内进行优化, 根据 RMSECV-Rank 关系图, 以 RMSECV 值最小确定最佳 Rank 值。

2.3.3 光谱的预处理方法筛选

采用矢量归一化 (VN)、减去一条直线 (constant offset elimination)、一阶导数 (FD)、二阶导数 (SD)、多元散射校正 (MSC) 等及其组合方法对光谱进行预处理^[12]。在特征谱段 9000 ~ 4000 cm^{-1} 范围内, 依次以不同预处理方法获得的光谱信息建立 siPLS 模

型, 评价模型的内部交叉验证结果, RMSECV 越小, R^2 越大, 预处理效果越好。

预处理方法为 VN 法时 (表 2), 模型的 RMSECV 最小, R^2 最大, 所建模型效果最好, 确定为本实验的预处理方法。

表 2 不同光谱预处理方法对定量模型的影响

Tab. 2 Effects of different preprocessing methods on the quantitative model

预处理方法	内部交叉验证		Rank
	R^2 / %	RMSECV / %	
未处理	25.61	7.34	6
VN	76.55	4.12	10
减去一条直线	31.66	7.04	4
FD 9 点平滑	40.75	6.55	9
SD 9 点平滑	30.48	7.10	1
MSC	43.64	6.39	3
FD+VN, 9 点平滑	33.00	6.97	6
FD+减去一条直线, 9 点平滑	50.77	5.97	9
FD+MSC, 9 点平滑	30.24	7.11	1

2.3.4 特征谱段的优选

在建立近红外定量分析模型时, 通常会对近红外的特征谱段进行筛选, 以消除干扰变量、简化模型以及提高预测的准确性。区间偏最小二乘法 (iPLS) 把谱区等分成 n 个子区间, 分别建立各子区间的回归模型以及该谱区的回归模型; 将该谱区的回归模型与各子区间的回归模型进行比较, 进而筛选出最佳特征谱段^[13]。但单一子区间的筛选可能会造成光谱信息的部分遗漏, 而区间组合偏最小二乘法 (siPLS) 可以弥补这一缺陷, 即通过不同区间数的任意组合而筛选出最优组合区间^[14]。

采用 VN 法对光谱进行预处理, 在 9000 ~ 4000 cm^{-1} 谱段范围内进行筛选, 将 9000 ~ 4000 cm^{-1} 谱段划分为 5、6、7、8、9 和 10 个区间, 依次进行 siPLS 运算, 选取各初始区间条件下所建模型的 RMSECV 值最小时为最佳组合区间, 所优选的特征谱段及其建模效果见表 3 以及图 2。由表 3 可确定区间数 $n = 10$ 的优化结果最佳。从图 2 RMSECV 与 Rank 关系图可知最佳 Rank 为 10。

表 3 siPLS 优选谱段及建模结果

Tab.3 The optimum spectrums and the modeling results by siPLS with VN

区间数 n	最佳谱段 / cm^{-1}	内部交叉验证		外部交叉验证		RPD	Rank
		R^2 / %	RMSECV / %	R^2 / %	RMSEP / %		
5	9000 ~ 7000	86.59	3.12	85.35	3.59	2.61	7
6	9000 ~ 7400, 6600 ~ 5800	85.97	3.19	85.72	3.55	2.66	8
7	9000 ~ 8900, 7600 ~ 6900, 6200 ~ 5500	88.68	2.86	79.10	4.29	2.22	8
8	9000 ~ 7800, 6600 ~ 6000	89.77	2.76	90.66	2.87	3.28	7
9	9000 ~ 7350, 6250 ~ 5700	87.30	3.03	89.14	3.09	3.07	7
10	9000 ~ 8500, 8000 ~ 7000, 6000 ~ 5500	92.16	2.38	90.67	2.87	3.27	10

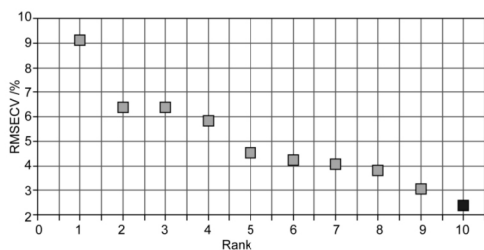
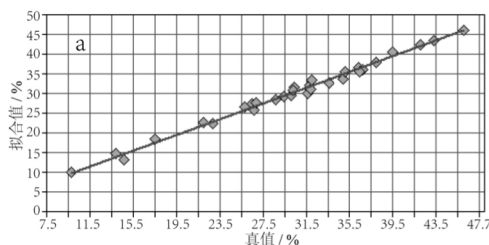


图 2 内部交叉验证均方差与主成分数关系

Fig.2 RMSECV vs Rank



a) 训练集; b) 验证集

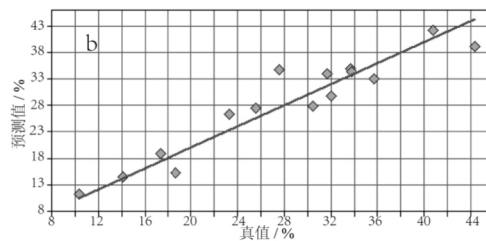


图 3 模型预测结果

Fig.3 Model prediction results

该模型对训练集预测结果 $RMSEE = 0.993\%$, 对验证集预测结果 $RMSEP = 2.87\%$, $R^2 = 90.67\%$, $RPD = 3.27 > 3$, 该模型结构合理, 预测能力强, 可满足实际应用的要求, 具有一定实用价值, 可用于藤茶中二氢杨梅素的含量测定. 15 批验证集藤茶样品中二氢杨梅素的含量预测结果见表 4.

表 4 验证集样品的预测结果

Tab.4 Prediction results of samples from test set

样品号	参考值 /%	预测值 /%	偏差 /%	预测回 收率/%	平均预测 回收率/%
1	23.30	24.92	1.62	106.95	
2	33.66	30.47	-3.19	90.52	
3	27.60	29.49	1.89	106.85	
4	31.65	32.72	1.07	103.38	
5	33.81	36.06	2.25	106.65	
6	18.69	17.87	-0.82	95.61	
7	32.03	29.53	-2.50	92.19	
8	25.59	27.93	2.34	109.14	97.57
9	30.46	29.80	-0.66	97.83	
10	35.71	33.72	-1.99	94.43	
11	17.35	13.24	-4.11	76.31	
12	44.32	38.95	-5.37	87.88	
13	10.39	11.12	0.73	107.03	
14	14.12	12.11	-2.01	85.76	
15	40.81	42.01	1.20	102.95	

3 结语

本研究通过对 45 批藤茶光谱数据的分析, 并以 HPLC 的测定结果作为参考, 采用 siPLS 法建立了二

2.3.5 模型评价和验证

根据上述结果, 以 VN 法预处理光谱, 并把 9000 ~ 4000 cm^{-1} 谱段划分为 10 个区间进行 siPLS, 确定的最佳特征谱段 9000 ~ 8500 cm^{-1} 、8000 ~ 7000 cm^{-1} 、6000 ~ 5500 cm^{-1} 组合作为建模谱段时, 所建二氢杨梅素的 siPLS 模型结构最合理, 预测能力最强. 将训练集和验证集样品光谱数据分别输入到所建模型中, 所得各样品集的预测结果见图 3.

氢杨梅素的近红外定量分析模型. 该模型具有较好的模型性能及预测能力, 可用于快速测定藤茶中二氢杨梅素的含量. 实验过程中, 与传统的高效液相色谱法相比, 近红外光谱法省去了繁琐的样品处理步骤, 大大提高了分析速度. 该方法为快速监控藤茶药材的质量提供了参考, 为藤茶的生产应用提供了依据.

参 考 文 献

- [1] 冯新玲, 陈科力, 李娟. 龙蔓藤茶增强免疫力功能的实验研究[J]. 中南民族大学学报(自然科学版), 2016, (1): 67-69.
- [2] 谢雪佳, 王晨光, 侯小龙, 等. 冷却结晶法纯化二氢杨梅素的工艺[J]. 医药导报, 2017, 8: 912-916.
- [3] 成凤桂, 欧知义. 鄂西藤茶中总黄酮的提取及含量测定[J]. 中南民族大学学报(自然科学版), 2005, (2): 19-22.
- [4] 李瑛琦, 于治国, 陆文超. HPLC 法测定藤茶中双氢杨梅黄素和杨梅黄素的含量[J]. 中草药, 2003, 34(12): 1098-1100.
- [5] 徐光通, 袁洪福, 陆婉珍. 现代近红外光谱技术及应用进展[J]. 光谱学与光谱分析, 2000, 20(2): 134-142.
- [6] 李彦周, 闵顺耕, 刘霞. 近红外光谱技术在中草药分析中的应用[J]. 光谱学与光谱分析, 2008, 28(7): 1549-1553.
- [7] 李燕, 吴然然, 于佰华, 等. 红外光谱在中药定性定

- 量分析中的应用[J]. 光谱学与光谱分析, 2006, 26(10):1846-1849.
- [8] 奉艳花,覃洁萍,王 淼,等. 不同干燥方法对藤茶主要活性成分二氢杨梅素的影响研究[J]. 药物分析杂志. 2015, 35(8):1483-1487.
- [9] 林诗瑶,阙金花,李 煌,等. HPLC 法测定藤茶的二氢杨梅素含量[J]. 福建中医药大学学报, 2014, 24(3):44-45.
- [10] Saptoro A, Tade M O, Vuthaluru H. A modified kennard-stone algorithm for optimal division of data for developing artificial neural network models: chemical product and process modeling [J]. Chemical Product & Process Modeling, 2012, doi:10.1515/1934-2659.1645.
- [11] Malley D F, McClure C, Martin P D, et al. Compositional analysis of cattle manure during composting using a field-portable near-infrared spectrometer [J]. Communications in Soil Science and Plant Analysis, 2005, 36(4-6):455-475.
- [12] 尼 珍,胡昌勤,冯 芳. 近红外光谱分析中光谱预处理方法的作用及其发展[J]. 药物分析杂志. 2008, 28(5):824-829.
- [13] 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.
- [14] 陈 龙,袁明洋,明 晶,等. 基于改进 siPLS 法建立 5 种含 CaCO_3 中药的拉曼光谱定量模型[J]. 中国中药杂志, 2015, 40(18):3608-3615.

(责任编辑 姚春娜)