

Bagging-PLS 的黄柏中试提取过程在线 近红外质量监测研究

周 正¹ 吴志生^{2,3} 史新元^{2,3} 王佳茜⁴ 乔延江^{1,2,3}

(1 福建中医药大学药学院,福州,350122; 2 北京中医药大学中药学院,北京,100102; 3 国家中医药管理局
中药信息工程重点研究室,北京,100102; 4 烟台市食品药品检验检测中心,烟台,264000)

摘要 目的:采用近红外光谱技术,建立黄柏中试提取过程偏最小二乘法(PLS)定量模型,快速分析盐酸小檗碱含量,以 Bagging-PLS 集成建模方法提高所建模型的预测性能。方法:以 HPLC 为参考方法测定黄柏提取液中盐酸小檗碱含量,运用 PLS 建立与近红外光谱(NIR)预测值之间的多元校正模型,并建立 Bagging-PLS 模型提高模型的稳定性与预测性能。结果:盐酸小檗碱定量模型的校正均方根误差 RMSEC 为 0.742 9,预测均方根误差 RMSEP 为 0.961 4,校正集相关系数 r_c 为 0.986 6,预测集相关系数 r_p 为 0.977 0;Bagging-PLS 模型预测均方根误差 RMSEP 均值为 0.413 6,预测集相关系数 r_p 均值为 0.998 3,表现出了良好的稳健性与预测性能。结论:所建立的近红外定量模型拥有较好的预测性能,能够实现快速实时检测黄柏提取液中盐酸小檗碱的含量;所建立的 Bagging-PLS 模型稳健性与预测性能良好,可应用于中药近红外在线监测过程中。

关键词 在线近红外;提取过程;黄柏;偏最小二乘;Bagging-PLS

Study on Online NIR for Monitoring the Extraction Process of Cortex Phellodendri in Pilot Scale Based on Bagging-PLS

Zhou Zheng¹, Wu Zhisheng^{2,3}, Shi Xinyuan^{2,3}, Wang Jiaqian⁴, Qiao Yanjiang^{1,2,3}

(1 Fujian University of Traditional Chinese Medicine, Fuzhou 350122, China; 2 Beijing University of Chinese Medicine, Beijing 100102, China; 3 Key Laboratory of Chinese Medicine Information Engineering, State Administration of Traditional Chinese Medicine, Beijing 100102, China; 4 Yantai Food And Drug Administration, Yantai 264000, China)

Abstract Objective: To establish a rapid quantitative analysis method for the quality control of Cortex Phellodendri extraction by using near-infrared (NIR) spectroscopy and improve the prediction performance and robustness of the model by using Bagging partial least square regression (Bagging-PLS). **Methods:** NIR spectroscopy was collected online, and a partial least square (PLS) model was developed for determination the content of berberine hydrochlorid with HPLC method as referential method. Besides, bagging algorithm was adopted for improving the prediction performance and robustness of the model. **Results:** The RMSEC and RMSEP were 0.7429 and 0.9614, the correlation coefficient of the calibration model was 0.9866, and the correlation coefficient of the prediction was 0.9770. The mean PMSEP of the Bagging-PLS model was 0.4136 and the correlation coefficient of the prediction was 0.9983. **Conclusion:** In this study, NIR spectroscopy was successfully applied to Cortex Phellodendri extraction process, which can rapidly determine the content of berberine hydrochloride. In addition, the Bagging-PLS provided a better performance and prediction accuracy than single PLS model. This indicates that Bagging-PLS could potentially be used for a better predictive control of the extraction process.

Key Words Online near-infrared; Extraction process; Cortex Phellodendri; Partial least square regression; Bagging-PLS

中图分类号:R284.2 文献标识码:A doi:10.3969/j.issn.1673-7202.2015.12.030

黄柏为芸香科植物黄檗(*Phellodendron amurense* Rupr.)及黄皮树(*Phellodendron chinense* Schneid.)的干糙树皮^[1],是一种常用中药,味苦性寒,归肾、膀

胱、大肠经,具有清热燥湿、泻火除蒸、解毒疗疮的功效。现代药理学研究表明,黄柏含有多种化学成分,具有抗菌、抗炎、降压、免疫抑制等药理作用^[2-4]。目

基金项目:北京中医药大学校级自主课题(编号:2015-JYB-XS081)

作者简介:周正(1990.09—),男,硕士在读研究生,研究方向:中药在线质量控制 Tel:(010)84738650,E-mail:saturday1226@163.com

通信作者:乔延江(1956.12—),男,博士,教授,博士生导师,研究方向:中药信息学,Tel:(010)84738661,E-mail:yjqiao@263.net;吴志生

(1983.12—),男,博士,副教授,研究方向:中药质量实时评价与微纳分析,Tel:(010)84738650,E-mail:wzs@bucm.edu.cn

前有关黄柏中指标成分盐酸小檗碱的含量测定多采用高效液相色谱法^[5-7],分析结果存在滞后性,难以满足在线分析的要求。近红外光谱技术具有分析速度快、绿色无污染、实时分析等优点,随着化学计量学的发展,在过程分析中得到了广泛应用^[8-11]。但近红外光谱存在光谱重叠及基线漂移等问题,需要借助化学计量学方法消除冗余信息,建立定量模型,而对于中药复杂体系,目前多建立的是单一 NIR 模型,模型的稳定性与预测性有待进一步提高。Bagging-PLS 作为一种集成建模方法,通过抽样重新选择样本集,不断重复抽样过程可得到多个回归模型,共同预测未知样本^[12],可以提高模型的稳健性与预测性能^[13-15]。

1 材料

XDS Rapid Liquid Analyzer 近红外光谱仪及其透射光纤(美国 Foss 公司),VISION 工作站(美国 Foss 公司);夹套式 100 L 多功能提取罐(天津隆业中药设备有限公司);Waters 2695 高效液相色谱仪及其 Waters 2996 二极管阵列检测器(美国 Waters 公司),赛多利斯十万分之一天平。黄柏药材(购自北京本草方源药业有限公司,批号 20130820);盐酸小檗碱标准品(中国食品药品研究院,批号 110713-201212);乙腈(色谱纯,美国 Fisher 公司);磷酸(分析纯);娃哈哈纯净水(杭州娃哈哈集团有限公司);提取用水为自制高纯水。

2 方法

2.1 样品制备 称取黄柏饮片 7 kg 置于 100 L 夹套式多功能提取罐中,提取两次,第 1 次加 12 倍水,提取 2 h,第 2 次加 10 倍水,提取 2 h。样品采集时间间隔为 5 min,按样品采集时间编号共得 65 个样品。

2.2 在线 NIR 采集光谱 光谱采集条件:在光程 2 mm 下,采用透射模式在线采集提取液吸收光谱,光谱范围为 800 ~ 2 200 nm,扫描次数 32 次,分辨率 0.5 nm,实验采用空气作为参比,采集的近红外光谱图见图 1。

2.3 HPLC 分析黄柏提取液中盐酸小檗碱含量 取盐酸小檗碱对照品配制成一系列不同浓度的标准溶液。DIKMA Diamonsil C₁₈ 色谱柱(4.6 mm × 250 mm, 5 μm);流动相 A 为 0.1% 磷酸水溶液, B 为乙腈,梯度洗脱,0 ~ 20 min, 10% ~ 30% B, 20 ~ 30 min, 30% ~ 50% B, 30 ~ 45 min, 50% ~ 90% B, 45 ~ 50 min, 90% ~ 10% B;柱温 30 ℃;流速为 1.0 mL/min;进样量 10 μL;检测波长 265 nm。将获得的样品用

25% 甲醇溶液稀释,过 0.45 μm 滤膜,取续滤液进行高效液相分析。

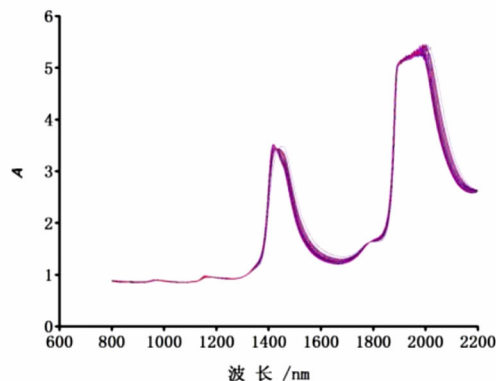


图 1 黄柏在线近红外原始光谱

2.4 数据处理 数据处理均在 Unscrambler 数据分析软件(version 9.6, 挪威 CAMO 软件公司)和 MATLAB(version 7.0, 美国 Math Works 公司)软件上完成。采用 Kennard-Stone 法(KS)将样本集划分为校正集与验证集。之后采用不同的光谱预处理方法,运用留一交叉验证法,以交叉验证验证均方根差 RMSECV 作为评价指标,选出最优预处理方法,建立全波段偏最小二乘(PLS)模型。采用组合间隔偏最小二乘法(SiPLS)对建模波段进行筛选,建立偏最小二乘模型,以校正均方根误差 RMSECV、校正集相关系数 r_c 、预测均方根误差 RMSEP、预测集相关系数 r_p 最为指标评价模型性能。在筛选出的最佳建模波段上采用 Bagging-PLS 集成建模方法建立 Bagging-PLS 模型,以预测均方根误差 RMSEP、预测集相关系数 r 最为指标评价模型性能。

3 结果与讨论

3.1 提取液中盐酸小檗碱 HPLC 测定结果 采用 2.3 中所建立的液相方法对样品中的盐酸小檗碱进行含量测定,图 2 为对照品与黄柏样品的色谱图。方法学考察主要参数参照药典与已经报道的文献,其分离度、精密度、重复性和稳定性均符合分析要求。所建立的线性方程为, R^2 为 0.999 5。

3.2 近红外校正模型的建立

3.2.1 光谱预处理 在近红外光谱采集过程中,近红外光谱受环境因素影响较大,温度、湿度等变化都会使其基线漂移噪音增大,因此再对近红外光谱进行分析之前,一般需要对原始光谱进行预处理来消除基线漂移,提高信噪比。本文比较了 Savitzky-Golay (SG) 平滑法,标准正态变换 (SNV),多元散射校正 (MSC),一阶导数 (1D) 以及它们的组合。表 1 为各预处理方法所建立的模型的性能

指标。采用留一交叉验证法，通过考察潜变量因子数对预测残差平方和（Press）的影响，确定最佳预处理方法，如图 3 所示。对于本文中黄柏提取

液，采用 SG11 点平滑预处理方法后 RMSECV 值最小，因此采用 SG11 点平滑法建立盐酸小檎碱的近红外定量模型。

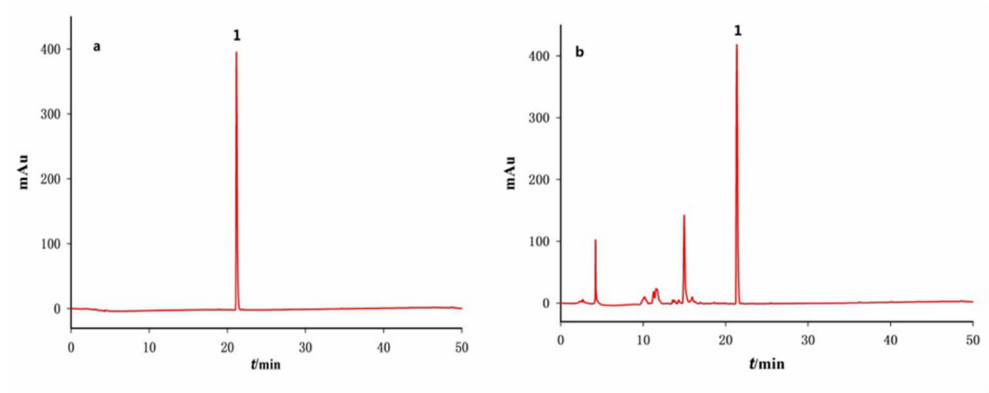


图2 对照品与黄柏提取液 HPLC 图

注:a. 盐酸小檎碱对照品,b. 黄柏样品,1. 盐酸小檎碱

表 1 不同光谱预处理方法的定量模型的性能指标

预处理方法	潜变量	r_c	RMSEC(mg/mL)	RMSECV(mg/mL)
Raw	5	0.9915	0.5921	1.3307
SG9	5	0.9901	0.6385	1.3052
SG11	5	0.9898	0.6436	1.2961
SG9 + 1D	7	0.9803	0.8988	4.0371
SG11 + 1D	4	0.9220	1.7048	4.2543
Normalize	5	0.9723	1.0660	1.6970
SNV	4	0.9738	1.0376	1.4895
MSC	4	0.9739	1.0352	1.4886

法为最佳预处理方法，SiPLS 筛选出的最优波段为建模波段，运用偏最小二乘法建立盐酸小檎碱的近红外定量模型，并采用内部样本集对建立的模型的预测性能进行验证。盐酸小檎碱 NIR 预测值与 HPLC 测定值的之间的线性关系良好，如图 5 所示。所建立的模型参数如下：交叉验证验证均方根差 RMSECV 为 0.952 4，校正均方根误差 RMSEC 为 0.742 9，预测均方根误差 RMSEP 为 0.961 4，校正集相关系数 r_c 为 0.986 6，预测集相关系数 r_p 为 0.977 0，表明所建立的模型预测性能良好。

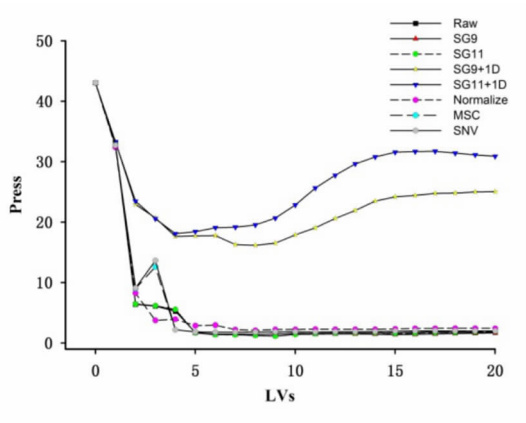


图3 不同预处理方法的 Press 值与潜变量数的相关关系图

3.2.2 波段选择 近红外光谱通常谱峰重叠严重，冗余信息过多，影响最终的建模结果，因此有必要对建模波段进行筛选来消除这些干扰因素。本文采用组合间隔偏最小二乘法（SiPLS）对黄柏提取液中盐酸小檎碱的建模波段进行筛选，间隔数 20，最大潜变量因子数 10，组合数 3。筛选出的最优波段为 1 080 ~ 1 150 nm，1 290 ~ 1 360 nm，1 710 ~ 1 780 nm，如图 4 所示。

3.2.3 模型的建立与校正 以 SG11 点预处理方

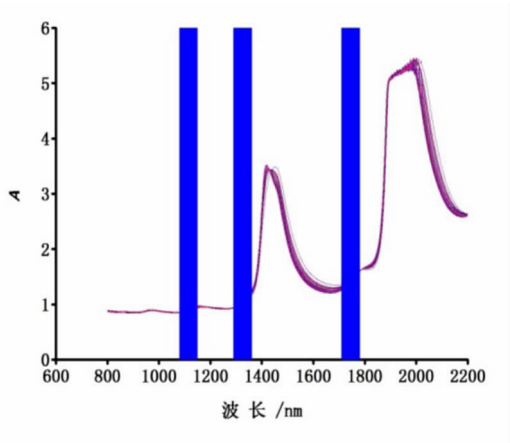


图4 盐酸小檎碱最优建模波段结果图

3.3 Bagging-PLS 应用于黄柏提取过程 NIR 定量模型 Bagging 是一种集成学习算法，应用在在分类与回归树种，能够提高模型的精度。本文中，在筛选出最佳建模波段的基础上建立 Bagging-PLS 模型，模型参数设置如下：校正集样本数与原始样本集相同，迭代次数为 1 000，潜变量为 6。图 6 为 Bagging-PLS 模型预测均方根误差 RMSEP 随迭代次数的变化趋势图，在迭代次数大于 500 的时候趋于稳定，均值为

0.413 6;图 7 为 Bagging-PLS 模型预测集相关系数 r_p 随迭代次数的变化趋势图,均值为 0.998 3,相较于 SiPLS 模型,Bagging-PLS 所建立的集成模型拥有更好的预测性能。

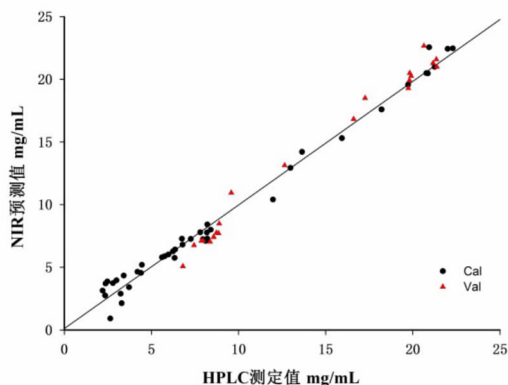


图 5 盐酸小檗碱 NIR 预测值与 HPLC 测定值相关关系图

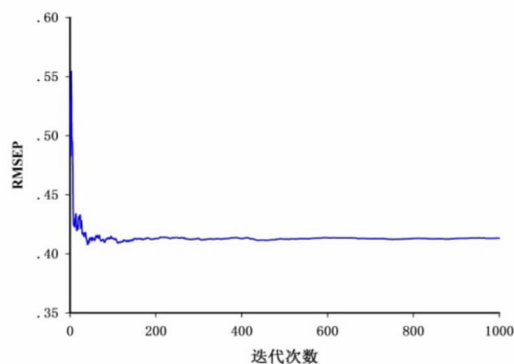


图 6 RMSEP 值变化趋势图

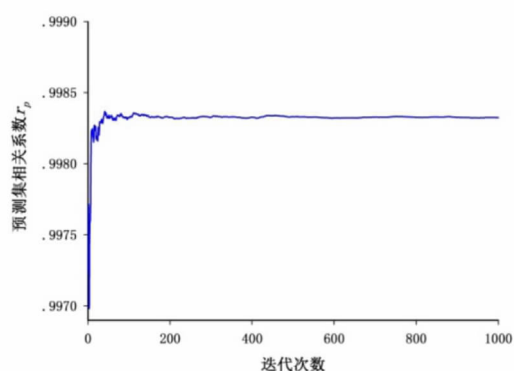


图 7 r_p 变化趋势图

4 结论

本研究采用近红外光谱技术对黄柏的中试提取过程进行了实时监测,建立了相应的近红外定量模型用于快速分析盐酸小檗碱的含量变化,且所得的模型预测性能良好。在此基础上所建立的 Bagging-PLS 集成模型预测均方根误差 RMSEP 均值 0.413 6,预测集相关系数 r_p 均值为 0.998 3,可以显

著提高单一 PLS 模型的预测性能,且表现出较强的稳健性。本研究结果为中药 NIR 在线监测与定量模型的建立提供了可供借鉴的新方法。

参考文献

- [1] 国家药典委员会. 中华人民共和国药典[M]. 一部. 北京:化学工业出版社,2010:286-287.
- [2] 李丹丹,江培,杨书美,等. 黄柏的化学成分,药理作用及临床应用研究进展[J]. 黑龙江医药,2014,27(3):601-605.
- [3] 陈兴文. 黄柏的化学成分分析和药理作用研究[J]. 内蒙古中医药,2014,33(31):69.
- [4] 吴嘉瑞,张冰,张光敏. 黄柏药理作用研究进展[J]. 亚太传统医药,2009,5(11):160-162.
- [5] 李红英,向极钎,龙澜,等. HPLC 法测定恩施黄柏药材中黄柏碱与小檗碱的含量[J]. 湖北民族学院学报:自然科学版,2013(1):101-103.
- [6] 李红玲,莫炫永. HPLC 法同时测定川黄柏中盐酸黄柏碱和盐酸小檗碱的含量[J]. 黑龙江医药,2014,27(3):601-605.
- [7] Yang Q, Zhang F, Gao S, et al. Determination of bioactive compounds in Cortex Phellodendri by high-performance liquid chromatography [J]. Journal of AOAC International, 2010, 93(3):855-861.
- [8] Li Y, Shi X, Wu Z, et al. Near-infrared for on-line determination of quality parameter of Sophora japonica L. (formula particles): From lab investigation to pilot-scale extraction process [J]. Pharmacognosy magazine, 2015, 11(41):8.
- [9] Wu Y, Jin Y, Ding H, et al. In-line monitoring of extraction process of scutellarein from Erigeron breviscapus (vant.) Hand-Mazz based on qualitative and quantitative uses of near-infrared spectroscopy [J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2011, 79(5):934-939.
- [10] Wu Y, Jin Y, Li Y, et al. NIR spectroscopy as a process analytical technology (PAT) tool for on-line and real-time monitoring of an extraction process [J]. Vibrational Spectroscopy, 2012, 58:109-118.
- [11] 李洋,吴志生,史新元,等. 中试规模 and 不同提取时段的黄芩配方颗粒质量参数在线 NIR 监测研究[J]. 中国中药杂志,2014,39(19):3753-3756.
- [12] 陈昭,吴志生,史新元,等. Bagging 偏最小二乘和 Boosting 偏最小二乘算法的金银花醇沉过程近红外光谱定量模型预测能力研究[J]. 分析化学,2014,42(11):1679-1686.
- [13] Shao X, Bian X, Cai W. An improved boosting partial least squares method for near-infrared spectroscopic quantitative analysis [J]. Analytica chimica acta, 2010, 666(1):32-37.
- [14] Shao X, Bian X, Liu J, et al. Multivariate calibration methods in near infrared spectroscopic analysis [J]. Anal. Methods, 2010, 2(11):1662-1666.
- [15] Pan X, Li Y, Wu Z, et al. A Online NIR Sensor for the Pilot-Scale Extraction Process in Fructus Aurantii Coupled with Single and Ensemble Methods [J]. Sensors, 2015, 15(4):8749-8763.

(2015-08-10 收稿 责任编辑:王明)