

基于波长优选的小柴胡颗粒黄芩苷近红外校正模型建立

李劲¹, 杨盼盼², 唐文旭¹, 李惠娟¹, 向琬¹,
蒋皖¹, 杨式华², 李伟², 侯英^{2*}

(1. 云南白药集团股份有限公司, 昆明 650500;
2. 云南同创检测技术股份有限公司, 昆明 650106)

[摘要] 目的:为提高小柴胡颗粒中黄芩苷近红外校正模型的准确性和预测精度。方法:基于模型的校正均方根误差(RMSEC)、预测均方根误差(RMSEP)、预测平均相对误差(PMRE)和模型对预测集的解释能力(Q^2)参数,对比评价竞争自适应重加权法(CARS),蒙特卡洛-无信息变量消除法(MC-UVE),遗传算法(GA),子窗口重排(SPA)算法筛选和全波长变量,采用模群集群分析(MPA)+偏最小二乘法(PLS)方法的建模效果。结果:校正模型准确性和预测精度:CARS>MC-UVE>GA>全波长变量>SPA;CARS算法所建立校正模型预测均方根误差为1.7004,决定系数 R^2 为0.9087,在 $\alpha=0.05$ 水平经配对 t 检验,50个外部验证样品实测值与预测值间无显著差异。结论:CARS算法筛选波长变量有效简化模型,提高模型预测的准确性和精度,适于小柴胡颗粒中黄芩苷含量的快速、无损检测。

[关键词] 小柴胡颗粒;黄芩苷;近红外技术;波长筛选;校正模型

[中图分类号] R284.1 **[文献标识码]** A **[文章编号]** 1005-9903(2016)18-0072-06

[doi] 10.13422/j.cnki.syfjx.2016180072

NIR Calibration Model to Detect Baicalin Content of Xiaochaihu Granules by Using Characteristic Spectrum Selection

LI Jin¹, YANG Pan-pan², TANG Wen-xu¹, LI Hui-juan¹, XIANG Wan¹,
JIANG Huan¹, YANG Shi-hua², LI Wei², HOU Ying^{2*}

(1. Yunnan Baiyao Group Co. Ltd., Kunming 650500, China;
2. Yunnan Comtestor Detection Technology Co. Ltd., Kunming 650106, China)

[Abstract] **Objective:** To improve the precision and accuracy of near-infrared spectroscopy (NIR) calibration model for determination of baicalin content in Xiaochaihu granules. **Method:** Four NIR characteristic spectrum selection methods were used including competitive adaptive reweighted sampling method (CARS), monte carlo uninformative variables elimination (MC-UVE), genetic algorithm (GA) and subwindow permutation analysis (SPA). Model population analysis (MPA) combined with partial least squares (PLS) was used to build five types of models, including CARS-PLS, MCUVE-PLS, GA-PLS, SPA-PLS and all NIR wavelength-PLS. Four model parameters including root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP), prediction mean relative error (PMRE) and the capability of model interpretation for testing set (Q^2) were analyzed to evaluate the precision and accuracy of those NIR calibration models. **Result:** precision and accuracy of the calibration models: CARS>MC-UVE>GA>all NIR wavelength>SPA. The values of RMSEP and decisive coefficients of the calibration models were 1.7004 and 0.9087 respectively, whose characteristic spectrum had been selected by CARS. According to the result of paired- t tests at the level of α of 0.05, there was no

[收稿日期] 20150825(002)

[基金项目] 云南白药集团自立科技项目(ZZ-2014-0076)

[第一作者] 李劲, 硕士, 正高级工程师, 从事药品技术质量研究, Tel:0871-68322467, E-mail:lijin@yunnanbaiyao.com.cn

[通讯作者] *侯英, 硕士, 正高级工程师, 从事近红外分析技术应用研究, Tel:0871-65337113, E-mail:houying@reascend.com

significant difference between prediction values and measure values in 50 samples. **Conclusion:** CARS algorithm to screen wavelength can be used to effectively simplify the models, and improve the precision and accuracy of the models, suitable for fast and nondestructive detecting the baicalin content of Xiaochaihu granules.

[Key words] Xiaochaihu granules; baicalin; NIR technology; spectrum selection; calibration model

小柴胡颗粒由柴胡、人参、黄芩、大枣、半夏、生姜、甘草 7 味药组成,具有解表散热、疏肝和胃的功效,被广泛用于寒热往来、胸胁苦满、心烦喜吐、口苦咽干等病证^[1]。黄芩苷是小柴胡颗粒主要有效成分之一,含量高低直接影响着小柴胡颗粒的药效,其具有抗氧化^[2]、预防缺血性中风^[3]、解热^[4]、防中暑^[5]等多种药理活性。目前,中药材及制剂中黄芩苷含量的测定方法包括高效液相色谱法^[6-7]、微乳液相色谱法^[8]、毛细管电泳法^[9]、胶束电动色谱-质谱法^[10]等,上述方法定量准确,但存在前处理繁琐,测定时间较长等不足。

近红外分析技术作为一种简便、快速、无损的检测方法,已被广泛应用于各行业产品品质的快速分析^[11-14]。近红外检测技术应用于黄芩^[15]及银黄颗粒^[16]中黄芩苷含量的快速测定已有文献报道,但利用其快速测定小柴胡颗粒中黄芩苷含量未见文献报道。为提高近红外分析方法的准确性,提升近红外校正模型的准确性和预测精度,可对波长变量进行筛选^[17],常用的波长变量筛选方法包括模拟退火算法^[18]、遗传算法^[19]、粒子群算法^[20]、连续投影算法及蒙特卡罗无信息变量等方法^[14]。近年来,梁逸曾教授研究团队提出的模型集群分析(Model Population Analysis, MPA)弥补了常规近红外一次性建模所采用的校正样本集难以完全表达问题样本总体信息的不足^[21]。MPA 结合常用波长变量筛选方法能最大限度的考察波长变量对校正模型建立的“有效性”,进一步提高波长筛选的效果^[22-23]。

为提高小柴胡颗粒中黄芩苷近红外校正模型的准确性和预测精度,本文应用模型集群分析,采用蒙特卡罗采样方法重复选择校正集和内部验证集,结合竞争自适应重加权法(Competitive Adaptive Reweighted Sampling, CARS)^[24-25],蒙特卡罗-无信息变量消除法(Monte Carlo-Uninformative Variable Elimination, MC-UVE)^[26-27],遗传算法(Genetic Algorithm, GA)^[28]和子窗口重排(Subwindow Permutation Analysis, SPA)^[25]筛选波长变量,建立偏最小二乘(Partial Least Square, PLS)校正模型,对比评价 4 种算法筛选的波长变量和全波长变量的建模效果,确定最优的波长变量筛选算法及最佳的建

模波长变量,评价校正模型预测的准确性和精度,结果满意。通过波长变量筛选,有效简化校正模型,同时提高预测的准确性和精度,适用于小柴胡颗粒中黄芩苷含量的快速、无损检测。

1 材料

小柴胡颗粒(国药准字 Z53021558),云南白药集团股份有限公司 2013 年 3 月至 9 月生产,14 个批次,共计 762 个,样品规格 10 g/袋。

AcQuity 型 UPLC(美国 Waters,包括 PDA 检测器,Empower 2 色谱工作站),Acquity UPLC BEH C₁₈ 色谱柱(2.1 mm × 50 mm, 1.7 μm, Waters)。黄芩苷对照品(中国食品药品检定研究院,纯度 93.3%,批号 110715-201016);甲醇、乙醇色谱纯,磷酸分析纯,蒸馏水(昆明物理研究所,≥18 MΩ)。

Antaris II 型 FT-NIR 光谱仪,配备 InGaAs 检测器的漫反射积分球,4.78 cm 石英杯,样品旋转台等采样附件,RESULTTM集成软件,TQ Analyst 8.6(美国 Thermo Fisher 公司);SCINO CT410 型旋风样品磨(美国 FOSS 公司),MATLAB 6.0 数据处理软件。

2 方法与结果

2.1 黄芩苷含量测定 样品测定参照 2015 年版《中国药典》^[1]方法。UPLC 色谱条件,流速 0.3 mL·min⁻¹,检测波长 315 nm,进样量 1.0 μL,柱温 35 °C,流动相甲醇-水-磷酸(47:53:0.2)。以对照品进样浓度为 X,峰面积为 Y,进行线性回归,得回归方程为 $Y = 7.53 \times 10^6 X - 21\,200$ ($r = 0.999\,993$),线性范围 0.059 2 ~ 0.295 9 g·L⁻¹,对照品及代表性样品 UPLC 色谱图,见图 1。

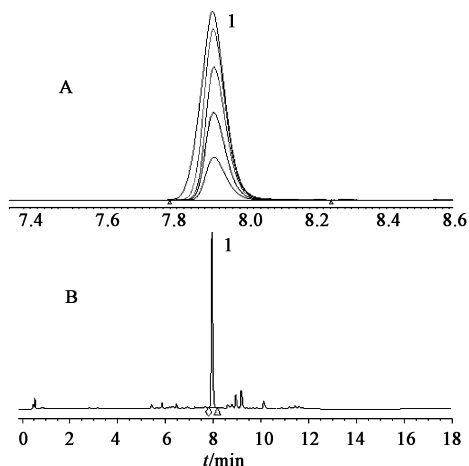
2.2 小柴胡颗粒近红外光谱的采集

2.2.1 样品前处理 将小柴胡颗粒粉碎过 60 目筛,制成粒径大小均匀、水分相对稳定的粉末样品(约 60 g),密封备用。

2.2.2 光谱数据采集 光谱扫描范围 4 000 ~ 10 000 cm⁻¹,分辨率 8 cm⁻¹,扫描次数 64 次,开机预热 2 h,采集样品的近红外光谱。

2.3 数据处理和波长变量筛选

2.3.1 光谱预处理和异常样本的剔除 多元散射校正(Multiplicative Scatter Correction, MSC)可消除近红外光谱分析中颗粒分布不均匀及颗粒大小产生



A. 不同浓度对照品重叠色谱; B. 小柴胡颗粒; 1. 黄芩苷

图 1 小柴胡颗粒代表样品 UPLC

Fig. 1 UPLC chromatograms representative baicalin sample

的散射影响, Karl Norris 滤波器 (Karl Norris Derivative Filter) 和二阶求导 (Second Derivative, SD) 可过滤噪声、消除基线漂移^[29]。因此采用 MSC + Norris (11, 3) 平滑 + SD 对样品近红外光谱进行预处理, 预处理前后的光谱见图 2。由图 2b 可知 9 000 ~ 10 000 cm^{-1} 存在高频噪声, 因此选择 4 000 ~ 9 000 cm^{-1} 作后续分析。

采用模群迭代奇异样本诊断法^[30] 多次循环建模, 剔除异常样本。采用模群迭代奇异样本诊断法对异常样本进行挑选, 当预测误差的均值绝对值或方差大于 $\mu_{\text{STD}} + 3\sigma_{\text{STD}}$ (其中 μ_{STD} 表示均值的均方差; σ_{STD} 表示 μ_{STD} 的均方差) 时, 可认定样品异常^[21]。光谱预测误差的均值-方差分布见图 3, 由图

表 1 4 种算法主要参数及变量重要性评价指标

Table 1 Main parameters and evaluation index of four algorithms

算法	算法主要参数	变量重要性评价指标
CARS	最大主成分数 15, 蒙特卡罗运算样本数 50, 交互验证样本数 5, 运算次数 100	频率运行 100 次, 统计每个变量选中频率
MC-UVE	交互验证的最大主成分数 15, 蒙特卡罗采样次数 100	变量稳定指数 (Reliability Index, RI) ^[31]
GA	运算次数 100, 种群个数 30, 一次模拟最大变量个数 30, 突变率 1%, 最大的主成分数 15	频率运行 100 次, 统计每个变量选中频率
SPA	蒙特卡罗采样次数 100, 训练样本占总样本的比例 0.85, 每次采样的样本数 20	条件协同得分 ^[21] (Conditional Synergetic Score, COSS) 变量基于 Mann-Whitney U 分布差异性检验的显著性 P 的负对数

采用 CARS, MC-UVE, GA 和 SPA 方法分别计算波长变量重要性参数, 见图 4。由图 4 可知, CARS 算法筛选出的波长变量均匀的分布在 4 000 ~ 9 000 cm^{-1} ; SPA 筛选出的波长变量较集中的分布在 4 000 ~ 5 300 cm^{-1} 和 6 800 ~ 6 950 cm^{-1} ; MC-UVE 和 GA 算法筛选出的波长变量多于上述 2 种方法, 4

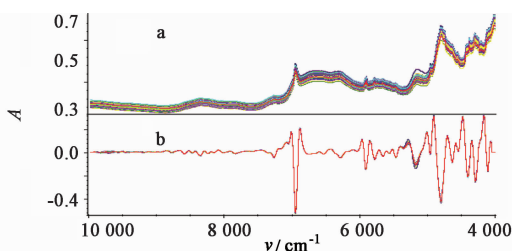


图 2 小柴胡的原始近红外光谱 (a) 和预处理后的光谱 (b)

Fig. 2 Raw (a) and pretreatment (b) near infrared spectroscopy of Xiaochaihu granules

3 可知异常样本个数 (outlier) 为 13 个, 剔除异常样品后, 样本集样品个数为 749, 黄芩苷含量为 24.70 ~ 47.80 $\text{mg}/\text{袋}$, 均值为 37.62 $\text{mg}/\text{袋}$ 。

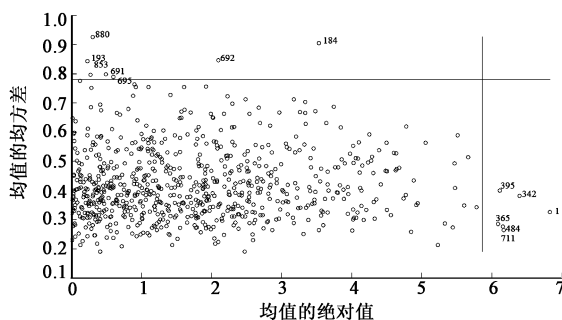


图 3 模群迭代奇异样本诊断

Fig. 3 Result of regression diagnosis and screening of outlier based on Monte-Carlo sampling

2.3.2 波长变量筛选 剔除异常样本后, 应用 CARS, MC-UVE, GA 和 SPA 方法计算每个波长变量对应的重要性指标参数。4 种算法主要参数和对应的波长变量重要性评价指标, 见表 1。

种方法筛选出的波长范围见表 2。

基于蒙特卡罗采样 (Monte Carlo Sampling) 方法, 采用向前选择法, 按变量重要性指标从高到低逐个引入波长变量 (最大变量数设为 250) 结合偏最小二乘法 (PLS) 重复建模, 采用蒙特卡罗交叉验证^[21] 评价模型的预测能力, 并以最小或变化趋势不明显

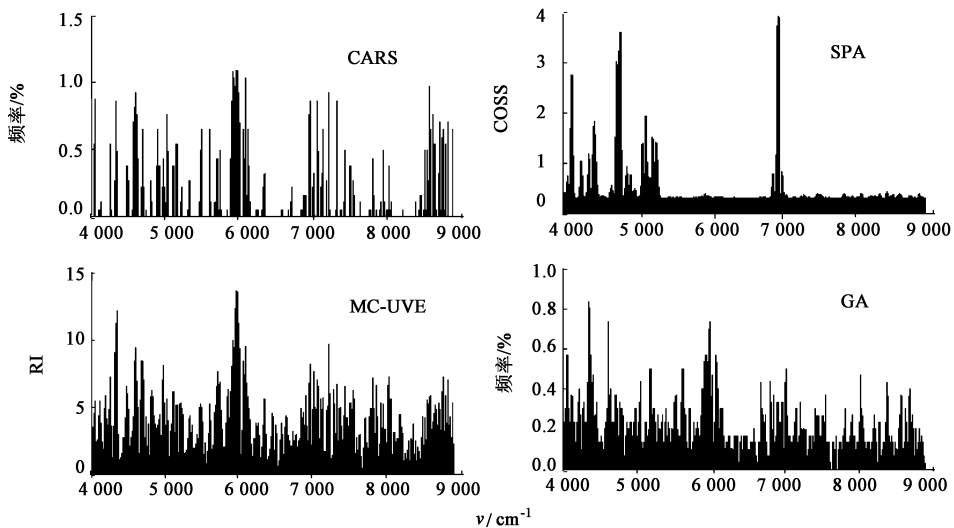


图 4 CARS,MC-UVE,GA 和 SPA 变量(波数)重要性
Fig. 4 Importance plot of spectrum variable of CARS, MC-UVE, GA and SPA

表 2 CARS,MC-UVE,GA 和 SPA 筛选的主要波长变量范围
Table 2 Screened spectrum range of CARS, MC-UVE, GA and SPA

波长筛选方法	筛选的波长范围/ cm^{-1}
CARS	4 042.07 ~ 4 250.34, 4 601.32 ~ 4 693.89, 5 881.82 ~ 5 997.53, 6 043.81 ~ 6 078.53, 7 100.61 ~ 7 197.04, 7 922.14 ~ 7 999.28, 8 512.25 ~ 8 612.53, 8 712.81 ~ 8 855.52
SPA	4 076.78 ~ 4 134.63, 4 223.34 ~ 4 246.48, 4 713.17 ~ 4 786.45, 4 836.59 ~ 4 898.31, 5 041.01 ~ 5 272.43, 6 807.49 ~ 6 957.91
MC-UVE	4 315.91 ~ 4 354.48, 4 470.19 ~ 4 493.33, 4 582.04 ~ 4 620.61, 4 659.18 ~ 4 705.46, 4 801.88 ~ 4 836.59, 5 677.41 ~ 5 731.4, 5 877.97 ~ 5 997.53, 6 039.96 ~ 6 090.1, 6 934.77 ~ 6 996.48, 7 181.61 ~ 7 200.89, 7 991.57 ~ 8 026.28, 8 512.25 ~ 8 562.39
GA	4 030.5 ~ 4 123.06, 4 204.06 ~ 4 277.34, 4 319.77 ~ 4 443.19, 4 582.04 ~ 4 609.04, 4 693.89 ~ 4 717.03, 4 755.6 ~ 4 794.17, 5 156.72 ~ 5 218.43, 5 511.56 ~ 5 607.98, 5 850.97 ~ 6 005.25, 6 043.81 ~ 6 078.53

的交互验证均方根误差 (Root Mean Square Error of Cross Validation, RMSECV) 确定最佳变量个数。计算所选变量数对应的交叉验证均方根误差值, 见图 5。

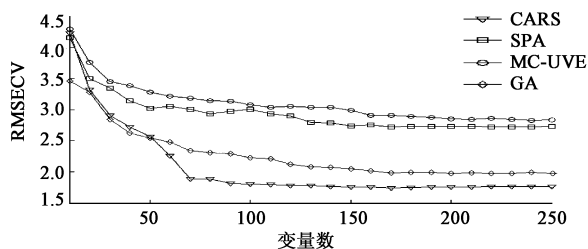


图 5 RMSECV 随变量增加的变化趋势
Fig. 5 Changing trend of RMSECV vs number of screened spectrum variable

由图 5 可知, 随着波长变量数的增加, 4 种方法筛选变量所建模型的 RMSECV 均逐渐减小。当变量数在 60 ~ 100 时, CARS 方法对应的 RMSECV 值减小的程度最大; 当变量数 > 60 时, CARS 方法对应

的 RMSECV 值最小, 采用 CARS, GA, MC-UVE 和 SPA 方法确定最优的建模变量数依次为 172, 170, 214, 215。

2.4 校正模型的建立及评价 基于模群集群分析 (MPA), 采用蒙特卡罗采样技术对 699 个样本随机抽样选取 560 个样品作为校正集, 139 作为内部验证集; 余下 50 个样品作为外部验证集样品, 不参与建模。分别采用 CARS, MC-UVE, GA, SPA 筛选的波长变量和全波段波长变量结合偏最小二乘法 (Partial Least Squares, PLS) 反复建模 500 次, 统计 500 次模型校正集的 RMSEC, 外部验证集的 RMSEP, PMRE 和 Q^2 的平均值和标准偏差, 作为校正模型预测准确性和精度的评价标准, 模型各参数的平均值和标准偏差见表 3。

由表 3 可知, 应用 CARS 算法筛选出建模波长建立模型 RMSEC, RMSEP 和 PMRE 的平均值和标准偏差在上述 5 种方法中均最小, Q^2 平均值最大的波长筛选算法为 CARS, 其对应的标准偏差为

表 3 CARS, MC-UVE, GA 和 SPA 筛选波长变量建模的 RMSEC, RMSEP, Q^2 和 PMRE

Table 3 Range of RMSEC, RMSEP, Q^2 , and PMRE of CARS, MC-UVE, GA and SPA algorithms

模型参数	CARS		SPA		MC-UVE		GA		全波长	
	平均值	标准偏差	平均值	标准偏差	平均值	标准偏差	平均值	标准偏差	平均值	标准偏差
RMSEC/(mg/袋)	1.63	0.023	2.49	0.031	1.66	0.038	1.77	0.027	1.66	0.04
RMSEP/(mg/袋)	1.77	0.090	2.75	0.140	1.88	0.097	2.02	0.100	2.00	0.092
Q^2	0.90	0.013	0.76	0.03	0.89	0.015	0.87	0.017	0.87	0.016
PMRE/%	3.85	0.220	6.06	0.371	4.11	0.237	4.38	0.248	4.36	0.236

0.013; 通过 500 次建模, 比较上述 4 种模型参数的标准偏差可知, CARS 方法筛选的波长变量所建立的校正模型具有更好的稳健性; CARS 算法筛选的波长变量为 172, 在 4 种波长算法中较小, 在减少模型复杂性方面具有较好的优势。因此, 竞争性自适应重加权法 (CARS) 在减少模型复杂性、提高模型预测精度和稳健性方面优于其他 3 种算法。

应用 CARS 筛选变量建立的小柴胡黄芩苷定量模型, 预测值与实测值的散点, 见图 6。所建立的小柴胡颗粒黄芩苷的近红外定量模型, 预测均方根误差为 1.700 4, 模型的决定系数 $R^2 = 0.908 7$ 。为进一步检验模型预测的准确性, 对 50 个外部验证集样品的预测值和实测值间进行配对 t 检验, $|t| = 0.678$ ($|t| < t_{0.05, 49}$, 由 t 分布表中查得 $t_{0.05, 49} = 2.010$), 显著性水平 $\text{sig} = 0.501$ 。说明两组数据在 $\alpha = 0.05$ 水平, 不存在显著差异。说明所建立的校正模型具有较好的预测准确性。

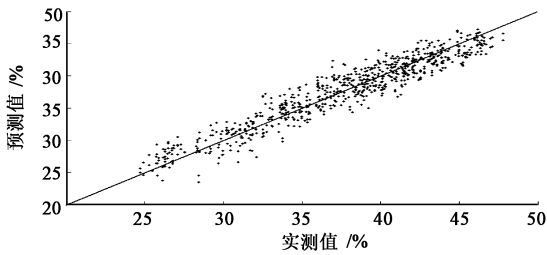


图 6 小柴胡黄芩苷 CARS-PLS 建模预测集实测值与预测值散点
Fig.6 Actual value versus calculated value of baicalin content of Xiaochaihu granules by CARS-PLS

3 讨论

波长筛选对提高近红外校正模型稳健性, 预测精度和准确性具有重要作用。本研究通过对比评价不同波长筛选方法结合蒙特卡洛采样和偏最小二乘法建立小柴胡颗粒中黄芩苷含量的校正模型, 结果显示, 模型预测准确性和精度 $\text{CARS} > \text{MC-UVE} > \text{GA} > \text{全波长变量} > \text{SPA}$, CARS 具有筛选出较优组合变量的能力, 与李洪东等^[23]得到的结论一致。CARS 算法有效简化模型, 提高模型预测的准确性

和精度, 适于小柴胡颗粒中黄芩苷含量的快速、无损检测。该分析方法为制药企业颗粒制剂有效成分的高效检测提供基础依据。

近红外光谱分析技术具有快速、无损、绿色、高效等优点, 被广泛应用于药品中有效成分的快速评价分析。该分析方法的稳定性和准确性取决于定性判别模型和定量校正模型的建立, 在实际的应用过程中, 需选择适宜的化学计量学方法有效识别异常样本、优选波长, 剔除冗余信息, 建立稳健的模型, 提高模型的稳定性和准确性。本研究将近红外光谱分析与化学计量学方法结合, 旨在构建小柴胡颗粒中黄芩苷含量的稳健校正模型, 为药品中有效成分含量的快速分析提供参考。

[致谢] 湖南师范大学工程与设计学院杨家红教授提供 MATLAB 6.0 软件支持。

[参考文献]

[1] 国家药典委员会. 中华人民共和国药典. 一部[S]. 北京: 中国医药科技出版社, 2010: 301.

[2] Akao T, Hanada M, Sakashitay Y, et al. Efflux of baicalin, a flavone glucuronide of Scutellariae Radix, on Caco-2 cells through multidrug resistance-associated protein 2 [J]. J Pharm Pharmacol, 2007, 59 (1): 87-93.

[3] Zhang Z J, Li P, Wang Z, et al. A comparative study on the individual and combined effects of baicalin and jasmionidin on focal cerebral ischemia-reperfusion injury [J]. Brain Res, 2006, 1123(1): 188-195.

[4] Tsai C C, Lin M T, Wang J J, et al. The antipyretic effects of baicalin in lipopolysaccharide- evoked fever in rabbits [J]. Neuropharmacology, 2006, 51 (4): 709-717.

[5] 刘彩红, 李玉琴, 贾宝秀, 等. 黄芩苷与人血清白蛋白的相互作用研究[J]. 分析测试学报, 2011, 30(5): 532-536.

[6] 原红霞, 王婷婷, 闫艳. HPLC 同时测定小柴胡汤乙酸乙酯部位中 5 种成分[J]. 中国实验方剂学杂志, 2015, 12(8): 64-66.

- [7] 吴冬梅, 刘卫红, 姚蓝, 等. 提取溶剂对柴黄颗粒中主要成分的影响[J]. 中国实验方剂学杂志, 2014, 20(24):71-73.
- [8] 陈莉婧, 张忠义, 张守尧. 清开灵注射液中黄芩苷含量的微乳液相色谱法测定[J]. 分析测试学报, 2009, 28(10):1217-1220.
- [9] 吴明花, 高言明, 梁光义, 等. 三黄泻心汤中黄芩苷含量的高效毛细管电泳内标法测定[J]. 时珍国医国药, 2014, 25(2):279-281.
- [10] 董喆, 黄湘鹭, 曹进, 等. 液相色谱-质谱检测 3 种不同属新鲜黄芩中 8 种成分[J]. 食品科学, 2014, 35(4):131-136.
- [11] 胡咏川, 田晓鑫, 刘蕾, 等. 近红外光谱技术鉴定中药的进展[J]. 中国中药杂志, 2012, 37(8):1066-1071.
- [12] 战皓, 柳梦婷, 方婧, 等. 近红外分析技术在中药鉴定和含量测定中的应用研究进展[J]. 中国实验方剂学杂志, 2015, 21(12):231-234.
- [13] 冯艳春, 胡昌勤. 近红外技术在我国药品流通领域的应用进展[J]. 光谱学与光谱分析, 2014, 34(5):1222-1228.
- [14] 褚小立, 陆婉珍. 近五年我国近红外光谱分析技术研究与应用进展[J]. 光谱学与光谱分析, 2014, 34(10):2595-2605.
- [15] 黄倩倩, 潘瑞乐, 魏建和, 等. 近红外漫反射光谱法测定黄芩中总黄酮及黄芩苷的含量[J]. 光谱学与光谱分析, 2009, 29(9):2425-2428.
- [16] 白雁, 张威, 王星. 近红外光谱法测定不同厂家银黄颗粒中黄芩苷含量[J]. 中国中药杂志, 2010, 35(2):166-168.
- [17] 褚小立, 袁洪福, 陆婉珍. 近红外分析中光谱预处理及波长选择方法进展与应用[J]. 化学进展, 2004, 16(4):528-542.
- [18] 石吉勇, 殷晓平, 邹小波, 等. 基于模拟退火波长优化的草莓坚实度近红外光谱检测[J]. 农业机械学报, 2010, 41(9):99-103.
- [19] 徐冰婉, 徐小雯, 黄家卫. 多目标遗传算法优化黄连的提取工艺[J]. 中成药, 2015, 37(4):768-773.
- [20] 朱华旭, 黄山, 陆文聪, 等. SVM 算法用于黄连解毒汤 3 种指标性成分配伍浓度与药效作用相关性的研究[J]. 中成药, 2011, 33(10):1709-1712.
- [21] 梁逸曾, 许青松. 复杂体系仪器分析-白、灰、黑分析体系及其多变量解析方法[M]. 北京: 化学工业出版社, 2012:259-260.
- [22] Li H D, Liang Y Z, Xu Q S, et al. Model population analysis for variable selection [J]. J Chemometr, 2010, 24(7/8):418-423.
- [23] Li H D, Liang Y Z, Cao D S, et al. Model-population analysis and its applications in chemical and biological modeling [J]. Trac-trends Anal Chem, 2012, (38):154-162.
- [24] Li H D, Liang Y Z, Xu Q S, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration [J]. Anal Chim Acta, 2009, 648(1):77-84.
- [25] 李江波, 郭志明, 黄文倩, 等. 应用 CARS 和 SPA 算法对草莓 SSC 含量 NIR 光谱预测模型中变量及样本筛选[J]. 光谱学与光谱分析, 2015, 35(2):372-378.
- [26] Cai W S, Li Y K, Shao X G. A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra [J]. Chemometr Intell Lab, 2008, 90(2):188-194.
- [27] Han Q J, Wu H L, Cai C B, et al. An ensemble of Monte Carlo uninformative variable elimination for wavelength selection [J]. Chim Acta, 2008, 612(2):121-125.
- [28] 蔡峰, 刘太昂, 张朝平, 等. 基于近红外数据预测烟草淀粉含量的 GA-SVR 模型[J]. 计算机与应用化学, 2014, 31(8):969-971.
- [29] 赵艳丽, 张霖, 袁天军, 等. 近红外光谱快速鉴别不同产地药用植物重楼的方法研究[J]. 光谱学与光谱分析, 2014, 34(7):1831-1835.
- [30] Cao D S, Liang Y Z, Xu Q S, et al. A new strategy of outlier detection for QSAR/QSPR [J]. J Comput Chem, 2010, 31(3):592-602.
- [31] 李洪东. 广义灰色分析体系建模的基本问题及其模型集群分析研究[D]. 长沙: 中南大学, 2012.

[责任编辑 顾雪竹]