

基于拉曼光谱实时监测甘草配方颗粒的提取过程

殷文俊¹, 唐建飞², 郑洁¹, 张璐¹, 张弘旭¹, 刘霄谊¹, 颜继忠^{1*}, 张慧^{1*}

1. 浙江工业大学药学院 中药研究所, 浙江 杭州 310014

2. 杭州朱养心药业有限公司, 浙江 杭州 310011

摘要: **目的** 应用拉曼光谱技术结合多种预处理算法和多种特征波段筛选方法建立数学模型, 对甘草配方颗粒提取过程中甘草苷和甘草酸含量实时监测。**方法** 以甘草配方颗粒为研究对象, 收集提取过程中各个时间点的提取液样本, 进行拉曼光谱检测。采集得到的光谱与液相色谱结果对应, 分别建立3种甘草苷和甘草酸定量校正模型, 考察不同预处理方法对模型的影响, 优选出最佳变量筛选方法。**结果** 甘草苷模型中标准正态变换(SNV)预处理方法和甘草酸模型中Savitzky-Golay 13点平滑方法对模型性能参数提升幅度最大。通过连续投影算法筛选的甘草苷和甘草酸模型分别只需要4个和3个光谱变量即可达到全光谱变量模型水平, 通过竞争性自适应重加权算法(CARS)筛选后建立的贝叶斯岭回归(BRR)甘草苷和甘草酸定量模型具有全局最优性能。**结论** 拉曼光谱技术应用于甘草配方颗粒提取过程中所建立模型性能良好, 为实现中药配方颗粒提取过程实时监测和快速分析提供了研究基础。

关键词: 拉曼光谱; 甘草配方颗粒; 提取过程; 甘草苷; 甘草酸; 实时监测

中图分类号: R283.6 文献标志码: A 文章编号: 0253-2670(2021)18-5560-09

DOI: 10.7501/j.issn.0253-2670.2021.18.010

Real-time monitoring of extraction process of *Glycyrrhizae Radix et Rhizoma* formula granule based on Raman spectroscopy technology

YIN Wen-jun¹, TANG Jian-fei², ZHENG Jie¹, ZHANG Lu¹, ZHANG Hong-xu¹, LIU Xiao-yi¹, YAN Ji-zhong¹, ZHANG Hui¹

1. Institute of Traditional Chinese Medicine Research, College of Pharmaceutical Sciences, Zhejiang University of Technology, Hangzhou 310014, China

2. Hangzhou Zhu Yang Xin Pharmaceutical Co., Ltd., Hangzhou 310011, China

Abstract: Objective To establish the mathematical models for a real-time Raman spectral analytical method combined with a variety of pretreatment algorithms and a variety of feature bands to monitor the content of liquiritin and glycyrrhizic acid during the extraction of Gancao (*Glycyrrhizae Radix et Rhizoma*, GRR) formula granules (GRRFG). **Methods** Taking GRRFG as the research object, different samples of extracts at various time points were collected in the extraction process and detected by Raman spectroscopy. The results of HPLC were used as a reference to build the quantitative calibration models for liquiritin and glycyrrhizic acid, respectively. A variety of spectral pre-processing methods and variable selection methods were applied to the establishment of models. **Results** The results showed that the standard normal variate transformation (SNV) pre-processing method in the liquiritin model and the Savitzky-Golay 13 point smoothing method in the glycyrrhizic acid model were suitable to improve the performance parameters of the mathematical models. Successive projection algorithm with only four bands and three bands were almost the same as the results of raw spectrums of liquiritin and glycyrrhizic acid. The bayesian ridge regression (BRR) model of these two compounds processed by the competitive adaptive reweighted sampling (CARS) showed the global optimal performance. **Conclusion** The Raman spectroscopy technology is applied to the GRRFG extraction process, and the established mathematical models have good performance, which provide a research foundation for realizing the real-time monitoring and rapid analysis of the

收稿日期: 2021-04-21

基金项目: 浙江省自然科学基金项目(LY20H280014); 中国学位与研究生教育学会研究课题(2019YX09)

作者简介: 殷文俊(1996—), 男, 硕士研究生, 主要从事药物质量评价方向的研究。E-mail: 913062306@qq.com

*通信作者: 张慧, 副教授, 硕士生导师, 主要从事中药质量控制与药物分析研究。E-mail: zh889@zjut.edu.cn

颜继忠, 教授, 博士生导师, 主要从事中药分离工程与中药质量控制研究。E-mail: yjz@zju.edu.cn

extraction process of traditional Chinese medicine formula granule.

Key words: Raman spectroscopy; *Glycyrrhizae Radix et Rhizoma* formula granule; extraction process; liquiritin; glycyrrhizic acid; real-time monitoring

中药配方颗粒是以优质的中药饮片为原料,采用先进工艺,经过提取、浓缩、干燥、制粒等工序精制而成的供医生临床配方使用的单味中药浓缩颗粒^[1]。提取过程是中药配方颗粒生产的首要环节,提取液的质量直接影响后续生产环节和最终产品质量的均一性和稳定性^[2]。目前,提取工艺操控主要依靠经验,精确度难以保证。传统的质量分析方法费时费力且滞后于生产过程,无法及时反馈生产现状。拉曼光谱是基于拉曼散射效应,对与入射光频率不同的散射光谱进行分析以得到分子振动、转动方面信息^[3]。拉曼光谱集光谱分析技术、化学计量学方法和计算机技术为一体,是一种极具发展前景的高新技术^[4],分析方法具备快速、经济和无损等特点^[5]。与近红外光谱相比,由于水引起的拉曼信号较弱,因此拉曼光谱更适合于中药配方颗粒提取过程中的分析^[6]。

本研究以甘草配方颗粒提取过程为例,依据《中国药典》2020年版中将甘草中甘草酸和甘草苷的含量作为检测指标,针对甘草提取过程中甘草苷和甘草酸含量的变化,利用拉曼光谱检测速度快的特点,结合液相参考分析,通过对预处理方法和特征提取方法的优选,分别建立甘草苷和甘草酸定量校正模型,实现对提取过程中的实时监测。

1 仪器与材料

1.1 仪器

Agilent 1290型超高效液相色谱仪,美国Agilent公司;CPA225D型1/10万分析天平,德国Sartorius公司;KQ-250DE型数控超声波清洗器,巩义市予华仪器有限公司;i-Raman Plus BWS 465-785S型便携式拉曼光谱仪,美国B&W Tek公司。

1.2 材料

甘草饮片建模集5个批次(GC1912029、GC1912030、GC1912032、GC1912034、GC1912036)和外部验证集3个批次(GC1912038、GC1912039、GC1912040)均在安徽亳州药材市场购买,产地为新疆。经浙江工业大学药学院张慧副教授鉴定为豆科植物甘草 *Glycyrrhiza uralensis* Fisch.的干燥根和根茎。对照品甘草苷(批号111610-201607,质量分数为93.1%)、甘草酸(批号110731-201619,质

量分数为93.0%),中国食品药品检定研究院;无水乙醇,分析纯,上海凌峰化学试剂有限公司;磷酸,分析纯,上海凌峰化学试剂有限公司;色谱甲醇(批号20085008)、色谱乙腈(批号20085249),美国TEDIA公司;纯净水,杭州娃哈哈集团有限公司。

2 方法与结果

2.1 甘草苷和甘草酸 HPLC 含量测定

2.1.1 色谱条件 色谱柱为Agilent Zorbax SB C₁₈柱(250 mm×4.6 mm, 5.0 μm);柱温25℃;流动相为乙腈-0.05%磷酸水溶液,梯度洗脱:0~8 min, 19%乙腈;8~35 min, 19%~50%乙腈;35~36 min, 50%~100%乙腈;36~40 min, 100%~19%乙腈;体积流量1.0 mL/min;检测波长237 nm;进样量10 μL。色谱图见图1。

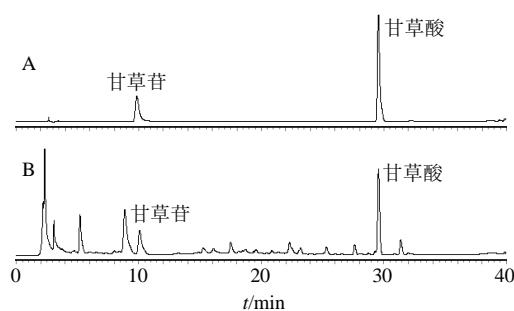


图1 混合对照品(A)和甘草提取液(B)的HPLC图
Fig. 1 HPLC of mixed reference substances (A) and extract of *Glycyrrhizae Radix et Rhizoma* (B)

2.1.2 供试品溶液的制备 精密移取甘草提取液1 mL,以13 000 r/min转速离心5 min,上清液过0.22 μm微孔滤膜,取续滤液,即得。

2.1.3 对照品溶液的制备 分别取经五氧化二磷干燥器中干燥24 h的甘草苷、甘草酸铵对照品各10 mg,精密称定,加70%乙醇分别制成含甘草苷20 μg/mL、甘草酸铵0.2 mg/mL的溶液,过0.22 μm微孔滤膜,即得对照品溶液(甘草酸质量浓度=甘草酸铵质量浓度/1.0207)。

2.1.4 线性关系考察 根据“2.1.1”项色谱条件,分别进样7个不同质量浓度的甘草苷和甘草酸混合对照品溶液,分别含甘草苷398.0、159.2、79.6、39.8、19.9、10.0、5.0 μg/mL和甘草酸3 922.8、1 569.1、784.6、392.3、196.1、98.1、49.0 μg/mL,按质量浓

度从低到高分分别重复进样3次,以峰面积值为纵坐标(y),甘草苷及甘草酸质量浓度为横坐标(x),计算线性回归方程及范围。结果显示甘草苷在5.0~398.0 $\mu\text{g/mL}$ 线性关系良好,回归方程为 $y=17\,747x-12.268$, $R^2=0.999\,9$;甘草酸在49.0~3\,922.8 $\mu\text{g/mL}$ 线性关系良好,回归方程 $y=5\,011.8x-84.883$, $R^2=0.999\,9$ 。

2.1.5 精密度考察 取混合对照品溶液1 d内连续进样6次,计算甘草苷和甘草酸色谱峰保留时间和峰面积的RSD来考察日内精密度。甘草苷和甘草酸保留时间和峰面积的RSD分别为0.72%和0.31%、0.20%和0.36%。连续3 d分别进样当天配制的供试品溶液,每天连续进样3针,分别计算甘草苷和甘草酸色谱峰保留时间和峰面积的RSD值来考察日间精密度。甘草提取液中甘草苷和甘草酸保留时间和峰面积的RSD值分别为2.11%和0.61%、0.16%和0.26%。

2.1.6 重复性考察 平行制备6份供试品溶液,连续进样分析,计算甘草苷和甘草酸的RSD值来考察重复性。甘草苷和甘草酸质量分数RSD值为0.64%和0.25%。该方法的精密度及重复性均良好。

2.1.7 稳定性考察 取同一份供试品溶液,分别在供试品溶液制备后0、2、4、8、12、24 h进样分析,通过计算甘草苷和甘草酸峰面积的RSD值来考察供试品溶液的稳定性。甘草苷和甘草酸峰面积的RSD为0.32%和0.11%,表明供试品溶液在24 h内稳定性良好。

2.1.8 加样回收率考察 精密移取已测定甘草苷和甘草酸质量浓度提取液2.5 mL,分别精密加入质量浓度为15.9 $\mu\text{g/mL}$ 的甘草苷对照品溶液0.85 mL,质量浓度为100.1 $\mu\text{g/mL}$ 的甘草酸铵对照品溶液1.15 mL定容至5 mL,平行配制6份,摇匀,取1 mL以13\,000 r/min离心5 min,过0.22 μm 微孔有机滤膜,即得供试品溶液。测定供试品溶液中甘草苷和甘草酸色谱峰面积,计算供试品溶液中甘草苷和甘草酸的质量浓度,再计算加样回收率。甘草苷的平均加样回收率为100.88%,RSD为3.09%;甘草酸的平均加样回收率为100.15%,RSD为2.67%。表明该方法可用于甘草水提取液中甘草苷和甘草酸的含量测定。

2.2 拉曼光谱采集

提取过程中在中试提取装置中进行,模拟甘草配方颗粒生产的提取过程。依据《中药饮片标准汤剂》

和《中药配方颗粒质量控制与标准制定技术要求》确定甘草配方颗粒的生产工艺参数:添加甘草饮片1 kg到提取罐中,加入7倍水浸泡30 min,加热提取90 min,收集提取液,并向提取罐中加入6倍水,进行第2次提取,持续30 min。

在浸泡完成后第1次收集样品,之后在提取过程中每10分钟收集1次样品,最后对2次提取液的混合液收集1次样品。收集的样品用于拉曼光谱测量和液相色谱分析。从每批可以收集得到15个样品,最后从5个不同批次(GC1912029、GC1912030、GC1912032、GC1912034、GC1912036)中共收集75个样品。

拉曼光谱可通过探头实现实时采集。光源的激发波长为785 nm,激光强度为100%(激光功率为300 mW),积分时间为10 s,采集3次取平均值,总采集时间30 s。拉曼位移的光谱范围为201.33~3100 cm^{-1} (1701个光谱变量),分辨率为1.71 cm^{-1} 。

2.3 拉曼光谱定量模型的建立

2.3.1 样本划分 通过Kennard-Stone算法以7:3的比例划分校正集与预测集。其中校正集52个样本,预测集23个样本。样本集划分结果见表1,可以看出,预测集的质量浓度包含在校正集的范围中,校正集和预测集的标准差与总样本之间无明显差异($P<0.05$)。

表1 校正集和预测集的样本划分结果

Table 1 Division results of calibration set and prediction set

指标成分	样品组	样品数	质量浓度范围/ ($\mu\text{g}\cdot\text{mL}^{-1}$)	$\bar{x} \pm s$ / ($\mu\text{g}\cdot\text{mL}^{-1}$)
甘草苷	校正集	52	18.3~350.3	174.8 \pm 94.2
	预测集	23	22.3~349.4	181.6 \pm 97.3
	总样本	75	18.3~350.3	176.9 \pm 94.6
甘草酸	校正集	52	113.9~1\,074.7	643.4 \pm 290.9
	预测集	23	138.6~1\,052.1	639.2 \pm 287.4
	总样本	75	113.9~1\,074.7	642.1 \pm 287.9

2.3.2 光谱的预处理 中药提取液成分复杂,分析物的光谱信息会出现相互重叠和掩盖现象,光谱采集过程中存在因仪器或环境而引入的噪声,且存在荧光背景和基线漂移现象^[7]。需要使用光谱预处理方法,提取有用信息,去除仪器噪声,减少荧光背景干扰^[8]。本研究中通过比较经Savitzky-Golay(SG)平滑、一阶导数、二阶导数、标准正态变换(standard normal variate transformation, SNV)、多元散射校正

(multiplicative scatter correction, MSC)等预处理的拉曼光谱数据和原始光谱数据的偏最小二乘回归(partial least squares regression, PLSR)定量模型参数,筛选出最适合甘草苷和甘草酸定量模型的预处理方法。5个不同批次提取液拉曼光谱原始光谱如图2-a所示,随着提取时间的增加,提取液拉曼光谱存在基线漂移现象。这可能是提取过程中提取液中指标成分含量的持续累积,使得荧光背景的增加所致^[9]。

光谱经SG平滑、一阶导数、二阶导数、MSC、SNV预处理的拉曼光谱数据和对应的甘草苷和甘草酸液相数据作为输入变量,得到不同预处理方法的PLSR定量模型与原始光谱比较结果见表2。从表2中可以看出,在提取液中甘草苷PLSR定量校

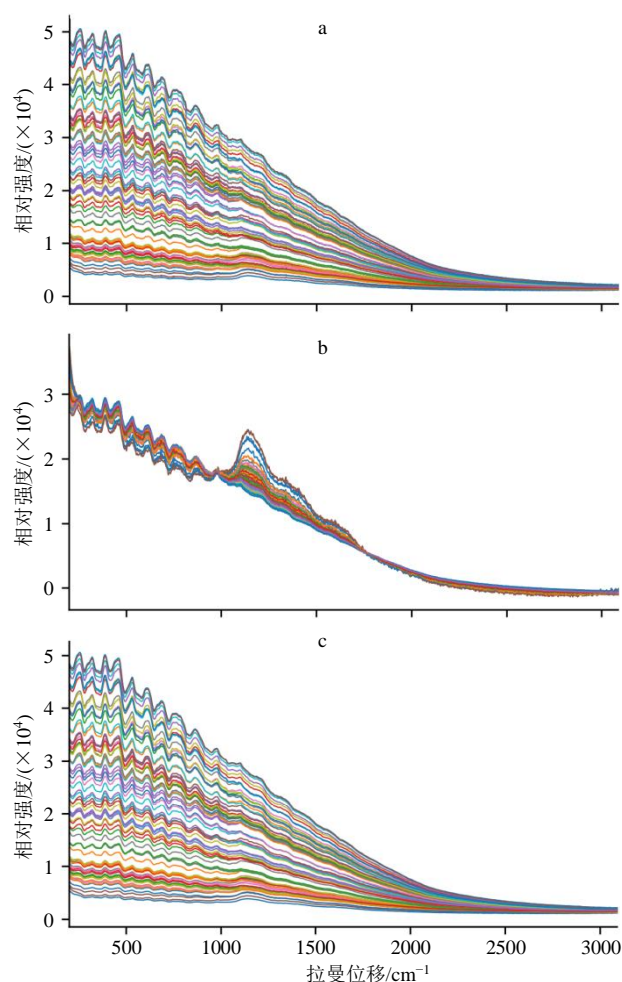


图2 甘草提取过程拉曼光谱的原始光谱(a)、SNV光谱(b)、SG13点平滑光谱(c)

Fig. 2 Raw spectra (a), spectra preprocessed with SNV (b), and spectra preprocessed with SG (window size 13) (c) of Raman spectra of extraction process of *Glycyrrhizae Radix et Rhizoma*

正模型中,SG-13点平滑、SG-17点平滑、一阶导数和SNV预处理后,定量模型均较原始光谱模型性能有所提升,其中SNV预处理方法,对模型性能参数提升幅度最大,光谱图如图2-b所示,预测集决定系数为0.8616,误差均方根为0.0338,选用作甘草苷定量校正模型的预处理方法;在提取液中甘草酸PLSR定量校正模型中,除MSC方法会对模型造成负面影响以外,其他预处理方法均会对模型性能有一定程度的提升,其中SG-13点平滑预处理后,光谱图如图2-c所示,对提取液中甘草酸含量预测效果最好,将其作为甘草酸定量校正模型的预处理方法。

2.3.3 光谱特征波段的筛选 原始光谱经过预处理后,一些相关因素带来的干扰信息能够得到消除,然而光谱中依然存在大量与指标成分的含量关联性较差的信息^[10],数据的冗杂不仅会造成模型运算时间增加,还会影响模型的预测结果。因此在构建定量模型之前,采用适当的方法进行特征波长的筛选,降低数据维度以实现模型快速、精准定量^[11]。从特征波段的筛选和特征波长点的筛选2个方面入手:针对特征波长点的筛选主要采用竞争性自适应重加权算法(competitive adaptive reweighted sampling, CARS)、连续投影算法(successive projection algorithm, SPA)、蒙特卡罗无信息变量消除(monte carlo elimination of uninformative variables, MC-UVE);针对特征波段的筛选主要采用联合区间偏最小二乘法(synergy interval partial least squares, siPLS)。

(1) CARS法: CARS法以迭代和竞争的方式进行蒙特卡罗采样,选择固定比例的变量子集,建立PLSR定量校正模型,采用指数递减函数和自适应加权采样技术选择较大权重变量。最后计算RMSECV值以评估子集选择出最关键变量^[12]。在甘草酸校正模型中设置主成分个数为7,算法进化次数为50次。CARS运行结果如图3所示,图3-a为筛选过程中变量个数的变化趋势,变量个数随采样次数的增加而减少;图3-b为交叉验证误差均方根(root mean square error of cross-validation, RMSECV)的变化趋势,29次采样之前随着运行次数的增加而减少,29次采样之后随着运行次数的增加而升高,表明可能剔除了与水提取液中甘草酸含量相关的变量;图3-c为回归系数的变化路径,图中“*”表示最小RMSECV所对应的采样次数,图中各线表示不同采样次数回归系数变化路径。根据交叉验证均

表 2 不同预处理方法对 PLSR 定量模型性能的影响

Table 2 Effect of different pretreatment methods on performance of PLSR models

预处理方法	甘草苷				甘草酸			
	校正集		预测集		校正集		预测集	
	R^2_c	RMSEC	R^2_p	RMSEP	R^2_c	RMSEC	R^2_p	RMSEP
原始光谱	0.908 3	0.028 0	0.857 6	0.034 6	0.900 5	0.089 5	0.878 6	0.099 3
SG (9 点)	0.896 5	0.029 8	0.844 4	0.036 0	0.999 8	0.003 7	0.943 4	0.066 9
SG (13 点)	0.905 2	0.028 5	0.859 9	0.034 6	0.978 0	0.042 0	0.954 2	0.060 6
SG (17 点)	0.904 1	0.028 7	0.858 9	0.034 7	0.987 9	0.031 1	0.948 9	0.063 9
一阶导数	0.894 4	0.030 1	0.861 0	0.033 8	0.992 6	0.024 2	0.926 3	0.077 1
二阶导数	0.881 6	0.031 9	0.822 1	0.038 9	0.997 4	0.014 4	0.899 7	0.089 8
MSC	0.859 9	0.034 7	0.827 9	0.038 9	0.933 2	0.073 0	0.862 3	0.105 5
SNV	0.873 9	0.032 9	0.861 6	0.033 8	0.976 2	0.043 6	0.909 1	0.084 3

R^2_c -校正集决定系数 RMSEC-校正集误差均方根 R^2_p -预测集决定系数 RMSEP-预测集误差均方根

R^2_c -coefficient of determination of calibration RMSEC-root mean square error of calibration R^2_p -coefficient of determination of prediction RMSEP-root mean square error of prediction

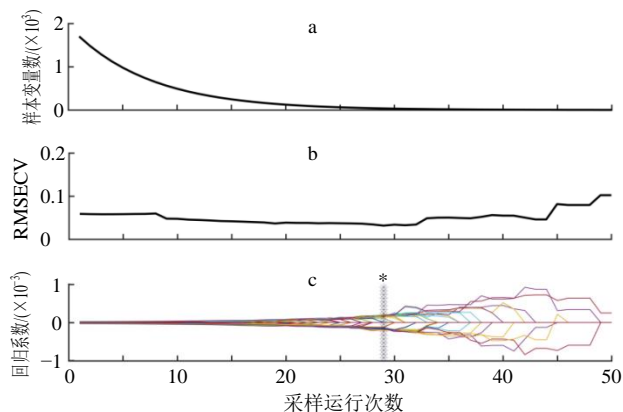


图 3 甘草酸定量校正模型 CARS 运行结果

Fig. 3 Results of CARS on quantitative calibration model of glycyrrhizic acid

方差值最小原则, 由图 3-a 可知特征波段变量个数为 36。在甘草苷校正模型中主成分个数为 11, 26 次采样后得到最小 RMSECV, 对应的最优波长变量为 54。

(2) SPA 法: SPA 法通过矢量空间共线性最小化筛选出最小的冗余信息变量组, 最大程度地减少建模变量的数量从而建立简化高效的模型^[13]。在甘草酸校正模型中如图 4-a 所示, 设定选定波长数在 1~25, 当模型的变量数为 3 时, RMSECV 最小, 表明此时筛选得到最高效模型, 得到的特征波段在全波段中的分布见图 4-b。在甘草苷校正模型中选定波长数在 1~25, 当模型的变量数为 4 时得到最优模型。

(3) MC-UVE 法: MC-UVE 算法通过实施蒙特

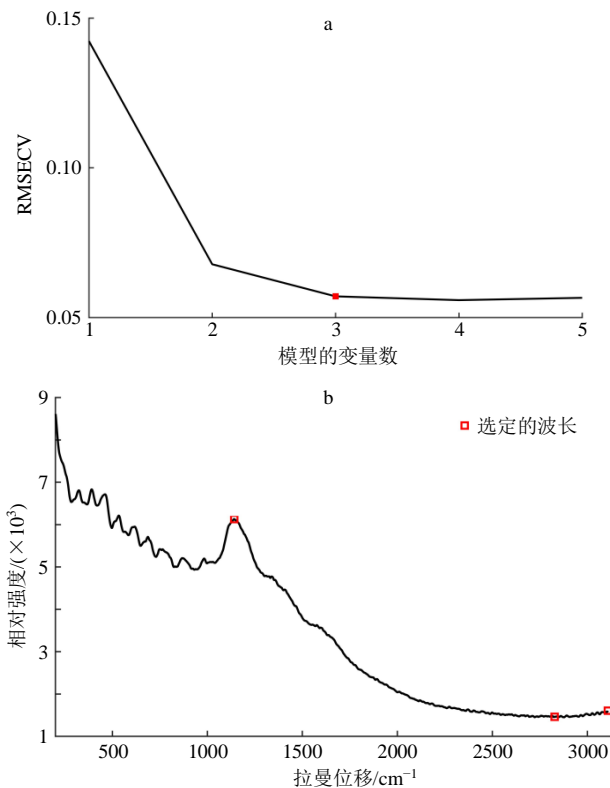


图 4 甘草酸定量校正模型 SPA 运行结果

Fig. 4 Results of SPA on quantitative calibration model of glycyrrhizic acid

卡罗随机采样, 对原无信息变量消除算法进行改进。其使用校正集的随机样本对大量模型进行校准, 通过模型相应系数来评估每个变量的可靠程度, 可靠性性低于临界值的变量被识别为无信息变量进行剔除^[14]。在甘草酸校正模型中设置 MC-UVE 运行次

数为 1000 次, 主成分数 15, 以变量可靠度指数 (reliability index, RI) 作为筛选依据, 将 RI 值按降序排列, 运行结果如图 5 所示, 确定最佳变量数为 50。在甘草苷校正模型中 MC-UVE 运行次数为 500 次, 主成分数 15, 最佳变量数为 18。

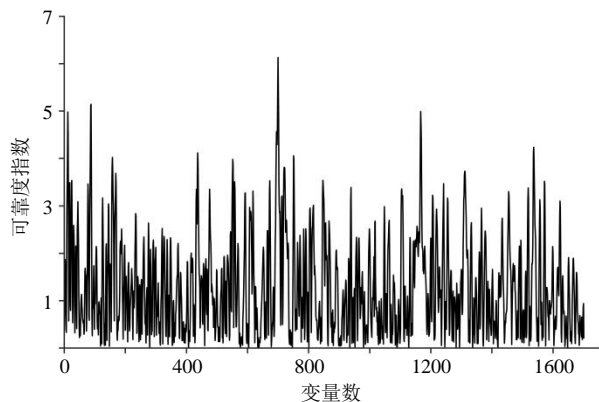


图 5 甘草酸定量校正模型 MC-UVE 运行结果

Fig. 5 Results of MC-UVE on quantitative calibration model of glycyrrhizic acid

(4) siPLS 法: siPLS 算法是将整个频谱平均分为 n 个间隔, 并对子间隔组合计算所有可能的 PLS 模型。选择具有最低 RMSECV 值的子间隔组合作为最佳变量^[15]。在甘草酸校正模型中通过 siPLS 如图 6-a 所示, 将全谱分为 10 个子区间, 联合其中 4 个子区间 [1, 4, 5, 8], 当主成分数为 12 时模型性能最佳, RMSECV 为 0.041 3, 此时选择波长数为 681。在甘草苷校正模型中通过 siPLS 算法, 将全谱分为 13 个子区间, 联合其中 4 个子区间 [1, 2, 5, 7], 当主成分数为 4 时模型性能最佳, RMSECV 为 0.029 7, 此时选择波长数为 524。

2.3.4 定量模型的建立 采用 PLSR、多层感知机 (multi-layer perceptron, MLP) 和贝叶斯岭回归 (bayesian ridge regression, BRR) 分别建立甘草水提取液样品的甘草苷和甘草酸拉曼光谱定量模型。定量校正模型验证指标包括: 交叉验证集决定系数 (coefficient of determination of cross-validation, R^2_{cv}) 和 RMSECV 作为模型优化的参考指标, R^2_c 、RMSEC、 R^2_p 和 RMSEP 作为模型性能的评价指标。其中 R^2_p 越接近 1, RMSEP 越小, 表明所对应模型性能更好。

采用多种特征波段筛选方法所建立的 PLSR、MLP 和 BRR 的甘草苷和甘草酸定量校正模型性能参数如表 3 所示。从表 3 中可以看出, 通过特征波段的筛选能够在确保模型性能的基础上, 大幅减少

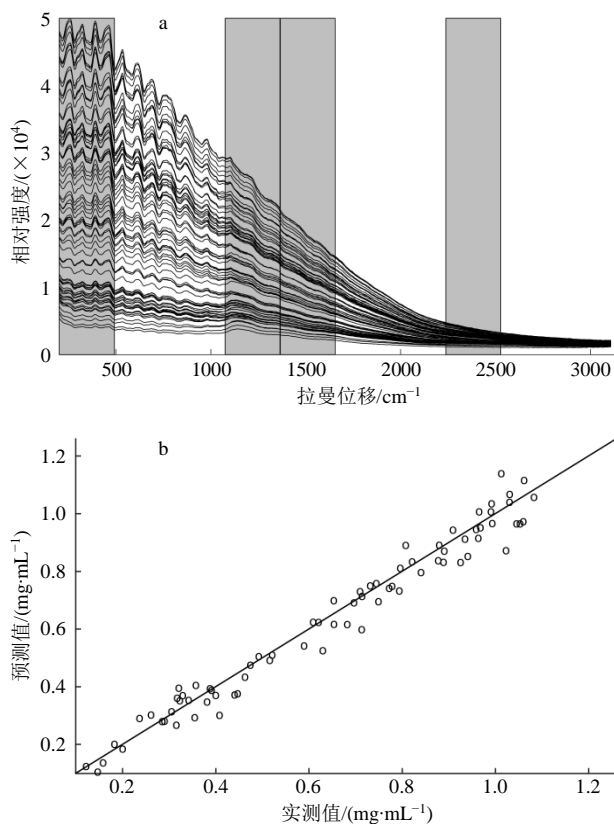


图 6 甘草酸定量校正模型 siPLS 运行结果

Fig. 6 Results of siPLS on quantitative calibration model of glycyrrhizic acid

对拉曼光谱变量的使用, 筛选变量后基本可以达到甚至超过全光谱变量所构建模型的性能。这表明利用特征波段提取方法能够排除冗余信息, 找到与指标成分密切相关的光谱变量, 从而大幅减少计算量, 显著提升模型运行速度。其中, 通过 SPA 筛选的甘草苷和甘草酸模型分别只需要 4 个和 3 个光谱变量即可达到全光谱变量模型水平, 通过 CARS 筛选建立的甘草苷和甘草酸模型均为各模型中最佳结果。3 种不同定量模型之间比较发现, 除 MLP 模型外, PLSR 和 BRR 建立的模型均能取得不错的效果, 其中 BRR 在不同筛选条件下能够得到最优结果。

CARS 筛选后建立的 BRR 甘草苷和甘草酸定量模型具有全局最优性能, 如图 7-a 和图 7-b 分别为该模型甘草苷和甘草酸含量预测值与参考真实值的散点图, 甘草苷和甘草酸含量液相参考值与拉曼光谱的预测值相关性较高, 甘草苷和甘草酸模型 R^2_p 分别为 0.987 6 和 0.990 2, RMSEP 分别为 0.010 4 和 0.027 6, 所对应的模型性能较高。

2.4 甘草苷和甘草酸含量实时监测

将已训练校正好的甘草苷 CARS-BRR 模型和

表 3 不同变量筛选方法建模结果比较

Table 3 Comparison on modeling results under different variables selection methods

特征波段	模型	甘草苷					甘草酸				
		波数/ cm ⁻¹	校正集		预测集		波数/ cm ⁻¹	校正集		预测集	
			R ² _c	RMSEC	R ² _p	RMSEP		R ² _c	RMSEC	R ² _p	RMSEP
FULL	PLSR	1701	0.972 0	0.015 5	0.861 6	0.034 8	1701	0.900 5	0.089 5	0.878 6	0.099 3
	MLP		0.989 3	0.007 4	0.857 0	0.035 3		0.936 4	0.071 5	0.861 7	0.105 6
	BRR		0.893 3	0.030 1	0.875 7	0.032 9		0.999 9	0.000 3	0.966 7	0.051 6
SPA	PLSR	4	0.873 9	0.032 9	0.863 1	0.034 5	3	0.964 2	0.053 6	0.960 2	0.056 6
	MLP		0.889 6	0.030 8	0.860 8	0.034 9		0.901 1	0.089 2	0.879 8	0.098 9
	BRR		0.889 6	0.030 8	0.864 2	0.034 4		0.964 2	0.053 6	0.959 9	0.056 9
CARS	PLSR	54	0.999 8	0.001 3	0.978 0	0.013 6	36	0.993 8	0.022 3	0.988 0	0.030 7
	MLP		0.999 9	0.000 9	0.589 4	0.049 4		0.945 4	0.065 7	0.886 9	0.092 3
	BRR		0.920 0	0.025 7	0.987 6	0.010 4		0.995 9	0.018 1	0.990 2	0.027 6
MC-UVE	PLSR	18	0.934 4	0.023 7	0.879 5	0.031 9	50	0.983 1	0.036 7	0.940 8	0.068 4
	MLP		0.931 3	0.024 2	0.870 5	0.033 5		0.934 7	0.070 0	0.899 4	0.086 7
	BRR		0.897 0	0.029 7	0.8951	0.030 0		0.978 4	0.041 5	0.951 1	0.062 6
siPLS	PLSR	524	0.991 2	0.008 6	0.847 4	0.036 5	681	0.996 1	0.017 6	0.947 8	0.063 9
	MLP		0.974 8	0.014 5	0.873 9	0.033 2		0.884 8	0.096 0	0.854 7	0.107 2
	BRR		0.896 5	0.029 8	0.857 4	0.035 1		0.999 9	0.000 9	0.970 5	0.048 2

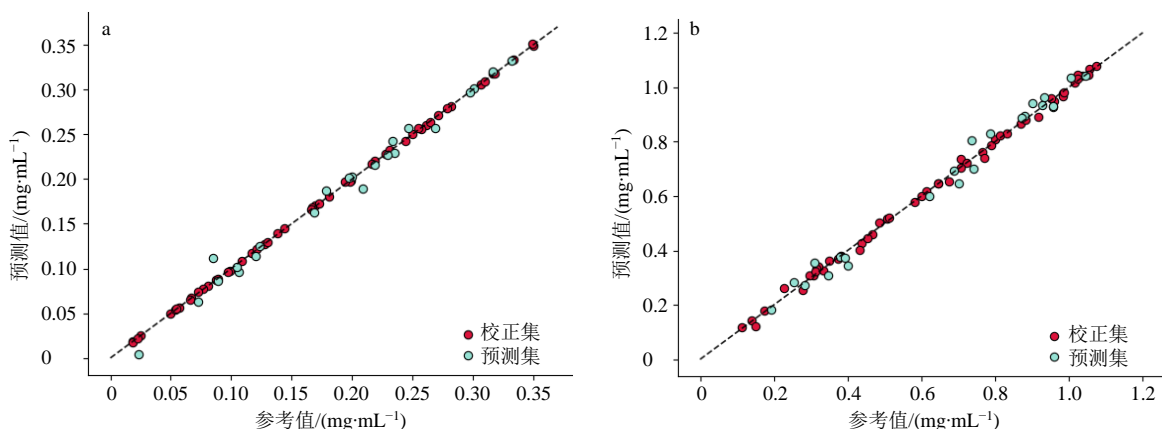


图 7 甘草苷 (a) 和甘草酸 (b) 的 BRR 模型预测值与参考值的相关图

Fig. 7 Correlation between predicted value and reference value of BRR quantitative models of liquiritin (a) and glycyrrhizic acid (b)

甘草酸 CARS-BRR 模型分别输出, 利用便携式拉曼光谱仪按照“2.2”项中方法对新的 3 批次 (GC1912038、GC1912039、GC1912040) 甘草提取过程进行实时监测, 每批次甘草可采集得到 15 个样本, 3 批次共可采集得到 45 个样本。图 8 为甘草提取过程中甘草苷 CARS-BRR 模型拉曼光谱预测值与实测值的趋势对比图, 图 9 为甘草酸 CARS-BRR 模型拉曼光谱预测值与实测值的趋势对比图。

由图 8、9 可知, 在 3 批甘草提取实时监测过程

中, 拉曼光谱甘草苷和甘草酸预测值与实测值结果能够很好地重叠在一起, 甘草苷模型 R²_p 为 0.891 0, 甘草酸模型 R²_p 为 0.969 2。上述结果显示, 所建立的模型能够很好地运用到实时监测甘草提取过程中甘草苷和甘草酸含量。

2.5 数据处理

本研究使用的光谱采集处理工具为美国 B&W Tek 公司 BWSpec4 软件, 采用 Spyder (Python 3.7) 进行光谱预处理和定量模型的构建等数据处理分析

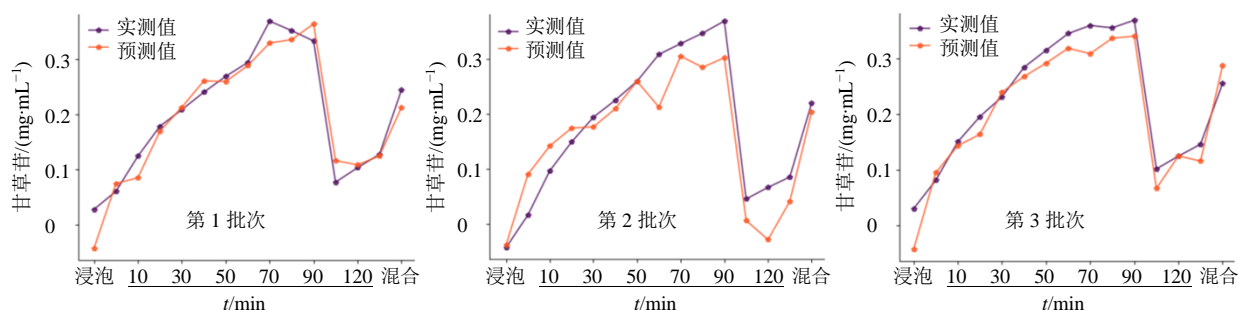


图8 甘草苷 CARS-BRR 模型拉曼光谱预测值与实测值的趋势对比图

Fig. 8 Comparison of Raman spectra prediction values with measured values of CARS-BRR quantitative model of liquiritin

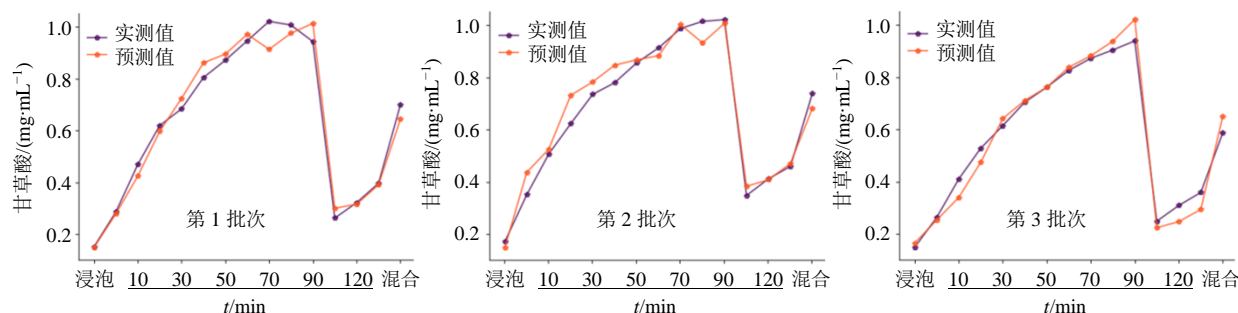


图9 甘草酸 CARS-BRR 模型拉曼光谱预测值与实测值的趋势对比图

Fig. 9 Comparison of Raman spectra prediction values with measured values of CARS-BRR quantitative model of glycyrrhizic acid

操作, 采用 Matlab 2017a 软件进行特征波段的筛选操作。

3 讨论

拉曼光谱应用在中药生产过程的质量控制中具有制样简易、准备迅速、非破坏和绿色环保等优点, 可利于稳定和提高中药产品质量, 从而促进中药生产的自动化进程, 推动中药生产迈进智能化时代^[16-17]。本研究针对甘草配方颗粒提取过程中甘草苷和甘草酸含量变化, 利用便携式拉曼光谱仪测样方便、检测时间短等特点^[18-19], 通过建立甘草苷和甘草酸定量校正模型, 实现对提取过程中实时监测。通过比较 5 种预处理方法和 4 种变量筛选方法, 分别建立了提取过程中甘草苷和甘草酸含量预测模型。各项评价指标表明, 所建立的甘草苷和甘草酸定量模型可以剔除干扰样本, 有效地筛选出相关变量, 预测效果较好, 应用于实际甘草配方颗粒提取生产过程中, 能准确监控提取过程甘草苷和甘草酸含量变化规律, 指导提取终点判定, 为工业化实时检测方法的开发提供技术保障。

在后续研究中, 可通过不断扩充校正集样本数量, 优化拉曼光谱采集过程和模型构建过程, 进一步提高模型的预测性能。同时工业上应用拉曼光谱

过程中^[20], 荧光现象对拉曼光谱的背景干扰和基线漂移现象仍值得深入的研究。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 徐玉玲, 雷燕莉, 曾立, 等. 中药配方颗粒品种统一标准的有关问题探讨 [J]. 中草药, 2020, 51(20): 5389-5394.
- [2] 杨越. 基于近红外光谱技术的中药生产过程质量控制方法研究 [D]. 杭州: 浙江大学, 2018.
- [3] Zhang C, Springall J S, Wang X Y, *et al.* Rapid, quantitative determination of aggregation and particle formation for antibody drug conjugate therapeutics with label-free Raman spectroscopy [J]. *Anal Chim Acta*, 2019, 1081: 138-145.
- [4] 陈秋媚, 张琼, 黄维, 等. 中药复方 S 洗剂不同保存环境下的激光拉曼光谱分析 [J]. 时珍国医国药, 2021, 32(1): 82-86.
- [5] Jin L, Wang S F, Cheng Y Y. A Raman spectroscopy analysis method for rapidly determining saccharides and its application to monitoring the extraction process of Wenxin granule manufacturing procedure [J]. *Spectrochim Acta A*, 2020, 241: 118603.
- [6] Zhang H, Wang J, Chen Y, *et al.* Establishing the chromatographic fingerprint of traditional Chinese

- medicine standard decoction based on quality by design approach: A case study of licorice [J]. *J Sep Sci*, 2019, 42(6): 1144-1154.
- [7] Yan X, Zhang S, Fu H, *et al.* Combining convolutional neural networks and on-line Raman spectroscopy for monitoring the *Cornu Caprae Hircus* hydrolysis process [J]. *Spectrochim Acta A*, 2020, 226: 117589.
- [8] Yan X, Li W L, Zhang X L, *et al.* Development of an on-line Raman spectral analytical method for monitoring and endpoint determination of the *Cornu Caprae Hircus* hydrolysis process [J]. *J Pharm Pharmacol*, 2020, 72(01): 132-148.
- [9] 万娜, 林环玉, 伍振峰, 等. 基于中药粒子设计结合近红外光谱技术研究清润丸原料粉末的混合均匀性 [J]. *中草药*, 2020, 51(17): 4425-4432.
- [10] 张永超, 徐芳芳, 张欣, 等. 腰痹通胶囊 4 种中间体粒径的近红外光谱通用定量模型研究 [J]. *中草药*, 2021, 52(01): 55-64.
- [11] 王承克, 张泽翔, 黄晓玮, 等. 高光谱成像的豆腐形成过程中组分含量变化检测 [J]. *光谱学与光谱分析*, 2020, 40(11): 3549-3555.
- [12] Liu R, Sun Q, Hu T, *et al.* Multi-parameters monitoring during traditional Chinese medicine concentration process with near infrared spectroscopy and chemometrics [J]. *Spectrochim Acta A*, 2018, 192: 75-81.
- [13] 刘雪松, 张丝雨, 赵曼茜, 等. 近红外光谱结合不同变量筛选方法用于黄芩提取过程中黄芩苷含量预测 [J]. *药学报*, 2019, 54(1): 138-143.
- [14] 赵艳丽, 张霁, 王元忠. MC - UVE 波长选择法在近红外光谱鉴别茯苓产地中的应用研究 [J]. *菌物学报*, 2017, 36(1): 112-125.
- [15] 王晴, 徐芳芳, 张欣, 等. 在线近红外光谱监测桂枝茯苓胶囊流化床干燥过程水分的方法研究 [J]. *中草药*, 2019, 50(22): 5429-5438
- [16] 王璐璐, 吴子建, 刘磊, 等. 表面增强拉曼光谱技术在中医药研究中的应用概况 [J]. *甘肃中医药大学学报*, 2018, 35(6): 90-93.
- [17] 刘丽, 徐宁. 基于光谱技术的中药质量控制研究进展 [J]. *发酵科技通讯*, 2019, 48(2): 85-93.
- [18] 温国基. 复杂体系气相混合物在线拉曼分析仪的开发及其应用 [D]. 杭州: 浙江大学, 2021.
- [19] 陈怡帆, 戴连奎, 朱欢银, 等. 基于拉曼光谱的聚酯过程酯化反应中清晰点的在线检测 [J]. *分析测试学报*, 2020, 39(4): 434-440.
- [20] 王忠. 拉曼光谱预处理及数据分析应用中的关键技术研究 [D]. 兰州: 兰州大学, 2021.

[责任编辑 郑礼胜]