

· 化学与分析 ·

# 近红外光谱技术结合竞争自适应重加权采样算法 用于中药定量分析

聂黎行<sup>1</sup>, 戴忠<sup>1</sup>, 马双成<sup>1\*</sup>, 张晓楠<sup>2</sup>, 解素花<sup>2</sup>

(1. 中国食品药品检定研究院, 北京 100050; 2. 北京同仁堂股份有限公司 北京 100051)

**[摘要]** 目的: 基质复杂、谱带重叠严重, 影响了中药近红外定量模型的准确性。为解决以上问题, 探讨竞争自适应重加权采样 (Competitive adaptive reweighted sampling, CARS) 变量筛选方法在中药材、中药提取物和中成药的定量分析中的应用。方法: 采集葛根药材、葛根提取物和愈风宁心滴丸的近红外漫反射光谱, 测定葛根素含量。分别优化光谱前处理方式, 剔除奇异样本后, 运用 CARS 法筛选出的相关变量, 建立偏最小二乘法 (PLS) 校正模型。结果: 原料、中间体和制剂的定量模型交互验证均方差 (RMSECV) 分别为 0.35%、1.76%、0.54%, 与基于全光谱建立的模型比较, 原料、中间体和制剂的 CARS-PLS 模型的预测准确度均有提高。结论: 竞争自适应重加权采样变量筛选方法可以提高模型的预测能力, 并有效简化运算过程, 为中药的快速、无损检测提供了新的思路。

**[关键词]** 近红外光谱; 愈风宁心滴丸; 葛根提取物; 葛根; 定量分析; 竞争自适应重加权采样

**[中图分类号]** R284.1 **[文献标识码]** A **[文章编号]** 1005-9903(2017)11-0045-05

**[doi]** 10.13422/j.cnki.syfjx.2017110045

**[网络出版地址]** <http://kns.cnki.net/kcms/detail/11.3495.R.20170309.1012.018.html>

**[网络出版时间]** 2017-03-09 10:12

## Quantitative Analysis of Traditional Chinese Medicine by Near-infrared Spectroscopy Combined with Competitive Adaptive Reweighted Sampling

NIE Li-xing<sup>1</sup>, DAI Zhong<sup>1</sup>, MA Shuang-cheng<sup>1\*</sup>, ZHANG Xiao-nan<sup>2</sup>, XIE Su-hua<sup>2</sup>

(1. National Institutes for Food and Drug Control, Beijing 100050, China;

2. Beijing Tongrentang Co. Ltd., Beijing 100051, China)

**[Abstract]** **Objective:** Accuracy of NIR model is restricted by complex matrix and overlapping bands of traditional Chinese medicine (TCM). To solve the above problem, the application of competitive adaptive reweighted sampling (CARS) in quantitative analysis of Chinese medicinal materials, TCM extract and patent medicines was discussed in this paper. **Method:** Near-infrared diffuse reflectance spectra of Puerariae Lobatae Radix, Puerarin Lobata Radix extract and Yufeng Ningxin dripping pills were collected to determine the contents of puerarin. Then the spectra were pre-processed with optimized methods to exclude uncommon samples, and CARS method was used to screen relevant variables and establish partial least squares (PLS) correction models. **Result:** Root mean square error of cross-validation (RMSECV) was 0.35%, 1.76% and 0.54%, respectively in models for raw material, semi-product and finished product. As compared with the models based on the full spectra, increased accuracy was obtained from the CARS-PLS models. **Conclusion:** CARS could increase prediction

**[收稿日期]** 20160730(009)

**[基金项目]** 国家自然科学基金项目(81303194); 国家发改委现代中药高技术产业发展专项[(2011)51]

**[第一作者]** 聂黎行, 硕士, 副研究员, 从事中药质量控制、中药化学对照品研制和光谱技术在药品质量控制中的应用等研究, Tel: 010-67095282, E-mail: nielixing@163.com

**[通讯作者]** \* 马双成, 博士, 研究员, 从事中药质量控制的研究, Tel: 010-67095272, E-mail: masc@nifdc.org.cn

accuracy of the models and simplify calibration, which would offer a new approach for rapid and non-destructive determination of TCM.

**[Key words]** near-infrared spectroscopy; Yufeng Ningxin dripping pills; Puerarin Lobata Radix Extract; Puerariae Lobatae Radix; quantitative analysis; competitive adaptive reweighted sampling

愈风宁心滴丸<sup>[1]</sup>是以葛根药材为原料,聚乙二醇和硅油为辅料,经提取、纯化、浓缩、干燥、熔融、滴制加工制成的中成药,具有解痉止痛,增强脑及冠脉血流量的作用,用于高血压头晕、头痛,颈项疼痛,冠心病与心绞痛,神经性头痛等。其主要活性成分为葛根素,标准及文献<sup>[2-3]</sup>均采用 HPLC 法控制其质量,为提高检测效率,笔者建立了 UPLC 测定愈风宁心滴丸中葛根素的方法<sup>[4]</sup>,以上方法均需要对样品进行前处理,耗时、费力,难以满足在线监测的要求。为对其生产过程进行控制,需要建立适合的在线分析方法,测定葛根药材、葛根提取物中间体及愈风宁心滴丸制剂中葛根素含量。近红外光谱(NIRS)技术具有快速、简便、无损的优势,已在中药的真伪鉴别、含量测定和过程分析等领域得到广泛的应用<sup>[5-18]</sup>。本文采用近红外光谱法对愈风宁心滴丸及其原料和中间体进行快速分析,在运用常规的偏最小二乘法(PLS)建立含量测定模型时发现,同样的成分,药材和提取物的预测结果较制剂好,推测制剂中的大量辅料干扰了葛根素的测定,需要建立变量筛选方法提取有用信息,提高模型准确性。

中药成分复杂,指标成分往往含量较低,如何从重叠的光谱中选择与化学值相关的特征变量,是建立中药含量测定模型的关键。目前多元校正分析中,波长变量方法主要包括回归系数法<sup>[19]</sup>、逐步回归法<sup>[20]</sup>、间隔偏最小二乘法<sup>[21]</sup>、移动窗口偏最小二乘法<sup>[22]</sup>、连续投影法<sup>[23]</sup>、无信息变量消除法<sup>[24]</sup>、蒙特卡罗无信息变量消除法<sup>[25]</sup>、遗传算法<sup>[26]</sup>等。竞争自适应重加权采样(Competitive adaptive reweighted sampling, CARS)是新近提出的一种用于变量筛选的方法,该方法模仿达尔文进化理论中的“适者生存”原则,通过自适应重加权采样(adaptive reweighted sampling, ARS)技术筛选出 PLS 模型中回归系数绝对值大的波长点,去掉权重小的波长点,利用交互检验选出模型交互验证均方差(lowest root mean square error of cross validation, RMSECV)最低的子集,可有效选择与所测性质相关的最优波长组合,简化模型并提高模型的预测能力<sup>[27]</sup>。自中南大学梁逸曾教授团队提出该算法以来,CARS 法在食

品的近红外分析中得到了成功的应用<sup>[28-32]</sup>,使用该方法筛选有效变量,预测中药活性成分的研究尚未见报道。本文以愈风宁心滴丸为载体,运用竞争自适应重加权采样优化波数变量,建立中药材、中药提取物和中成药的含量测定模型,为近红外在中药这一特殊、复杂对象的快速分析中的应用提供新的思路。

## 1 材料

Antaris 型近红外光谱仪(配有积分球附件, Result 操作软件与 TQ Analyst 分析软件,美国 Thermo 公司);Acquity UPLC 型超高效液相色谱仪(配有 PDA 检测器,美国 Waters 公司);AE240 型,FX-200 型电子天平(瑞士 Mettler 公司)。

甲醇、乙醇分析纯,购于北京化学试剂公司;甲醇色谱纯,购于 Merck 公司;水为超纯水。

葛根素对照品(批号 110752-200912)购自中国食品药品检定研究院。葛根(编号 1211001-1 ~ 1301034-2,经中国食品药品检定研究院康帅助理研究员鉴定,均为豆科植物野葛 *Pueraria lobata* 的干燥根。葛根提取物(中间体,编号 1 ~ 100),愈风宁心滴丸(编号 1-1 ~ 21-5),阴性样品由同仁堂股份有限公司提供。

## 2 方法与结果

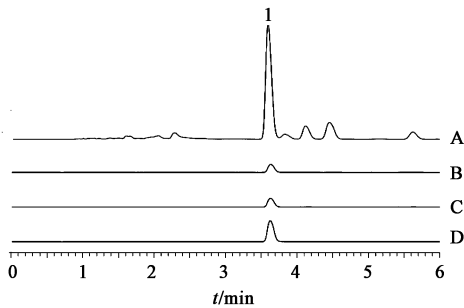
**2.1 光谱采集** 取葛根药材粉末,葛根提取物粉末,愈风宁心滴丸,阴性样品(均粉碎,过三号筛),置于内径 1 cm 的样品管内,采用积分球附件测量漫反射光谱。以仪器内置背景为参比,扫描范围 4 000 ~ 10 000  $\text{cm}^{-1}$ ,扫描次数 32 次,分辨率 8  $\text{cm}^{-1}$ 。光谱数据取 3 次采样的平均值。

**2.2 葛根素的测定**<sup>[4]</sup> 采用 UPLC 法测定药材、中间体和制剂中葛根素的含量。取愈风宁心滴丸 20 丸,精密称定,研细,取 33 mg,另取葛根提取物 20 mg,再取葛根药材粉末 0.5 g,以上样品精密称定,置 50 mL 量瓶中,加 70% 乙醇约 40 mL,超声处理(功率 250 W,频率 33 kHz)20 min,取出,放冷,加入 70% 乙醇至刻度,摇匀,用微孔滤膜(0.22  $\mu\text{m}$ )滤过,取续滤液,得供试品溶液。另精密称取葛根素对照品适量,分别加甲醇制成每 1 mL 含 1 mg(药材含测外标)和 0.1 mg(中间体、制剂含测外标)的溶液,

作为对照品溶液。色谱条件: AQUITY UPLC BEH C<sub>18</sub> 色谱柱(2.1 mm × 50 mm, 1.8 μm), 柱温 35 °C, 流动相甲醇-水(20:80), 流速 0.2 mL·min<sup>-1</sup>, 检测波长 250 nm, 进样量 1 μL。每个样品平行测定 2 次。

**2.3 数据处理与分析** 光谱预处理, 奇异点检测, PLS 建模通过 TQ Analyst 软件完成, CARS 程序在 Matlab 环境下运行。

**2.4 参比值测定结果** 药材、中间体和制剂的 UPLC 色谱, 见图 1。葛根素测定结果见表 1, 葛根提取物中被测成分含量分布范围最宽, 药材次之, 愈风宁心滴丸最窄。



1. 葛根素  
图 1 葛根药材(A), 葛根提取物(B), 愈风宁心滴丸(C)和葛根素对照品(D)的 UPLC

Fig. 1 UPLC chromatograms of Puerariae Lobatae Radix (A), Puerarin Lobata Radix extract (B), Yufeng Ningxin dripping pills (C) and puerarin reference substance (D)

表 1 不同样品中葛根素数据信息

Table 1 Contents of puerarin

分析对象	样本数 / 个	最小值 / %	最大值 / %	平均值 / %	相对标准偏差 / %
葛根药材	74	1.87	6.74	4.04	1.25
葛根提取物	100	1.29	81.88	33.38	20.46
愈风宁心滴丸	105	13.07	17.66	15.85	0.79

**2.5 光谱预处理** 葛根药材, 葛根提取物, 愈风宁心滴丸和阴性样品的近红外漫反射光谱, 见图 2。直观观察可发现, 药材和中间体的光谱较为相似。阴性样品为辅料聚乙二醇熔融后不加葛根提取物, 在硅油中滴制而成, 制剂的光谱与阴性样品更加相近, 说明葛根素的特征吸收被辅料的光谱掩蔽, 需要通过化学计量学方法加以提取。

在对变量进行筛选前需对光谱加以预处理。以 RMSECV 为指标, 组合比较了多重散射校正 (multiplicative scatter correction, MSC), 正交信号校正 (standard normal variate correction, SNV), 一阶导数光谱(1<sup>st</sup> D), 二阶导数光谱(2<sup>nd</sup> D), Norris 平滑和

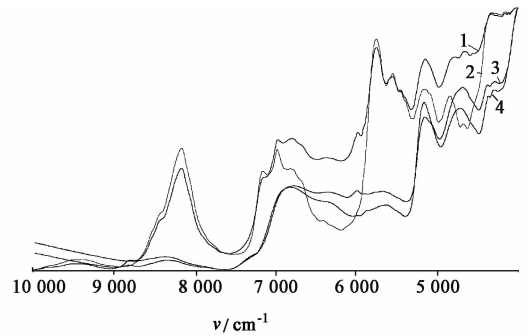


图 2 葛根药材(1), 葛根提取物(2), 愈风宁心滴丸(3)和阴性样品(4)的近红外漫反射光谱

Fig. 2 Near-infrared diffuse reflectance spectra of Puerariae Lobatae Radix (1), Puerarin Lobata Radix extract (2), Yufeng Ningxin dripping pills(3) and negative sample(4)

S-G 平滑对模型效果的影响, 确定了药材、中间体和制剂光谱各自的最佳前处理方式, 见表 2。

表 2 优化后的光谱前处理方式

Table 2 Optimized pre-processing methods

模型	光谱前处理方式
葛根药材	MSC
葛根提取物	MSC
愈风宁心滴丸	MSC + 1 <sup>st</sup> D + Norris 平滑

**2.6 CARS 变量筛选** 首先检测杠杆值和残差值较高的异常点<sup>[33]</sup>, 去掉奇异样本后, 按表 2 中方法分别对药材、中间体和制剂光谱进行前处理, 采用竞争自适应重加权采样 (CARS) 对葛根素的变量进行多次筛选, 最终分别选择 68, 30 和 8 个波数点用于建立定量分析模型, 见图 3。

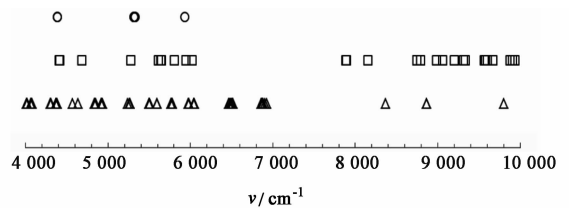


图 3 葛根药材(Δ), 葛根提取物(□), 愈风宁心滴丸(○) 筛选波数变量

Fig. 3 Selected wavenumbers of quantitative models for Puerariae Lobatae Radix (Δ), Puerarin Lobata Radix Extract (□) and Yufeng Ningxin dripping pills (○)

以葛根提取物为例说明变量筛选过程。首先采用蒙特卡罗采样 (Monte Carlo sampling, MCS) 法采样 50 次, 利用指数衰减函数 (exponentially decreasing function, EDF) 强行去掉权重相对较小的波数点。通过自适应重加权采样 (adaptive reweighted sampling, ARS) 技术筛选出 PLS 模型中回归系数绝对值大的波数点, 用每次产生的新变量子

集建立 PLS 回归模型, 计算各模型的交互验证均方差 (lowest root mean square error of cross validation, RMSECV), 最终选择 RMSECV 最小的变量子集<sup>[27]</sup>。CARS 变量筛选过程中波数变量数的变化趋势, 随着采样次数的增加, 被选择的波数变量数逐渐下降, 下降趋势由快到慢递减, 体现了粗选到精选的过程。随着采样次数的增加, RMSECV 曲线先缓慢下降, 表明变量选择过程中先剔除与葛根素无关的波数; 18 次采样之后曲线逐渐上升, 说明开始剔除与葛根素相关的波数, 丢失了有效信息。CARS 变量选择过程中各波长变量相关系数的变化趋势, 由最长的竖线对应了 50 次采样中的最佳变量选择结果, 即 RMSECV 最低点对应的 30 个波数点。见图 4。

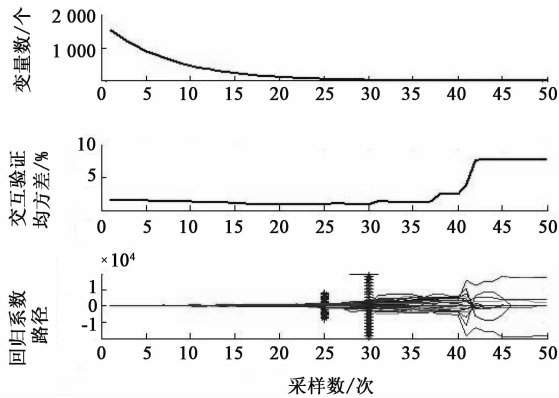


图 4 CARS 变量筛选过程示意  
Fig. 4 Plots of CARS variable selection

**2.7 PLS 模型的建立与预测** 按含量从低到高顺序排列, 分别从药材、中间体和制剂样本中均匀选择 10 个样本组成预测集, 其余组成校正集。分别选择优化的波数变量和全光谱 (4 000 ~ 10 000  $\text{cm}^{-1}$ , 1 557 个变量) 建立 PLS 模型。在采用近红外光谱对样品进行分析时, 通常采用相关系数 (correlation coefficient,  $R$ ), 校正均方差 (root mean square error of calibration, RMSEC), 验证均方差 (root mean square error of prediction, RMSEP) 和交互验证均方差 (root mean square error of cross validation, RMSECV) 等指标评价模型的预测能力。见表 3。经变量筛选后, 相比用全光谱参与计算, 3 类样本的预测结果都更加准确, 葛根药材、葛根素提取物和愈风宁心滴丸模型的 RMSECV 下降幅度分别为 28%, 14% 和 38%, 说明特征变量受其他成分干扰越强, CARS 法效果越显著。

### 3 结论

将近红外光谱技术结合竞争自适应重加权采样 (CARS) 变量选择算法应用于 3 种中药样品葛根药

表 3 模型参数

Table 3 Statistics of calibration models

模型	变量筛选方法	变量数 / 个	$R$	RMSEC / %	RMSEP / %	RMSECV / %
葛根药材	-	1 557	0.960 4	0.34	0.53	0.49
	CARS	68	0.981 3	0.24	0.24	0.35
葛根提取物	-	1 557	0.996 4	1.72	1.76	2.05
	CARS	30	0.997 8	1.35	1.66	1.76
愈风宁心滴丸	-	1 557	0.584 0	0.83	1.14	0.87
	CARS	8	0.769 2	0.48	0.51	0.54

材、葛根提取物和愈风宁心滴丸中葛根素含量的预测。结果表明, 采用 CARS 算法获得的特征变量建立的 PLS 模型性能优于全光谱 PLS 模型。CARS 算法能够非常好地克服中药样品光谱数据本身具有的噪声大、散射严重等问题, 可以有效地用于近红外光谱关键变量的提取, 不仅可以简化模型, 减少运算时间, 还提高了含量测定模型的预测精度, 能够满足在线检测的需要, 可进一步用于药品的过程控制。近红外光谱技术结合 CARS 变量筛选方法可快速、无损、准确地对中药材、中药提取物和中成药进行定量分析, 具有良好的应用前景。

#### [参考文献]

- [1] 国家食品药品监督管理局标准 (试行)[S]. YBZ00592004.
- [2] 施存元. 愈风宁心滴丸的质量标准研究[J]. 中草药, 2001, 32(10): 894-895.
- [3] 宋新波, 金华, 刘芄, 等. HPLC 法测定愈风宁心滴丸中葛根素的含量[J]. 天津中医学院学报, 2003, 22(2): 26.
- [4] 聂黎行, 戴忠, 林瑞超, 等. UPLC 法测定愈风宁心滴丸中葛根素的含量[J]. 中国药事, 2013, 27(10): 1075-1076.
- [5] 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.
- [6] 张继丹, 李化, 柯华香, 等. 近红外漫反射光谱法快速测定五味子中五味子醇甲和水分的含量[J]. 中国实验方剂学杂志, 2015, 22(21): 68-71.
- [7] 朱建光, 郝敏, 白雁, 等. 近红外漫反射光谱法快速鉴别不同连翘饮片[J]. 中国实验方剂学杂志, 2016, 22(8): 83-86.
- [8] 聂黎行, 王钢力, 李志猛, 等. 近红外光谱法对同仁乌鸡白凤丸的定性和定量分析[J]. 红外与毫米波学报, 2008, 27(3): 205-209.
- [9] BU H B, NIE L X, WANG D, et al. Rapid determination of *Panax ginseng* by near-infrared spectroscopy [J]. Anal Met, 2013, 5(23): 6715-6721.

- [10] 卜海博, 聂黎行, 王丹, 等. 近红外光谱法无损识别林下山参及其生长年限[J]. 光谱学与光谱分析, 2012, 32(7):1801-1805.
- [11] 李文龙, 瞿海斌. 近红外光谱用于中药定量分析的技术规范化研究进展[J]. 中国中药杂志, 2016, 41(19):3511-3514.
- [12] NIE L X, WANG G L, LIN R C. Development of near infrared spectroscopy for rapid quality assessment of red ginseng [J]. Chem Res Chin Univ, 2009, 25(5):633-637.
- [13] 徐冰, 罗赣, 林兆洲, 等. 基于过程分析技术和设计空间的金银花醇沉加醇过程终点检测[J]. 高等学校化学学报, 2013, 34(10):2284-2289.
- [14] 徐宁, 魏萱, 任冰, 等. 发酵冬虫夏草菌粉水分腺苷的近红外光谱定量分析及波段选择[J]. 光谱学与光谱分析, 2012, 32(7):1762-1765.
- [15] 丁长春, 方向京, 赵艳丽, 等. 近红外漫反射光谱法快速鉴别石斛属植物[J]. 光谱学与光谱分析, 2014, 34(1):82-86.
- [16] 李建宇, 徐冰, 张毅, 等. 近红外光谱用于大孔树脂纯化栀子提取物放大过程的监测研究[J]. 中国中药杂志, 2016, 41(3):421-426.
- [17] 刘桦, 赵鑫, 齐天, 等. 人参叶总皂苷大孔树脂分离纯化工艺近红外光谱在线监测模型及其含量测定[J]. 光谱学与光谱分析, 2013, 33(12):3226-3230.
- [18] 彭严芳, 史新元, 周璐薇, 等. 基于四种 NIR 仪器类型的清开灵注射液中黄芩苷成分的多变量检测限研究[J]. 光谱学与光谱分析, 2013, 33(9):2363-2368.
- [19] WU D, HE Y, FENG S, et al. Short-wave near-infrared spectroscopy analysis of major compounds in milk powder and wavelength assignment [J]. Analytica Chimica Acta, 2008, 610(2):232-242.
- [20] 褚小立, 袁洪福, 陆婉珍, 等. 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.
- [21] 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]. Appl Spectrosc, 2000, 54(3):413-419.
- [22] JIANG J H, Berry R J, Siesler H W, et al. Wavelength interval selection in multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data [J]. Analyt Chem, 2002, 74(14):3555-3565.
- [23] YE S F, WANG D, MIN S G. Successive projections algorithm combined with uninformative variable elimination for spectral variable selection [J]. Chemometr Intell Lab Syst, 2008, 91(2):194-199.
- [24] Centner V, Massart D L, Noord O E de, et al. Elimination of uninformative variables for multivariate calibration [J]. Anal Chem, 1996, 68(21):3851-3858.
- [25] CAI W S, LI Y K, SHAO X G, et al. A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra [J]. Chemometr Intell Lab Syst, 2008, 90(21):188-194.
- [26] Leardi R. Genetic algorithms in chemometrics and chemistry: a review [J]. J Chemometr, 2001, 15(7):559-569.
- [27] LI H, LIANG Y, XU Q, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration [J]. Anal Chim Acta, 2009, 648(1):77-84.
- [28] 张华秀, 李晓宁, 范伟, 等. 近红外光谱结合 CARS 变量筛选方法用于液态奶中蛋白质与脂肪含量的测定 [J]. 分析测试学报, 2012, 29(5):430-434.
- [29] 孙通, 许文丽, 林金龙, 等. 可见/近红外漫透射光谱结合 CARS 变量优选预测脐橙可溶性固形物 [J]. 光谱学与光谱分析, 2012, 32(12):3229-3233.
- [30] ZHANG X, LI W, YIN B, et al. Improvement of near infrared spectroscopic (NIRS) analysis of caffeine in roasted Arabica coffee by variable selection method of stability competitive adaptive reweighted sampling (SCARS) [J]. Spectrochim Acta A Mol Biomol Spectrosc, 2013, 114(12):350-356.
- [31] 孙通, 林金龙, 许文丽, 等. 可见/近红外半透射结合 CARS 方法在线检测脐橙 SSC [J]. 江苏大学学报: 自然科学版, 2013, 34(6):663-668.
- [32] 李江波, 彭彦昆, 陈立平, 等. 近红外高光谱图像结合 CARS 算法对鸭梨 SSC 含量定量测定 [J]. 光谱学与光谱分析, 2014, 34(5):1264-1269.
- [33] Mc Carthy W J. TQ Analyst user's guide [M]. Madison, WI. Thermo Nicolet Corp, 2000.

[责任编辑 顾雪竹]