

青蒿浓缩过程在线近红外快速检测模型的建立

徐芳芳^{1,2}, 冯双双^{2,3}, 李雪珂^{1,2}, 吴建雄^{2,3}, 毕宇安^{2,3}, 王振中^{2,3}, 萧伟^{1,2,3*}

1. 南京中医药大学, 江苏 南京 210000

2. 江苏康缘药业股份有限公司, 江苏 连云港 222001

3. 中药制药过程新技术国家重点实验室, 江苏 连云港 222001

摘要: **目的** 建立青蒿单效浓缩过程中近红外在线快速检测模型, 并讨论吸光度的变化对其模型建立的影响。**方法** 在线收集 9 批浓缩液样本, 偏最小二乘 (PLS) 法建立定量校正模型, 并用此模型对 1 批样品进行预测。**结果** 总酸和固含量 PLS 模型参数如下: 决定系数 (R^2) 分别为 0.967 9 和 0.962 3, 校正集均方根误差 (RMSEC) 分别为 0.783 5 和 0.948 8, 交叉验证集均方根误差 (RMSECV) 分别为 0.825 8 和 0.978 0。**结论** 青蒿浓缩液样品吸光度范围为 0~2.0, 该光谱建立的总酸和固含量的 PLS 模型的预测相对偏差 (RSEP) 值均在 10% 以内, 能够满足青蒿生产过程中质量要求, 说明当吸光度范围为 0~2.0 时, 通过增加样品数和运用化学计量学方法可消除吸光度太高 (即透过率太低) 对青蒿近红外模型建立的影响。

关键词: 青蒿; 单效浓缩过程; 近红外; 在线快速检测; 吸光度; 偏最小二乘法; 定量校正模型; 总酸; 固含量; 决定系数; 校正集均方根误差; 交叉验证集均方根误差; 预测相对偏差; 化学计量学; 透过率

中图分类号: R284.2 文献标志码: A 文章编号: 0253-2670(2016)10-1690-06

DOI: 10.7501/j.issn.0253-2670.2016.10.011

Construction of NIRS in-line rapid detection model of concentration process of *Artemisiae Annuae Herba*

XU Fang-fang^{1,2}, FENG Shuang-shuang^{2,3}, LI Xue-ke^{1,2}, WU Jian-xiong^{2,3}, BI Yu-an^{2,3}, WANG Zhen-zhong^{2,3}, XIAO Wei^{1,2,3}

1. Nanjing University of Chinese Medicine, Nanjing 210000, China

2. Jiangsu Kanion Pharmaceutical Co., Ltd., Lianyungang 222001, China

3. State Key Laboratory of New-tech for Chinese Medicine Pharmaceutical Process, Lianyungang 222001, China

Abstract: Objective In this study, an in-line near infrared spectroscopy (NIRS) rapid detection model of single-effect concentration process of *Artemisiae Annuae Herba* (AAH) was established and the impact of absorbance to model was discussed. **Methods** Nine batches of concentration process samples were collected inline and partial least square (PLS) method was applied to build its quantitative model which was used to predict an unknown batch of samples. **Results** The parameters of PLS model of total acid and soluble solid content were as follows: coefficient of determination (R^2) of 0.962 3 and 0.967 9, root mean square errors of calibration (RMSEC) of 0.783 5 and 0.948 8, and root mean square errors of cross-validation (RMSECV) of 0.825 8 and 0.978 0, respectively. **Conclusion** Absorbance of concentration process samples of AAH was between 0 and 2.0 and RSEP values of total acid and soluble solid content was less than 10%. It can meet the quality requirements of production process of AAH and indicate that the effect of high absorbance (namely low transmittance) on NIRS model of AAH can be eliminated by adding samples and using chemometrics when the absorbance is between 0.0 and 2.0.

Key words: *Artemisiae Annuae Herba*; single-effect concentration process; NIRS; in-line rapid detection; absorbance; partial least square; quantitative calibration model; total acid; soluble solid content; coefficient of determination; RMSEC; RMSECV; RSEP; chemometrics; transmittance

收稿日期: 2015-12-15

基金项目: 科技部重大新药创制现代中药创新集群与数字制药技术平台 (2013ZX09402203)

作者简介: 徐芳芳 (1990—), 女, 安徽宣城人, 南京中医药大学在读博士, 研究方向为新型中药制剂的研究与开发。

Tel: 13815665047 E-mail: shannaifen@163.com

*通信作者 萧伟, 男, 研究员级高级工程师, 博士, 研究方向为中药新药的研究与开发。Tel: (0518)81152367 E-mail: kanionlunwen@163.com

青蒿 *Artemisiae Annuae Herba* 是菊科 (Compositae) 青蒿属 *Artemisia* L. 植物黄花蒿 *Artemisia annua* L. 的干燥地上部分, 秋季花盛开时采收, 具有清热解暑、除蒸、截疟等功效^[1-2]。青蒿作为热毒宁注射液的君药, 提取工艺分为挥发油提取 (挥发性成分) 和加水煎煮 (非挥发性成分) 两部分。经过高温煎煮, 青蒿素类成分受到破坏, 其主要有效非挥发性成分为有机酸类^[3-4]。由于有机酸类的量较低, 在提取精制阶段并无明确的在线控制指标, 本实验首次将近红外光谱 (NIRS) 技术引入到热毒宁注射液原料青蒿的生产工艺过程中, 以总酸和固含量为在线监控指标, 建立非挥发性成分的快速定量测定模型。NIRS 技术作为一种快速、无损的过程分析技术^[5], 已经在食品、农产品等领域被较为广泛地成功运用^[6], 但在中药领域的应用还不是很成熟。由于中药体系的复杂性, NIRS 技术的应用针对不同的中药材有不同的特性。

在热毒宁注射液青蒿醇沉上清液的单效浓缩生产工艺过程中, 青蒿醇沉上清液未浓缩前为透明的淡黄色溶液, 由于浓缩过程中水醇溶液的挥发, 溶液颜色逐渐变为不透明的深棕色, 溶液颜色变化引起的 NIRS 吸光度 (A) 变化范围为 0~2.0, 变化的幅度比较大, 据文献报道, A 值大于 1.5 的光谱不宜建模^[7-8], 但是并无相关的明确研究确切地说明 A 值的大小对近红外建模效果的影响, 只是理论的推测。本实验以青蒿单效浓缩液为研究对象, 确切地说明了 A 值大于 1.5 的光谱建模的可行性, 并分析了原因, 为 NIRS 技术在中药领域的应用做了探索和铺垫。

1 仪器与材料

Luminar 3060 型 AOTF 近红外光谱分析仪, 美国 Brimrose 公司; BSA224S-CW 型万分之一电子天平, 德国 Sartorius 公司; 0.2、1.0 mL 移液枪, 德国 Eppendorf 公司; H1650 型台式高速离心机, 湖南湘仪实验室仪器开发有限公司; DHG-9143BS-III 型电热恒温鼓风干燥箱, 上海新苗医疗器械制造有限公司; 负压自动取样机, 无锡维邦工业设备成套技术有限公司; C22N/500 单效浓缩罐, 浙江温兄机械阀门有限公司; CoriolisMaster CE0045/0044 在线密度计, ABB 公司。

绿原酸对照品 (批号 110753-201415, 质量分数 96.2%), 中国食品药品检定研究院; 分析纯甲醇, 南京化学试剂有限公司; 青蒿单效浓缩液 (相对密

度 1.10~1.12, 温度 60~70 °C) 由江苏康缘药业股份有限公司中药现代化数字化提取工厂提供。

2 方法与结果

2.1 青蒿单效浓缩样品的收集

在取样点连接负压取样机, 调试取样参数, 在浓缩刚开始时启动取样程序, 终点浓缩前每 5 分钟取 1 个样, 进入终点浓缩后每分钟取 1 个样。共取样 9 批, 分别为 Z150417 (样 10 个)、Z150502 (样 18 个)、Z150602 (样 25 个)、Z150603 (样 20 个)、Z150702 (样 20 个)、Z150811 (样 22 个)、Z1508112 (样 24 个)、Z150818 (样 24 个)、Z150819 (样 26 个), 共计 189 个样品。

2.2 在线 NIRS 的采集

打开单效浓缩罐 T3015 的在线近红外检测 16 通道, Luminar 3060 近红外仪器有关参数设置如下: 光谱扫描范围 1 100~2 300 nm, 扫描次数 300 次, 分辨率 1 nm, 波长增量 2 nm, 光纤透射探头光程 2 mm, Ratio 模式。

2.3 单效浓缩过程中监控指标的测定

2.3.1 密度测定 单效浓缩罐上装有在线密度计, 可实时在线监测浓缩液的密度。

2.3.2 总酸测定 热毒宁注射液有效成分有机酸类成分在青蒿中的量较低, 低于近红外的检测限, 所以本实验中以总酸的量作为控制指标。

(1) 对照品溶液的制备 精密称量绿原酸 10.023 mg, 用 50% 甲醇定容到 50 mL 棕色量瓶中, 置于 4 °C 的冰箱中作为母液备用。移液枪移取绿原酸母液 0.125、0.200、0.250、0.400、0.500、0.800、1.000 mL, 用 50% 甲醇分别定容到 10 mL 棕色量瓶中, 得到系列对照品溶液。

(2) 供试品溶液的制备 将取得的生产样品用 50% 甲醇稀释到合适的质量浓度, 即得供试品溶液。

(3) 供试品溶液的测定^[12] 用 50% 甲醇作为空白, 根据热毒宁注射液质量标准, 在 327、400 nm 下分别测定对照品和样品的 A 值 A_1 、 A_2 , 建立对照品溶液质量浓度 (X , $\mu\text{g/mL}$) 和 ΔA 值 ($Y=A_1-A_2$) 之间的标准曲线 ($R^2 \geq 0.9998$), 标准曲线方程 $Y=0.05292X-0.00104$, $R^2=0.9999$, 结果表明总酸在 2.51~20.05 $\mu\text{g/mL}$ 线性关系良好。测定样品的量, 稀释的样品 A 值控制在 0.2~0.7。

2.4 固含量测定^[3]

固含量可以反映青蒿浓缩液中指标成分的总体变化。中药作为一个复杂的体系, 从中药整体性角

度出发,选取了固含量作为近红外实时监控指标。

将蒸发皿烘干至恒定质量 (M_0)。吸取青蒿单效浓缩过程中药液适量 (M_2)。水浴蒸干后, 105 °C 烘至恒定质量, 取出置干燥器中冷却 30 min, 迅速称定质量 (M_1), 固含量 = $(M_1 - M_0) / (M_2 - M_0)$ 。

2.5 NIRS 模型的建立及模型验证

青蒿 NIRS 在线监测模型采用经典的偏最小二乘 (PLS) 法建立, 优化模型主要采用相关系数 (R) 或决定系数 (R^2)、校正集均方根误差 (RMSEC) 和交叉验证集均方根误差 (RMSECV) 来评价, 模型预测准确性主要采用 R 或 R^2 、预测集相对偏差 (RSEP) 来验证。

2.5.1 光谱建模波段选择 青蒿单效浓缩液的原始光谱图见图 1。因为青蒿单效浓缩液的初始溶剂是 80% 左右的乙醇, 所以光谱在 1 450 nm 及 1 920 ~ 1 950 nm 附近水峰的干扰没有纯水明显。但还是比较了除去水峰的波段和含水峰波段建模的效果, 结果见表 1。由表 1 可知, 波段 1 100 ~ 2 300 nm 具有更高的 R 值和更低的 RMSEC、RMSECV 值, 因此, 本实验中直接采用的是仪器固有的扫描波段, 即 1 100 ~ 2 300 nm 为建模波段。

2.5.2 光谱预处理方法选择 在线采集光谱, 会有很多因素的干扰, 可能会造成光谱基线的偏移和噪音信号的放大, 因而需要对光谱进行预处理, 不同

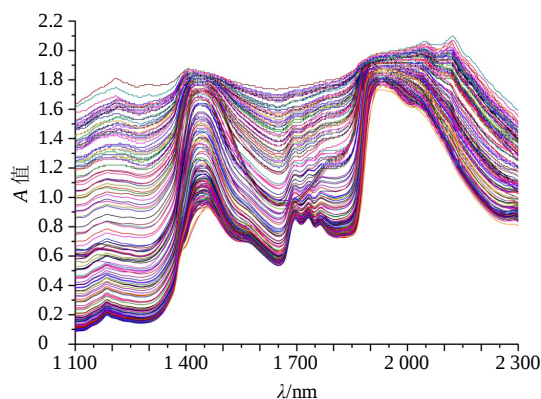


图 1 青蒿单效浓缩液的原始 NIRS

Fig. 1 Original NIRS of single-effect concentration solution of AAH

预处理方法得到的模型参数见表 2。结果表明, 1 阶导数结合 S-G 平滑的方法具有最高的 R 值和最低的 RMSEC、RMSECV 值, 因此光谱预处理方法选择 1 阶导数结合 S-G 平滑法, 预处理后 NIRS 见图 2。

2.5.3 异常样品剔除 总酸和固含量样品的残差-杠杆值分布图见图 3。杠杆值和残差值越接近于 0 越有利于模型的建立, 远远偏离空间样品相对集中分布区域的样品则为异常样品。因此样本 10、46、135、52、73 为总酸模型的异常样本, 样本 10、47、135、52、73 为固含量模型的异常样本。

2.5.4 建立 PLS 定量模型 将优化后的 9 批光谱数

表 1 不同波段总酸和固含量的模型参数比较

Table 1 Comparison on model parameters of total acid and soluble solid contents using different wavebands

波段/nm	总酸			固含量		
	校正集 R^2	RMSEC	RMSECV	校正集 R^2	RMSEC	RMSECV
1 100 ~ 1 350	0.863 2	1.385 1	1.430 5	0.844 9	2.055 1	2.287 8
1 500 ~ 1 850	0.901 4	0.901 8	0.981 8	0.912 6	1.235 6	1.348 7
1 100 ~ 1 500	0.893 5	1.012 3	1.136 7	0.886 7	1.565 0	1.647 8
1 100 ~ 2 000	0.941 8	0.884 5	0.983 2	0.938 9	1.043 8	1.198 4
1 100 ~ 2 300	0.967 9	0.783 5	0.825 8	0.962 3	0.948 8	0.978 0

表 2 不同光谱预处理方法所建立的模型参数

Table 2 Comparison on model parameters of total acid and soluble solid content using different preprocessing methods

预处理方法	总酸			固含量		
	校正集 R^2	RMSEC	RMSECV	校正集 R^2	RMSEC	RMSECV
原始光谱	0.913 6	0.974 5	1.035 6	0.907 8	1.154 4	1.230 5
1 阶导数	0.953 0	1.157 8	1.317 6	0.950 9	1.268 6	1.486 6
2 阶导数	0.927 1	2.203 8	2.578 4	0.938 4	2.495 0	2.862 8
1 阶导数 + S-G 平滑	0.967 9	0.783 5	0.825 8	0.962 3	0.948 8	0.978 0
2 阶导数 + S-G 平滑	0.938 5	1.855 5	1.917 6	0.940 5	2.171 0	2.356 3

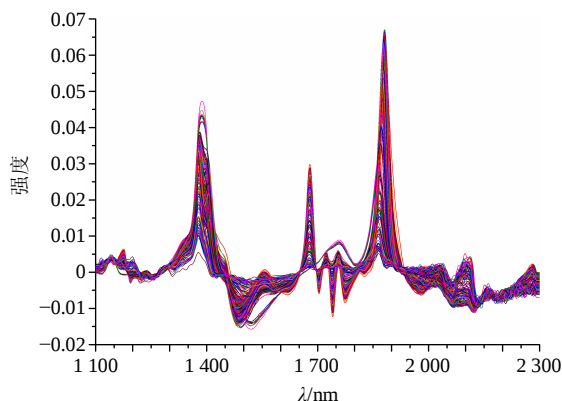


图 2 一阶导数+S-G 平滑处理后的 NIRS

Fig. 2 NIRS processed by treatment of first-order derivate and Savitzky Golay smooth

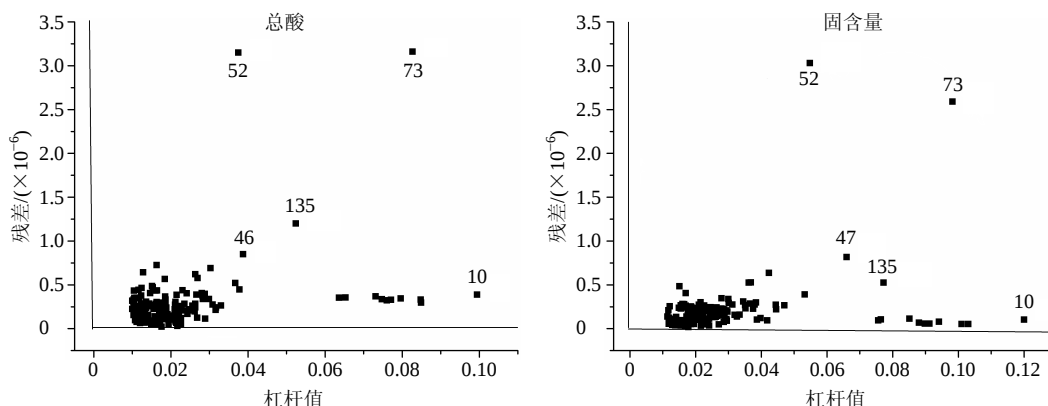


图 3 总酸和固含量样品的残差-杠杆值分布图

Fig. 3 Distribution of sample residual leverage values of total acid and soluble solid content

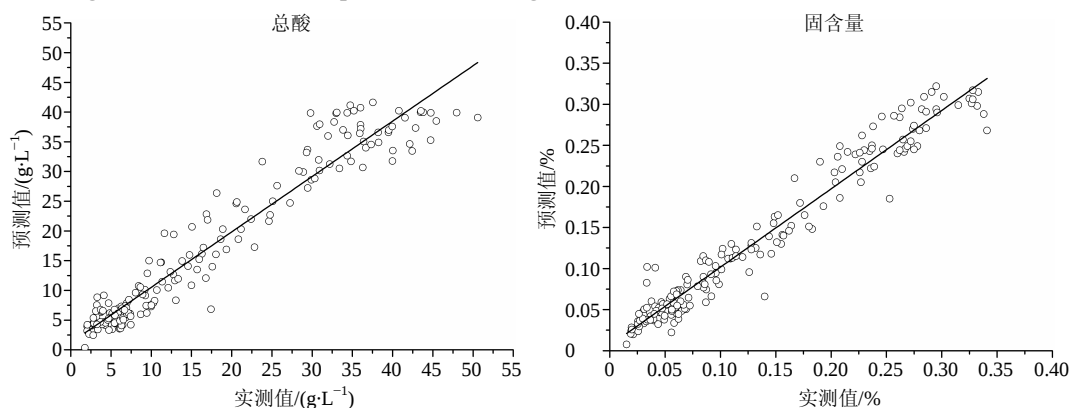


图 4 校正集总酸和固含量预测值与实测值的线性相关性图

Fig. 4 linear correlation figure of predicted and actual values of total acid and soluble solid contents of calibration set

总酸与固含量的值, 记录取样时间, 找出样品对应时间点的光谱及预测值, 计算验证集的 R^2 和 RSEP。总酸和固含量验证集 R^2 分别为 0.995 9 和 0.993 5, 其验证集总酸和固含量预测值与实测值的线性相关性见图 6。RSEP 分别为 6.15% 和 7.14%, 均在 10% 以内。

据进行合并, 导入 Unscrambler 建模软件, 并将光谱一一对应总酸和固含量的实测值, 采用留一交叉验证法建立青蒿总酸和固含量的定量校正模型。总酸和固含量的 PLS 模型的校正集 R^2 分别为 0.967 9 和 0.962 3, RMSEC 分别为 0.783 5 和 0.948 8, RMSECV 分别为 0.825 8 和 0.978 0。校正集预测值与真实值的线性关系见图 4。所建模型的 R 值较大, RMSEC 和 RMSECV 值较小, 可用于青蒿单效浓缩过程总酸和固含量的定量测定。

2.5.5 模型验证^[3,10] 将建好的模型文件导入近红外预测软件中, 在线收集单效浓缩液第 10 批 Z150829 的 NIRS, 总酸和固含量预测值的趋势图见图 5。在采集光谱的过程中, 在线取 10 个样测定其

3 讨论

单效浓缩过程不同于离线实验旋转蒸发瓶中的浓缩过程, 由于浓缩液量大, 分为补液浓缩和终点浓缩 2 部分。本实验中, 青蒿补液浓缩约 30 min, 终点浓缩约 15 min。在补液浓缩过程中, 浓缩达到一定液位, 就会继续加样, 其量的变化较为缓慢。

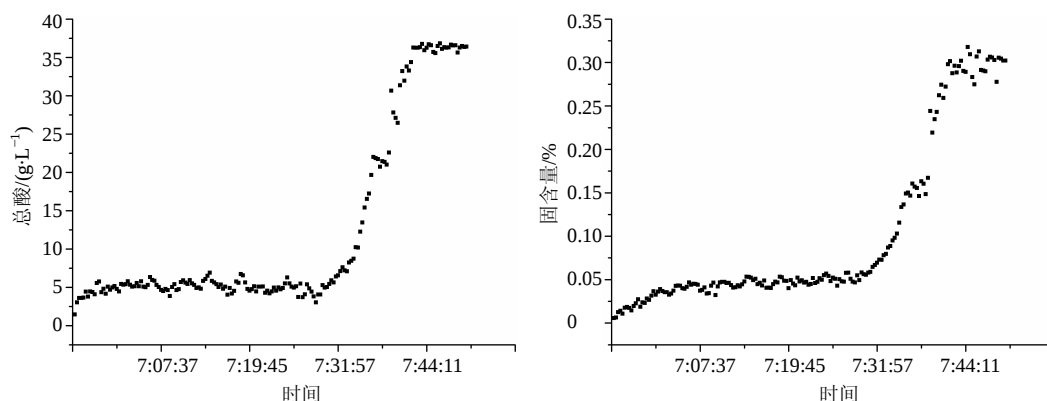


图 5 Z150829 批总酸和固含量的全过程预测值的趋势图

Fig. 5 Trendgram of predicted values of total acid and soluble solid contents in entire procedure of Z150829

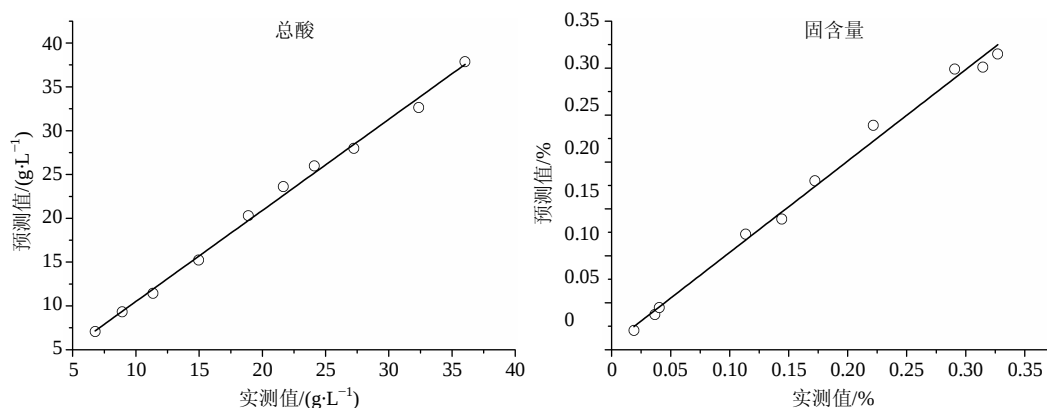


图 6 验证集总酸和固含量预测值与实测值的线性相关性

Fig. 6 linear correlation of predicted and actual values of total acid and soluble solid contents of validation set

进入终点浓缩后, 不再补液, 其量变化随体积的减少急剧增加, 直至达到终点。补液浓缩过程中, 青蒿浓缩液的颜色由浅黄色变为深黄色, A 值变化见图 7; 终点浓缩后, 溶液颜色由深黄色变为深棕色, A 值变化见图 8。

A 与透过率 (T) 的关系式为 $A = -\lg T$, 即 $T = 10^{-A}$, 当 A 值在 1~2, 即光的 T 在 1%~10%, 光

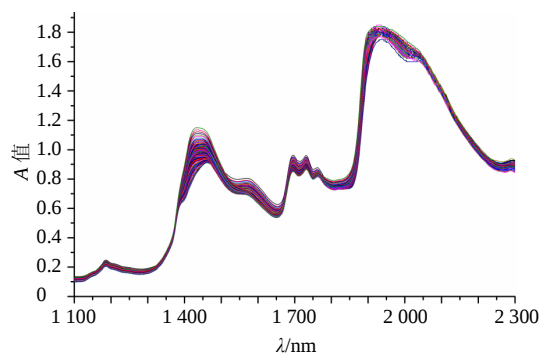
图 7 青蒿单效浓缩液补液浓缩过程 A 值变化图

Fig. 7 Absorbance variation of fluid infusion concentration process of single-effect concentration solution of AAH

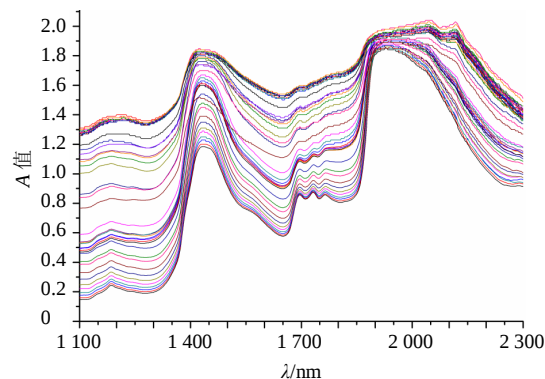


图 8 青蒿单效浓缩液终点浓缩过程吸光度变化图

Fig. 8 Absorbance variation of end point concentration process of single-effect concentration solution of AAH

的 T 值大小直接影响到有效信息的获取, 根据文献报道^[7-8], A 值大于 1.5 的光谱不宜建模, 主要原因可能是大部分物质如果 T 值太低, 其获取的有效信息也相对较少, 因而不提倡建模, 但是这个有效信息和物质对近红外的响应有关, T 值低不代表响应低, 如果运用合适的化学计量学方法, 就可以将物

质的有效信号再放大,经过化学计量学的计算,响应高的物质虽然透过率低,建模依然是可行的。此理论可通过青蒿的终点浓缩来说明,由图 8 可知,终点浓缩过程中光谱 A 值很快由 0.15 左右升至 1.3 左右,光谱波峰 A 值范围维持在 1.6~2.0。本实验中青蒿的终点浓缩过程持续时间约 15 min,约 12 min 的时间光谱波峰 A 值大于 1.5,但是建模效果良好,推测青蒿对近红外的响应良好,样本数和光谱预处理方法可以消除透过率不足造成的误差。在本实验中,补液浓缩过程是每 5 分钟取 1 个样,1 批约 6 个样,终点浓缩过程是每分钟取 1 个样,约 15 个样,终点浓缩的样本数是补液浓缩的 2.5 倍左右。预测集补液浓缩过程的总酸和固含量的 RSEP 值分别为 5.88% 和 8.64%,终点浓缩过程的总酸和固含量的 RSEP 值分别为 3.77% 和 3.01%,终点浓缩过程的 RSEP 值小于补液浓缩过程,由此推测,NIRS 模型中 1 个浓度点的样本数越多,对该浓度点的预测越准确。此外,青蒿 PLS 模型建立前,光谱预处理采用了 1 阶导数结合 S-G 平滑方法,可以降低外界因素对 NIRS 模型的干扰,提取更多的有效建模信息。本实验中青蒿单效浓缩 PLS 模型的建立说明样本数和光谱预处理可以消除吸光度过高带来的理论偏差,NIRS 模型的建立是可行的。

4 结论

本研究首次建立了青蒿单效浓缩过程总酸和固含量的快速定量测定模型,该模型能够实时监控青蒿单效浓缩过程中总酸和固含量的变化,增进了对该浓缩过程的了解和控制。此外,通过青蒿浓缩液颜色的变化导致 NIRS 吸光度的变化,阐述了增加样本数和优选光谱预处理方法可以消除吸光度过高即透过率太低带来的系统误差。在线的模型由于受

药液温度、流速、固体颗粒的影响,光谱噪音很大,需要进行光谱预处理,来提取更多的建模有效信息;同时在线模型样本数量要远大于实验室离线样本数,为了得到更精确的模型,需不断地补充样本数,对模型进行更新和校正,使预测的结果更加准确。

参考文献

- [1] 中国药典 [S]. 一部. 2015.
- [2] 张秋红,朱子微,李晋,等. 中药青蒿化学成分与种植研究现状 [J]. 中国医药导报, 2011, 8(19): 10-12.
- [3] 杜文俊,刘雪松,陶玲艳,等. 热毒宁注射液金银花和青蒿(金青)醇沉过程中多指标的近红外快速检测 [J]. 中草药, 2015, 46(1): 61-62.
- [4] 李海波. 热毒宁注射液药效物质基础研究 [D]. 南京: 南京中医药大学, 2013.
- [5] 张广仁,吴云,靳瑞婷,等. 基于近红外光谱分析技术的散结镇痛胶囊药物质量分数分析 [J]. 中国中药杂志, 2015, 40(16): 3204-3208.
- [6] 孙通,徐惠荣,应义斌. 近红外光谱分析技术在农产品/食品品质在线无损检测中的应用研究进展 [J]. 光谱学与光谱分析, 2009, 29(1): 122-125.
- [7] 金叶,丁海樱,吴永江,等. 近红外光谱技术用于血必净注射液提取过程的在线检测研究 [J]. 药物分析杂志, 2012, 32(7): 1214-1219.
- [8] Niu X Y, Shen F, Yu Y F, *et al.* Analysis of sugar in Chinese rice wine by fourier transform near-infrared spectroscopy with partial least-squares regression [J]. *J Agric Food Chem*, 2008, 56(16): 7271-7278.
- [9] 王雪,张伟,张亚飞,等. 热毒宁注射液中总有机酸含量的快速测定 [J]. 时珍国医国药, 2015, 26(4): 836-837.
- [10] 吴莎,刘启安,李云,等. 最小二乘支持向量机和偏最小二乘法在栀子中间体纯化工艺近红外定量分析中的应用 [J]. 中草药, 2015, 46(7): 990-991.