

# 栀子药材提取工艺的近红外光谱实时控制方法研究

严诗楷<sup>1,2</sup>, 罗国安<sup>2\*</sup>, 王义明<sup>2</sup>, 程翼宇<sup>1</sup>

1. 浙江大学药学院中药科学与工程学系, 浙江 杭州 310031

2. 清华大学化学系, 北京 100084

**摘要** 提出了用近红外光谱技术实时控制中药提取工艺的新方法。采用光纤近红外光谱法分析不同工艺条件下的栀子药材提取液样本, 通过 SIMCA 方法建立了工艺稳态监控模型, 用于评价 26 个未知样本的工艺状态, 成功地判别了其中的全部异常工艺样本。结果表明, 近红外光谱结合 SIMCA 分析方法是一种快速、方便、有效的工艺监测手段, 该方法无需对指标成分进行定量, 即可通过样品整体近红外光谱监控其工艺是否正常, 从而为中药提取工艺的实时控制提供了一条新的思路。

**主题词** 栀子; 近红外光谱; SIMCA; 实时控制

**中图分类号:** O657.3 **文献标识码:** A **文章编号:** 1000-0593(2006)06-1026-05

## 引言

现行中药生产工艺控制模式,基本上仍停留在传统的经验控制方法上,即仅针对生产过程的温度、压力、反应时间等物理参数进行控制,很少考虑工艺过程中的成分变化,控制过于简单粗略,难以表达出生产过程中复杂的物理和化学变化过程,使生产工艺很难得到有效控制,从而不能保证中药质量的稳定可控。色谱指纹图谱技术<sup>[1-4]</sup>针对中药复杂体系整体性和指纹性特点,是中药质量控制<sup>[5,6]</sup>和工艺控制<sup>[7,8]</sup>的一种有效技术手段,该技术通过指纹图谱整体定性和指标成分定量分析,可全面反映工艺过程中各类成分质的变化情况,克服了传统控制方法的不足。但色谱指纹图谱的获取的自动化程度低,需要消耗大量的时间、人力和物力,而且需要离线分析,难以实现工艺过程的实时控制,在一定程度上限制了指纹图谱技术的在中药生产工艺上的推广和应用。

近红外光谱是近年来迅速发展的一种快速分析技术,无需对样品作复杂的前处理即可直接分析,具有方便、快速、廉价、非破坏、无污染等优点,在石油、化工、农业、制药等领域应用广泛<sup>[9-12]</sup>,近年来又在中药领域成为研究热点<sup>[13-15]</sup>。近红外光谱是一种快速、有效的中药工艺实时控制技术,但近红外光谱用于工艺控制,往往是基于指标成分或有效成分的定量分析过程<sup>[16,17]</sup>,不能充分体现中药复杂性和整体性特点,因此,发展一种实时、快速、准确,且能从整体

上表达中药复杂体系特征的工艺控制新方法是十分必要的。

栀子茜草科植物山栀(*Gardenia jasminoides*, Ellis)的成熟干燥果实,具有清热利尿、泻火除烦、凉血解毒之功效。本文以栀子药材的提取工艺为研究对象,测定不同工艺条件下提取物样本的近红外图谱,无需定量测定其中的指标成分,直接用提取物样本的整体近红外光谱建立 SIMCA 工艺监测模型,用于评价栀子提取工艺,结果令人满意。

## 1 原理和方法

SIMCA 方法又称相似分析法,是 1976 年瑞典化学家 Wold 提出<sup>[18]</sup>。在相同的工艺条件下,栀子提取液具有相似的特征,因此 SIMCA 用于工艺评价的基本思路即是把样本按工艺归类,使每个类别代表一种工艺特征,并分别建立各类的主成分回归模型,用各模型对未知样本进行判别分析,即可根据未知样本的类别属性评价其工艺特征。

设训练集有  $N$  个样本,每个样本用  $p$  个变量描述,整个训练集样本数据矩阵可以记为  $\mathbf{M}_{p \times N}$ ,其中包含有类别  $q$ ,该类有  $n_q$  个样本,该类中的元素  $x_{ik}^{(q)}$  ( $i=1, 2, \dots, p; k=1, 2, \dots, n_q$ ),可用下述主成分模型表示:

$$x_{ik}^{(q)} = \bar{a}_i^{(q)} + \sum_{a=1}^A \beta_{ia} \cdot \theta_{ak} + \epsilon_{ik}^{(q)}$$

其中  $\bar{a}_i^{(q)}$  为  $q$  类中  $n_q$  个样本第  $i$  个变量的均值,  $A$  为该类样本数据矩阵的主成分数,  $\epsilon_{ik}$  是训练集光谱残差,  $\beta_{ia}$  和  $\theta_{ak}$  是为使  $\epsilon_{ik}$  达到极小时的参数。则该  $q$  类样本的总体标准偏差

收稿日期: 2005-03-06, 修订日期: 2005-06-16

基金项目: 国家“973”重点基础研究项目(G1999054404)及科技部重大专项(2002BA906A29-3)资助项目

作者简介: 严诗楷, 1974 年生, 浙江大学药学院博士研究生 \* 通讯联系人

可以表示为

$$S_0^{(q)} = \left[ \sum_i \sum_k (\epsilon_{ik}^{(q)})^2 / (n_q - A - 1)(p - A) \right]^{1/2}$$

模型建立后, 判别一个未知样本  $x_{ik}$  ( $i = 1, 2, \dots, p$ ) 是否属于  $q$  类, 使用多元线性回归法, 即用该样本拟合到上述主成分回归模型, 由于是未知样本, 得分与原训练集构造模型的不同, 以  $C_a$  表示,  $\epsilon_{ia}^{(q)}$  是拟合后的光谱残差,

$$x_u^{(q)} = \bar{a}_i^{(q)} + \sum_{a=1}^A \beta_{ia} \cdot C_a + \epsilon_{ia}^{(q)}$$

这个未知样本拟合  $q$  类回归模型的结果好坏, 可以用标准偏差表示,

$$S_u^{(q)} = \left[ \sum_{i=1}^p (\epsilon_{iu}^{(q)})^2 / (q - A) \right]^{1/2}$$

有了  $S_0^{(q)}$  和  $S_u^{(q)}$ , 就可以判断未知样本是否属于  $q$  类。为了判别未知样本的类别属性, SIMCA 方法引入了  $F$  检验统计量, 即

$$F = \frac{S_u^{(q)2}}{S_0^{(q)2}}$$

将  $F$  统计量与临界值  $F_0$  进行比较, 如果  $F > F_0$ , 则样

本  $u$  归于  $q$  类, 否则归属于其他类。

## 2 实验部分

### 2.1 实验仪器及参数设置

美国 Micron Es-512A 近红外光谱仪(波长范围 930~1 670 nm); 石英光导纤维透射式探头附件(光程 10 mm, 长度 1 m); 半导体冷冻 InGaAs 检测器; 测定过程中以空气作为参比, 积分时间 20 ms, 分辨率 1.4 nm。

NIR 光谱采集使用 Micron Optical Systems 软件(Micron Optical System Inc.); SIMCA 模型、留一交叉验证法、SNV、MSC 等相关程序参照文献方法<sup>[9-12]</sup>使用软件 MATLAB5.3(Mathworks Inc.)自编。

### 2.2 供试品制备

栀子药材产自江西新干, 经西南交通大学张兴国教授鉴定, 干燥, 粉碎, 过 40 目筛备用。精确称量药材粉末 46 份, 每份重 1.0 g, 然后按照表 1 所列不同提取工艺条件提取, 并经过离心、过滤, 得到 46 个栀子提取液样本用于 NIR 分析。

Table 1 46 samples extracted in different procedures

No.	Extraction method	Solvent	Solvent volume/mL
1-33	ultrasonic extraction 30 min	methanol	10
34	ultrasonic extraction 30 min	90% methanol	10
35	ultrasonic extraction 20 min	methanol	9
36	ultrasonic extraction 35 min	methanol	10
37	ultrasonic extraction 35 min	methanol	8
38	ultrasonic extraction 20 min	methanol	10
39	ultrasonic extraction 15 min	methanol	11
40	ultrasonic extraction 25 min	methanol	12
41	rough powder without sieved ultrasonic extraction 30 min	methanol	10
42	ultrasonic extraction 30 min	methanol	8
43	soaking for 10 min, then ultrasonic extraction 30 min	methanol	10
44	soaking for 15 min, then ultrasonic extraction 30 min	methanol	10
45	soaking for 20 min, then ultrasonic extraction 30 min	methanol	10
46	soaking for 200 min at 60 °C	methanol	10

### 2.3 NIR 光谱获取

按照 2.1 节方法测定各样本近红外光谱如图 1。样本累

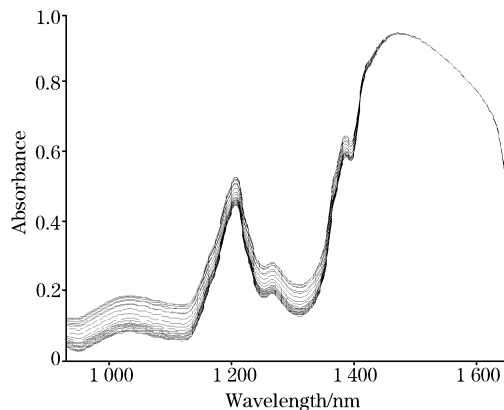


Fig. 1 Original NIR spectra of the samples of Gardenia

计扫描 50 次取平均值。

## 3 结果与讨论

### 3.1 光谱数据预处理

为从光谱量测数据中充分提取有效信息, 减少各种误差对模型的影响, 需对光谱信号进行预处理, 包括 S-G 卷积平滑、一阶或二阶求导(1st or 2nd derivative)、多元散射校正(MSC, Multiplicative Scatter Correction), 标准化(SNV, Standard Normal Variate)等方法处理, 以消除高频噪声、基线漂移、样品浓度及光散射等因素的影响。本文分别采用原谱和经 MSC、SNV、一阶微分、二阶微分等方法预处理后的量测数据建立模型, 用模型光谱残差平方和(PRESS)和对未知样品工艺的预测正确率为指标对模型进行了评价, 结果列于表 2。

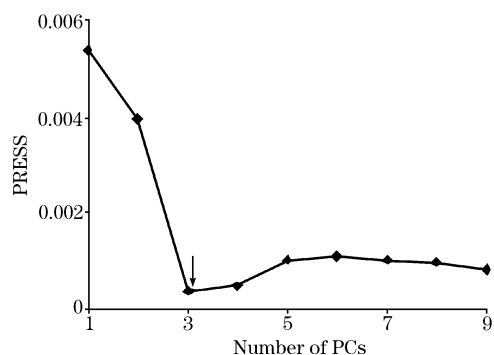
**Table 2 Model performances for different pretreatment methods**

Pretreatment method	Origin	MSC	SNV	1st de.	2nd de.	1st de. +MSC	2nd de. +MSC
PRESS/ $10^{-4}$	3.27	3.32	9.99	4.35	6.50	3.92	6.88
Prediction accurate rate	91%	94%	91%	98%	85%	100%	94%

结果表明,原始光谱经一阶微分和 MSC 校正后,所建模型的光谱残差较小,未知样本预测的结果全部正确, SIMCA 模型性能最优。

### 3.2 最佳主成分数的选择

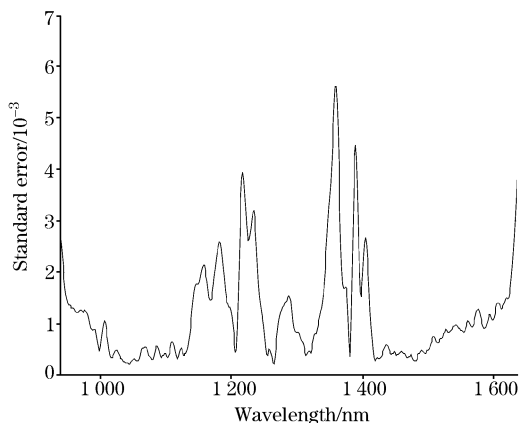
建模主成分数用交互验证法确定<sup>[19]</sup>,随着主成分增加, PRESS 逐渐减小,但主成分超过一定数值时,由于过拟合现象的出现, PRESS 反而增加。图 2 显示的是部分样品(1~20 号样本)光谱量测数据经一阶求导和 MSC 校正后,选取不同主成分数后用交互验证法得到的 PRESS 值,发现主成分数为 3 时 PRESS 最小,因而本文选取主成分数为 3 建立 SIMCA 模型。



**Fig. 2 Plot of press value vs number of principal components**

### 3.3 建模谱带的选择

尽管从理论上近红外仪器测得的所有谱段数据都可以用于建模,但不同谱段所体现的特征信息有较大差异,在实际建模过程中往往需要对光谱谱段进行筛选,以避免引入过多的冗余信息,改善模型性能。本文选用偏差权重法<sup>[19]</sup>选取谱段,即根据偏差大的变量比偏差小的变量特征更明显的原则,挑选偏差大的波长区域。图 3 所示是部分样品(1~20 号

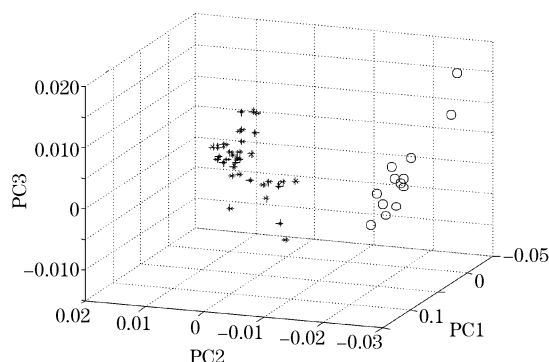


**Fig. 3 Standard deviations of the NIR spectra data of the samples**

样本)近红外光谱一阶求导后的偏差权重分布图,可知样品特征信息主要体现在 1 140~1 420 nm 范围内。

### 3.4 模型的建立

本文共分析了不同工艺条件下的 46 个梔子提取液样本,其中 1~33 号样本提取工艺一致,其余样本提取工艺各有差别。对全体样本的光谱矩阵进行主成分分析,结果见图 4(前三个主成分共解释了全部变量 86% 的方差)。从中可以看出,样本在空间中的分布反映了其提取工艺间的差别,样本 1~33 的提取工艺相同,在图中明显聚为一类,而其他工艺的样本聚为另一类。



**Fig. 4 3D plot of scores for all the samples (86% variance explained)**

\* : sample 1=33; O : sample 33-46

本文选用 1~20 号样本为训练集,并以其提取工艺作为参照工艺,21~46 号样本用作未知样本。训练集样本 1 140~1 420 nm 范围内光谱量测数据经 13 点平滑后,再经一阶微分和 MSC 校正,选取 3 个主成分数建立得到 SIMCA 工艺监测模型,用于实时评价未知样本的提取工艺是否偏离参照工艺。

### 3.5 SIMCA 模型对未知样品的工艺评价

图 5 是 SIMCA 模型预测各样本的  $F$  检验结果,其中 1~20 号训练集样本(黑色条框)用交互验证法求得,21~46 号未知样本(白色条框)用 SIMCA 模型预测得到。如设定显著性水平为 0.01(临界  $F_0$  值为 1.259 2),21~33 号样本的  $F$  检验值在临界值以下,表明其工艺与参照工艺没有显著性差异,工艺保持稳定,而 34~46 号样本的检验结果远远超出临界值,表明其提取工艺出现异常,该评价结果与实际情况完全相符。

从图 5 中还可发现 34 和 46 号样本的  $F$  检验值最大,表明其提取工艺偏离参照工艺最明显。与参照工艺相比,34~46 号未知样本的提取工艺各不相同,包括溶剂纯度、超声时间、溶剂用量、药材颗粒度、提取方式都有所差别,其中 46 号样本是在 60 °C 直接用溶剂浸提,34 号样本使用低纯度溶剂超声提取,说明在溶剂纯度和提取方式的不同对工艺的影

响最大。

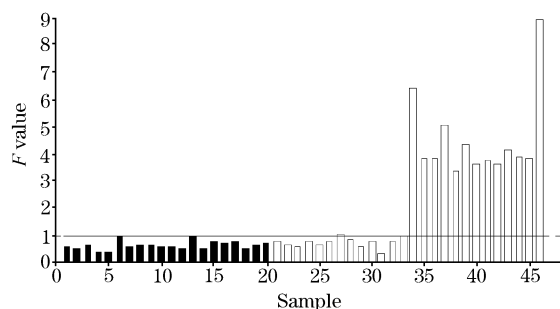


Fig. 5 Predicted results of unknown samples and training set samples by SIMCA model

Bars in black (■) denote the cross-validation predicted results of training set samples, while bars in white (□) for the predicted results of unknown samples, and the line represents  $F_0$  (99% confidence)

为了进一步考察未知样本之间的差异, 并对本模型的预测结果加以验证, 本文对 26 个未知样本进行了系统聚类分析, 以欧式距离作为类内距离测度, 采用 Ward 法计算类间距离, 所得系统聚类谱系图见图 6。由图可见, 随着提取工艺的不同, 各样本呈现出不同的类别特征, 即采用参照工艺提取所得的样本聚为一类, 其他工艺样本为另一类, 而且其中的 34 和 46 号样本较其他样本有明显的不同, 这一分析结果与 SIMCA 模型所得结论完全一致, 但 SIMCA 模型可以对任一未知样本的工艺状态进行实时评价, 且提供一个客观、量化的指标用以判别其工艺是否出现异常, 很显然更适于发

展应用成为一种有效而直观的工艺监控手段。

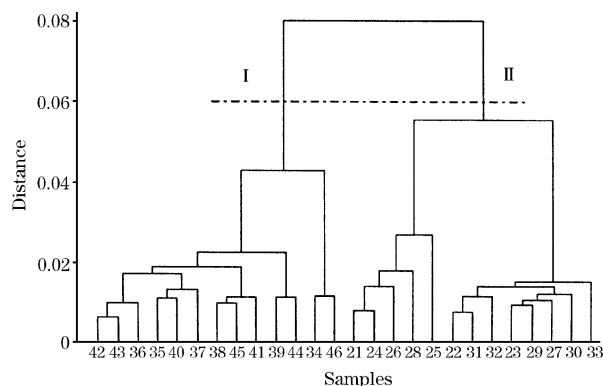


Fig. 6 Dendrogram for 26 unknown samples by Ward's clustering analysis

Numbers of 1-26 on X-axis represent the sample 21-46 respectively

## 4 结 论

本文提出一种基于近红外分析技术和 SIMCA 模型的工艺实时监控新方法。本方法无需对指标成分进行定量分析, 即可充分表达样品的工艺信息, 体现了符合中药复杂体系的研究特色。这种方法快速、准确、简便, 易于实现在线分析, 将为解决目前中药生产工艺中缺乏有效的实时控制手段的难题、保证产品质量、提高中药质量分析技术水平提供一条新的思路。

## 参 考 文 献

- [1] CAO Jin, RAO Yi, SHEN Qun, et al(曹进, 饶毅, 沈群, 等). World Science and Technology/Modernization of Traditional Chinese Medicine (世界科学技术——中药现代化), 2001, 3(4): 20.
- [2] CAO Jin, RAO Yi, SHEN Qun, et al(曹进, 饶毅, 沈群, 等). Traditional Chinese Drug Research and Clinical Pharmacology (中药新药与临床药理), 2001, 12(3): 200.
- [3] State Food and Drug Administration Bureau of China(中国国家食品药品监督管理局). Requirements for Studying Fingerprint of Traditional Chinese Medicine Injections (Draft)(中药注射剂指纹图谱研究的技术要求(暂行)), 2000.
- [4] CHEN Min-jun, QU Hai-bin, CHENG Yi-yu(陈闽军, 瞿海滨, 程翼宇). Chemical Journal of Chinese Universities(高等学校化学学报), 2003, 24: 2181.
- [5] LUO Guo-an, WANG Yi-ming, CAO Jin, et al(罗国安, 王义明, 曹进, 等). World Science and Technology/Modernization of Traditional Chinese Medicine(世界科学技术——中药现代化), 2002, 4(4): 5.
- [6] CAO Jin, WANG Yi-ming, LUO Guo-an, et al(曹进, 王义明, 罗国安, 等). World Science and Technology/Modernization of Traditional Chinese Medicine(世界科学技术——中药现代化), 2002, 4(5): 32.
- [7] YAN Yu-zhen, XIE Pei-shan, SONG Li-fei, et al(颜玉贞, 谢培山, 宋力飞, 等). Chinese Traditional Patent Medicine(中成药), 2004, 26(8): 603.
- [8] JIA Xiao-bin, CHEN Yan, CAI Bao-chang, et al(贾晓斌, 陈彦, 蔡宝昌, 等). Chinese Traditional Patent Medicine(中成药), 2003, 25(11): 866.
- [9] Macho M, Larrechi M S. Trends in Analytical Chemistry, 2002, 21(12): 799.
- [10] Lafargue M E, Feinberg M H, Daudin J J, et al. Anal. Bioanal. Chem., 2003, 375(4): 496.
- [11] McCaig T N. Food Res. Int., 2002, 35(8): 731.
- [12] Thosar S S, Forbess R A, Ebube N K. Pharm. Dev. Technol. 2001, 6(1): 19.
- [13] ZHAO Chen, QU Hai-bin, CHENG Yi-yu(赵琛, 瞿海斌, 程翼宇). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2004, 24(1): 50.

- [14] GU Xiao-yu, WANG Yan, XU Ke-xin, et al(谷筱玉, 汪燕, 徐可欣, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2004, 24(2): 155.
- [15] QU Hai-bin, LIU Xiao-xuan, CHENG Yi-yu(瞿海斌, 刘晓宣, 程翼宇). Chemical Journal of Chinese Universities(高等学校化学学报), 2004, 25(1): 39.
- [16] YANG Nan-lin, CHENG Yi-yu, QU Hai-bin(杨南林, 程翼宇, 瞿海斌). Acta Chimica Sinica(化学学报), 2003, 61(5): 742.
- [17] LIU Quan, QU Hai-bin, CHENG Yi-yu(刘全, 瞿海斌, 程翼宇). Journal of Chemical Industry and Engineering(化工学报), 2003, 54(11): 1586.
- [18] Wold S, Sjostrom M. Chemometrics; Theory and Application. Kowalski B R Ed. Washington D C: American Chemical Society, 1977. 242.
- [19] LIU Shu-shen, YI Zhong-sheng(刘树深, 易忠胜). Basic Chemometrics(基础化学计量学). Beijing: Science Press(北京: 科学出版社), 1999. 127, 145.

## Study on Real-Time Control of Extraction Procedure of Furitus Gardenia by Near Infrared Spectroscopy

YAN Shi-kai<sup>1,2</sup>, LUO Guo-an<sup>2\*</sup>, WANG Yi-ming<sup>2</sup>, CHENG Yi-yu<sup>1</sup>

1. Department of Chinese Medicine Science and Engineering, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310031, China
2. Department of Chemistry, Tsinghua University, Beijing 100084, China

**Abstract** A new approach to the real-time control of extraction procedures based on fiber optic near infrared spectroscopy (NIR) was described. Various extraction samples of gardenia were rapidly scanned by NIR, and the SIMCA monitoring model was developed. The model was applied to evaluate the extraction procedures of 26 unknown samples, and all the samples from abnormal procedures were successfully identified. It is suggested that near infrared spectroscopy, combined with SIMCA method, is a rapid, convenient and efficient tool for real-time procedure monitoring, which needs no quantitative determination and can effectively monitor whether the current procedure is in normal or not. The presented approach, therefore, offers a new way for the quality control in real-time of traditional Chinese medicine.

**Keywords** Furitus gardenia; Near infrared spectroscopy; SIMCA; Real-time control

(Received Mar. 6, 2005; accepted Jun. 16, 2005)

\* Corresponding author