

## 基于 $\beta$ -容度容忍区间的血塞通注射液近红外定量分析不确定度评估

徐冰<sup>1</sup>, 薛忠<sup>1</sup>, 罗赣<sup>1</sup>, 李建宇<sup>1</sup>, 史新元<sup>1,2\*</sup>, 乔延江<sup>1,2\*</sup>

1. 北京中医药大学 中药信息工程研究中心, 北京 100029

2. 教育部中药制药与新药开发关键技术工程研究中心, 北京 100029

**摘要:** **目的** 建立基于 $\beta$ -容度容忍区间( $\beta$ -CTI)的血塞通注射液近红外(NIR)定量分析不确定度评估方法。**方法** 采用“ $I \times J \times K$ ”型析因实验设计法获取 NIR 分析校正和验证实验数据, 通过 PLS 算法建立 NIR 定量模型, 通过 $\beta$ -CTI 估计 NIR 定量分析不确定度, 并与常规 $\beta$ -期望容忍区间( $\beta$ -ETI)进行比较。采用 $\ln U = a + b \ln Y$ 模型对相对扩展不确定度( $U$ )与相应质量浓度( $Y$ )之间的关系进行拟合, 设置不确定度可接受限( $\lambda$ ) = 20%, 并绘制不确定度决策曲线。**结果** 在考察的 7 个质量浓度水平中, 除 1.485 mg/mL 外, 由 $\beta$ -ETI 估计的不确定度均低于由 $\beta$ -CTI 估计的不确定度。由 $\beta$ -ETI 和 $\beta$ -CTI 确定的临界可接受不确定度质量浓度分别为 4.867、5.686 mg/mL。**结论**  $\beta$ -CTI 可充分估计 NIR 定量不确定度, 可用于 NIR 定量结果准确性和可靠性评价。

**关键词:** 近红外; 定量不确定度;  $\beta$ -容度容忍区间;  $\beta$ -期望容忍区间; 血塞通注射液; 三七总皂苷

**中图分类号:** R286.02 **文献标志码:** A **文章编号:** 0253-2670(2015)06-0832-08

**DOI:** 10.7501/j.issn.0253-2670.2015.06.009

## Quantitative analysis on uncertainty evaluation using near infrared spectroscopy based on $\beta$ -content tolerance interval

XU Bing<sup>1</sup>, XUE Zhong<sup>1</sup>, LUO Gan<sup>1</sup>, LI Jian-yu<sup>1</sup>, SHI Xin-yuan<sup>1,2</sup>, QIAO Yan-jiang<sup>1,2</sup>

1. Research Center of TCM Information Engineering, Beijing University of Chinese Medicine, Beijing 100029, China

2. Engineering Research Center of Chinese Medicine Production and New Drug Development, Ministry of Education, Beijing 100029, China

**Abstract: Objective** To establish the uncertainty evaluation for the quantitative analysis on Xuesaitong Injection by near infrared (NIR) spectroscopy based on  $\beta$ -content tolerance interval ( $\beta$ -CTI). **Methods** The data of NIR calibration and validation sets were obtained by “ $I \times J \times K$ ” full factorial experimental design. The NIR quantification model was established through the PLS algorithm. Based on the validation data, the  $\beta$ -CTI was introduced to estimate the quantitative analysis on uncertainty of NIR, which was compared with the regular  $\beta$ -expectation tolerance interval ( $\beta$ -ETI). **Results** Among the seven designed concentration levels, the uncertainty values estimated by the  $\beta$ -ETI are smaller than that estimated by the  $\beta$ -CTI, except for the concentration level of 1.485 mg/mL. Then, the relationship between the relative expanded uncertainty ( $U$ ) and corresponding concentration ( $Y$ ) was fitted by  $\ln U = a + b \ln Y$  model. When the uncertainty acceptance limit ( $\lambda$ ) was set to 20% and the decision profile of uncertainty was plotted, the critical concentration for  $\beta$ -ETI and  $\beta$ -CTI with the acceptable uncertainty could be read as 4.867 and 5.686 mg/mL, respectively. **Conclusion** This study has proved that  $\beta$ -CTI can be used to adequately estimate the uncertainty of quantitative analysis by NIR, which provides a new potential tool for the assessment of the accuracy and reliability of NIR spectroscopy.

**Key words:** near infrared; quantitative uncertainty;  $\beta$ -content tolerance interval;  $\beta$ -expectation tolerance interval; Xuesaitong Injection; total saponins of *Panax notoginseng*

近红外 (near infrared, NIR) 光谱因具有实时快速、无损、信息丰富等特点, 在制药、农业、化工等领域应用广泛<sup>[1-3]</sup>。NIR 定量分析采用间接校正的方法, 即建立反映 NIR 光谱和参考值之间关系的

收稿日期: 2014-11-03

基金项目: 国家自然科学基金资助项目 (81403112); 北京市重点实验室建设项目 (中药基础与新药研究实验室, 2014)

作者简介: 徐冰 (1985—), 男, 讲师, 研究方向为中药质量与过程控制。Tel: (010)84738650 E-mail: btcm@163.com

\*通信作者 乔延江, 男, 博士, 教授, 博士生导师, 主要从事中药信息学研究。E-mail: yjqiao@263.net

史新元, 女, 博士, 教授, 主要从事中药过程分析研究。E-mail: shixinyuan01@163.com

多变量校正模型,通过模型实现未知样品定量。建模过程包括光谱预处理、界外样本识别、特征变量筛选、模型参数优化和验证等多个步骤,如何评价所建模型的准确性和可靠性已成为NIR光谱法能否有效推广的关键。

目前,用于评价NIR定量模型性能的指标一般选择化学计量学参数,如反映模型校正性能的校正标准偏差(standard error of calibration, SEC)、交叉验证标准偏差(standard error of cross calibration, SECV)和相关系数( $r$ )等,反映模型预测性能的预测标准偏差(standard error of prediction, SEP)和相对预测偏差(relative prediction deviation, RPD)等<sup>[4]</sup>。然而,Zhang等<sup>[5]</sup>指出SEC和SEP等常规参数仅代表模型校正和验证结果的平均水平,无法反映模型在不同浓度水平的定量准确性和在实际应用中的预测能力。

2004年,Feinberg等<sup>[6]</sup>率先建立了 $\beta$ -期望容忍区间( $\beta$ -expectation tolerance interval,  $\beta$ -ETI)和不确定度之间关系,为基于验证数据的不确定度测算提供了便捷途径。随后, $\beta$ -ETI被应用于制剂水分、活性成分NIR定量分析的不确定度评估<sup>[7-9]</sup>。测量不确定度将NIR定量分析过程中的系统误差和随机误差集成<sup>[10]</sup>,其相对值形式可在同一尺度评价不同浓度水平下的定量可靠程度。然而, $\beta$ -ETI中仅包含了分析方法的真实性和精密度信息,因此采用 $\beta$ -ETI估计的不确定度值可能偏低<sup>[11]</sup>。针对此局限性,本实验拟采用 $\beta$ -容度容忍区间( $\beta$ -content tolerance interval,  $\beta$ -CTI)对NIR定量分析方法的不确定度进行充分估计,并比较 $\beta$ -ETI和 $\beta$ -CTI 2类容忍区间在NIR定量分析不确定度评估中的效果。

血塞通注射液主要含有三七总皂苷,具有活血祛瘀、通脉活络之功效,临床应用广泛。杨南林等<sup>[12]</sup>曾建立了三七总皂苷定量的NIR快速测定方法,但并没有进行方法可靠性评价。本实验拟以血塞通注射液中三七总皂苷定量的NIR分析为载体,将不确定度评估方法引入中药NIR定量分析,以期对中药NIR定量分析方法可靠性评价提供参考。

## 1 原理和方法

### 1.1 定量分析不确定度表示

根据LGC/VAM协议<sup>[13]</sup>和ISO/DTS 21748指南<sup>[14]</sup>,待测样品在某一浓度( $C$ )水平下,测量值 $Y$ 的不确定度评估模型可用下式表示:

$$u^2(Y) = S_R^2 + u^2(\delta) + \sum c_i^2 u^2(x_i) \quad (1)$$

式(1)中, $S_R$ 为方法中间精密度标准偏差, $u(\delta)$ 表示与方法偏差有关的不确定度, $\sum c_i^2 u^2(x_i)$ 表示其他影响因素( $x_i$ )造成的分析方法不确定度总和, $c_i$ 为参数灵敏度系数。SFSTP (Société Française des Sciences et Techniques Pharmaceutiques Commission)在采用基于总误差理论和验证实验设计的Accuracy Profile估计分析方法不确定度时,将公式(1)等号右边第3项忽略<sup>[15-16]</sup>,从而将不确定度评估模型简化为:

$$u^2(Y) = S_R^2 + u^2(\delta) \quad (2)$$

基于公式(2),Feinberg等<sup>[6]</sup>建立了容忍区间 $[L, H]$ 和不确定度 $u(Y)$ 之间的关系:

$$[L, H] = [\bar{Y} - Q_t u(Y), \bar{Y} + Q_t u(Y)] \quad (3)$$

式(3)中 $L$ 为容忍区间下限, $H$ 为容忍区间上限, $\bar{Y}$ 为某分析浓度 $C$ 的估计均值。 $Q_t$ 为当自由度为 $\nu$ 时Student's  $t$ 分布分位数。根据公式(3),在浓度水平 $C$ 时,定量分析不确定度可由下式计算。

$$u(Y) = (H - L) / 2Q_t \quad (4)$$

可见,容忍区间 $[L, H]$ 的计算是评估分析方法不确定度的关键, $\beta$ -ETI采用公式(5)计算。

$$P\{L < Y < H | \bar{Y}, \hat{\sigma}_M\} \geq \beta \quad (5)$$

式(5)表示在分析浓度 $C$ 时,给定测量值均值 $\bar{Y}$ 和中间精密度 $\hat{\sigma}_M$ ,测量值 $Y$ 落在区间 $[L, H]$ 内的概率为 $\beta$ ,测量值均值 $\bar{Y}$ 和中间精密度 $\hat{\sigma}_M$ 通过实验设计的方法设计验证实验以获取数据并估计。由于忽略公式(1)第3项,采用 $\beta$ -ETI估计的不确定度值偏低,因此本实验拟采用 $\beta$ -CTI对公式(5)进行校正:

$$P\{P_Y[L < Y < H | \bar{Y}, \hat{\sigma}_M] \geq \beta\} = \gamma \quad (6)$$

式(6)表示测量值 $Y$ 以一定的置信水平 $\gamma$ 落在区间 $[L, H]$ 内的概率为 $\beta$ ,测量值均值 $\bar{Y}$ 和中间精密度 $\hat{\sigma}_M$ 的计算同公式(5)。分析方法的扩展不确定度 $U(Y)$ 可由公式(7)计算:

$$U(Y) = k_p u(Y) \quad (7)$$

式(7)中, $k_p$ 为覆盖因子, $p$ 为置信水平,假定测量结果服从正态分布,则 $k_p$ 表示正态分布的 $1-p$ 分位数。扩展不确定度表示测量值平均水平落在 $U(Y)$ 内的概率为 $1-p$ 。

### 1.2 定量分析验证实验设计

在定量分析验证实验中,应尽可能考虑定量分析方法在实际应用时可能遇到的变异来源,并采用实验设计模拟方法的重复性方差和重现性方差。如经典的“ $I \times J \times K$ ”型析因实验设计,即在定量分析

验证实验中考虑 3 方面的影响：条件 ( $I$ )、重复次数 ( $J$ ) 和浓度水平 ( $K$ )<sup>[17]</sup>。 $K$  的设置决定定量分析浓度范围，不同  $K$  下的定量分析不确定度不同； $J$  用以评价定量分析方法的重复性； $I$  表示不同的测量时间、操作人员等影响因素。采用容忍区间估计分析方法不确定度时，通常将  $I$  设置为不同的测量时间，例如天数，以估计公式 (5) 和 (6) 中的中间精密度  $\hat{\sigma}_M$ <sup>[18]</sup>。

在某一浓度水平  $C$ ，对于“ $I \times J$ ”型验证实验数据，采用方差分析法 (ANOVA) 估计组间均方 ( $MS_B$ ) 和组内均方 ( $MS_E$ )。

若  $MS_E < MS_B$ ，则：

$$\hat{\sigma}_B^2 = \frac{MS_B - MS_E}{n} \quad (8)$$

$$\hat{\sigma}_E^2 = MS_E \quad (9)$$

若  $MS_E > MS_B$ ，则

$$\hat{\sigma}_B^2 = 0 \quad (10)$$

$$\hat{\sigma}_E^2 = \frac{1}{mn-1} \sum_{i=1}^m \sum_{j=1}^n (Y_{ij} - \bar{Y})^2 \quad (11)$$

式 (11) 中， $Y_{ij}$  表示在第  $i$  ( $i=1, 2, \dots, m$ ) 个条件下，浓度水平  $C$  的第  $j$  ( $j=1, 2, \dots, n$ ) 个测量值； $\bar{Y}$  的计算方法为：

$$\bar{Y} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n Y_{ij} \quad (12)$$

因此在浓度水平  $C$ ，可分别定义分析方法的重复性方差 ( $\sigma_{Re}^2$ ) 和中间精密度方差 ( $\sigma_M^2$ ) 如下：

$$\hat{\sigma}_{Re}^2 = \hat{\sigma}_E^2 \quad (13)$$

$$\hat{\sigma}_M^2 = S_R^2 = \hat{\sigma}_B^2 + \hat{\sigma}_E^2 \quad (14)$$

### 1.3 容忍区间

$\beta$ -ETI 和  $\beta$ -CTI 的估计为计算 NIR 不确定度的核心，以下简述其计算过程。

#### 1.3.1 $\beta$ -ETI $\beta$ -ETI 可用下式表示<sup>[19]</sup>：

$$[L, U] = [\bar{Y} - k_E \hat{\sigma}_M, \bar{Y} + k_E \hat{\sigma}_M] \quad (15)$$

式 (15) 中，

$$k_E = Q_t k_s = Q_{t(v, (1+\beta)/2)} \sqrt{1 + \frac{1}{mnB^2}} \quad (16)$$

式 (16) 中，通过  $k_E$  的计算可以一定概率获得测量值的期望分布范围，即容忍区间； $Q_t$  为当自由度为  $v$  时 Student's  $t$  分布的  $(1+\beta)/2$  分位数，即  $Q_t(v, \frac{1+\beta}{2})$ 。 $Q_t$  考虑了不确定度估计中的随机因素影响，而  $k_s$  项则考虑了不确定度估计中的系统因素的影响。

式 (16) 中，

$$B = \sqrt{\frac{R+1}{nR+1}} \quad (17)$$

$$R = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_E^2} \quad (18)$$

$$v = \frac{(R+1)^2}{(R+(1/n))^2 / (m-1) + (1-(1/n)) / mn} \quad (19)$$

#### 1.3.2 $\beta$ -CTI $\beta$ -CTI 又称为 $\beta$ -容度- $\gamma$ 置信容忍区间可用下式表示：

$$[L, U] = [\bar{Y} - k_C \hat{\sigma}_M, \bar{Y} + k_C \hat{\sigma}_M] \quad (20)$$

其中，

$$k_C = \sqrt{\frac{v' \chi_{v', \beta}^2(\tau)}{\chi_{v', 1-\gamma}^2}} \quad (21)$$

式 (21) 中， $\chi_{v', \beta}^2(\tau)$  指自由度 1 时的非中心卡方分布的  $\beta$  分位数， $\tau$  为非中心参数； $\chi_{v', 1-\gamma}^2$  表示自由度  $v'$  下的卡方分布的  $1-\gamma$  分位数；自由度  $v'$  的计算方法：

$$v' = \frac{(R'+1)^2}{(R'+(1/n))^2 / (m-1) + (1-(1/n)) / mn} \quad (22)$$

式 (22) 中，

$$R' = \max \left[ 0, \frac{1}{n} \left( \frac{F}{F_\eta} - 1 \right) \right] \quad (23)$$

其中， $F_\eta$  为自由度  $v_1=m-1$  和  $v_2=m(n-1)$  下的  $F$  分布的  $1-\eta$  分位数，即  $F_{(v_1, v_2, 1-\eta)}$ ；当  $\beta$ -CTI 的置信水平  $\gamma$  分别取 0.90、0.95 和 0.99 时，对应的  $\eta$  值分别为 0.85、0.905 和 0.976<sup>[20]</sup>； $F$  为  $MS_B$  和  $MS_E$  的比值。

式 (21) 中的非中心参数  $\tau$  的计算方法：

$$\tau = \frac{1}{mnB'} \quad (24)$$

$$B' = \frac{R'+1}{nR'+1} \quad (25)$$

采用  $\beta$ -CTI 估计不确定度时， $Q_t$  为当自由度为  $f$  时 Student's  $t$  分布的  $(1+\gamma)/2$  分位数，即  $Q_t(f, \frac{1+\gamma}{2})$ ，其中自由度  $f$  的计算方法<sup>[21]</sup>：

$$f = \frac{(\hat{R}+1)^2}{(\hat{R}+(1/n))^2 / (m-1) + (1-(1/n)) / mn} \quad (26)$$

$$\hat{R} = \max \left[ 0, \frac{1}{n} (F-1) \right] \quad (27)$$

## 2 实验部分

### 2.1 仪器与试剂

Antaris FT-NIR 光谱仪，美国 Thermo Nicolet 公司；Agilent 1100 HPLC 仪，美国 Agilent 公司，包括：四元泵、真空脱气泵、自动进样器、柱温箱、

DAD 检测器、ChemStation 数据处理工作站。

三七皂苷  $R_1$  (批号 110745-201318)、人参皂苷  $R_{g1}$  (批号 110703-201128)、人参皂苷  $R_e$  (批号 110754-201324)、人参皂苷  $R_{b1}$  (批号 110704-201223) 和人参皂苷  $R_d$  (批号 111818-201302) 对照品购自中国食品药品检定研究院。血塞通注射液, 批号 12EL03, 昆明制药集团股份有限公司; 乙腈, 色谱纯, 美国 Fisher 公司; 哇哈哈纯净水。

## 2.2 实验设计

根据“1.2”项下“ $I \times J \times K$ ”型实验设计要求, 选取“ $3 \times 6 \times 7$ ”型实验设计方案, 即实验分 3 d 进行, 每天 1 次, 每次实验包含 7 个浓度水平, 每个浓度水平进行 6 次平行重复实验, 将获得的 126 个样本采用 Kennard-stone (K-S) 算法划分为校正集 100 个和内部验证集 26 个, 校正集样本用于建立 NIR 定量分析模型。另选取“ $3 \times 3 \times 7$ ”型实验设计方案, 即实验分 3 d 进行, 每天 1 次, 每次实验包含 7 个浓度水平, 每个浓度水平进行 3 次平行重复实验, 获得的 63 个样本作为外部验证集并用于不确定度分析。7 个浓度水平包括血塞通注射液原液和 2.5、5、12.5、25、50、250 倍水稀释液。

NIR 定量分析目标为血塞通注射液三七总皂苷质量浓度, 以三七皂苷  $R_1$ 、人参皂苷  $R_{g1}$ 、人参皂苷  $R_e$ 、人参皂苷  $R_{b1}$  和人参皂苷  $R_d$  质量浓度之和表示。本实验设计未考虑血塞通注射液批次变化等信息, 为模拟方法开发初期的 NIR 定量模型不确定度评估。

## 2.3 NIR 分析

将血塞通样品盛装于内径 8 mm 的具塞石英管, 采用 NIR 透射模式进行分析, 光谱范围 10 000~4 000  $\text{cm}^{-1}$ , 分辨率 4  $\text{cm}^{-1}$ , 每个样品以空气为参比扫描 16 次后取平均光谱。

## 2.4 HPLC 分析

参考《中国药典》2010 年版一部三七总皂苷项下定量测定法, 采用 Agilent Zorbax  $C_{18}$  色谱柱 (250 mm $\times$ 4.6 mm, 5  $\mu\text{m}$ ); 流动相为乙腈-水, 梯度洗脱: 0~20 min, 20%乙腈; 20~45 min, 20%~46%乙腈; 45~55 min, 46%~55%乙腈; 55~60 min, 55%乙腈; 检测波长 203 nm, 体积流量 1.5 mL/min; 柱温 25  $^{\circ}\text{C}$ 。样品加甲醇适当稀释后, 过 0.45  $\mu\text{m}$  微孔滤膜, 进样 10  $\mu\text{L}$ 。

## 2.5 统计软件

本实验光谱预处理使用化学计量学软件

SIMCA P+11.5 (Umetrics, 瑞典), SiPLS 算法采用 iToolbox<sup>[22]</sup>在 Matlab 7.0 (Mathworks 公司) 平台实现, PLS 算法使用 PLS Toolbox 2.1 (Eigenvector Research Inc., USA) 在 Matlab 平台上实现, 不确定度曲线拟合采用 Matlab 曲线拟合工具箱, 其他程序自行编制。

## 3 结果和讨论

### 3.1 HPLC 分析结果

配制不同质量浓度的 5 种皂苷对照品溶液, 进样测定峰面积, 以峰面积为纵坐标 ( $Y$ ), 对照品质量浓度为横坐标 ( $X$ ), 绘制标准曲线, 结果见表 1。HPLC 定量分析结果显示校正集样品的三七总皂苷质量浓度分布在 0.070 4~19.85 mg/mL。内部验证集和外部验证集的三七总皂苷质量浓度分别为 0.070 4~18.23 mg/mL 和 0.070 4~19.85 mg/mL。结果表明校正集样品的三七总皂苷质量浓度具有较宽的分布范围, 验证集三七总皂苷质量浓度均在校正集范围之内。

表 1 血塞通注射液中 5 种皂苷 HPLC 定量标准曲线和线性范围 ( $n=6$ )

Table 1 HPLC quantitative calibration curves and ranges for five saponins in Xuesaitong Injection ( $n=6$ )

| 成分            | 标准曲线              | $r$    | 线性范围/<br>( $\mu\text{g}\cdot\text{mL}^{-1}$ ) |
|---------------|-------------------|--------|-----------------------------------------------|
| 三七皂苷 $R_1$    | $Y=43.25X+0.0503$ | 0.9999 | 40.5~486.5                                    |
| 人参皂苷 $R_{g1}$ | $Y=39.35X+1.7747$ | 0.9997 | 181.5~2177.6                                  |
| 人参皂苷 $R_e$    | $Y=39.30X-0.0463$ | 0.9997 | 22.7~272.9                                    |
| 人参皂苷 $R_{b1}$ | $Y=28.68X+0.0455$ | 0.9999 | 179.6~2155.8                                  |
| 人参皂苷 $R_d$    | $Y=44.79X+0.0695$ | 0.9999 | 25.5~306.3                                    |

### 3.2 NIR 光谱预处理

对校正集和验证集中的原始 NIR 光谱 (图 1), 分别采用 SG 平滑 (Savitzky-Golay smoothing)、一阶导数 (first derivative, 1std)、多元散射校正 (multiplicative scatter correction, MSC)、标准正态变换 (standard normal variate, SNV)、小波去噪 (wavelet denoising, WDS) 和基线校正 (baseline) 进行预处理, 建立预处理后的样品 NIR 光谱和相应 HPLC 参考值之间的偏最小二乘 (PLS) 回归模型, 结果见表 2。可见, 采用原始光谱建立的 PLS 模型的校正和预测性能良好, 采用 WDS 后的光谱建立的 PLS 模型性能与原始光谱相当, 因此本研究中采用原始 NIR 光谱建立模型。

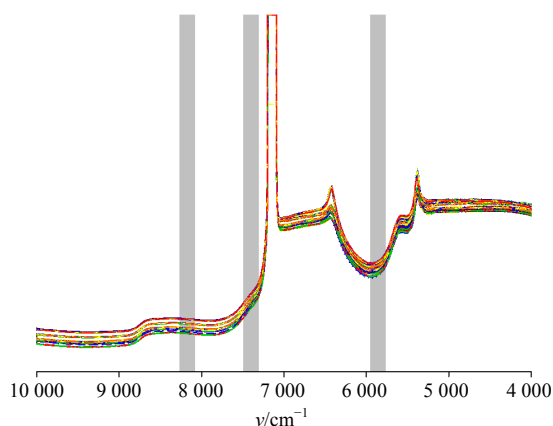


图 1 原始 NIR 光谱及 SiPLS 算法特征波段筛选结果  
Fig. 1 Original NIR spectra and results of featured wavebands selected by SiPLS algorithm

### 3.3 NIR 定量模型建立

采用组合间隔偏最小二乘 (SiPLS) 算法进行最佳建模波段筛选, 在 SiPLS 算法中, 设置波段组合数为 3, 波间隔数在 5~50 内优化, 结果当波段间隔数为 31 时, 波段 8 265.4~8 076.4、7 494.0~7 305.0、5 951.2~5 762.3  $\text{cm}^{-1}$  组合 (图 1)

建立的 PLS 模型具有最小的交叉验证误差均方根 (RMSECV) 值, 因此选择此 3 段波长建立血塞通三七总皂苷定量分析模型。最优潜变量因子的数目由留一交叉验证法 (leave one out cross validation, LOO) 获得。由图 2-A 可以发现, 校正误差均方根 (RMSEC)、RMSECV、预测误差均方根 (RMSEP) 和预测残差平方和 (PRESS) 在 6 个潜变量因子时趋于稳定, 因此选择 6 个潜变量因子建立校正模型。6 个潜变量因子下的校正集和内部验证集相关关系见图 2-B。

### 3.4 NIR 定量不确定度评估

采用  $\beta$ -CTI 计算不确定度时, Hoffman 等<sup>[23]</sup>将 FDA 生物样品分析方法验证的“4-6- $\lambda$ ”原则转换为  $\beta=0.667$ ,  $\gamma=0.90$ <sup>[24]</sup>, 本实验亦采用此设置。采用  $\beta$ -ETI 计算不确定度时, 令  $\beta=0.90$ <sup>[25]</sup>。采用建立的 NIR 定量模型对 63 个外部验证集样本进行预测, 并采用式 (28) 计算每一样本的相对预测偏差:

$$Y_{\text{bias}, ij, k} = \frac{\hat{Y}_{ij, k} - Y_{ij, k}}{Y_{ij, k}} \quad (28)$$

表 2 不同 NIR 光谱预处理方法的比较

Table 2 Comparison on different preprocessing methods for various NIR spectra

| 预处理方法    | 潜变量 | 校正集              |       |        |                            | 验证集              |       |      |                            |
|----------|-----|------------------|-------|--------|----------------------------|------------------|-------|------|----------------------------|
|          |     | $r_{\text{cal}}$ | RMSEC | RMSECV | $\text{BIAS}_{\text{cal}}$ | $r_{\text{val}}$ | RMSEP | RPD  | $\text{BIAS}_{\text{val}}$ |
| 原始       | 11  | 0.987 3          | 1.006 | 1.395  | 0.806                      | 0.989 7          | 0.800 | 7.07 | 0.603                      |
| SG 平滑    | 11  | 0.990 5          | 0.851 | 1.144  | 0.665                      | 0.986 6          | 1.022 | 6.14 | 0.804                      |
| 1 阶导数    | 14  | 0.993 1          | 0.740 | 3.605  | 0.583                      | 0.910 6          | 2.410 | 2.34 | 2.041                      |
| MSC      | 12  | 0.991 8          | 0.808 | 1.240  | 0.662                      | 0.988 0          | 0.881 | 6.40 | 0.656                      |
| SNV      | 12  | 0.991 9          | 0.805 | 1.240  | 0.660                      | 0.988 1          | 0.879 | 6.41 | 0.653                      |
| WDS      | 11  | 0.988 9          | 0.939 | 1.152  | 0.689                      | 0.991 1          | 0.798 | 7.06 | 0.610                      |
| baseline | 12  | 0.993 4          | 0.725 | 1.202  | 0.598                      | 0.988 9          | 0.885 | 6.37 | 0.697                      |

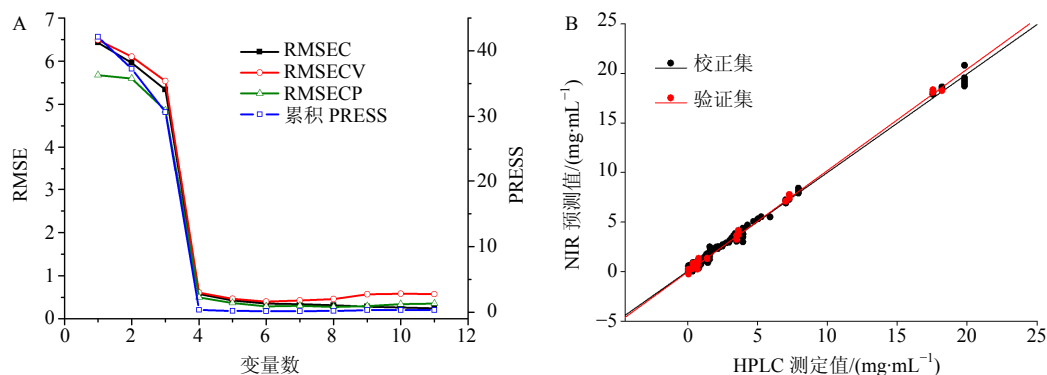


图 2 潜变量因子数优选 (A) 及三七总皂苷校正集和验证集相关系数图 (B)

Fig. 2 Selective preference for number of latent factors (A), and correlation coefficient diagram for calibration and validation sets of saponins in *Panax notoginseng* (B)

采用相对预测偏差值计算  $\beta$ -ETI 和  $\beta$ -CTI, 以利于在同一尺度下, 对各质量浓度水平下不确定度进行比较。以稀释 2.5 倍质量浓度水平 (7.423 mg/mL) 为例, 9 个样本的预测质量浓度和相对预测偏差见表 3。其  $\beta$ -ETI 的计算过程如下: 由实验设计可知,  $m=3$ ,  $n=3$ ; 由表 3 相对预测偏差结果, 由方差分析得  $MS_B=31.108\%$ ,  $MS_E=9.570\%$ ; 因  $MS_B>MS_E$ , 所以根据公式 (8) 和 (9) 可得  $\hat{\sigma}_B^2=7.179\%$ ,  $\hat{\sigma}_E^2=9.570\%$ ; 根据公式 (18),  $R=0.750\ 2$ ; 根据公式 (17),  $B=0.733\ 8$ ; 根据公式 (19), 自由度  $\nu=4.633\ 5$ ; 当  $\beta=0.90$  时,  $Q_t(4.6335, \frac{1+0.90}{2})=2.050\ 8$ , 根据公式 (16),  $k_E=2.252\ 5$ ; 根据公式 (14),  $\hat{\sigma}_M^2=\hat{\sigma}_B^2+\hat{\sigma}_E^2=16.749\ 6\%$ , 再根据公式 (15), 计算得  $[L, U]=[-9.900\%, 8.538\%]$ 。

稀释 2.5 倍质量浓度水平的  $\beta$ -CTI 的计算过程: 参数  $MS_B$ 、 $MS_E$ 、 $\hat{\sigma}_B^2$ 、 $\hat{\sigma}_E^2$  和  $\hat{\sigma}_M^2$  的计算同  $\beta$ -ETI; 当  $\gamma=0.90$  时,  $\eta=0.85$ , 自由度  $\nu_1=m-1=2$ , 和  $\nu_2=m(n-1)=6$ ,  $F_\eta=F_{(2,6,1-0.85)}=0.167\ 0$ ,  $F=MS_B/MS_E=3.250\ 6$ ; 根据公式 (23),  $R'=6.154\ 8$ ; 根据公式 (22),  $\nu'=2.423\ 6$ ; 由公式 (25),  $B'=0.367\ 6$ , 根据公式 (24) 非中心参数  $\tau=0.302\ 3$ ; 当  $\beta=0.667$ ,  $\gamma=0.90$  时,  $\chi_{1,\beta}^2(\tau)=\chi_{1,0.667}^2(0.302\ 3)=1.248\ 0$ ,  $\chi_{\nu',1-\gamma}^2=\chi_{5.4130,1-0.90}^2=0.362\ 4$ ; 根据公式 (21),  $k_C=2.929\ 7$ ; 根据公式 (20), 得  $\beta$ -CTI 为  $[L, U]=$

$[-12.671\%, 11.309\%]$ 。

仍以稀释 2.5 倍质量浓度水平为例, 根据公式 (4), 以  $\beta$ -ETI 为基础, NIR 定量分析相对不确定度:

$$u(Y)_{k=2}=(H-L)/2Q_t=[8.538\%-(-9.900\%)]/(2\times 2.050\ 8)=4.495\%$$

采用  $\beta$ -CTI 计算不确定度时, 根据公式 (27), 计算  $\hat{R}=0.750\ 2$ , 根据公式 (26), 计算  $f=4.633\ 5$ , 则  $Q_t(f, \frac{1+\gamma}{2})=Q_t(4.6335, \frac{1+0.90}{2})=2.050\ 8$ , NIR 定量分析相对不确定度为:

$$u(Y)_{k=2}=(H-L)/2Q_t=[11.309\%-(-12.671\%)]/(2\times 2.050\ 8)=5.846\%$$

按照与稀释 2.5 倍质量浓度水平相似的方法, 计算其他质量浓度水平的容忍区间和不确定度, 结果见表 4。由表 4 可见, 随着稀释倍数的增加, 不确定度值呈逐渐增大趋势。除稀释 12.5 倍质量浓度水平 (1.485 mg/mL) 外, 由  $\beta$ -CTI 计算的不确定度值均大于由  $\beta$ -ETI 计算的不确定度值, 表明采用  $\beta$ -ETI 估计的分析结果不确定度可能偏低, 与 Saffaj 等<sup>[20]</sup>报道一致。NIR 定量分析相对扩展不确定度可由公式 (7) 计算, 令  $p=0.95$ , 则  $k_p=2$ , 相对扩展不确定度  $U(Y)=2\times u(Y)$ 。

### 3.5 不确定度决策曲线

在实际应用中, 为估计测量范围内任意质量浓度水平下的不确定度, 通常建立不确定度和质量浓

表 3 NIR 预测质量浓度和相对预测偏差

Table 3 Predicted concentration by NIR and relative deviation of prediction

| 质量浓度水平/<br>(mg·mL <sup>-1</sup> ) | 实际质量浓度/<br>(mg·mL <sup>-1</sup> ) | t/d | 预测质量浓度/(mg·mL <sup>-1</sup> ) |       |       | 相对预测偏差/% |       |       |
|-----------------------------------|-----------------------------------|-----|-------------------------------|-------|-------|----------|-------|-------|
|                                   |                                   |     | 重复 1                          | 重复 2  | 重复 3  | 重复 1     | 重复 2  | 重复 3  |
| 7.423                             | 7.940                             | 1   | 7.938                         | 7.563 | 7.486 | -0.02    | -4.75 | -5.72 |
|                                   | 7.292                             | 2   | 7.355                         | 7.446 | 7.694 | 0.87     | 2.11  | 5.51  |
|                                   | 7.036                             | 3   | 6.825                         | 6.756 | 7.236 | -3.00    | -3.98 | 2.85  |

表 4 不同浓度水平不确定度评估

Table 4 Uncertainty evaluation for different concentration in validation design

| 质量浓度水平/<br>(mg·mL <sup>-1</sup> ) | 容忍区间/%              |                       | 不确定度 $u(Y)/\%$ |              | 扩展不确定度 $U(Y)/\%$ |              |
|-----------------------------------|---------------------|-----------------------|----------------|--------------|------------------|--------------|
|                                   | $\beta$ -ETI        | $\beta$ -CTI          | $\beta$ -ETI   | $\beta$ -CTI | $\beta$ -ETI     | $\beta$ -CTI |
| 18.56                             | [-6.794, 7.405]     | [-7.720, 8.151]       | 2.661          | 2.938        | 5.323            | 5.876        |
| 7.423                             | [-9.900, 8.538]     | [-12.671, 11.309]     | 4.495          | 5.846        | 8.990            | 11.692       |
| 3.711                             | [-14.045, 19.406]   | [-19.016, 24.377]     | 8.428          | 10.933       | 16.856           | 21.866       |
| 1.485                             | [-31.351, 36.020]   | [-28.649, 33.318]     | 18.028         | 16.582       | 36.056           | 33.163       |
| 0.742 3                           | [-80.677, 103.102]  | [-106.232, 128.657]   | 42.270         | 54.025       | 84.540           | 108.051      |
| 0.371 1                           | [-203.729, 192.046] | [-249.039, 237.356]   | 104.591        | 128.539      | 209.181          | 257.077      |
| 0.074 23                          | [-472.782, 990.363] | [-693.263, 1 210.844] | 359.941        | 468.421      | 719.883          | 936.841      |

度关系曲线, 本实验采用式(29)对相对扩展不确定度和质量浓度之间的相关关系进行拟合<sup>[26]</sup>。

$$\ln(U) = a + b \ln(Y) \quad (29)$$

式(29)中,  $U$  表示质量浓度  $Y$  对应的相对扩展不确定度。根据表3中由  $\beta$ -ETI 估计的相对扩展不确定度结果, 采用 Levenberg-Marquardt 算法对式(29)中的参数进行估计, 结果最佳拟合曲线为  $\ln U = 4.353 - 0.8577 \ln Y$  ( $r = 0.9981$ )。采用表3中  $\beta$ -CTI 估计的相对扩展不确定度结果, 最佳拟合曲线为  $\ln U = 4.538 - 0.8874 \ln Y$  ( $r = 0.9983$ )。在  $0.07423 \sim 18.56 \text{ mg/mL}$  内, 设定 NIR 分析方法的不确定度可接受限 ( $\lambda$ ) = 20%, 则可建立不确定度决策曲线, 如图3所示。在  $\lambda$  与不确定度曲线的交点对应的质量浓度以上, 表明分析方法的定量性能可以接受, 该交点代表定量分析方法有效与否的临界质量浓度。经计算,  $\beta$ -ETI 不确定度曲线与  $\lambda$  的交点为  $4.867 \text{ mg/mL}$ , 而  $\beta$ -CTI 不确定度曲线与  $\lambda$  的交点为  $5.686 \text{ mg/mL}$ , 说明由  $\beta$ -ETI 估计的临界质量浓度水平低于  $\beta$ -CTI。

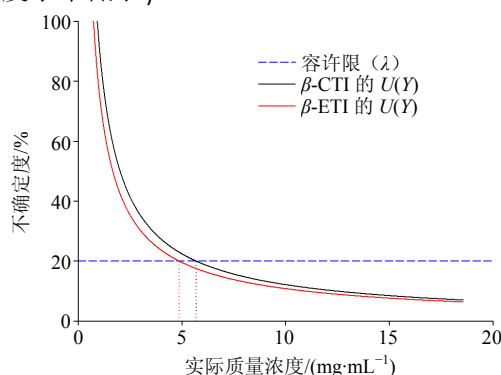


图3 NIR 定量不确定度决策曲线

Fig. 3 Uncertainty decision curve for NIR quantitation

#### 4 结论

NIR 定量分析方法的准确性在很大程度上依赖于所建多变量校正模型, 本实验将验证实验设计和  $\beta$ -CTI 引入 NIR 定量分析不确定度评估, 为 NIR 定量分析提供了可靠性评价方法及指标。血塞通注射液中三七总皂苷 NIR 定量分析不确定度评估结果显示, 与常规的  $\beta$ -ETI 相比,  $\beta$ -CTI 可实现 NIR 定量不确定度的充分估计。通过建立不确定度决策曲线, 可确定 NIR 定量分析可接受不确定度的临界质量浓度水平, 为 NIR 定量性能的提高和实际应用提供指导。然而, NIR 定量分析不确定度  $\lambda$  尚未有统一标准, 不确定度  $\lambda$  与  $\beta$ -CTI 参数 ( $\beta$ ,  $\gamma$ ) 之间的关系尚需进一步深入研究。

#### 参考文献

- [1] 汪小莉, 李 嫒, 秦昆明, 等. 近红外光谱学与化学计量学在中成药液体制剂过程分析中的应用 [J]. 中草药, 2013, 44(15): 2165-2171.
- [2] Sileoni V, Marconi O, Perretti G, *et al.* Evaluation of different validation strategies and long term effects in NIR calibration models [J]. *Food Chem*, 2013, 141(3): 2639-2648.
- [3] Kegel L, Boyne D, Booksh K. Sensing with prism-based near-infrared surface plasmon resonance spectroscopy on nanohole array platforms [J]. *Anal Chem*, 2014, 86(7): 355-3364.
- [4] Bellon-Maurel V, Fernandez-Ahumada E, Palagos B, *et al.* Critical review of chemometric indicators commonly used for the quality of the prediction of soil attributes by NIR spectroscopy [J]. *TrAC-Trends Anal Chem*, 2010, 29(9): 1073-1081.
- [5] Zhang L, Garcia-Munoz S. A comparison of different methods to estimate prediction uncertainty using Partial Least Squares (PLS): A practitioner's perspective [J]. *Chemometr Intell Lab Syst Lab Inf Manage*, 2009, 97(2): 152-158.
- [6] Feinberg M, Boulanger B, Dewè W, *et al.* New advances in method validation and measurement uncertainty aimed at improving the quality of chemical data [J]. *Anal Bioanal Chem*, 2004, 380(3): 502-514.
- [7] Mantanus J, Ziémons E, Lebrun P, *et al.* Moisture content determination of pharmaceutical pellets by near infrared spectroscopy: Method development and validation [J]. *Anal Chim Acta*, 2009, 642(1/2): 186-192.
- [8] Wu Z, Xu B, Du M, Sui C, *et al.* Validation of a NIR quantification method for the determination of chlorogenic acid in *Lonicera japonica* solution in ethanol precipitation process [J]. *J Pharmaceut Biomed*, 2012, 62: 1-6.
- [9] Schaefer C, Clicq D, Lecomte C, *et al.* Process Analytical Technology (PAT) approach to control a new API manufacturing process: Development, validation and implementation [J]. *Talanta*, 2014, 120: 114-125.
- [10] Rozet E, Rudaz S, Marini R, *et al.* Models to estimate overall analytical measurements uncertainty: Assumptions, comparisons and applications [J]. *Anal Chim Acta*, 2011, 702(2): 160-171.
- [11] González A, Herrador M. Accuracy profiles from uncertainty measurements [J]. *Talanta*, 2006, 70(4): 896-901.
- [12] 杨南林, 瞿海斌, 程翼宇. 近红外光谱法快速测定三七总皂苷的方法研究 [J]. 浙江大学学报: 工学版, 2002, 36(4): 463-466.

- [13] Barwick V, Ellison L. VAM Project 3.2.1 Development and Harmonization of Measurement Uncertainty Principles. Part (d): Protocol for uncertainty evaluation from validation data [S]. 2000.
- [14] International Organization of Standardization (ISO). *ISO/DTS 21748 Guide to the Use of Repeatability, Reproducibility and Trueness Estimates in Measurement Uncertainty Estimation* [S]. 2004.
- [15] Hubert P, Nguyen-Huu J, Boulanger B, *et al.* Harmonization of strategies for the validation of quantitative analytical procedures: A SFSTP proposal Part IV. Examples of application [J]. *J Pharm Biomed*, 2008, 48(3): 760-771.
- [16] González A, Herrador M. A practical guide to analytical method validation, including measurement uncertainty and accuracy profiles [J]. *TrAC-Trends Anal Chem*, 2007, 26(3): 227-238.
- [17] Govaerts B, Dewé W, Maumy M, *et al.* Pre-study analytical method validation: comparison of four alternative approaches based on quality-level estimation and tolerance intervals [J]. *Qual Reliab Eng Int*, 2008, 24(6): 667-680.
- [18] Feinberg M. Validation of analytical methods based on accuracy profiles [J]. *J Chromatogr A*, 2007, 1158(1/2): 174-183.
- [19] Mee R.  $\beta$ -expectation and  $\beta$ -content tolerance limits for balanced one-way ANOVA random model [J]. *Technometrics*, 1984, 26(3): 251-254.
- [20] Saffaj T, Ihssane B, Jhilal F, *et al.* An overall uncertainty approach for the validation of analytical separation methods [J]. *Analyst*, 2013, 138(16): 4677-4691.
- [21] Saffaj T, Ihssane B. Uncertainty profiles for the validation of analytical methods [J]. *Talanta*, 2011, 85(3): 1535-1542.
- [22] Nørgaard L. iToolbox for MATLAB [R/OL]. 2014. <http://www.models.life.ku.dk/iToolbox>.
- [23] Hoffman D, Kringler R. A total error approach for the validation of quantitative analytical methods [J]. *Pharm Res*, 2007, 24(6): 1157-1164.
- [24] Food and Drug Administration. *Guidance for Industry: Bioanalytical Method Validation* [S]. 2013.
- [25] Saffaj T, Ihssane B. Remarks on "Reply to the responses to the comments on "uncertainty profiles for the validation of analytical methods "by Saffaj and Ihssane" [J]. *Talanta*, 2013, 106: 155-157.
- [26] Rozet E, Marini R, Ziemons E, *et al.* Total error and uncertainty: Friends or foes? [J]. *TrAC-Trends Anal Chem*, 2011, 30(5): 797-806.