

扩展近红外通用性定量模型应用范围的可行性研究

雷德卿^{1,2}, 胡昌勤^{1*}, 冯艳春¹, 冯芳²

(1. 中国药品生物制品检定所, 北京 100050; 2. 中国药科大学, 江苏 南京 210009)

摘要: 近红外分析模型建立耗费大量的人力和物力, 是一项复杂的工作, 因此, 探索扩展现有模型应用范围的方法对近红外方法的推广具有重要意义。本文以通用性头孢拉定胶囊定量模型为例, 利用斜率/截距校正或分段直接标准化校正, 实现了利用通用性胶囊定量模型预测生产中混合或制粒后的药物填充物含量。结果表明, 这些校正方法扩展近红外通用性定量模型应用范围是可行的; 虽然校正效果较模型更新法略差, 但能满足过程控制的应用; 斜率/截距法较分段直接标准化法校正结果略好。

关键词: 近红外通用性模型; 扩展应用; 斜率/截距校正; 分段直接标准化校正; 模型更新

中图分类号: R917 文献标识码: A 文章编号: 0513-4870 (2010) 11-1421-06

Feasibility of the extended application of near infrared universal quantitative models

LEI De-qing^{1,2}, HU Chang-qin^{1*}, FENG Yan-chun¹, FENG Fang²

(1. National Institute for the Control of Pharmaceutical and Biological Products, Beijing 100050, China;

2. China Pharmaceutical University, Nanjing 210009, China)

Abstract: Construction of a successful near infrared analysis model is a complex task. It spends a lot of manpower and material resources, and is restricted by sample collection and model optimization. So it is important to study on the extended application of the existing near infrared (NIR) models. In this paper, cephradine capsules universal quantitative model was used as an example to study on the feasibility of its extended application. Slope/bias correction and piecewise direct standardization correction methods were used to make the universal model to fit to predict the intermediates in manufacturing processes of cephradine capsules, such as the content of powder blend or granules. The results showed that the corrected NIR universal quantitative model can be used for process control although the results of the model correction by slope/bias or piecewise direct standardization were not as good as that of model updating. And it also indicated that the model corrected by slope/bias is better than that by piecewise direct standardization. Model correction provided a new application for NIR universal models in process control.

Key words: near infrared universal model; extended application; slope/bias correction; piecewise direct standardization correction; model updating

近红外 (near infrared, NIR) 光谱分析具有方便快捷, 无需对样品进行预处理, 适用于在线、现场分析等特点, 已逐渐被药物分析领域的专家所关注, 近

年来应用近红外技术进行药物的生产过程控制成为药物分析的又一热点问题^[1-4]。

中国药品生物制品检定所自 2004 年以来研究药品通用性模型, 并装备在药品检测车上用于流通领域药品质量筛查。所谓通用性模型可以分析来自不同厂家含有同一种活性成分的同剂型的药品^[5]。例如

收稿日期: 2010-07-16.

*通讯作者 Tel: 86-10-67095308, Fax: 86-10-65115148,

E-mail: hucq@nicpbp.org.cn

通用性罗红霉素片剂定量分析模型, 共有 18 个厂家 78 批样品参与建模, 模型中样品浓度范围 19.5%~73.9%, 经过验证该模型几乎能对目前国内市场上所有厂家生产的罗红霉素片剂进行快速的定量分析^[6]。目前已经建立了 400 多种常见药品的通用性定性分析模型和 50 多种药品的通用性定量分析模型, 已经装备到全国近 400 辆检测车上应用。如果能将通用性模型经过简单的校正直接应用到药品的生产过程控制, 再逐渐进行模型完善, 既解决了由于初始建模繁琐导致近红外技术应用于过程控制起步慢的难题, 又为通用性模型找到新应用。

常用的模型校正方法主要包括模型更新^[7]、斜率/偏差 (slope/bias correction, S/B) 算法^[8]、直接标准化 (DS) 算法^[9]以及分段直接标准化 (piecewise direct standardization, PDS) 算法^[10]等。众所周知, 模型更新在这些方法中效果最好, 但并不能避免模型应用之前样品补充工作。因此, 本文以通用性头孢拉定胶囊定量模型为例, 尝试使用 S/B 和 PDS 算法对模型进行校正, 并与通过模型更新得到的校正结果进行比较, 探索能否找到一种简单易行的方法将通用性模型的应用范围拓展到过程控制中。

材料与方 法

仪器 MATRIX-F 傅立叶变换近红外光谱仪 (德国 Bruker 公司), 附有光纤探头, 铟镓砷 (InGaAs) 检测器, OPUS6.5 光谱分析软件。岛津 HPLC 色谱系统 (LC-20AT 泵, DGU-20A3 在线脱气机, SIL-20AC 自动进样器, CTO-10Asvp 柱温箱, SPD-M20A 二极管阵列检测器)。

样本 92 个厂家 102 批次头孢拉定胶囊为 2009 年全国抽验样品, 由中国药品生物制品检定所收集。头孢拉定胶囊内容物包括粉末混合或制粒后颗粒 (以下简称胶囊内容物), 20~40 粒同批次头孢拉定胶囊在实验室中去除囊壳, 内容物装入西林瓶中混合均匀, 即得, 浓度范围为 65.9%~94.6%。头孢拉定对照品 (批号: 130427-200806) 由中国药品生物制品检定所提供。

光谱测量 用光纤探头采集样品漫反射近红外光谱, 扫描范围: 12 000~4 000 cm^{-1} , 扫描次数 32 次, 分辨率 8 cm^{-1} , 每个样品重复测定 6 次, 计算平均光谱。胶囊/空胶囊: 直接测量胶囊壳单层端; 胶囊内容物: 取样本约 3 g, 装入西林瓶中, 摇匀轻磕使样本填实, 将光纤探头从瓶口插入接触样本测量。

化学分析 按照《中国药典》(2005 年版) 头孢拉定胶囊各论项下的要求^[11], 采用 HPLC 方法测定样品的含量, 含量按 $\text{C}_{16}\text{H}_{19}\text{N}_3\text{O}_4\text{S}$ 计算, 以纯度 (mg/mg) 表示。

建模 将头孢拉定胶囊剂通用性近红外定量分析模型^[12]作为源模型 (primary model), 其模型参数为: 训练集样本数为 100; 谱段范围为 9 751~7 498.4 cm^{-1} , 光谱预处理方法为一阶导数结合矢量归一化, 回归方法为偏最小二乘法 (PLS); 外部验证 (验证集样本数为 46) 均方差 (RMSEP) 为 2.24%, 与 HPLC 参比方法的最大预测偏差 (预测偏差为预测值与参比值之差的绝对值) 为 3.8%, 分别采用 S/B 法、PDS 法对源模型进行校正, 并与传统的模型更新 (model updating) 法比较。

S/B 法 S/B 校正是基于头孢拉定胶囊和胶囊内容物的预测值之间存在线性关系。使用一元线性回归方程进行拟合, 求得该线性方程的最小二乘解, 作为校正模型的斜率 (slope) 和截距 (bias):

$$\text{Slope} = \frac{\sum (y_{s,i} - \bar{y}_s)(y_{m,i} - \bar{y}_m)}{\sum (y_{s,i} - \bar{y}_s)^2} \quad (1)$$

$$\text{Bias} = \bar{y}_m - \text{slope} * \bar{y}_s \quad (2)$$

有了回归系数矩阵 b 和 slope/bias 校正项, 就可以直接利用胶囊内容物的光谱通过源模型计算样品的预测值。

PDS 法 PDS 校正是使用转换矩阵 F 将目标光谱矩阵 (objective spectra, S_m) 即头孢拉定胶囊光谱和转化光谱矩阵 (transformation spectra, S_s) 即对应的胶囊内容物光谱进行关联:

$$S_m = S_s F \quad (3)$$

$$F = S_s^{-1} S_m \quad (4)$$

其中 S_s^{-1} 为 S_s 的广义逆。待测的胶囊内容物光谱 X_s 经过转化为胶囊光谱 $X_{s, \text{std}}$ 后即可用于分析:

$$X_{s, \text{std}} = X_s F \quad (5)$$

对于一个实际的光谱数据, 波长点的漂移通常只局限于一个小的区域, 因此, S_m 中的每个波长点只需要与其对应 S_s 相应波长点附近的谱段相关联。具体为: 首先在 S_s 的第 i 个波长点左右扩展一个窗口 ($i-k \sim i+w$), 令 Z_i 表示 S_s 从 $i-k$ 到 $i+w$ 共 $k+w+1$ 个波长点的吸收度矩阵,

$$Z_i = [A_{s, i-k}, A_{s, i-k+1}, \dots, A_{s, i+w-1}, A_{s, i+w}] \quad (6)$$

然后将相对应的 S_m 的第 i 个波长点吸收度 $A_{m, i}$ 与 Z_i 构建一个多元回归模型:

$$A_{m, i} = Z_i b_i + e_i \quad (7)$$

此方程由 PLS 方法求解。将所有的回归系数 b_i 置于转换矩阵 F 的主对角线上, 并将其他元素置为 0, 这样就得到对角形矩阵 F :

$$F = \text{diag}(b_1^T, b_2^T, \dots, b_i^T, \dots, b_p^T) \quad (8)$$

(假设有 p 个波长点)

有了转换矩阵 F , 就可以对待测胶囊内容物样本的光谱 X_s 转换成头孢拉定胶囊光谱 $X_{s, \text{std}}$ 。

$$X_{s, \text{std}} = X_s F \quad (9)$$

结果与讨论

1 源模型的预测结果

利用源模型直接预测 102 批头孢拉定胶囊的含量, RMSEP 为 3.03%, 最大偏差为 5.69%, 与文献[12]结果基本相似; 提示源模型适用于对目前市场中流通的头孢拉定胶囊含量的预测。但如用源模型直接预测 102 批头孢拉定胶囊内容物的含量, 则 RMSEP 为 14.83%, 最大偏差为 19.89%; 可见无法直接利用源模型预测胶囊内容物的含量。

2 胶囊内容物样本聚类分析

按文献^[13]聚类分析的方法对 102 个胶囊内容物样本进行聚类分析。光谱的距离表明了谱图的相似度, 两张光谱 (在被比较的谱段范围内) 距离为零则表明它们一样, 两张光谱的距离值随谱图差别的增加而增大。聚类分析时首先计算所有谱图之间的光谱距离, 两张最相似 (距离最小) 的谱图被聚成一类, 计算该类与其他所有谱图的距离。最小距离的两个目标 (谱图/谱图、谱图/类) 再被聚成一类, 计算这个新类与其他目标 (谱图、类) 的距离, 最小距离的两个目标 (谱图/谱图或谱图/类或类/类) 被聚成一个新类, 这个过程一直重复到只剩下一个大类。选择 4 100~11 000 cm^{-1} 波长, 一阶导数+矢量归一化法, 9 点平滑, 因子化法, 因子数 4, 在马氏距离约 1.8 处, 所有样本可明显分为 5 类 (图 1)。在 5 类样本中按样本数比例选择出 31 个含量值均匀分布于整个含量

预测范围的样本作为训练集备用; 剩下的 71 个样本作为验证集备用。

3 S/B 法校正

分析源模型预测头孢拉定胶囊和胶囊内容物中头孢拉定含量的误差。以验证集样本 HPLC 测定的头孢拉定含量参考值为横坐标, 用 NIR 预测值为纵坐标, 源模型对头孢拉定胶囊和胶囊内容物中头孢拉定预测值与 HPLC 参考值均呈线性关系, 两条直线的斜率分别为 0.844 3 和 0.911 3, 对斜率作 F 检验 ($\alpha = 0.05$), 二者无显著性差异, 即两条直线为平行线。提示源模型预测头孢拉定胶囊与胶囊内容物中头孢拉定含量的误差为系统误差, 对胶囊内容物的预测偏差可以通过 S/B 校正的方法消除系统误差, 使得预测结果与参考值相近。

实际校正中, 首先根据图 1 的聚类分析结果, 按马氏距离 (约 1.8) 将全部胶囊内容物样本分为 5 类; 从每类样本中各选择 1 个校正样本, 5 个校正样本的含量均匀覆盖样本总体的含量范围; 且用源模型对 5 个校正样本的预测值 (y_s) 与参比 HPLC 值的预测偏差接近样本总体的平均预测方差 14.53% (由于样本总体中预测偏差在 11.0%~18.0% 的样本数占 80% 以上, 能保证绝大多数样本校正后的预测偏差在可接受的较小范围内)。所选择的 5 个校正样本的 HPLC 参考值分别为 68.51%, 75.42%, 82.45%, 87.54%, 93.61%, 预测值 (y_s) 分别为 52.60%, 60.48%, 69.94%, 73.11%, 79.91%。用校正样本的 HPLC 参考值作为 y_s 的校正值 ($y_{s, \text{corr}}$):

$$y_{s, \text{corr}} = 0.914 4 y_s + 20.05 \quad (r = 0.991 8) \quad (10)$$

用方程 (10) 校正 2 项下的 71 批验证集样本的含量, RMSEP 为 2.56%, 最大偏差为 5.52%。

4 PDS 法校正

PDS 法校正系对测定光谱的校正。从 102 个头孢拉定胶囊样本中随机选取 44 个样本, 分别测定胶囊、胶囊内容物和空胶囊的 NIR 光谱图, 光谱经一阶

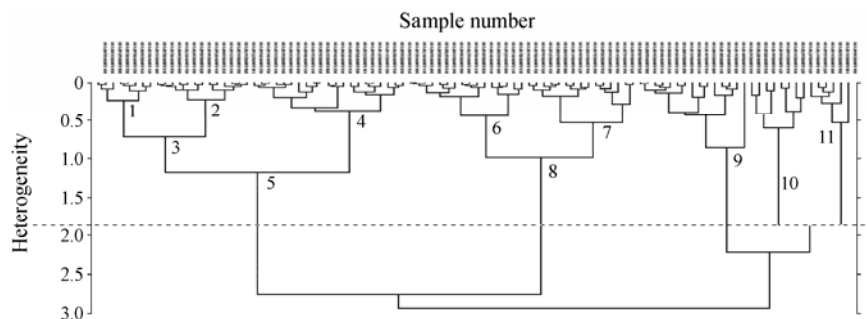


Figure 1 Cluster analysis for 102 batches of samples without capsule shell (5 classes)

导数、21点平滑处理后,比较三者光谱的差异(图2)。可见,头孢拉定胶囊光谱相当于空胶囊光谱与胶囊内容物光谱的加和,即头孢拉定胶囊光谱与胶囊内容物光谱的差异为系统误差;提示可以采用PDS法对胶囊内容物的光谱先进行校正,再利用通用性头孢拉定胶囊含量预测模型预测胶囊内容物的含量。

由图2可见,在属于C-H倍频与合频的4100~5000 cm⁻¹谱段(图中I区)和其二倍频的5500~7500 cm⁻¹谱段(图中II区),空胶囊的吸收峰较多且峰强度较高,不同基质和颜色的空胶囊光谱的差异主要表现在这两个谱段;而属于C-H三倍频的7500~9750 cm⁻¹谱段(图中III区)的吸收峰单一且强度较弱,在图中基本成一直线,差异性不明显。故光谱经PDS法校正,III区的校正效果应比I区和II区好。44批胶囊内容物光谱经PDS法校正后与相应头孢拉定胶囊光谱图比较(图3),I区、II区略有一定差异,III区基本一致。分别选择这三个谱段范围,计算校正后胶囊内容物光谱与相应胶囊光谱的相似度,结果I区、II区和III区相似度分别为:98.41%~99.96%、90.11%~99.72%和99.23%~99.97%,全范围校正效果都较好,而III区的校正效果最好。源模型的谱段

范围为7498.4~9751.0 cm⁻¹,胶囊内容物光谱经转换后与对应的头孢拉定胶囊光谱的相似度大于99.23%,说明通过PDS法对光谱进行校正,可以校正光谱的系统误差。

根据图1的聚类分析结果,从5类样本中选择12个胶囊内容物光谱作为标准S_s光谱,以相应的12个胶囊光谱作为标准S_m光谱。具体选择原则:①在每一类中按样本数比例选择标准样本,样本数较小的类选取的标准样本少,样本数较大的类选取的标准样本多,即按样本数量赋予不同类样本不同的权重,进而保证样本数较大的类在转换矩阵F中的贡献较大。②选择10~15个标准样品为宜。标准样品过多,转换矩阵F中矩阵元过多,计算量大,且有效信号的重复不能改善校正方法,反而易引入过多噪声信号;标准样品过少,转换矩阵F中矩阵元过少,不能充分提取有效信息。经预实验,选择12个标准样本的校正效果较好。③按方程(11),根据S_s光谱与S_m光谱的平均差异(M.D.)选择标准样本。

$$\text{M.D.} = \frac{1}{n} \sum_{i=1}^n |A_{i,S_s} - A_{i,S_m}| \quad (11)$$

S_s光谱与S_m光谱的平均差异M.D.值主要在

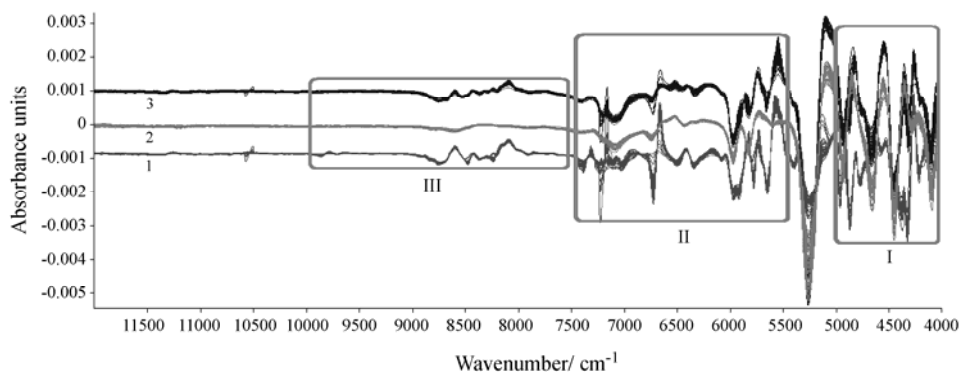


Figure 2 The comparison of the spectra of the samples without capsules shell, the capsules shell and the samples with capsule shell. 1: Samples without capsules shell; 2: Capsules shell; 3: Samples with capsule shell

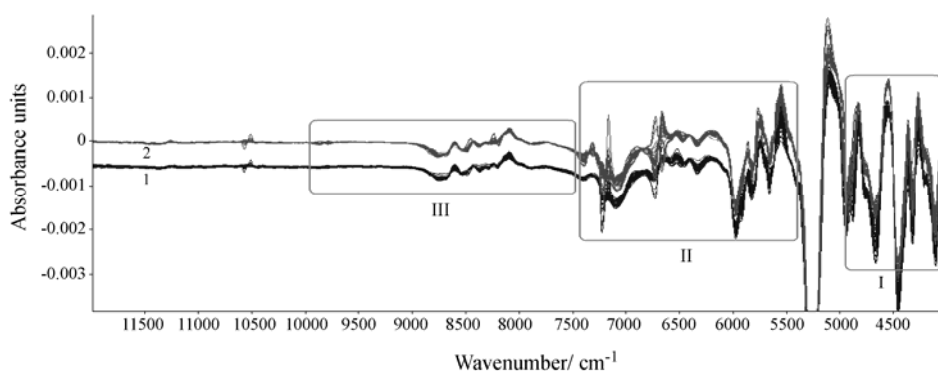


Figure 3 The results of the spectra after correction. 1: Samples with capsule shell; 2: Samples without capsule shell corrected by PDS

0.01~0.06 之间。实验显示, 选择 M.D. 值大于 0.055 或小于 0.015 的样本光谱作为标准 S_s 光谱, 校正效果变差。实验中选择的 12 个标准 S_s 光谱与标准 S_m 光谱的平均差异 M.D. 值在 0.018 8~0.054 0。此外, 由于 PDS 算法采用移动窗口技术, 故计算中在光谱的末端会出现由于数据点的缺失而不足以形成一个完整窗口的情况, 因此窗口不宜太宽; 但窗口太窄, 噪声信号可能不会被充分滤除, 仍保留在转换矩阵 F 中。经预实验, 窗口选择 7 点效果较好。

12 张标准 S_s 光谱及其转化成与源模型相匹配的 $X_{T, \text{std}}$ 光谱后与标准 S_m 光谱比较 (图 4), 可见, 标准 S_s 光谱与标准 S_m 光谱差异性较大, 而 $X_{T, \text{std}}$ 光谱与标准 S_m 光谱基本一致。

利用源模型, 采用上述 PDS 校正法预测 2 项下的 71 个验证集样本的含量, RMSEP 为 3.03%, 最大偏差为 6.02%。

5 模型更新法验证校正效果

模型更新法是成熟的校正技术。通过向源模型中添加新样本的信息, 可以扩充建模样本所覆盖的空间 (主成分) 范围。采用模型更新法对源模型进行校正, 并与上述两种校正结果进行比较, 进而评价上述校正方法的实用性。选择适宜的新样本建立新模型直接关系到预测结果的准确性。将 2 项下的 31 个训练集加入到源模型训练集中, 利用新训练集的 131 张光谱, 不改变源模型参数 (模型谱段: 7 498.4~9 751.0 cm^{-1} ; 光谱预处理方法: 一阶导数结合矢量归一化法, 21 点平滑), 得到新模型, 内部交叉验证, RMSECV (内部交叉验证均方差) 为 1.70%, R^2 为 93.29%; 用新模型预测上述验证集中 71 批次胶囊内容物样本的含量, RMSEP 为 1.84%, 与 HPLC 参比方法的最大偏差为 4.54%。

比较相同验证集中 71 批头孢拉定胶囊内容物预测值的校正结果及相应胶囊的预测结果: 胶囊预测结果、模型更新法、S/B 法及 PDS 法校正结果的 RMSEP 分别为 3.03%、1.84%、2.56% 和 3.03%, 最大偏差分别为 5.69%、4.54%、5.52% 和 6.02%。可见, S/B 法的校正效果更接近于模型更新法, PDS 法的校正结果与源模型预测头孢拉定胶囊的结果相当。在 S/B 校正法中, 采用了 HPLC 参考值而不是用胶囊预测值建立校正方程, 较好地消除了由胶囊壳和光谱仪等引入的系统误差和随机误差, 因此校正效果较好。PDS 校正法选用头孢拉定胶囊光谱作为目标光谱对待测光谱进行校正, 由于目标光谱本身存在多种变异, 故校正结果不能优于源模型的预测结果。

结果显示, 将通用性头孢拉定模型经过校正后用于企业生产过程控制, 预测药物混合或制粒后活性成分含量存在一定的系统误差, 但对于特定企业在一定时间内, 原辅料、处方和工艺都较稳定的样品而言, 由于这些样品本身含量变异范围较小, 且系统误差作用的方向相同, 因此并不影响校正后模型的整体应用。

结论

通用性头孢拉定胶囊剂定量模型的校正结果显示通用性模型经过校正后有应用到过程控制的可能性, S/B 算法和 PDS 算法都是比较简单易行的方法。两种算法校正的关键在于标准样品的选择, 利用系统聚类分析方法将样本分类, 再按一定的规则选择标准样本, 可保证校正效果比较理想。由于 S/B 法校正比 PDS 法校正算法更简单, 选择标准样本也更容易, 因此 S/B 法可作为首选的校正方法。

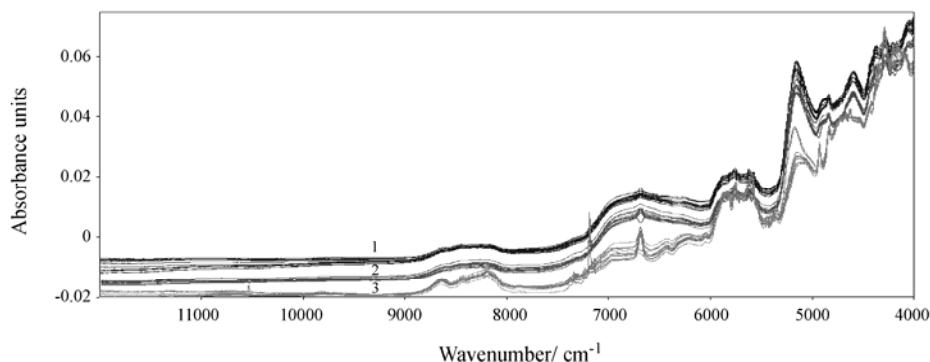


Figure 4 Spectra comparison between standard S_s and $X_{T, \text{std}}$ and standard S_m . 1: The spectrum of standard S_s ; 2: The spectrum of $X_{T, \text{std}}$; 3: The spectrum of standard S_m

致谢: 在本文的写作过程中, 得到了北京化工大学袁洪福老师的悉心指导, 在此表示衷心感谢!

References

- [1] Blanco M, Alcalá M. Content uniformity and tablet hardness testing of intact pharmaceutical tablets by near infrared spectroscopy — A contribution to process analytical technologies [J]. *Anal Chim Acta*, 2006, 557: 353–359.
- [2] Zheng YW, Lai XX, Bruun SW, et al. Determination of moisture content of lyophilized allergen vaccines by NIR spectroscopy [J]. *J Pharm Biomed Anal*, 2008, 46: 592–596.
- [3] Moes JJ, Ruijken MM, Gout E, et al. Application of process analytical technology in tablet process development using NIR spectroscopy: blend uniformity, content-uniformity and coating thickness measurements [J]. *Int J Pharm*, 2008, 357: 108–118.
- [4] Rosa SS, Barata PA, Martins JM. Near-infrared reflectance spectroscopy as a process analytical technology tool in *Ginkgo biloba* extract qualification [J]. *J Pharm Biomed Anal*, 2008, 47: 320–327.
- [5] Hu CQ, Feng YC. Scenario and feasibility of applying NIR in counterfeit drug recognition [J]. *Chin Pharm Affairs (中国药事)*, 2004, 18: 250–252.
- [6] Feng YC, Hu CQ. Construction of universal quantitative models for determination of roxithromycin and erythromycin ethylsuccinate in tablets from different manufacturers using near infrared reflectance spectroscopy [J]. *J Pharm Biomed Anal*, 2006, 41: 373–384.
- [7] Roychoudhury P, O’Kennedy R, McNeil B, et al. Multiplexing fibre optic near infrared (NIR) spectroscopy as an emerging technology to monitor industrial bioprocesses [J]. *Anal Chim Acta*, 2007, 590: 110–117.
- [8] Wang Y, Veltkamp DJ, Kowalski BR. Multivariate instrument standardization [J]. *Anal Chem*, 1991, 63: 2750–2756.
- [9] Wang Y, Lysaght MJ, Kowalski BR. Improvement of multivariate calibration through instrument standardization [J]. *Anal Chem*, 1992, 64: 562–564.
- [10] Bouveresse E, Hartmann C, Massart DL, et al. Standardization of near-infrared spectrometric instruments [J]. *Anal Chem*, 1996, 68: 982–990.
- [11] Chinese Pharmacopoeia Commission. Pharmacopoeia of People’s Republic of China (中华人民共和国药典) [S]. 2005 ed. Vol II. Beijing: Chemical Industry Press, 2005: 146–147.
- [12] Liu XP, Feng YC, Hu CQ, et al. Construction of universal models for determination of cefradine capsules [J]. *Chin J Pharm Anal (药物分析杂志)*, 2008, 28: 722–726.
- [13] Yan YL. Basis and Applying of Near Infrared Spectroscopy (近红外光谱分析基础与应用) [M]. Beijing: Chinese Light Industry Press, 2004: 359–364.