

近红外光谱法快速测定乳块消糖衣片的包衣厚度^{*}

刘冰¹, 陈晓辉¹, 史新元², 孙立新¹, 毕开顺¹, 乔延江^{2*}

(1. 沈阳药科大学, 沈阳 110016 2 北京中医药大学中药学院, 北京 100102)

摘要 目的: 利用近红外光谱分析技术, 建立乳块消糖衣片包衣厚度快速、无损测定方法。方法: 采用 Kennard-Stone法对校正集样本和验证集样本进行分类, 比较不同的建模方法, 并对光谱的预处理方法、建模波段、主因子数的选择进行详细地讨论。结果: 采用偏最小二乘回归方法 (PLSR) 建立包衣厚度的近红外光谱定量分析校正模型。所建模型的相关系数 $R = 0.9961$, 校正误差均方根 (RMSEC) 为 0.0149, 预测误差均方根 (RMSEP) 为 0.0194。结论: 预测结果表明, 本文所建方法快速、无损、可靠, 可推广应用于中药生产包衣过程的在线检测。

关键词: 近红外光谱; 乳块消片; 包衣厚度; Kennard-Stone法; 偏最小二乘回归方法; 在线检测

中图分类号: R917 文献标识码: A 文章编号: 0254-1793(2009)09-1435-05

Fast determination of coating thickness of Rukuaixiao tablets' by NIRS^{*}

LU Bing¹, CHEN Xiao-hui¹, SHI Xin-yuan², SUN Li-xin¹,
BI Kai-shun¹, QIAO Yan-jiang^{2*}

(1. Shenyang Pharmaceutical University, Shenyang 110016, China)

2. School of Chinese Pharmacy, Beijing University of Chinese Medicine, Beijing 100102, China)

Abstract Objective Near infrared spectroscopy (NIRS) analytical technique is used for establishing a rapid and nondestructive method to determine coating thickness of Rukuaixiao tablets. **Methods** Calibration and validation sets were partitioned by Kennard-Stone algorithm. Different calibration methods were compared with other method. Spectral pretreating methods, wavenumber selection and factors of NIRS were discussed in detail. **Results** Partial least squares regression (PLSR) was used for building the quantitative calibration model. Correlation coefficients (R), root mean square error of calibration (RMSEC), and root mean square error of prediction (RMSEP) obtained by PLSR model were 0.9961, 0.0149 and 0.0194 respectively. **Conclusion** The predictive result shows that the method of this study is rapid, non-destructive and credible, which can be applied to the on-line monitor of Chinese medicine tablets' coating process.

Key words near infrared spectroscopy; Rukuaixiao tablets; coating thickness; Kennard-Stone algorithm; partial least squares regression (PLSR); online control

乳块消片由丹参、橘叶、地龙、王不留行、皂角刺、川楝子 6 味中药组成, 是用于治疗乳腺增生的纯中药制剂。包衣厚度是其质量控制的重要指标, 直接影响到药品质量的好坏。与薄膜衣相比由于糖衣片较厚, 不能通过增重法测定厚度, 而采用螺旋测微仪测定厚度的方法麻烦、耗时, 生产过程中通过经验用包衣时间来控制包衣厚度, 从而产生很大误差, 影响药品的质量。因此, 建立包衣厚度的快速测定方法对于产品的质量控制在非常重要的。

近红外光谱法 (near infrared reflectance spectroscopy, NIRS) 是一种具有分析速度快, 样品不需特别预处理, 操作简单, 分析成本低等优点的分析技术。产生初期, 由于其吸收谱带宽, 重叠较严重, 吸收信号弱, 信息解析复杂等原因, 一直未能得到广泛应用。近年来, 随着计算机技术和化学计量学的发展, NIR 光谱技术在中药领域中的应用日趋广泛, 在定性、定量测定中显示了巨大的潜力, 被广泛应用于原药材的真伪鉴别^[1~3], 中间

* 北京市科技计划课题资助课题 (D02050040040111); 北京市“中药基础与新药研究重点实验室”资助课题

** 通讯作者 Tel: (010) 84738622 E-mail: yjia@263.net

体的质量控制^[4,5], 中成药中有效成分的含量测定^[6,7]。目前已有文献报道近红外光谱分析技术用于片剂的薄膜衣厚度测定^[8], 但尚未见用于片剂糖衣片的研究报道。

本文运用 Kennard-Stone 法对乳块消糖衣片的近红外光谱样本进行校正集和验证集分类, 然后采用偏最小二乘回归方法 (PLSR) 建立模型, 测定其包衣厚度。研究结果表明, 所建方法快速、无损、可靠。可望推广用于片剂糖衣片厚度的快速测定。

1 仪器与试剂

Antaris 傅里叶变换近红外光谱仪 (美国 Thermo Nicolet 公司), 积分球漫反射检测器附件, 该仪器配有软件包: Resulh 软件用于光谱采集及 TQ Analyst 软件用于数据处理。螺旋测微仪 (0~25 mm, 南京赛峰科技仪器实业有限公司)。

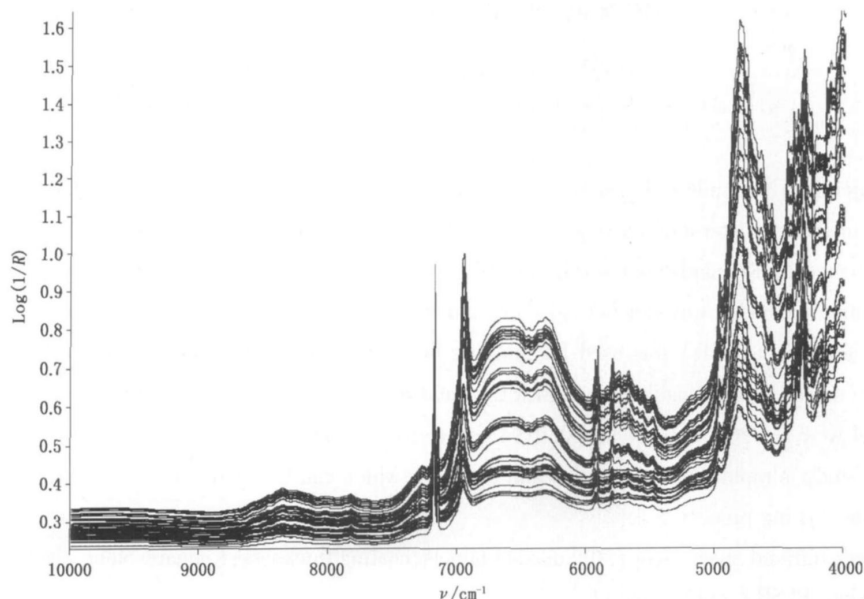


图 1 样品的原始 NIR 漫反射光谱

Fig 1 Original NIR diffuse reflectance spectra

2.3 数据处理软件 TQ Analyst 用于光谱的处理和计算 (近红外光谱仪自带); Kennard-Stone 算法工具包由 Michal Daszykowski 等人提供网络共享 (<http://www.chemometria.us.edu.pl/RobustToolbox/>) 下载, 各计算程序均自行编写, 采用 MATLAB 软件工具 (Mathwork Inc.) 计算。

2.4 校正模型评价参数 光谱经过预处理后, 采用 PLSR 建立乳块消糖衣片包衣厚度与 NIR 光谱之间的定量校正模型, 以模型的相关系数 (R)、内部交叉验证均方根差 (RMSECV)、校正误差均方根 (RM-

乳块消糖衣片 (批号: 83091-83093) 包衣样品由北京中医药大学附属药厂提供。共 3 批, 每批 7 个不同时间取样点, 制得不同厚度的糖衣片。

2 实验方法

2.1 样品的包衣厚度测定 每个取样点取样品 4~6 片, 共 104 个样本, 用螺旋测微仪测定厚度。首先用螺旋测微仪测定片剂正中顶端厚度, 然后剥去糖衣及隔离层, 再测片心正中顶端厚度, 取两厚度之差的 1/2 即为该片包衣厚度。

2.2 近红外光谱的采集 光谱采集条件: 采用积分球漫反射检测系统; NIR 光谱扫描范围 4000~10000 cm^{-1} ; 扫描次数 32 分辨率 8 cm^{-1} ; 以内置背景为参照。每份样品重复测定 6 次, 取其平均光谱用于数据处理, 样品的原始 NIR 光谱见图 1, 其中横坐标为波数值, 纵坐标为吸收度。

SEC)、预测均方根差 (RMSEP) 为指标, 优化模型参数。计算公式见 (1)、(2) 和 (3)。

$$R = \sqrt{1 - \frac{\sum_{i=1}^m (\hat{c}_i - c_i)^2}{\sum_{i=1}^m (c_i - \bar{c})^2}} \quad (1)$$

$$\text{RMSECV 或 RMSEC} = \sqrt{\frac{\sum_{i=1}^m (\hat{c}_i - c_i)^2}{n}} \quad (2)$$

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^m (\hat{c}_i - c_i)^2}{m}} \quad (3)$$

其中 $\bar{c}_i$ 表示包括校正集和验证集所有样本实际值的均值, c_i 为实际值, $\hat{c}_i$ 为预测值, n 为校正集样本数, m 为验证集样本数。

3 结果与讨论

3.1 Kennard–Stone法选取校正集和验证集

Kennard–Stone法是把所有的样本都看作校正集候选样本,依次从中挑选样本进校正集^[9~12]。首先,选择欧氏距离最远的2个向量对进入校正集;定义 d_{ij} 为从第 i 个样本向量到 j 样本向量的欧氏距离,假设已有 k 个样本向量被选进校正集,这里 k 小于样本总数 n ,针对第 v 个待选样本向量,定义最小距离:

$$D_{kv} = \min(d_{1v}, d_{2v}, \dots, d_{kv}) \quad (1)$$

所有待选样本向量的 D_{kv} 最大值: $D_{mkv} = \max(D_{kv})$,拥有最大最小距离 D_{mkv} 的待选样本进入校正集。依此类推,达到要求的样本数目。该方法优点是能保证校正集样本按照空间距离分布均匀,缺点是需要进行数据转换和计算样本两两空间距离,计算量大。

采取 PLSR 对进行回归,选择有代表性的校正集不但可以减少建模的工作量,而且直接影响所建模型的适用性和准确性。本文共 104 个样本,取其中的 2/3 作为校正集,1/3 作为验证集。表 1 是通过 Kennard–Stone 法依次挑选出 34 个验证集的样本号。

表 1 通过 Kennard–Stone 法依次挑选出验证集的样本号

Tab 1 Validation set selected from the whole primary set using Kennard–Stone algorithm

样本序号 (serialNo)	样本号 (sampleNo)						
1 [→] 7	3	12	14	15	16	17	22
8 [→] 14	26	30	33	35	41	42	43
15 [→] 21	44	45	46	53	54	56	62
22 [→] 28	67	69	70	81	84	87	92
29 [→] 34	94	95	99	100	103	104	

图 2 为通过 Kennard–Stone 法挑选 104 个样本的主成分得分图。因前 2 个主成分 (PC) 已经代表了大于 90% 的光谱信息, 所以将 PC1 对 PC2 作图, 可得样本在主成分空间的分布情况。可以看出, 所选样本的分布比较均匀, 基本没有异常点存在, 而且 34 个验证集样本也都被包容在校正集的空间内, 校正集和验证集的取舍也比较合理。

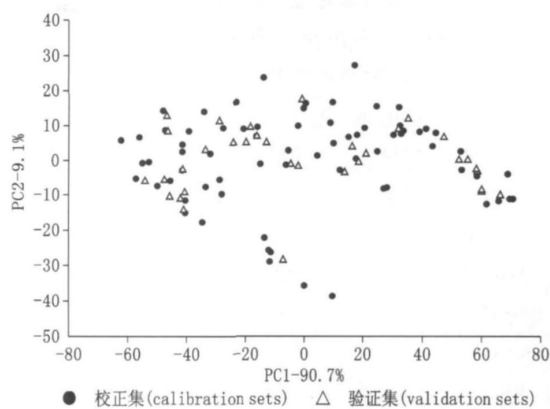


图 2 样品的主成分得分图

Fig 2 PC score plot of PCA for samples

3.2 建模方法的选择 研究中分别比较了逐步多元线性回归 (SMLR)、偏最小二乘回归 (PLSR)、主成分回归 (PCR) 所建模型的性能见表 2 结果表明以 PLSR 建立的模型性能较好。

表 2 不同建模方法的比较

Tab 2 Comparison of different calibration methods

建模方法 (methods)	相关系数 (R)	校正误差均方根 (RM SEC)	预测均方根 (RMSEP)
SMLR	0.9717	0.0399	0.0292
PLSR	0.9900	0.0238	0.0267
PCR	0.9735	0.0386	0.0273

3.3 光谱预处理方法 研究中考察了多元散射校正 (MSC) 与标准正态变换 (SNV) 分别与导数法 (First Derivation, FD 法; Second Derivation, SD 法) 及平滑法 [Savitzky–Golay (SG) 滤波器平滑、Norris Derivative (ND) 滤波器平滑] 结合, 作为光谱预处理方法, 对模型性能的影响, 结果见表 3, 经比较, SNV 法、FD 和 SG 平滑法 (五点三次多项式平滑) 相结合作为光谱的预处理方法效果较好。这是因为在进行 NIR 漫反射光谱采集时, 由于样品颗粒尺寸、均匀性等的影响, 光程无法保持恒定, 使用 MSC 与 SNV 对光谱进行处理, 可以消除这些因素的影响。而导数处理既可以消除基线漂移, 还可以起到一定的放大和分离重叠信息, 但在对光谱进行微分处理时, 由于噪声信号也同时被放大, 因此需要和平滑处理结合。平滑处理的作用是提高信噪比, 减少随机噪音。

表 3 不同光谱预处理方法对 PLSR 模型的结果

Tab 3 Effects of various pretreatment methods on PLSR model				
光谱预处理方法 (pretreatments)	相关系数 (R)	校正误差均方根 (RMSEC)	预测均方根 (RMSEP)	主因子数 (factors)
MSC	0.9882	0.0259	0.0249	9
MSC+SG	0.9882	0.0259	0.0240	10
MSC+FD	0.9929	0.0201	0.0243	7
MSC+FD+SG	0.9924	0.0208	0.0237	7
MSC+FD+ND	0.9877	0.0264	0.0265	7
MSC+SD	0.9888	0.0252	0.0289	4
MSC+SD+SG	0.9934	0.0193	0.0257	6
MSC+SD+ND	0.9869	0.0272	0.0302	7
SNV	0.9887	0.0253	0.0239	10
SNV+SG	0.9883	0.0257	0.0240	10
SNV+FD	0.9930	0.0199	0.0242	7
SNV+FD+SG	0.9961	0.0149	0.0194	8
SNV+FD+ND	0.9902	0.0236	0.0231	9
SNV+SD	0.9888	0.0252	0.0288	4
SNV+SD+SG	0.9934	0.0193	0.0255	6
SNV+SD+ND	0.9919	0.0214	0.0233	9

3.4 光谱波段选择范围 建立模型时可采用全谱或部分波段,应该视情况而定,虽然全谱包涵的信息丰富,但存在各成分谱图重叠严重,光谱信息冗余,特征吸收区域不明显的问题,需要对采集到的波长进行优选,以达到提高模型预测精度和简化模型的目的,如果选择不当将影响模型的准确性。本文比较了全谱和信息量最丰富,受背景干扰最少的 2 个波段,如表 4 所示,结果表明采用后者建模效果较好。从近红外光谱角度分析,5800.83~4994.73 cm^{-1} 是糖衣材料蔗糖中 CH , CH_2 的一级倍频吸收,6861.48~6001.39 cm^{-1} 是蔗糖中的羟基的二级倍频吸收。

表 4 不同的波段范围对模型的影响

Tab 4 Effect of wavelength ranges on PLSR model				
光谱范围 (ranges) / cm^{-1}	相关系数 (R)	校正误差均方根 (RMSEC)	预测均方根 (RMSEP)	主因子数 (factors)
1000~4000	0.9925	0.0206	0.0291	9
6861.48~6001.39	0.9961	0.0149	0.0194	8
5800.83~4994.73				

3.5 PLSR 主因子数的确定 在采用 PLSR 建模过程中,应对主因子数进行合理选择。通过留一交叉验证方法,当 RMSECV 最小时的主因子数,即最佳

主因子数。如果主因子数太大,会出现“过拟合”现象,反之,则是“欠拟合”。本实验在 6861.48~6001.39 cm^{-1} 和 5800.83~4994.73 cm^{-1} 波段,确定 SNV 法、FD 和 SG 平滑法作为预处理方法,采用留一交叉验证法确定 PLSR 建模中主因子数,结果见图 3。当主因子数为 8 时,RMSECV 值最小,确定其最佳主因子数为 8,此时 RMSECV 为 0.0252。

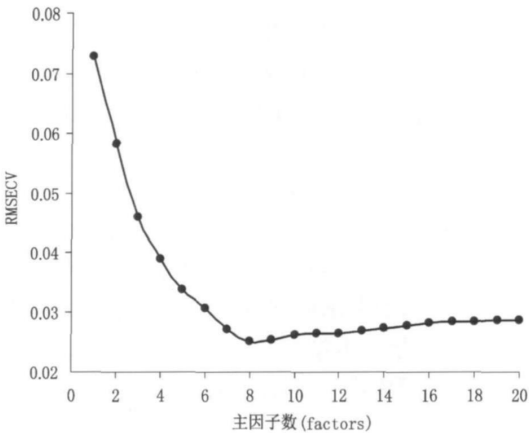


图 3 最佳主因子数和交叉验证结果
Fig 3 Optimal PLSR factors and results obtained by cross-validation

3.6 预测结果 光谱经 SNV 法、FD 和 SG 平滑法相结合预处理后,在 6861.48~6001.39 和 5800.83~4994.73 cm^{-1} 2 个波段,采用 PLSR 建立包衣厚度最优的校正模型并进行预测,结果见图 4。校正模型的相关系数 R 为 0.9961,校正误差均方根 (RMSEC) 为 0.0149,预测误差均方根 (RMSEP) 为 0.0194。糖衣片的近红外光谱预测值能够准确地逼近实测值。

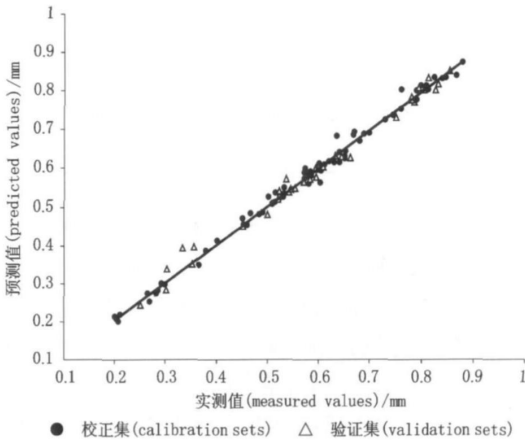


图 4 验证集样品 NIRS 预测值和包衣厚度相关图
Fig 4 Correlation between NIRS predicted and coating thickness
© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

4 结论

本文采用近红外漫反射光谱法建立了乳块消糖衣片包衣厚度的快速测定方法。上述研究结果表明,根据 PLS回归方法建立预测模型,对乳块消片包衣厚度的测定是可行的,而且结果令人满意,同其他方法相比,不需复杂的样品前处理,具有快速、无损等优点。

本方法具有一定的普适性,有望推广用于其他片剂糖衣厚度的在线检测。

参考文献

- 1 YANG Hai- lei(杨海雷), LU Xue- song(刘雪松), QU Hai- bin(瞿海斌), *et al*. Fast quality assessment of Radix Ginseng Rubra by near- infrared spectroscopy(一种基于近红外的红参药材质量快速评价方法). *Chin Tradit Herb Drugs*(中草药), 2005, 36(6): 912
- 2 HU Gang- liang(胡钢亮), CHEN Rui- zhen(陈瑞珍), CHENG Ke(程柯), *et al*. Rapid quantitative determination of *Fritillaria thunbergii* Miquel into *Fritillaria cirrhosa* D. Don by near- infrared diffuse reflectance spectroscopy(近红外漫反射光谱快速检测川贝母中浙贝母的掺入量). *Chin J Pharm Anal*(药物分析杂志), 2005, 25(2): 150
- 3 ZHANG Xiao- fang(张孝芳), ZHANG Zhuo- yong(张卓勇), FAN Guo- qiang(范国强). Identification of rhubarb samples with support vector machine and near- infrared spectrometry(支持向量机与近红外光谱法鉴定大黄). *Chin J Pharm Anal*(药物分析杂志), 2006, 26(7): 914
- 4 NILI- jun(倪力军), SHI Xiao- hao(史晓浩), GAO Xin- jiao(高秀蛟), *et al*. Research on the application of on- line detection and analytical technique by NIR in quality control of water extracting process of *Salvia miltiorrhiza* (NIR在线检测、分析技术在丹参水提过程质量监控中的应用). *Chin Pharm J*(中国药杂志), 2004, 39(8): 628
- 5 QU Hai- bin(瞿海斌), LI Bin(李斌), LIU Xue- song(刘雪松), *et al*. On- line determination for concentrating process of red ginseng alcohol extract using near- infrared spectroscopy(红参醇提液浓缩过程近红外光谱在线分析方法). *Chin Pharm J*(中国药杂志), 2005, 40(24): 1897
- 6 JIA Nuan(贾暖), LIU Li- li(刘荔荔), WU Yu- tian(吴玉田). Determination of norfloxacin in capsules using near infrared diffuse reflectance spectrophotometry(近红外漫反射光谱法测定诺氟沙星胶囊的含量). *Chin J Pharm Anal*(药物分析杂志), 2002, 22(2): 122
- 7 YU Ke(虞科), HU Chu- chu(胡楚楚), CHENG Yi- yu(程翼宇). A novel method for fast determination of three active components in compound Danshen dropping pills by near infrared spectroscopy(近红外光谱法测定复方丹参滴丸中的3种有效成分). *Chin Pharm J*(中国药杂志), 2006, 41(3): 226
- 8 KE Bo- ke(柯博克), LIU Xue- song(刘雪松), CHENG Yi- yu(程翼宇), *et al*. Film coating thickness of compound Danshen dropping pills by NIRs fast detecting(近红外光谱快速测定复方丹参滴丸的包衣厚度). *Chin Tradit Herb Drugs*(中草药), 2006, 37(5): 685
- 9 Kennard RW, Stone LAS. Computer aided design of experiments. *Technometrics*, 1969, 11: 137
- 10 Kocjančič R, Zupan J. Modelling of the river flow rate: the influence of the training set selection. *Chemometr Intell Lab Syst*, 2000, 54: 21
- 11 Siano GG, Goicoechea HC. Representative subset selection and standardization techniques. *Chemometr Intell Lab Syst*, 2007, 88: 204
- 12 Chen Da- Cai Wen- sheng Shao Xue- guang. Representative subset selection in modified iterative predictor weighting (mIPW) - PLS models for parsimonious multivariate calibration. *Chemometr Intell Lab Syst*, 2007, 87: 312

(本文于 2008 年 10 月 23 日收到)