

## 基于偏最小二乘法建立大株红景天片素片硬度近红外光谱预测模型

陈露萍<sup>1</sup>, 徐芳芳<sup>2,3\*</sup>, 张欣<sup>2,3</sup>, 吴云<sup>2,3</sup>, 王振中<sup>2,3</sup>, 肖伟<sup>1,2,3\*</sup>

1. 南京中医药大学, 江苏 南京 210023

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 基于近红外光谱 (near infrared spectrum, NIRS) 技术, 建立一种快速预测大株红景天片 (*Rhodiola grandiflora* Tablets, RGT) 素片硬度的方法。方法 采集共 600 个生产样本与自制样本的 NIRS, 通过比较不同光谱预处理方法与不同特征变量筛选条件下模型的优劣, 建立偏最小二乘 (partial least-square, PLS) 算法模型, 另采集 120 个样本的 NIRS 对模型进行外部验证, 预测 RGT 素片硬度。结果 建立的素片硬度 PLS 模型中, 无预处理的光谱采用全波段建模的模型效果最佳, 预测模型的校正集相关系数 (correlation coefficient of training set,  $R_{cal}$ ) 与验证集相关系数 (correlation coefficient of verification set,  $R_{pre}$ ) 分别为 0.971 9 与 0.988 7, 相关性良好, 预测均方根误差 (root mean square error of prediction, RMSEP) 为 2.03 N, 性能偏差比 (ratio of performance to deviation, RPD) 为 6.68, 预测相对偏差 (relative standard to deviation, RSEP) 为 4.24%, 模型内部验证的平均相对预测误差为 2.82%, 外部验证的平均相对预测误差为 4.59%, 均 < 5%, 对不合格素片的检出率高达 97.33%。结论 NIRS 分析技术结合 PLS 算法建立的 RGT 素片硬度预测模型具有良好的模型性能与预测能力, 为 RGT 素片硬度的无损检测提供了一种新方法。

**关键词:** 近红外光谱技术; 偏最小二乘法; 大株红景天片; 硬度; 素片; 无损检测

中图分类号: R283.6 文献标志码: A 文章编号: 0253-2670(2023)08-2446-07

DOI: 10.7501/j.issn.0253-2670.2023.08.011

## Near infrared prediction model for hardness of *Rhodiola grandiflora* Tablets based on partial least square method

CHEN Lu-ping<sup>1</sup>, XU Fang-fang<sup>2,3</sup>, ZHANG Xin<sup>2,3</sup>, WU Yun<sup>2,3</sup>, WANG Zhen-zhong<sup>2,3</sup>, XIAO Wei<sup>1,2,3</sup>

1. Nanjing University of Traditional Chinese Medicine, Nanjing 210023, China

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

3. State Key Laboratory of New Technologies in Traditional Chinese Medicine Pharmaceutical Process, Lianyungang 222001, China

**Abstract: Objective** To establish a rapid method for predicting the hardness of *Rhodiola grandiflora* Tablets (RGT) based on near infrared spectroscopy (NIRS). **Methods** The NIRS of 600 production samples and self-made samples were collected. The partial least square (PLS) algorithm model was established by comparing the advantages and disadvantages of the models under different spectral pretreatment methods and different characteristic variable screening conditions. The NIRS of 120 samples were collected for external verification of the model to predict the hardness of RGT. **Results** Among the PLS models for plain tablet hardness, the full band model was the best for the spectrum without pretreatment. The correlation coefficient of training set ( $R_{cal}$ ) and the correlation coefficient of verification set ( $R_{pre}$ ) of the prediction model were 0.971 9 and 0.988 7, respectively. The root mean square error of prediction (RMSEP) is 2.03 N, the ratio of performance to deviation (RPD) is 6.68, and the relative standard to deviation (RSEP) is 4.24%. The average relative prediction error of the internal validation of the model is 2.82%, and the average relative prediction error

收稿日期: 2022-10-25

基金项目: 连云港市科学技术局重大技术攻关“揭榜挂帅”项目 (CGJBGS2101); 中药口服固体制剂智能化连续制造关键技术研究

作者简介: 陈露萍, 女, 硕士研究生, 研究方向为中药制药技术与产品开发。Tel: 18852008496 E-mail: 1793829067@qq.com

\*通信作者: 肖伟, 中国工程院院士, 研究员级高级工程师, 博士生导师, 研究方向为中药新药研发及过程控制研究。

E-mail: kanionlunwen@163.com

徐芳芳, 女, 博士, 研究方向为中药新药研发及过程控制研究。E-mail: 879164331@qq.com

of the external validation is 4.59%, both of which are less than 5%. The detection rate of unqualified tablets is as high as 97.33%.

**Conclusion** The hardness prediction model of RGT established by near-infrared spectroscopy combined with partial least squares algorithm has good model performance and prediction ability. This study provides a new method for nondestructive testing of the hardness of RGT.

**Key words:** near-infrared spectroscopy; partial least square method; *Rhodiola grandiflora* tablets; hardness; plain tablets; non destructive testing

大株红景天片 (*Rhodiola grandiflora* Tablets, RGT) 是由大株红景天药材经提取精制加工而成的中药成方制剂, 具有活血化瘀、通脉止痛等作用<sup>[1]</sup>。硬度是 RGT 质量控制中最重要指标之一, 影响片剂的外观质量和内在属性, 硬度过大会造成片剂崩解迟缓, 从而影响药物的释放与疗效<sup>[2]</sup>; 而硬度不够会造成素片脱粉, 包衣粗糙, 运输过程中易断裂等问题。目前, 片剂硬度通常需要操作人员在压片过程中不断取样后使用硬度测定仪测定而得, 这种方法劳动强度大且具有破坏性, 造成生产资源的浪费。因此, 亟需建立一种绿色无损的检测方法, 应用于 RGT 生产过程中的质量控制。

近年来, 近红外光谱 (near infrared spectrum, NIRS) 分析技术<sup>[3-6]</sup> 凭借其快速无损的优点已被广泛应用于中药制药行业中, 在预测有效成分含量<sup>[7]</sup>、颗粒性质<sup>[8]</sup>、药材鉴别<sup>[9]</sup> 等方面取得了显著成果, 而片剂物理属性的预测模型相对较少。刘秋安等<sup>[10]</sup> 利用分类与回归树算法建立了天舒片崩解时间预测模型, 表明 NIRS 技术预测片剂物理属性是可行的。邱素君等<sup>[11]</sup> 利用近红外漫反射光谱建立了 NIRS 与硬度的校正模型, 相关系数达到 0.977 8。林翔等<sup>[12]</sup> 利用 NIRS 技术, 结合化学计量学方法, 建立了辛伐他汀片压片过程水分含量、制片压力、片剂硬度与主药含量 4 个参数的定量分析模型, 所建模型具有良好的预测能力。上述研究表明, NIRS 技术已在预测片剂物理属性方面取得了一定成果。

目前, 国内外对于 NIRS 预测片剂硬度的报道相对较少, 本研究拟采用偏最小二乘法 (partial least-square, PLS) 建立 RGT 素片硬度定量预测模型, 为片剂质量控制提供一种新方法。

## 1 仪器与材料

### 1.1 仪器

Antaris II 型傅里叶变换近红外分析仪, 配积分球漫反射检测器, 美国 Thermo Fisher Scientific 公司; YPD-300D 型片剂硬度仪, 上海黄海药检仪器有限公司; DP-30A 型单冲压片机, 北京国药龙立科技有限公司。

### 1.2 材料

RGT 素片与总混颗粒样本均由江苏康缘药业股份有限公司提供, 9 批 RGT 素片 (生产合格样本) 批号 220401、220402、220501、220502、220503、220601、220602、220603、220701; 9 批 RGT 总混颗粒 (对应批次自制不合格样本) 批号 220401、220402、220501、220502、220503、220601、220602、220603、220701。

## 2 方法

### 2.1 不合格素片的制备

参照 RGT 企业内控标准, 素片硬度的质量控制要求为大于 34.3 N, 在实际生产过程中, 物料的差异常常导致不同批次素片硬度差异较大, 而同一批次素片容易受压片机冲头压力变化、装量差异等影响, 导致素片硬度产生波动, 不合格素片由此产生。在压片机高速运作的过程中, 车间人员需要不断取样检测以确保素片硬度合格, 而取样方法多采用抽样检测, 具有随机性与破坏性, 不仅造成资源浪费, 且无法确保成品质量。本研究在分析 9 批生产合格素片样本基础上, 取同批次 9 批总混颗粒, 模拟了生产过程中可能出现的各种异常情况, 通过采取降低冲头压力, 在装量合格范围内改变片重等操作, 制备了部分不合格素片, 与生产合格素片共同参与模型建立, 以扩大样本分布范围, 提高模型的泛化能力。

### 2.2 NIRS 采集

将合格样本与不合格样本采用同一参数进行 NIRS 数据信息的采集, 将 RGT 素片置于近红外分析仪圆形药片固定装置卡槽内, 波数范围为 10 000~4000  $\text{cm}^{-1}$ , 分辨率为 8  $\text{cm}^{-1}$ , 4 倍增益, 参比背景为空气, 每个样本平行采集 3 条光谱, 每条光谱为 32 次扫描的平均值, 求平均光谱。

本研究共采集 600 个素片样本的 NIRS, 结果如图 1 所示, 样本吸光度略有差异但整体趋势相近。

### 2.3 素片硬度测定

素片 NIRS 扫描后, 使用片剂硬度仪测定各素片硬度, 按照编号记录硬度数据。376 个生产合格

样本与 224 个自制不合格样本的硬度数据分布情况如表 1 所示, 2 种样本共计 600 个。

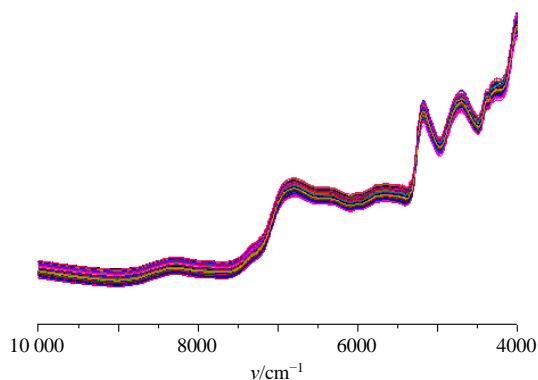


图 1 600 个素片样本的 NIRS 图

Fig. 1 NIRS of 600 plain film samples

表 1 合格样本与不合格样本的硬度数据分布情况

Table 1 Distribution of hardness data between qualified and unqualified samples

样本	数量/ 个	最大值/ N	上四分位 数/N	下四分位 数/N	最小值/ N
合格	376	76.24	58.92	45.98	34.89
不合格	224	33.91	28.71	7.59	0.15

## 2.4 PLS 定量模型的建立

**2.4.1 建模方法及评价指标的确定** PLS 算法<sup>[17]</sup>是经多次迭代映射分解, 组合成一个  $Y$  与  $X$  的线性回归, 将每次迭代过程中的映射变换参数进行线性运算从而得到最终模型。PLS 采用成分分解的思路来进行特征降维, 关注各个变量对目标变量的解释, 会保留共线性的变量, 先拟合大致轮廓, 再逐步拟合细节, 考虑当前的最优拟合决策进行局部最优化。

本研究以 PLS 算法建立素片硬度定量模型, 通过优选主成分数、光谱预处理以及波段筛选等方式建立最优模型, 以校正集相关系数 (correlation coefficient of training set,  $R_{cal}$ )、验证集相关系数 (correlation coefficient of verification set,  $R_{pre}$ )、校正均方根偏差 (root mean square error of correction, RMSEC)、预测均方根偏差 (root mean square error of prediction, RMSEP)、交叉验证均方根误差 (root mean square error of validation, RMSECV)、性能偏差比 (ratio of performance deviation, RPD)、预测相对偏差 (relative standard to deviation, RSEP) 为指标评价模型优劣。

上述指标中, 一般以  $R_{cal}$  与  $R_{pre}$  判断模型的拟合程度, 这两个指标越趋近于 1, 对应的模型拟合

效果越好。RMSEC、RMSECV、RMSEP 与 RSEP 通常用来判断模型的预测误差, 这 4 个指标越小模型的预测精度越高, 性能越好。RPD 则用来评价模型的预测效果, 一般认为  $RPD \geq 3$  时, 模型的预测精度较高, 可以用来定量检测某些指标;  $2.5 \leq RPD < 3$  时, 认为能够粗略检测某些指标, 此时可采取相关措施提高模型的预测精度;  $RPD < 2.5$  时, 认为此时的模型不适合检测该指标, 需要重新建模或者采取相关措施<sup>[13]</sup>。相关指标计算公式如下。

$$R_{cal} = [1 - \sum(C_i - C_i')^2 / \sum(C_i - C_n)]^{1/2}$$

$$R_{pre} = [1 - \sum(C_i - C_i')^2 / \sum(C_i - C_m)]^{1/2}$$

$$RMSEC = [\sum(C_i - C_i')^2 / n]^{1/2}$$

$$RMSEP = [\sum(C_i - C_i')^2 / m]^{1/2}$$

$$RSEP = [\sum(C_i - C_i')^2 / \sum C_i^2]^{1/2}$$

$C_i$  和  $C_i'$  分别为测定的参考值和 NIRS 预测值,  $C_n$  和  $C_m$  分别为校正集和验证集样本参考值的平均值,  $n$  和  $m$  分别为校正集和验证集样本的个数

**2.4.2 数据处理** 采用 UnscrambleX 10.4 (Camo software AS, 挪威) 软件进行主成分分析 (principal component analysis, PCA) 与光谱预处理, 采用 Matlab 2020 (MathWorks 公司, 美国) 软件进行 PLS 模型的样本集划分、变量筛选及模型构建, 采用 Origin 8 (OriginLab 公司, 美国) 绘制 NIRS 图, 采用 GraphPad Prism 8 (GraphPad Software 公司, 美国) 软件绘制校正集与验证集线性关系图。

**2.4.3 样本集划分** 本研究采用 Kennard-Stone (K-S) 算法<sup>[10,14]</sup>划分样本集。K-S 算法是通过最大化所选样本的仪器响应变量 ( $x$ ) 之间的欧氏距离进行样本集划分, 以统一的方式覆盖多维空间。首先, 计算所有样本之间的欧式距离, 将具有最大距离的 2 个样本选入校正集, 并在后续的每一次迭代中, 不断计算剩余样本与校正集样本中光谱之间的最小欧式距离, 选择 2 个样本的最小欧式距离最大的样本加入校正集, 直至样本数量达到设定的数目。

PCA 是利用数据降维的思想, 在损失较少数据信息的前提下, 将多个指标转化为少数几个综合指标的多元分析方法<sup>[15]</sup>, 本研究基于原始光谱对合格样本与不合格样本进行 PCA, 作为掌握 2 类样本分布情况与为后续建模剔除异常样本的辅助手段。对合格样本与不合格样本进行 PCA, 基于第 1 主成分 (83%) 与第 2 主成分 (13%) 的 PCA 得分图如图 2 所示, 前 2 个主成分代表了 96% 的光谱信息, 且 2 类样本分布均匀, 无明显差异, PCA 无法作为样本

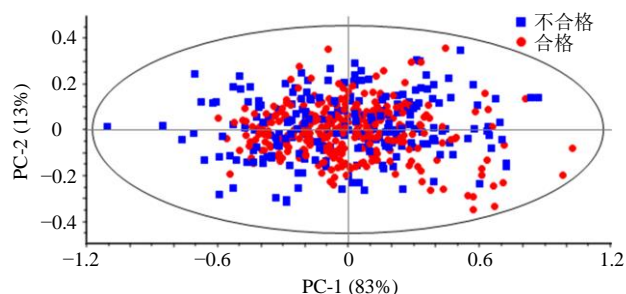


图2 合格样本与不合格样本的PCA得分图

Fig. 2 PCA scores of qualified and unqualified samples

硬度是否合格的直接判别依据,因此还需建立定量模型进一步研究;而样本分布较为集中,无明显的离群值,因此本研究后续素片硬度PLS定量模型的建立不进行异常样本的剔除。

采用K-S算法将600个样本按4:1的比例划分为480个校正集和120个验证集,样本硬度数据分布情况如表2所示。

对校正集样本与验证集样本进行PCA,得分图如图3所示,2类样本分布均匀,且校正集覆盖了验证集的范围,说明校正集与验证集的样本选择合理,有利于模型的建立<sup>[14]</sup>。

2.4.4 主成分数的选择 对于PLS建立校正模型,

表2 校正集与验证集样本硬度数据分布情况

Table 2 Distribution of sample hardness data in calibration set and verification set

样本	数量/ 个	最大值/ N	上四分位 数/N	下四分位 数/N	最小值/ N
校正集	480	76.24	54.00	16.37	0.15
验证集	120	68.40	55.45	45.66	1.90

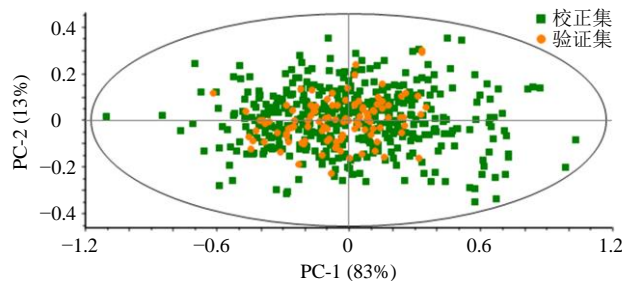


图3 校正集样本与验证集样本的PCA得分图

Fig. 3 PCA scores of calibration set samples and validation set samples

参与线性回归的最佳主成分数的选择非常重要。若选取的主成分数过少,会导致模型拟合不充分,丢失光谱的重要信息;若选取的主成分数过多,则可能出现过拟合现象而影响模型结果。本研究采取留一交叉验证法<sup>[8]</sup>,以RMSECV为评价指标选取最佳主成分数,结果见表3。

2.4.5 光谱预处理 采集光谱时,由于受到环境温度湿度、仪器状态以及样本状态等各种因素的影响,NIRS会产生噪声信号与基线漂移等情况,常常需要对光谱进行预处理,以消除干扰,提升模型稳健性。常见的光谱预处理方法<sup>[10]</sup>中,矢量归一化法(vector normalize)可增强光谱差异,多元散射校正(multiplicative scatter correction, MSC)和标准正态变换(standard normal variate transformation, SNV)可消除样本分布不均带来的干扰,导数法可消除基线漂移,M-A平滑(moving average smoothing)和S-G平滑(Svaitzky-Golay smoothing)可滤除噪声。

本研究考察了以下9种预处理方法:基线校正、M-A平滑、MSC、矢量归一化法、一阶导数法、S-G

表3 不同预处理方法对PLS模型性能的影响

Table 3 Effect of different pretreatment methods on PLS model performance

预处理方法	校正集				验证集			
	主成分数	$R_{cal}$	RMSEC/N	RMSECV/N	$R_{pre}$	RMSEP/N	RPD	RSEP/%
无预处理	17	0.971 9	4.78	5.29	0.988 7	2.03	6.68	4.24
基线校正	17	0.972 8	4.71	5.33	0.987 1	2.17	6.26	4.53
M-A平滑	18	0.971 9	4.78	5.19	0.987 8	2.12	6.40	4.43
MSC	16	0.968 4	5.06	5.59	0.987 6	2.12	6.38	4.44
矢量归一化法	17	0.969 1	5.01	5.53	0.988 0	2.09	6.50	4.37
一阶导数	11	0.972 0	4.78	5.84	0.969 6	3.41	3.98	7.13
S-G平滑	17	0.971 0	4.86	5.29	0.988 7	2.07	6.56	4.33
S-G平滑+一阶导数	11	0.972 5	4.73	5.85	0.977 6	2.85	4.75	5.97
S-G平滑+二阶导数	12	0.974 6	4.55	6.36	0.949 3	4.30	3.16	8.99
SNV	16	0.968 4	5.06	5.58	0.987 6	2.12	6.38	4.44

平滑、S-G 平滑+一阶导数、S-G 平滑+二阶导数、SNV。

不同预处理方法的模型比较结果如表 3 所示。综合比较各预处理方法所得模型的 RMSEC、RMSECV、RMSEP、RPD 以及 RSEP 可知,不同预处理方式所建模型性能区别不大,表明样本较为稳定,受环境温湿度、基线漂移及噪声的影响较小。无预处理的光谱所建 PLS 模型的 RMSEP 为 2.03 N, RPD 为 6.68, RSEP 为 4.24%, 模型性能最好。

**2.4.6 特征变量筛选** 由于 NIRS 中通常会包含冗余的变量,有时需要对光谱区间进行筛选以获取关键的信息。常见的变量筛选方法<sup>[10,14]</sup>有间隔偏最小二乘法(interval partial least squares, iPLS)、组合间隔偏最小二乘法(synergy interval partial least squares, siPLS)和移动窗口偏最小二乘法(moving window partial least squares, mwPLS)。合适的变量筛选方法有利于从 NIRS 中获取最有用的信息,舍弃不必要的信息,提高模型的性能。

iPLS 是将全波段划分成若干个区间,在每个划分出的区间上都建立 PLS 模型,比较每个区间 PLS 模型的性能,选出最优模型,那么该模型对应的区间即为最佳的建模区间。本研究考察了将全光谱分成 20 个区间,筛选最优区间。

siPLS 与 iPLS 相似,不同的是,它是将全区间光谱进行划分,再把划分的若干个子区间任意组合起来建立 PLS 模型,选取性能最好的模型,那么该模型对应的光谱区间组合就是最优的建模区间。本研究考察将全光谱划分为 20 个区间,再分别比较选取 2、3、4 个子区间随机组合后建立的 PLS 模型性能,筛选最优组合区间。

mwPLS 是从全光谱的第 1 个波长点开始移动,沿波长变化的方向截取适当窗口宽度的区间,建立一系列的 PLS 模型。NIRS 共有 1557 个波数点,本研究分别考察了占全光谱总波数 5% (即 77)、10% (即 155) 与 15% (即 233) 的窗口宽度,以 1 为步长进行截取,筛选最佳建模波段。

不同建模波段筛选方式对模型的影响如表 4 所示,综合比较采用各变量筛选方法所得模型的 RMSEC、RMSECV、RMSEP、RPD 以及 RSEP 可知,经过波段筛选后的模型性能均劣于全光谱建模,原因可能是 NIRS 中蕴含着较多信息,筛选后的波段丢失了部分重要信息而使样本代表性变差。采用全波长建模的模型, RMSEP 最小, RPD 最大, RSEP 最小,模型性能优于其他模型,故选取全波长进行建模,建模范围为 3 999.640~10 001.030  $\text{cm}^{-1}$ 。

**2.4.7 模型验证** 本研究采用内部验证与外部验证相结合的模型验证方法。对所建 PLS 模型进行内部验证,模型的平均相对预测误差为 2.82%,小于 3%,其中 17 个硬度不合格素片样本被全部检出,说明所建模型对素片硬度的预测准确性较高。但内部验证样本中不合格素片较少,为了进一步验证模型对不合格样本的检出能力,本研究新增 120 个样本进行外部验证,其中 75 个为自制的合格样本,外部验证的平均相对预测误差为 4.59%,与内部验证相比,效果略差一些,但仍在 5% 以内,表明模型具有较好的预测效果,其中 75 个硬度不合格样本有 73 个被检出,2 个未被检出的样本在 34.3 N 的控制限附近,2 种方法对模型的验证结果如表 5 所示。

模型校正集与验证集的相关系数分别为 0.971 9 和 0.988 7,相关性良好,线性关系见图 4 所示。

表 4 不同变量筛选方法对 PLS 模型的影响

Table 4 Impact of different variable screening methods on PLS model

筛选方法	变量区间/ $\text{cm}^{-1}$	$R_{\text{cal}}$	RMSEC/N	RMSECV/N	$R_{\text{pre}}$	RMSEP/N	RPD	RSEP/%
全光谱	3 999.640~10 001.030	0.971 9	4.78	5.29	0.988 7	2.03	6.68	4.24
iPLS	3 999.640~4 296.624	0.952 3	6.20	6.66	0.954 6	4.33	3.13	9.05
siPLS 20-2	3 999.640~4 296.624、5 203.003~5 499.987	0.966 6	5.21	5.90	0.968 6	3.39	4.00	7.10
siPLS 20-3	3 999.640~4 296.624、5 203.003~5 499.987、 6 406.366~6 703.351	0.970 3	4.91	5.46	0.977 2	2.91	4.66	6.09
siPLS 20-4	3 999.640~4 296.624、5 203.003~5 499.987、 5 804.685~6 101.669、6 406.366~6 703.351	0.972 7	4.71	5.28	0.978 6	2.84	4.77	5.94
mwPLS 5%	4 177.059~4 474.043	0.911 6	8.36	8.79	0.854 2	7.28	1.86	15.22
mwPLS 10%	4 258.055~4 855.879	0.925 3	7.71	8.15	0.920 5	5.46	2.48	11.43
mwPLS 15%	4 362.191~5 260.857	0.953 0	6.16	6.89	0.945 7	4.54	2.98	9.51

表5 PLS模型对120个验证集样本的验证结果

Table 5 Verification results of PLS model for 120 validation set samples

验证方法	样本数	平均相对预测误差/%	不合格样本数	检出率/%
内部验证	120	2.82	17	100.00
外部验证	120	4.59	75	97.33

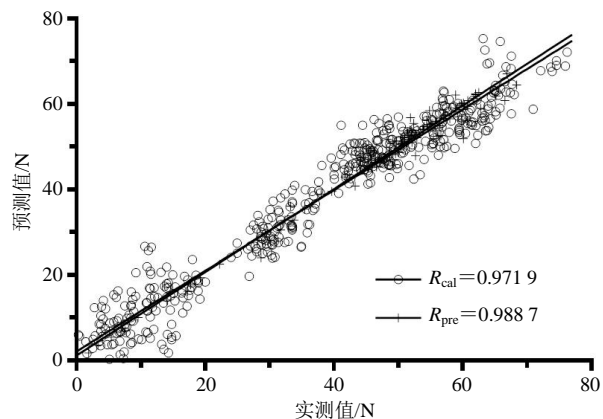


图4 校正集与验证集线性关系

Fig. 4 Linear relationship between correction set and verification set

### 3 讨论

本研究建立了一种基于NIRS分析技术预测素片硬度的方法，该方法与具有破坏性的传统方法相比，可以绿色无损的预测素片的硬度，降低了生产过程的损耗。在研究生产样本的基础上，本研究制备部分不合格素片与生产样本共同建立PLS模型，扩大了样本的分布范围，提高了模型的适用性，使其可以更好的满足生产需求，应对实际生产过程中由于压片机参数调整与装量差异等原因导致的素片硬度波动情况，协助车间人员检出不合格素片。所建PLS模型内部验证的平均相对预测误差为2.82%，外部验证的平均相对预测误差为4.59%，均小于5%，预测准确性较高，对不合格素片的检出能力在95%以上，模型具有较高的实用价值。

硬度是片剂重要的物理属性，与原辅料种类与性质、颗粒水分含量、压片机参数等密切相关，而NIRS可以深入样品内部进行分析，其吸收光谱反映大多数含氢基团的倍频与合频吸收<sup>[11]</sup>，由于药物和辅料通常含有C-H、O-H和N-H等含氢基团，在原辅料种类与比例不变的情况下，硬度大的素片往往具有更为紧密的分子排列结构，通过NIRS信息的转化表现为吸光度的差异，因而NIRS与素片硬度间存在着密切的联系，本研究所建模型良好的预

测性能也佐证了这一点，采用NIRS分析技术预测片剂硬度是可行的。

NIRS分析技术具有绿色、无损等优点，可满足中药制药生产过程中许多指标的测定，本研究将NIRS分析技术结合PLS算法应用到RGT素片硬度的测定中，解决了传统测定方法破坏性大，造成资源浪费的问题，降低了生产成本。从研究结果来看，素片硬度与NIRS存在良好的线性关系，而PLS算法可以有效建立变量内部高度线性相关的模型<sup>[16]</sup>，因而模型对素片硬度有着良好的预测效果，但研究尚处于离线检测阶段，若要将模型应用到RGT生产在线分析<sup>[18]</sup>中，还需考虑自动取样、样本传输等问题<sup>[19]</sup>，真正实现自动化。本研究为后续研究积累了样本与数据，对片剂生产过程质量控制提供了指导意义。

**利益冲突** 所有作者均声明不存在利益冲突

### 参考文献

- [1] 李集正, 卢玉平. 大株红景天治疗不稳定型心绞痛患者的疗效及对hs-CRP、血脂、血液流变学指标的影响 [J]. 云南中医中药杂志, 2020, 41(1): 40-43.
- [2] 刘超, 宗剑飞. 影响片剂溶出度的因素分析 [J]. 中国医药工业杂志, 2019, 50(3): 252-259.
- [3] 胡杨, 李先芝, 刘洋, 等. 近红外光谱技术在中药质量监控中的研究进展 [J]. 安徽农业科学, 2022, 50(1): 8-11.
- [4] Rosas J G, Blanco M, González J M, et al. Real-time determination of critical quality attributes using near-infrared spectroscopy: A contribution for process analytical technology (PAT) [J]. *Talanta*, 2012, 97: 163-170.
- [5] 张路, 冯明建, 朱海芳, 等. 近红外光谱技术在中成药质量控制中的应用进展 [J]. 中国药物评价, 2013, 30(4): 204-206.
- [6] Pauli V, Roggo Y, Kleinebudde P, et al. Real-time monitoring of particle size distribution in a continuous granulation and drying process by near infrared spectroscopy [J]. *Eur J Pharm Biopharm*, 2019, 141: 90-99.
- [7] 祁梅, 顾志荣, 李芹, 等. 近红外漫反射光谱法快速测定麦冬冬指标成分含量 [J]. 中国中医药信息杂志, 2023, 30(3): 114-120.
- [8] 张永超, 徐芳芳, 张欣, 等. 腰痹通胶囊4种中间体粒径的近红外光谱通用定量模型研究 [J]. 中草药, 2021, 52(1): 55-64.
- [9] 沈国芳, 黄俊航, 许麦成, 等. 基于近红外高光谱成像鉴别不同产地的红参 [J]. 世界中医药, 2021, 16(23):

- 3419-3423.
- [10] 刘秋安, 徐芳芳, 张欣, 等. 基于近红外光谱技术和分类与回归树算法建立天舒片崩解时间预测模型 [J]. 中草药, 2021, 52(16): 4837-4843.
- [11] 邱素君, 罗晓健, 张国松, 等. 近红外漫反射光谱无损预测片剂硬度研究 [J]. 中国药学杂志, 2016, 51(11): 904-909.
- [12] 林翔, 彭熙琳, 陈晓春, 等. 基于近红外光谱技术的辛伐他汀片剂生产过程多参数的质量监控 [J]. 四川大学学报: 工程科学版, 2015, 47(4): 192-197.
- [13] 王闽予, 刘丽萍, 朱德全, 等. 甘草配方颗粒多指标成分的近红外快速检测方法 [J]. 今日药学, 2020, 30(6): 385-393.
- [14] 夏春燕, 徐芳芳, 张欣, 等. 近红外光谱分析技术快速测定天舒片的包衣薄膜厚度 [J]. 中国实验方剂学杂志, 2020, 26(1): 118-124.
- [15] 郭云香, 陈龙, 李晓瑾, 等. 基于 NIRS 技术和 PCA-SVM 算法快速鉴别国产和进口啤酒花 [J]. 药学实践杂志, 2019, 37(4): 322-331.
- [16] 潘杰, 刘德胜, 颜贵明. 数据分析应用于中药谱效关系中的研究进展 [J]. 中医药学报, 2018, 46(4): 119-122.
- [17] 陈琪, 徐芳芳, 张欣, 等. 基于不同算法对桂枝茯苓胶囊内容物吸湿性预测建模研究 [J]. 中草药, 2021, 52(11): 3216-3223.
- [18] 何倩梅, 林坤河, 朱运锐, 等. 三七提取液中三七皂苷 R<sub>1</sub> 近红外在线检测方法构建 [J]. 大众科技, 2021, 23(3): 23-26.
- [19] 尹凡, 陈壮宇, 唐春平, 等. 基于在线近红外技术的片烟烟碱均质化系统研究 [J]. 科技风, 2020(29): 188-190.

[责任编辑 郑礼胜]