

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

张永超¹, 徐芳芳^{2,3,4*}, 张欣^{2,3,4}, 徐冰⁵, 赵月梅^{2,3,4}, 朱文灯^{2,3,4}, 吴云^{2,3,4}, 王振中^{2,3,4}, 肖伟^{1,2,3,4*}

1. 南京中医药大学, 江苏 南京 210000
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
4. 中成药智能制造国家地方联合工程研究中心, 江苏 连云港 222001
5. 北京中医药大学, 北京 100029

摘要: 目的 旨在通过近红外光谱 (NIRS) 信息与腰痹通胶囊 (Yaobitong Capsules, YC) 中间体中值粒径 (D_{50}) 的关系分析, 探讨对该品种生产过程中 4 种中间体建立 D_{50} 近红外通用定量模型的可行性。方法 采集 YC 生产过程中的原料细粉、干燥颗粒、整粒颗粒和总混颗粒 4 种中间体的 NIRS, 考察不同预处理方法对模型的影响, 并采用间隔偏最小二乘法 (iPLS)、组合间隔偏最小二乘法 (siPLS) 和移动窗口偏最小二乘法 (mwPLS) 优选 NIRS 波段, 采用偏最小二乘法 (PLS) 对 4 种中间体建立 1 个 D_{50} 通用定量模型。结果 通用模型的交叉验证均方根误差 (RMSECV) 为 3.918 μm , 预测均方根误差 (RMSEP) 为 2.832 μm , 预测相对偏差 (RSEP) 为 2.26%, 小于 5%, 性能偏差比 (RPD) 为 4.60, 大于 3, 该模型可以用于定量测定, 且与独立模型比, 预测准确性相差不大。结论 NIRS 通用定量模型可用于 YC 4 种中间体 D_{50} 的测定。

关键词: 腰痹通胶囊; 近红外光谱技术; 偏最小二乘法; 粒径; 中值粒径; 通用模型

中图分类号: R283.6 文献标志码: A 文章编号: 0253-2670(2021)01-0055-10

DOI: 10.7501/j.issn.0253-2670.2021.01.008

Research on near-infrared universal quantitative model of four intermediate particle sizes of Yaobitong Capsules

ZHANG Yong-chao¹, XU Fang-fang^{2,3,4}, ZHANG Xin^{2,3,4}, XU Bing⁵, ZHAO Yue-mei^{2,3,4}, ZHU Wen-deng^{2,3,4}, WU Yun^{2,3,4}, WANG Zhen-zhong^{2,3,4}, XIAO Wei^{1,2,3,4}

1. Nanjing University of Chinese Medicine, Nanjing 210000, China
2. Jiangsu Kanion Pharmaceutical Co., Ltd., Lianyungang 222001, China
3. State Key Laboratory of New-tech for Chinese Medicine Pharmaceutical Process, Lianyungang 222001, China
4. The National & Local Joint Engineering Research Center on the Intelligent manufacturing of TCM, Lianyungang 222001, China
5. Beijing University of Chinese Medicine, Beijing 100029, China

Abstract: Objective To explore the feasibility of establishing a D_{50} near-infrared universal quantitative model for the four intermediates in the productive process of Yaobitong Capsules (YC) by analyzing the relationship between near infrared spectrum (NIRS) information and the median particle size (D_{50}). **Methods** The NIRS of the four intermediates in the production of YC, including raw material fine powder, dry particles, sieving particles and total mixed particles, were collected. Different pretreatment methods were investigated. The interval partial least squares (iPLS), the synergy interval partial least squares (siPLS), and moving window partial least squares (mwPLS) were used to screen the optimal NIRS interval. The partial least squares (PLS) method was employed to build a D_{50} universal quantitative model for the four intermediates. **Results** In the universal model, the root mean square error of cross-validation (RMSECV) was 3.918 μm , the root mean square error of prediction (RMSEP) was 2.832 μm ; The relative standard error of prediction (RSEP) was 2.26%, which was less than 5%; The ratio of prediction to deviation (RPD) was 4.60, which

收稿日期: 2020-08-24

基金项目: 国家科技重大专项: 基于功效成分群的中药口服固体制剂先进制药与信息化技术融合示范应用 (2018ZX09201010-004); 康缘中药智能化固体制剂工厂项目 (KYYY190820)

作者简介: 张永超 (1995—), 女, 硕士研究生, 研究方向为中药制药技术与工程。E-mail: zyc020896@163.com

*通信作者: 肖伟, 研究员级高级工程师, 博士, 研究方向为中药新药的研究与开发。Tel: (025)87181851 E-mail: kanionlunwen@163.com
徐芳芳, 博士, 研究方向为过程分析技术。E-mail: 879164331@qq.com

was greater than 3. The universal model could be used for quantitative determination, and its prediction accuracy was close to that of the independent model. **Conclusion** The NIRS universal quantitative model can be used to determine the D_{50} of four intermediates in YC.

Key words: Yaobitong Capsules; near infrared spectroscopy; partial least squares; particle size; median diameter; universal model

腰痹通胶囊 (Yaobitong Capsules, YC) 是由三七、川芎、延胡索、白芍、牛膝、狗脊、熟大黄、独活 8 味中药组成, 具有活血化瘀、祛风除湿的功效, 主要用于治疗腰椎间盘突出症^[1]。颗粒粒径是评价制粒所得颗粒是否合格的关键指标^[2-3], 并对多个关键质量属性有影响, 例如吸湿稳定性^[4]、流动性^[5]、可压缩性^[6]、溶出度^[7], 它还影响制粒工艺参数的选择^[8]及胶囊的填充^[9]。在实际生产应用中, 中值粒径 (D_{50}) 是评价颗粒粒径的主要指标, D_{50} 是指粒径分布累积达到 50% 时对应的粒径值^[10-11], 它能较准确的反映样本的粒径情况。目前, YC 正在进行工艺优化提升质量标准, D_{50} 拟纳入质量内控标准。

近红外光谱 (near infrared spectrum, NIRS) 技术与化学计量法结合, 能够实现快速检测化学成分含量及物理性质指标, 在药材鉴别与化学成分含量快速测定中应用较多^[12-14], 但较少用来快速检测物性指标。粒径信息可以通过 NIRS 基线的变化反映出来^[15], 已有研究者将 NIRS 技术用于粒径的快速检测。Rosas 等^[16]以实验室自制尼美舒利为研究对象, 成功开发出 NIRS 快速检测颗粒粒径分布 ($< 125 \mu\text{m}$ 、 $125 \sim 250 \mu\text{m}$ 、 $> 250 \mu\text{m}$) 的模型, 模型的预测均方根误差 (root mean square error of prediction, RMSEP) 为 2.5%。Otsuka 等^[17]采用 NIRS 技术实时监控实验室规模的扑热息痛散流化床制粒过程, 建立了颗粒 D_{50} 的快速检测方法, 模型的校正集相关系数达 0.961。Pauli 等^[18]开发了 3 个在线 NIRS 定量分析方法, 用于测定双氯芬酸钠湿法制粒和流化床干燥过程中的颗粒粒径分布, 预测值在允许的误差范围内。潘晓宁^[19]以党参与玄参药材为研究对象, 分别构建了粒径分布累积为 10%、50%、90% 对应的粒径值 (D_{10} 、 D_{50} 、 D_{90}) 的 NIRS 定量预测模型。上述研究表明, NIRS 技术结合化学计量法是可以测定药物颗粒粒径的, 但多数研究都是针对某一类样本进行分析, 且都集中于化学药领域。通用模型是指针对某个指标建立一个模型, 可以用于分析 2 种及以上的样本, 通用性强, 可以节约较多成本。建立一个稳定、可靠的 NIRS 定量预测模型是一项较为复杂的工作, 耗时长且成本

高。如果能针对两种及以上的样本建立一个通用的模型, 将会提升检测效率, 节约成本。本研究以 YC 生产过程中的 4 种中间体为研究对象, 通过采集 4 种中间体的近红外漫反射光谱, 以中间体 D_{50} 为参考值, 采用偏最小二乘法 (partial least squares, PLS) 建立一个检测 4 种中间体 D_{50} 的通用方法。

1 仪器与材料

1.1 仪器

Antaris II 型傅里叶近红外变换光谱仪, 配有积分球漫反射采样系统、Result 光谱采集软件, 美国 Thermo 公司; BT-2600 激光粒度分布仪, 丹东百特仪器有限公司。

1.2 材料

YC 生产过程中的中间体, 包含原料细粉、干燥颗粒、整粒颗粒和总混颗粒, 由江苏康缘药业股份有限公司提供。本研究收集的样品为 2019 年 8 月至 2020 年 5 月生产的中间体样品, 收集到的中间体样品批次为 Z190801~Z190820、Z190901~Z190918、Z200301~Z200320、Z200401~Z200420、Z200501~Z200502, 包含 4 种中间体共 320 个样品。

2 方法

2.1 NIRS 采集

取 4 g 左右样品, 置于配备的样品杯里, 轻轻压实, 采用积分球漫反射方式采集 NIRS。扫描范围为 $10\ 000 \sim 4000 \text{ cm}^{-1}$, 分辨率为 8 cm^{-1} , 2 倍增益, 扫描次数 64 次, 以空气为背景, 每小时扫描 1 次背景。每个样品扫描 3 次, 平均值用于分析。

2.2 参考值测定

激光粒度分布仪法测粒径代表性强且适用于粒径小的颗粒^[10], “1.2” 项下的样本粒径小于 $550 \mu\text{m}$, 采用该法测粒径结果较为准确。取适量样品置于进料斗中, 使用 BT-2600 激光粒度分布仪测定样品的粒径分布, 并计算 D_{50} (粒径分布累积为 50% 时对应的粒径值)。参数设置: 分散介质为空气, 空气压力为 0.25 MPa, 遮光率为 3%~12%。

2.3 光谱预处理方法

采集光谱时, 由于受到环境温湿度、仪器状态以及颗粒状态等各种因素的影响, NIRS 会产生噪声信号、基线漂移等, 获取的光谱中包含了自身的

信息以及其他不必要的信号。对 NIRS 进行合适的预处理,可以减少噪声,滤过无关信息,提高模型稳健性。常见的预处理方法有矢量归一化法、导数法(一阶导数、二阶导数)、平滑法[Savitzky-Golay (SG)平滑、Norris Derivative (ND)平滑]、多元散射校正(multiplicative scatter correction, MSC)、标准正态变量变换法(standard normal variate transformation, SNV)、基线校正以及上述几种方法的结合。矢量归一化法能增强光谱差异,可以校正由光程或样品稀释等导致的光谱变化;MSC和SNV可以消除颗粒大小、分布不均带来的干扰;导数法可以消除基线漂移;平滑法能够滤除噪声^[20-21]。

2.4 数据处理与评价方法

采用 Unscramble X 10.4 (Camo software AS, Norway) 软件对 NIRS 进行预处理,采用 Matlab 2016a (Mathwork Inc., USA) 软件对样本进行变量筛选与建模,采用留一交叉验证法以交叉验证均方根误差(root mean square errors of cross validation, RMSECV)为评价指标确定主因子数。以中间体 D_{50} 为因变量,以对应的 NIRS 值为自变量建立 PLS 定量模型。本研究以下列参数来评价模型性能,以优选最佳模型。校正集相关系数(R_{cal})、验证集相关系数(R_{pre})分别表示校正模型与验证模型的拟合

程度, R_{cal} 和 R_{pre} 越大,模型拟合效果越好;校正均方根误差(root mean square errors of calibration, RMSEC)、RMSEP 分别指校正模型与验证模型中参考值与预测值之间的均方根误差,越小模型预测性能越高;校正集偏差($BIAS_{cal}$)和验证集偏差($BIAS_{pre}$)分别表示校正模型与验证模型中参考值与预测值之间的偏差,其越小,模型预测精度越高^[22-23]。预测相对偏差(relative standard error of prediction, RSEP)是模型中参考值与预测值之间的相对偏差,一般 RSEP 越小模型预测性能越好;性能偏差比(ratio of performance to deviation, RPD)表示模型预测性能,是标准差(SD)与 RMSEP 的比值,当 $RPD > 3$ 时,表示模型预测精度高^[24]。各项评价参数不能孤立参考,需要综合起来评价。

3 结果与分析

3.1 样本划分

采用随机抽样(random sampling, RS)法划分校正集与验证集,并保证验证集中参考值范围包含于校正集中^[12]。每种中间体按照 3:1 的比例划分校正集与验证集,随机选取 60 个样品作为校正集,20 个样品作为验证集;通用模型的校正集为 4 种中间体校正集的总和,共 240 个样品,验证集为 4 种中间体验证集的总和,共 80 个样品,划分结果见表 1。

表 1 样品校正集与验证集参考方法测定值的统计结果

Table 1 Reference measurement results of sample calibration and prediction set

中间体	校正集			验证集		
	样本数	参考值范围/ μm	平均值/ μm	样本数	参考值范围/ μm	平均值/ μm
原料细粉	60	90.56~140.53	112.10	20	92.24~135.87	110.38
干燥颗粒	60	115.77~152.07	131.05	20	116.17~151.17	129.61
整粒颗粒	60	112.00~154.40	131.39	20	114.20~152.33	130.06
总混颗粒	60	107.37~151.53	131.86	20	114.33~148.30	128.50
总样品	240	90.56~154.40	126.60	80	92.24~152.33	124.64

3.2 光谱预处理方法的选择

YC 中间体的近红外原始光谱见图 1。本研究考察了以下预处理方法:移动窗口平滑(9 点)、SNV、S-G 1st(9 点)、基线校正、MSC、归一化法。使用上述方法对光谱进行预处理后,建立 D_{50} 的 PLS 定量模型,不同预处理方法对模型性能的影响如表 2 所示。以 RPD 与 RSEP 为评价标准,筛选最优的预处理方法。由表 2 可知,原料细粉模型中,一阶求导结合 S-G 平滑对光谱进行预处理建模效果最优,RPD 为 4.50, RSEP 为 2.30%;干燥颗粒模型中,

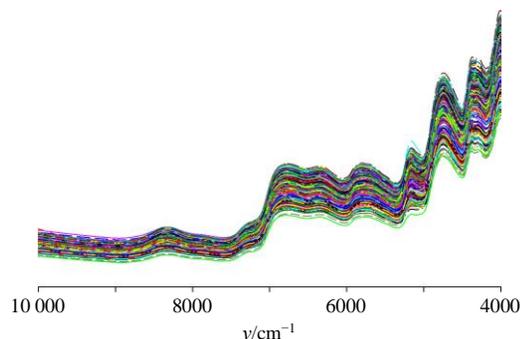


图 1 YC 中间体近红外原始光谱

Fig. 1 Original NIRS of YC intermediates

表2 不同预处理方法对模型的影响

Table 2 Effects of different pretreatment methods on model

项目	预处理方法	校正集				验证集				
		R_{cal}	RMSEC/ μm	RMSECV/ μm	BIAS $_{cal}$	R_{pre}	RMSEP/ μm	BIAS $_{pre}$	RPD	RSEP/%
原料细粉模型	无预处理	0.939	4.099	4.608	3.464	0.979	2.598	2.134	4.41	2.34
	移动窗口平滑	0.939	4.102	4.611	3.466	0.979	2.599	2.136	4.41	2.34
	SNV	0.969	2.925	5.811	2.418	0.923	4.317	3.628	2.66	3.89
	S-G 1 st	0.952	3.635	4.756	2.964	0.974	2.548	2.063	4.50	2.30
	基线校正	0.933	4.257	4.621	3.579	0.972	2.684	2.148	4.27	2.42
	MSC	0.969	2.922	5.805	2.418	0.923	4.310	3.620	2.66	3.88
	归一化法	0.922	4.586	5.982	3.716	0.952	3.482	3.038	3.29	3.14
干燥颗粒模型	无预处理	0.943	2.944	3.748	2.553	0.972	2.452	2.053	4.11	1.89
	移动窗口平滑	0.942	2.970	3.767	2.576	0.972	2.441	2.025	4.12	1.88
	SNV	0.918	3.497	4.322	2.841	0.965	2.689	2.232	3.75	1.88
	S-G 1 st	0.957	2.579	5.022	2.126	0.949	3.169	2.435	3.18	2.44
	基线校正	0.932	3.225	5.032	2.520	0.959	2.816	2.267	3.58	2.17
	MSC	0.918	3.494	4.317	2.838	0.965	2.684	2.218	3.75	2.07
	归一化法	0.964	2.309	4.512	1.845	0.961	2.745	2.317	3.67	2.11
整粒颗粒模型	无预处理	0.970	2.110	2.572	1.678	0.977	2.342	1.970	4.15	1.80
	移动窗口平滑	0.970	2.112	2.570	1.679	0.977	2.344	1.972	4.15	1.80
	SNV	0.970	2.112	2.570	1.679	0.977	2.343	1.972	4.15	1.80
	S-G 1 st	0.944	3.203	3.911	2.563	0.960	2.750	2.053	3.71	2.01
	基线校正	0.961	2.470	2.973	1.884	0.966	2.624	2.132	3.71	2.01
	MSC	0.974	1.955	4.413	1.613	0.939	3.443	2.838	2.82	2.64
	归一化法	0.972	2.027	3.761	1.499	0.943	3.083	2.708	3.15	2.36
总混颗粒模型	无预处理	0.970	2.405	2.935	1.998	0.978	2.027	1.602	4.74	1.57
	移动窗口平滑	0.974	2.212	2.880	1.850	0.978	1.986	1.642	4.84	1.54
	SNV	0.967	2.499	3.760	2.015	0.959	2.705	1.990	3.55	2.10
	S-G 1 st	0.912	4.022	4.322	3.100	0.977	2.280	1.807	4.20	1.77
	基线校正	0.965	2.586	3.109	2.065	0.974	2.144	1.693	4.48	1.66
	MSC	0.967	2.499	3.761	2.014	0.959	2.713	2.000	3.54	2.11
	归一化法	0.943	3.281	3.913	2.662	0.957	2.743	2.384	3.50	2.13
通用模型	无预处理	0.965	3.392	3.918	2.666	0.977	2.832	2.309	4.60	2.26
	移动窗口平滑	0.963	3.472	3.972	2.738	0.976	2.839	2.331	4.53	2.30
	SNV	0.963	3.473	3.968	2.738	0.976	2.876	2.331	4.53	2.30
	S-G 1 st	0.963	3.545	4.216	2.907	0.974	2.978	2.417	4.38	2.38
	基线校正	0.960	3.680	4.298	2.892	0.976	2.845	2.308	4.58	2.27
	MSC	0.955	3.832	4.495	3.043	0.950	4.030	3.268	3.24	3.22
	归一化法	0.965	3.379	4.036	2.626	0.969	3.201	2.534	4.07	2.55

采用移动窗口平滑（9点）预处理方法最佳，RPD为4.12，RSEP为1.88%；整粒颗粒模型里，采用原始光谱建模比进行预处理后建模效果好，RPD为4.15，RSEP为1.80%；总混颗粒模型采用移动窗口

平滑（9点）方法对光谱进行预处理后建模效果最好，RPD为4.84，RSEP为1.54%；通用模型中，采用原始光谱建模性能最佳，RPD为4.60，RSEP为2.26%。

3.3 特征变量筛选

筛选特征变量可以剔除无关信息，提高模型性能。本研究在上述筛选出的最佳预处理方法基础上进一步筛选特征变量。

3.3.1 基于间隔偏最小二乘法 (interval PLS, iPLS) 筛选变量 iPLS 是把全光谱划分成若干个子区间，然后在每个子区间进行建模^[25]。本研究将全光谱划分成 20 个子区间，以 RMSECV 为评价指标，筛选最佳建模波段。

3.3.2 基于组合间隔偏最小二乘法 (synergy interval PLS, siPLS) 筛选变量 siPLS 是基于 iPLS，它是将全光谱划分成若干个子区间后，再把子区间任意组合起来建模。本研究是把全光谱划分成 20 个子区间，再以子区间组合数为 4 建立模型，以 RMSECV 为评价指标，筛选最佳建模波段。

3.3.3 基于移动窗口偏最小二乘法 (moving

window PLS, mwPLS) 筛选变量 mwPLS 是从整个光谱的第一个波长点开始移动，沿波长变化的方向截取选定窗口宽度的区间，建立一系列的 PLS 模型^[26]。本研究以初始窗口宽度为 31，以 10 为步长依次增加窗口宽度，建立了窗口宽度为 31~311 的 PLS 模型，并根据 RMSECV 选取最佳建模波段。

3.3.4 筛选特征变量 采用 3 种方法建立模型的性能参数如表 3~7 所示。分别以 RMSEC、RMSECV、RMSEP 为评价指标，综合评价筛选出建模的最佳波段。

原料细粉模型采用全光谱建模与 mwPLS 优选的波段建模性能相差不大，但是筛选波段后变量减少至 115 个，建模时间会缩短，因此选择建模波段 3 999.64~4 018.92、5 739.12~6 136.38 cm^{-1} 。

干燥颗粒模型采用 siPLS 和 mwPLS 筛选变量后，模型性能均有所提高，变量数分别减少到 392、

表 3 原料细粉优选区间建模与全光谱建模的比较

Table 3 Comparison of optimal interval modeling and full spectrum modeling of raw material fine powder

筛选方法	变量区间/ cm^{-1}	R_{cal}	RMSEC/ μm	RMSECV/ μm	R_{pre}	RMSEP/ μm	RPD
全光谱	3 999.64~10 001.03	0.952	3.635	4.756	0.974	2.548	4.50
siPLS	3 999.64~4 296.62、4 902.16~5 199.15、 6 406.37~6 703.35、6 707.21~7 004.19	0.948	3.776	4.464	0.972	2.680	4.28
iPLS	3 999.64~4 296.62	0.916	4.761	5.221	0.944	3.708	3.09
mwPLS	3 999.64~4 018.92、5 739.12~6 136.38	0.941	4.010	4.934	0.975	2.542	4.51

表 4 干燥颗粒优选区间建模与全光谱建模的比较

Table 4 Comparison of optimal interval modeling and full spectrum modeling of dry particles

筛选方法	变量区间/ cm^{-1}	R_{cal}	RMSEC/ μm	RMSECV/ μm	R_{pre}	RMSEP/ μm	RPD
全光谱	3 999.64~10 001.03	0.942	2.970	3.767	0.972	2.441	4.12
siPLS	4 601.32~4 898.31、6 406.37~6 703.35、 7 008.05~7 305.03、8 813.09~9 110.08	0.975	1.953	3.231	0.970	2.404	4.19
iPLS	8 512.25~8 809.24	0.922	3.413	4.063	0.921	4.276	2.35
mwPLS	4 030.50~4 219.49、4 443.19~4 504.90、 6 957.91~7 767.86	0.966	2.261	3.603	0.979	2.214	4.55

表 5 整粒颗粒优选区间建模与全光谱建模的比较

Table 5 Comparison of optimal interval modeling and full spectrum modeling of sieving particles

筛选方法	变量区间/ cm^{-1}	R_{cal}	RMSEC/ μm	RMSECV/ μm	R_{pre}	RMSEP/ μm	RPD
全光谱	3 999.64~10 001.03	0.970	2.110	2.572	0.977	2.342	4.15
siPLS	3 999.64~4 296.62、5 203.00~5 499.99、 5 503.84~5 800.83、8 813.09~9 110.08	0.959	2.448	3.068	0.974	2.313	4.20
iPLS	3 999.64~4 296.62	0.927	3.275	3.785	0.948	3.014	3.23
mwPLS	4 933.02~4 971.59、5 982.10~6 213.52、 7 359.03~7 397.60	0.976	1.905	3.158	0.975	2.163	4.49

表6 总混颗粒优选区间建模与全光谱建模的比较

Table 6 Comparison of optimal interval modeling and full spectrum modeling of total mixed particles

筛选方法	变量区间/cm ⁻¹	<i>R</i> _{cal}	RMSEC/μm	RMSECV/μm	<i>R</i> _{pre}	RMSEP/μm	RPD
全光谱	3 999.64~10 001.03	0.974	2.212	2.880	0.978	1.986	4.84
siPLS	4 300.48~4 597.46、5 203.00~5 499.99、 7 910.57~8 207.55、8 813.09~9 110.08	0.969	2.422	3.062	0.978	2.029	4.73
iPLS	3 999.64~4 296.62	0.965	2.569	3.142	0.962	2.859	3.36
mwPLS	4 015.07~4 242.63、4 396.90~4 524.18、 7 667.58~7 698.44、8 369.54~8 454.40	0.977	2.073	2.692	0.980	1.870	5.13

表7 通用模型优选区间建模与全光谱建模的比较

Table 7 Comparison of optimal interval modeling and full spectrum modeling of universal model

筛选方法	变量区间/cm ⁻¹	<i>R</i> _{cal}	RMSEC/μm	RMSECV/μm	<i>R</i> _{pre}	RMSEP/μm	RPD
全光谱	3 999.64~10 001.03	0.965	3.392	3.918	0.977	2.832	4.60
siPLS	4 300.48~4 597.46、5 203.00~5 499.99、 3 308.89~7 605.87、7 910.57~8 207.55	0.964	3.450	3.793	0.973	3.034	4.30
iPLS	5 203.00~5 499.99	0.963	3.496	4.285	0.953	4.000	3.26
mwPLS	4 686.17~4 740.17、4 855.88~5 488.42	0.966	3.339	3.887	0.975	2.924	4.46

278个,但mwPLS筛选变量后的波点数更少且*R*_{cal}、*R*_{pre}较大, RMSEC、RMSEP较小,故采用波段4 030.50~4 219.49、4 443.19~4 504.90、6 957.91~7 767.86 cm⁻¹建模。

整粒颗粒模型采用siPLS与mwPLS筛选变量模型性能有所提升,变量数分别减少至312、83个,但mwPLS的*R*_{cal}、*R*_{pre}更大, RMSEC、RMSEP也较小,且波点数也较少,故选用波段4 933.02~4 971.59、5 982.10~6 213.52、7 359.03~7 397.60 cm⁻¹建模。

总混颗粒模型采用mwPLS法进行波段筛选后,*R*_{cal}、*R*_{pre}均变大, RMSEC、RMSEP均变小,模型性能提高,因此选用波段4 015.07~4 234.26、4 396.90~4 524.18、7 667.58~7 698.44、8 369.54~8 454.40 cm⁻¹建模。

通用模型筛选变量后虽然降低了模型复杂程度,但是模型性能却不及全光谱模型,可能是因为通用模型是包含4个中间体信息的共用模型,采用波段筛选可能会使重要信息的流失,造成模型性能降低,故通用模型采用全波段建模。

3.4 主因子数的确定

以RMSECV值为评价指标确定主因子数,主因子数选择不当会导致拟合不充分或拟合过度,一般以RMSECV值最小时对应的主因子数为最佳。本研究采用留一交叉验证法,以RMSECV为指标,

考察主因子数对模型的影响。最终确定原料细粉模型、干燥颗粒模型、整粒颗粒模型、总混颗粒模型和通用模型的最佳主因子数分别为6、15、12、11、14。

3.5 *D*₅₀定量预测模型的建立

采用筛选出来的最佳预处理方法及最优NIRS波段建立*D*₅₀近红外定量预测模型,见图2。优选出来的原料细粉、干燥颗粒、整粒颗粒、总混颗粒和通用模型的评价参数见表8。各项模型的*R*_{cal}、*R*_{pre}接近于1,说明参考值与预测值相关性较高; RMSEC、RMSECV、RMSEP、RSEP较小, RPD大于3,说明模型预测性能较高,可以用于*D*₅₀的定量预测。

3.6 模型验证与评价

将验证集样品的NIRS导入校正模型中,预测中间体的*D*₅₀,并与参考值进行比较,结果见表9~10。原料细粉模型、干燥颗粒模型、整粒颗粒模型、总混颗粒模型及通用模型*D*₅₀的平均相对预测误差分别为1.94%、1.43%、1.46%、1.21%、1.87%,均小于5%,说明2种模型的预测准确性均较高。由表11可知,独立模型与通用模型对各中间体*D*₅₀的预测误差的差值均小于1%,说明二者的预测性能相近。

对原料细粉、干燥颗粒、整粒颗粒、总混颗粒和通用模型的验证集中参考值与预测值进行配对 *t*

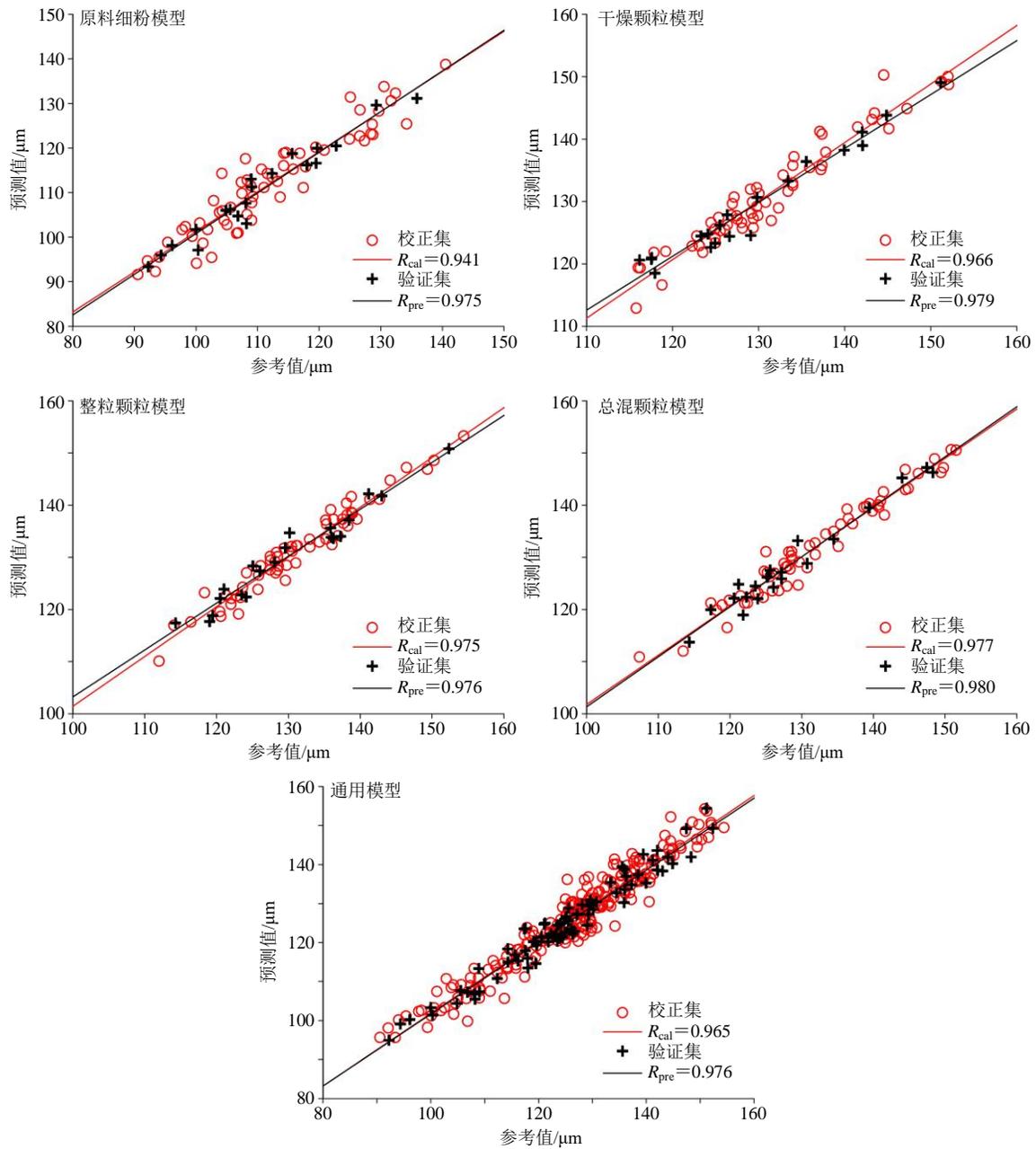


图 2 D_{50} 预测值与参考值的相关性

Fig. 2 Correlation between predicted value and reference value of D_{50}

表 8 最佳模型的评价参数

Table 8 Evaluation parameters of best model

项目	R_{cal}	R_{pre}	RMSEC/ μm	RMSECV/ μm	RMSEP/ μm	RSEP/%	RPD
原料细粉模型	0.941	0.975	4.010	4.934	2.542	2.29	4.51
干燥颗粒模型	0.966	0.979	2.261	3.603	2.214	1.70	4.55
整粒颗粒模型	0.976	0.975	1.905	3.158	2.163	1.66	4.49
总混颗粒模型	0.977	0.980	2.073	2.692	1.870	1.45	5.13
通用模型	0.965	0.977	3.392	3.918	2.832	2.26	4.60

检验, 结果 t 值依次为 0.22、0.17、-0.42、-0.38、0.75, P 值依次为 0.825、0.866、0.678、0.706、0.454,

P 值均 >0.05 , 说明参考值与预测值之间无显著性差异。

表9 NIRS 参考值与预测值的比较 (独立模型)

Table 9 Comparison of NIRS reference value and predicted value (independent model)

编号	原料细粉			干燥颗粒			整粒颗粒			总混颗粒		
	参考值	预测值	相对误差/%									
1	115.63	118.76	2.71	126.33	127.86	1.21	136.00	133.86	1.58	130.77	128.81	1.50
2	104.87	105.97	1.05	116.17	120.60	3.82	120.50	122.10	1.33	122.30	122.36	0.05
3	135.87	131.14	3.48	142.03	141.14	0.63	130.17	134.71	3.49	117.33	119.95	2.23
4	122.77	120.47	1.87	144.87	143.83	0.71	141.17	142.20	0.73	148.30	146.29	1.36
5	105.57	106.19	0.58	139.93	138.24	1.21	143.00	141.79	0.84	134.47	133.51	0.71
6	96.11	98.12	2.09	117.57	120.98	2.90	125.07	128.32	2.60	120.57	122.17	1.32
7	92.24	93.35	1.20	126.60	124.40	1.74	123.50	122.79	0.58	123.53	124.49	0.77
8	106.80	104.72	1.95	124.43	122.63	1.45	129.60	131.84	1.73	129.43	133.20	2.91
9	99.98	101.72	1.74	133.40	133.29	0.08	138.47	137.17	0.94	121.83	118.92	2.39
10	108.10	107.60	0.46	117.93	118.50	0.48	152.33	150.83	0.99	125.60	127.52	1.53
11	109.03	111.23	2.02	125.50	126.21	0.57	121.03	123.88	2.35	114.33	113.73	0.52
12	108.20	103.02	4.79	129.07	124.54	3.51	136.27	133.61	1.95	147.43	147.23	0.14
13	100.33	97.09	3.23	117.53	120.73	2.72	114.30	117.38	2.70	144.00	145.22	0.85
14	108.93	113.02	3.75	135.53	136.41	0.65	124.13	122.39	1.40	123.87	122.06	1.46
15	94.34	95.96	1.72	124.10	124.77	0.54	128.10	129.05	0.74	121.20	124.84	3.00
16	118.03	116.15	1.59	142.07	138.97	2.18	126.03	127.37	1.06	126.07	124.23	1.46
17	112.33	114.28	1.73	129.80	130.63	0.64	119.03	117.68	1.13	125.17	126.19	0.81
18	129.27	129.63	0.28	123.30	124.46	0.94	135.87	135.64	0.16	127.13	127.11	0.02
19	119.50	116.61	2.42	124.93	123.27	1.33	119.47	118.83	0.54	127.17	125.87	1.02
20	119.67	119.93	0.22	151.17	149.07	1.39	137.27	134.03	2.36	139.43	139.51	0.06

表10 NIRS 参考值与预测值的比较 (通用模型)

Table 10 Comparison of NIRS reference value and predicted value (universal model)

编号	原料细粉			干燥颗粒			整粒颗粒			总混颗粒		
	参考值	预测值	相对误差/%									
1	115.63	116.80	1.01	126.33	122.14	3.32	136.00	133.63	1.74	130.77	130.46	0.23
2	104.87	104.35	0.50	116.17	115.34	0.72	120.50	121.29	0.66	122.30	122.26	0.04
3	135.87	130.28	4.11	142.03	143.61	1.11	130.17	128.50	1.28	117.33	123.54	5.29
4	122.77	121.60	0.95	144.87	140.23	3.20	141.17	141.17	0.00	148.30	141.95	4.28
5	105.57	107.66	1.98	139.93	135.26	3.34	143.00	138.34	3.26	134.47	132.76	1.27
6	96.11	100.27	4.33	117.57	123.59	5.12	125.07	126.17	0.88	120.57	121.47	0.75
7	92.24	94.95	2.93	126.60	122.89	2.93	123.50	120.39	2.52	123.53	124.45	0.75
8	106.80	107.26	0.43	124.43	121.31	2.51	129.60	130.89	1.00	129.43	129.40	0.02
9	99.98	103.31	3.32	133.40	135.40	1.50	138.47	137.45	0.74	121.83	120.19	1.35
10	108.10	106.77	1.23	117.93	115.92	1.70	152.33	149.28	2.00	125.60	128.80	2.55
11	109.03	107.36	1.53	125.50	125.53	0.02	121.03	124.76	3.08	114.33	114.97	0.55
12	108.20	105.45	2.54	129.07	124.44	3.58	136.27	136.99	0.53	147.43	149.22	1.21
13	100.33	101.40	1.06	117.53	117.91	0.32	114.30	118.36	3.55	144.00	141.91	1.45
14	108.93	113.31	4.02	135.53	139.47	2.90	124.13	121.96	1.75	123.87	121.05	2.27
15	94.34	99.11	5.05	124.10	124.91	0.65	128.10	129.70	1.25	121.20	124.96	3.10
16	118.03	113.48	3.86	142.07	138.57	2.46	126.03	123.65	1.89	126.07	123.10	2.36
17	112.33	110.76	1.40	129.80	130.12	0.25	119.03	120.18	0.97	125.17	126.66	1.19
18	129.27	127.20	1.60	123.30	122.17	0.92	135.87	138.80	2.16	127.13	127.15	0.02
19	119.50	114.60	4.10	124.93	122.67	1.81	119.47	120.20	0.61	127.17	127.34	0.13
20	119.67	119.38	0.24	151.17	154.40	2.14	137.27	134.81	1.79	139.43	142.63	2.29

表 11 独立模型与通用模型相对预测误差比较

Table 11 Comparison of relative prediction error between independent model and universal model

中间体	平均相对预测误差/%		差值/%
	独立模型	通用模型	
原料细粉	1.94	2.31	0.37
干燥颗粒	1.43	2.03	0.60
整粒颗粒	1.46	1.58	0.12
总混颗粒	1.21	1.56	0.35

4 讨论

本研究以 YC 为研究对象, 收集其生产过程中的原料细粉、干燥颗粒、整粒颗粒及总混颗粒共 4 种中间体, 采用 NIRS 技术结合 PLS 建立了预测中间体 D_{50} 的通用定量模型。各项评价指标表明, 通用定量模型预测准确性高、能满足实际需求, 且与单一模型比预测性能差异不大, 可以实现对 4 种中间体 D_{50} 的测定。

本研究中的通用模型是基于一个品种中差异不大的中间体建立起来的, 通用模型预测效果较好, 可能是因为中间体的 D_{50} 有一定差异但还在可接受范围内, 且由于这种差异使得 D_{50} 的分布范围变宽, 更利于建模, 但对不同品种建立相关指标的通用模型的可行性还有待探索。相比激光粒度分布仪法, NIRS 法测粒径会更快, 若应用于在线研究将无需取样, 提高分析效率。目前, NIRS 通用模型在制药领域研究较少, 若将 NIRS 技术与通用模型结合起来, 建立的 NIRS 通用模型将具备快速准确、无损绿色、节约成本等优点, 可以为中药生产过程实现在线检测相关指标提供更好的方法。

本研究是初步证明了 YC 4 种中间体近红外 D_{50} 通用定量模型的可行性, 后续完善该模型还应当加入更多的样本, 以及后续生产的持续验证, 并朝着在线监测方向上研究。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 中国药典 [S]. 一部. 2015.
 [2] 臧恒昌, 臧立轩, 张惠, 等. 近红外光谱分析技术在制药领域中的应用研究进展 [J]. 药学研究, 2014, 33(3): 125-128.
 [3] 姜爽, 杨海龙, 臧恒昌. 近红外光谱分析技术在中药生产过程质量控制领域的应用 [J]. 食品与药品, 2013, 15(2): 135-138.
 [4] 李琼, 赵磊, 王权. 不同粒度黄芪三七胶囊粉体学性质

及质量的初步对比 [J]. 现代中药研究与实践, 2019, 33(6): 32-35.
 [5] Goh H P, Heng P W S, Liew C V. Comparative evaluation of powder flow parameters with reference to particle size and shape [J]. *Int J Pharm*, 2018, 547(1/2): 133-141.
 [6] Fu X W, Huck D, Makein L, et al. Effect of particle shape and size on flow properties of lactose powders [J]. *Particuology*, 2012, 10(2): 203-208.
 [7] 梁兆昌, 褚洪标, 肖琳, 等. 杜仲超微粉体理化特性及体外溶出性能研究 [J]. 中草药, 2015, 46(11): 1609-1614.
 [8] 李淞明, 周永全, 王玲, 等. 中药浸膏粉物理性质及其表征方法的研究进展 [J]. 中成药, 2015, 37(11): 2491-2494.
 [9] Faulhammer E, Llusà M, Radeke C, et al. The effects of material attributes on capsule fill weight and weight variability in dosator nozzle machines [J]. *Int J Pharm*, 2014, 471(1/2): 332-338.
 [10] 许俊男, 涂传智, 陈颖翀, 等. 激光粒度测定法在中药粉体粒径测定中的应用与思考 [J]. 世界科学技术—中医药现代化, 2016, 18(10): 1776-1781.
 [11] 宋晓娇, 夏俊, 陈峰, 等. 粒度检测方法研究及其在口服吸入制剂研发中的应用 [J]. 药物评价研究, 2019, 42(12): 2319-2324.
 [12] 雷晓晴, 王秀丽, 李耿, 等. 近红外光谱法快速测定当归中 7 种成分的含量 [J]. 中草药, 2019, 50(16): 3947-3954.
 [13] 王晴, 徐芳芳, 张欣, 等. 在线近红外光谱监测桂枝茯苓胶囊流化床干燥过程水分的方法研究 [J]. 中草药, 2019, 50(22): 5429-5438.
 [14] 高鸿彬, 瞿敏, 刘浩, 等. 不同产地覆盆子的近红外漫反射快速无损鉴别 [J]. 世界中医药, 2020, 15(10): 1386-1390.
 [15] Liu R, Li L, Yin W, et al. Near-infrared spectroscopy monitoring and control of the fluidized bed granulation and coating processes-A review [J]. *Int J Pharm*, 2017, 530(1/2): 308-315.
 [16] 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.
 [17] Otsuka M, Koyama A, Hattori Y. Real-time release monitoring for water content and mean particle size of granules in lab-sized fluid-bed granulator by near-infrared spectroscopy [J]. *RSC Adv*, 2014, 4(34): 17461.
 [18] 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.
- [19] 潘晓宁. 基于 Kubelka-Munk 理论的中药近红外漫反射定量研究 [D]. 北京: 北京中医药大学, 2016.
- [20] 范剑. 近红外光谱技术在干姜和桂枝混合蒸馏提取过程中的应用研究 [D]. 济南: 山东大学, 2017.
- [21] 杨越. 基于近红外光谱技术的中药生产过程质量控制方法研究 [D]. 杭州: 浙江大学, 2018.
- [22] 李春燕. 近红外光谱法在葛根活性成分含量快速测定中的应用研究 [D]. 新乡: 新乡医学院, 2018.
- [23] 王永香, 郑伟然, 米慧娟, 等. 热毒宁注射液青蒿金银花浓缩过程近红外快速定量检测方法的建立 [J]. *中草药*, 2017, 48(1): 102-108.
- [24] 张娜, 徐冰, 贾帅芸, 等. 丹参提取过程多源信息融合建模方法研究 [J]. *中草药*, 2018, 49(6): 1304-1310.
- [25] 吴静珠, 石瑞杰, 陈岩, 等. 食用油油酸的近红外特征谱区优选 [J]. *中国粮油学报*, 2015, 30(2): 118-121.
- [26] 陈玥洋, 高志山, 郁晓晖, 等. 基于区间组合移动窗口法筛选近红外光谱信息 [J]. *应用光学*, 2017, 38(1): 99-105.

[责任编辑 郑礼胜]