

# 基于大数据与模式识别技术对中药材品质的快速评价研究

张立军<sup>1</sup>, 张新玥<sup>2</sup>, 马冬妮<sup>3</sup>, 张转平<sup>1\*</sup>, 谭艳萍<sup>1</sup>, 钟海婷<sup>1</sup>, 熊贤锋<sup>1</sup>, 刘丽<sup>1</sup>

<sup>1</sup>陕西省安康市食品药品检验检测中心, 安康 725000;

<sup>2</sup>北京中医药大学中药学院, 北京 100029; <sup>3</sup>甘肃省敦煌市医院, 敦煌 736200

**摘要:** 基于大数据背景下, 通过多项指标的快速检测与整合分析, 建立北柴胡药材品质综合评价模型与等级划分的评价方法。收集本中心 2015 年~2020 年抽检样品中采用《中华人民共和国药典》(以下简称《中国药典》) 2015 年版检测的 130 批北柴胡药材各项检测数据; 通过近红外光谱(NIRS)技术采集各批药材对应的光谱图, 采用模式识别技术建立北柴胡快速溯源分析模型, 同时构建各检测指标与光谱之间的拟合模型, 开展多批次北柴胡药材的系统检测; 采用数学建模方法构建北柴胡药材品质综合评价指数( $F_q$ )计算方法, 建立其品质综合评价与等级划分数数据库。建立了北柴胡药材溯源分析快速定性鉴别模型; 建立了北柴胡药材水分、灰分、酸不溶性灰分、浸出物、柴胡皂苷 a 和 d 总量及茎秆(地上部分)占比 6 项指标的近红外预测模型, 应用模型完成了对 20 批次北柴胡药材的快速检测; 建立了北柴胡品质快速评价与等级划分数数据库。该方法基于大数据背景下利用现行《中国药典》监管体系产生的原始数据实现了对北柴胡药材品质的快速综合评价和等级划分, 为中药材科学监管研究提供了新的解决方法。

**关键词:** 大数据; 品质综合评价指数; 北柴胡; 近红外光谱技术; 系统建模; 等级划分

中图分类号: R286

文献标识码: A

文章编号: 1001-6880(2023)4-0630-10

DOI: 10.16333/j.1001-6880.2023.4.010

## Study on rapid evaluation of Chinese herbal medicine quality based on big data and pattern recognition technology

ZHANG Li-jun<sup>1</sup>, ZHANG Xin-yue<sup>2</sup>, MA Dong-ni<sup>3</sup>,

ZHANG Zhuan-ping<sup>1\*</sup>, TAN Yan-ping<sup>1</sup>, ZHONG Hai-ting<sup>1</sup>, XIONG Xian-feng<sup>1</sup>, LIU Li<sup>1</sup>

<sup>1</sup>Ankang Inspection and Detection Center of Food and Drug Control, Ankang 725000, China;

<sup>2</sup>Beijing University of Chinese Medicine, Beijing 100029, China;

<sup>3</sup>Dunhuang Hospital of Gansu Province, Dunhuang 736200, China

**Abstract:** Based on the background of big data, through the integration analysis of several testing indexes, an integrated evaluation method of quality evaluation and a grade identification system of the roots of *Bupleurum chinense* DC. (BCD) were developed. Collected the test data of 130 batches of the roots of BCD detected by the 2015 edition of "Chinese Pharmacopoeia" in the random inspection samples of our center from 2015 to 2020. The near-infrared spectroscopy (NIRS) technology were collected the spectra corresponding to each batch of samples. The pattern recognition technology was used to establish a fast traceability analysis model of samples, and developed to establish fitting model between the NIRS and the content with detection index of samples respectively. Carry out the system detection of multiple batches of the roots of BCD samples. The mathematical modeling method was used to construct the calculation method of the quality comprehensive evaluation index factor ( $F_q$ ), and established the comprehensive evaluation and grade division database of medicinal materials. The novel rapid qualitative analysis the roots of BCD identification model were established. The NIRS model of 6 indexes including water, total ash, acid-insoluble ash, extract, the total amount of saikosaponins a and d, and the proportion of stem of BCD were established, and then applied to detect of 20 of batches of the roots of BCD samples. Established a database for rapid quality evaluation and grade division of the roots of BCD samples. This method is based on the background of big data and uses the original

data generated by the current "Chinese Pharmacopoeia" supervision system to realize the rapid comprehensive evaluation and grading of the quality of the roots of BCD samples, and provides a new solution for the scientific supervision and research of Chinese herbal medicine.

**Key words:** big data; quality comprehensive evaluation index factor; *Bupleurum chinense* DC.; NIRS; system modeling; grade division

中药是中医药大健康产业的主体,中药质量关系大健康产业的发展。目前中药质量仍存在品种多、差异大、追溯过程长、检测指标繁琐及缺乏科学评价方法等问题<sup>[1]</sup>。此外,在中药质量监管过程中产生和积累的海量原始数据,并没有从数据中发现规律、有效利用率低,存在数据丰富(data rich)、知识匮乏(knowledge poor)的问题<sup>[2]</sup>。在国务院 2015 年发布《促进大数据发展行动纲要》推动下,借助大数据融合、云计算及区块链等现代信息技术应用于中药产业科学监管成为未来的必然发展趋势<sup>[3,4]</sup>。中药产业大数据中隐含的模式和规则往往无法靠经验或直觉发现,需要借鉴大数据分析的思维和将药材的生长环境、种植加工、炮制、化学成分演变及传统质量控制等低质、碎片化的数据转变成高质、高价值密度信息,然后借助数据挖掘、机器学习、人工智能等计算方法,并结合中药专业知识和判断,建立与应用相关联的数学模型,将实体关系透明化,为中药材品质的科学监管提供重要依据。

随着化学计量学的快速发展与应用,针对食品药品安全检测仪器和技术得到了迅速发展。检测方法也由早期的理化(physicochemical)、显微(microscopical)、色谱(chromatography)及质谱(mass spectrometry, MS)等技术,发展到非接触的近红外光谱(near infrared spectroscopy, NIRS)、拉曼(raman spectrometry, RS)、太赫兹(tera hertz, THz)及高光谱成像(hyperspectral imaging, HI)等技术<sup>[5,6]</sup>。其中 NIRS 是近几年发展起来的,是以采集被检测物质含氢基团 X-H(X=C、N、O)在近红外谱区振动、转动的合频及倍频信息用于物质定性和定量分析,广泛应用于农业、石油、化工、烟草及食品中,目前已被逐步应用于制药及药监行业,在药品快速定性和定量检测分析方面具有很大的发展潜力<sup>[6-8]</sup>。

柴胡 *Bupleuri Radix* 为伞形科植物柴胡 *Bupleurum chinense* DC. 或狭叶柴胡 *B. scorzonerifolium* Willd. 的干燥根,是我国常用的大宗药材,始载于《神农本草经》,列为上品,具有和解退热、疏肝解郁、升举阳气的功效,前者习称“北柴胡”,后者习称

“南柴胡”。北柴胡主产于河北省、黑龙江省、辽宁省等,南柴胡主产于辽宁省、吉林省、内蒙古自治区等<sup>[9,10]</sup>。因柴胡原植物品种复杂,本属约有 100 余生物种,我国有 36 个生物种、17 个变种及 7 个变型,此外生长环境变迁、品种变异、加工炮制及非药用部位去除不净等,药材品质差异很大<sup>[11]</sup>。目前《中国药典》2015、2020 年版对其水分、灰分、酸不溶性灰分、浸出物、柴胡皂苷 a 和 d 等含量进行了限定,对柴胡质量控制发挥了重要作用<sup>[10]</sup>。然而质量标志物(Q-marker)研究发现,柴胡中挥发油及脂肪油、三萜皂苷及黄酮类成分是其解热镇静、镇痛抗炎及镇咳的主要活性成分<sup>[12]</sup>。目前对于柴胡质量评价虽然多项指标并存,但含测指标单一、限量过低,不能满足对其品质评价与等级划分的需求。此外,监管过程中产生和积累了大量碎片化原始数据未得到充分利用、缺乏系统的整合。故本研究基于大数据背景下,以北柴胡药材为例,收集本中心采用《中国药典》2015 年版检测的 130 批次北柴胡药材各项检测数据,采用 NIRS 分析技术、化学计量法及数学建模方法对北柴胡药材各项检测数据进行模式识别研究,建立北柴胡药材快速溯源分析模型、含量预测模型、药材品质综合评价与等级划分数据库,系统地评价北柴胡药材品质,为中药材品质的科学监管提供新思路。

## 1 材料与方法

### 1.1 仪器

傅里叶变换 Tensor 37 NIRS 仪, OPUS 5.0 分析软件(德国 Bruker 公司); LC-20AT 高效液相色谱仪, LC Solution 色谱工作站(日本岛津公司); Phenomenex ODS C<sub>18</sub> 色谱柱(250 mm × 4.6 mm, 5 μm); BS224S 电子分析天平; FW-400A 高建万能粉碎机(北京科伟永兴仪器有限公司)。

### 1.2 试药

收集本中心 2015 年~2020 年抽检样品中采用《中国药典》2015 年版检测的 130 批次北柴胡(BCD)药材,包括北柴胡及其炮制品、不合格品,均为全检;柴胡皂苷 a(批号 110777-201912,纯度

94.8%)、柴胡皂苷 d (批号 110778-201912, 纯度 96.3%) 均购于中国食品药品鉴定研究院; 乙腈和甲醇均为色谱纯; 氨水、乙醇均为分析纯。

### 1.3 收集北柴胡药材 6 项指标的检测数据

收集本中心 2015 年~2020 年抽检样品中采用《中国药典》2015 年版检测的 130 批次北柴胡药材 6 项指标检测数据, 包括水分、灰分、酸不溶性灰分、浸出物、柴胡皂苷 a 和 d 的总量及茎秆占比(中心内部规定茎秆占比不得超过 10%)。

### 1.4 NIRS 采集与预处理

取训练集北柴胡药材粉末(过 6 号筛), 置于石英杯中(直径 2 cm), 装量为石英杯体积的三分之二, 振摇混匀。以空石英杯为空白参照, 采用积分球漫反射测样, 扫描波段范围 12 000~4 000  $\text{cm}^{-1}$ , 扫描时间 64 s, 分辨率 16  $\text{cm}^{-1}$ , 温度 10~30  $^{\circ}\text{C}$ , 湿度 (35±5)%。为了提高准确度, 每个样品重复测定 5 次求平均光谱用于建模分析。研究采用 NIRS 仪自带的 OPUS 软件对每个样品光谱进行预处理。

### 1.5 北柴胡药材溯源分析模型建立与应用

为了实现对不同产区北柴胡药材溯源分析, 以识别率为指标, 采用“1.4”项预处理方法优化光谱数据; 根据方差光谱选取方差较大的波段作为特征波段, 结合 SIMCA 14.1 统计软件中(principal components analysis, PCA)和(partial least square discriminant analysis, PLS-DA)分析进行建模, 并对构建模型进行验证。

### 1.6 北柴胡药材多指标定量分析模型建立及应用

采用“1.4”项光谱预处理方法, 以决定系数(correlation coefficient,  $R^2$ )、训练集均方根误差(root mean square error of cross-validation, RMSECV)、验证集均方根误差(root mean square error of prediction, RMSEP)以及相对分析误差(relative analysis error, PRD)为评价指标确定最佳光谱预处理方法。通过 PLS-DA, 采用各自最佳光谱预处理方法、光谱区域及主因子数建立定量分析模型, 并对构建模型进行验证<sup>[13]</sup>。

### 1.7 北柴胡药材品质综合评价模型的建立与应用

#### 1.7.1 原始数据无量纲化处理

由于样品各检测指标间的单位与量纲不同, 及考虑检测指标与药材品质正负相关性, 为体现样品的差异性和方便计算, 采用最大优越型( $R_{ij}$ )和最小优越型( $R'_{ij}$ )分别对各指标原始数据进行同度无量

纲化处理, 计算公式如下:

$$R_{ij} = X_{ij} / \max X_{ij}; R'_{ij} = \min X_{ij} / X_{ij}$$

式中,  $X_{ij}$  指第  $j$  个样品的第  $i$  项指标的测定结果;  $\max X_{ij}$  指在  $m$  个样品中第  $i$  项正相关指标的含量最大值;  $\min X_{ij}$  指在  $m$  个样品中第  $i$  项负相关指标的含量最小值 ( $1 \leq i \leq n; 1 \leq j \leq m$ )。

#### 1.7.2 指标权重系数计算

采用变异系数计算  $W_i$ , 计算过程如下:

$$W_i = \delta_i / \sum_{i=1}^n \delta_i; \delta_i = D_i / \bar{X}_i;$$

$$\bar{X}_i = \frac{1}{m} \sum_{j=1}^m X_{ij}; D_i = \sqrt{\frac{1}{m} \sum_{j=1}^m (X_{ij} - \bar{X}_i)^2}$$

式中,  $W_i$  指第  $i$  项指标的权重系数;  $\delta_i$  指第  $i$  项指标的变异系数;  $\bar{X}_i$  指第  $i$  项指标的平均值;  $D_i$  指第  $i$  项指标的均方差。

#### 1.7.3 综合评价指数计算

综合评价指数 ( $F_q$ ) 是反应样品与标准样品间相互接近的程度, 其值越大越接近。根据  $F_q$  大小来综合评价样品的品质和等级划分。同时基于各检测指标对药材品质贡献度的不同, 赋予不同的权重系数  $W_i$ , 建立体现药材整体质量的综合评价数学模型:

$$F_{qj} = \left[ \sum_{i=1}^a W_i \times R_{ij} + \sum_{i=1}^b W_i \times R'_{ij} \right] \times 100$$

式中,  $F_{qj}$  指第  $j$  个样品的综合评价指数得分;  $a$  指正相关指标的个数,  $b$  指负相关指标的个数。

## 2 结果与分析

### 2.1 训练集、验证集与预测集样品选取

采用 Kennard-Stone 方法将样品划分为训练集和验证集, 研究中选用 90 批作为训练集, 20 批为验证集用于建模研究, 剩余的作为预测集<sup>[14]</sup>。训练集和验证集样品 6 项指标的含量分布见表 1。

### 2.2 NIRS 采集与预处理

采用“1.4”项方法采集每个样品光谱。因采集 NIRS 时, 存在随机噪声、基线漂移、光散射、样本颗粒大小等随机因素的影响, 导致光谱的差异, 干扰 NIRS 与样本目标成分含量的关系, 影响建模的可靠性和稳定性。研究分别采用消除常数偏移 (elimination of constant offset, ECO)、多散射校正 (multiplication signal correction, MSC)、标准正态变换 (standard normal variate, SNV)、一阶导数 (first derivative, FD)、二阶导数 (second derivative, SD) 及平滑 (smoothing points, SP) 等组合方法处理光谱中存在的无用信息, 以期提高模型判别的准确度 (见图 1)。

表 1 训练集和验证集北柴胡样品 6 项指标的含量分布

Table 1 Contents distribution of the 6 detection index of the roots of BCD samples in the calibration set and validation set

检测指标 Detection index	训练集 Calibration set (n = 90)		验证集 Validation set (n = 20)	
	质量分数 Content (%)	平均值 Mean (%)	质量分数 Content (%)	平均值 Mean (%)
水分 Water	4.0 ~ 13.3	7.7 ± 1.86	5.9 ~ 10.9	7.9 ± 1.38
灰分 Total ash	3.8 ~ 11.3	6.0 ± 1.56	3.9 ~ 8.9	5.4 ± 1.02
酸不溶性灰分 Acid-insoluble ash	0.9 ~ 5.1	1.8 ± 0.81	0.9 ~ 3.9	1.7 ± 0.66
浸出物 Extract	10.8 ~ 21.5	14.8 ± 2.96	11.5 ~ 20.9	15.7 ± 2.79
柴胡皂苷 a + d Saikosaponins a and d	0.33 ~ 1.93	0.92 ± 0.34	0.56 ~ 1.92	1.01 ± 0.26
茎秆占比 Proportion of stem	0.5 ~ 10.7	4.7 ± 2.2	0.5 ~ 9.6	4.6 ± 2.1

注:平均值采用均数 ± 标准差表示。

Note: Mean is expressed as  $x \pm s$ .

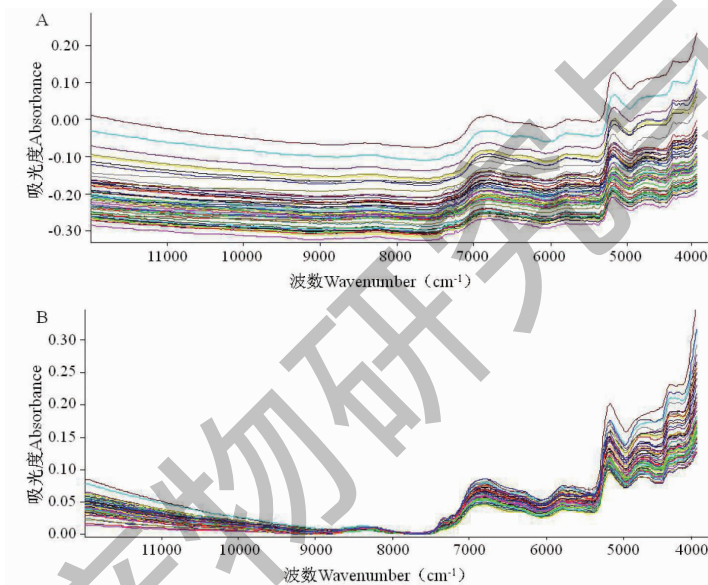


图 1 训练集中 60 批次北柴胡样品的近红外光谱图

Fig. 1 NIRS spectra of 60 batches of the roots of BCD samples in calibration set

注:A-原始光谱;B-消除常数偏移+标准正态转换。Note:A-Raw spectrum;B-ECO + SNV.

## 2.3 北柴胡药材溯源分析模型建立与应用

### 2.3.1 模型建立

PCA 分析模型识别率结果见表 2, 特征波段见图 2A。由表 2 和图 2A 可知, 采用最优预处理方法 ECO + SNV + SP 处理光谱、根据方差光谱选取方差

较大的波段(4 204.15 ~ 5 608.51  $\text{cm}^{-1}$ 、6 325.51 ~ 7 606.04  $\text{cm}^{-1}$ 、7 612.50 ~ 8 101.00  $\text{cm}^{-1}$ 、8 120.10 ~ 9 609.66  $\text{cm}^{-1}$ )作为特征波段结合 PCA 分析, 基本可以实现 5 个产区北柴胡药材的产区划分, 但效果不佳, 陕西省、河北省与四川省均有部分重叠(见图 2B)。

表 2 通过 SIMCA 和不同的预处理方法获得的识别率

Table 2 Identification accuracy obtained by SIMCA and different pretreatment methods

预处理方法 Pretreatment method	识别率 Identification accuracy (%)					
	陕西省 Shanxi	河北省 Hebei	辽宁省 Liaoning	四川省 Sichuan	内蒙古自治区 Inner Mongolia	整体 Overall
ECO + SNV + SP	70	80	90	100	90	86
ECO + SNV + FD	60	70	70	100	80	76

续表 2 (Continued Tab. 2)

预处理方法 Pretreatment method	识别率 Identification accuracy (%)					
	陕西省 Shanxi	河北省 Hebei	辽宁省 Liaoning	四川省 Sichuan	内蒙古自治区 Inner Mongolia	整体 Overall
ECO + SNV	60	50	70	90	80	70
ECO	40	60	70	90	80	68
SNV	50	70	80	100	70	74
MSC	40	50	70	80	90	66
SD	40	80	70	80	60	66
FD	50	70	80	90	80	74

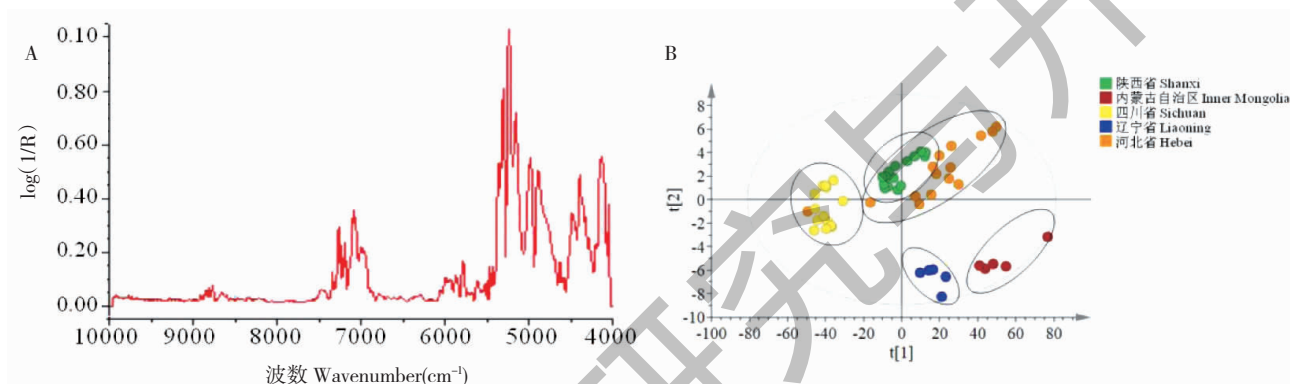


图 2 北柴胡药材 PCA 模式识别模型的方差光谱图 (A) 及得分图 (B)

Fig. 2 PCA model variance spectrum (A) and score diagram (B) of the roots of BCD samples

为进一步提高识别准确率,在 PCA 分析基础上,按照主成分贡献率大于 1%,累计贡献率大于 75% 的原则选择 3 个主成分,见表 3。采用 PLS-DA 分析进行建模,结果 5 个产区北柴胡药材被较好地分开,PLS-DA 得分图见图 3A。对得分图进行分析,该模型自变量累计解释能力  $R^2X$  为 0.983,因变量累计解释能力  $R^2Y$ 、 $Q^2Y$  分别为 0.901、0.820,均大于 0.5,且  $0 < R^2Y - Q^2Y = 0.081 < 0.3$ 。 $R^2$  与  $Q^2$  与纵轴和横轴交点均小于 0.5,表明建立模型没有出现过度拟合(见图 3B),说明模型预测能力较好<sup>[15]</sup>。

对波段变量的重要性投影(VIP)值进行排序,VIP 值越大,说明该波段对产地区分的贡献越大。因此,该法选择波段并不连续,由许多波段累计组合,主要特征波段为  $4\ 104.35 \sim 5\ 708.00\ \text{cm}^{-1}$ 、 $6\ 101.51 \sim 7\ 382.33\ \text{cm}^{-1}$ 、 $7\ 768.95 \sim 8\ 701.25\ \text{cm}^{-1}$  及  $8\ 120.10 \sim 9\ 609.66\ \text{cm}^{-1}$ 。特征波段的区间与 PCA 特征波段区间基本一致。NIRS 分析显示,上述波段是表征挥发油、皂苷及黄酮类物质中 O-H、C=O、C-H 键振动的主要区间,继而反映出北柴胡药材所含化学成分的差异。

表 3 主成分及贡献率

Table 3 Principal components and contribution rate

主成分数 Principal component	特征值 Eigenvalue	贡献率 Contribution rate (%)	累计贡献率 Cumulative contribution rate (%)
1	28.7	95.070	95.070
2	15.6	2.195	97.265
3	2.05	1.047	98.312

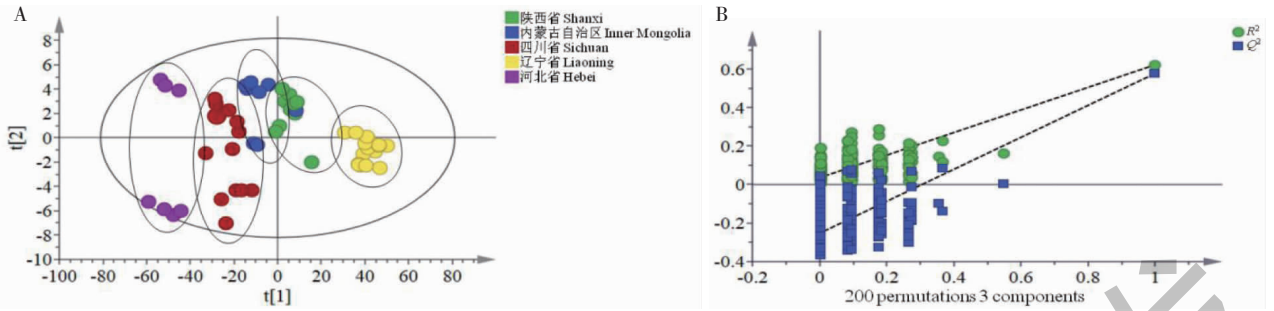


图3 北柴胡药材 PLS-DA 模式识别模型的得分图 (A) 与模型检验图 (B)

Fig. 3 PLS-DA model score diagram (A) and model test diagram (B) of the roots of BCD samples

2.3.2 模型的验证与应用

为了验证 PLS-DA 模型的可靠性,分别将验证集和预测集样品的 NIRS 输入溯源模型中进行内、外部验证。结果内部验证识别率均为 100%,外部

验证 5 个产区样品中除了陕西省和河北省识别率为 90%,其余 3 个产区识别率均为 100%,整体识别率为 90%(见表 4)。

表 4 北柴胡样品产地预测结果

Table 4 Prediction results of the roots of BCD samples

项目 Project	验证集 Validation set					预测集 Predicted set				
	陕西省 Shanxi	河北省 Hebei	辽宁省 Liaoning	四川省 Sichuan	内蒙古自治区 Inner Mongolia	陕西省 Shanxi	河北省 Hebei	辽宁省 Liaoning	四川省 Sichuan	内蒙古自治区 Inner Mongolia
样品个数 Sample size	6	3	5	4	2	5	5	4	4	2
预测准确个数 Predicted size	6	3	5	4	2	4	4	4	4	2

2.4 北柴胡药材多指标定量分析模型建立及应用

2.4.1 模型建立

结果分别得到北柴胡药材 6 项指标的 NIRS 定量分析模型,其水分、灰分、酸不溶性灰分、浸出物、柴胡皂苷 a 与 d 总量及茎秆占比的  $R^2_{cal}$  分别为

0.982、0.934、0.942、0.915、0.938 及 0.955,均大于 0.85;RMSECV 分别为 0.151、0.215、0.156、0.123、0.210 及 0.175,均为各自模型的最小值,说明模型预测性能良好(见表 5)。

表 5 不同预处理方法以及光谱区段对定量分析模型的影响

Table 5 Effects of different spectra pretreatments and spectral section on the models

检测指标 Detection index	预处理方法 Pretreatment method	谱区选择 Select band ( $cm^{-1}$ )	主因子数 Factor number	$R^2_{cal}$	RMSECV	RPD
水分 Water	FD + SNV	7 501.9 ~ 5 450.6; 4 601.0 ~ 4 414.0	7	0.982	0.151	5.120
	FD + SNV	5 152.7 ~ 4 415.0	5	0.816	0.211	3.860
	SNV	7 501.9 ~ 4 420.0	6	0.756	0.251	2.670
灰分 Total ash	SD + SNV	7 501.7 ~ 5 350.6; 4 701.0 ~ 4 404.0	10	0.934	0.215	5.630
	SD + SNV	5 453.8 ~ 4 597.6	7	0.865	0.235	4.910
	SD	6 230.8 ~ 4 120.0	5	0.815	0.242	2.910
酸不溶性灰分 Acid-insoluble ash	SD + MSC	8 100.1 ~ 5 751.0; 4 623.0 ~ 4 210.0	7	0.942	0.156	5.310
	SD + MSC	5 453.8 ~ 4 597.6	6	0.915	0.220	4.420
	SD	6 230.8 ~ 4 120.0	6	0.822	0.352	3.110

续表 5 (Continued Tab. 5)

检测指标 Detection index	预处理方法 Pretreatment method	谱区选择 Select band ( $\text{cm}^{-1}$ )	主因子数 Factor number	$R^2_{\text{cal}}$	RMSECV	RPD
浸出物 Extract	SNV	7 929.0 ~ 5 325.0; 4 512.1 ~ 4 221.0	8	0.915	0.123	4.740
	SNV	5 325.0 ~ 4 213.0	8	0.856	0.176	4.012
	FD	7 956.0 ~ 4 327.0	7	0.830	0.182	3.202
柴胡皂苷 a + d Saikosaponins a and d	FD + MSC	8 123.0 ~ 5 522.2; 4 912.0 ~ 4 324.5	7	0.938	0.210	4.370
	SNV	5 925.0 ~ 4 280.0	8	0.852	0.226	2.120
	FD	7 986.0 ~ 4 314.0	6	0.831	0.280	2.203
茎秆占比 Proportion of stem	FD + SNV	7 511.2 ~ 5 444.2; 4 810.3 ~ 4 656.1	6	0.955	0.175	4.450
	SNV	5 656.0 ~ 4 251.0	7	0.877	0.179	3.813
	FD	7 875.0 ~ 4 221.0	7	0.811	0.189	3.116

### 2.4.2 模型的验证与应用

将验证集样品的 NIRS 输入到上述模型中, 预测样品各项检测指标的含量, 结果 6 项指标的预测

值和实测值的  $R^2_{\text{test}}$  分别为 0.958、0.928、0.951、0.947、0.901 及 0.956, 表明所建立的模型具有很好的预测能力(见图 4)。

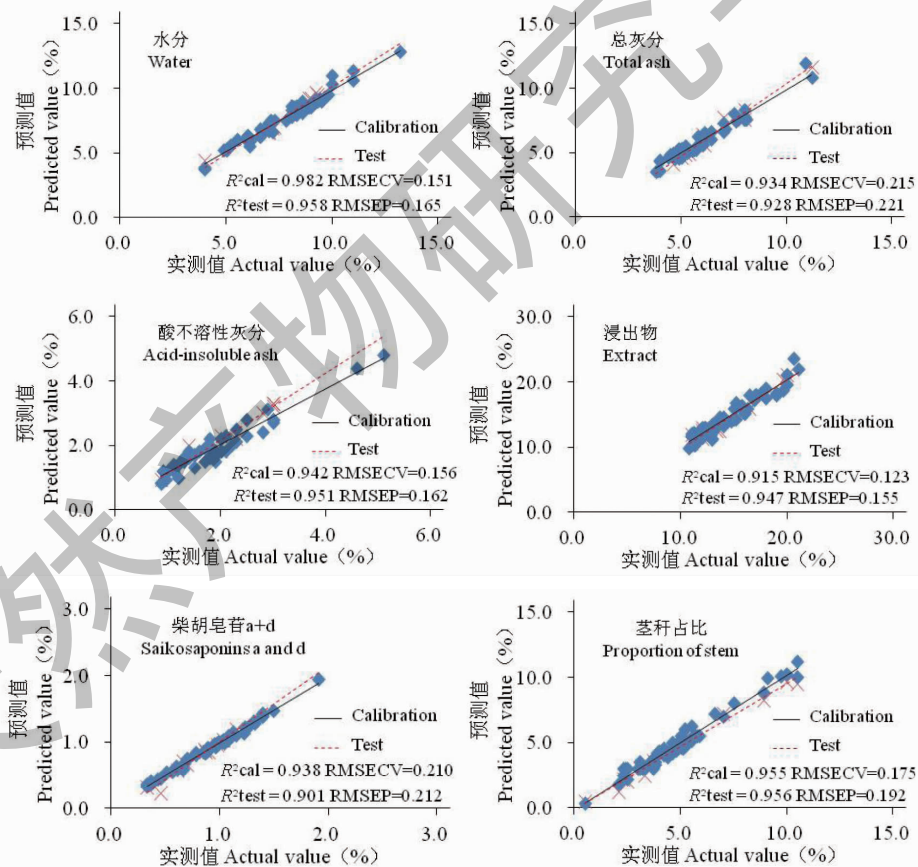


图 4 北柴胡药材 6 项检测指标的 NIRS 检测模型

Fig. 4 NIRS quantitative model of six detection indexes for the roots of BCD samples

利用上述所建立的模型, 对预测集 20 批北柴胡样品进行 6 项检测指标快速分析, 并建立快速评价数据库。含量预测结果显示, 不符合《中国药典》规

定的样品为 4 批, 主要集中在水分、总灰分、酸不溶性灰分、浸出物及茎秆占比(中心内部规定茎秆占比不得超过 10%)项下(见图 5)。

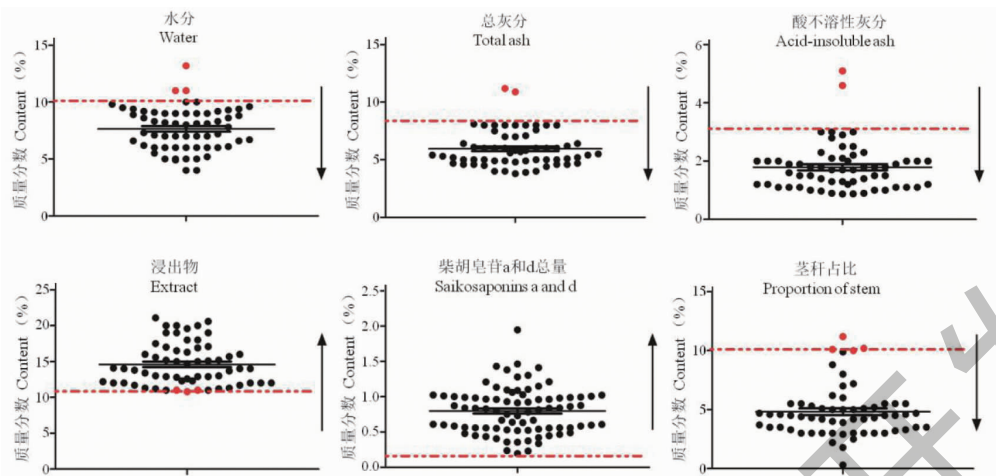


图5 北柴胡样品6项检测指标的近红外检测值

Fig. 5 NIRS predicted values of six detection indexes of the roots of BCD samples

注:黑点表示合格样品,红点表示不合格样品;黑色实线为检测均值与误差棒,红色虚线为检测限量;箭头方向代表满足限量要求的合格样品趋势。Note:Black dots represent qualified samples,red dots represent rejected samples;The black solid line is the mean value and error bars,the red dotted line is the detection limit;The arrow direction represents the trend of qualified samples meeting the limit requirements.

## 2.5 北柴胡药材品质综合评价模型的建立与应用

### 2.5.1 北柴胡药材品质的综合评价与等级划分模型建立

分别将90批北柴胡样品的6项指标数据代入“1.7”项综合评价模型各计算公式中,得水分、灰分、酸不溶性灰分、浸出物、柴胡皂苷a和d总量及茎秆占比最佳权重系数 $W_1$ 、 $W_2$ 、 $W_3$ 、 $W_4$ 、 $W_5$ 、 $W_6$ 分别为0.112 3、0.121 1、0.207 3、0.092 1、0.155 3和0.203 0,该条件下 $F_q$ 计算公式如下。

$$F_{qj} = [0.112\ 3 \times R_{1j} + 0.121\ 1 \times R_{2j} + 0.207\ 3 \times$$

$$R_{3j} + 0.092\ 1 \times R_{4j} + 0.155\ 3 \times R_{5j} + 0.203\ 0 \times R_{6j}] \times 100$$

通过对90批北柴胡样品的 $F_q$ 值计算,结果 $F_q$ 值范围在25.55%~80.08%,且 $F_q$ 值呈高斯分布(Sig=0.25),见图6A。依据正态分布区域的样品数与实际“辨状论质、理化检测”结果建立等级划分数据库。将其分为4个区域, $F_q$ 值<30%为不合格; $F_q$ 值在30%~50%为三等品; $F_q$ 值在50%~70%为二等品; $F_q$ 值在70%~100%为一等品(见图6B)。

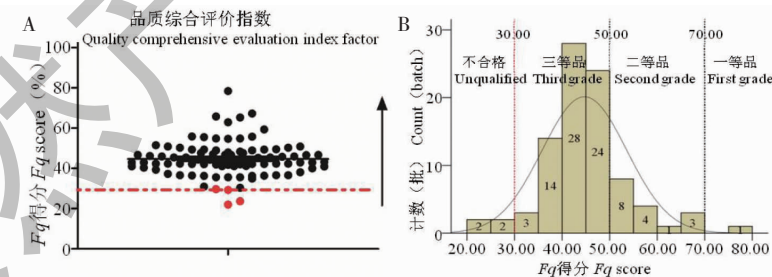


图6 北柴胡药材的综合品质评价与等级划分

Fig. 6 Comprehensive quality evaluation and grade identification of the roots of BCD samples

注:A-90批北柴胡样品 $F_q$ 值的散点分布;B-正态分布区域中不合格区域及等级划分。Note:A-Scatter distribution diagram of  $F_q$  value of 90 batches of the roots of BCD samples;B-Unqualified area and grade division in normal distribution area.

### 2.5.2 等级划分模型的验证与应用

依据上述建立的等级划分数据库,对40批北柴胡药材NIRS定量模型预测数据进行分析。结果显示,不合格为4批、三等品为33批、二等品为3批。

4批不合格样品按照《中国药典》2020年版一部检测,结果分别为茎秆占比(>10%)、浸出物(<11.0%)不合格;合格样品均分布在二等及三等品的区域,其中3批二等品 $F_q$ 值分别为55.73%、51.50%、



50.30%,通过检测各项指标均较优,与实际“辨状论质、理化检测”质量相符,表明该模型用于北柴胡

药材等级划分具有一定的合理性(见图7)。

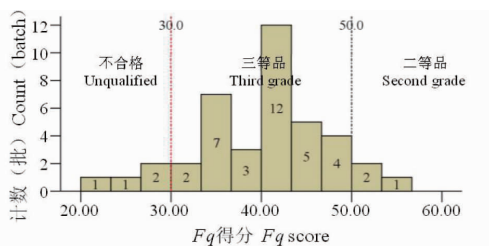


图7 北柴胡药材等级划分数据库预测结果

Fig. 7 Prediction results of the grade identification database of the roots of BCD samples

### 3 讨论与结论

参照德国“工业4.0”,目前我国中药产业总体处于工业2.0水平,相对较为滞后<sup>[16]</sup>。Sigma差距指产品性能与消费者期望之间的差距,在六西格玛管理法中(6 Sigma是一个目标),目前我国中药产业处于2.0~3.0 Sigma水平,产生上述差距的原因因为产品难以追溯、过于依赖最终产品检验(Quality testing),缺乏过程理解<sup>[17]</sup>。2018年至2020年,我国中药饮片抽检不合格率远高于中成药与化药<sup>[18,19]</sup>。中药饮片质量是中药临床安全有效的基础,关系中药大健康产业的发展,开展中药监管方法研究,实现中药质量科学监管显得尤为重要。影响中药材质量因素较多,如品种、生境、采收、加工炮制、储藏等,具有一定复杂性,但又存在必然的内在联系。要实现科学监管,需要相应交叉学科技术、信息科学及大数据科学的整合应用。近年来,我国对中药质量控制做了大量工作,但仍不能满足日益提高的质量控制要求。对此,有研究提出中药质量标志物(Q-marker)的概念,为中药质量评价指明了方向<sup>[20]</sup>。然而中药质量综合评价涉及多个Q-marker,并且不同Q-marker对中药质量的贡献度也不相同,缺乏科学的综合评价方法。本实验提出中药综合评价指数 $F_q$ 的概念,结合现行《中国药典》的评价体系,基于大数据分析的结果,通过对各检测指标 $R_i$ 数据无量纲化处理,运用变异系数法优化计算中药质量评价指标的权重系数 $W_i$ ,避免专家赋权的主观偏好性,较为客观地反映了中药质量评价各指标的相对重要程度。建立了一种适合中药材品质综合评价和等级划分的快速科学评价方法,为中药材质量的科学监管提供了新的方法。

科学监管是药物安全性、有效性及质量可控性

的保证,中药材质量科学监管不仅要结合传统的经验,现行的评价体系及先进的技术方法,也要引入现代的信息技术和统计方法,通过对大数据提取分析建立中药材品质综合评价方法和等级划分标准。模式识别技术具有“整体性”和“模糊性”的特点,是一种综合的、可量化的鉴别手段,是当前符合中药特色,能很好地反映中药真实性。 $F_q$ 是一个发展的体系,随着大数据的不断积累及Q-marker的不断引入,该体系也会不断迭代和完善,可以更好地应用于中药质量的科学监管。

### 参考文献

- 1 Wang J, Li FC, Song HM, et al. Influencing factors and quality control measures of Chinese medicinal materials and decoction pieces[J]. Food Drug (食品与药品), 2020, 22: 406-411.
- 2 Yang M. Quality process control system of Chinese medicine preparation[J]. China Food Drug Admin Mag (中国食品药品监管), 2019, 7: 87-90.
- 3 Chen W. Interpretation of the "Outline of Action for Promoting the Development of Big Data" [J]. Zhongguo Xinxihua (中国信息化), 2015, 10: 11-14.
- 4 Xu B, Shi XY, Luo G, et al. Key technologies and applications of industrial big data in manufacturing of Chinese medicine[J]. China J Chin Mater Med (中国中药杂志), 2020, 45: 221-223.
- 5 Wu GP, Zhou YH, Cao CX, et al. Detection of pioglitazone and rosiglitazone illegally added in health care foodstuffs by nano-gold surface-enhanced[J]. Phys Test Chem Anal: Chem Anal (理化检验-化学分册), 2019, 55: 1211-1223.
- 6 Liu JX. Modern Near Infrared Spectroscopy Analysis Technology (现代近红外光谱分析技术) [M]. Beijing: Science Press, 2008: 102-106.

- 7 Wu XY, Hou XL, Su Y, et al. Determination of water-soluble extracts and gentiopicroside in *Gentianae Radix et Rhizoma* by portable near-infrared spectrometers [J]. *Nat Prod Res Dev* (天然产物研究与开发), 2020, 32:1363-1369.
- 8 Zhang LJ, Jia JZ, Zhang XY, et al. Determination of sex-enhancing drugs illegally added in health care products by TLC-NIRS combined technology [J]. *Nat Prod Res Dev* (天然产物研究与开发), 2021, 33:1572-1581.
- 9 State Administration of Traditional Editorial Board. *Chinese Herbal Medicine: Volume V (中华本草: 第五卷)* [M]. Shanghai: Shanghai Science and Technology Press, 1999: 909.
- 10 Chinese Pharmacopoeia Commission. *Pharmacopoeia of the People's Republic of China: Vol I (中华人民共和国药典: 一部)* [M]. Beijing: China Medical Science Press, 2020: 293.
- 11 Meng J, Yao RY, Chen XF, et al. Advances in studies on classification of *Bupleurum* [J]. *China J Chin Mater Med* (中国中药杂志), 2012, 37:1523-1526.
- 12 Li CN, Liu Y, Liu YY, et al. Advances in research of chemical constituents and active constituents of *Dictamnus dasycarpus* DC [J]. *Chin Arch Tradit Chin Med* (中华中医药学刊), 2014, 32:2674-2677.
- 13 Yi C, Xie MY, Yan Y, et al. Discrimination of *Ganoderma lucidum* according to geographical origin with near infrared diffuse reflectance spectroscopy and pattern recognition techniques [J]. *Anal Chim Acta*, 2008, 618:121-130.
- 14 Claeys DD, Verstraelen T, Pauwels E, et al. Conformational sampling of macrocyclic alkenes using a Kennard-Stone-based algorithm [J]. *J Phys Chem A*, 2010, 114:6879-6887.
- 15 Suhanandy D, Yulia M, Kuroki S, et al. The use of SIMCA method and NIR spectroscopy with hand-held spectrometers equipped with integrating sphere for classification of two different Indonesian specialty coffees [J]. *J Physics: Conf Ser*, 2021, 1751:012080.
- 16 Chen JY, Zhai HB, Zhang BL. Chinese medicine industry 4.0: advancing digital pharmaceutical manufacture toward intelligent pharmaceutical manufacture [J]. *China J Chin Mater Med* (中国中药杂志), 2016, 41:1-5.
- 17 Blackburn TD, Mazzuchi TA, Sarkani S. Overcoming inherent limits to pharmaceutical manufacturing quality performance with QbD (quality by design) [J]. *J Pharma Innov*, 2011, 6: 69-76.
- 18 Circular of the State Food and Drug Administration on 25 batches of drugs that do not meet the requirements (No. 55 in 2020) (国家药监局关于 25 批次药品不符合规定的通告 (2020 年第 55 号)) [EB/OL]. (2020-08-20) [2022-09-28]. <https://www.nmpa.gov.cn/yaopin/ypjgdt/20200820172832135.html>.
- 19 Liu P, Li X, Dai Y. Analysis of the quality results of drug sampling and testing in Shaanxi institute of food and drug control from 2016 to 2018 [J]. *China Pharm* (中国药业), 2020, 29:64-66.
- 20 Liu CX. Improve the quality standard and quality control theory of traditional Chinese medicine and promote the scientific development of traditional Chinese medicine industry [J]. *Chin Tradit Herb Drugs* (中草药), 2019, 50:4517-4518.