

基于系统建模思想的脑心通胶囊中丹酚酸 B 近红外定量建模

高瑞琳¹, 杨鹏硕¹, 许刚², 武晓文¹, 杨畅¹, 史新元^{1*}

1. 北京中医药大学中药学院, 北京 102488

2. 陕西步长制药有限公司, 陕西 西安 710075

摘要 近红外定量建模过程常以模型评价指标对样本集选择、光谱预处理方法、潜变量因子数、变量选择等参数进行逐步优选, 然后采用优选参数建立模型, 此建模过程并未充分考虑参数间的协同作用对模型的影响, 导致模型的质量存在风险。系统建模思想将定量建模过程看作一个整体, 将样本集选择、光谱预处理、变量筛选、校正等作为要素, 基于要素间的关联性进行参数轨迹全局优化, 优选出一组最佳建模参数来建立模型, 从而保障模型质量。该研究基于系统建模思想开展了脑心通胶囊中丹酚酸 B 的近红外定量建模。采用高效液相色谱法对 56 份丹参样品中的丹酚酸 B 进行含量测定, 同时采集其近红外光谱, 采用 D-最优设计样本集划分方法、光谱预处理方法、潜变量因子数及变量筛选方法作为自变量, 以模型评价指标为因变量优选最佳建模参数轨迹并建立定量分析模型。结果表明, 采用 Kennard-Stone(K-S)算法从 56 份样本中选择 3:1 的校正集样本、建模光谱波段为 $4\ 000\sim 10\ 000\ \text{cm}^{-1}$ 、预处理方法为二阶求导和标准正态变量变换校正, 潜变量因子数为 7 时所建立的定量分析模型效果最优。模型的校正标准偏差(RMSEC)、预测标准偏差(RMSEP)均为 0.001 8, 表明模型稳健且预测准确度高。校正决定系数(R_{cal}^2)为 0.994 0, 预测决定系数(R_{pre}^2)为 0.995 2, 相对分析误差(RPD)为 9.19, 表明模型可用于高质量的定量分析和预测。本研究基于系统建模思想, 采用 D-最优设计实现了建模参数轨迹全局优化, 建立了丹酚酸 B 的定量分析模型并且显著提升了模型的稳健性和预测性, 为高效、快速测定脑心通胶囊中丹酚酸 B 含量提供了方法, 对脑心通胶囊中间体及成品的质量控制具有重要意义。

关键词 系统建模; D-最优设计; 近红外光谱; 脑心通胶囊; 丹酚酸 B

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)11-3573-06

引言

脑心通胶囊为黄芪、赤芍、丹参、全蝎、水蛭等 16 味药材粉碎加工制成的中药复方制剂, 具有益气活血、化瘀通络的功效, 临床上治疗心脑血管疾病具有显著疗效^[1]。目前, 脑心通胶囊的质量控制主要采用指纹图谱分析及指标成分丹酚酸 B 定量分析的方法^[2], 检测过程繁琐费时, 且难以实现脑心通胶囊制剂过程中间体的快速评价和质量控制。

近年来, 近红外光谱(near-infrared spectroscopy, NIR)分析技术凭借其快速、可靠、无损、环保等优势, 在中药材、中药制剂^[3]、中药生产过程以及药物质量快速评价^[4]中得到日益广泛的应用^[5]。稳健可靠的定量模型是近红外技术成功

应用的前提和保证。常规近红外定量建模过程采用逐步优化的思路^[6], 这样的建模思路并未充分考虑建模参数间的关联性对模型性能的影响, 致使定量模型的质量存在风险^[7]。系统建模思想^[7]将建模过程看作一个整体, 将样本集选择、光谱预处理、变量筛选、校正等作为要素, 基于要素间的关联性对建模参数轨迹进行全局优选, 采用最优参数轨迹构建模型, 保障了模型质量。本研究基于这一思想, 采用 D-最优设计以关键建模参数为自变量, 模型评价指标为因变量, 基于参数间关联性进行多次建模试验, 根据定量建模结果拟合预测模型, 以最优模型性能为目标对建模参数轨迹进行全局优选, 优选出最佳参数轨迹, 建立了稳健的丹酚酸 B 偏最小二乘法(partial least square, PLS)定量模型, 为脑心通胶囊生产过程及产品的快速质量评价提供有效方法。

收稿日期: 2019-11-12, 修订日期: 2020-03-07

基金项目: 国家中药标准化项目(ZYBZH-C-QIN-45)资助

作者简介: 高瑞琳, 女, 1993 年生, 北京中医药大学中药学院硕士研究生 e-mail: gaoruilin02@163.com

* 通讯联系人 e-mail: xyshi@126.com

1 实验部分

1.1 仪器

Antaris 傅里叶变换近红外光谱仪(美国 Thermo Nicolet 公司)配备 InGaAs 检测器,积分球漫反射采样系统,Result 操作软件;Agilent 1260 型高效液相色谱仪(美国安捷伦科技公司);Agilent Eclipse XDB-C₁₈ 色谱柱(5 μm , 4.6 mm \times 250 mm);拜杰高速多功能粉碎机 BJ-400A 型, BP211D 型电子天平(德国赛多利斯科学仪器有限公司)。

1.2 材料与试剂

乙腈(色谱纯,批号 175157,赛默飞世尔科技(中国)有限公司),蒸馏水(广州屈臣氏食品饮料有限公司),甲醇(分析纯,批号 20170515,北京化工厂),丹酚酸 B(上海源叶生物科技有限公司,纯度 $\geq 98\%$,批号 P10J9F52565)。丹参药材由陕西步长制药有限公司提供。将丹参药材随机编号,粉碎,过六号筛,低温干燥后备用。

1.3 丹酚酸 B 的含量测定

1.3.1 色谱条件

采用 Agilent Eclipse XDB-C₁₈ 色谱柱(5 μm , 4.6 mm \times 250 mm),流动相为乙腈(A)和 0.1% 磷酸水溶液(B),梯度洗脱(0~20 min, 23% A; 20~25 min, 23%~95% A; 25~35 min, 95%~23% A; 35~40 min, 23% A),检测波长 286 nm,流速 1.0 mL \cdot min⁻¹,柱温 30 $^{\circ}\text{C}$,进样量 10 μL 。

1.3.2 对照品溶液制备

精密称取丹酚酸 B 对照品适量,加 80% 甲醇配制成浓度为 2.002 mg \cdot mL⁻¹ 对照品溶液。

1.3.3 供试品溶液制备

取样品细粉约 0.5 g,精密称定,置于具塞锥形瓶中,精密加入 80% 的甲醇 50 mL,称定重量,超声提取 30 min,放冷后,再称定重量,用 80% 甲醇补足减失的重量,摇匀,滤过,取续滤液,即得。

1.3.4 方法学考察

1.3.4.1 线性关系考察

精密量取丹酚酸 B 对照品溶液 0.05, 0.25, 1.0, 2.0 和 2.5 mL,分别用 80% 甲醇稀释于 5 mL 容量瓶中,即得各系列浓度对照品溶液。对照品溶液进样 10 μL ,测定峰面积,作为线性关系考察。

1.3.4.2 精密度、稳定性、重复性、加样回收率考察

精密吸取丹酚酸 B 对照品溶液,重复进样 6 次,测定丹酚酸 B 峰面积值,作为精密度考察;取同一份供试品溶液,分别于制备后 0, 2, 4, 8, 12 和 24 h 进样分析,测定丹酚酸 B 峰面积值,作为稳定性考察;精密称取同一批丹参样品 6 份,按照 1.3.3 节所述方法制备供试品溶液,进样分析,测定丹酚酸 B 含量,作为重复性考察;精密称取 6 份已知含量的丹参样品约 0.3 g,分别精密加入对照品溶液 3.5 mL,按供试品制备方法制备,进样分析,测定丹酚酸 B 峰面积,作为回收率考察。

1.3.5 样品测定

分别精密吸取对照品和供试品溶液 10 μL 注入液相色谱

仪,记录 40 min 色谱图。见图 1。

1.4 近红外漫反射光谱采集

以积分球漫反射方式采集近红外光谱,光谱采集范围为 4 000~10 000 cm^{-1} ,分辨率 8 cm^{-1} ,扫描 32 次,增益 8,以仪器内置背景为参照,对 56 份丹参样品进行光谱采集,每份样品重复测定 3 次,取其平均光谱作为丹参各样品的近红外光谱。丹参样品的原始近红外光谱图见图 2。

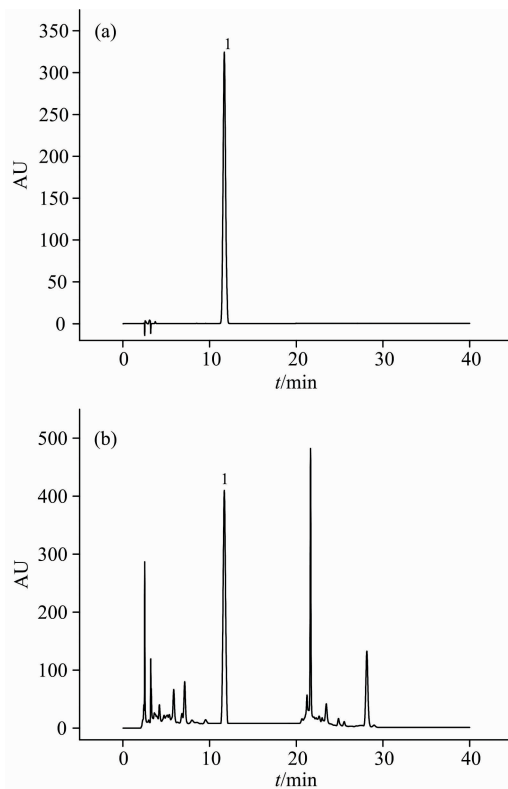


图 1 对照品(a)和丹参样品(b)的 HPLC 图

1: 丹酚酸 B

Fig. 1 The chromatograms of salviolic acid B reference standard (a) and sample solution (b)

1: Salviolic acid B

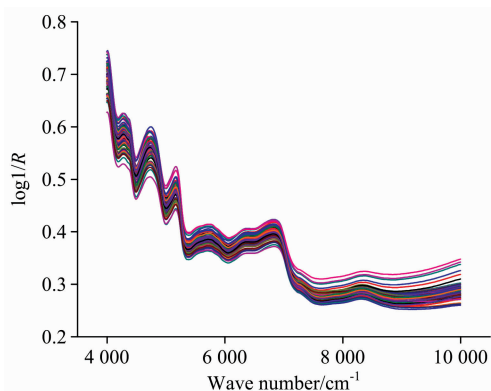


图 2 丹参样品的原始近红外光谱图

Fig. 2 The raw spectra of *Salvia miltiorrhiza* samples

1.5 数据处理

运用 Unscrambler7.0(挪威 CAMO 公司)软件对光谱进

行预处理。采用 Matlab7.0(美国 Mathwork 公司)软件进行变量筛选及模型计算。

1.6 建模参数及评价指标确定

为构建稳健的定量模型,选择多元散射校正(MSC)、标准正态变换散射校正法(SNV)、SG平滑(SG)、二阶导数法(2D)、标准正态变换组合二阶导数法(2D-snv)等不同的光谱预处理方法对原始光谱进行预处理,再分别采用间隔偏最小二乘法(iPLS)、组间间隔偏最小二乘法(siPLS)、移动窗口偏最小二乘法(mwPLS)对光谱变量^[8]进行筛选,建立丹酚酸 B 定量模型。以校正决定系数(R_{cal}^2)、校正标准偏差(RMSEC)、预测决定系数(R_{pre}^2)、预测标准偏差(RMSEP)、相对分析误差(RPD)为综合指标对定量模型的稳健性及预测性进行评价。

1.7 基于 D-最优试验设计的参数轨迹优选及模型建立

D-最优设计^[9]具有试验次数少、多目标同步优化、预测参数精密度高等特点,可用于建模参数轨迹全局优化。采用 Design Expert 8.0.6 软件进行 D-最优试验设计,优选出最佳

建模参数轨迹并建立丹酚酸 B 定量模型。以关键建模参数样本集划分比例(X_1)、潜变量因子数(X_2)、光谱预处理方法(X_3)、变量筛选方法(X_4)为自变量,以 R_{cal}^2 (Y_1)、RMSEC(Y_2)、 R_{pre}^2 (Y_3)、RMSEP(Y_4)及 RPD(Y_5)为因变量,根据课题组前期对中药 NIR 定量建模过程解析^[10]的工作基础,对各因素的水平进行设置。D-最优试验设计因素水平表见表 1。

2 结果与讨论

2.1 HPLC 的方法学考察

依据 1.3.4 节所述方法对系列对照品溶液进行含量测定,以浓度($\text{mg} \cdot \text{mL}^{-1}$)为横坐标,峰面积为纵坐标绘制标准曲线,计算得标准曲线为 $y=11\ 712x-124.35$,相关系数 $R^2=0.999\ 9$,表明丹酚酸 B 对照品浓度在 $0.020\ 0\sim 1.001\ 0\ \text{mg} \cdot \text{mL}^{-1}$ 的范围内线性关系良好;精密度考察中各对照品峰面积的 RSD 为 0.80% ,表明仪器的精密度良好;稳定性考察中供试品丹酚酸 B 峰面积的 RSD 为 0.95% ,表明供试品溶液在 24 h 内稳定性良好;重复性考察测得供试品中丹酚酸 B 的含量分别为 $0.065\ 42, 0.066\ 48, 0.067\ 19, 0.066\ 38, 0.065\ 38$ 和 $0.066\ 21\ \text{mg} \cdot \text{mg}^{-1}$,RSD 为 1.04% ,表明方法的重复性良好;测得丹酚酸 B 的回收率处于 $99\%\sim 103\%$ 范围内,RSD 为 1.38% ,表明方法准确度良好。方法学考察结果表明,线性关系、精密度、稳定性、重复性及加样回收率均符合要求,表明丹酚酸 B 含量测定方法可靠。

2.2 D-最优设计方案及结果

根据 D-最优试验设计因素水平表设计进行 55 次试验,按照试验表中建模参数轨迹建立定量模型,并对模型的性能进行评价。D-最优试验设计及结果见表 2。由表 2 可知,不同的建模参数轨迹构建的 PLS 定量模型在性能方面表现出显著的差异。因此可以看出,优选最佳建模参数对于构建稳健的定量模型尤为重要。

表 1 D-最优试验设计因素水平表

Table 1 Experimental factor levels and response variables for the D-optimal design

Variable	Type	Levels
X_1	K-S	Continuous 1 : 1 4 : 1
X_2	LVs	Discrete 1 2 3 4 5 6 7 8 9 10
X_3	P	Nominal Raw 2D MSC SNV SG9 2D-snv
X_4	V	Nominal siPLS iPLS mwPLS Full
Response variables		R_{cal}^2 RMSEC R_{pre}^2 RMSEP RPD

注: K-S 为样本集划分比例(校正集:验证集); LVs 为潜变量因子数; P 为光谱预处理方法; V 为变量筛选方法

Note: K-S: Partition ratio of calibration set and validation set; LVs: Latent variables; P: spectra pre-processing; V: variable selection

表 2 D-最优试验设计及结果

Table 2 Experimental results of D-optimal design

run	X_1	X_2	X_3	X_4	R_{cal}^2	RMSEC	R_{pre}^2	RMSEP	RPD
1	1.45	1	2D-snv	sipls	0.176 8	0.016 0	0.080 1	0.017 0	0.944 4
2	3.10	7	SG9	full	0.983 0	0.002 9	0.990 0	0.002 4	6.823 7
3	4.00	10	2D	sipls	0.994 1	0.001 8	0.998 7	0.003 2	5.239 6
4	2.58	5	MSC	mwpls	0.986 3	0.002 7	0.987 7	0.003 3	4.905 4
5	1.66	8	RAW	ipls	0.980 4	0.003 3	0.969 4	0.004 3	3.937 4
6	1.00	1	SG9	full	0.647 8	0.017 9	0.460 0	0.015 6	1.126 2
7	3.50	9	MSC	full	0.993 4	0.001 9	0.995 1	0.001 7	9.868 5
8	2.56	6	2D-snv	full	0.992 2	0.002 0	0.993 2	0.002 1	7.467 3
9	3.33	3	MSC	ipls	0.966 2	0.004 2	0.990 2	0.002 3	7.075 0
10	3.92	8	2D	mwpls	0.981 3	0.003 1	0.995 6	0.003 0	5.478 0
11	1.00	1	SNV	sipls	0.259 4	0.017 8	0.309 7	0.014 7	1.229 3
12	4.00	2	2D-snv	ipls	0.831 5	0.008 8	0.865 3	0.009 2	1.753 3
13	1.00	1	2D	mwpls	0.322 9	0.016 8	0.441 5	0.016 8	0.993 8
14	3.10	7	SG9	full	0.983 0	0.002 9	0.990 0	0.002 4	6.823 7
15	1.00	10	MSC	ipls	0.988 7	0.002 7	0.958 0	0.004 4	4.137 8
16	2.50	6	2D-snv	mwpls	0.986 0	0.002 6	0.991 0	0.002 4	6.596 6

续表 2

17	1.08	2	SNV	ipls	0.469 2	0.015 6	0.429 0	0.013 3	1.356 9
18	2.81	1	SNV	mwpls	0.220 0	0.016 2	0.104 9	0.016 5	0.998 6
19	2.47	5	SG9	ipls	0.982 5	0.003 0	0.961 5	0.004 8	3.365 0
20	1.00	9	RAW	full	0.994 0	0.001 9	0.982 3	0.003 0	5.798 0
21	2.37	5	SG9	sippls	0.946 2	0.005 3	0.943 6	0.006 1	2.722 3
22	4.00	5	SNV	full	0.976 6	0.003 5	0.988 0	0.002 7	6.167 0
23	2.26	9	2D	mwpls	0.981 9	0.003 1	0.992 6	0.003 0	5.499 7
24	2.71	2	2D	sippls	0.776 4	0.010 2	0.830 0	0.009 7	1.694 9
25	1.23	2	MSC	full	0.714 4	0.012 0	0.732 5	0.010 5	1.668 6
26	2.47	5	SG9	ipls	0.982 5	0.003 0	0.961 5	0.004 8	3.365 0
27	3.85	10	2D-snv	sippls	0.996 0	0.001 4	0.995 3	0.002 4	6.605 1
28	4.00	1	2D	mwpls	0.334 9	0.016 8	0.782 7	0.016 7	0.987 3
29	4.00	1	RAW	ipls	0.368 0	0.016 5	0.382 4	0.018 0	0.899 4
30	1.08	9	2D-snv	ipls	0.941 3	0.005 6	0.956 6	0.005 0	3.332 7
31	2.42	6	2D	full	0.990 0	0.001 9	0.992 8	0.002 1	7.880 1
32	1.65	4	2D-snv	ipls	0.881 1	0.007 4	0.939 2	0.006 0	2.665 0
33	2.50	6	2D	ipls	0.974 8	0.003 6	0.988 7	0.003 4	4.804 2
34	3.66	9	SNV	ipls	0.982 7	0.003 0	0.985 2	0.002 9	5.761 3
35	3.45	8	SNV	sippls	0.992 6	0.002 0	0.992 6	0.002 2	7.628 7
36	4.00	1	MSC	sippls	0.126 0	0.016 2	0.033 1	0.016 9	0.966 1
37	3.22	1	RAW	full	0.573 0	0.016 4	0.504 3	0.018 0	0.900 0
38	1.23	3	SNV	full	0.945 4	0.005 6	0.948 9	0.004 8	3.659 8
39	1.68	2	RAW	sippls	0.356 1	0.015 6	0.360 8	0.014 8	1.140 1
40	1.00	10	RAW	mwpls	0.993 8	0.001 9	0.965 6	0.004 2	4.217 6
41	1.00	8	SNV	mwpls	0.993 7	0.002 0	0.961 3	0.004 5	4.017 4
42	1.52	7	MSC	sippls	0.978 7	0.003 5	0.989 7	0.002 8	6.111 6
43	3.63	10	RAW	sippls	0.991 7	0.002 1	0.992 9	0.001 9	8.640 9
44	3.71	1	SG9	mwpls	0.372 9	0.016 5	0.388 8	0.017 8	0.915 5
45	1.00	10	2D	sippls	0.996 8	0.001 3	0.986 6	0.003 2	5.201 0
46	2.37	5	SG9	sippls	0.946 2	0.005 3	0.943 6	0.006 1	2.722 3
47	1.08	8	SG9	mwpls	0.984 8	0.003 0	0.967 8	0.004 0	4.416 5
48	2.02	10	2D-snv	sippls	0.996 7	0.001 3	0.993 3	0.002 9	5.368 7
49	2.27	7	RAW	sippls	0.975 0	0.003 6	0.988 5	0.002 5	6.652 9
50	3.15	10	SNV	full	0.994 3	0.001 7	0.994 5	0.001 8	8.978 0
51	1.89	1	2D	ipls	0.278 3	0.016 0	0.526 4	0.018 0	0.877 9
52	4.00	5	RAW	mwpls	0.982 3	0.003 0	0.956 2	0.004 9	3.329 1
53	2.71	2	2D	sippls	0.776 4	0.010 2	0.830 0	0.009 7	1.694 9
54	2.42	6	2D	full	0.990 0	0.001 9	0.992 8	0.002 1	7.880 1
55	4.00	10	SG9	mwpls	0.985 1	0.002 8	0.969 4	0.004 1	3.925 8

2.3 丹酚酸 B 定量模型的建模参数轨迹优选

采用 Design Expert 8.0.6 软件对试验数据进行拟合, 建立响应指标与建模参数间的拟合模型, 确定丹酚酸 B 定量模型的最佳建模参数轨迹。各模型拟合与分析结果见表 3。可知各拟合模型 P 值均小于 0.000 1, 决定系数 R^2 均在 0.99 以上, 表明拟合模型对不同建模参数轨迹下的模型性能拟合良好, 可用于实际预测。以最大 R_{cal}^2 、 R_{pre}^2 、RPD 及最小 RM-SEC、RMSEP 为目标进行建模参数轨迹优选, 拟合模型预测结果显示采用 K-S 算法从 56 份样本中选择 3/4 的校正集样本、4 000~10 000 cm^{-1} 的建模光谱波段、二阶求导和标准正则变量变换校正的预处理方法, 潜变量因子数为 7 时所建立的定量模型性能最优。

表 3 模型拟合与分析结果

响应值	模型	P 值	R^2	R_{adj}^2
Y_1	二次模型	<0.000 1	0.995 8	0.977 4
Y_2	二次模型	<0.000 1	0.999 0	0.994 9
Y_3	二次模型	<0.000 1	0.994 7	0.971 5
Y_4	二次模型	<0.000 1	0.998 7	0.993 2
Y_5	二次模型	<0.000 1	0.990 6	0.949 0

2.4 定量模型的建立与评价

以优选的建模参数轨迹建立丹酚酸 B 定量模型, 模型的 RMSEC, RMSEP 较小且均为 0.001 8, 表明模型的预测误差小, 精度高并具有较好的拓展性和稳健性; 校正决定系数

R_{cal}^2 为 0.994 0, 预测决定系数 R_{pre}^2 为 0.995 2, RPD 为 9.19, 表明模型的预测性能好, 能够准确的进行定量分析, 可确定该定量模型为最优模型。丹酚酸 B 含量的预测值和真实值的相关性良好, 见图 3。为进一步探讨全局最优定量模型的性能, 基于常规建模思路构建了丹酚酸 B 的定量模型, 采用模型评价指标对建模参数进行逐步优化, 结果显示采用 K-S 算法从 56 份样本中选择 2/3 的校正集样本、4 000~10 000 cm^{-1} 的建模光谱波段、二阶求导光谱预处理方法, 潜变量因子数为 6 时构建的定量模型性能最优。模型的 R_{cal}^2 和 R_{pre}^2 分别为 0.992 6 和 0.989 8, RMSEC 和 RMSEP 分别为 0.002 1 和 0.002 2, RPD 为 7.60, 结果表明基于系统建模思想构建的丹酚酸 B 定量模型性能更优。

3 结 论

基于系统建模思想, 采用科学系统的 D-最优设计以关键建模参数为自变量, 模型评价指标为因变量, 基于参数间的关联性进行多次建模试验。试验充分考察了建模参数间协同关系对模型性能的影响, 以最优模型性能为目标优选出最佳

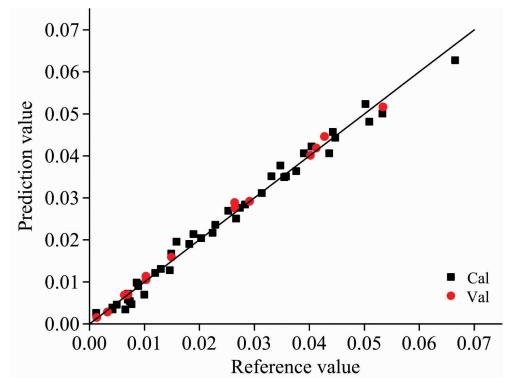


图 3 丹酚酸 B 含量预测值与真实值相关关系图

Fig. 3 Correlation between the prediction and reference values of salvianolic acid B

的建模参数轨迹, 构建了全局最优的丹酚酸 B 近红外定量模型。该模型的稳健性和预测准确度高, 可实现丹酚酸 B 含量的快速测定, 为高效、快速、无损地获取脑心通胶囊中丹酚酸 B 含量提供保证, 对脑心通胶囊的质量控制具有重要意义。

References

- [1] LIU Yan-ru, TANG Zhi-shu, SONG Zhong-xing, et al(刘妍如, 唐志书, 宋忠兴, 等). Chinese Traditional and Herbal Drugs(中草药), 2018, 49(12): 2775.
- [2] LI Geng, MENG Fan-yun, YANG Hong-jun, et al(李 耿, 孟繁蕴, 杨洪军, 等). Chinese Journal of Pharmaceutical Analysis(药物分析杂志), 2013, 33(3): 414.
- [3] TANG Hai-jiao, KUANG Yan-hui, ZHANG Si-si, et al(唐海蛟, 匡艳辉, 张偲偲, 等). China Journal of Chinese Materia Medica(中国中药杂志), 2018, 43(14): 2857.
- [4] WANG Xiao-liang, SU Ke, ZHANG Bing-hua, et al(王小亮, 苏 珂, 张秉华, 等). Chinese Journal of Pharmaceutical Analysis(药物分析杂志), 2018, 38(3): 527.
- [5] SUN Fei, XU Bing, DAI Sheng-yun, et al(孙 飞, 徐 冰, 戴胜云, 等). China Journal of Traditional Chinese Medicine and Pharmacy(中华中医药杂志), 2017, 32(12): 5316.
- [6] LEI Xiao-qing, LI Geng, WANG Xiu-li, et al(雷晓晴, 李 耿, 王秀丽, 等). Chinese Traditional and Herbal Drugs(中草药), 2018, 49(11): 2653.
- [7] Zhao N, Wu Z S, Zhang Q, et al. Scientific Reports, 2015, 5: 11647.
- [8] DU Chen-zhao, ZHAO An-bang, WU Zhi-sheng, et al(杜晨朝, 赵安邦, 吴志生, 等). Chinese Traditional and Herbal Drugs(中草药), 2017, 48(16): 3317.
- [9] ZHANG Ting-ting, LI Yi-dan, ZHENG Kai-xuan, et al(张婷婷, 李艺丹, 郑凯旋, 等). Chinese Traditional and Herbal Drugs(中草药), 2018, 49(15): 3639.
- [10] Du C Z, Dai S Y, Zhao A, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 222: 117267.

Study on Establishment of Near-Infrared Quantitative Model for Salvianolic Acid B in Naoxintong Capsule Based on the System Modeling Idea

GAO Rui-lin¹, YANG Peng-shuo¹, XU Gang², WU Xiao-wen¹, YANG Chang¹, SHI Xin-yuan^{1*}

1. School of Chinese Pharmacy, Beijing University of Chinese Medicine, Beijing 102488, China

2. Shanxi Buchang Pharmaceutical Co., Ltd., Xi'an 710075, China

Abstract NIR quantitative modeling mainly involves four process parameters: the selection of sample set, spectra pre-processing, latent variables and variable selection methods. And in the traditional method, the optimization of the PLS modeling parameters was step by step according to the model evaluation indexes and the inter-influence among the parameters was rarely considered in the model development. This is risky because the modeling path is not necessarily the best approach to step by step optimization. The system modeling method is global trajectory parameter optimization idea based on the correlation between the parameters, which is a systematic approach to improve the efficiency and accuracy in the development of a quantitative model. The purpose of this study is based on the system modeling idea to develop a NIR quantitative model for the analysis of salvianolic acid B in Naoxintong capsule. The content of salvianolic acid B in 56 samples of *Salvia miltiorrhiza* was determined by high-performance liquid chromatography (HPLC), and the near-infrared spectra were also collected. D-optimization design method was used to optimize the modeling parameters including the sample set partition, spectral pre-processing, latent variable factor number and variable selection. The global optimal parameter trajectory was applied in the preparation of the quantitative analysis model. The results showed that the best quantitative model was developed by the spectral data pretreatment of 2D+SNV (2nd Derivative + Standard Normal Variate), choosing the latent variable factor number of 7 and spectral region of 4 000~10 000 cm^{-1} in combination with the 56 samples were randomly divided into a calibration set and a validation set according to the proportion of 3 : 1 by a K-S algorithm. Its standard deviation of calibration (RMSEC) and prediction (RMSEP) were both 0.001 8, which demonstrated the high analytical accuracy and robust fitting. The calibration coefficient of determination (R_{cal}^2) and the prediction coefficient of determination (R_{pre}^2) were 0.994 0 and 0.995 2, respectively. The ratio of the standard error of prediction to the standard deviation (RPD) was 9.19, further confirming that the model can be used for high-quality quantitative analysis. Based on the system modeling idea, the D-optimal design approach was adopted to implement the global trajectory parameter optimization in the process of establishing the quantitative model in this paper. Furthermore, a quantitative model for salvianolic acid B was developed with robust predictability and high accuracy. The model provided an efficient and rapid method to quantify salvianolic acid B of Naoxintong capsule, which was of great significance for the quality control of intermediates and products of Naoxintong capsule.

Keywords System modeling; D-optimal design; Near-infrared spectroscopy(NIR); Naoxintong capsule; Salvianolic acid B

(Received Nov. 12, 2019; accepted Mar. 7, 2020)

* Corresponding author