·药剂与炮制·

近红外光谱结合主成分分析和聚类分析鉴别炉甘石生品、伪品和炮制品

张晓冬[1]，陈龙[2]，白玉[3]，陈科力[1*]

（1. 湖北中医药大学 教育部中药资源和中药复方重点实验室，武汉 430065；
2. 南漳县人民医院，湖北 襄阳 441500；3. 马应龙药业集团股份有限公司，武汉 430065）

［摘要］ **目的**：利用主成分判别分析和聚类分析建立炉甘石生品、伪品和炮制品的近红外光谱鉴别模型。**方法**：采集炉甘石生品、伪品和炮制品的近红外光谱，每类样品随机划分为训练集和测试集。对光谱预处理方法和建模谱段进行筛选，分别建立主成分判别分析模型及聚类分析模型。**结果**：光谱经一阶导数预处理，主成分判别分析模型的特征谱段为 4 800～4 000 cm^{-1}，聚类模型的特征谱段为 7 300～7 000，4 800～4 000 cm^{-1}。在主成分判别分析模型中，预测准确率 94.34%；在聚类模型中，模型的预测准确率 96.23%。**结论**：所建立的近红外主成分判别分析模型和聚类分析模型均可用于炉甘石生品、伪品和炮制品的鉴别。

［关键词］ 炉甘石；解毒明目；近红外光谱；主成分分析；聚类分析；炮制品；伪品

［中图分类号］ R22；R283；R284；R943.1；O433 ［文献标识码］ A ［文章编号］ 1005-9903(2018)12-0001-08

［doi］ 10.13422/j.cnki.syfjx.20181004

［网络出版地址］ http://kns.cnki.net/kcms/detail/11.3495.R.20180309.1030.025.html

［网络出版时间］ 2018-03-09 11:34

Identification of Crude Products, Counterfeit Products and Processed Products of Calamina by Near Infrared Spectroscopy, Principal Component Analysis and Cluster Analysis

ZHANG Xiao-dong[1], CHEN Long[2], BAI Yu[3], CHEN Ke-li[1*]

(1. *Key Laboratory of Traditional Chinese Medicine Resource and Compound Prescription, Ministry of Education, Hubei University of Chinese Medicine, Wuhan* 430065, *China*;
2. *Nanzhang People's Hospital, Xiangyang* 441500, *China*;
3. *Mayinglong Pharmaceutical Group Co. Ltd., Wuhan* 430065, *China*)

［**Abstract**］ **Objective**: To establish a near infrared spectral discriminant model of crude products, counterfeit products and processed products of Calamina by principal component analysis and cluster analysis. **Method**: Near infrared spectra of crude products, counterfeit products and processed products of Calamina were collected. Each category sample was randomly divided into training set and testing set. The spectral preprocessing methods and modeling spectral bands were screened, the principal component discriminant analysis model and the cluster analysis model were established respectively. **Result**: Spectrs were preprocessed by the first derivative. The characteristic spectral band of the principal component discriminant analysis model was 4 800-4 000 cm^{-1}, and the characteristic spectral bands of the cluster analysis model were 7 300-7 000, 4 800-4 000 cm^{-1}. In the

［收稿日期］ 20171013(005)

［基金项目］ 武汉市 2012 年高新技术产业发展行动计划生物技术与新医药专项(201260523193)；国家中药标准化项目(1399)

［第一作者］ 张晓冬，在读硕士，从事中药资源及其品质研究，E-mail：2352874015@qq.com

［通信作者］ *陈科力，教授，从事中药资源及其品质研究，Tel：027-68890106，E-mail：Kelichen@126.com

principal component discriminant analysis model, the prediction accuracy rate was 94.34%. In the cluster analysis model, the prediction accuracy rate was 96.23%. **Conclusion**: The principal component analysis model and cluster analysis model of near infrared spectra can be used for identification of crude products, counterfeit products and processed products of Calamina.

[**Key words**] Calamina; detoxification and eyesight; near infrared spectroscopy; principal component analysis; cluster analysis; processed products; counterfeit products

炉甘石来源于碳酸盐类矿物方解石族菱锌矿或碳酸盐类矿物水锌矿[1],炮制后具有解毒明目退翳、收湿止痒敛疮的功效,是一种常用的外用药[2]。据文献报道[3-4]和本课题组进行的市场调查验证,市售炉甘石存在严重的假冒伪劣现象。基于此,本实验拟建立炉甘石生品、伪品和炮制品的近红外光谱鉴别方法,对炉甘石的质量进行控制。

传统的矿物药鉴别主要依据性状和化学方法,而性状鉴别的主观性较强且需一定的专业知识和经验,适用性不强;化学方法的专属性不强,往往相似的物质都可以发生同样的反应。而近红外光谱分析技术在矿物药的鉴别上已取得了很好的效果[5-6],具有快速、无损、环保的优点[7]。因此,本实验利用近红外光谱技术结合主成分判别分析和聚类分析建立炉甘石生品、伪品和炮制品的鉴别模型,以实现对三者的快速准确鉴别。

1 材料

XPertPro 型 X 射线粉晶衍射仪(荷兰帕纳科公司),MPA 型傅里叶变换近红外光谱仪(德国 Bruker 公司,OPUS 7.5 光谱分析软件)。MDI Jade 6.0 软件(美国 Materials Data 公司),Unscrambler 9.7 数据分析软件(挪威 CAMO 软件公司)。

本课题组前期收集了各大药材市场的炉甘石样品,共计 32 批次,建立了炉甘石生品、伪品和炮制品的鉴别模型,取得了较好的效果[8]。在上述建模样品的基础上,继续从炉甘石矿产区(湖南、广西、云南、四川、贵州等地)收集了 18 批样品,从亳州、禹州、安国等药材市场收集了共 34 批样品,另外武汉马应龙药业股份有限公司提供了 6 批样品。所有的样品均经 X 射线粉晶衍射仪进行物相分析并鉴别其真伪[9],采用 2015 年版《中国药典》[2]中含量测定方法测定氧化锌的含量。在收集到的 58 批样品中,经鉴定发现 36 批为炉甘石生品,其中仅 2 批含菱锌矿,其余均为水锌矿;22 批为伪品。为了增加模型样品的代表性,经配制又得到 6 批含有菱锌矿的炉甘石生品。另外,从生品中随机挑选出 38 批次,从伪品样品中随机挑选出 13 批次,按 2015 年版《中国药典》[2]中炉甘石的炮制方法进行炮制。将上述所得的炉甘石生品、伪品和炮制品样品随机划分为训练集和测试集,并保证每类炉甘石训练集样品数要大于测试集样品数,炉甘石样品信息详见表 1。

表 1 炉甘石样品的信息

Table 1 Information of Calamina samples

No.	来源	物相	ZnO/%	鉴别结果	No.	来源	物相	ZnO/%	鉴别结果
1	安徽亳州	水锌矿、异极矿、白云石	69.95	生品	12	马应龙	水锌矿、异极矿、方解石、石膏	68.95	生品
2	湖南矿区	水锌矿、异极矿	59.25	生品	13	马应龙	水锌矿、异极矿、白云石	69.82	生品
3	湖南矿区	水锌矿	67.79	生品	14	河北安国	水锌矿	69.59	生品
4	广西矿区	水锌矿、异极矿、方解石	61.83	生品	15	河北安国	水锌矿、石英、黑磷铁钠石	70.08	生品
5	广西矿区	水锌矿、异极矿、方解石	67.41	生品	16	广东深圳	水锌矿、异极矿、方解石、石膏	66.92	生品
6	贵州矿区	水锌矿、菱锌矿、白云石	65.50	生品	17	广东深圳	水锌矿、异极矿、白云石、方解石、石膏	54.27	生品
7	四川矿区	水锌矿、异极矿	54.74	生品	18	贵州矿区	水锌矿、异极矿、白云石、方解石	38.12	生品
8	云南矿区	水锌矿、异极矿	66.57	生品	19	安徽亳州	水锌矿、异极矿、白云石、方解石、石英、云母、石膏	42.41	生品
9	云南矿区	水锌矿、异极矿	64.88	生品					
10	广西矿区	水锌矿、异极矿、白云石	43.21	生品	20	6+73+生	-	-	生品
11	马应龙	水锌矿、异极矿、白云石	69.24	生品	21	6+73+生	-	-	生品

续表 1

No.	来源	物相	ZnO/%	鉴别结果
22	6 + 73 + 生	–	–	生品
23	江西樟树	方解石	0	伪生品
24	河南禹州	方解石	0	伪生品
25	84 + 制	–	–	伪制品
26	27 + 制	–	–	伪制品
27	江西樟树	方解石	0	伪生品
28	广西玉林	方解石	0	伪生品
29	陕西西安	方解石	0	伪生品
30	河南禹州	方解石	0	伪生品
31	安徽亳州	方解石	0	伪生品
32	33 + 制	–	–	伪制品
33	陕西西安	方解石、石英	0	伪生品
34	35 + 制	–	–	伪制品
35	安徽亳州	方解石、石英	0	伪生品
36	89 + 制	–	–	伪制品
37	江西樟树	方解石、石英、云母	0	伪生品
38	陕西西安	方解石	0	伪生品
39	安徽亳州	方解石	0	伪生品
40	河南禹州	方解石	0	伪生品
41	安徽亳州	方解石、石英	0	伪生品
42	广西矿区	高岭石、方解石	1.60	伪生品
43	14 + 制	–	–	炮制品
44	17 + 制	–	–	炮制品
45	16 + 制	–	–	炮制品
46	72 + 制	–	–	炮制品
47	19 + 制	–	–	炮制品
48	64 + 制	–	–	炮制品
49	63 + 制	–	–	炮制品
50	1 + 制	–	–	炮制品
51	65 + 制	–	–	炮制品
52	2 + 制	–	–	炮制品
53	73 + 制	–	–	炮制品
54	3 + 制	–	–	炮制品
55	4 + 制	–	–	炮制品
56	66 + 制	–	–	炮制品
57	7 + 制	–	–	炮制品
58	10 + 制	–	–	炮制品
59	64 + 制	–	–	炮制品
60	66 + 制	–	–	炮制品
61	7 + 制	–	–	炮制品
62	69 + 制	–	–	炮制品
63	湖南矿区	水锌矿、异极矿	62.91	生品
64	广西矿区	水锌矿、方解石	67.55	生品
65	广西矿区	水锌矿、方解石	66.88	生品
66	贵州矿区	水锌矿、异极矿、白云石	53.02	生品
67	四川矿区	水锌矿、异极矿、白云石	51.57	生品
68	云南矿区	水锌矿、异极矿	63.72	生品
69	广西矿区	水锌矿、异极矿	64.20	生品
70	马应龙	水锌矿、异极矿、白云石	52.92	生品
71	马应龙	水锌矿	70.51	生品
72	马应龙	水锌矿、异极矿、白云石	69.98	生品
73	河北安国	水锌矿、菱锌矿、石膏	68.58	生品
74	广东深圳	水锌矿、长石	70.42	生品
75	广东深圳	水锌矿、异极矿、方解石、云母	45.33	生品
76	广东深圳	水锌矿、异极矿、石膏	66.91	生品
77	广东深圳	水锌矿、石膏	66.35	生品
78	安徽亳州	水锌矿、异极矿、方解石、石膏、白云石	48.70	生品
79	安徽亳州	水锌矿、石英、白云石、石膏	47.09	生品
80	6 + 73 + 生	–	–	生品
81	6 + 73 + 生	–	–	生品
82	6 + 73 + 生	–	–	生品
83	24 + 制	–	–	伪制品
84	河南禹州	方解石	0	伪生品
85	27 + 制	–	–	伪制品
86	30 + 制	–	–	伪制品
87	江西樟树	方解石	0	伪生品
88	广东深圳	方解石	0	伪生品
89	江西樟树	方解石、石英	0	伪生品
90	河北安国	方解石	0	伪生品
91	39 + 制	–	–	伪制品
92	90 + 制	–	–	伪制品
93	安徽亳州	方解石、石英	0	伪生品
94	41 + 制	–	–	伪制品
95	河南禹州	方解石、石英	0	伪生品
96	42 + 制	–	–	伪制品
97	36 + 制	–	–	伪制品
98	12 + 制	–	–	炮制品
99	68 + 制	–	–	炮制品
100	71 + 制	–	–	炮制品

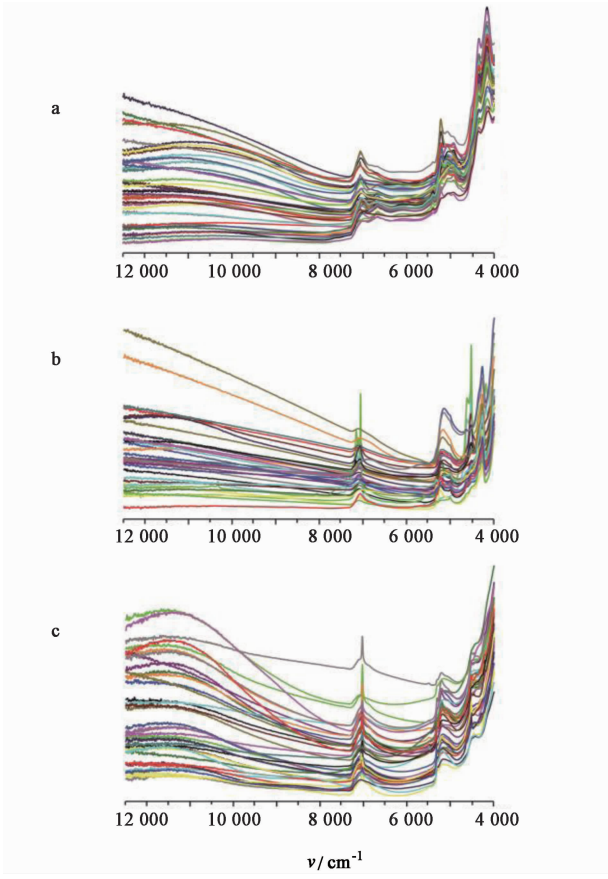
续表 1

No.	来源	物相	ZnO/%	鉴别结果	No.	来源	物相	ZnO/%	鉴别结果
101	78 + 制	–	–	炮制品	109	65 + 制	–	–	炮制品
102	69 + 制	–	–	炮制品	110	6 + 制	–	–	炮制品
103	63 + 制	–	–	炮制品	111	8 + 制	–	–	炮制品
104	1 + 制	–	–	炮制品	112	3 + 制	–	–	炮制品
105	64 + 制	–	–	炮制品	113	5 + 制	–	–	炮制品
106	2 + 制	–	–	炮制品	114	6 + 制	–	–	炮制品
107	73 + 制	–	–	炮制品	115	8 + 制	–	–	炮制品
108	2 + 制	–	–	炮制品					

注:“数字 + 数字 + 生”表示由这 2 种编号的样品粉末混合均匀得到;“数字 + 制”表示由该编号的样品经炮制得到;马应龙全名为马应龙药业集团股份有限公司。炮制品. 由生品炮制所得;伪制品. 由伪品炮制所得;伪生品. 伪品未经炮制;No. 1 ~ 62. 训练集样品;No. 63 ~ 115. 测试集样品。

2 方法与结果

2.1 近红外光谱采集 将上述 115 批样品粉碎并过 60 目筛,分别取 2 g 置于样品杯中,采用积分球漫反射测试模式扫描近红外光谱。光谱扫描范围 12 500 ~ 4 000 cm^{-1},扫描数 32 次,仪器分辨率 8 cm^{-1}。每个样品重复扫描 3 次,取平均光谱作为该样品的分析光谱,见图 1。观察各类炉甘石样品的近红外图谱,对于正品的同类别生品和炮制品,样品间的图谱相近;对于伪品样品,各样品的图谱差异较大。但总体而言,炉甘石生品、伪品和炮制品的图谱差异主要在谱段 7 300 ~ 7 000,5 550 ~ 4 800,4 800 ~ 4 000 cm^{-1}。



a. 生品;b. 伪品;c. 炮制品

图 1 炉甘石样品的近红外光谱

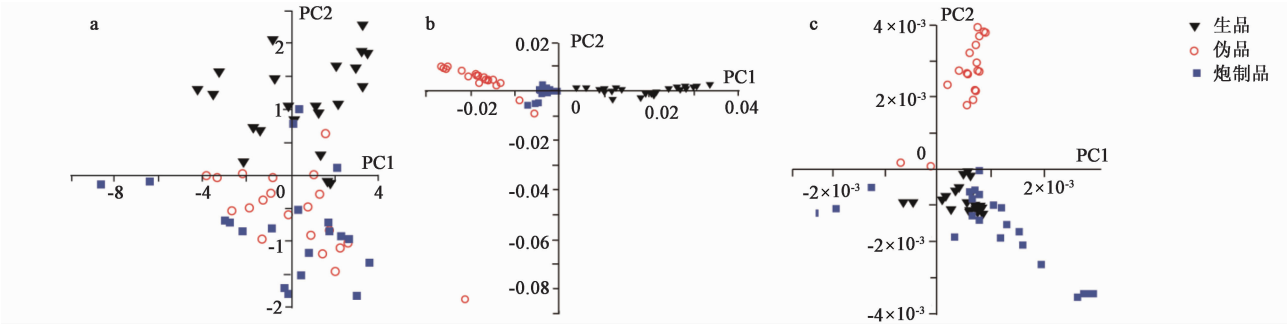
Fig. 1 Near-infrared spectra of Calamina samples

2.2 主成分判别分析

2.2.1 预处理方法筛选及主成分分析 由于近红外光谱除了包含自身信息外,还包含了其他无关信息和噪声,因此在采用化学计量学方法建立模型时,需要对光谱进行预处理以消除无关信息和噪声的干扰[10]。常用的光谱预处理方法有一阶导数法(first derivative, FD),二阶导数法(second derivative, SD),多元散射校正法(multiplicative scatter correction, MSC),矢量归一化法(vector normalization, VN)等。另外,由于原始光谱是高维数据集,光谱中具有多重共线性、信息重叠等不利因素,若直接用来建立模型,会出现过拟合现象,模型分析精度也随之降低。故采用主成分分析(PCA)对光谱进行降维处理[11]。

为确定最佳的光谱预处理方法,利用 Unscrambler 9.7 数据分析软件,在训练集样品的全谱段范围,分别对未处理光谱,FD 及 SD 预处理后的光谱进行 PCA 降维,得到各主成分得分值。分别将训练集各样品的第 1 个主成分(PC1)和第 2 个主成分(PC2)的主成分得分值作为该样品在二维坐标系下的坐标值,训练集样品未预处理及不同预处理后的光谱 PC1 和 PC2 得分散点图见图 2。结果发现训练集样品只有经 FD 预处理后,同类样品点彼此

靠近,异类样品点彼此分离,即生品、伪品和炮制品可以相互区分。故确定 FD 为最佳预处理方法。



a. 未处理;b. 一阶导数预处理;c. 二阶导数预处理

图 2　炉甘石训练集样品在全谱段下第 1,2 个主成分的得分散点

Fig. 2　Score scatter plots of first and second principal components in full spectrum of Calamina training set samples

2.2.2　特征谱段的筛选　在利用近红外光谱建立定性校正模型时,筛选特征谱段,一方面可以简化模型,另一方面由于剔除了不相关的变量,可以增强模型的预测能力和稳健性[12]。由图 1 可知,结果发现炉甘石生品、伪品和炮制品的区别主要在谱段 7 300 ~ 7 000 cm^{-1}(A),5 550 ~ 4 800 cm^{-1}(B)和 4 800 ~ 4 000 cm^{-1}(C),但并不能确定这 3 个谱段均为特征谱段。训练集样品光谱经 FD 预处理,在这 3 个谱段以及三者的组合谱段内分别进行 PCA,以各样品的 PC1 和 PC2 的得分值作为该样品在二维坐标系的坐标值,得不同谱段下 PC1 和 PC2 得分散点图,见图 3。结果发现在谱段 C,A + C,B + C,A + B + C 下,同类样品点彼此靠近,异类样品点彼此分离,即生品、伪品和炮制品可以相互区分;而在其他谱段下,不同类别样品点的分布存在交叉重叠,即不能相互区分;另外,由于在谱段 C,A + C,B + C,A + B + C 的样品主成分得分散点图分类效果类似,为了简化模型,确定 C 谱段为特征谱段。

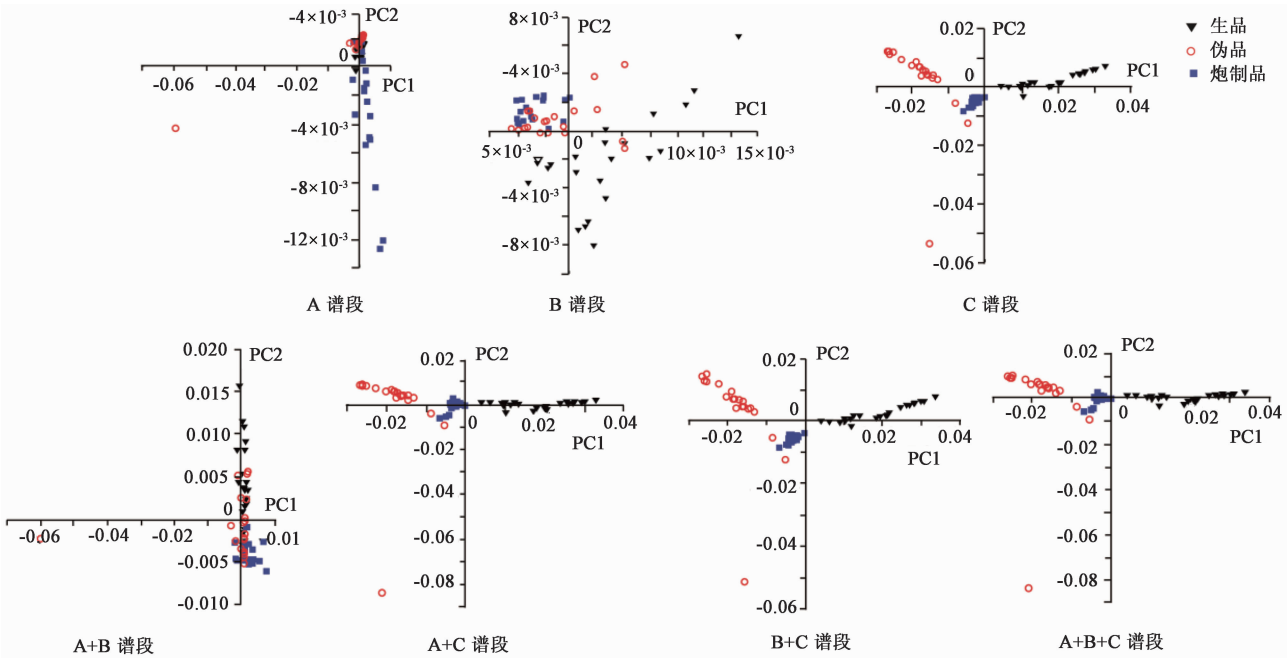
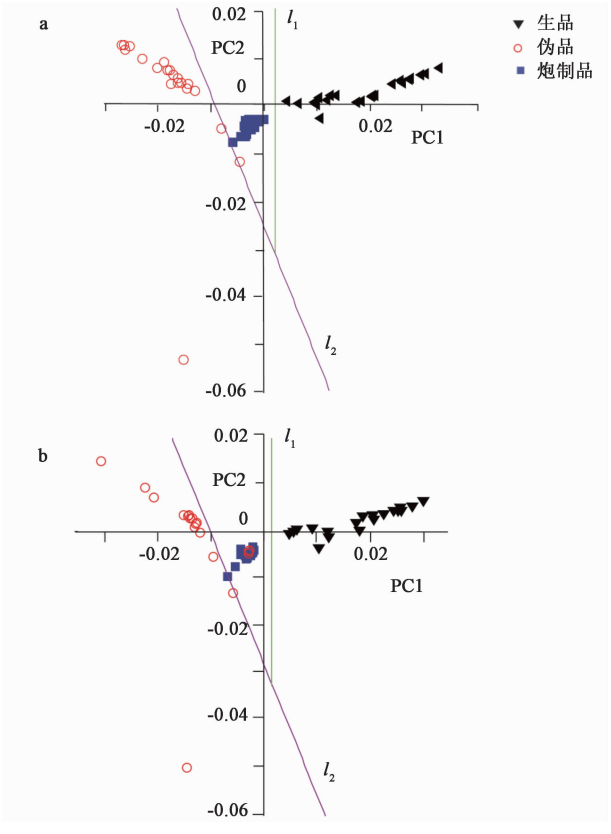


图 3　炉甘石训练集样品在不同谱段下第 1,2 个主成分的得分散点

Fig. 3　Score scatter plots of first and second principal components under different spectral segments of Calamina training set samples

2.2.3　主成分判别模型的建立　在上述分析中,确定了 FD 为最佳预处理方法,4 800 ~ 4 000 cm^{-1}为建模谱段。为了建立 PC1 和 PC2 得分判别模型,需建立区分生品、伪品和炮制品在二维坐标系下的判别方程。参照支持向量机分类的原理[13],在解决线性可分的问题时,寻找 1 个超平面,使得样品集到分

类超平面的距离最大;在二维坐标系下,则是寻找 1 条线使得样品集到分界线的距离最大。具体计算过程是在 PC1 和 PC2 得分散点图上,找到生品与炮制品最靠近的 2 个样品点,将这 2 个样品点的中垂线(l_1)作为区分生品和炮制品的分界线,结果分界线的方程为PC1 = 2.165 8 × 10^{-3};另外,找到伪品与炮制品最靠近的 2 个样品点,连接 2 点得到 1 条直线,同时找到炮制品最靠近伪品的样品点,以该点作上述直线的平行线,将这 2 条平行线的中线(l_2)作为区分伪品和炮制品的分界线,分界线的方程为 PC2 = −2.780 2PC1 −0.026 1。具体的判别条件为当PC1 > 2.165 8 × 10^{-3} 且 PC2 > −2.780 2 PC1 −0.026 1,判断为生品;当 PC1 < 2.165 8 × 10^{-3} 且 PC2 > −2.780 2PC1 −0.026 1,判断为炮制品;当 PC2 < −2.780 2PC1 −0.026 1,判断为伪品。见图 4(a)。



a. 训练集;b. 测试集

图 4　炉甘石样品的主成分判别

Fig. 4　Principal component discrimination diagrams of Calamina samples

2.2.4　模型的验证及评价　将表 1 中验证集样品光谱经 FD 预处理,在谱段 4 800 ~ 4 000 cm^{-1} 进行 PCA,将样品的 PC1 和 PC2 得分值作为坐标值,带入所建炉甘石生品、伪品和炮制品的主成分判别模型中,对模型进行验证。测试集样品主成分判别图见图 4(b)。结果发现测试集样品中生品和炮制品全部预测准确,伪品有 3 个预测错误(样品 83 ~ 85),模型的预测准确率为 94.34%。显然,模型的准确度高、预测能力强,可以用于炉甘石生品、伪品及炮制品的鉴别。

2.3　聚类分析

2.3.1　预处理方法的筛选　选取炉甘石训练集样品,共计 62 批次,在全谱段范围,对未处理光谱,FD 预处理和 SD 预处理光谱分别进行聚类分析,见表 2。结果发现训练集样品光谱经 FD 预处理后聚类分析的正确率最高,故确定 FD 为最佳预处理方法,其聚类效果见图 5。

表 2　炉甘石训练集样品在全谱段范围不同预处理方法下的聚类分析

Table 2　Cluster analysis of Calamina training set samples under different pretreatment method of full spectrum

预处理方法	类别	数目/个	分类正确数/个	正确率/%
未处理	生品	22	16	58.06
	伪品	20	11	
	炮制品	20	9	
FD	生品	22	22	95.16
	伪品	20	17	
	炮制品	20	20	
SD	生品	22	19	91.94
	伪品	20	18	
	炮制品	20	20	

2.3.2　特征谱段的筛选　训练集样品光谱经过 FD 预处理后,在 A,B,C 谱段以及三者的组合谱段下,分别进行聚类分析,见表 3。结果发现训练集样品的光谱经过 FD 预处理后在谱段 5 550 ~ 4 800,4 800 ~ 4 000 cm^{-1}(B + C)或 7 300 ~ 7 000,4 800 ~ 4 000 cm^{-1}(A + C)或 7 300 ~ 7 000,5 500 ~ 4 800,4 800 ~ 4 000 cm^{-1}(A + B + C)下进行聚类分析,准确率达 95.16%,相较于全谱段下,模型的准确率虽然没有提高,但是模型的输入变量都有所减少,在模型效果相同的情况下,优选需要输入变量最少的谱段,有利于保持模型的稳健性,故确定 7 300 ~ 7 000,4 800 ~ 4 000 cm^{-1}(A + C)为特征谱段。

2.3.3　模型的建立、验证及评价　通过对预处理方法和特征谱段的优化,确定光谱的最佳预处理方法为 FD 预处理,特征谱段为 7 300 ~ 7 000,4 800 ~ 4 000 cm^{-1},采用 Ward's algorithm 标准算法,建立的

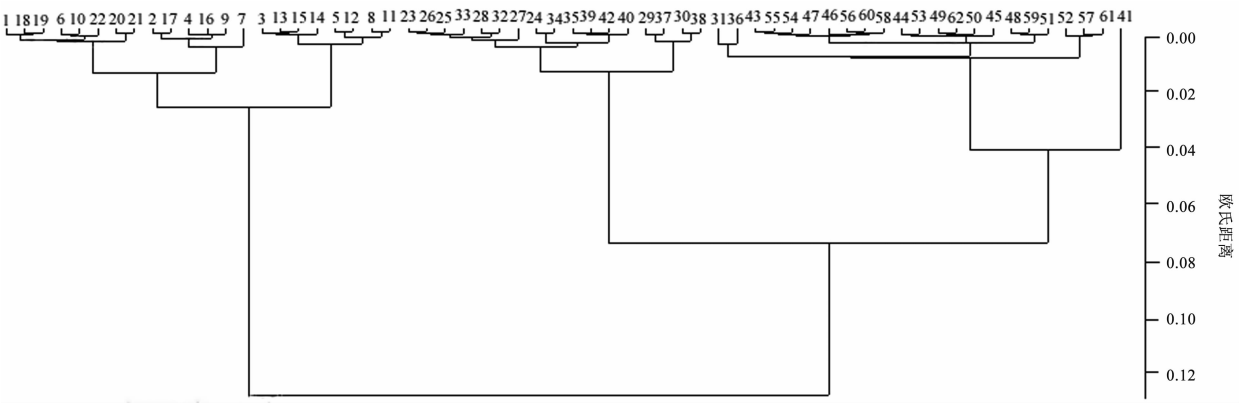


图 5　炉甘石训练集样品在全谱段一阶导数预处理下的聚类分析

Fig. 5　Cluster analysis of Calamina training set samples in first derivative preprocessing of full spectrum

表 3　炉甘石训练集样品在不同谱段下的聚类分析

Table 3　Cluster analysis of Calamina training set samples under different spectral segments

建模谱段/cm^{-1}	类别	数目/个	分类正确数/个	正确率/%
A	生品	22	20	74.19
	伪品	20	13	
	炮制品	20	13	
B	生品	22	17	74.19
	伪品	20	14	
	炮制品	20	15	
C	生品	22	22	90.32
	伪品	20	14	
	炮制品	20	20	
A + B	生品	22	17	87.10
	伪品	20	17	
	炮制品	20	20	
B + C	生品	22	22	95.16
	伪品	20	17	
	炮制品	20	20	
A + C	生品	22	22	95.16
	伪品	20	17	
	炮制品	20	20	
A + B + C	生品	22	22	95.16
	伪品	20	17	
	炮制品	20	20	

炉甘石生品、伪品和炮制品的鉴别模型效果良好。将测试集的 53 批样品的近红外光谱调入模型中，验证模型的准确性。结果发现 53 批样品中有 2 批伪品聚类错误（样品 87 和 92），其他样品均聚类正确，见图 6。模型的预测准确率 96.23%，说明模型的预测能力强，可用于鉴别炉甘石生品、伪品和炮制品。

3　讨论

本课题组前期收集了各大药材市场的炉甘石样品，共计 32 批次，采用近红外光谱法结合多参考相关系数法和人工神经网络算法建立了炉甘石生品、伪品和炮制品的鉴别模型，模型的预测准确率为 95.00%，模型的效果好[8]。本文在上述建模样品的基础上，继续收集了炉甘石主要矿产区的样品，同时新增了一些药材市场的样品，使所收集的炉甘石样品来源更广泛，更具有代表性，使建立的模型适用性增强。同时，本文采用 PCA 和聚类分析 2 种方法建立了炉甘石的近红外光谱鉴别模型，预测准确率分别为 94.34% 和 96.23%，模型的效果均良好，均可以用于炉甘石生品、伪品和炮制品的鉴别，同时可以使近红外光谱数据的多种化学计量学方法相互印证。

从建模的难度上来看，人工神经网络模型建立的难度更大。首先需要对输入变量进行压缩，其中压缩变量的方法就包括了 PCA 等；然后还要对建模参数进行选择和设置，例如隐含层节点数、学习速率、传播函数等；并且需要经过反复的训练，最终才能得到较好的神经网络模型。其突出的优点就是有自学习、自组织、自适应能力，很强的容错能力，分布储存与并行处理信息的功能及高度非线性表达能力[14]。本文采用的 PCA 和聚类分析则只需要对光谱的预处理方法及建模谱段进行筛选，选取最佳的预处理方法和建模谱段，从而建立最佳的模型，相比较而言，建模的难度大为降低。

从预测的结果上来看，人工神经网络模型测试集样品数为 20 个，其中 19 个样品预测正确，1 个样品预测错误，把 1 个炮制品误判为伪品[8]。PCA 模型测试集样品预测错误的 3 个，编号为 83 ~ 85；聚类分析模型测试集样品预测错误的 2 个，编号为 87 和 92，均把伪品误判为炮制品。相互印证结果，说明这 2 种方法判断不一致的样品 83 ~ 85，87 和 92

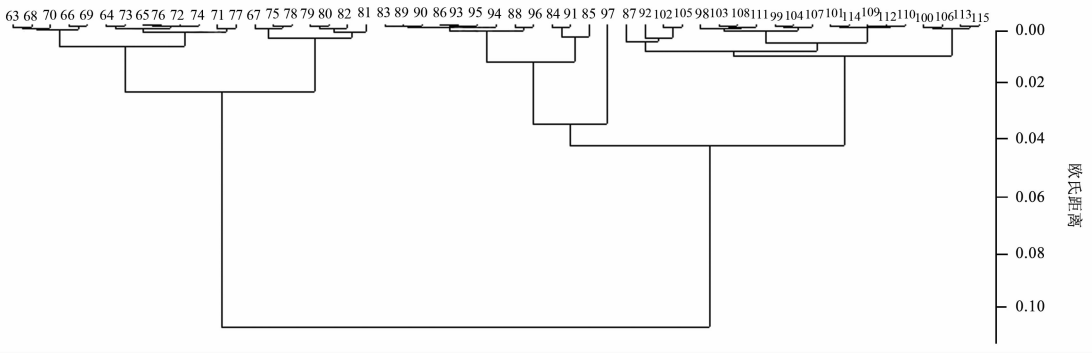


图 6　炉甘石测试集样品的聚类分析

Fig. 6　Cluster analysis of Calamina testing set samples

等可能被误判，因此可通过进一步分析（例如含量测定或 X 衍射）来判定其真伪，这将有效缩小滴定分析范围，极大地提高分析的可靠性。以上分析结果也说明伪品用所建立的近红外模型判断失误的可能性较大，因为伪品的来源较广，而作者收集到的伪品样品总是有限的。因此，用 2 种或多种化学计量学方法建立近红外光谱鉴别模型，对待测样品的近红外光谱数据进行预测，可以相互印证，避免 1 种近红外鉴别模型的不足，以保证对样品最终判断结果的准确度。

PCA 属于有监督的模式识别方法，是从已知的各种分类结果中总结出规律，训练得到判别函数，当新样品进入时，判断其与判别函数之间的相似程度，从而确定归属类别。本文提取了炉甘石近红外图谱的前 2 个主成分，建立了主成分得分二维图，并在图中确定了判别函数，当进入新样品时，样品以样品点的形式出现在二维图中，与判别函数的关系显而易见，结果的直观性强，适用于大量样品的定性鉴别。聚类分析是一种无监督的模式识别方法，开始各个样品自成一类，计算各样品之间的距离，并将距离最近的 2 个样品聚为一类。然后选择并计算类与类之间的距离，将距离最近的两类合并，直到所有样品聚为一类。聚类分析也很简单、直观，但是聚类分析的结果直接依赖于聚类变量，一些特殊变量可能会对聚类结果产生很大的影响，并且当样本量较大时，聚类分析会比较困难。上述 2 种方法的结合应用，可以避免各自的一些弊端。

［参考文献］

［1］ 国家中医药管理局《中华本草》编委会. 中华本草［M］. 上海：上海科学技术出版社，1999：382-383.
［2］ 国家药典委员会. 中华人民共和国药典. 一部［M］. 北京：中国医药科技出版社，2015：227.
［3］ 周灵君，张丽，路长珍，等. 市售生、煅炉甘石的成分分析及质量评价［J］. 中国药房，2010，21（27）：2534-2536.
［4］ 张杰红，刘友平，施学娇，等. 市售炉甘石的化学成分及抑菌活性研究［J］. 中药与临床，2011，2（6）：16-18.
［5］ 徐子杰，陈龙，刘义梅，等. 基于多参考相关系数法和 BP-ANN 建立紫石英的近红外光谱定性模型［J］. 中国实验方剂学杂志，2017，23（22）：37-42.
［6］ 陈科力，陈龙. 矿物类中药的近红外光谱鉴别方法［J］. 中南民族大学学报：自然科学版，2014，33（4）：52-56.
［7］ 杨哲萱，周立红，章顺楠，等. NIRS 技术在中药生产中的应用及其验证方法探讨［J］. 中草药，2013，44（10）：1342-1348.
［8］ SUN Y B，CHEN L，HUANG B S，et al. A rapid identification method for Calamine using near-infrared spectroscopy based on multi-reference correlation coeffient method and back propagation artificial neural network［J］. Appl Spectrosc，2017，71（4）：1447-1456.
［9］ 杨连菊，张志杰，李娆娆，等. 基于物相与成分分析的炉甘石基源研究［J］. 光谱学与光谱分析，2011，31（11）：3092-3097.
［10］ 褚小立，袁洪福，陆婉珍. 近红外分析中光谱预处理及波长选择方法进展与应用［J］. 化学进展，2004，16（4）：528-542.
［11］ Jailais B，Pinto R，Barros A S，et al. Out-product analysis using PCA to study the influence of temperature on NIR spectra of water［J］. Vib Spectrosc，2005，39（1）：50-58.
［12］ 杨皓旻，卢启鹏，黄富荣. 近红外光谱分析中建模光谱宽度的选择［J］. 红外与毫米波学报，2011，30（6）：522-525.
［13］ Vapnik V N. *The Nature of Statistical Learning Theory*［M］. Berlin：Springer，1999：988-999.
［14］ HUA X，ZHANG G，YANG J W，et al. Theory study and application of the BP-ANN method for power grid short-term load forecasting［J］. 中兴通讯技术：英文版，2015，13（3）：2-5.

［责任编辑　刘德文］