

FT-NIR 与主成分分析在中药贝母 鉴别和聚类中的应用研究

邓波^① 周玉荣 刘志宏 庞小峰

(电子科技大学生命科学与技术学院 成都市 610054)

摘 要 近红外漫反射光谱和主成分分析法相结合,选取不同范围内的近红外光谱数据,对川贝、浙贝以及东北平贝共 48 个样品进行了主成分分析,作二维线性投影图和三维线性投影图,比较了 48 个样品在红外光谱上的差异程度,发现基于傅里叶变换近红外光谱的主成分分析投影图能够较好的表征 48 个样品的类别关系。结果表明:应用近红外漫反射光谱法能够鉴别贝母及其近缘物种,具有较好的聚类 and 鉴别作用,作为一种药材质量检测手段,该方法具有一定的应用价值。

关键词 主成分分析, 近红外光谱, 贝母, 系统聚类。

中图分类号: O 657. 33

文献标识码: A

文章编号: 1004-8138(2006)05-1103-04

1 引言

川贝母属百合科植物川贝母 *Fritillaria cirrhosa* D. Don 的干燥鳞茎,主要产于四川。性味苦寒,清热润肺,化痰止咳。而浙贝母属百合科植物浙贝母 *F. thunbergii* Miq., 大者除去芯芽,习称大贝,小者不去芽,习称珠贝。东贝母为 *F. thunbergii* Miq. var. *chekiangensis* Hsiao et K. C. Hsia 的干燥鳞茎^[1]。实际上我国贝母属植物已超过 50 种(变种)^[2]。根据中医传统用药习惯,临床处方一般只分川贝母和浙贝母两大类。由于贝母原植物种类较多,商品贝母来源复杂,市场上伪品混杂,影响了临床用药的疗效。贝母普遍采用人工性状鉴别,既需要大量的人力,又需要鉴别人员具有较高的经验。不同种类的药材,其主要成分不相同,在红外光谱中产生不同的吸收;而同种中药材,其主要药用成分的含量也不相同。目前,红外光谱法测定中药多采用溶剂萃取法和透射法进行测定^[3-5],其缺点是制样过程中增加了一些不确定因素。本文采用漫反射近红外光谱技术结合主成分分析法对川贝、浙贝、东贝进行鉴别和分类,为保证临床用药的有效性 and 证实传统贝母分类的合理性提供了一定的简便快捷的方法。

2 实验材料和方法

2.1 材料与仪器设备

川贝(珠贝与松贝)购自四川杏林医药股份公司,浙贝与东北平贝购自成都荷花池中药材市场,所有样品均经成都中医药大学鉴定。Nexus670 傅里叶变换红外光谱仪(美国 Thermo Electron 公司),配备 Drift-diffuse 漫反射附件、InGaSn 2.6 μm 检测器、CaF₂ 分束器;恒温干燥箱(上海科析实验设备制造厂)。使用软件为 omnic6.2 和 matlab6.5。

^① 联系人,手机: (0)13551351195; E-mail: dengbo1975@tom.com

作者简介: 邓波(1975—),男,四川省自贡市人,硕士,电子科技大学生物医学工程专业在读博士,主要从事分子光谱学与生物物理方面的研究。

收稿日期: 2006-04-18; 接受日期: 2006-05-15

2.2 实验条件

光谱分辨率为 4cm^{-1} , 测量范围 $10000\text{--}4000\text{cm}^{-1}$, 扫描信号累加 64 次, 扫描速度为 0.6329cm/s , 增益为自动增益。近红外光谱测量采用漫反射法。各类贝母置于恒温干燥箱内 80°C 干燥 1h 后, 研磨成细粉过 200 目筛, 每类取 12 个样品, 其中 8 个组成模型集, 8 个组成检验集。将贝母粉末置于样品槽内, 刮平压紧, 采集光谱。为减小实验误差, 每个样品测定 3 次。将各次测定得到的光谱曲线加和取平均得到各样品的数据曲线, 并用 omnic 软件进行归一化处理。

2.3 数据处理方法

各类贝母的中、近红外光谱如图 1、图 2 所示, 从图中可看出, 各类贝母的中近红外谱图较为相似, 难于相互区分。比较 4 类贝母的近红外谱时发现各类贝母的谱图有细微差异, 并具有一定的特征性, 虽然难于用肉眼进行鉴别, 但这一差异为贝母的主成分分析鉴别和聚类奠定了一定的数学基础。主成分分析法^[6]是一种有效的数据降维分析法, 它能将原始变量线性组合成一组新的变量, 即一组主成分, 仅用部分主成分就可表达原有变量的主要信息。利用这一特性, 可实现红外光谱数据的线性降维投影显示, 从而使人们能直观地从二维或三维图中观察到光谱的主要特性和聚类情况。本文采用主成分分析法, 选取不同波数范围的中近红外光谱数据进行处理, 从而对各类贝母加以鉴别和聚类。

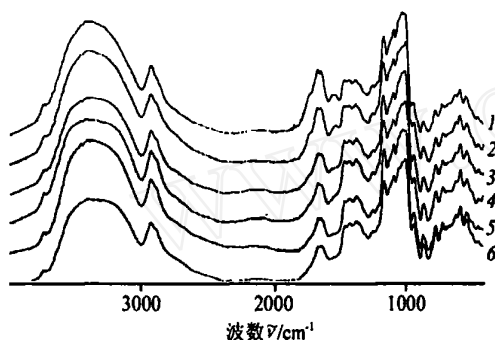


图 1 贝母的中红外光谱

1——东贝; 2——浙贝; 3——川松贝 1 级;
4——川松贝 2 级; 5——川珠贝 1 级; 6——川珠贝 2 级。

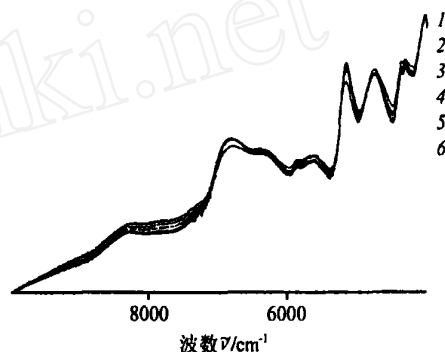


图 2 贝母的近红外光谱

1——东贝; 2——浙贝; 3——川松贝 1 级;
4——川松贝 2 级; 5——川珠贝 1 级; 6——川珠贝 2 级。

3 结果及讨论

3.1 主成分分析结果

首先使用 matlab6.5 软件对各谱段范围的数据集进行主成分分析, 在全谱 $10000\text{--}4000\text{cm}^{-1}$ 的范围内, 并非所有的谱图数据都含有有效的信息, 因此确定对于产地分类最重要的特征谱段就显得很重要。文献[7]中列出特征谱段的选择方法, 指出中药光谱的某些谱段具有更好的鉴别与分类效果。本文中特征谱段的选择根据光谱中各主要吸收峰的位置, 前 4 个主成分贡献率数据见表 1。一般, 选取的主成分总的方差贡献率大于 85% 即可。对表 1 个谱段范围的数据, 由于前 4 个主成分的贡献率已超过 85%, 故仅用前 4 个主成分完全可表示原近红外光谱的主要信息。

表 1 主成分贡献率

波数 $\nu(\text{cm}^{-1})$	10000—4000	4485—4188	4956—4485	5384—4956	5970—5384	7544—5970	8876—7544
PC1	53.560	97.758	85.556	78.343	97.559	82.248	97.248
PC2	23.405	1.543	11.323	14.998	1.279	11.944	2.041
PC3	16.019	0.505	2.8465	5.959	0.506	4.594	0.355
PC4	4.704	0.066	0.148	0.334	0.360	0.720	0.095

对各谱段的数据做主成分分析, 并作二、三维投影图。结果表明主成分分析在各个谱段对 4 类

贝母的鉴别与分类能力明显不一致。图 3a 是全谱数据的二维投影, 由图根据样本分布能够识别东贝、浙贝以及川贝(松贝和珠贝), 各类贝母样本均各自聚集分布在一定区间内, 因此利用主成分分析能够对贝母进行鉴别和分类。从全谱的三维投影(图 3b)也能得到这样的结论。松贝与珠贝的样本之所以难于区分, 是由于两者基于性状形态分类, 实际上都属于川贝母, 其化学成分极其相似, 因而主成分投影图上二者难于区分。5384—4956 cm^{-1} (图 3c 和图 3d)与 4485—4188 cm^{-1} 谱段, 但二维投影图上, 东贝与浙贝的样影可一定程度区分各类贝母; 谱段的分类效果一般, 三维投影上各类贝母完全能够加以区分; 而 5970—5384 cm^{-1} 、7544—5970 cm^{-1} 、8876—7544 cm^{-1} 谱段分类效果很差, 以 7544—5970 cm^{-1} 为例, 如图 3e、图 3f 所示, 无论二维投影还是三维投影上各类贝母的样本都难于区分。对 4956—4485 cm^{-1} 谱段, 虽然少数东贝与浙贝的样本在二维投影图上比较接近, 但绝大多数样本在二维与三维投影上均能清楚的加以鉴别与分类, 同时通过增加样本数能够提高分类的准确性。比较图 3a、3b、3c、3d 发现, 使用全谱与 4956—4485 cm^{-1} 谱段数据对贝母的鉴别与分类效果接近, 但全谱使用的数据过多, 为数据处理带来困难, 所以 4956—4485 cm^{-1} 是主成分分析鉴别与分类贝母的最佳谱段范围。

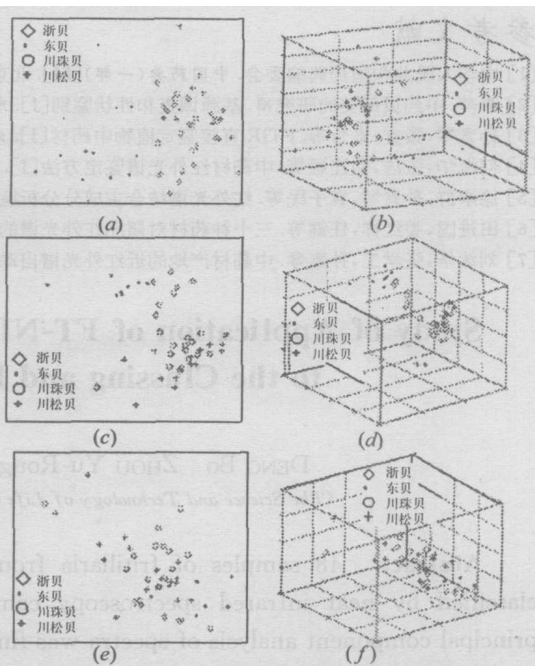


图3 贝母各谱段的二维和三维主成分投影图

3.2 系统聚类分析

本文用系统聚类法中的类平均法对样本的红外光谱进行聚类分析。该方法是以两类元素两两之间的平均平方距离为聚类依据, 即: $D_{PQ}^2 = \frac{1}{n_P n_Q} \sum_i \sum_j d_{ij}^2$ 。式中类与类之间的距离采用马氏(Mahalanobis)距离, 第 i 个和第 j 个样本之间的马氏距离为 $d_{ij}(M) = (X^{(i)} - X^{(j)})^T S^{-1} (X^{(i)} - X^{(j)})$ 。式中 S 为数据矩阵, $X^{(i)}, X^{(j)}$ 分别代表两个样品的向量。根据上述公式, 样本为各样品在 4956—4485 cm^{-1} 谱段所对应的红外光谱吸收率, 计算由 Matlab 6.5 编程实现。本文选取东贝、浙贝、川松贝与川珠贝样品共 6 个, 聚类结果如图 4 所示。由聚类图可以看出川贝、东贝与浙贝在种属关系上的远近。

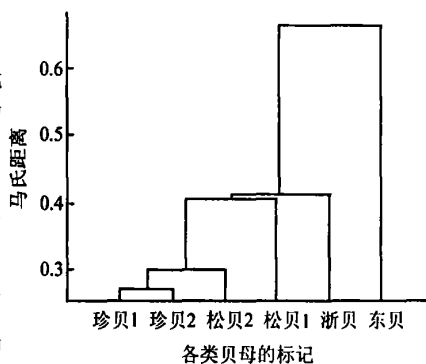


图4 贝母的聚类图

4 结论

本文川贝母、珠贝和小东贝同为百合科属植物, 但不同种, 其近红外光谱极其相似, 但可获得分类特征, 用采用主成分分析和系统聚类法对贝母样本光谱数据进行处理, 可以对其进行鉴别与分类, 不仅可以用于鉴定伪品, 准确区分不同种类(产地)的贝母, 而且与中医临床用药习惯相符合。而且本方法直接、简便、准确。在目前市场正品川贝母货源紧缺, 各种代用品使用不规范的情况下, 作为一种药材质量检测手段, 该方法具有一定的应用价值。

参考文献

- [1] 中华人民共和国药典编委会 中国药典(一部)[M]. 北京: 化学工业出版社, 2000
- [2] 李萍 中药贝母类的研究Ⅻ 基源调查和性状鉴别[J]. 中国药科大学学报, 1993, 24(6): 360
- [3] 孙素琴, 张宣, 秦竹等 FTIR 直接鉴定植物中药材[J]. 光谱学与光谱分析, 1999, 19(4): 542—544
- [4] 何淑华, 张洁, 曲连颖等 中药材红外光谱鉴定方法[J]. 吉林大学自然科学学报, 1999, 4: 103—106
- [5] 徐永群, 孙素琴, 袁子民等 红外光谱结合主成分分析鉴别道地山药[J]. 分析化学, 2002, 30(10): 1231—1233
- [6] 田进国, 娄红祥, 任键等 三十种药材对照品红外光谱的研究[J]. 中国药科大学学报, 1996, 27(1): 24—28
- [7] 刘沛华, 张学工, 孙素琴 中药材产地的近红外光谱自动鉴别和特征谱段选择[J]. 科学通报, 2005, 50(4): 393—398

Study of Application of FT-NIR and Principal Component Analysis to the Classing and Distinguishing of Fritillaria

DENG Bo ZHOU Yu-Rong LU Zhi-Hong PANG Xiao-Feng

(The Science and Technology of Life College of UESTC, Chengdu 610054, P. R. China)

Abstract 48 samples of fritillaria from Sichuan, Zhejiang and Liaoning were collected and classified by near infrared spectroscopy combined with principal component analysis. Then the principal component analysis of spectra was finished within different area and it is found that the class of 48 samples could be distinguished according to difference of the 2 or 3 dimensional projective figures. Finally, the relation of fritillaria can be classified by system class. The method is valuable for detecting the quality of fritillaria and distinguishing the different classification of the fritillaria.

Key words Principal Component Analysis, FT-NIR, Fritillaria, System Class

廉价光谱分析资料讯息

1. 《光电光谱分析》, 周开亿主编, 4 册, 16 开, 1236 页, 《光谱实验室》编辑部增刊, 每套 55 元。
2. 《光谱实验室》2004 年 1—6 期, 16 开, 1248 页, 每套 75 元。
3. 《数理统计在化学、光谱分析中的应用》, 纳利莫夫著, 余生等译, 《光谱实验室》编辑部编, 16 开, 396 页, 每册 10 元。
4. 《邮票上的光谱学和化学史》, 周开亿等编, 《科学出版社》出版, 16 开, 158 页, 每册 10 元。
5. 《空心阴极放电及其应用》, 周开亿主编, 《真空科学与技术》杂志社出版, 32 开, 798 页, 上下册, 每套 10 元。
6. 《邮票上的化学、光学和光谱学史》, 周开亿主编, 《光谱实验室》2006 年第 1 期彩色抽印本(收藏本), 16 开, 64 页, 每册 30 元。

以上各书售价中均含普通印刷品邮费。

联系地址: 北京市 81 信箱 66 分箱《光谱实验室》编辑部联络处 刘建林,
邮编: 100095