

# 基于近红外光谱快速鉴别木材种类的研究

崔宏辉, 房桂干\*, 梁龙, 吴珽, 刘雯雯

(中国林业科学研究院林产化学工业研究所, 江苏省生物质能源与材料重点实验室, 国家林业局林产化学工程重点开放性实验室, 生物质化学利用国家工程实验室, 江苏南京210042)

**摘要:**通过采集5种木材近红外光谱特征曲线, 结合化学计量学技术建立预测模型, 对不同种木材进行快速无损检测。设计正交试验 $L_{24}(2 \times 4 \times 3 \times 8)$ 对光谱特征曲线进行优化预处理。利用TQ8.6软件对正交试验结果进行判别分析。结果显示, 选择多元散射校正(MSC)、Savitzky-Golay平滑对波段 $10\ 000 \sim 4\ 000\ \text{cm}^{-1}$ 进行光谱预处理, 提取主成分为8时, 光谱判别正确率为100%。将22种已知种类木材光谱信息导入预测模型, 进一步评估预测模型的性能, 验证通过率为100%, 获得了较理想的预测结果。该方法的建立为木材种类鉴别提供了一个新方法, 同时为其他木种植物的鉴别提供参考。

**关键词:**木材; 近红外漫反射光谱; 主成分分析; 判别分析

中图分类号: O675.3; TS721

文献标志码: A

文章编号: 0253-4320(2015)02-0169-03

## A new method based on near infrared spectra for rapid identification of timber species

CUI Hong-hui<sup>1</sup>, FANG Gui-gan\*, LIANG Long, WU Ting, LIU Wen-wen

(1. Institute of Chemical Industry of Forestry Products, CAF, Key Lab. of Biomass Energy and Material, Jiangsu Province, National Engineering Lab. for Biomass Chemical Utilization; Key and Open Lab. on Forest Chemical Engineering, SFA, Nanjing 210042, China)

**Abstract:** Near infrared diffuse reflection spectra of 5 timber species combined with chemometric statistical analysis are used to build a prediction model, in order to identify different species of timber quickly and non-destructively. Orthogonal experiment  $L_{24}(2 \times 4 \times 3 \times 8)$  is designed to optimize the spectrum characteristic curve. The result of orthogonal experiment is analyzed by the discriminant analysis using TQ8.6. It reveals that when multiple scattering correction (MSC) and the Savitzky-Golay smooth are used to treat  $10\ 000 - 400\ \text{cm}^{-1}$  band, by extracting 8 kinds of principal extracting components, the spectral discriminant accuracy is 100%. The performance of prediction model is evaluated by inputting 22 kinds of known species of wood to prediction model. The test pass rate also reaches 100% and the results are satisfied. This method provides a new method for the identification of timber species, and a reference for the identification of other wood species of plants.

**Key words:** wood; near infrared diffuse reflection spectroscopy; principal component analysis; discrimination analysis

我国是个树木资源丰富、品种繁多的国家, 又是个森林覆盖率较低的贫林国。合理利用木材需要了解和认识木材性质<sup>[1]</sup>。木材原料由于化学结构和成分含量的差别而具有不同的制浆造纸性能。传统的木材识别是根据木材的化学分析、颜色、气味等作为识别依据以及通过制作显微切片来观察木材结构特征, 并一一对照木材检索表来确定木材种类。但是该识别方法复杂、费时且成本高, 不能满足林木培育过程及制浆过程的快速分类及评价要求。

现代近红外光谱技术利用全谱段或者多波长下的光谱数据进行定量或定性分析。近红外光谱技术作为一种先进的检测技术, 具有测试方便、速度快、成本低、无损识别等优点, 已在众多领域得到了广泛的应用<sup>[2-6]</sup>, 在木材识别方面也取得了一些成效<sup>[7-10]</sup>。

在近红外光谱的判别分析中, 主成分分析是较常用的分析方法, 他可以结合基于误差反向传播算法(back propagation, BP)多层前馈神经网络、偏最小二乘法-判别分析(PLS-DA)等化学计量学方法, 都是在品种识别上较成熟的处理方法<sup>[10-12]</sup>, 但是这些方法往往忽略对原始光谱的预处理选择。笔者通过设计正交试验对比不同的预处理方法对判别分析模型的影响, 最终择优建立木材识别预测模型。

## 1 材料与与方法

### 1.1 仪器设备

Antaris II型近红外光谱仪(美国, Thermo Fisher Scientific), 配置漫反射模块, 使用Result 2.1软件采集光谱图, 设定其采样光谱范围为 $10\ 000 \sim 4\ 000\ \text{cm}^{-1}$ ,

收稿日期: 2014-08-21

基金项目: 国家林业局948项目(2014-4-31)

作者简介: 崔宏辉(1990-), 女, 在读研究生, 主要从事近红外检测木材材性研究工作, cui-honghui@foxmail.com; 房桂干(1966-), 男, 博士, 研究员, 国际木材科学院院士, 主要从事制浆造纸清洁生产、环保和生物质利用研究, 通讯联系人, fangguigan@icifp.cn。

分辨率为  $8\text{ cm}^{-1}$ , 扫描次数为 64 次, 光谱数据格式为  $\text{Log}(1/R)$ , 光谱图经 TQ8.6 软件进行主成分分析; 微型植物粉碎机(天津, 泰斯特仪器有限公司生产); 40、60 目不锈钢筛盘(浙江, 道墟路达仪器厂生产); SPSS19.0 分析软件。

### 1.2 样品来源及光谱获取

实验材料分别是尾巨桉、蓝桉、马占相思、厚荚相思、竹柳 5 种木材, 共计 78 个样本, 如表 1 所示。这 5 种木材均为实验室常用制浆原料。所有样品均削片风干后, 置入微型植物粉碎机中磨成细末, 过筛, 截取能通过 0.38 mm 筛孔(40 目)而不能通过 0.25 mm 筛孔(60 目)的细末。全部样本随机分成建模集和预测集, 建模集有 56 个样本, 预测集有 22 个样本。木粉混合均匀后置于采样杯中, 使用 Result 2.1 软件编写流程采集光谱, 样品均采集 3 次, 取其平均光谱。

表 1 不同种类的木材样本

序号	种名	来源	建模集/个	预测集/个
M1	尾巨桉	雷州林业局	12	4
M3	蓝桉	四川	7	2
M4	马占相思	广州	12	6
M5	厚荚相思	广西	15	6
M7	竹柳	江苏	10	4

### 1.3 光谱数据预处理

为消除光程不一致、光谱偏移或飘移、噪声信号等影响, 需要对光谱进行预处理。根据样本光谱信息及其之间的差异, 将近红外光谱波段分为 4 个不同范围:  $4\ 000 \sim 10\ 000$ 、 $4\ 000 \sim 8\ 500$ 、 $4\ 000 \sim 7\ 500$ 、 $4\ 000 \sim 6\ 000\text{ cm}^{-1}$ 。通过  $L_{24}(2 \times 4 \times 3 \times 8)$  正交试验设计, 对光程类型、光谱波段、导数和平滑 4 个因素进行优化处理, 如表 2 所示。

表 2 光谱预处理的正交试验设计

序号	光谱类型选择	光谱波段选择	导数处理	平滑处理
1	SNV	6000 ~ 4000	无处理	无平滑
2	SNV	8500 ~ 4000	一阶	无平滑
3	MSC	7500 ~ 4000	二阶	无平滑
4	MSC	10000 ~ 4000	无处理	无平滑
5	MSC	10000 ~ 4000	二阶	3 点平滑与波长段为 4
6	MSC	7500 ~ 4000	无处理	3 点平滑与波长段为 4
7	SNV	8500 ~ 4000	无处理	3 点平滑与波长段为 4
8	SNV	6000 ~ 4000	一阶	3 点平滑与波长段为 4
9	SNV	8500 ~ 4000	无处理	5 点平滑与波长段为 5
10	SNV	6000 ~ 4000	二阶	5 点平滑与波长段为 5

11	MSC	7500 ~ 4000	无处理	5 点平滑与波长段为 5
12	MSC	10000 ~ 4000	一阶	5 点平滑与波长段为 5
13	MSC	10000 ~ 4000	无处理	7 点平滑与波长段为 6
14	MSC	7500 ~ 4000	一阶	7 点平滑与波长段为 6
15	SNV	8500 ~ 4000	二阶	7 点平滑与波长段为 6
16	SNV	6000 ~ 4000	无处理	7 点平滑与波长段为 6
17	MSC	6000 ~ 4000	一阶	段长为 3 与段间距为 5
18	SNV	10000 ~ 4000	二阶	段长为 3 与段间距为 5
19	SNV	7500 ~ 4000	二阶	段长为 5 与段间距为 7
20	MSC	8500 ~ 4000	一阶	段长为 5 与段间距为 7
21	MSC	8500 ~ 4000	二阶	段长为 7 与段间距为 9
22	SNV	7500 ~ 4000	一阶	段长为 7 与段间距为 9
23	SNV	10000 ~ 4000	一阶	段长为 9 与段间距为 11
24	MSC	6000 ~ 4000	二阶	段长为 9 与段间距为 11

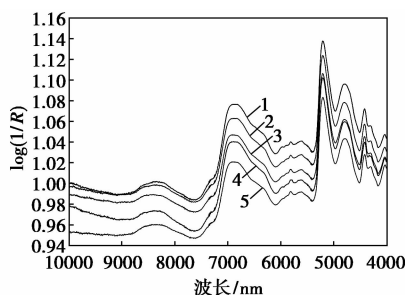
注: MSC: 多元信号修正(multiplicative scatter correction); SNV: 标准正态变换(standard normal variate); 5 ~ 16 号为 Savitzky-Golay filter; 17 ~ 24 号为 Norris filter。

化学计量方法中的主成分分析(Principal Component Analysis, PCA)是统计学中经典的数据降维算法。由于木材原始光谱是高维的数据集, 如果直接利用原始光谱建立判别分析模型, 光谱中具有的多重共线性、信息重叠等不利因素的影响会造成所建模型出现过拟合现象, 降低模型分析精度。TQ8.6 软件提供的判别分析采用主成分分析与马氏距离相结合的方法。经主成分分析之后, 可以用数量较少的主成分解释原有的绝大部分信息, 大大提高了建模速度。

## 2 结果与分析

### 2.1 木材样本的近红外漫反射光谱

5 种木材典型的近红外漫反射光谱曲线如图 1 所示。从图 1 可知, 由于木材不同的化学成分含量和结构, 不同种木材的光谱曲线有明显差别, 这一差异为木材不同种的鉴别奠定了基础。



1—厚荚相思; 2—尾巨桉; 3—马占相思; 4—蓝桉; 5—竹柳

图 1 5 种木材的近红外漫反射光谱

### 2.2 主成分分析对不同种木材进行定性分析

采用 TQ8.6 分析软件对正交试验结果进行判别分析,结果如表 3 所示。由表 3 可知,24 种优化处理方法中较好的为 4、13、21、22、24,判别正确率都为 100%。对这 5 组较为满意的处理进行可视化分析,结合优化处理的光谱波段选择、主成分数及光谱处理方式综合考虑,选取 13 号为建模最优预处理方法。当主成分数选为 8 时,累积贡献率达 97.6%,且增加趋势趋于平缓,如图 2 所示。13 号

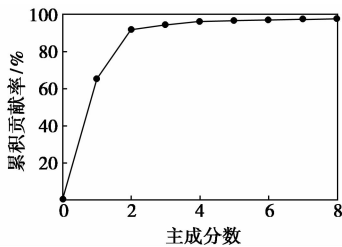


图 2 主成分贡献率

表 3 主成分分析处理结果

处理次数序号	主成分数	主成分贡献率/%	错误个数	处理次数序号	主成分数	主成分贡献率/%	错误个数
1	8	99.6	0	13	8	97.6	0
2	8	25.6	28	14	8	35.4	14
3	9	35.8	17	15	8	25.2	43
4	9	96.7	0	16	8	99.6	0
5	8	27.3	37	17	10	94.2	0
6	8	99.4	0	18	10	43.4	39
7	10	99.2	0	19	8	53.5	0
8	8	31.2	14	20	8	91.7	0
9	8	98.9	0	21	10	91.6	0
10	8	27.5	39	22	10	90.8	0
11	8	99.4	0	23	10	78.5	0
12	8	29.2	27	24	10	98.9	0

优化方法(见表 2)在三维空间中能将 56 种木材样本较好的分为 5 类,分类结果如图 3 所示。从图 3 可以看出,5 种木材在主成分空间中呈现较好的聚合度,5 个种间的分界清楚分明。同时可以看到,马占相思(M4)、竹柳(M7)与其余种分界线较为明显,这与种的亲缘关系有关。

### 2.3 建立不同种木材快速鉴别模型

将 2.2 中光谱数据的优化结果 13 号作为建立模型的输入值,建立了木材快速预测模型。将 22 种已知种类木材光谱信息一起导入预测模型,进一步评估预测模型的性能,并设定距离 < 3 为通过验证,最终验证输出结果如表 4 所示,由表 4 可知,木材种类验证通过率为 100%,获得了较理想的预测结果。

表 4 木材定性分析模型的验证结果

序号	真实	预测	通过	马氏距离	序号	真实	预测	通过	马氏距离
M1-1	M1	M1	Pass	0.80	M4-6	M4	M4	Pass	1.53
M1-2	M1	M1	Pass	1.07	M5-1	M5	M5	Pass	0.81
M1-3	M1	M1	Pass	1.06	M5-2	M5	M5	Pass	0.47
M1-4	M1	M1	Pass	0.93	M5-3	M5	M5	Pass	0.69
M3-1	M3	M3	Pass	2.57	M5-4	M5	M5	Pass	1.01
M3-2	M3	M3	Pass	2.53	M5-5	M5	M5	Pass	0.92
M4-1	M4	M4	Pass	0.85	M5-6	M5	M5	Pass	0.72
M4-2	M4	M4	Pass	0.68	M7-1	M7	M7	Pass	0.93
M4-3	M4	M4	Pass	0.83	M7-2	M7	M7	Pass	1.20
M4-4	M4	M4	Pass	0.79	M7-3	M7	M7	Pass	0.70
M4-5	M4	M4	Pass	1.98	M7-4	M7	M7	Pass	1.03

### 3 结论

通过实验获得了 5 种木材的近红外光谱特征曲线,为准确鉴别不同种木材,设计正交试验对所采集光谱进行优化处理,并结合主成分分析建立了木材品种鉴别模型。结果显示,选择多元散射校正(MSC)、Savitzky-Golay 平滑对波段 10 000 ~ 4 000  $\text{cm}^{-1}$  进行光谱预处理,提取主成分为 8,对木材样品进行判别分析,其中主成分数的累积贡献率达 97.6%,建模集及验证集识别率均可达到 100%。说明运用近红外光谱技术结合现代化学计量学,可以快速、准确、无损地对木材品种进行鉴别,是一种很有前景的分析方法,对造纸用原料的定向培育选材具有很大的促进作用。同时,下一步可以研究同一物种不同产

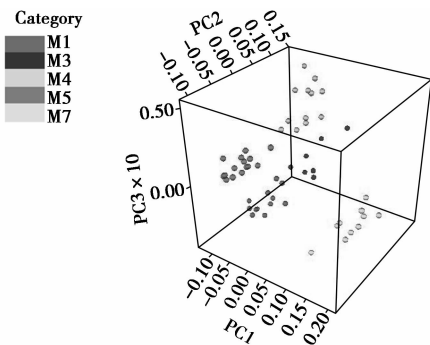


图 3 5 种木材分类的 3 维效果图

或2个以上的苯环以线性排列、弯接或簇聚的方式构成的一类有机物<sup>[4]</sup>,如萘、蒽、菲和芘等。环烷芳烃为多环芳烃的芳烃部分饱和化合物,如四氢萘、八氢菲等。

## 2 全二维气相色谱的优势

全二维气相色谱因其峰容量大、灵敏度好、定性准确等特点,在食品、环境、农残、香烟、土壤及沉积物、甾醇和甾族化合物等多个领域有着极广的应用<sup>[5]</sup>。GC×GC第2维柱的柱长比第1维短,固定相厚度更薄,导致其分离速度比第1柱快,避免了前后流出组分重叠。与一维气相色谱相比,全二维气相色谱具有诸多优势<sup>[6]</sup>:峰容量大,其峰容量为组成他的2根柱各自峰容量的乘积;灵敏度高,较通常的一维色谱的灵敏度提高20~50倍;定性定量的准确性高。

GC×GC的突出优点是根据化合物所属族类型的不同进行族分离。不同族类化合物,在GC×GC二维谱图上分布在不同的区带内,每一区带内化合物按照沸点进行分离,所有的族均呈现“瓦片效应”,即不同碳数的同族物平行排列,相同碳数的异构体呈线性排列,为定性分析提供依据<sup>[7]</sup>。Vendevre等<sup>[8]</sup>研究了GC×GC的正交分离、族分离等特性。结果表明,与传统一维气相色谱相比,GC×GC是分离复杂混合物及分析复杂基体中目标化合物的

有效手段。王汇彤等<sup>[9-10]</sup>相继建立了原油中饱和烃和芳烃组分的GC×GC-TOFMS分析方法。利用所建立的方法对饱和烃组分中三环萜烷类、甾烷类和藿烷类等典型生物标记物进行了定性分析,检测到普通GCMS分析中检测不到的C<sub>31</sub>~C<sub>35</sub>三环萜烷;对原油芳烃组分中二环到五环芳烃系列、氧芴、芴系列、单芳甾和三芳甾烃系列进行了分区域描述,将一些低含量的杂原子化合物与干扰的芳烃组分较好分开,验证了GC×GC-TOFMS在复杂混合物体系和痕量物质分析上的巨大优势。李水福等<sup>[11]</sup>建立了生物降解原油饱和烃与芳烃组分的分析方法,得到了萜烷、甾烷和芳烃组分的二维点阵图和三维立体图。

全二维气相色谱在含硫、氮化合物分析上也具有广泛应用。Adam等<sup>[12]</sup>利用GC×GC-NCD建立了分离和鉴定石油中间馏分中含氮化合物的分析方法,结果表明,当第2根极性柱子上有自由电子对时能极大地提高含氮化合物的分离能力。Wang等<sup>[13]</sup>研究表明GC×GC-NCD能很好检测柴油中的含氮化合物,可以用来监控分析油品加氢脱氮过程。

## 3 全二维气相色谱在重馏分油中的应用

尽管近年来GC×GC已应用到重馏分油分析中,但相对于中间馏分油而言,重馏分油的分析应用还较少。由于二维极性柱固定相对温度的不稳定

(上接第171页)

地的原料分类并开发相应生产线上的在线检测设备,提高木材在线检测精度。

### 参考文献

- [1] 杨淑慧. 植物纤维化学[M]. 北京:中国轻工业出版社,2011:4-6.
- [2] 丁长春,方向京,赵艳丽,等. 近红外漫反射光谱法快速鉴别石斛属植物[J]. 光谱学与光谱分析,2014,34(1):82-86.
- [3] Foca G, Salvo D, Cino A, et al. Classification of pig fat samples from different subcutaneous layers by means of fast and non-destructive analytical techniques [J]. Food Research International, 2013, 52(1):185-197.
- [4] 李晓丽,何勇,裴正军. 一种基于可见-近红外光谱快速鉴别茶叶品种的新方法[J]. 光谱学与光谱分析,2007,27(2):279-282.
- [5] 谈爱玲,毕卫红. 基于KPCA和LSSVM的蜂蜜近红外光谱鉴别

分析[J]. 激光与红外,2011,41(12):1331-1336.

- [6] 温珍才,孙通,耿响,等. 可见/近红外联合 UVE-PLS-LDA 鉴别压榨和浸出山茶油[J]. 光谱学与光谱分析,2013,33(09):2354-2358.
- [7] 杨金勇,李学春,黄安民,等. 基于主成分分析与 Fisher 判别的 NIR 木材识别[J]. 东北林业大学学报,2013,41(12):032.
- [8] 张蓉. 基于光谱分析的木材树种识别方法的研究[D]. 南京:南京林业大学,2011.
- [9] Schimleck L R, Michell A J, Vinden P. Eucalypt wood classification by NIR spectroscopy and principal components analysis[J]. Appita Journal, 1996, 49(5):319-324.
- [10] 姜笑梅,殷亚方,刘波. 木材树种识别技术现状、发展与展望[J]. 木材工业,2010,24(4):36-39.
- [11] 马明宇,王桂芸,黄安民,等. 人工神经网络结合近红外光谱用于木材树种识别[J]. 光谱学与光谱分析,2012,32(09):2377-2381.
- [12] 杨忠,吕斌,黄安民,等. 近红外光谱技术快速识别针叶材和阔叶材的研究[J]. 光谱学与光谱分析,2012,32(7):1785-1789. ■