

SIMCA 法判别分析木材生物腐朽的研究

杨 忠¹, 江泽慧^{1,2*}, 费本华¹, 覃道春²

1. 中国林业科学研究院木材工业研究所, 北京 100091

2. 国际竹藤网络中心, 北京 100102

摘要 木材是一种生物质材料, 容易受到各种微生物的危害, 生物腐朽可以迅速导致木材结构的破坏, 因此, 对木材生物腐朽的快速、准确地检测或鉴定具有重要意义。近几年来, 近红外光谱和 SIMCA 方法正被用于识别或检测食品、药品和农产品等研究中, 因此, 本研究尝试利用近红外光谱结合 SIMCA 方法来检测木材的生物腐朽。研究表明, 应用近红外光谱和 SIMCA 方法能有效地判别木材的生物腐朽类型, 通过培训集样本建立的基于 PCA 分析的 SIMCA 判别模型对未腐朽、白腐和褐腐三种类型样本进行回判, 判别准确率分别为 100%, 82.5% 和 100%; 而对未知腐朽类型的样本(包括未腐朽、白腐和褐腐样本), 判别准确率分别为 100%, 85% 和 100%; SIMCA 方法对未腐朽和褐腐类型的判别准确率均达到 100%, 但对白腐样本都有错判, 造成这种错判的主要原因可能是由于样本包括的信息不够丰富以及腐朽初期白腐和褐腐试样的性质差异太小等。

关键词 近红外光谱; SIMCA; PCA; 木材; 生物腐朽; 判别

中图分类号: O657.3; S782.3 **文献标识码**: A **文章编号**: 1000-0593(2007)04-0686-05

引言

木材是一种生物性特征明显的植物材料, 因而容易受到各种微生物的侵袭。真菌腐朽是对木材破坏最严重的一种方式, 因为真菌腐朽可以迅速导致木材结构的破坏, 即使在真菌腐朽的初期, 木材的力学强度也会发生急剧下降^[1,2]。在美国, 由于真菌腐朽、败坏木材造成的损失估计每年超过 52 亿美元^[3]。因此, 对木材生物腐朽的快速、准确地检测或鉴定可以为木材的及时保护和合理利用提供参考, 具有重要的现实意义。

检测生物腐朽的方法和技术有很多种, 如表面观察、显微镜观察、力学强度试验、电导率测定和声波探测(声波发射和应力波计时器、超声波脉动测试)等。但是, 这些方法一般很难快速、准确地检测和判别木材的生物腐朽。近十几年, 有关人员已开始将一些先进的无损检测与现代仪器分析技术引入到木材生物腐朽的检测与评估研究中^[4,5], 但所有的方法均有各自的局限性, 没有一种方法可以同时对多种条件下的木材腐朽进行检测。

近年来, 近红外光谱(NIR)和 SIMCA(soft independent modeling of class analog)方法在食品、医药和农产品等领域

中的检测或识别方面得到了许多应用^[6-9], 而生物腐朽木材与健康材相比, 不仅受木腐菌侵染, 木材内部有菌丝生长, 木材颜色参数和化学成分也有所变化, 而这些变化基本上都可以在近红外光谱中得到反映^[10]。因为, NIR 包含了键强度、化学组成与试样的散射、表面光泽、折光指数等重要信息, 这为利用 NIR 检测与评估木材的腐朽提供重要的依据。因此, 本研究尝试利用近红外光谱结合 SIMCA 方法来检测木材的生物腐朽。

1 材料与方法

1.1 生物腐朽试样的准备

木材的生物腐朽试验采用试菌为一种褐腐菌和一种白腐菌, 接种后放入培养室内进行生物腐朽, 经过不同时间的腐朽后, 取出试样干燥恒重, 按照生物腐朽前后试样的质量变化计算失重率(失重率数据的分析将在以后的文章中分析)。从腐朽试样中随机选取 140 个木材腐朽试样, 其中未腐朽试样 20 个, 白腐 60 个、褐腐 60 个。

1.2 近红外光谱的采集

试验用设备是由美国 ASD 公司(分析光谱设备公司)生产的 Field Spec 近红外光谱仪。光谱仪的波长范围在 350 ~

收稿日期: 2006-04-21, 修订日期: 2006-08-06

基金项目: 国家“948”项目(2006-45, 2003-4-27)和中国林业科学院重点预研课题专项补助基金(2005-M-01)资助

作者简介: 杨 忠, 1976 年生, 中国林业科学院木材工业研究所助理研究员

e-mail: zyang@caf.ac.cn

*通讯联系人

2 500 nm, 采用光纤探头采集试样表面的近红外光谱, 光纤探头在试样表面的垂直上方。光谱的空白校准采用商用聚四氟乙烯制成的白色材料。照射光源用 DC 直流灯, 在样品的上方与样品成 30° 角度, 并与样品的纵轴保持平行。将腐朽木材试样磨成木粉后装入直径约 2 cm 的样品杯, 样品杯置于可旋转的转盘上。样品经过光纤探头采集每扫描 30 次并自动平均为一条光谱后保存起来, 得到的近红外光谱经 ASD 提供的专业软件转换成光谱数据文件。经过综合分析比较, 在本试验中拟采用 500 ~ 2 500 nm 的光谱信息进行 SIMCA 分析。

1.3 SIMCA 分析方法

化学计量学中定性分析的方法是判别分析法, 判别分析的基本思想是根据已知样品集特征, 选定适合的判别准则, 建立定性分析模型, 最后用于预测判定未知样品。近红外光谱判别分析法主要有命中指数法、PCA-马氏距离法、SIMCA、PLS 判别分析 (PLS, discriminant analysis) 等, 其中 SIMCA 和 PLS 判别分析是光谱分析软件 Unscrambler 中主要的分析方法。

SIMCA 是光谱分析中分类、判别等定性分析的主要方法, 是一种二值判定方法。SIMCA 判别方法是基于每一个类别培训集的 PCA 分析建立的模型, 然后计算未知样本点与培训集的 PCA 模型的距离, 根据距离判别方法判别未知样本的类别。化学计量学中, 通常用样品间的距离, 来描述样本间的差异, 而判别分析的基本思想就是比较样本到各类总体的距离, 寻找最近距离的类, 如图 1 中如果未知样本 Samples 距 G_i 类总体的距离比 G_j 类总体的短, 则判属于 G_i 类。SIMCA 方法中使用的距离是马氏距离 (mahalanobis distance, MD)。

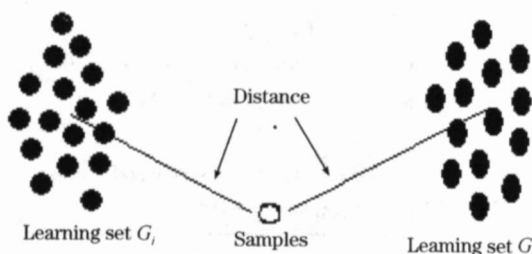


Fig 1 The principle of distance-discrimination analysis

2 结果与讨论

2.1 SIMCA 判别模型的建立

近红外光谱 SIMCA 定性分析, 主要分为两步: (1) 是对培训集样本的光谱数据进行 PCA 分析, 为每一个类建立一个 PCA 模型; (2) 计算未知样本距各培训集 PCA 模型的距离, 找出最小距离的类。SIMCA 判别结果有三类, 分别是未知样本属于已知的某一类、未知样本同时属于已知的多个类, 以及未知样本不属于任何已知类。当出现未知样本同属于多个类时, 说明判别分析模型不够精确, 需要改进, 一般可以通过增加培训集样本数或是增加变量数来解决。

本试验在所有试样中按不同生物腐朽类别分别随机抽取了 2/3 的试样用于培训建立 SIMCA 判别模型, 剩余 1/3 样

本作评价培训集模型有效性的检测集。本研究分别对未腐朽、白腐和褐腐三类试样的培训集样本进行 PCA 分析, 并通过交互验证建立各腐朽类别的 PCA 模型, 即 SIMCA 模型包括了未腐朽、白腐和褐腐试样的三个 PCA 模型, 结果见图 2 ~ 图 4。

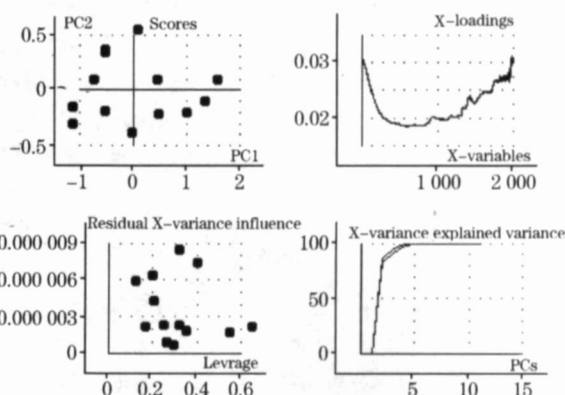


Fig 2 PCA model of the non-decay samples in SIMCA

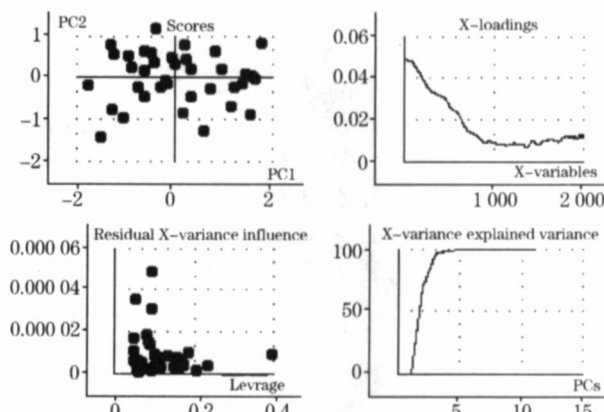


Fig 3 PCA model of the white-rot samples in SIMCA

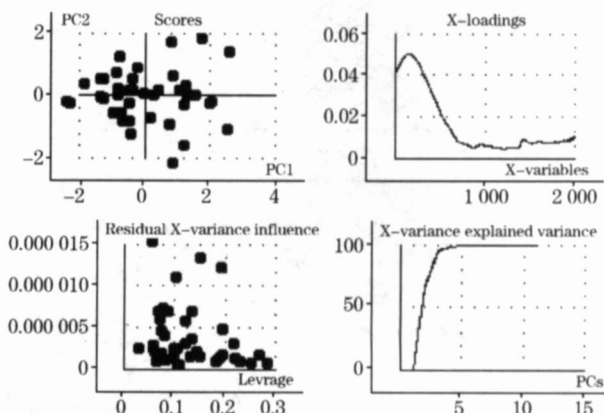


Fig 4 PCA model of the brown-rot samples in SIMCA

图 2, 图 3 和图 4 中 Scores 图提示了样本点分散和差异, 具有相同或相近性质的样本聚集在一起, 而差异较明显的样本相互远离。判别分析要求培训集样本能代表所属类的特征, 否则其构建的模型不完善, 导致模型预测和判别能力的降低; 而 Influence 图则表示是各样本点对相应 PCA 模型的

影响程度, 由样本点的杠杆值 (Leverage) 和残差 (Residual) 来决定, 高杠杆值和高残差值的样本点被认为是异常值 (Outlier)。杠杆值是样本点或变量在模型中投影点距模型中心的距离, 表示的是单个样本或变量与模型中其他样本或变量的区别, 和样本点或变量对建立的模型的影响程度, 值越大表示对模型的影响越大。残差是样本点或变量的观察值与拟合值之差, 表示了模型未能解释的样本点或变量的特征的量, 其值越小, 模型拟合越好。从图 2~图 4 可知, 各模型的样本在主成分排序图中散开, 且样本点的残差值和杠杆值都较小, 表明各模型中培训集选取的样本具有相应腐朽特性的代表性, 且各模型的拟合性较好。

2.2 SIMCA 判别模型的验证

由于 SIMCA 方法是二值判别法, 样本特征的判别结果只有两种, 即是或否, 试验中腐朽样本的类别有未腐朽、白腐和褐腐三类, 因此 Unscrambler 光谱分析软件 SIMCA 分析时分别判断所有试样的未腐朽、白腐和褐腐特性。本试验采用了培训集样本回判来验证 SIMCA 判别模型的判别效率, 图 5~图 7 是利用培训集样本建立的 SIMCA 判别模型对培训集样本进行验证的结果。图中纵轴是样本距模型的距离, 其值是样本的残差的平方根, 值越小表示对应的分类模型能更好的描述样本, 样本属于该类的可能性高; 横轴为样本杠杆值 (Sample leverage), 表示在不考虑类对样本描述程度的前提下, 比较样本与类中其他样本点的不同, 以及对模型的影响程度; 图中分别用一条垂直线和一条水平的直线给出了样本到模型的距离和杠杆值在 $\alpha = 0.05$ 显著水平下的域值, 两直线与坐标轴构成的区域即为具有相应属性的有效区域。

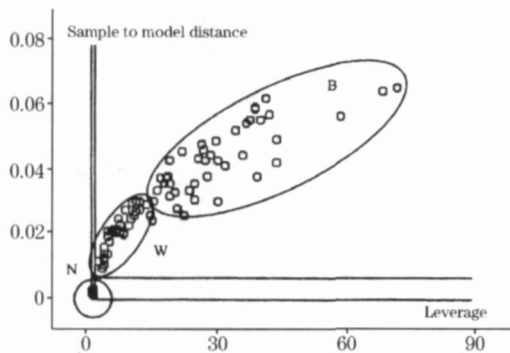


Fig 5 The discrimination results of non-decay samples of training set by SIMCA

N, W, and B stands for non decay, white-rot and brown-rot decay samples, respectively

从图 5~图 7 可知, 除了图 7 中有 7 个白腐试样坐标值落在样本到褐腐模型距离和杠杆值的域值里 ($\alpha = 0.05$), 其他所有样本均被正确判别, 落在褐腐判别域内有 7 个白腐样本被判为同属于白腐和褐腐, 即产生了一个样本同属于多个类的判别结果。表 1 列出了利用培训集样本对 SIMCA 判别模型的验证结果, 从表中可以看出, SIMCA 方法能有效的判别试样的腐朽类型, SIMCA 模型对未腐朽、白腐和褐腐三种类型的判别准确率分别为 100%, 82.5% 和 100% ($\alpha =$

0.05), 其中白腐样本虽然都被 SIMCA 模型判为白腐类型, 但同时也有被错判为褐腐类型的样本 (错判率为 17.5%), 造成这种判别结果的原因可能是由于培训集样本包括的信息不够丰富以及腐朽初期白腐和褐腐试样的性质差异太小等, 从而造成 SIMCA 判别分析中培训集 PCA 模型不够精确, 可以通过增加样本集数量或包含的信息量等方法提高 PCA 模型的精度, 从而提高判别效率。

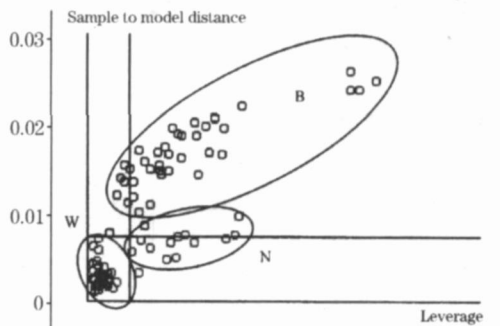


Fig 6 The discrimination results of white-decay samples of training set by SIMCA

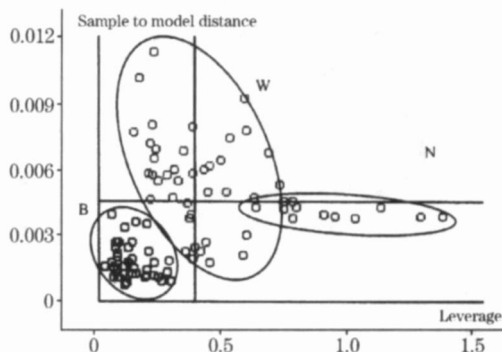


Fig 7 The discrimination results of brown-decay samples of training set by SIMCA

Table 1 The discriminant results of wood biological decay of test set by SIMCA

Training set (n = 93)	Non-decay / %	White-rot decay/ %	Brown-rot decay/ %
Discriminant probability ($\alpha = 0.05$)	100	82.5	100

2.3 SIMCA 模型对未知样本的检测

本试验中用于检测的未知样本为未参与建模的检测集样本, 检测集样本有 47 个。图 8~图 10 是利用培训集样本建立的 SIMCA 判别模型来检测未知样本的结果, 经过与实际结果的核对发现, 图 8、图 9 和图 10 分别是 SIMCA 模型对未腐朽、白腐和褐腐样本的判别结果 ($\alpha = 0.05$)。

从图 8~10 可知, 图 10 中有 3 个白腐试样坐标值落在样本到褐腐模型距离和杠杆值的域值里 (被判为同属于白腐和褐腐), 其他所有样本均被正确判别。表 2 列出了利用 SIMCA 判别模型对未知样本的检测结果, 从表中可以看出, SIMCA 模型能有效判别木材的生物腐朽类型, SIMCA 模型

对未腐朽、白腐和褐腐三种类型的判别准确率分别为 100 % , 85 % 和 100 % ($\alpha = 0.05$) , 其中白腐样本虽然都被 SIMCA 模型判为白腐类型, 但同时也有被错判为褐腐类型的样本, 错判率为 15 % 。即使我们将原 SIMCA 模型中的错判样本剔除重新建立判别模型, 也没能提高对未知样本的判别准确率, 因此, 造成这种错判的原因可能是由于样本包括的信息不够丰富以及腐朽初期白腐和褐腐试样的性质差异太小等。

3 结 论

应用近红外光谱和 SIMCA 方法能有效的判别木材的生

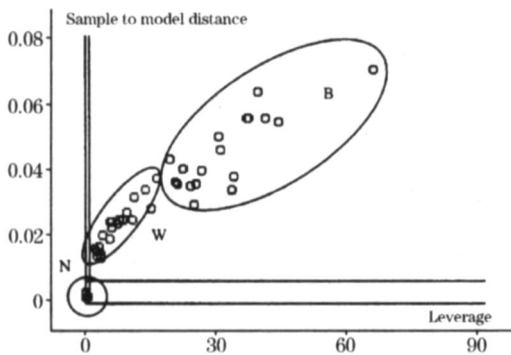


Fig 8 The discrimination results of non-decay samples in tset set by SIMCA

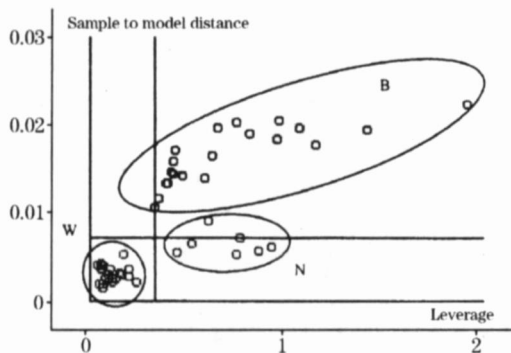


Fig 9 The discrimination results of white-decay samples in tset set by SIMCA

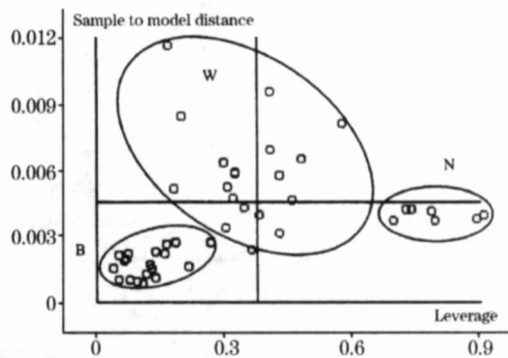


Fig 10 The discrimination results of brown-decay samples in tset set by SIMCA

Table 2 The discrimination results of wood biological decay of test set by SIMCA

Test set (n = 47)	Non-decay / %	White-rot decay/ %	Brown-rot decay/ %
Discriminant probability ($\alpha = 0.05$)	100	85	100

物腐朽类型。通过培训集样本建立的 SIMCA 判别模型对未腐朽、白腐和褐腐三种类型样本进行回判, 判别准确率分别为 100 %、82.5 % 和 100 % ; 对未知腐朽类型的样本 (包括未腐朽、白腐和褐腐样本), 判别准确率分别为 100 % , 85 % 和 100 % 。

在本研究中, SIMCA 方法对未腐朽和褐腐类型的判别准确率均达到 100 % , 没有错判, 但对白腐样本都有错判, 造成这种错判的主要原因可能是由于样本包括的信息不够丰富以及腐朽初期白腐和褐腐试样的性质差异太小等, 我们可以通过增加样本集数量或包含的信息量等方法提高 PCA 模型的精度, 从而提高判别效率。本研究的重要意义在于通过试验证明了近红外光谱技术在实验室范围内可以应用于木材生物腐朽的检测。

致谢: 本研究得到了美国农业部林务局的首席科学家 Chung-Yun Hse 博士和技术员 Dale 的帮助, 在此一并表示感谢。

参 考 文 献

- [1] Winandy J E, Morrell J J. *Wood and Fiber Science*, 1993, 25(3) : 278.
- [2] Curling S F, Clausen C A, Winandy J E. *Forest Product Journal*, 2002, 52(7/8) : 34.
- [3] Lee H L, Chen G C, Rowell R M. *Holzforchung*, 2004, 58 : 311.
- [4] Backa S, Brodin A, Nilsson T. *Holzforchung*, 2001, 55(3) : 225.
- [5] Müller U, Bammer R, Teischinger A. *Holzforchung*, 2002, 56(5) : 529.
- [6] WANG LI, ZHUO Lin, HE Ying, et al(王 丽, 卓 林, 何 鹰, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2004, 24(12) : 1537.
- [7] LI Qing-bo, YANG Li-min, LING Xiao-feng, et al(李庆波, 杨丽敏, 凌晓峰, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2004, 24(4) : 414.
- [8] YANG Zhong, JIANG Ze-hui, FEI Ben-hua, et al(杨 忠, 江泽慧, 费本华, 等). *Scientia Silvae Sinicae(林业科学)*, 2005, 41(4) : 177.
- [9] JIANG Ze-hui, FEI Ben-hua, YANG Zhong(江泽慧, 费本华, 杨 忠). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2007, 27(3) : 435.
- [10] YANG Zhong, JIANG Ze-hui, FEI Ben-hua(杨 忠, 江泽慧, 费本华). *Scientia Silvae Sinicae(林业科学)*, 2006, 42(3) : 99.

Discrimination of Wood Biological Decay by Soft Independent Modeling of Class Analogy (SIMCA) Pattern Recognition Based on Principal Component Analysis

YANG Zhong¹, JIANG Ze-hui^{1,2*}, FEI Ben-hua¹, QIN Dao-chun²

1. Research Institute of Wood Industry, Chinese Academy of Forestry, Beijing 100091, China
2. International Center for Bamboo and Rattan, Beijing 100102, China

Abstract Wood, as a biomass materials, tends to be attacked by microorganisms, and its structure could be rapidly destroyed by biological decay. Therefore, it is significant to rapidly and accurately detect or identify biological decay in wood. Recently, extensive research has demonstrated that near infrared spectroscopy (NIR) and soft independent modeling of class analogy (SIMCA) can be used to discriminate or detect a wide variety of food, medicine and agricultural products. The use of NIR coupled with principal component analysis (PCA) and SIMCA pattern recognition to detect wood biological decay was investigated in the present paper. The results showed that NIR spectroscopy coupled with SIMCA pattern recognition could be used to rapidly detect the biological decay in wood. The discrimination accuracy by the SIMCA model based on the training set for the non-decay, white-rot and brown-rot decay samples were 100%, 82.5% and 100%, respectively; and that for the samples for the test set were 100%, 85% and 100%, respectively. However, some white-rot decay samples were mis-discriminated as brown-rot decay, for which the main reasons might be that the training set does not have enough typical samples, and there is a slight difference between white-rot and brown-rot decay during the early stage of decay.

Keywords Near infrared spectroscopy (NIR); Soft independent modeling of class analogy (SIMCA); Principal component analysis (PCA); Wood; Biological decay; Discrimination

(Received Apr. 21, 2006; accepted Aug. 6, 2006)

* Corresponding author