

文章编号: 1672-8785(2012)11-0039-05

用近红外光谱法快速测定 欧美杨木质素含量

刘栋梁 刘 胜

(北京林业大学理学院, 北京 100083)

摘 要: 以欧美杨的木质素含量为研究对象, 给出了建立近红外光谱分析模型的一种多模型方法。所建模型可用于快速预测欧美杨中的木质素含量, 其预测值与实验值之间的相关系数为 0.9702, 预测结果的平均相对误差为 0.0100, 模型的拟合优度为 0.9126。将所建模型与用偏最小二乘法建立的欧美杨木质素含量预测模型进行了对比, 前者的预测效果优于后者。

关键词: 近红外光谱; 多模型; 木质素; 欧美杨

中图分类号: O657. 33 **文献标识码:** A **DOI:** 10.3969/j.issn.1672-8785.2012.11.008

Rapid Determination of Lignin Content in Populus Euramericana by Near-infrared Spectroscopy

LIU Dong-liang, LIU Sheng

(College of Science, Beijing Forestry University, Beijing 100083, China)

Abstract: A near-infrared spectral analysis model is established by using a multi-model method. The model can be used to rapidly determine the lignin content in populus euramericana. The correlation coefficient between the predicted value and the experimental value is 0.9702. The mean relative error of the prediction is 0.0100. The goodness of the fit model is 0.9126. Compared with the determination model established by using the partial least squares (PLS) method, the prediction of this model is better.

Key words: near-infrared spectroscopy; multi-model; lignin; populus

0 引言

近红外光谱是分子吸收一部分近红外线后原子的振动能级与转动能级发生跃迁而产生的分子吸收光谱。由于近红外光谱具有快速、准确、无损等特点, 它已经被广泛应用到食品、化工、医药等领域^[1-6]。木质素是植物体中含量较高的天然高分子聚合物。在制浆造纸过程中, 木质素很大程度上决定了制浆性能和纸浆质量, 因此人们在制定蒸煮和漂白工艺时都需要参考

木质素的含量。由于传统的实验室木质素测量方法操作繁琐, 耗时多, 且会破坏样品, 近年来, 在木材科学领域, 近红外光谱得到了越来越多的应用, 从而弥补了传统测试方法成本高、效率低等的不足^[7-16]。文献[17]使用迭代法对相思树的克拉森木质素含量进行了快速预测。

常用的由近红外光谱数据建立数学模型的方法有偏最小二乘回归、主成分回归、多元线性回归和支持向量机等, 其中以偏最小二乘回归方

收稿日期: 2012-09-20

作者简介: 刘栋梁(1987-), 男, 河北承德人, 硕士研究生, 主要从事应用生物数学研究工作。

E-mail: lshlxc@263.net

法的应用最为广泛。本文以欧美杨的木质素含量为研究对象,给出了建立近红外光谱分析模型的一种多模型方法:首先利用近红外光谱数据建立若干个线性子数学模型,然后采用加权平均的方法得出总的数学模型,实现对欧美杨木质素含量的预测。本文还将上述建模方法与现在常用的偏最小二乘法进行了对比,结果表明多模型方法优于偏最小二乘法。本文所采用的建模方法有望用于建立其它近红外光谱分析模型。

1 仪器与数据

1.1 实验原料

从北京顺义地区选取 4 个不同品系 5~6 年生的欧美杨试材用于实验。每株欧美杨试样从胸高 1.3 米处开始,向上每隔 2 米截取 1 个圆盘,每棵树取 4~6 个圆盘,圆盘采集按照国家标准 GB1930~41-91 进行。将圆盘磨粉后筛取 40~60 目之间的合格木粉用于实验室测量和近红外光谱采集,共制得 42 个样本。

1.2 实验数据的测定

实验时取 1 克经苯醇抽提过的样本,把它移入具有磨口玻塞的 100 ml 锥形瓶中,再加入 72% 的硫酸 15 ml,在 18 °C ~20 °C 的恒温槽中保温 2h。然后将瓶内容物移入到容量为 1000 ml 的锥形瓶中,加水稀释至 3%,煮沸 4h。在此过程中加水保持总体积不变,将处理后的溶液先过滤,用热蒸馏水洗涤至中性为止,再烘干,称重,计算克拉森木质素含量。

1.3 仪器设备与近红外光谱采集

实验采用德国 Bruker 公司生产的带有 RT-PbS 检测器、OPUS/OVP 自检功能的傅里叶近红外光谱仪,光谱仪内置镀金漫反射积分球。仪器参数设定:波数范围为 4000~12000 cm^{-1} ;分辨率为 8 cm^{-1} ;扫描次数为 64 次,取平均值构成一个光谱数据;用旋转台增加采样面积,采集木粉的漫反射光谱。

2 数学模型的建立

2.1 数据的分组

为了使校正集和验证集的划分变得较为合理,先将 42 个样本按照木质素含量实验值从小到大排序,将序号为 4+6N (N=0,1,2,3,4,5,6) 的 7 个样本挑出来构成验证集,其余的 35 个样本构成校正集。然后将校正集中的样本重新编号为 1~35,将验证集中的样本重新编号为 36~42,并设 $Y^C = (y_1, y_2, \dots, y_{35})$ 为校正集的木质素含量向量,设 $Y^V = (y_{36}, y_{37}, \dots, y_{42})$ 为验证集的木质素含量向量。其中 y_i 代表第 i 个样本的木质素含量实验值 ($i=1, 2, \dots, 42$)。将校正集光谱数据中每个波长处吸光值构成的吸光值向量按波数从小到大排列,依次记为 $X_{0001}^C, X_{0002}^C, \dots, X_{2074}^C$,其中每个吸光值向量 $X_i^C (i=1, 2, \dots, 2074)$ 都是 35 维的向量。相应地,将验证集的吸光值向量记为 $X_{0001}^V, X_{0002}^V, \dots, X_{2074}^V$,其中每个吸光值向量 $X_i^V (i=1, 2, \dots, 2074)$ 都是 7 维的。由于波长较小的数据噪音成分较大,建模时只选取下标不大于 1400 的吸光值向量,即 $X_{0001}^C, X_{0002}^C, \dots, X_{1400}^C$ 。由于巨大的光谱数据量会降低运算速度,为减小计算量,只选取十分之一的吸光值向量,即 $X_{0001}^C, X_{0011}^C, X_{0021}^C, X_{0031}^C, \dots, X_{1391}^C$ 来建模,然后用验证集中与之对应的吸光值向量 $X_{0001}^V, X_{0011}^V, X_{0021}^V, X_{0031}^V, \dots, X_{1391}^V$ 来检验模型。

为了更好地与偏最小二乘法进行比较,建模前未进行光谱数据预处理。首先将校正集中的吸光值向量 $X_{0001}^V, X_{0011}^V, X_{0021}^V, X_{0031}^V, \dots, X_{1391}^V$ 划分为 10 个小组,用于建立 10 个子模型,第 $k+1$ 组由 $X_{00k1}^C, X_{01k1}^C, X_{02k1}^C, X_{03k1}^C, \dots, X_{13k1}^C$ 构成 ($k=0, 1, 2, \dots, 9$)。为便于数学表达,将它们简记为 $W_{k1}^C, W_{k2}^C, \dots, W_{k14}^C$ 。对验证集的吸光值向量做相同的分组,并简记为 $W_{k1}^V, W_{k2}^V, \dots, W_{k14}^V$ 。然后根据 10 个子模型的预测结果,通过加权平均的方法得出总的预测值。

2.2 模型的初步建立

下面用 $W_{k1}^C, W_{k2}^C, \dots, W_{k14}^C$ 来建立第 $k+1$ 个子模型 ($k=0, 1, 2, \dots, 9$)。参考物理光学中的朗伯-比尔定律, 本文设 Y^C 可由线性表达式 $Y_k^C = c_{k0}I^C + \sum_{i=1}^{14} c_{ki}W_{ki}^C$ 近似表示, 即设

$$Y^C = Y_k^C + \varepsilon_k \quad (1)$$

式中 c_{ki} ($i=0, 1, 2, \dots, 14$) 为待定常数, ε_k 为误差向量, $\varepsilon_k = (\varepsilon_{k1}, \varepsilon_{k2}, \dots, \varepsilon_{k35})$, I^C 为 35 维的单位向量。

通过多元回归方法求出常数项 c_{ki} 的值, 第 $k+1$ 个子模型的平均相对误差为

$$\overline{\varepsilon}_k = \frac{1}{35} \sum_{m=1}^{35} \frac{|\varepsilon_{km}|}{y_m} \quad (2)$$

对 Y_k^C 进行加权平均, 即可得到 Y^C 的估计值向量 $\sum_{k=1}^{10} q_k Y_k^C$, 其中 q_k 为权重系数

$$q_k = \frac{(\overline{\varepsilon}_k)^{-2}}{\sum_{r=0}^9 (\overline{\varepsilon}_r)^{-2}} \quad (3)$$

对于验证集, 令

$$Y_k^V = c_{k0}I^V + \sum_{i=1}^{14} c_{ki}W_{ki}^V \quad (4)$$

式中 I^V 是 7 维的单位向量。则 $\sum_{k=1}^{10} q_k Y_k^V$ 为 Y^V 的初步预测值向量。

2.3 模型的完善

由于每个子模型中待定常数的数量会直接影响模型的最终预测效果, 所以待定常数数量的选择至关重要。利用 MATLAB 软件通过大量的尝试发现, 对于本文所使用的光谱数据和木质素含量实验数据, 当每个子模型中含有 8 个待定常数时, 预测效果最好。将每个子模型中的待定常数个数从 15 减少为 8 的具体做法如下:

在第 $k+1$ 个子模型中 ($k=0, 1, 2, \dots, 9$), 先用多元回归的方法求出常数 c_{ki} ($i=0, 1, 2, \dots, 14$) 的值, 设使 $|c_{ki}|$ ($i=1, 2, \dots, 14$) 最小的系数为 c_{ki_1} ($1 \leq i_1 \leq 14$), 令 $c_{ki_1} = 0$; 重新使用多元回归的方法求出剩余待定常数的值, 设使 $|c_{ki}|$ ($1 \leq i \leq 14, i \neq i_1$) 最小的系数为 c_{ki_2} ($1 \leq i_2 \leq 14$), 令 $c_{ki_2} = 0$; 重复使用上述方法直至待定常数的个数减少到 8 个为止。使用多元回归方法重新求出第 $k+1$ 个子模型中所剩 8 个待定常数的值, 并计算 Y_k^V 及 q_k , 则 $\sum_{k=1}^{10} q_k Y_k^V$ 即为 Y^V 的最终预测值向量。

表 1 木质素含量预测值及其误差

样本编号	实验值 (%)	预测值 (%)	绝对误差 (%)	相对误差
36	21.12	21.36	0.24	0.0115
37	21.82	21.55	0.27	0.0123
38	22.25	22.68	0.43	0.0194
39	22.66	22.83	0.17	0.0074
40	22.97	23.22	0.25	0.0108
41	23.30	23.25	0.05	0.0021
42	23.80	23.96	0.16	0.0068

3 结果与对比

3.1 模型的预测结果

表 1 列出了向量的最终预测值向量及其误差情况。由表 1 的可知, 实验值与预测值之间的相关系数 $r=0.9702$, 平均相对误差为 0.0100, 模

型的拟合优度 $R^2=0.9126$ 。预测值与实验值的散点图如图 1 所示。

3.2 建模方法的对比

下面将本文所用的建模方法与大家常用的偏最小二乘建模方法进行对比。用偏最小二乘法建模时所使用的数据以及校正集、验证集的划

分与第 2 节完全相同, 所得木质素含量预测值及其误差情况见表 2。

由表 2 的结果可知: 用偏最小二乘法建模所得木质素含量预测值与实验值之间的相关系数 $r=0.9626$, 平均相对误差为 0.0107, 模型的拟合优度 $R^2=0.9011$ 。从各项指标的对比看出: 本文所用建模方法优于偏最小二乘建模方法, 所建模型具有更好的预测效果。

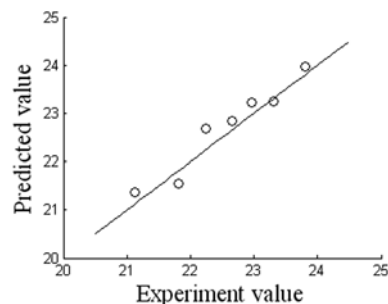


图 1 模型的预测结果

表 2 PLS 方法建模所得木质素含量预测值及误差

样本编号	预测值 (%)	绝对误差 (%)	相对误差
36	21.31	0.19	0.0090
37	21.40	0.42	0.0194
38	22.63	0.38	0.0172
39	22.76	0.10	0.0044
40	23.23	0.26	0.0113
41	23.19	0.11	0.0045
42	24.01	0.21	0.0088

4 结论

本文给出了建立近红外光谱分析模型的一种多模型方法, 并将该方法用于建立欧美杨木质素含量的预测模型。模型的预测值与实验值之间的相关系数为 0.9702, 平均相对误差为 0.0100, 模型的拟合优度为 0.9126; 所建模型比用偏最小二乘法建立的模型具有更好的预测效果。为便于与偏最小二乘法进行对比, 文中用多模型方法建模时未对光谱数据进行预处理。如果做某种预处理, 模型的预测精度有望进一步提高。此外, 建立数学模型的过程中只选取了十分之一的光谱数据, 如果选取所有光谱数据用于建模, 模型的预测结果也可能进一步提高。文中所用的多模型建模方法有望用于建立其它近红外光谱分析模型。

5 致谢

感谢北京林业大学材料科学与技术学院提供数据, 感谢张勇老师、姚胜博士在本文的完成过程中提供的帮助!

参考文献

- [1] 陆婉珍. 现代近红外光谱分析技术 (第二版) [M]. 北京: 中国石化出版社, 2006, 6.
- [2] 严衍禄, 赵龙莲, 韩东海, 等. 近红外光谱分析基础与应用 [M]. 北京: 中国轻工业出版社, 2005, 1.
- [3] 李艳坤, 邵学广, 蔡文生. 基于多模型共识的偏最小二乘法用于近红外光谱定量分析 [J]. 高等学校化学学报, 2007, 28(2):246-249.
- [4] 聂黎行, 王钢力, 李志猛. 近红外光谱法对同仁乌鸡白凤丸的定性和定量分析 [J]. 红外与毫米波学报, 2008, 27(3):205-209.
- [5] 李文龙, 邢丽红, 薛东升, 等. 一种基于近红外光谱技术的熊胆粉鉴别方法 [J]. 光谱学与光谱分析, 2011, 31(3):673-676.
- [6] 徐云, 王一鸣, 吴静珠, 等. 用近红外光谱检测牛奶中的三聚氰胺 [J]. 红外与毫米波学报, 2010, 29(1):53-56.
- [7] Kelley S S, Rials T G, Snell R, et al. Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood [J]. Wood Science and Technology, 2004, 38(4):257-276.
- [8] 杨忠, 江泽慧, 费本华, 等. 近红外光谱技术及其在木材科学中的应用 [J]. 林业科学, 2005, 41(4):177-182.
- [9] Schimleck L R, Kube P D, Raymond C A. Extending Near Infrared Reflectance(NIR) Pulp Yield Calibrations to New Sites and Species [J]. Journal of Wood Chemistry and Technology, 2006, 26(4):299-311.

- [10] 李改云, 黄安民, 王戈, 等. 近红外光谱法快速测定毛竹 Klason 木质素的含量 [J]. *光谱学与光谱分析*, 2007, **27**(10):1977-1980.
- [11] Sykes R, Li B, Hodge G, et al. Prediction of loblolly pine wood properties using transmittance near-infrared spectroscopy [J]. *Canadian Journal of Forest Research*, 2005, **35**(10):2423-2431.
- [12] Poke F S, Raymond C A. Predicting Extractives, Lignin, and Cellulose Contents Using Near Infrared Spectroscopy on Solid Wood in Eucalyptus globules [J]. *Journal of Wood Chemistry and Technology*, 2006, **26**(2):187-199.
- [13] 聂志东, 韩建国, 玉柱, 等. 近红外反射光谱法测定苜蓿干草主要纤维成分的研究 [J]. *光谱学与光谱分析*, 2008, **28**(5):1045-1048.
- [14] Hodge G R, Woodbridge W C. Global near infrared models to predict lignin and cellulose content of pine wood [J]. *Journal of Near Infrared Spectroscopy*, 2010, **18**(6):367-380.
- [15] Green B, Jones P. D, Nicholas D D, et al. Non-destructive assessment of Pinus spp. wafers subjected to Gloeophyllum trabeum in soil block decay tests by diffuse reflectance near infrared spectroscopy [J]. *Wood Science and Technology*, 2011, **45**(3):583-595.
- [16] Schwanninger M, Rodrigues J C, Gierlinger N, et al. Determination of lignin content in norway spruce wood by fourier transformed near infrared spectroscopy and partial least squares regression. Part 1: Wavenumber selection and evaluation of the selected range [J]. *Journal of Near Infrared Spectroscopy*, 2011, **19**(5):319-329.
- [17] 刘胜, 张文杰. 迭代法在相思树近红外光谱分析中的应用 [J]. *分析化学*, 2011, **39**(1):129-132.

新闻动态 News

加拿大 L-3 WESCAM 公司推出适合地面应用的 MX-RSTA 电光 / 红外成像系统

据 www.asdnews.com 网站报道, 加拿大 L-3 WESCAM 公司最近发布了一款专为地面车辆军事行动设计的 MX-RSTA 电光 / 红外成像系统。用户可以将这种系统配置成指挥官专用观察仪、主炮手瞄准系统或者是安装在支杆上的车辆侦察与监视系统。

相关人员表示, MX-RSTA 电光 / 红外成像系统是一种可以满足陆军目前及未来需求的非常先进的地面系统。该系统专门设计用来为地面军用车辆的驾驶员提供高度稳定的高清晰成像能力, 以帮助人们增强作战平台的杀伤力、军用车辆的生存能力以及对战士的保护效果。

经历两年的研制工作, 模块化的 MX-RSTA 系统成功地将最新的稳定技术与通用系统万向支架及可互换的传感器 / 激光载荷集于一身。这种设计能够实现灵活的系统解决方案, 比如在项目要求、应用范围和预算约束等方面发生变化时, 该系统可以方便地更换传感器载荷。

MX-RSTA 系统的基本配置是一个充分加固型载荷外加一台高清晰昼夜成像仪、一台致冷型中波红外成像仪以及一台对人眼安全的激光测距仪。此外, 该系统的可选载荷包括致冷型或非致冷型长波红外成像仪、微光电视成像仪以及对人眼安全的激光照明器。MX-RSTA 系统中采用的先进图像处理技术可以使地面干扰对系统的影响降至最低, 并可为用户提供更加清晰的电光 / 红外图像。



□ 岳桢干