

毛竹化学成分光谱分析的快速建模方法研究

李改云, 黄安民, 秦特夫*

中国林业科学研究院木材工业研究所, 北京 100091

摘要 研究了毛竹综纤维素和木质素含量近红外光谱定量分析方法的建立。选用不同竹龄、纵向和横向部位的54个竹材粉末样品,用湿化学方法测定其综纤维素和木质素含量。在综纤维素和木质素含量的分布范围内,从低值、中间值和高值中挑选11个代表性样品,11个样品按预定比例混合得到21个混合样品,混合样品的综纤维素和木质素含量计算得到,再挑选22个目标成分含量不同的样品组成54个样品的校正集。用偏最小二乘法分别建立样品综纤维素含量、木质素含量和近红外漫发射光谱之间的相关模型。结果表明,综纤维素含量的预测模型的相关系数(R_p)为0.92,标准偏差(SEP)为1.04%;木质素含量的预测模型的 R_p 为0.93,SEP为0.91%,与常规方法建立的模型预测精度相似。说明利用样品混合的方式可快速增加校正集样品的数量、改善校正集样品的分布状况,继而建立稳定可靠的近红外定量分析模型。

关键词 近红外光谱;综纤维素;木质素;混合样品;建模方法

中图分类号: S795.9 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2009)07-1868-04

引言

近红外光谱技术(NIR)是一种高效快速的现代分析技术,具有分析时间短、样品需要量少、操作简单和重现性好等优点,已经在石油化工、农业、医药、造纸、烟草等众多领域得到了广泛的应用^[1,2]。近年来在木材科学方面也日益受到重视,国外已有将NIR用于木材原料化学组成、物理力学性质等的研究报道^[3-11],国内中国林业科学研究院也已从事NIR在木材科学上的应用研究^[12,13]。从用NIR技术预测木材的化学成分国内外文献来看^[14-19],为了采集足够多的代表性样品,人们在建立木材等植物纤维原料中化学成分的定量分析模型时,常依据原料化学成分的变异特点,采集不同年龄、种植密度、生长部位等材质影响因子下的大量样品。众所周知,测量上述化学成分含量的方法一般是通过常规实验室湿化学方法,这种方法操作手续繁琐,测量一个样品往往需要耗费2~3d的时间,需要消耗大量的人力、物力及财力,样品的采集和湿化学分析已严重制约了近红外光谱技术在植物纤维材料上的应用。另外,作者在实际研究中发现,湿化学方法测完大量样品的参考值后,有时出现样品化学成分含量分布密集、样品数分布不均匀的情况,因此需进行下一轮的实地采样,无疑增加了建立模型的时间和成本。

为降低建模成本及缩短建模时间,本研究以毛竹粉末状样品中综纤维素含量和木质素含量的分析为例,采用样品混合的方式得到混合样品,经计算得到混合样品的综纤维素含量和木质素含量,以此来探讨快速、可靠的植物纤维材料化学成分定量分析模型的建立方法。

1 实验部分

1.1 样品

54个毛竹样品采自浙江省杭州市萧山区,采样方法详见文献^[20]。综纤维素含量和木质素含量分别按照GB/T 2677.10—1995和GB/T 2677.8—94测定。

1.2 仪器与软件

LabSpec® Pro近红外光谱仪,美国ASD公司,低噪声512阵元PDA检测器,Indico pro图形界面接口软件,杯光源探头,波长范围350~2500nm,分辨率:3nm@700nm,10nm@1400,2100nm。

Unscrambler®(V9.2)化学计量学软件包,挪威CAMO公司。

1.3 数据采集和处理

样品于装有空调的恒温室(20±2)内敞口放置2周,含水率调节到10%以下。用杯光源探头采集样品的近红外光

收稿日期: 2008-08-10, 修订日期: 2008-11-20

基金项目: 国家“十一五”科技攻关项目(2006BAD18B10)和中央级公益性科研院所基金项目(CAFINT2007C04)资助

作者简介: 李改云,女,1974年生,中国林业科学研究院木材工业研究所助理研究员 e-mail: ligy@caf.ac.cn

*通讯联系人 e-mail: qintefu@caf.ac.cn

谱,扫描 30 次,每个样品装样 3 次。

利用 Unscrambler 化学计量学软件包对每个样品的 3 个光谱平均,并对光谱数据进行微分和平滑处理,用偏最小二乘法和完全交互验证方式建立样品化学成分含量和光谱数据之间的定量分析模型。

2 结果与分析

2.1 混合样品的制备

Table 1 Summary of device for samples mixture and computed results

混合样品	来源	综纤维素含量/ %	木质素含量/ %
1	E + F	61.14	21.82
2	E + F	61.36	22.15
3	F + G	61.70	20.69
4	G + H	62.69	21.07
5	A + E	62.95	21.92
6	I + J	63.75	20.74
7	H + K	63.77	21.35
8	A + B	65.54	21.90
9	B + D + E	65.60	23.93
10	B + C	66.81	22.70
11	B + C	66.91	23.09
12	B + C	67.01	23.50
13	B + C	67.14	24.00
14	B + D	67.44	24.40
15	B + D	67.58	24.80
16	B + D	67.69	25.40
17	B + D	67.72	25.20
18	B + D	67.93	25.80
19	B + D	68.03	26.10
20	B + D	68.18	26.50
21	B + D	68.29	26.84

实验用 54 个样品选自 3 个竹龄、2 个纵向高度和竹壁的 3 个横向部位,综纤维素含量和木质素含量按照国标方法分析得到实测值,综纤维素含量分布范围为 59.68% ~ 70.49%,木质素含量分布范围为 19.67% ~ 28.48%。在兼顾木质素含量的同时,主要依据综纤维素含量,在 54 个样品中挑选包含低值、中间值到高值的部分样品。确定均匀分布在这部分样品范围内的综纤维素含量数值,计算参与混合样品的质量比,样品按比例称重后计算混合样品的实际综纤维素和木质素含量。

本研究从 54 个样品中挑选 11 个样品(编号 A ~ K),11 个样品按比例混合得到 21 个混合样品(编号 1 ~ 21),具体混合配样方案和计算结果见表 1。

在建立植物材料化学成分含量的定量分析模型时,常依据植物材料化学成分的变异特点,采集各个材质影响因素下的样品,大量样品的采集和湿化学分析增加了建模的时间和成本。Yeh 等^[21]尝试用木材中分离出的综纤维素和磨木木质素,按比例混合得到不同木质素含量的样品,结果表明该方

法建立的定量分析模型可靠性差,预测效果不理想。原因是纯的综纤维素和木质素的混合物与木材的背景信息不同,即光谱特征不同。而植物纤维材料自身之间的混合则保证了光谱特征的相似性。

2.2 校正集样品的选择

在近红外光谱分析中,校正集样品的选择及样品基础数据测量的准确性直接决定所建立校正模型的适用性和准确性^[22]。为了确保校正集样品的代表性和样品数量,除参与混合样品制备的 11 个样品和 21 个混合样品外,从 54 个实测样品中再挑选与上述样品的综纤维素含量不同的 22 个样品,共同组成校正集样品,样品的综纤维素含量和木质素含量的累计分布见图 1。对比实地采集的 54 个样品的综纤维素含量和木质素含量的分布(图 2),包含混合样品的校正集不仅覆盖了整个采集样品的分布区域,而且由于混合样品的存在,使校正集中的样品分布更均匀。

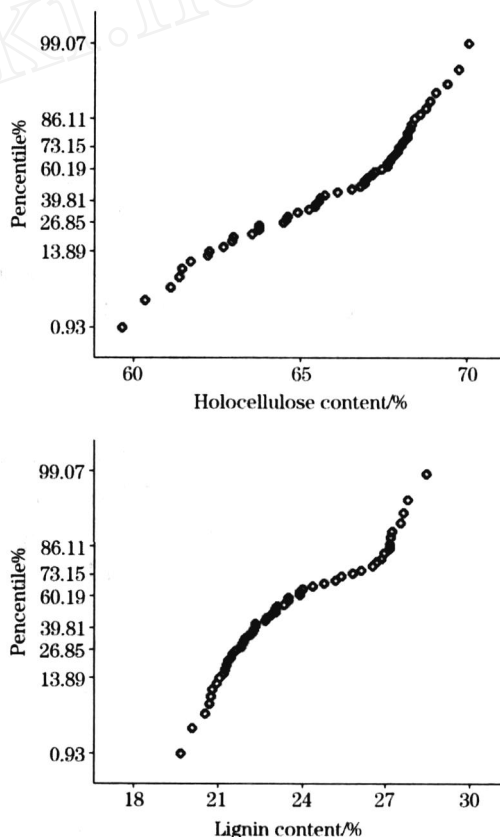


Fig. 1 Cumulative distribution of holocellulose and lignin contents of 33 collected samples and 21 mixed samples

2.3 定量分析模型的建立

用杯光源探头采集样品的漫反射光谱,然后用 Savitsky-Golay 卷积平滑方法对光谱数据同时进行 1 阶微分和 15 点平滑处理。用偏最小二乘法和完全交互验证方式建立综纤维素和木质素含量的校正模型和预测模型。此外,用同样的方法建立 54 个实测样品的定量分析模型,结果见表 2。

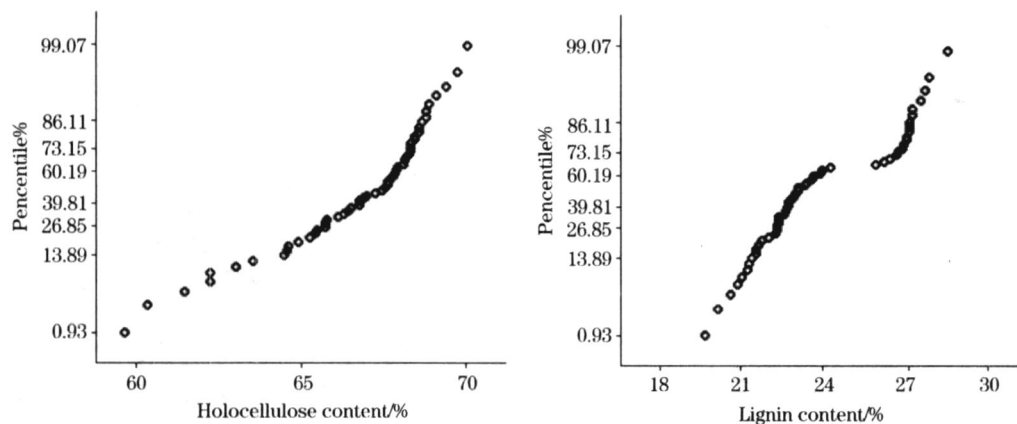


Fig 2 Cumulative distribution of holocellulose and lignin contents of 54 collected samples

Table 2 Statistics of calibration model and prediction model for holocellulose content and lignin content

样品	样品数	化学成分	主成分数	校正模型		预测模型	
				R_c	SEC	R_p	SEP
混合 + 采集	54	综纤维素	6	0.96	0.72	0.92	1.04
		木质素	6	0.97	0.56	0.93	0.91
采集	54	综纤维素	6	0.94	0.80	0.88	1.13
		木质素	6	0.97	0.62	0.94	0.85

从表 2 可知, 在样品数和建模方法等完全相同的条件下, 混合样品参与建模的模型预测效果均与完全由实测样品建立的模型预测效果相似, 说明利用混合样品参与建模的方法具有可行性。并且, 在建立综纤维素含量定量分析模型的同时, 也可建立木质素含量的定量分析模型。与完全由实测样品建立模型的方法相比, 该方法只需要采集少量的目标化学成分含量不同的样品, 即可获得建模所需的较多化学成分含量不同的其它样品, 有效减少了样品的实地采集量及实验室分析工作, 提高了建模速度。通过控制化学成分含量的均匀分布还可有效地提高模型的精度, 对推广近红外光谱定量分析方法在植物纤维材料中的应用具有重要的理论和实际意义。

3 结 论

从 54 个综纤维素和木质素含量已知的样品中, 挑选了 11 个具有代表性的样品, 按预定的比例混合得到目标成分均匀分布的 21 个混合样品, 混合样品的存在使校正集样品在综纤维素含量和木质素含量的分布范围内更均匀。用偏最小二乘法分别建立了综纤维素含量和木质素含量的定量分析模型, 取得了与常规方法建立的模型相似的预测效果, 说明可以通过样品混合方式快速得到建模用不同化学成分含量的其它样品, 继而建立稳定可靠的近红外定量分析模型。

参 考 文 献

- [1] CHU Xiao-li, YUAN Hong-fu, LUO Xian-hui, et al (褚小立, 袁洪福, 骆献辉, 等). Spectroscopy and Spectral Analysis (光谱学与光谱分析), 2008, 28(6): 1227.
- [2] CHEN Jian, CHEN Xiao, LI Wei, et al (陈建, 陈晓, 李伟, 等). Spectroscopy and Spectral Analysis (光谱学与光谱分析), 2008, 28(8): 1806.
- [3] Schimleck L R, Mora C, Daniels R F. Canadian Journal of Forest Research, 2003, 33(12): 2297.
- [4] Martin M E, Aber J D. Journal of Near Infrared Spectroscopy, 1994, 2(1): 25.
- [5] Taylor A, Lloyd J. Forest Products Journal, 2007, 57(12): 116.
- [6] Kelley S S, Rials T G, Snell R, et al. Wood Science and Technology, 2004, 38(4): 257.
- [7] Hodge G R, Woodbridge W C. Journal of Near Infrared Spectroscopy, 2004, 12(6): 381.
- [8] Schimleck L R, Evans R, Llic J, et al. Canadian Journal of Forest Research, 2002, 32(1): 129.
- [9] Schimleck L R, Kube P D, Raymond C A, et al. Canadian Journal of Forest Research, 2005, 35(12): 2797.
- [10] Rials T G, Kelley S S, So C L. Wood and Fiber Science, 2002, 34(3): 398.
- [11] So C L, Via B K, Groom L H. Forest Products Journal, 2004, 54(3): 3.

- [12] HUANG An-min, JIANG Ze-hui, LI Gai-yun (黄安民, 江泽慧, 李改云). Spectroscopy and Spectral Analysis (光谱学与光谱分析), 2007, 27(7): 1328.
- [13] JIANG Ze-hui, LI Gai-yun, WANG Ge, et al (江泽慧, 李改云, 王戈, 等). Chemistry & Industry of Forest Products (林产化学与工业), 2007, 27(1): 15.
- [14] Huang A M, Li G Y, Fu F, et al. Journal of Wood Chemistry and Technology, 2008, 28(3): 194.
- [15] Jones P D, Schimleck L R, Peter G F, et al. Canadian Journal of Forest Research, 2005, 35(1): 85.
- [16] Poke F S, Wright J K, Raymond C A. Journal of Wood Chemistry and Technology, 2004, 24(1): 55.
- [17] Raymond C A, Schimleck L R. Canadian Journal of Forest Research, 2002, 32(1): 170.
- [18] Gierlinger N, Schwanninger M, Hinterstoisser B, et al. Journal of Near Infrared Spectroscopy, 2002, 10(3): 203.
- [19] Schimleck L R, Wright P J, Michell A J, et al. Appita Journal, 1997, 50(1): 40.
- [20] LI Gai-yun, HUANG An-min, WANG Ge, et al (李改云, 黄安民, 王戈, 等). Spectroscopy and Spectral Analysis (光谱学与光谱分析), 2007, 27(10): 1977.
- [21] Yeh Tingfeng, Chang Houmin, Kadla J F. Journal of Agricultural and Food Chemistry, 2004, 52(6): 1435.
- [22] LU Wan-zhen, YUAN Hong-fu, XU Guang-tong, et al (陆婉珍, 袁洪福, 徐广通, 等). Modern Analysis Technique for Near-Infrared Spectra (现代近红外光谱分析技术). Beijing: China Petrochemical Press (北京: 中国石化出版社), 2000. 62.

Rapid Modeling Method for Spectroscopic Analysis of Chemical Components of Bamboo

LI Gai-yun, HUANG An-min, QIN Te-fu*

Research Institute of Wood Industry, Chinese Academy of Forestry, Beijing 100091, China

Abstract A rapid modeling method for predicting the chemical components contents of bamboo was presented. The holocellulose contents and lignin contents of 54 samples from three growth years, two longitudinal positions and three radial positions were analyzed according to traditional chemical methods. Eleven samples were selected based on their holocellulose content and lignin content from these 54 samples to cover the range of holocellulose content and lignin content. Eleven samples were mixed at pre-set ratio with each other to give 21 mixed samples, the holocellulose content and lignin contents of which were computed. Another 22 samples with different chemical component contents were selected from the same 54 samples. The relationship between the chemical component contents and the diffuse reflectance NIR spectra of these samples was established using partial least squares regression. The correlation coefficient of prediction model for holocellulose content and lignin content was 0.92 and 0.93, respectively. The standard error of prediction for holocellulose content and lignin content was 1.04% and 0.91%, respectively. The prediction results were similar to those from the prediction models developed by traditional methods. The results presented in this study demonstrate that samples can be prepared rapidly by the mixture of samples with each other and their chemical component contents can be computed. The technique will significantly reduce sampling time and analyzing time without adversely affecting the quality of the model.

Keywords Near infrared spectra; Holocellulose; Lignin; Mixed sample; Modeling method

(Received Aug. 10, 2008; accepted Nov. 20, 2008)

* Corresponding author