

DOI: 10.13332/j.1000-1522.20150299

基于小波压缩的木材密度近红外光谱的预处理研究

李耀翔¹ 李颖¹ 姜立春²

(1 东北林业大学工程技术学院 2 东北林业大学林学院)

摘要:近红外光谱数据维数多、数据量大,直接保存需要庞大储存空间,且海量数据会对网络化在线检测的分析速度和准确性产生影响。为探讨应用小波压缩进行近红外光谱预处理的可行性及其对枫桦木材密度预测精度的影响,通过强光探头采集木材圆盘的近红外光谱,在Matlab软件中应用小波变换法对枫桦木材密度近红外光谱数据进行压缩。结果表明:当小波基sym2分解层为6时,基于均衡稀疏标准形式的全局硬阈值压缩效果最好,将2151个变量压缩成38个小波系数,其能量保留成分、零系数成分、压缩比分别为99.66%、98.34%、56.61%。用未处理光谱数据和压缩后的38个小波系数分别建立偏最小二乘定标分析模型,同时做内部交叉验证,并用未处理和压缩后的预测集做外部检验,得知压缩后校正模型对压缩后样品预测能力较好,预测决定系数为0.9139。因此,小波压缩可有效简化近红外光谱数据,提高近红外光谱对枫桦木材密度的预测精度。

关键词:近红外光谱;小波压缩;枫桦;木材密度

中图分类号:S781.31 **文献标志码:**A **文章编号:**1000-1522(2016)03-0089-06

LI Yao-xiang¹; LI Ying¹; JIANG Li-chun². **Pretreatment of near-infrared spectroscopy of wood based on wavelet compression.** *Journal of Beijing Forestry University* (2016)38(3) 89-94 [Ch, 14 ref.]

1 College of Engineering and Technology, Northeast Forestry University, Harbin, Heilongjiang, 150040, P. R. China;

2 College of Forestry, Northeast Forestry University, Harbin, Heilongjiang, 150040, P. R. China.

Due to the multi-dimension of near infrared spectrum (NIRS) and large volume of data, huge storage space is needed for data processing, which directly affects the speed and accuracy of online data analysis. This study aims to discuss the feasibility of pretreatment of near-infrared spectroscopy of wood based on wavelet compression as well as its effect on prediction accuracy of *Betula costata* Trautv wood density using NIR technology. The NIRS data of *B. costata* wood were compressed using wavelet transform algorithm with the aid of Matlab. Results showed that the global threshold value based on balance sparsity norm and the heuristic threshold value were observed to be the best with decomposition layer of 6 for the sym2 wavelet. With the method, the 2151 variables were compressed into 38 wavelet coefficients, and the corresponding energy reserved component, zero coefficient component and compression ratio were 99.66%, 98.34% and 56.61%, respectively. The partial least squares (PLS) models were developed based on both the original NIRS and the 38 wavelet coefficients after compression. The inner cross validation was used and the external validation was applied to both the original and the compressed dataset. The best prediction results were associated with the calibration model developed with the compressed NIR data with determination coefficient (R^2) of 0.9139. This study indicates that the wavelet compression method could effectively simplify NIRS data and improve the prediction accuracy.

Key words NIRS; wavelet compression; *Betula costata* Trautv; wood density

收稿日期: 2015-08-08 修回日期: 2015-09-14

基金项目: 林业公益性行业科研专项(201504508)、国家自然科学基金项目(31570624)。

第一作者: 李耀翔,教授,博士生导师。主要研究方向:木材检测与材性改良。Email: yaoxiangli@nefu.edu.cn 地址: 150040 黑龙江省哈尔滨市香坊区和兴路26号东北林业大学工程技术学院。

责任编辑: 姜立春,教授,博士生导师。主要研究方向:森林经理。Email: jlichun@nefu.edu.cn 地址: 150040 黑龙江省哈尔滨市香坊区和兴路26号东北林业大学林学院。

本刊网址: <http://j.bjfu.edu.cn>; <http://journal.bjfu.edu.cn>

近红外光谱(NIRS)技术由于分析速度快、预测精度高、绿色无污染等优点,在食品加工、医药检测、工业生产等领域^[1-3]得到深入研究和广泛应用。在林业方面,NIRS技术不但可以检测木材的物化性质,如含水率、强度、综纤维素含量、抽取物质量分数等^[4-6];还可以对微纤丝角、纤维结晶度等木材生长特性进行估测^[7-8]。应用NIRS技术进行无损检测已成为一种趋势,但由于NIRS维数多、数据量大,直接对数据进行保存或处理,不仅会占用庞大存储空间,而且会延长光谱预处理和建模时间,并大大增加模型后期传递和维护的工作量。同时,在对木材及木加工产品进行在线检测时,多台网络化仪器的协同工作依赖于信息资源的实时共享,海量数据无疑会对网络传输产生一定影响,不利于检测的准确性。如何在少丢失甚至不丢失光谱有效信息、保证预测模型精度的前提下,尽可能多地压缩NIRS、提高分析速度,成为困扰广大林业工作者的一道难题。Nicolai等^[9]对db4小波压缩后的苹果糖度NIRS建立了4个核偏最小二乘校正模型,其决定系数均大于0.8;汤彦丰等^[10]利用小波压缩技术将大黄生药NIRS变量由700个减少到44个,计算得到41种大黄生药的鉴别准确率为82.9%;田高友等^[11]对柴油NIRS进行3次小波分解压缩,最终将储存空间减少了7/8。国内外研究报告表明小波压缩可实现NIRS数据的有效压缩,但小波压缩NIRS在林业领域的应用鲜有报道。

本文以枫桦(*Betula costata*)木材密度NIRS为研究对象,在Matlab(R2011b)软件的小波工具箱中分析不同小波基、小波分解层数和阈值量化对光谱数据压缩效果的影响,通过能量保留成分、零系数成分等指标确定最优压缩方案,建立枫桦木材密度定标分析模型,以期为木材在采伐阶段的分流利用提供理论依据与技术支撑。

1 材料与方法

1.1 样品制备

枫桦样品采自吉林省白山市靖宇县景山实验林场,该林场位于东经126°37'~127°、北纬42°29'~42°40',在林地中选取8棵标准木,以1 m为间隔,沿树干方向向上截取厚度为5 cm的圆盘,共获取枫桦圆盘94个,按3:1比例,随机分配校正集、预测集样品,校正集样品数为70,预测集样品数为24。

1.2 数据获取

使用美国ASD公司生产的LabSpec® Pro FR/A114260便携式光谱扫描仪,在样品采伐完成当天基于全波段光谱范围,用强光探头对枫桦圆盘进行

近红外光谱采集,每1 nm记录1个信息点,样品扫描30次后自动平均为1个光谱,每个圆盘采集光谱3次,保证3个光斑尽可能多地覆盖到圆盘全部面积,取其平均作为最终值。

采用排水法测量木材密度,具体步骤详见GB/T1933—2009^[12]。

1.3 小波压缩原理与效果评价

小波变换进行数据压缩处理的基本方法是选择适合的小波及分解层数对原始NIRS进行小波分解,得到不同小波系数,对小波域中各分解尺度下的高频系数进行阈值量化处理,剔除微小系数,低频系数保持不变,最后对处理后的系数做逆变换完成光谱重构。压缩后光谱以小波系数形式保存,即用适当的小波系数替代原始NIRS进行存储,并进行后续处理,从而达到压缩目的。小波压缩的实质可归结为在小波分解域内对信号进行小波系数量化。

对枫桦密度近红外光谱做小波压缩不仅要尽可能多地减少数据的储存空间,更要保证NIRS的有效信息少丢失甚至不丢失,因此本文采用能量保留成分、零系数成分、压缩比作为小波压缩效果的评价指标。能量保留成分越大,压缩过程中失真越小,压缩后保留的有效数据越完整;零系数成分大,表明压缩过程中置零的系数多,压缩量大,冗余信息剔除的更彻底;压缩比直接体现压缩后数据存储空间的大小,压缩比越大,需要的存储空间越小,反之则大。在评价小波压缩效果时,不能过度追求单一指标最大化,要综合考虑上述3项指标,避免出现光谱失真或压缩不完全等不良现象。

2 结果与分析

2.1 确定小波基和分解层数

为确定不同小波基对NIRS压缩效果的影响,先设定小波分解层为4,在均衡稀疏标准规则下对NIRS进行全局软阈值压缩,考察db4、sym4、coif4、demy4种常用小波的压缩指标,选出最优小波基。由表1可知,当小波为sym4时,能量保留成分、零系数成分、压缩比均达到最大,此时压缩效果最好。

表1 不同小波函数的压缩结果

Tab. 1 Compression results with different wavelets %

小波 Wavelet	能量保留成分 Energy reserved component	零系数成分 Zero coefficient component	压缩比 Compression ratio
db4	99.80	93.53	15.04
sym4	99.82	93.53	15.04
coif4	99.65	93.05	13.61
demy	99.59	91.03	9.31

由于 $\text{sym}N$ 是由一组小波够成的小波系,因此需要确定不同 N 值对压缩效果的影响。同样,设定小波分解层为 4,在均衡稀疏标准规则下进行全局软阈值压缩,并确定最优小波基。由表 2 可知:经 $\text{sym}2$ 小波压缩后得到的能量保留成分、零系数成分、压缩比最大,此时压缩效果最好。

表 2 $\text{sym}N$ 小波系的压缩结果Tab. 2 Compression results with a series of $\text{sym}N$ %

SymN	能量保留成分		压缩比
	Energy reserved component	Zero coefficient component	
sym2	99.82	93.66	15.47
sym3	99.81	93.59	15.26
sym4	99.82	93.53	15.04
sym5	99.81	93.50	14.94
sym6	99.79	93.43	14.73
sym7	99.77	93.37	14.53
sym8	99.74	93.30	14.34

NIRS 经小波分解成低频和高频 2 个部分,低频部分对应近似系数,高频部分对应细节系数。低频系数含有原始光谱信号绝大部分有效信息,且波形幅值大、数量少;而以噪音为代表的冗余信息多存在于细节系数中,波形幅值小且数量多。随着分解层数的增加,细节系数呈增加趋势,冗余信息剔除得更彻底;而近似系数在逐步减少,光谱有效信息易被剔除掉,最终将影响小波压缩的效果,因此需要进一步确定最佳分解层数。

由表 3 可知:随着分解层的增加能量保留成分呈递减趋势,零系数呈递增趋势,压缩比呈先增加后递减的趋势。造成这种现象的原因为随着分解层数的增加被剔除掉的细节系数增加,而近似系数减少;当分解层为 9 时,近似系数剔除过多导致压缩比下降。小波压缩的目的是保证有效信息完整的情况下,尽可能提高压缩比。据此观能量保留成分,当分解层为 7 时,能量保留成分较分解层为 6 时出现最大幅度下降;因此,分解层为 6 时, $\text{sym}2$ 小波压缩效果最好,零系数成分、压缩比均分别为

98.34%、56.61%。

表 3 $\text{sym}2$ 小波不同分解层数下的压缩结果

Tab. 3 Compression results with varied wavelet

分解层 Layer	decomposition layers		压缩比 Compression ratio
	能量保留成分 Energy reserved component	零系数成分 Zero coefficient component	
1	99.99	50	1.99
2	99.93	74.97	3.97
3	99.87	87.45	7.88
4	99.82	93.66	15.47
5	99.77	96.77	29.88
6	99.66	98.34	56.61
7	99.24	99.10	102.43
8	98.51	99.34	113.21
9	97.71	99.39	107.55
10	97.78	99.45	102.43

2.2 压缩方法的选择

在小波压缩 NIRS 信号过程中,阈值的选择与量化一定程度上关系到信号压缩的质量。为选择合适的压缩方法,采用均衡稀疏标准和去近零值 2 种阈值选取方法,在全局情况下进行软阈值和硬阈值压缩,压缩小波为 $\text{sym}2$ 、分解层数为 6。

由表 4 可知:经均衡稀疏标准阈值和去近零值阈值压缩后,各自得到的零系数成分和压缩比相同;但在硬阈值下的能量保留成分大于软阈值,这表明硬阈值压缩效果好于软阈值。在软情况下,经均衡稀疏标准阈值压缩后的能量保留成分、零系数成分、压缩比是去近零值的 0.99 倍、1.97 倍、39.31 倍;在硬阈值下,经均衡稀疏标准阈值压缩后的能量保留成分、零系数成分、压缩比是去近零值的 1.00 倍、1.97 倍、39.31 倍。这表明均衡稀疏标准阈值压缩效果优于去近零值,而均衡稀疏标准硬阈值压缩效果最好。

图 1 为未处理的原始 NIRS,小波变换将原始 NIRS 数据(2 151 个变量)压缩成 38 个小波系数,并以这 38 个小波系数代替原始 NIRS 进行后续处理

表 4 不同压缩方式压缩效果的比较

Tab. 4 Comparison of compression results with different compression methods

去噪方式 Denoising	软阈值 Soft threshold			硬阈值 Hard threshold			%
	能量保留成分		零系数成分	压缩比	能量保留成分		零系数成分
	Energy reserve	Zero coefficient	Compression ratio	Energy reserved	Zero coefficient	Compression ratio	
均衡稀疏标准 Balance sparsity-norm	99.41	98.34	56.61	99.66	98.34	56.61	
去近零值 Remove near 0	99.99	50.02	1.44	100	50.02	1.44	

得到图2,即小波压缩后的NIRS。图2光谱吸收峰保留较完整,与图1无显著差异。

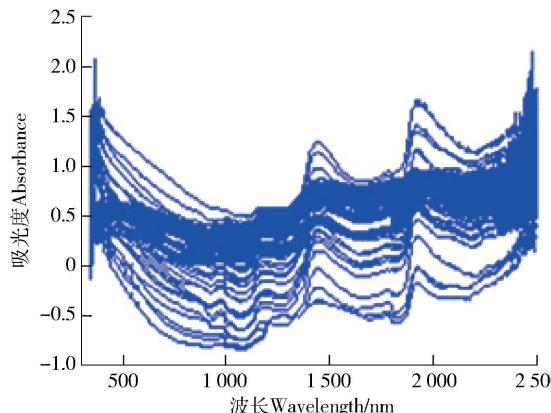


图1 枫桦木材密度的原始近红外光谱

Fig. 1 Raw NIRS of *Betula costata* wood

2.3 模型建立与预测

将经sym2小波6层分解,均衡稀疏标准形式全局硬阈值压缩后的NIRS和枫桦密度实测值导入分析软件Unscrambler中,应用偏最小二乘法(PLS)构建枫桦木材密度近红外校正模型,并用交叉验证法进行模型检验,最后用建好的校正模型对经未处理和小波压缩后的预测集样品密度进行预测。

表5 枫桦木材密度近红外模型的预测结果

Tab. 5 Model statistics for *Betula costata* wood density prediction

预处理方法 Pretreatment	模型 Model	R^2	SEE	RSD/%	未处理预测集 Untreated prediction set			小波压缩预测集 Compression prediction set		
					R^2	SEE	RPD	R^2	SEE	RPD
未处理 Untreated	校正模型 Calibration	0.9017	0.0178	1.8579	0.8753	0.0166	2.8313	0.9117	0.0140	3.3656
	验证模型 Validation	0.8289	0.0235	2.4508						
小波压缩 Compression	校正模型 Calibration	0.8965	0.0183	1.9057	0.8852	0.0158	2.9509	0.9139	0.0138	3.4069
	验证模型 Validation	0.8546	0.0216	2.2589						

注: R^2 为决定系数, SEE 为标准分析误差, RSD 为相对标准差, RPD 为相对分析误差。Notes: R^2 refers to coefficient of determination, SEE refers to the standard error of estimate, RSD refers to the relative standard deviation, and RPD is the ratio of standard deviation to SEE.

用所建模型对预测集样品木材密度进行预测得到:对所有校正模型,压缩预测集决定系数均大于未处理预测集决定系数;同样,不论预测集是否经过处理,压缩校正模型进行预测得到的决定系数均大于未处理校正模型决定系数。综上,压缩校正模型对压缩预测集预测效果最好(图3、4),其RPD为3.4069大于3^[14],定量说明了压缩后校正模型有良好的稳定性能和精准的预测能力。

3 结论与讨论

在Matlab软件中对枫桦木材密度NIRS进行小波压缩,讨论不同小波基、分解层数、阈值对压缩效果的影响,最后建立枫桦木材密度NIRS模型,对未

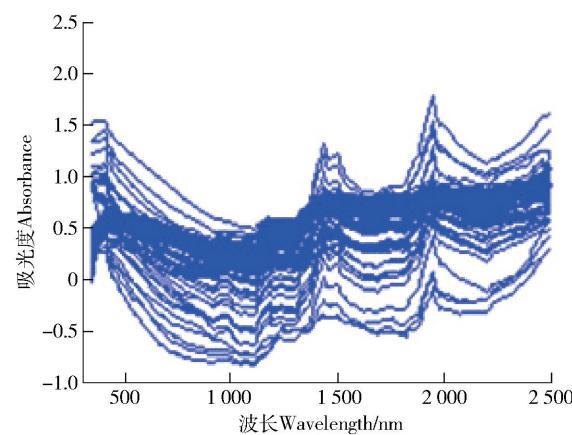


图2 重构后枫桦木材密度的近红外光谱

Fig. 2 Reconstructed NIRS of *Betula costata* wood

由表5可知:通过未处理NIRS建立的校正模型决定系数略大于经小波压缩建立的校正模型决定系数,对校正模型精度的影响甚微。未处理的校正与验证模型标准差之间的差值比经小波压缩后的校正与验证模型标准差之间的差值大73%,表明经小波压缩后的模型信息提取更充分。RSD的值均小于10%^[13],说明模型可用于对未知样品进行密度检测。

知样品进行预测,得到结论如下:1)基于sym2小波6层分解的均衡稀疏标准形式全局硬阈值压缩效果最好。将原始NIRS数据变量由2151个压缩成38个小波系数,各项压缩参数达到最优配置,能量保留成分、零系数成分、压缩比分别为99.66%、98.34%、56.61%,在尽可能多地保留原始NIRS有效信息的前提下,实现了压缩最大化,减少了储存空间。说明小波变换可对枫桦木材密度NIRS进行有效压缩,合理选择相关参数能有效改善压缩效果。2)比较有无小波压缩处理模型的相关参数大小,小波压缩对近红外校正模型决定系数的影响甚微,可忽略不计。应用模型预测样品密度时,压缩后校正模型对压缩后样品的预测决定系数最大,标准分析

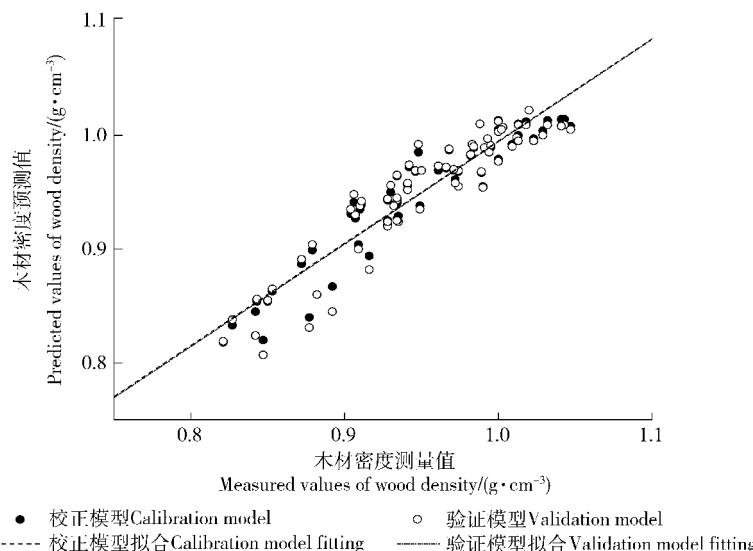


图3 枫桦木材密度实测值与预测值相关性

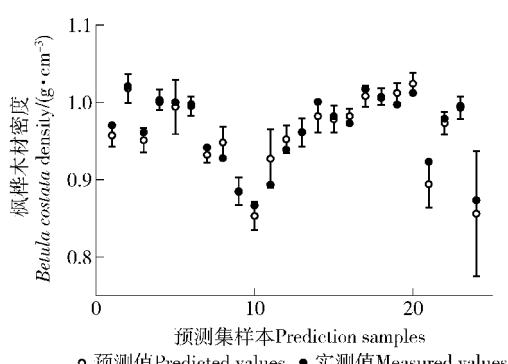
Fig. 3 Relationship between measured and predicted values of *Betula costata* wood density

图4 枫桦木材密度 NIRS 校正模型预测结果

Fig. 4 Prediction results of NIRS calibration model on *Betula costata* wood density

误差最小, 分别为 0.913 9 和 0.013 8, 其预测精度优于基于原始近红外光谱构建的模型。说明小波压缩用于枫桦木材密度 NIRS 的检测是可行的, 可有效提高模型的预测能力。

小波变换在压缩枫桦密度 NIRS 过程中有效地剔除了微小系数, 在减小储存空间的同时, 并未造成校正模型精度大幅度下降, 反而提高了验证模型精度, 增加了模型稳定性, 使预测精度由 0.875 3 提高到 0.913 9。在建立模型时, 只讨论了小波压缩对线性 PLS 模型的影响, 对其他线性和非线性模型的影响有待进一步论证。由于试验样品来自小区域, 在应用本文所建模型对未知样品进行预测时, 要及时扩充、更新数据, 以保证模型具有广泛的适用性。后续研究中, 可以考虑通过 X 射线密度仪测定木材表面微密度作为近红外定标模型的实测值, 进一步探讨采用表面密度作为真值, 是否可以提高木材整体平均密度的近红外预测精度。此外, 因木材密度具

有不均匀性, 通常边材较心材密度更大, 本文旨在寻求一种在木材采伐阶段通过近红外光谱技术获取原木密度的方法, 为木材初期的分流利用提供决策依据; 因此在光谱获取过程中, 没有使用传统的二分光纤, 而是通过强光探头形成光斑采集光谱。所获光谱信息基本覆盖了整个木材圆盘, 可以很好地表征整个圆盘的密度信息度, 而并未对心、边材密度进行分别探讨。在后续的研究中, 可以采用二分光纤法分别采集木材心材、边材的光谱信息, 分心、边材分别构建近红外预测模型, 以期为木材的精细化分流利用提供基础数据。同时, 由于近红外仪器标准化未统一及测试环境的影响, 在使用前要对已建模型进行修正及待检测 NIRS 校正, 以保障模型的预测能力不受外界因素的干扰。

参 考 文 献

- [1] HAUGHEY S A, GRAHAM S F, CANCOUET E, et al. The application of near-infrared reflectance spectroscopy (NIRS) to detect melamine adulteration of soya bean meal [J]. Food Chemistry, 2013, 136(3): 1557-1561.
- [2] NKANSAH K. Rapid characterization of biomass: the use of near infrared and fluorescence spectroscopy as process analytical technology (PAT) method [M]. West Virginia: West Virginia University, 2009.
- [3] DIAZ J T, VEAL M W, CHINN M S. Development of NIRS models to predict composition of enzymatically processed sweet potato [J]. Industrial Crops and Products, 2014, 59: 119-124.
- [4] 郝斯琪, 宋博骐, 李湃, 等. 基于近红外光谱与 BP 神经网络预测落叶松木屑的含水率 [J]. 森林工程, 2012, 28(4): 9-11.
- HAO S Q, SONG B Q, LI P, et al. Prediction of sawdust water content of Dahurian larch based on NIRS and BP neural network [J]. Forest Engineering, 2012, 28(4): 9-11.

[5] 张鹏, 李耀翔. 近红外光谱分析技术在木材机械性能检测中的研究进展[J]. 森林工程, 2014, 30(3): 68-70.
ZHANG P, LI Y X. Thereseach progress on application of near-infrared spectroscopy in wood mechanical properties prediction [J]. Forest Engineering, 2014, 30(3): 68-70.

[6] RAMADEVI P, MEDER R, VARGHESE M. Rapid estimation of kraft pulp yield and lignin in *Eucalyptus camaldulensis* and *Leucaena leucocephala* by diffuse reflectance near-infrared spectroscopy (NIRS)[J]. Southern Forests, 2010, 72(2): 107-111.

[7] 霍淑媛, 姚春丽, 王娜. 近红外光谱法测定欧美杨纤维形态和结晶度[J]. 纸和造纸, 2012, 31(1): 28-31.
HUO S Y, YAO C L, WANG N. Nondestructive estimation of the fiber length and crystallinity of *Populus × euramericana* by near-infrared spectroscopy [J]. Paper and Paper Making, 2012, 31(1): 28-31.

[8] 孔文杨, 刘镇波, 刘一星, 等. 近红外光谱技术在木材材性分析及木质复合材料生产中的应用[J]. 世界林业研究, 2012, 25(4): 45-51.
KONG W Y, LIU Z B, LIU Y X, et al. Research status and prospect of acoustic vibration properties modification of wood used for soundboard[J]. World Forestry Research, 2012, 25(4): 45-51.

[9] NICOLAI B M, THERON K I, LAMMERTYN J. Kernel PLS regression on wavelet transformed NIR spectra for prediction of sugar content of apple [J]. Chemometrics and Intelligent Laboratory Systems, 2007, 85(2): 243-252.

[10] 汤彦丰, 侯占忠, 王志宝, 等. 中草药大黄小波变换的近红外光谱的聚类分析[J]. 安徽农业科学, 2012, 40(30): 14726-14727.

TANG Y F, HOU Z Z, WANG Z B, et al. Cluster analysis of rhubarb from different habitats based on near-infrared spectrometry by wavelet transform[J]. Journal of Anhui Agricultural Sciences, 2012, 40(30): 14726-14727.

[11] 田高友, 袁洪福, 刘慧颖, 等. 小波变换用于近红外光谱数据压缩[J]. 分析测试学报, 2005, 24(1): 17-20.
TIAN G Y, YUAN H F, LIU H Y, et al. Application of wavelet transform to compressing near infrared spectra data [J]. Journal of Instrumental Analysis, 2005, 24(1): 17-20.

[12] GB/T1933—2009 木材密度测定方法 [S]. 北京: 中国标准出版社, 2009.
GB/T1933—2009 Method for determination of the density of wood [S]. Beijing: Standards Press of China, 2009.

[13] 张小超, 吴静珠, 徐云. 近红外光谱分析技术及其在现代农业中的应用 [M]. 北京: 电子工业出版社, 2012.
ZHANG X C, WU J Z, XU Y. Modern NIR analysis technology and its applications in modern agriculture [M]. Beijing: Electronic Industry Press, 2012.

[14] 邱素君, 何雁, 张国松, 等. 近红外光谱快速测定柴胡总皂苷肠溶片包衣膜厚度研究[J]. 中国药学杂志, 2013, 48(24): 2128-2133.
QIU S J, HE Y, ZHANG G S, et al. Fast determination of coating thickness of the total saponin of radix bupleuri enteric coated tablets by NIRS[J]. Chinese Pharmaceutical Journal, 2013, 48(24): 2128-2133.

(责任编辑 李文军
责任编委 赵广杰)