JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Rapid Screening of Wood Chemical Component Variations Using Transmittance Near-Infrared Spectroscopy

TING-FENG YEH,† TATSUHIKO YAMADA,† EWELLYN CAPANEMA,†
HOU-MIN CHANG,† VINCENT CHIANG,‡ AND JOHN F. KADLA*,§

Department of Wood and Paper Science, North Carolina State University, Raleigh, North Carolina 276095-8005, Department of Forestry, North Carolina State University, Raleigh, North Carolina 276095-7247, and Biomaterials Chemistry, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4

A rapid transmittance near-infrared (NIR) spectroscopy method was developed to predict the variation in chemical composition of solid wood. The effect of sample preparation, sample quantity (single versus stacked multiple wood wafers), and NIR acquisition time on the quantification of α-cellulose and lignin content was investigated. Strong correlations were obtained between laboratory wet chemistry values and the NIR-predicted values. In addition to the experimental protocol and method development, improvements in calibration error associated with utilizing stacked multiple wood wafers as opposed to single wood wafers are also discussed.

KEYWORDS: Loblolly pine (*Pinus taeda*); aspen *(Populus trichocarpa*); transmittance near-infrared spectroscopy (NIR); increment cores; wood wafer; α-cellulose content; lignin content; screening

INTRODUCTION

To ensure the global competitiveness of the pulp and paper industry in the southeastern United States, more wood with targeted characteristics has to be produced more efficiently on less land. One viable solution to meet future industrial wood demands is to greatly increase the productivity of current pine plantations, leaving natural forests to be managed at low intensity, primarily for saw timber, conservation, aesthetics, and recreational ends. To achieve efficient utilization of the fast growing plantation wood, tree breeders need to accurately and rapidly screen the large breeding populations for a variety of phenotypic traits.

Wood properties including density, tracheid diameter and length, cell-wall thickness, and chemical composition have been shown to be related to product quality. For example, paper properties such as burst, tear strength, and tensile strength are closely related to fiber morphology (*1*). While processing costs and resultant profitability are more significantly affected by chemical composition (*2*), specifically α-cellulose and lignin content. Traditional wet chemistry methods for the determination of α-cellulose and lignin content are quite costly and time-consuming (*3*).

Recently, we reported a rapid transmittance near-infrared (NIR) spectroscopic method for the determination of lignin content in solid wood (*4*). The lignin content of wood wafers taken from 12 mm increment cores were statistically analyzed using multiple regression and partial least-squares analysis. Strong correlations were obtained between the predicted NIR results and those obtained from traditional chemical methods. This method satisfactorily predicted lignin content for samples not included in the model development. Sykes et al. (*5*) utilized this model to predict fiber length, coarseness, and α-cellulose and lignin content of loblolly pine. However, lignin content could not be adequately predicted using this model because of the large error associated with the lignin measurements.

The single wood wafer NIR method (*4*) enables good prediction of lignin content, but it requires collecting 15 single-wafer NIR spectra from each ring of an increment wood core. These spectra are then averaged to produce a single NIR spectrum. When considering screening a tree-breeding project where the amount of samples is enormous, the time required to analyze a single sample is crucial. In this paper, a new method is proposed wherein a single NIR spectrum is collected utilizing several wafers from the same year ring stacked together. The results obtained from the single stacked wafer NIR spectrum model and the correlation between the wet chemistry and NIR measurements for α-cellulose and lignin content are compared to the averaged multiple single-wafer NIR spectra model.

MATERIALS AND METHODS

Materials. Wood increment core samples were collected from 13 9-year-old loblolly pines (*Pinus taeda*) received from the Tree Breeding Program, Department of Forestry, North Carolina State University, NC, and from 37 4-year-old aspen (*Populus trichocarpa*) received from Oak Ridge National Laboratory, TN. The increment wood core extractives were removed by acetone extraction as described previously (*6*). The

* To whom correspondence should be addressed. Telephone: (604) 827-5254. Fax: (604) 822-9104. E-mail: john.kadla@ubc.ca.
† Department of Wood and Paper Science, North Carolina State University.
‡ Department of Forestry, North Carolina State University.
§ University of British Columbia.

10.1021/jf0480647 CCC: $30.25
Published on Web 04/02/2005

extractive-free increment wood cores were then soaked in deionized water overnight and microtomed into wood wafers (13 mm in diameter and 200 μm in thickness) (*4*). The wood wafers were dried under vacuum over P_2O_5 overnight prior to NIR measurement. All NIR spectra were recorded prior to chemical analysis.

Near-Infrared Spectroscopy. A Foss NIRSystems near-infrared spectrometer equipped with an InTact Single Tablet Module (NR-1650) and a monochromator (NR-6500-V/H) was used to analyze the wood wafers. Absorbance spectra totaling 32 scans were collected at 2.0 nm intervals over the range of 600−1900 nm.

NIR Sample Preparation and Measurement. The dried wood wafers were analyzed using a modified sample holder as reported previously (*4*). The NIR and wet chemistry measurements were conducted on a ring-by-ring basis. All wood wafers collected from a single growth ring were considered as one sample. For both the pine and aspen, only a limited amount of woody material was available for analysis. Pine α-cellulose content was determined using wood wafers collected from the springwood rings 2, 4, 6, and 8, whereas lignin content was determined from springwood rings 3, 5, and 7. A total of 55 and 59 samples were analyzed for α-cellulose and lignin content, respectively. For the aspen samples, the amount of available wood was much lower and only lignin content was determined using the springwood from ring 3. A total of 62 aspen samples were analyzed. In a typical experiment, 10 wood wafers for the pine (corresponding to about 80 mg of wood) or 14 wood wafers for the aspen (corresponding to about 100 mg of wood) were stacked together, placed on the NIR sample holder, and scanned. For the averaged single-wafer model, the 10 (pine) or 14 (aspen) wood wafers obtained per ring were individually scanned and averaged to represent a single sample spectrum. As a result, the regression models developed for both the averaged single-wafer spectrum and the stacked-wafer spectrum were obtained from the same wood wafers and therefore correspond to the same reference data obtained from the wet chemistry analyses.

Holocellulose Preparation. The isolation of holocellulose was carried out according to the protocol of Yokoyama et al. (*6*) utilizing a total of 10 wood wafers per analysis. Specifically, ~100 mg (oven-dried) of wood wafers were suspended in 4 mL of deionized (DI) water at 90 °C and reacted with 200 mg of 80% sodium chlorite and 0.8 mL of acetic acid for 1 h. The reaction mixture was then filtered using a course crucible, washed, and dried at 105 °C until no change in weight was observed. For specimens where 10 wafers were less than 100 mg, the amount of the applied chemicals was reduced proportionally.

α-Cellulose Preparation. α-Cellulose was prepared as per the protocol of Yokoyama et al. (*6*), wherein 50 mg of the holocellulose (outlined above) was reacted with 4 mL of 17.5% sodium hydroxide for 30 min, and then diluted with 4 mL of DI water, and the reaction mixture was left for 30 min. After a total reaction time of 1 h, the fiber suspension was filtered with a coarse crucible, washed thoroughly with DI water, and soaked in 1.0 M acetic acid for 5 min. The neutralized α-cellulose was then washed with deionized water. The yield was calculated after drying at 105 °C.

Lignin Content Determination. The lignin content was determined using a modified Klason lignin method. The wood wafers (~100 mg oven-dried) were reacted with 1.5 mL of 72% H_2SO_4 at room temperature with occasional stirring for 2 h. The solution was then diluted with DI water to a 3% H_2SO_4 concentration and heated at 121 °C and 2 atm for 1 h in a commercial pressure cooker. The reaction was filtered, and the acid-insoluble lignin was determined gravimetrically. The filtrate was diluted to 100 mL with DI water, and the acid-soluble lignin was calculated from the UV absorbance at 205 nm using an extinction coefficient of 110 (AU L)/(g cm) (*7*). The acid-insoluble and acid-soluble lignins were combined and reported as the total lignin content. No statistically significant difference was observed in the total lignin content obtained between the classical Klason lignin method and our pressure-cooking method.

Calibration Development and Statistics. The calibration models were developed using Foss NIRSystems Vision software (version 2.51). First, outliers were identified using the Mahalanobis distance algorithm to measure how far a sample was from the cluster center of the spectra. A sample is considered to be an outlier when its probability level exceeds a threshold value of 0.95 (*8*). Once the outliers were removed, the remaining samples are split 75% for the calibration set and 25% for the prediction set using an algorithm that measures a Euclidean distance between samples. Redundant samples are moved into the prediction set (*8*). A statistical summary of the calibration and prediction sets is given in **Table 1**.

Table 1. Summary Statistics for α-Celluloses and Total Lignin Contents for the Calibration and Prediction Sets

chemical compositions (%)[a]	calibration set					prediction set				
	n	min	max	avg	std[b]	*n*	min	max	avg	std
pine α-cellulose										
stacked-wafer model	38	35.6	47.3	42.4	2.4	12	38.2	46.4	42.0	2.5
single-wafer model	38	35.6	47.3	42.4	2.4	12	38.2	43.8	41.0	2.0
pine total lignin	39	28.0	32.0	30.0	0.9	14	28.5	32.0	29.8	0.9
aspen total lignin	39	20.9	28.6	25.3	2.3	14	21.4	27.1	24.6	1.9

[a] On the basis of extractive-free, OD wood weight. [b] Standard deviation.

Prior to any calibration development, the original spectra were converted to 2nd derivative spectra with a 10 nm segment and 0 nm gap. Calibration equations were developed using a partial least squares (PLS) regression with four cross validation segments and a maximum of 16 factors. The best number of PLS factors for the model was determined by the PRESS (prediction residual error sum of squares) value, which is the sum of all squared differences between the lab and predicted values (*9*). The PLS factors that yield the lowest PRESS values were then chosen to establish a model (*9*, *10*).

The coefficient of determination (R^2), the standard error of calibration (SEC), and the standard error of cross validation (SECV) were used to evaluate the calibration performance. SEC is the standard deviation for the residuals because of the difference between the actual lab values and the fitted values of samples within the calibration set (*11*, *12*). SECV is an indication of how well an equation will predict samples that were not used to generate the calibration equation in cross validation (*8*, *13*, *14*).

The standard error of prediction (SEP) was used to evaluate how well the calibration predicts the interested constituent value for a set of unknown samples that are different from the calibration set (*15*). The predictability of the calibration was evaluated by the ratio of performance to deviation (RPD). The RPD was calculated from the ratio of standard deviation of the reference data of prediction data set to the SEP (*16*). The RPD should be as high as possible; values between 5 and 10 are adequate for quality control, values >2.5 are satisfactory for screening breeding programs (*11*, *16*), and values of ~1.5 can be used as initial screening tools (*15*, *17*, *18*).

RESULTS AND DISCUSSION

Stacked-Wafer Model versus Averaged Single-Wafer Model. Our previous model (*4*) was established by averaging 15 single-wafer NIR spectra from a single growth ring of an increment wood core. The large number of samples (wafers) analyzed ensured that the NIR spectra obtained was representative of the entire year of growth and that enough material was available for the subsequent wet chemistry measurements. However, the analysis of 15 spectra per ring was quite laborious and time-consuming. In an attempt to minimize data collection time but analyze a representative amount of wood, 10 wafers from the same year ring were stacked together and one NIR spectrum was taken. **Figure 1** shows the 2nd derivative NIR spectra obtained from a single NIR spectrum of 10-stacked wafers and the averaged spectrum of 10-single wafer spectra. The intensity of the 2nd derivative NIR spectrum from the 10 stacked wafers was far more intense than that obtained from the averaged spectrum of 10 single wood wafers. In addition to the improved signal-to-noise and reduced calibration error (*19*), the intense NIR absorption bands of the stacked wafer spectra enhance regression development.

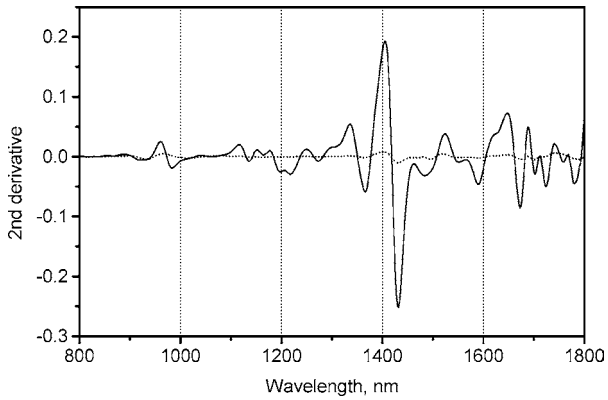

Figure 1. Second derivative NIR spectra of (—) 10-stacked wafers and (•••) averaged 10-single wafer.

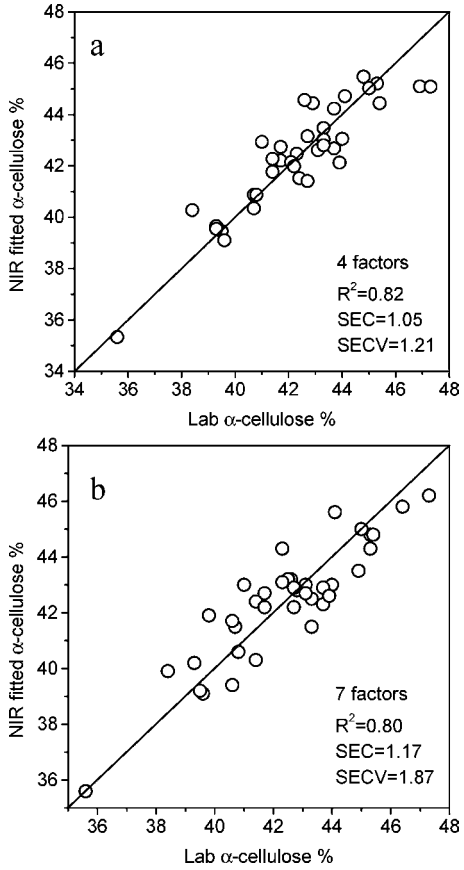

Figure 2. Correlation between the lab α-cellulose content and the NIR-fitted α-cellulose content of the (a) stacked-wafer model and (b) averaged single-wafer model.

The pine wood wafers of the α-cellulose data set were used to develop two calibration models, the stacked-wafer model and the averaged single-wafer model. **Figure 2** illustrates the calibration results of α-cellulose content of loblolly pine for the stacked-wafer model (**Figure 2a**) and the averaged single-wafer model (**Figure 2b**). A stronger correlation was obtained for the stacked-wafer model ($R^2 = 0.82$) than that of the averaged single-wafer model ($R^2 = 0.80$). The SEC was 1.05 for the stacked-wafer model and 1.17 for the averaged single-wafer model. Thus, a better fit was obtained for the regression models of the stacked-wafer model than the averaged single-wafer model. The SECV, which is a better measurement of the calibration error (*18*), exhibited the same trend between both models, i.e., smaller calibration error in the stacked-wafer model,

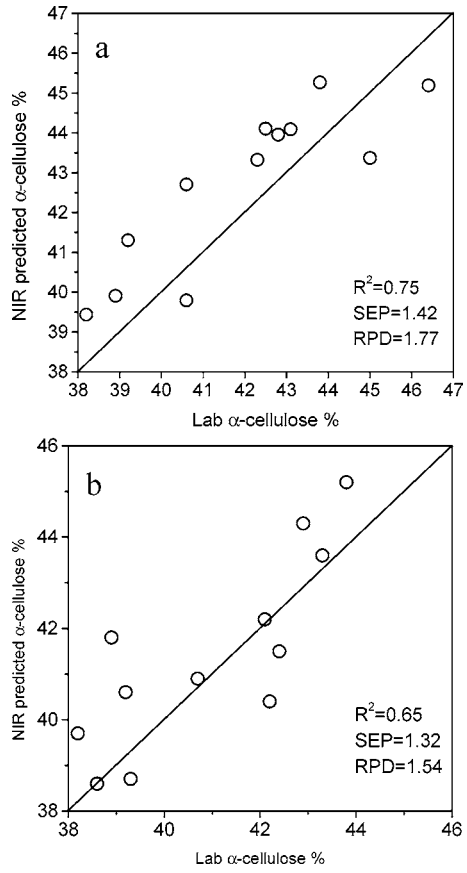

Figure 3. Correlation between lab-measured α-cellulose content and the NIR-predicted α-cellulose content using (a) stacked-wafer model and (b) averaged single-wafer model.

SECV = 1.21 versus 1.87 for the averaged single-wafer model. The considerably larger SECV than SEC in the averaged single-wafer model could possibly be due to over-fitting of the data (*18*). The lower signal-to-noise ratio in the averaged single-wafer spectra (**Figure 1**) could result in some of the noise being modeled during calibration development, thereby reducing the SEC. Therefore, using stacked wafers as opposed to averaging single-wafer measurements is not only faster but results in a better signal-to-noise ratio, which leads to a reduction in the calibration error. The stacked-wafer model will be utilized in all of the proceeding analyses.

Prediction of α-Cellulose Content of Loblolly Pine. The α-cellulose content calibration models were tested using the prediction sample sets (12 loblolly pine wood wafer samples). The relationship between the wet chemistry α-cellulose content measurements and the NIR predicted α-cellulose contents are quite good using the stacked-wafer model ($R^2 = 0.75$). As shown in **Figure 3a**, the SEP is 1.42, which is slightly higher than the SECV (1.21). The RPD is 1.77 indicating that the stacked-wafer model could be used as a screening tool for estimating the α-cellulose content of increment core samples. **Figure 3b** illustrates the correlation between the wet chemistry α-cellulose values and the NIR-predicted values for the averaged single-wafer model ($R^2 = 0.65$), which is not as good as the stacked-wafer model. Furthermore, the SEP is closer to the SEC than SECV, and the RPD (1.54) is lower than that of the stacked-wafer model. Thus, the predictability of the single-wafer model is weaker than the stacked-wafer model. However, they both fulfill the initial screening criterion (RPD = ~1.5).

PLS Calibrations Based on Lignin Content. The stacked-wafer method was also applied to develop two lignin content

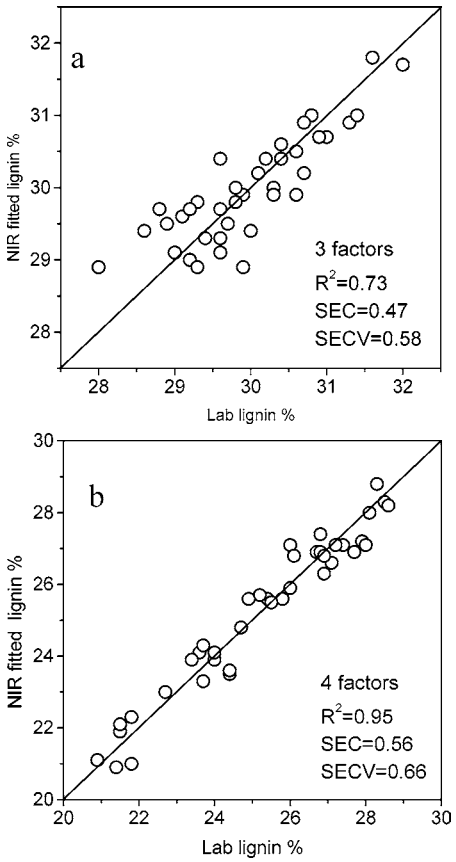

Figure 4. Correlation between the wet-chemistry-measured total lignin content and the NIR-fitted lignin content for (a) loblolly pine and (b) aspen.

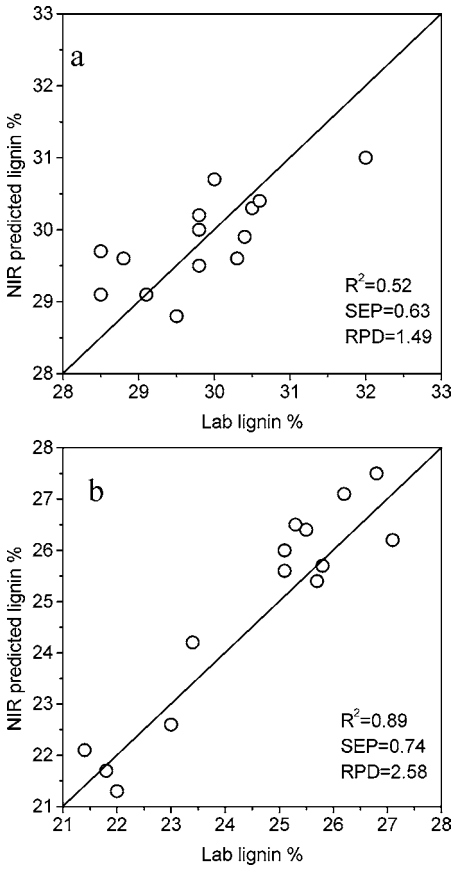

Figure 5. Correlation between lab-measured total lignin content and the NIR-predicted lignin content for (a) loblolly pine and (b) aspen.

calibration models based on loblolly pine and aspen. The calibration results are shown in **Figure 4**. The correlation of the pine wet chemistry lignin values and the NIR-predicted lignin values was quite good ($R^2 = 0.73$, **Figure 4a**), where the SEC and SECV were 0.47 and 0.58, respectively. These considerably low values as compared to the α-cellulose models (**Figure 2a**) may be due to the lower laboratory error for the reference methods, 0.55 for lignin and 1.05 for α-cellulose. Interestingly, the correlation obtained for the aspen data set was very strong ($R^2 = 0.95$, **Figure 4b**), where the SEC and SECV was 0.56 and 0.66, respectively. The enhanced calibration performance of the aspen data set as compared to the pine data set ($R^2 = 0.73$) may be due to the range of lignin content present (*11*, *19*). The aspen has a broad lignin content range (21−29% lignin) in which the samples are uniformly distributed as compared to the pine samples (28−32% lignin) (*20*). Therefore, to improve the calibration performance of loblolly pine, more samples with greater differences in lignin content over a wider range of lignin content values are needed.

Prediction of Lignin Content. Both lignin content calibration models were tested on the loblolly pine and aspen prediction sets (14 samples in each). The results are shown in **Figure 5**. The correlation ($R^2 = 0.52$, **Figure 5a**) for the pine lignin prediction was considerably lower than that of the calibration set ($R^2 = 0.73$). However, the RPD was 1.49; therefore, the calibration model can still be used for initial screening (*15*, *17*, *18*). The relationship between the wet chemistry lignin values and the NIR-predicted lignin values of the aspen data set is very strong ($R^2 = 0.89$, **Figure 5b**), and the RPD is 2.58. This high RPD value indicates that this calibration model can be used successfully for screening lignin variation in aspen (*11*, *16*).

Improving the Sources of Error. The work presented in this study demonstrates that it is feasible to develop a good calibration model using stacked wood wafers. The calibration error can be reduced by some technical improvements, such as increasing the signal intensities, increasing the accuracy of the reference method, or broadening the variability of the calibration set. Albeit, the greatest source of error in any calibration is generally from the error associated with the reference laboratory data (*11*, *19*).

The wet chemistry methods involved in the determination of α-cellulose or lignin content rely on first breaking down the wood into fine wood meals. This provides a more uniform material and increases the chemical accessibility during the respective reactions. However, to facilitate rapid screening, minimal processing of the wood is required. Further, to reduce the introduction of unnecessary variation between the NIR and wet chemistry measurements, our method involves performing the wet chemistry analysis directly on the wafers used in the NIR analyses. Otherwise, these variations would cause a negative influence on the calibration model (*18*), Therefore, complete dispersion and mixing of chemicals throughout the sample is crucial, particularly for the Klason lignin measurements (*21*), where care needs to be taken to thoroughly knead and stir the sample mixtures.

In α-cellulose determination, the system is more susceptible to error because of the two-step reaction procedure utilized. First, holocellulose must be isolated from the wood wafers by acetic acid and sodium chlorite. As with the lignin analysis, care must be taken to ensure sufficient mixing and introduction of reagent chemicals. The resulting holocellulose is then further reacted with $NaOH_{(aq)}$ to produce α-cellulose. These processes involve

very frequent weighing, kneading, and stirring. In each step, care must be taken to minimize the possible sources of error.

LITERATURE CITED

(1) Kellog, R. M.; Thykeson, E. Influence of wood and fiber properties on kraft converting-paper quality. *Tappi J.* **1975**, *58*, 131−135.

(2) Michell, A. J.; Schimleck, L. R. Developing a method for the rapid assessment of pulp yield of plantation eucalypt trees beyond the year 2000. *Appita J.* **1998**, *51*, 428−432.

(3) Fengel, D.; Wegener, G. Chemical composition and analysis of wood. In *Wood Chemistry, Ultrastructure, Reactions*; de Gruyter, W., Ed.: New York, 1989; pp 26−65.

(4) Yeh, T. F.; Chang, H.-M.; Kadla, J. F. Rapid prediction of solid wood lignin content using transmittance near-infrared spectroscopy. *J Agr. Food Chem.* **2004**, *52*, 1435−1439.

(5) Sykes, R.; Li, B.; Hodge, G.; Goldfarb, B.; Kadla, J.; Chang, H.-M. Rapid prediction of wood properties of loblloly pine using transmittance near infrared spectroscopy. *Can. J. For. Res.* **2004**, manuscript accepted.

(6) Yokoyama, T.; Kadla, J. F.; Chang, H.-M. Microanalytical method for the characterization of fiber components and morphology of woody plants. *J. Agr. Food Chem.* **2002**, *50*, 1040−1044.

(7) Dence, C. W. The determination of lignin. In *Methods in Lignin Chemistry*; C. W. Dence, Ed.; Springer-Verlag: New York, 1992; pp 34−35.

(8) Foss NIRSystems. *Manual for VISION Software*: Silver Spring, MD, 2001.

(9) Gierlinger, N.; Schwanninger, M.; Hinterstoisser, B.; Wimmer, R. Rapid determination of heardwood extractives in *Larix* sp. by means of Fourier transform near infrared spectroscopy. *J. Near Infrared Spectrosc.* **2002**, *10*, 203-214.

(10) Fardim, P.; Ferreira, M. M. C.; Duran, N. Multivariate calibration for quantitative analysis of eucalypt kraft pulp by NIR spectrometry. *J Wood Chem. Technol.* **2002**, *22*, 67−81.

(11) Bailleres, H.; Davrieux, F.; Ham-Pichavant, F. Near infrared analysis as a tool for rapid screening of some major wood characteristics in a eucalyptus breeding program. *Ann. For. Sci.* **2002**, *59*, 479−490.

(12) Mark, H.; Workman, J. *Statistics in Spectroscopy*; Academic Press: San Diego, CA, 1991; pp 287−302.

(13) Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. *Chemometrics: A Practical Guide*; John Wiley and Sons: New York, 1998; p 348.

(14) Kramer, R. *Chemometric Techniques for Quantitative Analysis*; Marcel Dekker: New York, 1998; pp 99−110.

(15) Schimleck, L. R.; Doran, J. C.; Rimbawanto, A. Near infrared spectroscopy for cost-effective screening of foliar oil characteristics in a *Melaleuca cajuputi* breeding population. *J. Agr. Food Chem.* **2003**, *51*, 2433−2437.

(16) Williams, P. C.; Sobering, D. C. Comparison of commercial near infrared transmittance and reflectance instruments for the analysis of whole grains and seeds. *J. Near Infrared Spectrosc.* **1993**, *1*, 25−33.

(17) Schimleck, L. R.; Evans, R. Estimation of *Pinus radiata* D. Don tracheid morphological characteristics by near infrared spectroscopy. *Holzforschung* **2004**, *58*, 66−73.

(18) Schimleck, L. R.; David Jones, P.; Peter, G. F.; Daniels, R. F.; Clarklll, A. Nondestructive estimation of tracheid length from sections of radial wood strips by near infrared spectroscopy. *Holzforschung* **2004**, *58*, 375−381.

(19) Workman, J. NIR spectroscopy calibration basics. In *Handbook of Near-Infrared Analysis*; E. W. Ciurczak, Ed.; Marcel Dekker: New York, 1992; pp 247−280.

(20) Sykes, R.; Isik, F.; Li, B. L.; Kadla, J.; Chang, H.-M. Genetic variation of juvenile wood properties in a loblolly pine progeny test. *Tappi J.* **2003**, *2*, 3−8.

(21) Schwanninger, M.; Hinterstoisser, B. Klason lignin: Modifications to improve the precision of the standardized determination. *Holzforschung* **2002**, *56*, 161−166.

Received for review November 18, 2004. Revised manuscript received March 5, 2005. Accepted March 9, 2005. We would like to thank the US DOE (Grant DE-PS36-00GO 10480) and USDA (Award Number 2001-52104-11224) for financial support of this study.

JF0480647