JOURNAL OF
AGRICULTURAL AND FOOD CHEMISTRY
ARTICLE

Rapid Determination of Lignin Content of Straw Using Fourier Transform Mid-Infrared Spectroscopy

YUKIHIRO TAMAKI AND GIUSEPPE MAZZA*

Pacific Agri-Food Research Centre, Agriculture and Agri-Food Canada, 4200 Highway 97, Summerland, BC, Canada V0H 1Z0

To determine lignin content in triticale and wheat straws, calibration models were built using Fourier transform mid-infrared spectroscopy combined with partial least-squares regression. The best model for triticale and wheat straws was built using averaged spectra with raw spectrum in spectrum format and constant in path length as spectral pretreatments. The values of r^2, root-mean-square error of prediction (RMSEP), and residual predictive deviation (RPD) for the triticale straw model were 0.935, 0.305, and 3.89, respectively. The r^2, RMSEP, and RPD values for the wheat straw model were 0.985, 0.163, and 8.50, respectively. Both models showed good predictive ability. A model built using both triticale and wheat straws indicated that the values of r^2, RMSEP, and RPD were 0.952, 0.27, and 4.63, respectively. This model also showed good predictive ability and could predict lignin contents in triticale and wheat straws with the same high accuracy.

INTRODUCTION

In recent years, due to fossil fuels depletion and environmental problems, lignocellulosic biomass such as wood, straw, and bagasse has attracted considerable interest as a renewable energy source. It is available in considerable quantity at low cost and does not compete directly with food production (*1*, *2*). It can be utilized for biofuel and value-added biochemical production in a lignocellulosic feedstock biorefinery (*3*−*5*).

Analysis of the chemical components of lignocellulosic biomass is essential to understand its utilization potential because lignocellulose structure and composition vary greatly depending on plant species, plant tissues, production location, harvest date, and storage time (*6*, *7*). For example, information on chemical composition is needed to develop pretreatment and fermentation processes for bioethanol production and reactor design and process adjustments for the fast pyrolysis of lignocellulosic biomass (*7*). The conventional wet chemical analysis to determine chemical components is generally based on gravimetric, colorimetric, and chromatographic techniques, and these techniques are time-consuming, labor-intensive, and expensive; involve numerous complicated steps (*7*, *8*); and result in the production of hazardous waste. Biomass feedstock improvement programs and optimization of biorefinery processes would therefore benefit from the replacement of conventional wet chemical analysis with rapid and inexpensive methods for determining the physical and chemical properties of biomass (*9*).

The development of both multivariate statistical techniques and instrumental tools has attracted increasing attention as a means for achieving simple and rapid analyses to determine the concentration of specific chemical constituents (*10*). Mid- and near-infrared spectroscopies have the advantages of being rapid, noninvasive techniques that require minimal sample processing prior to the analysis and are also relatively inexpensive and environmentally benign (*10*). These infrared techniques combined with chemometric tools have been used for qualitative and quantitative analysis in various fields including food, soil, pharmaceutical, and biomedical applications (*11*−*15*). In addition, mid- and near-infrared spectroscopic techniques have been used for rapid measurement of the chemical composition of lignocellulosic biomass such as woody plants and grasses (*7*−*9*, *16*−*21*).

Infrared spectroscopy is based on the absorption by molecules of specific frequencies that are characteristic of their structure on the basis of the bond or group that vibrates. The infrared region can be divided into near ($14000-4000$ cm^{-1}), mid ($4000-400$ cm^{-1}), and far ($400-10$ cm^{-1}) zones (*22*). The near-infrared absorption spectrum is based on molecular overtones and combinations of fundamental vibrations (*23*). Mid-infrared is used to study the fundamental vibrations and associated rotational−vibrational structure (*22*). Near-infrared spectroscopy requires a minimum of sample preparation compared to mid-infrared spectroscopy, whereas the molar absorptivity is smaller and the spectrum is more complex (*23*, *24*). The mid-infrared spectrum exhibits a high degree of spectral resolution, leading to effective assignment to specific chemical entities or individual product constituents (*25*).

Publications on lignin content prediction for wood have been published using mid-infrared spectroscopy (*19*, *20*, *26*−*30*). However, to our knowledge, there is no publication on the prediction of lignin content in straw using mid-infrared spectroscopy. Here we report lignin content measurement in triticale and wheat straws using

*Author to whom correspondence should be addressed [phone (250) 494-6376; fax (250) 494-0755; e-mail giuseppe.mazza@agr.gc.ca].

Table 1. Location, Year, and Irrigation of Triticale and Wheat Straws Used for Modeling

location	year	irrigation	triticale			wheat		subtotal
			AC Ultima	Pronghorn	Tyndal	AC Andrew	Hoffman	
Lethbridge, AB	2007	not irrigated	3	3	3	3	3	15
		irrigated	3	3	2	3	3	14
	2008	not irrigated	3	3	3	3	3	15
		irrigated	1	3	1	2	3	10
Sprit River, AB	2008	not irrigated	3	3	3	3	3	15
Brandon, MB	2008	not irrigated	3	3	3	3	3	15
Normandin, QC	2008	not irrigated	3	3	3	3	3	15
Charlottetown, PEI	2008	not irrigated	3	3	3	3	3	15
		subtotal	22	24	21	23	24	total 114

Fourier transform mid-infrared spectroscopy coupled with a chemometric method, partial least-squares (PLS). An attenuated total reflectance (ATR) accessory was used for the mid-infrared measurement. The ATR technique requires little sample preparation and is one of the most versatile sampling techniques compared to transmission and diffuse reflectance techniques (*25*).

The straw samples used were collected over two production seasons, from different cultivars, and from different locations, with or without irrigation treatment. Spectral pretreatments and selection of wavenumber regions are generally performed to improve the accuracy of the calibration models used for infrared predictive techniques (*11*, *31*). For example, pretreatments are used to remove spectral variation unrelated to sample components (e.g., significant noises, baselines, and other undesirable factors) (*31*). Wavenumber region selection is used for an efficient prediction model of the response variable by reducing the data-space dimension and the identification of the relevant spectral features (*31*). The calibration models for this investigation were built using different spectral pretreatments including first or second derivatives, standard normal variate (SNV), and multiplicative signal correction (MSC). Three different wavenumber regions were used for model development, and the various models were then compared to generate the best model. In addition, the models were built using triplicate spectra or the averaged spectrum from the triplicate spectra for each sample to select the best model. This was done because the absorbance intensity varied in replicates, and the use of replicate spectra or averaged spectrum for building a calibration model has been done previously (*9*, *16*, *32*−*34*). Predictive techniques using infrared spectroscopy have been frequently developed for only one species; however, for a broad-based model, calibration models have been investigated using different biomass feedstocks including woody plants and grasses and chemically pretreated feedstocks (*7*, *16*, *18*, *21*). Therefore, for this work, calibration models using both triticale and wheat straws were also generated to develop a flexible and broad-based model for simultaneous lignin prediction of both kinds of straw.

MATERIALS AND METHODS

Plant Materials. Triticale (×*Triticosecale* Wittm. ex A. Camus.; cultivars AC Ultima, Pronghorn, and Tyndal) and wheat (*Triticum aestivum* L.; cultivars AC Andrew and Hoffman) straws were collected from various areas across Canada: Lethbridge, AB; Sprit River, AB; Brandon, MB; Normandin, QC; and Charlottetown, PEI. The straws were harvested in 2007 (Lethbridge only) and 2008. Part of the straw harvested in Lethbridge came from an irrigated field (**Table 1**).

Sample Preparation. Raw samples were milled using a Retsch SM 2000 cutting mill (Retsch GmbH, Haan, Germany) with a 2 mm discharge screen. Milled samples were sieved using a Retsch AS 200 tap sieve shaker

Table 2. Lignin Content (%, wt/wt, Oven-Dry Basis) of Triticale and Wheat Straws Measured by Wet Chemical Analysis

| straw sample	no. of samples	mean	standard deviation	minimum	maximum
triticale	67	15.52	1.02	12.98	17.69
calibration	51	15.54	0.99	12.98	17.69
validation	16	15.44	1.19	13.25	17.45
wheat	47	15.91	1.28	12.76	18.47
calibration	36	15.94	1.26	12.76	18.47
validation	11	15.81	1.39	13.49	17.82
triticale + wheat	114	15.68	1.15	12.76	18.47
calibration	87	15.71	1.12	12.76	18.47
validation	27	15.59	1.26	13.25	17.82

(Retsch GmbH, Haan, Germany) with 20-mesh (850 μm) and 80-mesh (180 μm) sieves. The fraction retained on the 80-mesh sieve was subjected to wet chemical analysis. For mid-infrared spectroscopy measurement, the fraction was ball-milled for 3 min at 30 Hz (Mix Mill MM301, Retsch GmbH) prior to analysis.

Wet Chemical Determination of Lignin. All samples were analyzed using NREL Laboratory Analytical Procedures (*35*−*37*). Extractives in samples were removed using water and ethanol extractions for 16 h each prior to hydrolysis. The extractives-free samples were hydrolyzed sequentially utilizing 72% H_2SO_4 at 30 °C for 1 h followed by 4% H_2SO_4 at 121 °C. Acid-insoluble lignin was determined gravimetrically as ash-free acid insoluble residue resulting from the hydrolyses. Acid-soluble lignin was calculated from the absorbance (Cary 50 Bio, Varian, Inc., Palo Alto, CA) at 320 nm of the liquid phase resulting from hydrolysis. An absorptivity of 30 (L g^{-1} cm^{-1}) was used to convert absorbance readings to mass values. Lignin content was determined as amount of acid-insoluble lignin and acid-soluble lignin.

Fourier Transform Mid-Infrared Measurement. All of the Fourier transform mid-infrared measurements were recorded in triplicate using a Nicolet 380 spectrometer (Thermo Fisher Scientific Inc., Madison, WI) with SMART iTR diamond attenuated total reflectance (ATR) with a 45° incident angle generating one bounce. The spectrometer was equipped with a deuterated triglycine sulfate detector scanning over the wavenumber range of 4000−650 cm^{-1} at a resolution of 4 cm^{-1}. For each spectrum, a total of 32 repetitive scans were accumulated using OMNIC 8.0 software (Thermo Fisher Scientific Inc., Madison, WI). Approximately 2−3 mg of ball-milled samples was placed on the head of the ATR crystal (2 mm diameter) and then pressed using a pressure tower to apply the same pressure. Spectra were collected in triplicate for each sample and then averaged to one spectrum.

Chemometric Analysis. All multivariate analyses of mid-infrared spectra for quantitative lignin content prediction were performed using TQ Analyst 8.0 software (Thermo Fisher Scientific Inc.). Partial least-squares (PLS) regression was used to generate models. About 75% of samples were used to develop the calibration models, and the other 25% were used for validation (**Table 2**). The samples were selected randomly but to reflect the range and mean values of the calibration and validation data.

Mathematical pretreatments of the spectrum were used prior to generating the calibration model: raw spectrum, first derivative, or second
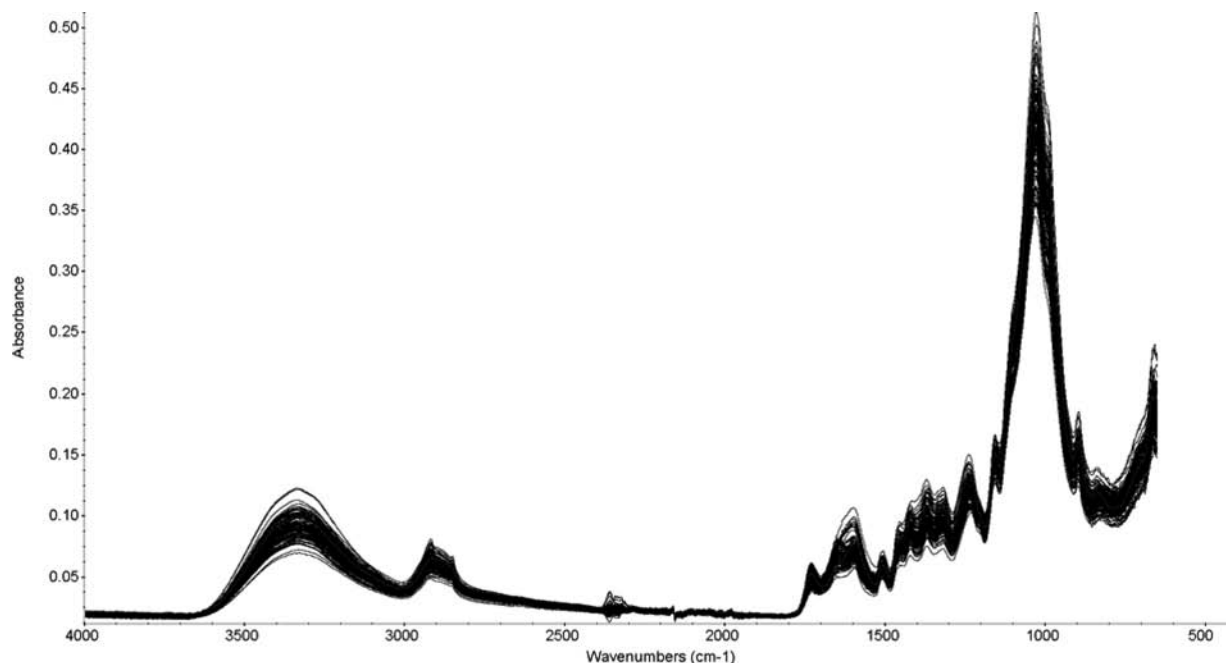
Figure 1. Averaged Fourier transform mid infrared spectra (n = 114) of triticale and wheat straws.

derivative was used in spectrum format (SF); constant, standard normal variate (SNV), or multiplicative signal correction (MSC) was used in path length (PL). MSC was used only with raw spectrum in SF because the results were similar to those of SNV. Calibration models were generated for three wavenumber regions (WR): 4000−650, 1800−700, or both 3700−2700 and 1800−700 cm^{-1}. The TQ Analyst software was allowed to determine the optimum number of PLS factors based on the predicted residual error sum of squared (PRESS) value to avoid under- or overfitting of the model.

The predictive performance of the models was evaluated by several standards such as correlation coefficient (r), root-mean-square error of calibration (RMSEC), root-mean-square error of cross-validation (RMSECV), root-mean-square error of prediction (RMSEP), coefficient of regression of predicted values against reference data (i.e., slope), coefficient of determination (r^2), standard error of performance (SEP), residual predictive deviation (RPD), R/SEP, and relative error. The following equations were used:

$$\text{RMSE (C, CV, or P)} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \quad (1)$$

$$\text{SEP} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i - \text{bias})^2}{n-1}} \quad (2)$$

$$\text{bias} = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)}{n} \quad (3)$$

$$\text{RPD} = \frac{\text{SD}}{\text{RMSEP}} \quad (4)$$

$$\text{R/SEP} = \frac{\text{data range}}{\text{SEP}} \quad (5)$$

$$\text{relative error} = \frac{\text{RMSEP}}{\text{mean}} \times 100 \quad (6)$$

y_i is the actual value by wet chemical analysis for the ith sample; $\hat{y}_i$ the predicted value by IR spectra for the same sample; n is the number of samples used in each set; SD is the standard deviation in each set; and data range is the difference between the maximum and minimum values in each set.

The model validation was performed by cross-validation and validation. In cross-validation, a single sample or subsets of samples for a calibration set are removed from the set and then predicted using a model built without those samples (*16*). A single sample was removed and predicted to obtain RMSECV (leave-one-out cross-validation). In validation, the samples for a validation set, not included in the calibration set, are predicted using the model built with calibration set samples, and then RMSEP can be obtained. These RMSECV and RMSEP values provide the average uncertainty that can be expected for predictions of future samples (*11*).

To evaluate the accuracy of the models, the slope, r^2, RPD, and R/SEP values were used. In slope, values of < 0.8 or > 1.2, 0.8−1.2, and 0.9−1.1 are assessed as less reliable, reliable, and very reliable, respectively (*13*). In r^2, a value between 0.66 and 0.80 indicates approximate quantitative predictions, whereas a value between 0.81 and 0.90 indicates good prediction. Calibration models having a value of > 0.90 are considered to be excellent (*38, 39*). In RPD, values of < 2.0 are considered to be insufficient for application, whereas values between 2.0 and 2.5 make approximate quantitative predictions possible. For values between 2.5 and 3.0, the prediction can be classified as a good prediction, and values of > 3, indicate excellent prediction (*38, 39*). In R/SEP, values of ≥4 are qualified for screening calibration, those of ≥10 are acceptable for quality control, and those of ≥15 are very good for research quantification (*7*).

RESULTS AND DISCUSSION

Wet Chemical Analysis of Lignin Content. The lignin content of the triticale and wheat straws harvested under different conditions as determined by wet chemical analysis are summarized in **Table 2**. Triticale and wheat straws had similar lignin contents of 15.52% (range, 12.98−17.69%) and 15.91% (range, 12.76−18.47%), respectively. The range of values for wheat straw was greater than that for triticale straw. The overall mean lignin content in both straws was 15.68%, with a difference of 5.71% between the minimum and maximum values.

Mid-Infrared Spectra. After the mid-infrared spectrum measurements in triplicate, the three spectra for each sample were
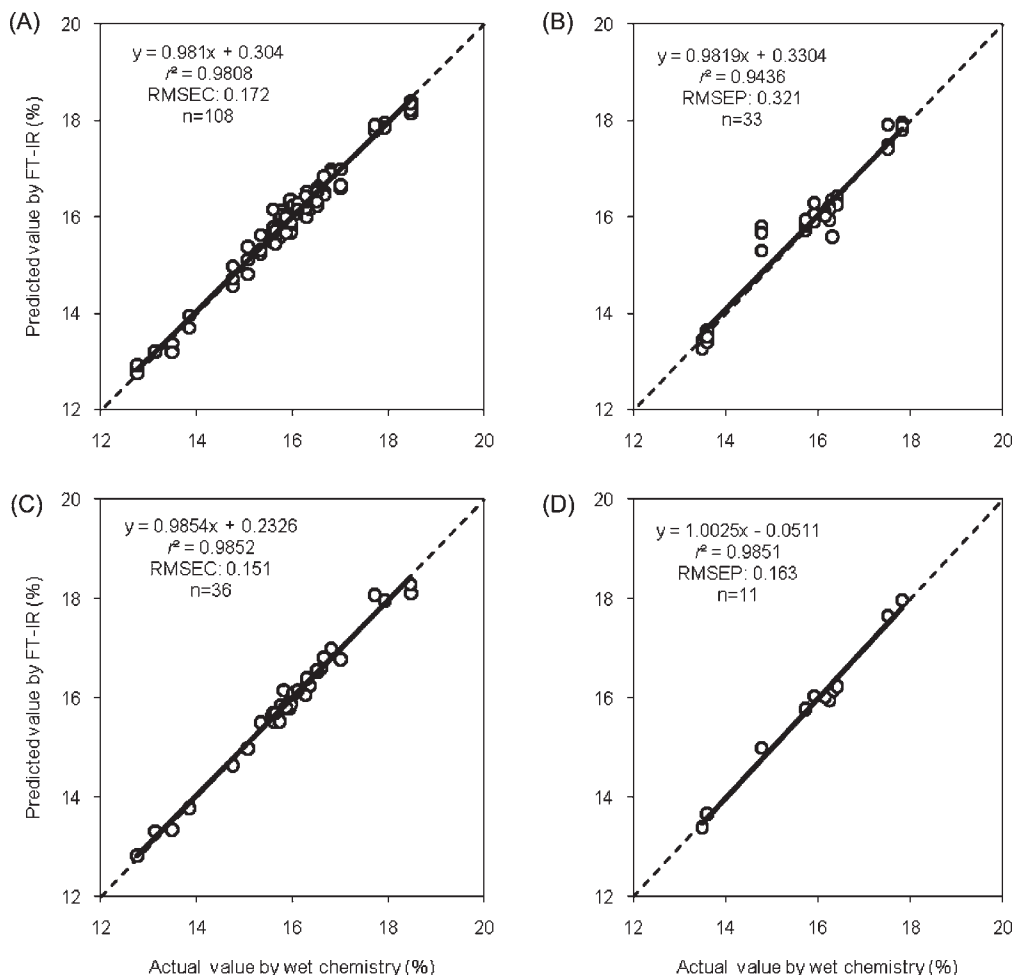
Article

J. Agric. Food Chem., Vol. 59, No. 2, 2011   **507**
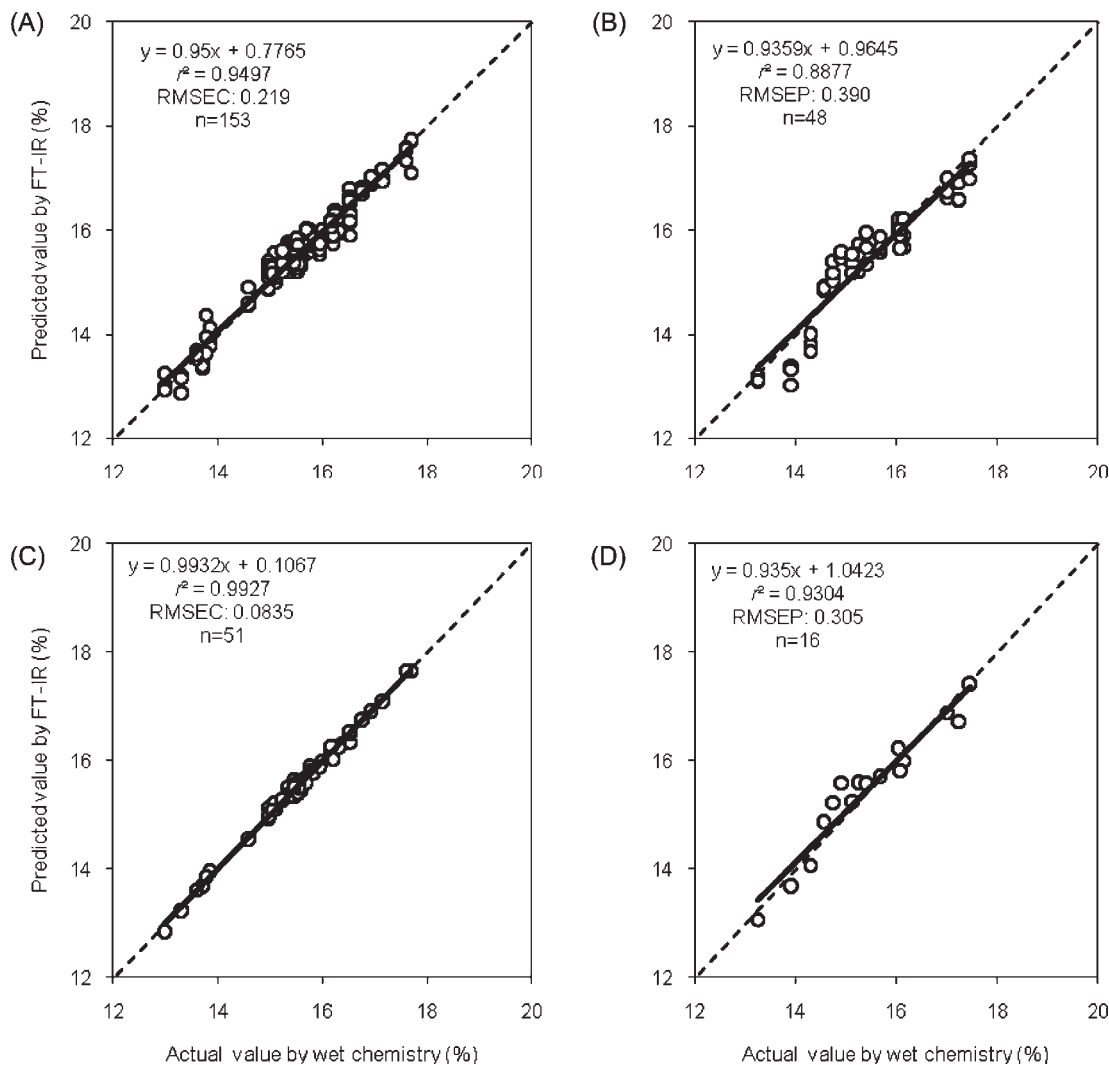
Figure 2. Plots of predicted versus actual value of lignin content (%) in triticale straw: (**A**) calibration and (**B**) validation using triplicate spectra (data format, raw spectrum; path length, constant; wavenumber region, 3700−2700 and 1800−700 cm^{-1}); (**C**) calibration and (**D**) validation using averaged spectrum (data format, raw spectrum; path length, constant; wavenumber region, 3700−2700 and 1800−700 cm^{-1}). Solid line, trend line; dashed line, X = Y.

averaged. All of the averaged infrared spectra of triticale and wheat straws are shown in **Figure 1**. All of the spectra were similar, and there was no distinguishable difference between triticale and wheat straws in the spectra. The absorbances related to functional groups of lignin were assigned for selection of the wavenumber region to use in calibration model development (*20, 27, 40−43*). Two prominent absorbances around 3300 and 2900 cm^{-1} are attributed to stretching vibrations of O─H and C─H, respectively. The absorbance at 1730 cm^{-1} is attributed to carbonyl C═O stretch of the ester linkage of the ferulic and *p*-coumeric acids of lignin. The absorbances at 1600 and 1505 cm^{-1} are attributed to the aromatic C═C stretch of aromatic rings of lignin. It is known that the spectrum assigned to isolated lignin has absorbances at 3400−2800 and 1750−800 cm^{-1} (*44, 45*). On the basis of these results, three wavenumber regions of full spectrum (4000−650, 1800−700, and both 1800−700 and 3700− 2700 cm^{-1}) were selected for generating a calibration model.

Calibration Model Development for Triticale and Wheat Straw with Various Pretreatments. PLS calibration models were built for triticale or wheat straw with triplicate or averaged spectra. Prior to building a calibration model, several mathematical pretreatments of the spectrum were applied to improve the predictive accuracy of the calibration model. The spectral pretreatments

were raw spectrum, first or second derivative in spectrum format (SF), and constant, SNV, or MSC in path length (PL). In addition, three different wavenumber regions (WR) of 4000− 650, 1800−700, or both 1800−700 and 3700−2700 cm^{-1} were applied as well. All of the results of PLS calibration models are shown in Tables S1 and S2 [see the Supporting Information (SI)] for triticale straw and Tables S3 and Table S4 (SI) for wheat straw. The selection of the best model in each data set was based on the lowest RMSEP value. **Figures 2** and **3** show the plots of predicted versus actual value of lignin content in the best model for triticale and wheat straws, respectively.

In the models for triticale straw (SI, Tables S1 and S2), the SF pretreatment showed that the RMSEP values calculated in the models built with raw spectrum were lower compared to the values with first and second derivatives. The pretreatments in PL and choices of WR did not influence RMSEP improvement significantly compared to the SF pretreatment. The best calibration models were built with raw spectrum in SF, constant in PL, and both 3700−2700 and 1800−700 cm^{-1} in WR using triplicate and averaged spectra data sets, indicating RMSEP values of 0.390 and 0.305, respectively. The r^2 value in the validation set in the best models was 0.888 in the triplicate model and 0.930 in the averaged model with slopes of 0.936 and 0.935, respectively (**Figure 2B,D**).

Figure 3. Plots of predicted versus actual value of lignin content (%) in wheat straw: (**A**) calibration and (**B**) validation using triplicate spectra (data format, raw spectrum; path length, SNV; wavenumber region, 4000−650 cm^{-1}); (**C**) calibration and (**D**) validation using averaged spectra (data format, raw spectrum; path length, constant; wavenumber region, 4000−650 cm^{-1}). Solid line, trend line; dashed line, $X = Y$.

Table 3. Performance of Selected Models in Predicting Lignin Content Using Triticale and Wheat Straws and Both Straw Samples

straw sample	pretreatment				validation						
	spectrum format	path length	wavenumber (cm^{-1})	slope[a]	$r^{2\,[b]}$	RMSEP	SEP[c]	RPD[d]	R/SEP[e]	relative error[f]
triticale	triplicate	raw	constant	3700−2700 and 1800−700	0.936	0.888	0.390	0.394	2.98	10.67	2.53
	averaged	raw	constant	3700−2700 and 1800−700	0.935	0.930	0.305	0.313	3.89	13.41	1.98
wheat	triplicate	raw	SNV	4000−650	0.982	0.944	0.321	0.323	4.18	13.40	2.03
	averaged	raw	constant	4000−650	1.003	0.985	0.163	0.171	8.50	25.36	1.03
triticale + wheat	triplicate	raw	SNV	4000−650	0.906	0.923	0.344	0.346	3.62	13.19	2.21
	averaged	raw	constant	3700−2700 and 1800−700	0.925	0.952	0.272	0.277	4.63	16.50	1.74

[a] Slope, coefficient of regression of predicted values against reference data. [b] r^2, coefficient of determination. [c] SEP, standard error of performance. [d] RPD, residual predictive deviation. [e] R/SEP, ratio of data range (R) to SEP. [f] Relative error, percentage of RMSEP to mean.

In the models for wheat straw (SI, Tables S3 and S4), the RMSEP values calculated for the models built with raw spectrum were lower than those with the other two treatments as well as those for triticale straw. The best model in the triplicate spectra data set with the lowest RMSEP value of 0.321 was built with raw spectrum in SF, SNV in PL, and 4000−650 cm^{-1} in WR, whereas the best model in the averaged data set with the lowest value of 0.163 was built with raw spectrum in SF, constant in PL, and 4000−650 cm^{-1} in WR. The values of r^2 and slope in the validation set were estimated to be 0.944 and 0.982 in the triplicate model and 0.985 and 1.003 in the averaged model (**Figure 3B,D**).

The best models for triticale and wheat straws were built with spectral pretreatments of raw spectrum in SF, of constant or SNV in PL, and of 4000−650 cm^{-1} or both 3700−2700 and 1800−700 cm^{-1} in WR. The models built with raw spectrum in SF for triticale and wheat straws indicated the lower RMSEP prediction compared to the models built with first and second derivatives. The first- and second-derivative treatments have the drawback of decreasing the signal-to-noise ratio as well as the benefit of highlighting subtle band shape (*31*). The drawback resulted in less accurate predictions for the models built with the derivative treatments. The RMSEP values did not improve substantially
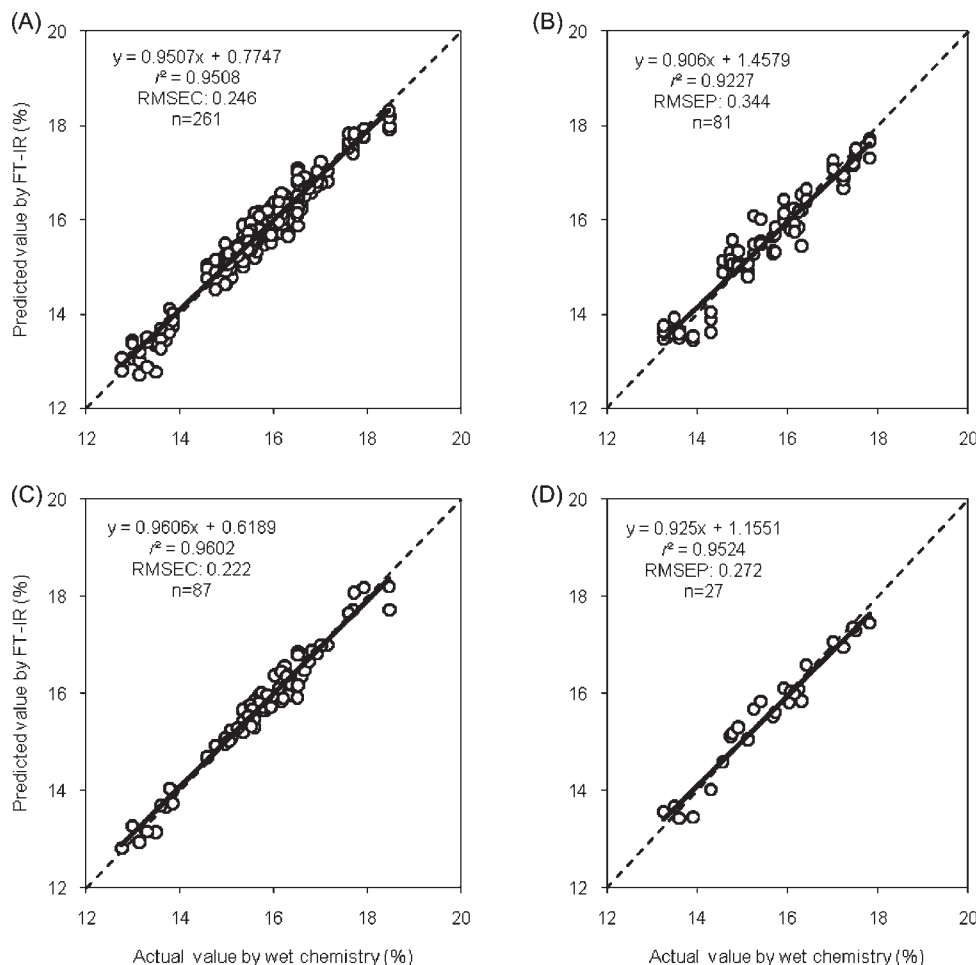
Figure 4. Plots of predicted versus actual value of lignin content (%) in both triticale and wheat straws: (**A**) calibration and (**B**) validation using triplicate spectra (data format, raw spectrum; path length, SNV; wavenumber region, $4000-650$ cm^{-1}); (**C**) calibration and (**D**) validation using averaged spectra (data format, raw spectrum; path length, constant; wavenumber region, $3700-2700$ and $1800-700$ cm^{-1}). Solid line, trend line; dashed line, $X = Y$.

with SNV or MSC pretreatments used for eliminating baseline shift and tilt effects due to light scattering and with wavenumber region selection used for choosing informative regions (*11, 31*). The averaged models predicted lower RMSEP values than the triplicate models.

Evaluation of the Best Models Selected for Triticale and Wheat Straws. Table 3 shows the accuracy of the best models selected from each set based on RMSEP. The performance evaluation was based on the values of slope, r^2, RPD, and R/SEP.

All models gave slope values between 0.9 and 1.1, indicating that the models were very reliable for prediction. Almost all models had r^2 values of >0.90, indicating that the models were excellent. The model built for triticale straw with triplicate spectra had an r^2 value of 0.888, which is still considered to be good for prediction. All models showed RPD values of >3.0, indicating that the models are classified as excellent for prediction, except for the model built using triticale straw with triplicate spectra, which is classified as good for prediction. In R/SEP, the model built using wheat straw with averaged spectra had values of ≥15, which is almost very good for research quantification. The other models had values between 10 and 15, and the prediction was found to be acceptable for quality control. Relative error values in all of the models were relatively small (<3.0%). The models for triticale and wheat straws showed excellent or good capacity for prediction.

The accuracy of the models for triticale and wheat straws was comparable to those reported previously for wood lignin prediction using mid-infrared spectroscopy based on RMSEC,

RMSECV, and RMSEP values (*19, 27, 29*). In addition, the accuracy was also comparable to the model built for wheat straw lignin prediction using near-infrared spectroscopy based on slope, r^2, and RMSEP values (*17*). These results show the practicality of rapid lignin content determination of straws using mid-infrared spectroscopy.

Calibration Model Development and Evaluation of the Selected Models Using Both Straws. In the models built using both triticale and wheat straw samples (SI, Tables S5 and S6), the RMSEP values obtained for the models built with raw spectrum were also lower than those with the first- and second-derivative treatments, as well as for the values obtained for the triticale and wheat straw models. The best model built using the triplicate spectra data set gave the lowest RMSEP value of 0.344 with raw spectrum in SF, SNV in PL, and $4000-650$ cm^{-1} in WR, and the model built using the averaged spectrum data had the lowest value of 0.272 with raw spectrum in SF, constant in PL, and both $3700-2700$ and $1800-700$ cm^{-1} in WR. Plots of predicted versus actual value of lignin content in the best model using both straws are shown in **Figure 4**.

The model using triplicate spectra showed values of 0.906, 0.923, 3.62, and 13.19 in slope, r^2, RPD, and R/SEP, respectively (**Table 3**). The model using averaged spectra showed values of 0.952, 0.952, 4.63, and 16.50 in slope, r^2, RPD, and R/SEP, respectively. The averaged spectrum model provided better predictions than the triplicate spectra model as well as the triticale and wheat straw models. The model using averaged spectra was classified as being very reliable and provided excellent prediction with four criteria: slope, r^2, RPD, and R/SEP.

(A)

(B)

(C)
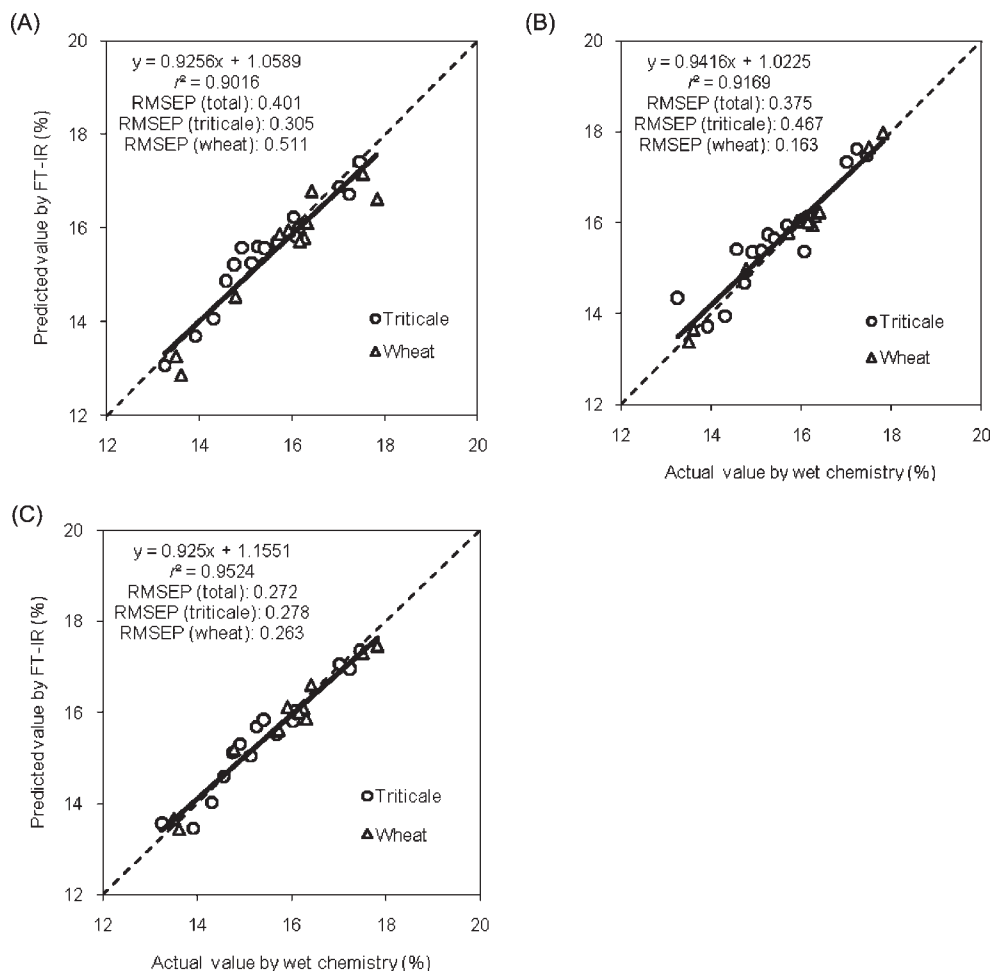
Figure 5. Comparison of prediction of triticale and wheat validation samples using (**A**) triticale straw, (**B**) wheat straw, and (**C**) both straw samples with averaged spectrum. Solid line, trend line; dashed line, *X* = *Y*.

Comparison of the Models Built Using Triticale Straw, Wheat Straw, and Both Straw Samples with Averaged Spectra. The lignin contents of triticale and wheat straws were similar, and there was no distinguishable difference between triticale and wheat straws in the mid-infrared spectra. This was expected as triticale is a hybrid crop that was developed by crossing wheat and rye. We have demonstrated here that a model developed using triticale straw samples can work for lignin content prediction of wheat samples and vice versa.

Figure 5 shows the results for the validation samples for predicting triticale and wheat lignin content on the basis of the models built using triticale straw, wheat straw, and both straw samples with averaged spectra. In the triticale straw model, the value of RMSEP for both the triticale and the wheat validation samples was 0.401, and the value for the wheat sample was 0.511. In the wheat straw model, the value of RMSEP for both validation samples was 0.375, and the value for the triticale validation sample was 0.467. These results indicate that prediction for different biomass samples not included in calibration model can be less accurate compared with the predictions for the same biomass samples included in the calibration. The model built using both straws showed that the RMSEP values for triticale and wheat were 0.278 and 0.263, respectively. The RMSEP value, 0.278 for triticale, was lower than the value of 0.305 predicted by the model using triticale straw. These results showed the potential of the broad-based model for predicting lignin content in triticale and wheat straws.

In conclusion, the best models for triticale and wheat straws were built using averaged spectra with raw spectrum in SF and constant in PL as spectral pretreatments. The values of RMSEP, slope, r^2, RPD, and R/SEP in the validation of the triticale straw model were 0.305, 0.935, 0.930, 3.89, and 13.41, respectively. The values of RMSEP, slope, r^2, RPD, and R/SEP for the wheat straw model were 0.163, 1.003, 0.985, 8.50, and 25.36, respectively. Both models showed the ability to predict excellent or good results on the basis of slope, r^2, RPD, and R/SEP. The best model using both straws for a broad-based model was built using averaged spectra with raw spectrum in SF, constant in PL, and both 3700−2700 and 1800−700 cm^{-1} in WR. The best model using both straws indicated that the values of RMSEP, slope, r^2, RPD, and R/SEP were 0.272, 0.925, 0.952, 4.63, and 16.50, respectively. The model could predict the lignin content of both triticale and wheat straw with the same accuracy. RMSEPs for triticale and wheat straw were 0.278 and 0.263, respectively. The model built using both straws also showed excellent predictive ability based on the criteria used. Mid-infrared spectroscopic techniques can be used for rapid lignin content prediction of triticale and wheat straw samples.

Article

J. Agric. Food Chem., Vol. 59, No. 2, 2011 **511**

Supporting Information Available: Six additional tables. This material is available free of charge via the Internet at http://pubs.acs.org.

LITERATURE CITED

(1) Reddy, N.; Yang, Y. Biofibers from agricultural byproducts for industrial applications. *Trends Biotechnol.* **2005**, *23*, 22−27.

(2) Hallac, B. B.; Sannigrahi, P.; Pu, Y.; Ray, M.; Murphy, R. J.; Ragauskas, A. J. Biomass characterization of *Buddleja davidii*: a potential feedstock for biofuel production. *J. Agric. Food Chem.* **2009**, *57*, 1275−1281.

(3) Ohara, H. Biorefinery. *Appl. Microbiol. Biotechnol.* **2003**, *62*, 474−477.

(4) Kamm, B.; Kamm, M. Principles of biorefineries. *Appl. Microbiol. Biotechnol.* **2004**, *64*, 137−145.

(5) Carvalheiro, F.; Duarte, L. C.; Gírio, F. M. Hemicellulose biorefineries: a review on biomass pretreatments. *J. Sci. Ind. Res.* **2008**, *67*, 849−864.

(6) Zhang, Y.-H. P.; Ding, S.-Y.; Mielenz, J. R.; Cui, J.-B.; Elander, R. T.; Laser, M.; Himmel, M. E.; McMillan, J. R.; Lynd, L. R. Fractionating recalcitrant lignocellulose at modest reaction conditions. *Biotechnol. Bioeng.* **2007**, *97*, 214−223.

(7) Liu, L.; Ye, X. P.; Womac, A. R.; Sokhansanj, S. Variability of biomass chemical composition and rapid analysis using FT-NIR techniques. *Carbohydr. Polym.* **2010**, *81*, 820−829.

(8) Jiang, W.; Han, G.; Zhang, Y.; Wang, M. Fast compositional analysis of ramie using near-infrared spectroscopy. *Carbohydr. Polym.* **2010**, *81*, 937−941.

(9) Nkansah, K.; Dawson-Andoh, B.; Slahor, J. Rapid characterization of biomass using near infrared spectroscopy coupled with multivariate data analysis: part 1 yellow-poplar (*Liriodendron tulipifera* L.). *Bioresour. Technol.* **2010**, *101*, 4570−4576.

(10) Sinelli, N.; Spinardi, A.; Di Egidio, V.; Mignani, I.; Casiraghi, E. Evaluation of quality and nutraceutical content of blueberries (*Vaccinium corymbosum L.*) by near and mid-infrared spectroscopy. *Postharvest Biol. Technol.* **2008**, *50*, 31−36.

(11) Nicolaï, B. M.; Beullens, K.; Bobelyn, E.; Peirs, A.; Saeys, W.; Theron, K. I.; Lammertyn, J. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biol. Technol.* **2007**, *46*, 99−118.

(12) Cen, H.; He, Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends Food Sci. Technol.* **2007**, *18*, 72−83.

(13) Nduwamungu, C.; Ziadi, N.; Parent, L.-É.; Tremblay, G. F.; Thuriès, L. Opportunities for, and limitations of, near infrared reflectance spectroscopy applications in soil analysis: a review. *Can. J. Soil Sci.* **2009**, *89*, 531−541.

(14) Wang, L.; Mizaikoff, B. Application of multivariate data-analysis techniques to biomedical diagnostics based on mid-infrared spectroscopy. *Anal. Bioanal. Chem.* **2008**, *391*, 1641−1654.

(15) Burns, D. A.; Ciurczak, E. W. *Handbook of Near-Infrared Analysis*; CRC Press: Boca Raton, FL, 2008.

(16) Wolfrum, E. J.; Sluiter, A. D. Improved multivariate calibration models for corn stover feedstock and dilute-acid pretreated corn stover. *Cellulose* **2009**, *16*, 567−576.

(17) Lomborg, C. J.; Thomsen, M. H.; Jensen, E. S.; Esbensen, K. H. Power plant intake quantification of wheat straw composition for 2nd generation bioethanol optimization − a near infrared spectroscopy (NIRS) feasibility study. *Bioresour. Technol.* **2010**, *101*, 1199−1205.

(18) Kelley, S. S.; Rowell, R. M.; Davis, M.; Jurich, C. K.; Ibach, R. Rapid analysis of the chemical composition of agricultural fibers using near infrared spectroscopy and pyrolysis molecular beam mass spectrometry. *Biomass Bioenergy* **2004**, *27*, 77−88.

(19) Meder, R.; Gallagher, S.; Mackie, K. L.; Böhler, H.; Meglen, R. R. Rapid determination of the chemical composition and density of *Pinus radiata* by PLS modelling of transmission and diffuse reflectance FTIR spectra. *Holzforschung* **1999**, *53*, 261−266.

(20) Ferraz, A.; Baeza, J.; Rodriguez, J.; Freer, J. Estimating the chemical composition of biodegraded pine and eucalyptus wood by DRIFT

(21) spectroscopy and multivariate analysis. *Bioresour. Technol.* **2000**, *74*, 201−212.

(21) Sanderson, M. A.; Agblevor, F.; Collins, M.; Johnson, D. K. Compositional analysis of biomass feedstocks by near infrared reflectance spectroscopy. *Biomass Bioenergy* **1996**, *11*, 365−370.

(22) Luykx, D. M. A. M.; van Ruth, S. M. An overview of analytical methods for determining the geographical origin of food products. *Food Chem.* **2008**, *107*, 897−911.

(23) Teixeira, A. P.; Oliveira, R.; Alves, P. M.; Carrondo, M. J. T. Advances in on-line monitoring and control of mammalian cell cultures: supporting the PAT initiative. *Biotechnol. Adv.* **2009**, *27*, 726−732.

(24) Gottlieb, D. M.; Schultz, J.; Bruun, S. W.; Jacobsen, S.; Søndergaard, I. Multivariate approaches in plant science. *Phytochemistry* **2004**, *65*, 1531−1548.

(25) Roychoudhury, P.; Harvey, L. M.; McNeil, B. The potential of mid infrared spectroscopy (MIRS) for real time bioprocess monitoring. *Anal. Chim. Acta* **2006**, *571*, 159−166.

(26) Friese, M. A.; Banerjee, S. Lignin determination by FT-IR. *Appl. Spectrosc.* **1992**, *46*, 246−248.

(27) Rodrigues, J.; Faix, O.; Pereira, H. Determination of lignin content of *Eucalyptus globulus* wood using FTIR spectroscopy. *Holzforschung* **1998**, *52*, 46−50.

(28) Costa e Silva, J.; Nielsen, B. H.; Rodrigues, J.; Pereira, H.; Wellendorf, H. Rapid determination of the lignin content in Sitka spruce (*Picea sitchensis* (Bong.) Carr.) wood by Fourier transform infrared spectroscopy. *Holzforschung* **1999**, *53*, 597−602.

(29) Raiskila, S.; Pulkkinen, M.; Laakso, T.; Fagerstedt, K.; Löija, M.; Mahlberg, R.; Paajanen, L.; Ritschkoff, A.-C.; Saranpää, P. FTIR spectroscopic prediction of Klason and acid soluble lignin variation in Norway spruce cutting clones. *Silva Fenn.* **2007**, *41*, 351−371.

(30) Chen, H.; Ferrari, C.; Angiuli, M.; Yao, J.; Raspi, C.; Bramanti, E. Qualitative and quantitative analysis of wood samples by Fourier transform infrared spectroscopy and multivariate analysis. *Carbohydr. Polym.* **2010**, *82*, 772−778.

(31) Moros, J.; Garrigues, S.; de la Guardia, M. Vibrational spectroscopy provides a green tool for multi-component analysis. *Trends Anal. Chem.* **2010**, *29*, 578−591.

(32) Ferreira, D.; Barros, A.; Coimbra, M. A.; Delgadillo, I. Use of FT-IR spectroscopy to follow the effect of processing in cell wall polysaccharide extracts of a sun-dried pear. *Carbohydr. Polym.* **2001**, *45*, 175−182.

(33) Ozen, B. F.; Weiss, I.; Mauer, L. J. Dietary supplement oil classification and detection of adulteration using Fourier transform infrared spectroscopy. *J. Agric. Food Chem.* **2003**, *51*, 5871−5876.

(34) Nunes, A.; Martins, J.; Barros, A. S.; Calvis-Sánchez, A. C.; Delgadillo, I. Estimation of olive oil acidity using FT-IR and partial least squares regression. *Sens. Instrum. Food Qual.* **2009**, *3*, 187−191.

(35) Sluiter, A.; Ruiz, R.; Scarlata, C.; Sluiter, J.; Templeton, D. *Determination of Extractives in Biomass*; National Renewable Energy Laboratory: Golden, CO, July 17, 2005; http://www.nrel.gov/biomass/pdfs/42619.pdf (accessed July 2010).

(36) Sluiter, A.; Hames, B.; Ruiz, R.; Scarlata, C.; Sluiter, J.; Templeton, D.; Crocker, D. *Determination of Structural Carbohydrates and Lignin in Biomass*; National Renewable Energy Laboratory: Golden, CO, April 25, 2008; http://www.nrel.gov/biomass/pdfs/42618.pdf (accessed July 2010).

(37) Tamaki, Y.; Mazza, G. Measurement of structural carbohydrates, lignins, and micro-comoponents of straw and shives: effects of extractives, particle size and crop species. *Ind. Crops Prod.* **2010**, *31*, 534−541.

(38) Zornoza, R.; Guerrero, C.; Mataix-Solera, J.; Scow, K. M.; Arcenegui, V.; Mataix-Beneyto, J. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soil. *Soil Biol. Biochem.* **2008**, *40*, 1923−1930.

(39) Saeys, W.; Mouazen, A.; Ramon, H. Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy. *Biosystems Eng.* **2005**, *91*, 393−402.

(40) Bjarnestad, S.; Dahlman, O. Chemical compositions of hardwood and softwood pulps employing photoacoustic Fourier transform

Tamaki and Mazza

infrared spectroscopy in combination with partial least-squares analysis. *Anal. Chem.* **2002**, *74*, 5851−5858.

(41) Sain, M.; Panthapulakkal, S. Bioprocess preparation of wheat straw fibers and their characterization. *Ind. Crops Prod.* **2006**, *23*, 1−8.

(42) Alemdar, A.; Sain, M. Isolation and characterization of nanofibers from agricultural residues − wheat straw and soy hulls. *Bioresour. Technol.* **2008**, *99*, 1664−1671.

(43) Kumar, R.; Mago, G.; Balan, V.; Wyman, C. E. Physical and chemical characterizations of corn stover and poplar solids resulting from leading pretreatment technologies. *Bioresour. Technol.* **2009**, *100*, 3948−3962.

(44) Buranov, A. U.; Ross, K. A.; Mazza, G. Isolation and characterization of lignins extracted from flax shives using pressurized aqueous ethanol. *Bioresour. Technol.* **2010**, *101*, 7446−7455.

(45) Yang, Q.; Wu, S.; Lou, R.; Lv, G. Structural characterization of lignin from wheat straw. *Wood Sci. Technol.* **2010**, in press.