Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biomass and Bioenergy 27 (2004) 77–88
BIOMASS & BIOENERGY
www.elsevier.com/locate/biombioe

Rapid analysis of the chemical composition of agricultural fibers using near infrared spectroscopy and pyrolysis molecular beam mass spectrometry

Stephen S. Kelley[a,*], Roger M. Rowell[b], Mark Davis[b], Cheryl K. Jurich[a], Rebecca Ibach[b]

[a] *National Bioenergy Center, National Renewable Energy Laboratory, 1617 Cole Blvd., Golden, CO 80401, USA*
[b] *USDA Forest Products Laboratory, Madison, WI, USA*

Received 17 May 2002; received in revised form 4 November 2003; accepted 11 November 2003

Abstract

The chemical composition of a variety of agricultural biomass samples was analyzed with near infrared spectroscopy and pyrolysis molecular beam mass spectroscopy. These samples were selected from a wide array of agricultural residue samples and included residues that had been subjected to a variety of different treatments including solvent extractions and chemical modifications. This analysis showed that both spectroscopic tools, coupled with multivariate analytical techniques, could be used to differentiate the samples and accurately predict the chemical composition of this disparate set of agricultural biomass samples.

Keywords: Rapid chemical analysis; Near infrared spectroscopy; Pyrolysis-molecular beam mass spectrometry

1. Introduction

Understanding the chemical composition of biomass is a key feature in determining potential uses for, and the value of, a specific biomass resource. The chemical composition and material properties of biomass can be dramatically changed by extraction with aqueous reagents or organic solvents, or by chemical modification. Traditionally, chemical analyses of the individual components (e.g., sugars, lignin) of lignocellulosics have been performed by acid hydrolysis followed by gravimetric determination of lignin and chromatographic determination of sugars [1–4]. These methods can provide highly precise data, but because they are laborious, time-consuming, and, consequently, expensive to perform, sample throughput is limited. Thus, there is a great deal of interest in developing analytical tools that can be used to rapidly, inexpensively measure the chemical composition of biomass.

There are many differences in structure, and physical and chemical properties between wood and other agricultural resources, including differences in fiber length and width, and cell wall architecture. The major physical differences are water sorption, free volume, permeability, and strength. Chemically, they differ in

* Corresponding author. Tel.: +1-303-384-6123; fax: +1-303-384-6363.
E-mail address: steve_kelley@nrel.gov (S.S. Kelley).

0961-9534/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.biombioe.2003.11.005

lignin, cellulose, and hemicellulose content as well as the actual chemical structure and make up of both the lignin and the hemicelluloses. Many of the plants contain large quantities of proteins that are sometimes mistaken for lignin in analysis. Plants also may contain large quantities of starch that can be mistaken for cellulose while most woods contain very little if any starch. The extractive type and content also vary widely between plants and wood. Finally, the inorganic content can be quite variable in both woods and plants. Because of these differences, it is difficult to use one set of analytical tools to analyze all types of agricultural resources.

A variety of spectroscopy tools have been used to measure the chemical composition of wood and other forms of biomass. In one early study, Schultz and coworkers [5] used diffuse reflectance infrared spectroscopy (DRIFT) to determine the glucose, lignin and xylose content of wood. They suggested this technique could be used to accurately predict the chemical composition of wood. Since the advent of commercially available packages capable of performing MVA techniques on complex spectral data sets, several researchers have used DRIFT and transmission Fourier transform infrared spectroscopy (FTIR) techniques to quantitatively analyze wood composition [6,7]. In one recent study, DRIFT and FTIR were used to measure the complete chemical composition, e.g., lignin, glucose, xylose, mannose, galactose, arabinose and extractives, of a series of *Pinus radiata* samples [6]. Both spectroscopic techniques produced high quality predictions of the chemical composition of the wood samples, and due to the ease of sample preparation, DRIFT was recommended as the preferred method. This work also compared the merits of using different projection to latent structures (PLS) techniques that can be used to correlate chemical composition measured by traditional techniques with the spectral data. A similar study [7] also showed that transmission FTIR could be used to accurately predict the carbohydrate composition of *Eucalyptus globulus*. This work was based on the use of pre-selected wavelengths that were known to be related to the carbohydrates of interest. Both univariate and MVA techniques were used to predict the carbohydrate content, and both methods produced high quality predictions. Prediction of the chemical composition of wood with FTIR may be viewed as the "bench-mark standard" against which any new rapid analysis technique should be compared.

While FTIR is capable of providing accurate predictions of the chemical composition of wood, it has some significant drawbacks in terms of sample preparation. Even with a relatively rapid technique such as DRIFT, the samples must be ground to a homogeneous powder and sampling conditions are critical. In contrast, near infrared (NIR) spectroscopy can be used to collect spectra rapidly on a wide variety of samples. One paper highlights DRIFT as a rapid technique at 8–10 samples/h [6] while more than one hundred samples/h may be processed by NIR. NIR spectroscopy has one additional significant advantage; since it is a non-contact measurement it can be used for process control applications [8,9]. The initial applications of NIR to forestry and the forest products industry focused on forest health and analyzing nitrogen, cellulose and lignin content of fresh and dried leaves and needles [10–15]. Taken together, these studies showed that the correlations between NIR and nitrogen content were very strong, $r > 0.95$, while the correlations between NIR spectra and the cellulose, starch or lignin content were generally very good, $r > 0.90$. McLellan and coworkers [13] concluded that there were greater errors for interlaboratory analysis of foliage samples using traditional wet chemistry, e.g., extraction/digestion and chromatography than for NIR analysis.

A number of groups have shown the value of using NIR for predicting properties of interest to the pulp and paper industry [16–29]. Their work showed that NIR could be used to predict pulp yield [17–21], the properties of paper [22,23], or the kappa number of pulps [25]. Several reports also have demonstrated the value of using NIR for directly measuring the chemical composition of wood and biomass [9,26–29]. In several of these studies the complete chemical composition of wood, e.g., lignin, glucose, xylose, mannose, galactose, arabinose and extractives, was measured with NIR.

Analytical pyrolysis techniques have been used to measure the chemical composition of many types of biomass and isolated biomass components. Early work by Evans and Milne showed that pyrolysis molecular beam mass spectrometry (py-MBMS) could be used to characterize the complex suite of reaction products produced from pyrolysis of wood and its

individual components [30,31]. This work is very useful since it provides mass spectral assignments for many of the individual fragments produced in this complex fast pyrolysis environment. The syringyl/guaiacyl ratio of lignin from different families of *E. globulus* has been studied in detail using pyrolysis gas chromatography mass spectroscopy (py-GC-MS) [32,33]. This study was able to distinguish trees with different types of lignins and the technique was recommended for screening Eucalyptus lines for pulp production.

A similar study showed that py-GC-MS could be used to study the structure of lignin [34]. Specifically this work focused on the role of cinnamyl alcohol end-groups in lignins isolated from different sources. The chemical composition of pine needles has been studied with py-GC-MS [35]. This work showed that both carbohydrate and lignin fragments could be identified using this technique. To increase the utility of py-GC-MS, a number of authors also have created derivatives in situ by conducting the pyrolysis in the presence of tetramethylammonium hydroxide or similar reagents [36–39]. This approach increases the volatility of many of the pyrolysis fragments, and can be used to distinguish between hydroxyl and acid groups that were present in the original samples and those that were formed during the pyrolysis process. However, none of this work demonstrated the use of pyrolysis mass spectroscopy techniques to measure the chemical composition of biomass.

The goal of this study was to test the effectiveness of NIR and py-MBMS for measuring the chemical composition of biomass that has been subjected to a wide variety of extractions and chemical treatments. The focus of this work is to demonstrate that these tools could be used to measure the chemical composition of a wide variety of agricultural residues, and not to optimize either the wet chemical analysis or the rapid analysis models. Additional work on the wet chemistry and rapid analysis should allow for the production of high quality, predictive models.

2. Experimental

The protocols for wet chemical analysis are based on methods developed for woody materials that are not optimized for fibers or agricultural residues. The variance in the wet chemistry is small, less than 1% for all components, but many components of these samples, acetyl groups, uronic acids, protein, are not completely captured by these analysis protocols. Samples subjected to solvent extraction or chemical modification are particularly problematic for the general wet chemical analysis. Thus, the mass closure for many of the samples is well below the normal range of 98–102%. But since the analyses are reproducible these results can still be used to demonstrate the potential for using these rapid analysis tools on a wide variety of fibers and agricultural residues.

2.1. Biomass samples

The biomass samples were obtained by the US Forest Products Laboratory from a wide variety of sources. They were air-dried and used as received. The biomass samples are listed in Table 1. The different treatments used to modify the samples are also listed in Table 1.

2.2. Chemical analysis by reference method

Samples were milled until they passed through a 1.00-mm screen and vacuum dried at 45°C. Primary hydrolysis of 80–120 mg subsamples was performed with 1.00 ml 72% (w/w) H_2SO_4 for 1 h at 30°C. Hydrolysates were diluted to 4% (w/w) H_2SO_4 with distilled water, fucose added as an internal standard, and a secondary hydrolysis performed for 1 h at 120°C. Following secondary hydrolysis, samples were immediately filtered through tared Gooch porcelain crucibles containing Whatmanr 934-AH glass fiber filters. Residues were extensively washed with hot water, dried, and measured gravimetrically. Residue ash contents were determined gravimetrically following combustion and deducted from total residue masses to yield Klason lignin contents.

Sugar contents of hydrolysates were determined by high pH anion exchange chromatography with pulsed amperometric detection (HPAEC/PAD) [40]. The chromatographic system consisted of an AS50 autosampler, a GS50 quaternary gradient high pressure pump, and an ED50 pulsed amperometric detector (Dionex Corporation, Sunnyvale, CA). Sugar separation was achieved with Carbo-Pac PA1 guard

Table 1
Sample types, treatments, and chemical compositions. (Lig—Klason Lignin, Glu—Glucan, Xyl—Xylan, Mann—Mannan, Arab—Abrabinan, Galac—Galactan, Rham—Rhaninan and Clos—Total Mass Closure.)

Biomass source	Scientific name	Treatment	Lig	Glu	Xyl	Mann	Arab	Galac	Rham	Clos
Abaca	*Musa textilis*	None	8.0	62.5	12.1	0.3	1.1	0.7	0.1	84.9
Agava	*Agave tequilana*	None	9.3	47.4	15.1	0.3	0.3	0.6	0.2	73.4
Babacu	*Orbignya phalerata*	None	32.8	29.9	20.9	0.4	0.9	0.5	0.2	86.3
Banana	*Musa species*	None	18.5	44.6	11.1	0.6	2.2	0.7	0.2	82.9
Coconut Coir	*Cocos nucifera*	None	33.8	31.5	15.9	0.2	1.3	0.6	0.2	83.9
Coconut Coir	*Cocos nucifera*	None	27.2	41.8	6.1	10.5	1.1	1.9	0.6	88.8
Coconut Coir	*Cocos nucifera*	Hot Water	26.7	44.3	6.3	10.4	0.9	1.6	0.1	90.3
Coconut Coir	*Cocos nucifera*	EtOH	25.5	43.7	6.1	11.2	1.0	1.9	0.1	89.5
Coconut Coir	*Cocos nucifera*	Ether	26.5	42.4	6.1	10.8	1.1	1.8	0.2	88.9
Coconut Coir	*Cocos nucifera*	1% NaOH	27.4	49.9	6.3	8.3	1.0	1.2	0.1	94.2
Coconut Coir	*Cocos nucifera*	HNO_3	20.7	48.8	5.0	9.8	0.1	1.2	0.1	85.7
Coconut Coir	*Cocos nucifera*	HNO_3	26.9	22.0	14.8	0.1	0	0.2	0.1	64.1
Coconut Coir	*Cocos nucifera*	ECH	31.4	24.1	9.1	0.1	0.2	0.4	0	65.3
Cotton	*Gossypium spp.*	ECH	0.2	69.7	0.3	0	0	0.1	0	70.3
Curaua	*Ananas lucidus*	None	7.5	66.4	11.6	0.1	0.5	0.5	0	86.7
Flax	*Linum usitatissimum*	None	6.9	64.3	1.8	4.9	0.6	3.4	0.6	82.7
Hemp	*Cannabis sativa*	None	5.4	70.2	1.1	5.4	0.5	1.9	0.6	85.3
Kenaf bast	*Hibiscus cannabinus*	Hot Water	10.8	51.9	12.4	1.1	1.2	0.9	0.5	78.8
Kenaf bast	*Hibiscus cannabinus*	1% NaOH	11.9	52.9	1.4	1.3	1.0	0.9	0.4	69.8
Kenaf bast	*Hibiscus cannabinus*	HNO_3	4.0	62.2	10.3	1.7	0	0.6	0.2	79.0
Kenaf bast	*Hibiscus cannabinus*	ECH	7.5	32.7	6.5	0.5	0.3	1.0	0.1	48.6
Kenaf bast	*Hibiscus cannabinus*	Phenol	11.2	49.3	11.5	1.2	1.4	1.1	0.5	76.2
Kenaf bast	*Hibiscus cannabinus*	Pyridine	9.0	42.3	9.8	1.2	1.5	1.0	0.5	65.3
Kenaf core	*Hibiscus cannabinus*	None	21.0	39.2	17.7	1.3	0.3	0.6	0.4	80.5
Kenaf core	*Hibiscus cannabinus*	Hot Water	35.2	40.0	19.2	1.2	0.2	0.5	0.4	96.7
Kenaf core	*Hibiscus cannabinus*	EtOH	19.9	35.0	18.3	1.2	0.3	0.6	0.4	75.7
Kenaf core	*Hibiscus cannabinus*	Ether	20.7	39.4	17.9	1.2	0.3	0.6	0.4	80.5
Kenaf core	*Hibiscus cannabinus*	1% NaOH	21.5	51.2	16.5	0.4	0.2	0.4	0.1	90.3
Kenaf core	*Hibiscus cannabinus*	ECH	13.7	19.4	6.3	0.3	0.1	0.3	0.1	40.2
Kenaf core	*Hibiscus cannabinus*	Phenol	20.9	40.2	18.8	1.2	0.2	0.5	0.4	82.2
Kenaf core	*Hibiscus cannabinus*	Pyridine	20.0	39.5	18	1.3	0.2	0.6	0.4	80.0
Loofa	*Luffa*	None	15.2	52.6	15.3	0.2	0.3	0.4	0.2	84.3
Palm	*Palmae spp.*	None	18.1	38.5	23.4	0.5	1.3	0.7	0.2	84.1
Sisal	*Agave sisalana*	None	6.5	60.5	15.6	0.1	0.2	0.4	0.1	83.7
Sugar cane	*Saccharum officinanum*	None	18.6	41.9	22.5	0.1	1.2	0.5	0.1	84.9
Sugar cane	*Saccharum officinanum*	Hot Water	18.9	44.3	24.3	0.1	1.8	0.5	0.1	90.0
Sugar cane	*Saccharum officinanum*	EtOH	18.5	44.0	24.1	0.1	1.9	0.5	0.1	89.2
Sugar cane	*Saccharum officinanum*	Ether	18.5	42.5	22.5	0.1	1.8	0.5	0.1	86.0
Sugar cane	*Saccharum officinanum*	1% NaOH	6.3	59.7	25.7	0	2.3	0.3	0	94.3
Sugar cane	*Saccharum officinanum*	HNO_3	14.8	53.2	17.1	0.1	0.5	0.3	0	86.0

and analytical columns (Dionex) connected in series. Sugars were eluted with distilled H_2O at a flow rate of 1.0 ml/min and a temperature of 18°C. For detection, 300 mM NaOH was added as a post-column reagent at a flow rate of ca. 0.3 ml/min. Prior to each injection, the anion exchange columns were conditioned with 240 mM $NaC_2H_3O_2$ in 400 mM NaOH, and then equilibrated with distilled H_2O. The ongoing inter-batch precision of the method is tracked by analysis of a quality assurance (QA) sample with each analytical batch. The data used herein were derived from 34 analyses of subsamples of a QA sample of loblolly pine (*Pinus taeda*, L).

2.3. Near infrared analysis

The visible/NIR spectra (500–2400 nm) were acquired with an Analytical Spectral Devices FieldSpec FR Spectrometer. The samples were illuminated with a DC light source and the reflected NIR signal was collected with a fiber optic probe. The samples were held in a 2.5 cm diameter cup. Thirty individual scans were averaged for each reflectance spectra (0.1 s/spectrum) and two spectra were taken from each sample. The reflectance spectra were converted to absorbance spectra and subjected to MVA. Due to limitations on the amount of sample (less than one gram) only 23 of the samples were analyzed with NIR.

2.4. Pyrolysis molecular beam mass spectrometry

The py-MBMS analyses were conducted using a pyrolysis furnace coupled to a free-jet molecular beam mass spectrometer (MBMS). Ground samples (20–30 mg) were pyrolyzed in the furnace that was preheated to 550°C. The molecular fragments are swept out of the furnace into the MBMS with an argon gas stream. The gas stream is expanded in a series of three vacuum chambers to quench most intermolecular collisions. A low-energy electron beam (23 eV) in the triple quadrupole mass spectrometer produces a positive ion mass spectrum. The MBMS experiment is described in detail elsewhere [30,31]. All 41 of the samples were analyzed in duplicate, except for the 1% NaOH extracted kenaf core, and the 2 N nitric acid treated kenaf core and kenaf bast where there was not enough sample for duplicate analysis. The duplicate spectra were evaluated for reproducibility and then averaged to give one py-MBMS spectrum for each sample. The single spectra of samples from the 1% NaOH extracted kenaf core, and the 2 N nitric acid treated kenaf core and kenaf bast were used directly.

2.5. Multivariate analysis (MVA)

While a complete description of MVA can be found elsewhere [41,42] the following summary describes the steps used to construct PLS models in this work. MVA was performed using The Unscrambler ® version 7.6 (CAMO, Corvallis, OR) [43]. The package has the capability to perform both principal component analysis (PCA) and projection to latent structures (PLS) (also known as partial least squares) analyses. The NIR or py-MBMS spectra for all of the samples are combined into a single data matrix (X-matrix), while the chemical composition measured by HPAEC/PAD are combined into a response matrix (Y-matrix). The PCA software is used to systematically extract (decompose) variation in the data matrix (X-matrix) while principal component regression is used to regress each response variable (Y-matrix) onto the decomposed spectra (X-matrix), and make a projection to latent structures.

The Unscrambler® can be used for PLS-2 analysis where all of the Y-variables, e.g., concentrations all six individual sugars and lignin, are projected into the X-matrix (NIR or py-MBMS spectral intensities). Both the X- and Y-matrices were mean centered variance normalized prior to performing the PLS analysis. The number of principal components (factors) used for a model was selected by observing the response of the residual Y-variance with added factors. When additional factors did not substantially decrease the residual Y-variance, the model was completed. All of the PLS models were cross-validated. Cross-validation systematically removes a single sample from the data set, constructs a model with the remaining samples and uses that model to predict the value(s) of the Y-variable(s) for the extracted sample. This process continues until each individual sample has been removed from the data set and a fully cross-validated model is constructed [41,42]. All of the results reported here are based on PLS-2 analysis.

3. Results

3.1. Sample properties

This set of biomass samples is very diverse in terms of the original source of the biomass, the secondary treatments applied to the samples, and thus the chemical composition of the samples varies widely. The source of the biomass, the type of treatment and the chemical composition of the samples are listed in Table 1. Sample sources include a very pure fiber (e.g., cotton), perennial crops (e.g., sugar cane, flax, sisal, and hemp), and tropical monocotyledons (e.g., palm and banana). Samples from sugar cane bagasse, kenaf core, kenaf bast, and coconut coir were subjected to a series of secondary treatments also listed in Table 1. These treatments included relatively mild extractions such as hot water, diethyl ether and ethanol, and more severe extractions such as NaOH, phenol and pyridine, and chemical modifications such as nitric acid and epichlorohydrin. The chemical modifications represent some of the treatments that might be used for modifying the fibers prior to incorporating them into composites. Samples from this diverse series of biomass sources and secondary treatments provided a wide range of chemical compositions (Table 1).

The quality of the PLS models is heavily dependent on the accuracy and precision of the method used to generate reference values. Any errors in reference values used for the calibration will directly reduce the quality of the PLS model and increase the standard deviation of values predicted with the PLS model. Precise and reliable results have been obtained for the agricultural lignocellulosic kenaf [4], although the analytical precision for these types of highly modified agricultural fiber samples is not well established.

Of note are the lower mass closure values for these non-wood samples. The closure values obtained for these components in wood range from 85% to 98% [40]. The range for non-treated samples examined in the present study is 73–89%. It is unlikely that the HPAEC/PAD method underestimates sugar components, as the carbohydrate polymers of agricultural materials are more amenable to acid hydrolysis than are their more crystalline counterparts in wood. The expected bias of Klason lignin values is not helpful in explaining this lower mass closure, as they are expected to be overestimated by use of the reference method with agricultural samples due to condensation of the relatively high concentrations of protein typically found in these types of samples. These low mass closure values for the non-treated biomass samples are likely due to higher levels relative to woody biomass of protein, oils, uronic acid, acetyl groups, and acid soluble lignin.

The mass closure for some of the samples subjected to the chemical treatments with nitric acid or epichlorohydrin was below 60%. These treatments result in chemical modifications of sample constituents. The lack of mass closure therefore does not imply a failing on the part of the analytical methods, but rather simply that the desired chemical modification has occurred. The constituent sugars of sample carbohydrates are no longer measured as sugars because in fact they no longer are sugars.

3.2. PCA for classification of samples

The results of PCA of the py-MBMS analysis of the set of 41 samples are shown in Fig. 1. PCA groups the samples based on their similarities and differences, and provides information on the mass spectral features (loadings) that are the basis of the chemical similarities and differences [41,42]. For example, PCA allows one to distinguish samples that have been subjected to a chemical treatment from “normal” samples or samples with a high concentration of one component from samples with a low concentration of that component.

Fig. 1a shows a plot of principal component (PC)1 and PC2. These two PCs contain 81% of the mass spectral variation, 69% for PC1 and 12% for PC2. These principal components are comprised of highly correlated mass fragments that are indicative of chemical differences between samples. There are several distinct groups of samples highlighted in Fig. 1a. There is one group of samples that is strongly positive along PC1. This group contains all the samples that were subjected to nitric acid treatment. The untreated Curava sample also is included in this group. A second set of samples is positive along PC2. These are the samples that were treated with epichlorohydrin. As one might expect, these two relatively harsh chemical treatments resulted in significant changes in the chemical composition of the samples. There is a third

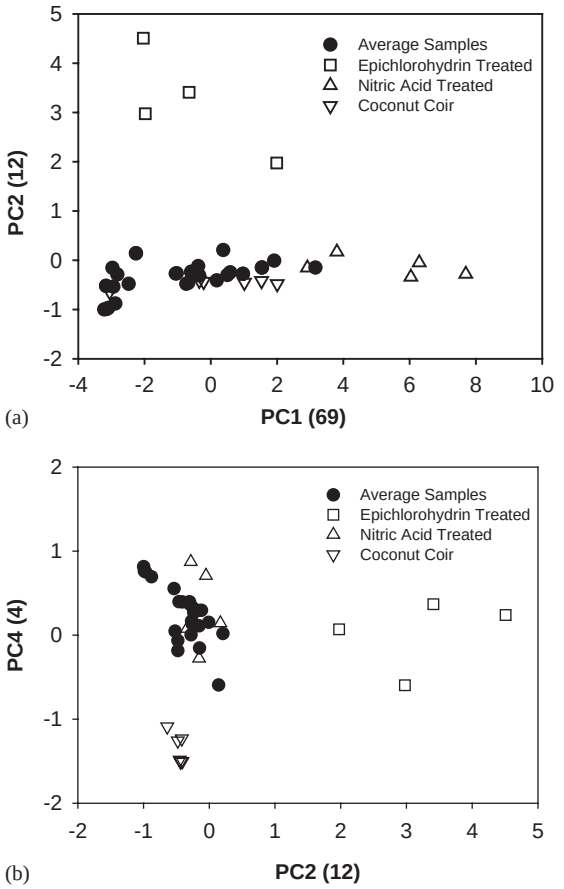

Fig. 1. Results of the principal component analysis of py-MBMS data for all on the agricultural fiber samples: (a) PC1 vs. PC2 and (b) PC2 vs. PC4.

group of samples that is slightly negative along PC1. This third set includes samples that were extracted with phenol, and a number of untreated samples.

The chemical features that distinguish the nitric acid treated samples from the remainder of the samples are related to high concentrations of carbohydrate fragments, e.g., masses 57, 60, 73, 97 and 144 [29]. The epichlorohydrin treated samples have very high mass spectral intensity at masses 58 and 86. These masses have been assigned to carbohydrate degradation products [29] that may be formed during the reaction with epichlorohydrin. The third set of samples that was slightly negative along PC1 is rich in masses 137, 154, 168, 180, and 210. All of these fragments are related to syringyl lignin moieties. The samples in this set include untreated coconut, palm, kenaf, and flax samples, and kenaf treated in different ways. Syringyl lignin is common to angiosperms but also has been found in annual plants such as flax [44].

Fig. 1b shows results from the same PCA projected along PC2 and PC4. Again the samples treated with epichlorohydrin are positive along PC2. But a second group of samples can be distinguished along PC4, which accounts for 4% of the mass spectral variation. These samples are all from one biomass source, coconut coir. The only coconut coir samples that are not contained in this group are the two samples that were treated with nitric acid. The mass spectral features that distinguish the coconut coir samples are masses 138, 154, 164 and 272. These fragments have been assigned to guaiacyl lignin fragments [30], suggesting that relative to any of the other biomass samples, coconut coir is deficient in monomethoxy aromatics.

3.3. PLS analysis for predicting chemical composition

Both the py-MBMS and NIR spectral data were used to predict the chemical composition of the samples using a PLS-2 analysis. This type of analysis simultaneously predicts a number of the individual chemical components present in the sample. In this case, the lignin content and six individual sugars, glucose, xylose, mannose, galactose, arabinose and rhamnose, were all predicted simultaneously. The results of these predictions are shown in Tables 2 and 3, for the NIR and py-MBMS sample sets, respectively. The PLS-2 models were constructed with full cross validation techniques, which allows us simultaneously to evaluate the calibration model and predict unknown samples [41,42].

Table 2 shows the results for the 23 samples subjected to analysis with NIR spectroscopy. Given the limited number of samples available for NIR analysis, the correlation coefficient (r) for the calibration model of the major components (lignin, glucose and xylose), is very high. The correlations between the measured and predicted concentration of the major biomass components are shown in Fig. 2. The correlations (r CALB) between the measured and predicted concentrations of lignin, glucose, and xylose are all above 0.85 (Table 2). The r CALB value for the measured and predicted concentrations of mannose is 0.80.

Table 2
Summary of the PLS-2 predictions of chemical composition from NIR spectra using 6 principal components

	Lignin	Glucose	Xylose	Mannose	Galactose	Arabinose	Rhamnose
r (CALB)	0.88	0.94	0.87	0.80	0.84	0.87	0.72
r (VALD)	0.71	0.87	0.71	0.44	0.52	0.57	0.36
RMSEC	4.00	3.60	3.50	2.60	0.40	0.30	0.10
RMSEP	6.10	5.20	5.80	4.50	0.70	0.50	0.20

Table 3
Summary of the PLS-2 predictions of chemical composition from py-MBMS using 6 principal components

	Lignin	Glucose	Xylose	Mannose	Galactose	Arabinose	Rhamnose
r (CALB)	0.85	0.85	0.87	0.92	0.83	0.70	0.80
r (VALD)	0.77	0.75	0.81	0.86	0.65	0.54	0.71
RMSEC	4.60	6.20	3.40	1.40	0.40	0.50	0.10
RMSEP	5.50	8.00	4.10	1.80	0.50	0.60	0.10

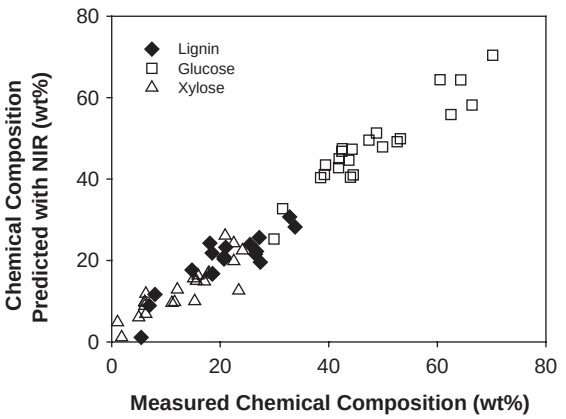

Fig. 2. Plot of the measured weight percent Glucose, Lignin and Xylose and Weight Percent Predicted from the NIR Data.

The r CALB -values for the minor components, e.g., galactose, arabinose and rhamnose, also are shown in Table 3, and as expected they are lower than those of the three major components. The lower r CALB values for these minor components are expected since these sugars are, on average, present in concentrations of less than 1 wt%.

The root mean square error of calibration (RMSEC) and root mean square error of prediction (RMSEP) are measures of the accuracy of the models in the units used for the original model. The RMSEP is similar to the standard deviation of the prediction. The RMSEP for the major components, lignin, glucose, xylose and mannose are 6.1, 5.2, 5.8 and 4.5 percent, respectively. These values are greater than the standard deviations obtained with the reference method but still allow for detection of differences between the samples. These RMSEP values also are higher than those obtained for a homogeneous set of pine samples [29], reflecting the heterogeneous nature of these agricultural fibers. However, these results suggest that NIR can be used to rapidly rank and compare samples based on composition of their major components. The NIR method is not suitable for evaluation of the minor components for this highly varied set of samples, although minor sugars have been accurately predicted for more homogeneous sets of samples [9,27–29].

The accuracy of the NIR predictions as a function of the measured concentration of the major components is shown in Fig. 3. The results in Fig. 3 show that the accuracy of the NIR predictions increase as the concentration of the component of interest increases. The accuracy of the NIR predictions at low xylose concentrations, below 10%, is poor, but the accuracy for all components improves above 10% concentration of that component.

A total of 41 samples were analyzed with py-MBMS. The results of this analysis are shown in

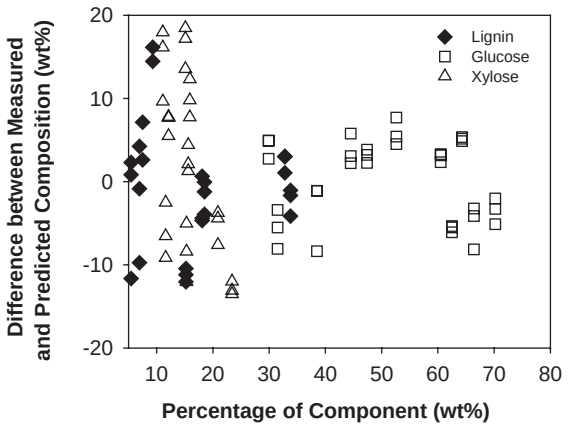

Fig. 3. Plot of the percent deviation of the predicted weight percent of Glucose, Lignin and Xylose as a function of the amount of the different components.

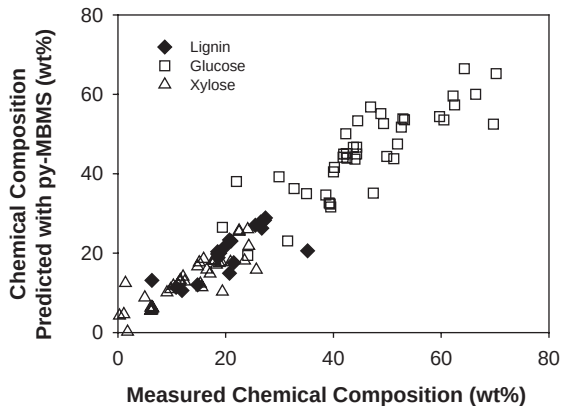

Fig. 4. Plot of the measured weight percent Glucose, Lignin and Xylose and Weight Percent predicted from the py-MBMS data.

Table 3. As seen with the NIR analysis, there is a good correlation between the measured and predicted composition for the four major components, lignin, glucose, xylose and mannose. Fig. 4 shows the correlation between the measured and predicted lignin, glucose, and xylose contents for the py-MBMS predictions. The correlations for the major components are promising with r CALB values above 0.85. Again, the r CALB values for the three minor sugars that are, on average, present at concentration below 1% are relatively low. Nevertheless, considering that the correlations are based on differences in the fragmentation patterns for sugars with very similar chemical structures, these correlations are promising. The RMSEC and RMSEP are comparable to those seen with the NIR sample set, with a slight decrease for glucose and a slight increase for mannose.

4. Discussion

Results from the PLS models produced from the NIR and py-MBMS spectra are essentially the same. This is somewhat surprising considering the distinct differences in the experiments. In the py-MBMS experiment the individual sugar isomers that make up the polysaccharides are broken into a complex suite of fragments [30,31]. The PLS models are based on correlations between the relative composition of these fragments and the concentration of the individual sugar present in the original biomass sample. Most of these fragments do not resemble the original sugar and it is not obvious how differences in their fragments would allow us to distinguish isomers such as glucose, mannose and galactose, although the results presented in Table 3 show that it is possible to distinguish these hexose isomers. The lignin macromolecule also is fragmented by the py-MBMS experiment, although many of these fragments closely resemble the compounds we would expect to find in lignin, e.g. phenol, mono and dimethoxylated phenols, and alkylated phenols [30]. Thus it is more intuitive that the concentration of methyl guaiacol, for example, might be correlated with the concentration of lignin in a sample.

In the NIR experiments, subtle differences in the structure of the different sugars and lignin cause subtle changes in hydroxyl and C–H vibrations. In the NIR experiment these sugars and lignin are intact and the relative intensity of distinct, albeit highly overlapped, vibrations are directly related to the concentration of the sugars and lignin. Even with these substantial differences in the physical phenomena underlying the spectral measurement, both the NIR and py-MBMS spectra can be used to accurately predict the concentration of lignin and six individual sugars.

In terms of the quality of the correlations, there does not seem to be an advantage of one technique relative to the other. In practical terms, the NIR is much faster (more than one hundred samples/h) and easier

to run, but the py-MBMS experiment requires very small sample sizes (as little as 5 mg), which could be a distinct advantage if one were investigating very small samples such as individual growth rings or fine roots. By way of comparison, traditional wet chemical methods (extraction, acid hydrolysis, and Klason lignin and sugars analysis) require a sample size of only approximately 100 mg, but the complete analysis takes several days.

The speed of NIR and py-MBMS could give it an analytical advantage over the much more precise, but labor-intensive reference method. Because of the limited throughput of traditional wet chemical methods, adequate replication of experimental samples is sometimes not possible. As a result, natural variation within a sample set could introduce substantial error into determination of the mean compositional value of the set. A more intensive sampling regime combined with NIR or py-MBMS analysis could be very useful in measuring the natural variation within a biomass sample.

Great care must be taken when comparing results from different studies. Many variables such as the spectral range [9], preprocessing of the data (e.g., second derivatives [27] or orthogonal signal correction [24]), and the number of principal components [41,42] used for the models will all have a significant impact on the correlations. Recognizing these issues, we have tried to put our results into context with the prior work. The correlations reported here are generally slightly poorer than those reported from other studies that predict the complete chemical composition of biomass with NIR [9,27–29]. This is not surprising given the very heterogeneous nature of this set of samples. In addition, it should be noted that non-wood tissues are compositionally more complex than wood, as is reflected by the relatively poor mass closure in the present study. If only the 14 untreated samples are analyzed, then the correlations for both the NIR and py-MBMS models (not shown) improve slightly. However, the correlations presented in all of the NIR studies are not as strong as those derived with either DRIFT or transmission FTIR. This is reasonable since the FTIR spectra contains information on the fundamental vibrations and has more separation between similar signals, e.g., hydroxyls for carbohydrate isomers. The NIR spectra are limited to the overtones of these fundamental vibrations.

Correlations from both NIR and py-MBMS are promising given the tremendous diversity of the biomass samples. The PCA of the py-MBMS results shown in Figs. 1a and b demonstrated that there were distinct differences between samples subjected to different chemical treatments and samples from different sources. Even with these distinct differences, one PLS-2 model can be used to predict the chemical composition of this very diverse array of samples. These PLS numbers can distinguish samples with low, medium and high concentrations of lignin, glucose and xylose.

5. Conclusions

The chemical composition of a very diverse array of biomass samples was predicted from NIR and py-MBMS spectra using PLS modeling. The effects of different chemical treatments could be detected with py-MBMS. Good correlations between the measured and predicted concentrations of the three major components, lignin, glucose and xylose could be obtained with both spectroscopic techniques. The correlations between the four minor sugars, mannose, galactose, arabinose and rhamnose, were weaker but promising. Improvements in the protocols used for obtaining the wet chemistry, along with more samples from a specific feedstock of interest should allow for the construction of high quality predictive models for all biomass components.

Acknowledgements

Support for this work from DOE's Office of Industrial Technologies and the USDA Forest Products Laboratory is gratefully acknowledged.

References

[1] Pettersen RC, Schwandt VH, Effland MJ. An analysis of the wood sugar assay using HPLC: a comparison with paper chromatography. J Chromatographic Sci 1984;22: 478–84.

[2] Theander O, Westerlund E. Quantitative analysis of cell wall components, Chapter 4 in forage cell wall structure and digestibility, ASA-CCA-SSSA, 1993.

[3] Carbohydrate Composition of Extractive-free Wood and Wood Pulp by Gas–liquid Chromatography. Method T 249 cm-85 in Tappi Test Methods, Tappi Press, Atlanta, GA, 1992.

[4] Davis MW. A rapid modified method for compositional carbohydrate analysis of lignocellulosics by HPAEC/PAD. J Wood Chem Technol 1998;18:235–52.

[5] Schultz T, Templeton M, McGinnis G. Rapid determination of lignocellulose by diffuse reflectance Fourier transform infrared spectroscopy. Anal Chem 1985;57:209–12.

[6] Meder R, Gallagher S, Mackie KL, Bohler H, Meglen R. Rapid determination of the chemical composition and density of *Pinus radiata* by PLS modeling of transmission and diffuse reflectance FTIR spectra. Holzforschung 1999;53:261–6.

[7] Rodrigues J, Puls J, Faix O, Pereira, Pereira H. Determination of monosaccharides composition of *Eucalyptus globulus* wood by FTIR spectroscopy. Holzforschung 2001;55: 265–9.

[8] Rantanen J, Rasanen E, Antikainen O, Mannermaa JP, Yliruusi J. In-line moisture measurement during granulation with a four-wavelength near-infrared sensor: an evaluation of process related variables and a development of non-linear calibration model. Chemo Intell Lab Syst 2001;56:51–8.

[9] Axrup L, Markides K, Nilsson T. Using miniature diode array NIR spectrometers for analyzing wood chips and bark samples in motion. J Chemo 2000;14:561–72.

[10] Newman S, Soulia M, Aber J, Dewey B, Ricca A. Analyses of forest foliage I: laboratory procedures for proximate carbon fractionation and nitrogen determination. J Near Infrared Spectros 1994;2:5–14.

[11] Martin M, Aber J. Analyses of forest foliage III: determining nitrogen, lignin and cellulose in fresh leaves using near infrared reflectance data. J Near Infrared Spec 1994;2: 25–32.

[12] McLellan T, Martin M, Aber J, Melillo J, Nadelhoffer K, Dewey B. Comparison of wet chemistry and near infrared reflectance measurements of carbon-fraction chemistry and nitrogen concentration of forest foliage. Can J For Res 1991;21:1689–93.

[13] McLellan T, Aber J, Martin M, Melillo J, Nadelhoffer K. Determination of nitrogen, lignin, and cellulose content of decomposing leaf material by near infrared reflectance spectroscopy. Can J For Res 1991;21:1684–8.

[14] Couteaux M, McTinernan K, Berg B, Szuberla D, Dardenne P, Bottner P. Chemical composition and carbon mineralization potential of Scots Pine needles at different stages of decomposition. Soil Biol Biochem 1997;30(5):583–95.

[15] Hiukka R. A multivariate approach to the analysis of pine needle samples using NIR. Chemo Intell Lab Sys 1998;44: 395–401.

[16] Wright J, Birkett M, Gambino M. Prediction of pulp yield and cellulose content from wood using near infrared reflectance spectroscopy. Tappi J 1990;73:164–6.

[17] Michell AJ. Vibrational spectroscopy—a rapid means of estimating plantation pulpwood quality. Appita J 1993;47(1):29–37.

[18] Michell AJ. Pulpwood quality estimation by near infrared spectroscopic measurements on eucalypt woods. Appita J 1995;48(6):425–8.

[19] Schimleck L, Michell A, Vinden P. Eucalpt wood classification by NIR spectroscopy and principal component analysis. Appita J 1996;49:319–24.

[20] Schimleck L, Michell A. Determination of within-tree variation of kraft pulp yield using near-infrared spectroscopy. Tappi J 1998;81(5):229–36.

[21] Raymond C, Schimleck L, Muneri A, Michell A. Nondestructive sampling of *Eucalyptus globulus* and *E. nitens* for wood properties. III. Predicted pulp yield using near infrared reflectance analysis. Wood Sci Tech 2001;35(3): 203–15.

[22] Wallbacks L, Edlund U, Norden B, Iversen T, Mohlin UB. Multivariate characterization of pulp, Part 2. Interpretation and prediction of beating processes. Nord Pulp Paper Res J 1991;6(3):104–9.

[23] Wallbacks L, Edlund U, Lindgren T, Agnemo R. Multivariate characterization of pulp, Part 3. Evaluation of physical and spectroscopic data on unbeaten and beaten kraft pulp samples. Nord Pulp Paper Res J 1995;2:88–93.

[24] Marklund M, Hauksson J, Sjostrom M. Prediction of strength parameters for softwood kraft pulps: Multivariate data analysis based on orthogonal signal correction and near infrared spectroscopy. Nord Pulp Pap Res J 1999;14(2): 140–8.

[25] Malkavaara P, Alen R. A spectroscopic method for determining lignin content of softwood and hardwood kraft pulps. Chemo Intell Lab Syst 1998;44:287–92.

[26] Sanderson MA, Agblevor F, Collins M, Johnson DK. Compositional analysis of biomass feedstocks by near infrared reflectance spectroscopy. Biomass Bioenergy 1996;11(5): 365–70.

[27] Schimleck L, Wright P, Michell A, Wallis A. Near Infrared spectra and chemical compositions of *E. globulus* and *E. nitens* plantation woods. Appita J 1997;50(1):40–6.

[28] Hiukka R. A multivariate approach to the analysis of pine needle samples using NIR. Chemo Intell Lab Syst 1998;44:395–401.

[29] Kelley SS, Rials TG, Groom L, Snell R, Sluiter A. Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood. J Wood Sci Tech 2004, in press.

[30] Evans RJ, Milne TA. Molecular characterization of the pyrolysis of biomass; 1. Fundamentals. Energy Fuels 1987;1(2):123–37.

[31] Agblevor FA, Evans RJ, Johnson KD. Molecular-beam mass spectrometric analysis of lignocellulosics materials, I Herbaceous biomass. J Anal Appl Pyrolysis 1994;30: 125–44.

[32] Rodrigues J, Meier D, Faix O, Pereira H. Determination of tree to tree variations in syringyl/guaiacyl ratio of *Eucalyptus globulus* wood lignin by analytical pyrolysis. J Anal Appl Pyrolysis 1999;48:121–8.

[33] Izumi A, Kuroda K-I. Pyrolysis-mass spectrometry analysis of dehydrogenation lignin polymers with

various syringyl/guaiacyl ratios. Rapid Comm Mass Spec 1997;11:1709–15.
[34] Kuroda K. Analytical pyrolysis products derived from cinnamyl alcohol end groups in lignin. J Anal Appl Pyrolysis 1999;53:123–34.
[35] Statheropoulos M, Liodakis S, Tzamtzis N, Pappa A, Kyriakou S. Thermal degradation of *Pinus halepnsis* pine needles using various analytical methods. J Anal Appl Pyrolysis 1997;43:115–23.
[36] Gonzalez-Vila FJ, Almendros G, del Rio JC, Martin F, Gutierrez A, Romero J. Ease of delignification assessment of wood from different *Eucalyptus* species by pyrolysis(TMAH)-GC/MS and CP/MAS ^{13}C-NMR spectrometry. J Anal Appl Pyrolysis 1999;49:295–305.
[37] Fabbri D, Helleur R. Characterization of the tetramethylammonium hydroxide thermochemolysis products of carbohydrates. J Anal Appl Pyrolysis 1999;49:277–93.
[38] Ishida Y, Wakamatsu S, Yokoi H, Ohtani H, Tsuge S. Compositional analysis of polyunsaturated fatty acid oil by one-step thermally assisted hydrolysis and methylation in the presence of trimethylsulfonium hydroxide. J Anal Appl Pyrolysis 1999;49:267–76.
[39] Hardell HL, Nilvebrant NO. A rapid method to discriminate between free and esterified fatty acids by pyrolytic methylation using tetramethylammonium acetate or hydroxide. J Anal Appl Pyrolysis 1999;52:1–14.
[40] Pettersen RC. The Chemical Composition of Wood, Chapter 2 in The Chemistry of Solid. In: Wood RR, editor. Advances in Chemistry Series 207. Washington, DC: American Chemical Society; 1984.
[41] Martens, H. Naes T. Multivariate calibration. Chichester: Wiley; 1991. p. 419.
[42] Vandeginste BGM, Massart DL, Buydens LMC, De Jong S, Lewi PJ, Smeyers-Verbeke J. Handbook of Chemometrics and Qualimetrics Part A. New York: Elsevier; 1998. p. 349–79.
[43] Anonymous, The Unscrambler 7.6, User Manual, Corvallis, OR: CAMO Inc.; 2000.
[44] Akin DE, Gamble GR, Morrison WH, Rigsby LL, Dodd RB. Chemical and structural analysis of fibre and core tissues from flax. J Sci Food Agric 1996;72:155–65.