Carbohydrate Polymers 81 (2010) 820–829

Contents lists available at ScienceDirect

Carbohydrate Polymers

journal homepage: www.elsevier.com/locate/carbpol

Variability of biomass chemical composition and rapid analysis using FT-NIR techniques

Lu Liu [a], X. Philip Ye [a,*], Alvin R. Womac [a], Shahab Sokhansanj [b]

[a] Department of Biosystems Engineering and Soil Science, The University of Tennessee, 2506 E. J. Chapman Dr., Knoxville, TN 37996, USA
[b] Department of Chemical and Biological Engineering, University of British Columbia, 2360 East Mall, Vancouver, BC, Canada V6T 1Z3

ARTICLE INFO

Article history:
Received 13 January 2010
Received in revised form 26 March 2010
Accepted 29 March 2010
Available online 3 April 2010

Keywords:
Biomass
Chemical composition
FT-NIR
Corn stover
Switchgrass
Wheat straw
Broad-based model

ABSTRACT

A quick method for analyzing the chemical composition of renewable energy biomass feedstock was developed by using Fourier transform near-infrared (FT-NIR) spectroscopy coupled with multivariate analysis. The study presents the broad-based model hypothesis that a single FT-NIR predictive model can be developed to analyze multiple types of biomass feedstock. The two most important biomass feedstocks – corn stover and switchgrass – were evaluated for the variability in their concentrations of the following components: glucan, xylan, galactan, arabinan, mannan, lignin, and ash. A hypothesis test was developed based upon these two species. Both cross-validation and independent validation results showed that the broad-based model developed is promising for future chemical prediction of both biomass species; in addition, the results also showed the method's prediction potential for wheat straw.

1. Introduction

Lignocellulosic plant biomass shows great potential as a renewable energy source both because of its widespread and ample availability and because of its environmental and economic benefits. It can be used for various industrial purposes (Demirbas, 2001; Grover, 1997), such as bio-fuel and the production of value-added bio-chemicals. However, the heterogeneous nature of biomass feedstocks—their chemical compositions vary by species, location, harvest and storage time, and even the botanic fractions—make assessment of their composition essential for bio-energy conversion processes. For example, development of the pretreatment and fermentation process for bio-ethanol production requires this chemical compositional information; reactor design and process adjustment for the fast pyrolysis of biomass need this information as well. Unfortunately, conventional wet chemistry analysis has many drawbacks for chemical composition assessment of biomass; it is time-consuming, labor-intensive, and expensive, all of which hinder at-line or online operation. By contrast, near-infrared (NIR) spectroscopy, as a fast analytical alternative, has been successfully used in the food (Hong, Ikeda, Kreft, & Yasumoto, 1996; Kim & Kays, 2009; Manley, Van Zyl, & Osborne, 2002; Wang et al., 2009) and wood industries (Kelley, Rials, Snell, Groom, & Sluiter, 2004; Michell & Schimleck, 1996; Schimleck et al., 2000) and is potentially feasible in the bio-energy industry as well.

Corn (*Zea mays* L.) stover, the stalk residue left after corn production, is a potentially important renewable energy feedstock species due to its abundance. Because corn is a primary crop product and also the feedstock for current bio-ethanol production, corn stover is a major field crop residue in the US; over 238 million tons (Sokhansanj, Turhollow, Cushman, & Cundiff, 2002) of corn stover are produced annually.

A study using dispersive NIR spectroscopy (Hames, Thomas, Sluiter, Roth, & Templeton, 2003) has shown the feasibility of NIR techniques for biomass chemical composition analysis. However, the use of Fourier transform near-infrared (FT-NIR) has several advantages over dispersive NIR, including that (1) FT-NIR measures all the frequencies simultaneously, compared to dispersive NIR, which only provides an individual single-frequency scan at a time; given a certain amount of time, FT-NIR can provide better spectral representation by averaging more scans (McCarthy & Kemeny, 2001); (2) because there is no degradation of optical throughput for FT-NIR, higher resolution is possible without compromising signal-to-noise ratio (Griffiths, De Haseth, & James, 2007); (3) FT-NIR performs internal wavelength calibration using the laser's constant wavelength, leading to better repeatability. Therefore, it would seem advisable to upgrade NIR analysis of corn stover by replacing it with FT-NIR.

* Corresponding author. Tel.: +1 865 974 7266; fax: +1 865 974 4514.
E-mail address: xye2@utk.edu (X.P. Ye).

0144-8617/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carbpol.2010.03.058

Another biomass species, switchgrass (*Panicum virgatum* L.), has garnered national attention in recent years (Boateng, Hicks, & Vogel, 2006; Christian, Riche, & Yates, 2002; Hopkins, Vogel, Moore, Johnson, & Carlson, 1995; Lemus et al., 2002; McLaughlin & Kszos, 2005). Switchgrass is one of the dominant perennial tall grass species growing on the central North American prairie (Towne & Owensby, 1984). The US Department of Energy's Bioenergy Feedstock Development Program focused on this species as an energy crop after screening more than 30 herbaceous crops species because of its high yield, excellent conservation attributes, and good compatibility with conventional farming practices (McLaughlin & Kszos, 2005). Although studies suggest that switchgrass is a promising feedstock for the bio-fuel industry, information regarding switchgrass is still being developed, especially the compositional information associated with the various cultivars and their botanic fractions (mainly nodes, internodes, and leaves). Specific NIR studies on this species have been lacking, which is unfortunate given switchgrass's promise in the bio-fuel industry and given the advantages of using NIR for this application. The significance of these two biomass species to the US and the domestic growing interest in biomass conversion to bio-fuel justify choosing them for FT-NIR investigation.

To achieve reliable and accurate prediction of chemical composition using FT-NIR spectra, a model that correlates the NIR spectra with the biomass chemical composition should be developed based on a calibration with a large variability. Several studies (Lestander & Rhen, 2005; Montross & Crofcheck, 2004) have provided evidence for compositional variation among plant botanic fractions, and our previous research (Ye et al., 2008) has proved that wide variability exists among the botanic parts of corn stover. Our approach efficiently replaces the arduous and costly process of sample collection over different times and from different locations. In this study, botanic fractions were utilized to increase the calibration variability of corn stover and switchgrass, respectively.

Promising results in the previous FT-NIR study of corn stover composition (Ye et al., 2008) have led to increased efforts to develop a rapid FT-NIR model for switchgrass chemical compositional analysis, as well. Our hypothesis is that the chemical composition of multiple biomass feedstocks can be predicted from one single broad-based FT-NIR model instead of compositional prediction having to be conducted via each individual model. NIR predictive methods have been much more frequently developed for one species than for a variety of species, as can be seen throughout the literature. For example, animal feeds, lucernes (Marten, Brink, Buxton, Halgerson, & Hornstein, 1984), and legumes (Smith & Flinn, 1991) were investigated individually through different studies; woods, pulpwood (Schimleck et al., 2000), eucalyptus (Michell & Schimleck, 1996) and Norway spruce (Lestander & Rhen, 2005) were also separately studied. In studies of corn stover, great efforts have been made in the past to increase the variability for the calibration (Albanell, Plaixats, Ferret, Bosch, & Casanas, 1995; Hames et al., 2003; Liu & Chen, 2007); samples were collected over years, from different locations, and from different cultivars. In fact, most researchers appear to think that NIR model calibration should be conducted respectively for each species, as can be seen by their practice of modeling individual species while enlarging the sample variability within each species. Although Sanderson et al. (1996) used NIR to predict the chemical composition of a broad range of biomass feedstocks, including herbaceous and woody species; the study showed the feasibility of NIR modeling over multiple species. But in the end, the researchers suggested calibrating for narrower populations, such as for herbaceous or woody species alone, or for a single species, to improve the accuracy. Kelley, Rowell, Davis, Jurich & Ibach (2004) stated that due to many differences in structure and differences in the physical and chemical properties between wood and agricultural species, including differences in fiber length and width, cell-wall architecture, free volume, permeability, and strength, it is difficult to analyze both types of species using a single NIR model. However, they suggested, it is rational to investigate the feasibility and reliability of a broad-based model for agricultural species or for wood species individually. While spectral pretreatment algorithms are able to remove some of the physical interferences caused by the dissimilarity of physical properties and to expose chemical differences, the capability of spectral pretreatment is limited. For example, it is not practical to develop an NIR predictive model to quantify lipid content in such dissimilar species as oats and tomatoes, even though the lipid variation can be created in this way.

The objectives of this study are to investigate the chemical compositional variability of corn stover and switchgrass, and also to examine the potential of the quantitative prediction of these two biomass varieties using FT-NIR techniques coupled with chemometrics. A broad-based FT-NIR predictive model for varieties of agricultural biomass species is hypothesized, and work on the hypothesis test for these two important biomass species is begun. A third species, wheat straw, was utilized as an attempt to test the generality of the broad-based model. Glucan, xylan, galactan, arabinan, mannan, lignin and ash were identified as the target analytes to be quantified and were chosen either for their industrial value (mainly as sugars for bio-ethanol production and lignin for co-firing or aromatics chemical production) or for their detrimental qualities (e.g., ash's fouling problems in combustion).

2. Materials and methods

Corn stover samples of the cultivar DeKalb DK64-10RR were collected from East Tennessee Research and Education Center in May 2006. Manually separated botanical fractions (nodes, piths, rinds, sheaths, husks and leaves) of corn stover were included to introduce large variability (Fig. 1), as our previous study (Ye et al., 2008) had verified our assumption that different amounts of glucose could be released from different plant botanic fractions. The investigation of botanic fractions is both meaningful and potentially beneficial to industry, since the aerodynamic partition of these botanic parts has been proved to be feasible (Klasek, 2006). Switchgrass samples were collected in August 2006 from the same site. Both upland (Cave-in-Rock and Shelter) and lowland (Alamo, Kanlow, NC1-16, and NC2-16) ecotype cultivars were covered in this study to make the variability investigation more representative. In addition to the investigation of the whole stalks of the six switchgrass cultivars, the Alamo and Kanlow cultivars were investigated by their botanical fractionations (Fig. 1). These two cultivars are particularly good energy feedstock choices due to their high biomass dry matter yield (Porter, 1966) and good stability over time (Fuentes & Taliaferro, 2002). Table 1 shows the data collection scheme, and the numbers show the replications performed for each variety. All of the samples shown in the calibration dataset in Table 1 (71 samples in total) were used to develop the broad-based model.

Table 1 also lists the dataset for the independent validation of the broad-based model. The five corn stover and switchgrass samples used for validation were not included in the calibration, having been collected either at different times or from different sites, thus providing variations to represent future prediction. Furthermore, five wheat-straw samples (Table 1), harvested from the East Tennessee Research and Education Center in the summer of 2005, were collected and analyzed to investigate the potential of using the broad-based model we developed to predict the chemical composition of a third species (wheat straw). Wheat straw was selected for this purpose because of the similarity of its chemical composition to that of corn stover (see Department of Energy database, http://www1.eere.energy.gov/biomass/feedstock_databases.html).

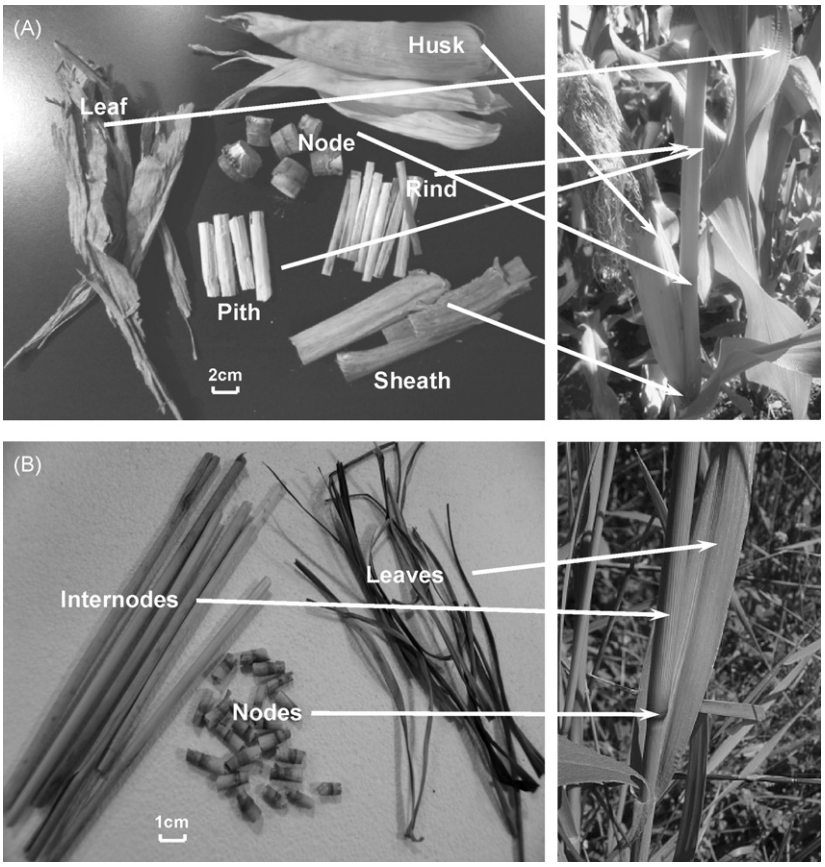

Fig. 1. Manually separated botanical fractions of corn stover (A) and switchgrass (B).

A relatively large amount (several pounds) of each biomass variety was cut into less than 1×1 by 0.5 cm pieces, well mixed, and was dried at 105 °C using the primary oven-dry method, as defined in ASTM D4442 (ASTM, 1996) until a constant weight was achieved. Samples (according to Table 1) were then randomly selected, and were ground with a Wiley Mini-Mill (Thomas Scientific, Swedesboro, NJ) and allowed to pass through a #40 mesh screen (~425 μm). The samples were then cooled to room temperature in a desiccator before the FT-NIR spectra were acquired.

The FT-NIR diffuse reflectance spectrum was collected through an FT-NIR spectrometer (Excalibur 3100, Varian Inc. Palo Alto, CA) equipped with a high-speed, low-noise, indium–gallium–arsenide (InGaAs) detector. A light compartment IntegratIR™ Accessory (PIKE Technologies, Madison, WI) was installed for the diffuse reflectance. Spectra from 4000 cm^{-1} to 10,000 cm^{-1} were collected at a resolution of 8 cm^{-1}. Each spectrum was generated by averaging 64 scans. Immediately after FT-NIR spectral acquisition, the samples were subjected to wet chemistry analysis for the measurement of glucan, xylan, galactan, arabinan, mannan, lignin, and ash. The wet chemistry procedures are summarized in Fig. 2; the experimental procedure was detailed in our previous publication (Ye et al., 2008). Briefly, the samples were subjected to two-step acid hydrolysis, and were fractionated into monomeric units of polysaccharides, which were readily to be quantified by HPLC using

Table 1
Description of the experimental samples including dataset for both calibration and validation. The listed numbers are the number of replications.

Calibration dataset						
Switchgrass (36 samples in total) from East Tennessee Research and Education Center, August 2006						
	Cave-in-rock	Alamo	Kanlow	Shelter	NC1-16	NC2-16
Whole stalk	3	3	3	3	3	3
Leaf	–	3	3	–	–	–
Node	–	3	3	–	–	–
Internode	–	3	3	–	–	–
Corn stover (35 samples in total) From East Tennessee Research and Education Center, May 2006						
Whole	Husk	Sheath	Leaf	Node	Rind	Pith
5	5	5	5	5	5	5
Independent validation dataset						
Corn Stover East Tennessee Research and Education Center, summer 2006	5	Switchgrass Research and Education Center at Milan, TN, summer 2005	5	Wheat straw East Tennessee Research and Education Center, summer 2006	5	

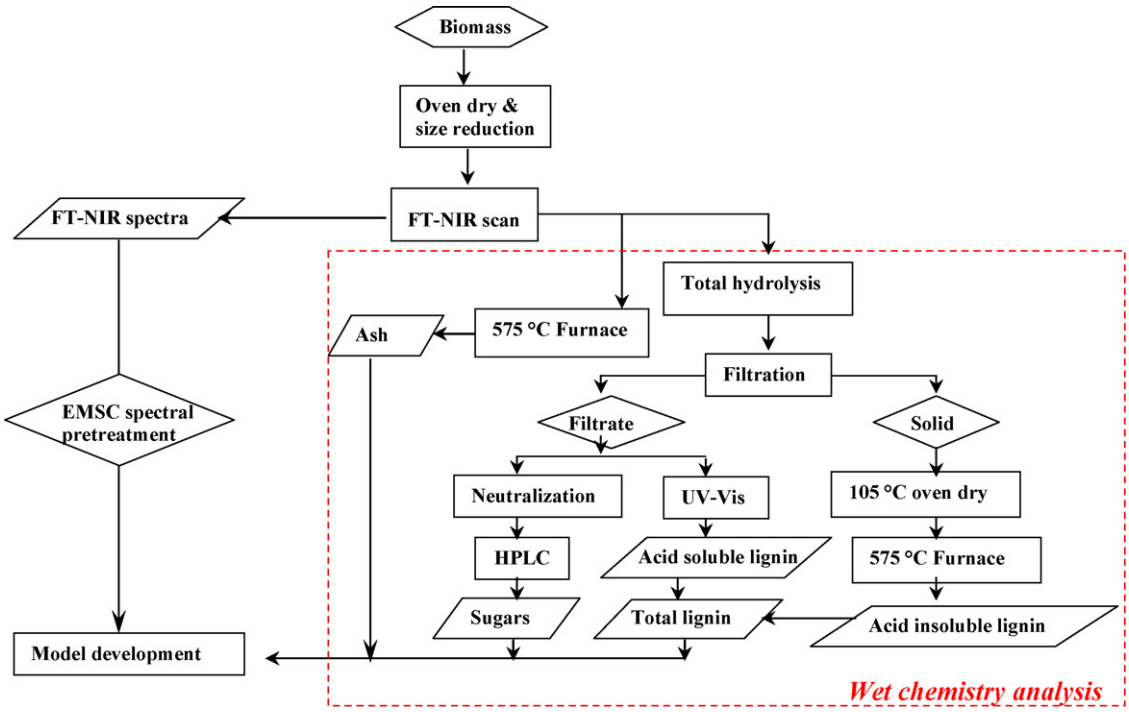

Fig. 2. Schematic representation of experimental procedures.

a Biorad Aminex HPX-87P column and a Waters 410 differential refractive index detector. Anhydro correction was applied to the quantification results of monomeric sugars to calculate the quantities of glucan, xylan, galactan, arabinan, and mannan. The total lignin content was the sum of acid soluble lignin (ASL) and acid insoluble lignin (AIL), where ASL was measured by a Shimadzu UV-1700 UV–vis spectrophotometer, and AIL was calculated as the weight loss of the dried solid hydrolysate before and after calcination in a 575 °C furnace. Ash content was measured according to the ASTM protocol (ASTM, 2002). All the results of chemical composition were reported on the 105 °C oven-dry weight basis.

Upon the acquisition of the complete dataset for both chemical and FT-NIR spectral information, a partial least squares regression (PLS) was used to correlate the seven chemical constituents (in percent weight based on dry biomass sample, %w/w) with the spectra after extended multiplicative signal correction (EMSC). EMSC was previously determined to be the most efficient spectral pretreatment method to remove the spectral variation due to physical interferences. The number of latent variables in the calibration model was determined based on the occurrence of the minimal cross-validation residual variance to prevent the over-fitting problem (Shao, 1993; Wold, 1978). The entire procedure is illustrated in Fig. 2, and all the spectral calculations were performed in the Unscrambler 9.6 (Camo Software Inc., Woodbridge, NJ).

The model validation in this study included both leave-one-out cross-validation (referred to hereinafter as "cross-validation") and independent validation. Cross-validation was employed to evaluate both of the two single models for corn stover and switchgrass, respectively. The broad-based model was first evaluated by cross-validation, and then by the independent validation using a new sample dataset, denoted as the independent validation set in Table 1. Cross-validation technique is briefly described as followings. For a dataset of n samples, one sample is left out and the rest of $(n-1)$ samples (the subset) are calibrated using PLS to generate a subset model. The left out sample is then used to validate the subset model, and the prediction residual is calculated. This procedure is repeated until each of the n samples has been left out once. As a result, n subset calibration models are generated afterwards, as well as n sets of residual variance. Finally, an overall model is to be generated combining the n sets of the subset models, and also all the n sets of residual variance is to be combined to calculate overall validation residual variance and root mean square error for prediction. Cross-validation is a well-accepted conservative method for model validation when there are limited numbers of data in the calibration set (Ben-Dor et al., 2000; Curda & Kukackova, 2004; Goodchild, El Haramein, El Moneim, Makka, & Williams, 1998; Hames et al., 2003; Hildrum, Nilsen, Mielnik, & Næs, 1996; Martens & Dardenne, 1998; Meuret, Dardenne, Biston, & Poty, 1993; Sivakesava & Irudayaraj, 2001; Togersen, Isaksson, Nilsen, Bakker, & Hildrum, 1999; Westa & Martens, 2000). As for the independent validation, the developed broad-based model was applied to the spectra of the independent validation set, and the predicted values were compared with their wet chemical measurements to evaluate the model's predictive performance.

3. Results and discussion

3.1. Wet chemistry results

Table 2 presents the chemical compositions of switchgrass and corn stover as measured by means of wet chemistry analysis. The result shows the mean and the standard deviation of the replicates measured for each variety. Corn stover was observed to have the following variations among its botanic fractions: 13.31% in glucan, 13.48% in xylan, 3.91% in arabinan, 1.89% in galactan, 2.24% in mannan, 14.26% in lignin, and 7.61% in ash. This degree of variation was observed within one corn stover cultivar, and it was larger than the variation in the switchgrass data for six cultivars and three botanic parts. The variations for the switchgrass dataset were as follows: 8.42% in glucan, 6.03% in xylan, 2.60% in arabinan, 1.87% in galactan, 0.89% in mannan, 4.78% in lignin, and 3.79% in ash. These observations suggest that as a biomass species, the chemical composition of switchgrass is more consistent than that of corn stover. Also, the variation of the botanic fractions in one corn stover cultivar was observed to be much greater than that of any one culti-

Table 2
Chemical composition of switchgrass and corn stover measured via wet chemistry analysis in terms of essential components for biomass conversion processes.

Switchgrass	Cave-in-rock	Alamo	Kanlow	Shelter	NC1-16	NC2-16	Leaf		Node		Internode	
							Alamo	Kanlow	Alamo	Kanlow	Alamo	Kanlow
Glucan	41.56[a] ± 0.32[b]	36.37 ± 0.25	36.31 ± 0.28	37.37 ± 0.35	34.52 ± 0.62	40.24 ± 0.46	34.73 ± 0.42	36.40 ± 0.08	35.18 ± 0.38	34.02 ± 0.20	40.98 ± 0.14	39.11 ± 0.12
Xylan	17.29 ± 0.14	20.15 ± 0.64	18.22 ± 0.05	20.91 ± 0.18	19.92 ± 0.84	19.04 ± 0.61	17.23 ± 0.20	21.10 ± 0.18	22.76 ± 0.15	22.61 ± 0.44	21.43 ± 0.23	22.67 ± 0.19
Galactan	1.95 ± 0.05	3.13 ± 0.27	3.01 ± 0.16	2.80 ± 0.10	2.54 ± 0.21	2.35 ± 0.10	2.21 ± 0.59	1.68 ± 0.11	2.30 ± 0.38	1.93 ± 0.26	1.93 ± 0.12	1.81 ± 0.12
Arabinan	3.21 ± 0.50	3.69 ± 0.32	3.62 ± 0.28	4.19 ± 0.10	3.19 ± 0.11	2.99 ± 0.04	3.74 ± 0.80	2.95 ± 0.02	4.75 ± 0.46	4.04 ± 0.20	2.79 ± 0.09	3.17 ± 0.03
Mannan	0.85 ± 0.12	0.56 ± 0.07	0.56 ± 0.05	0.87 ± 0.07	1.15 ± 0.28	1.00 ± 0.20	0.82 ± 0.19	0.96 ± 0.10	0.90 ± 0.02	0.60 ± 0.04	0.63 ± 0.10	1.03 ± 0.10
Lignin	21.73 ± 0.57	22.89 ± 0.37	20.64 ± 0.01	21.91 ± 0.25	22.83 ± 0.55	21.49 ± 0.57	24.43 ± 0.33	25.10 ± 0.13	21.43 ± 0.12	22.18 ± 0.24	21.43 ± 0.21	21.36 ± 0.12
Ash	2.62 ± 0.04	3.90 ± 0.19	3.21 ± 0.07	2.64 ± 0.25	3.29 ± 0.23	3.26 ± 0.16	4.90 ± 0.08	3.74 ± 0.01	2.62 ± 0.12	1.49 ± 0.28	2.19 ± 0.44	1.60 ± 0.07
Closure[c]	89.20 ± 0.85	90.68 ± 0.30	85.56 ± 0.15	90.69 ± 0.47	87.43 ± 0.90	90.37 ± 0.85	89.66 ± 0.66	91.93 ± 0.11	89.94 ± 0.46	86.88 ± 1.08	91.40 ± 0.30	90.75 ± 0.18

Corn stover	Whole	Husk	Sheath	Leaf	Node	Rind	Pith	NREL-1[d]	NREL-2	NREL-3	NREL-4
Glucan	33.18 ± 1.71	37.60 ± 2.58	39.63 ± 0.77	30.78 ± 1.88	29.15 ± 1.01	37.83 ± .57	39.03 ± 1.34	35.42	37.00	34.34	30.97
Xylan	18.94 ± 1.72	22.23 ± 4.29	19.44 ± 0.83	16.11 ± 1.13	15.93 ± 0.93	16.64 ± 1.16	17.10 ± 0.78	22.66	23.35	22.38	20.42
Galactan	2.17 ± 0.47	2.45 ± 0.28	2.05 ± 0.44	2.24 ± 0.51	2.17 ± 0.23	1.52 ± 0.31	1.67 ± 0.18	0.88	0.66	0.85	0.92
Arabinan	3.13 ± 0.96	4.77 ± 0.63	3.96 ± 0.56	3.10 ± 0.24	3.86 ± 0.19	2.32 ± 0.51	2.90 ± 0.33	2.67	2.3	2.7	2.75
Mannan	1.12 ± 0.37	1.59 ± 0.92	1.28 ± 0.43	1.19 ± 0.09	1.01 ± 0.15	1.06 ± 0.28	0.84 ± 0.28	0.33	0.31	0.27	0.29
Lignin	22.1 ± 2.25	16.14 ± 1.79	16.31 ± 6.02	23.95 ± 0.46	23.6 ± 0.41	22.68 ± 1.71	19.88 ± 0.63	17.12	18.25	17.35	17.56
Ash	3.37 ± 0.38	2.42 ± 1.19	5.42 ± 1.15	7.39 ± 1.87	3.70 ± 0.28	3.8 ± 0.46	4.86 ± 1.00	5.60	2.77	5.74	5.76
Closure	84.01 ± 3.61	87.20 ± 1.79	88.09 ± 5.14	84.77 ± 2.85	79.42 ± 2.18	85.85 ± 2.19	86.28 ± 2.30	84.68	84.64	77.89	78.67

[a] Mean (%w/w).
[b] Standard deviation (%w/w).
[c] Some less significant components, such as extractives, uronic acid, proteins, and acetyl compounds, were not measured in this study; this accounts for the mass closure less than 100.
[d] NREL-1–4 are the reference data available at US Department of Energy website [http://www.afdc.energy.gov/biomass/progs/search1.cgi]. NREL-1 is a set of data for Alamo switchgrass; NREL-2 is a set of Kanlow switchgrass; and NREL-3 and NREL-4 denote two sets of corn stover references.

var of switchgrass, and this result again suggests that the chemical composition of the corn stover is more heterogeneous than that of the switchgrass. As bio-fuel feedstock, switchgrass showed advantages over corn stover: a larger proportion of sugar content was observed in the switchgrass samples than the corn stover samples, indicating the greater potential of switchgrass in fermentation for a higher bio-ethanol yield; switchgrass also was seen to have a lower proportion of ash, portending fewer operational problems such as slagging and fouling of boilers.

A large variability was observed among the different corn stover botanic fractions, proving that manual separation can efficiently provide the large variability needed for chemometric analyses. The corn husks were determined to be the most valuable botanic fraction for bio-ethanol production, since (1) the overall sugar content of corn husks is comparatively high, and previous studies (Bothast, Saha, Flosenzier, & Ingram, 1994; Van Zyl, Prior, & Du Preez, 1988) have proved that both cellulose and hemicellulose can be converted to ethanol and (2) cornhusk also has the lowest lignin and ash content among the botanic fractions. Compared to corn leaves, nodes, and internodes, cornhusk is much easier to separate (Kracl, 1986); in addition, corn husks are routinely separated from the corn when it is prepared for use as food. Thus, it is both promising and meaningful that cornhusk is well adapted for bio-ethanol production.

The glucan content was significantly higher ($p < 0.05$) in the pith, sheath, and rind (all components of internodes), where it ranged from 37.83 to 39.63%w/w, than in the nodes and leaves, where it ranged from 29.15 to 30.78%w/w. Our result of finding a higher glucan content in the internodes than in the other botanic fractions is consistent with the results of Duguid et al. (2007), indicating that the internodes are favorable for ethanol production as well. Our results show that the leaves have the highest ash content; therefore, although they have a high lignin proportion, the leaves are not ideal for co-firing due to concerns about fouling.

Among the six switchgrass cultivars, Cave-in-Rock had the highest glucan content (41.56 ± .32%w/w), followed by NC2-16 (40.24 ± .46%w/w), while NC1-16 had the lowest glucan content (34.52 ± .62%w/w). Among the three botanic fractions, the internodes were clearly observed to have a higher glucan content than the nodes or leaves, while the leaves were observed to have a higher lignin and ash content than the other two botanic fractions. The Alamo and Kanlow cultivars were not outstanding in terms of glucan content, although they had been reported as top-yield cultivars and recommended for planting (Lemus et al., 2002). Therefore, we conclude that, in terms of overall efficiency for bio-fuel production, the proportion of sugar content should be taken into account in making cultivar choices. NC2-16 was observed to have high glucan and low ash and lignin content, and its robust lowland ecotype suggests good dry matter yield, compared to the smaller, less robust, upland ecotype cultivars (Porter, 1966; Sladden, Bransby, & Aiken, 1991). Given that there have been only limited studies on this cultivar, further research is recommended on NC2-16. The internodes of switchgrass had higher glucan content than the nodes or leaves, consistent with the results for corn stover, and the same trend was also found in wheat straw (Duguid et al., 2007).

3.2. FT-NIR spectroscopic modeling

3.2.1. Development of separate models for corn stover and switchgrass

The NIR spectra arise from the overtones and combination bands of the chemical bonds (O–H, C–H, C–C, etc.) that compose the chemical constituents of biomass. This is the rationale for correlating the spectra with the chemical quantities. In contrast to mid-infrared spectra, however, the peaks in NIR do not serve as "fingerprints" for certain bonds or certain compounds; instead, the signal is a cumulative effect of the chemical bonds at the specific wave numbers. Therefore, the entire NIR spectrum region scanned was used in the PLS analysis instead of the characteristic bands typically used in the mid-infrared spectroscopy. Corn stover and switchgrass were first modeled separately with PLS, and two sets of calibration model were thus developed. The quantified cross-validation results for each single model are listed in Table 3.

Table 3
Statistics for model evaluation and cross-validation of the individual corn stover and switchgrass models as well as the broad-based model.

	Glucan	Xylan	Galactan	Arabinan	Mannan	Lignin	Ash
Mean (%w/w)							
Corn stover	35.31	17.75	2.04	3.43	1.16	20.67	4.48
Switchgrass	37.23	20.28	2.39	3.56	0.85	22.28	2.95
S + G	36.55	19.13	2.21	3.56	0.99	21.49	3.37
Model evaluation							
Corn stover							
SEP (%w/w)	0.93	1.04	0.19	0.28	0.23	1.04	1.08
R/SEP	14.31[a]	12.97[a]	9.95[b]	13.96[a]	9.74[b]	13.71[a]	7.05[b]
Switchgrass							
SEP (%w/w)	0.66	0.58	0.13	0.20	0.12	0.49	0.26
R/SEP	12.75[a]	10.41[a]	14.37[a]	13.01[a]	7.39[b]	9.77[b]	14.56[a]
S + G							
SEP (%w/w)	0.85	0.88	0.26	0.34	0.18	1.32	0.62
R/SEP	16.31[c]	15.06[c]	9.12[b]	11.38[a]	12.73[a]	10.84[a]	12.16[a]
Model validation							
Corn stover							
Correlation	0.97	0.93	0.92	0.96	0.91	0.94	0.90
RMSEP (%w/w)	0.97	1.03	0.17	0.27	0.21	1.12	0.57
Switchgrass							
Correlation	0.97	0.96	0.97	0.95	0.87	0.93	0.96
RMSEP (%w/w)	0.65	0.57	0.12	0.19	0.11	0.48	0.26
S + G							
Correlation	0.97	0.94	0.88	0.91	0.89	0.90	0.88
RMSEP (%w/w)	0.84	0.87	0.26	0.33	0.18	1.31	0.61

Note: "S + G" denotes the broad-based model, including both corn stover (S) and switchgrass (G). SEP = standard error of performance (%w/w); *R* = data range (%w/w); RMSEP = root mean square error of prediction (%w/w).

[a] The calibrations that meet the requirements for quality control.
[b] The calibrations that meet requirements for industrial screening.
[c] The calibrations that meet quantitative analysis requirements.

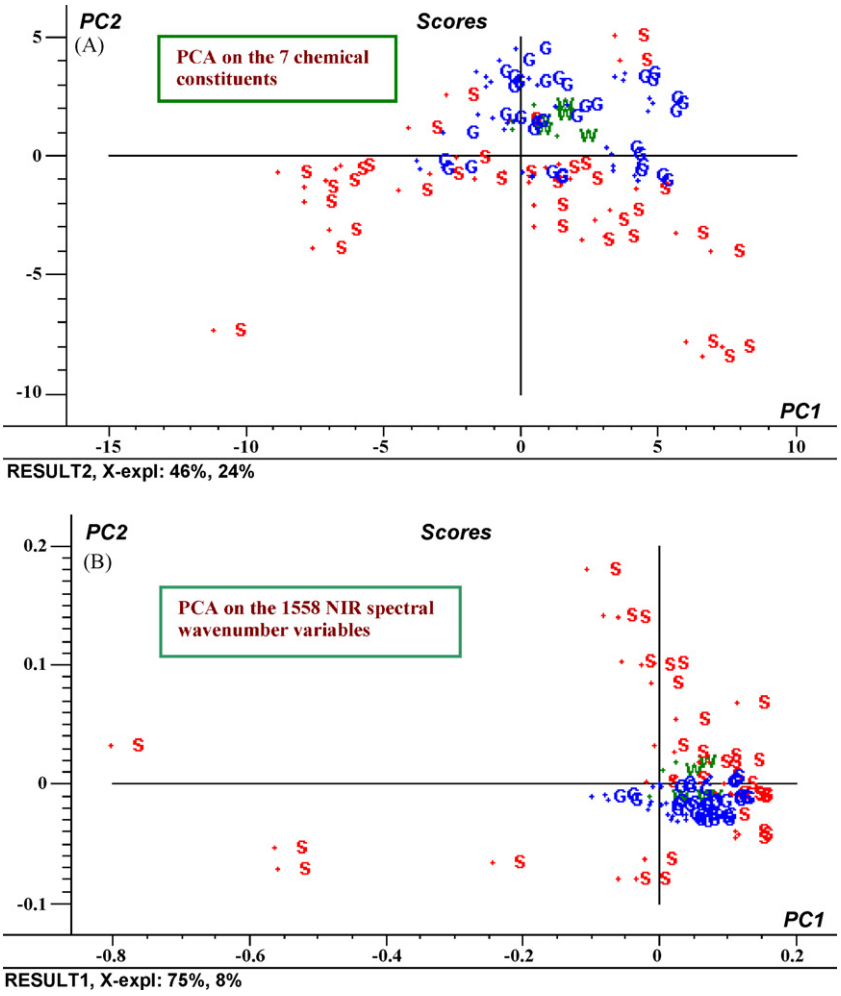

Fig. 3. Score plot of the first two principle components of PCA on chemical composition (A) and on FT-NIR spectra (B).

Both models had good correlations (>0.9) between the predicted and the measured value, especially for glucan (both are 0.97), which is a key constituent for sugar fermentation. As for the corn stover model, the root mean square error of prediction (RMSEP) was 0.97%w/w for glucan, 0.57%w/w for xylan, and 0.48%w/w for lignin, which together make up over 70% of the weight percentage of the dry biomass. The RMSEP for the constituents of corn stover compared with its average content in biomass is low; for glucose, especially, the relative error expressed as RMSEP/mean (Versari, Parpinello, Mattioli, & Galassi, 2008) was only 2.75%. Even smaller RMSEPs were observed for the switchgrass single model, as well as smaller relative errors (e.g., 1.75% for glucan prediction, 2.82% for xylan, and 2.15% for lignin).

As stated earlier, the purpose of using FT-NIR predictive modeling is to facilitate industrial monitoring processes; therefore, it is worthwhile to evaluate the model's performance with the industrial application in mind. However, for the bio-energy industry, there has not yet been any evaluation standard established for biomass feedstocks. One parameter, *R*/SEP (*R* being the range of the validation dataset and SEP being the standard error of performance), that has been used in the cereal industry was adopted to further evaluate the developed model's performance. The American Association of Cereal Chemists (AACC) stated in *AACC Method 39-00* (AACC, 1999) that any model with *R*/SEP ≥ 4 is qualified for screening calibration, ≥10 is acceptable for quality control, and ≥15 is very good for research quantification. Based on these criteria, the prediction of corn stover's galactan, mannan, and ash contents by the individual model reached the industrial screening level, and the prediction of corn stover's glucan, xylan, arabinan and lignin by the individual model was found to be good for quality control. The developed switchgrass individual model can predict mannan and lignin for screening purposes and can predict all the other five constituents for quality-control purposes. Our method of using FT-NIR techniques to acquire chemical composition information works for both corn stover and switchgrass via individual modeling.

3.2.2. *Rationale and development of the broad-based model*

As noted, earlier research indicated that if a calibration involved more than a single species, the dissimilarities among the species would sacrifice accuracy. However, this study demonstrates a rationale for combining corn stover and switchgrass into one calibration before proceeding to the development of a broad-based model. Principle component analysis (PCA) was utilized to justify this point. Fig. 3 shows the samples' relocation in the dimensions composed of principle component (PC)1 and PC2, where the new structure is able to explain the maximum variance of the original data (Esbensen, 2002). Fig. 3A is the result of PCA conducted solely on wet chemical data. The variations explained by PC1 and PC2 are 46% and 24% respectively, which together explain 70% of variance. In this plot, which is representative of the chemical dataset, it is hard to distinguish switchgrass samples from corn stover. Likewise, PCA results based on only FT-NIR spectra of the corn stover and switchgrass samples again showed that corn stover and switchgrass samples were not differentiable, representing 83% of total

Table 4
Performance of broad-based model in predicting 15 independent validation samples, including five corn stover, five switchgrass and five wheat-straw samples.

	Glucan	Xylan	Galactan	Arabinan	Mannan	Lignin	Ash
Mean (%w/w)	38.89	19.44	1.90	3.45	0.90	22.49	3.02
RMSEP (%w/w)	0.77	0.46	0.21	0.26	0.06	0.81	0.42
SEP (%w/w)	0.78	0.48	0.21	0.28	0.07	0.82	0.43
R/SEP	12.58	12.87	7.24	11.25	8.21	11.23	11.73
Relative error (%)	1.99	2.37	10.96	7.53	6.65	3.62	13.95

SEP = standard error of performance (%w/w); *R* = data range (%w/w); RMSEP = root mean square error of prediction (%w/w); relative error = RMSEP/mean.

variance (Fig. 3B). Thus, the PCA analysis justifies combining both species together to start development of a broad-based model. Further, wheat-straw samples (denoted as "W") were also included in the PCA; in both Fig. 3A and B, the wheat-straw samples could not be distinguished from the corn stover and switchgrass used for calibration. The result again justified the selection of wheat straw as part of the independent validation of the broad-based model.

Complementary to the statistics shown in Table 3, the accuracy of the FT-NIR broad-based predictive model is illustrated in Fig. 4, which compares the composition as determined by wet chemistry analysis with that predicted by the model. The diagonal line represents perfect agreement. Fig. 4A presents the data from the cross-validation, which provides a conservative estimation of the accuracy of the predictive model, and Fig. 4B presents the data from the 15-sample independent validation. The distribution of points along the diagonal line represents the range of calibration or validation, while the distribution of the points deviating from the diagonal line represents errors in the prediction. Fig. 4 visually shows the accuracy of the developed broad-based model. The cross-validation statistics for the broad-based model are also listed in Table 3 for the convenience of comparison. Although the correlations slightly decreased for some constituents compared to both of the two individual models, they were still satisfying, especially for the three major constituents (the correlation being greater than 0.9). Considering the prediction errors (RMSEPs and SEPs), this broad-based model shows an improvement over the corn stover individual model. Comparing *R*/SEP among the three developed models in Table 3, the predictive accuracy for glucan and xylan by the broad-based model has been raised to the research quantification level ($R/SEP \geq 15$), a significant improvement over the accuracy of the individual models. Except for the galactan prediction, which meets the screening criteria ($R/SEP \geq 4$), the predictions of all the remaining four constituents are at a level that can be used for industrial quality control ($R/SEP \geq 10$). Comparing the cross-validation results of the three models, the broad-based model for both corn stover and switchgrass did not lose prediction accuracy, and actually made the predictive model application more robust.

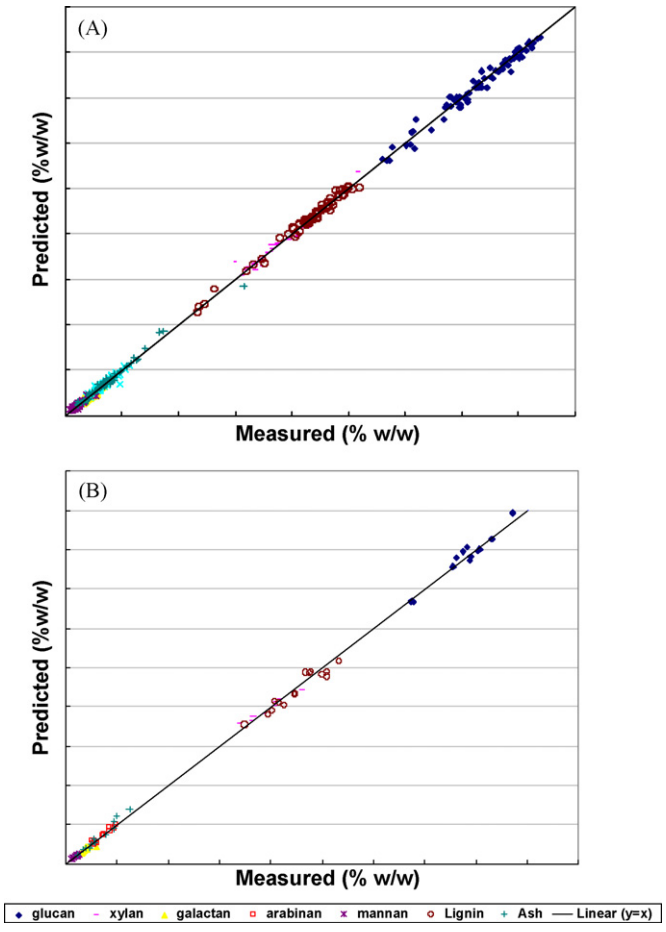

Fig. 4. Overall results of FT-NIR/PLS of the broad-based model (predicted vs. measured in %w/w). (A) Cross-validation and (B) independent validation using 15 samples.

To further verify the predictive accuracy of the broad-based model, we performed an independent validation using a dataset composed of five corn stover samples, five switchgrass samples, and five wheat-straw samples, all of which were randomly selected and were not included in the calibration. Compared to the cross-validation results, RMSEP and SEP decreased for all seven constituents (Table 4), showing that the cross-validation technique is a conservative and reliable substitute for the independent validation technique. The relative errors for the three major and important constituents, glucan, xylan, and lignin, were 1.99%, 2.37% and 3.62%, respectively, showing that they can be well predicted. The poor relative error result for ash is due to the fact that inorganic compounds do not absorb NIR and ash content was rather indirectly determined through correlation with other NIR absorbing species (Esbensen, 2002; Lestander & Rhen, 2005). According to AACC criteria, this independent validation shows that the predictions for most constituents meet the criteria for quality control, except for the prediction rates for galactan and mannan, which can only serve for screening purposes. Compared to those for the major constituents (i.e., glucan, xylan and lignin), the relatively poor predictive results for these minor sugars could be possibly explained by (1) their weight percentages being comparatively too small while the PLS regression is inclined to the variation of large-proportional constituents; (2) the small proportions resulting in low concentrations in the hydrolyzate, and thus the HPLC measurement not being as accurate as that for glucan and xylan. Nevertheless, the independent validation results showed consistency with the cross-validated results and further confirmed the good predictive performance of the broad-based model.

4. Conclusions

Our wet chemistry analyses showed that (1) great variability of chemical composition exists among different botanic fractions of corn stover, while this variability is much lower among botanic fractions of switchgrass; (2) for both corn stover and switchgrass, internodal fractions contain higher glucan content than nodes and

leaves, and leaves have higher ash content; (3) greater variability in chemical composition exists within corn stover species than within switchgrass species; (4) switchgrass has higher sugar content and lower ash content than corn stover; (5) corn husk is promising for fermentation due to its high total sugar content and low ash and lignin content; (6) among the six switchgrass cultivars, Cave-in-Rock and NC2-16 have significantly higher glucan content than other cultivars, while NC1-16 has the lowest glucan content.

The model validation results demonstrated that coupling the FT-NIR technique with chemometrics is a practical and promising tool for the analysis of biomass chemical composition. Our results showed the potential for the development of a broad-based FT-NIR predictive model across multiple agricultural biomass species, and the developed model achieved good accuracy in predicting the chemical composition of corn stover, switchgrass, and also a third species, wheat straw. Further development should include more biomass species/samples in the broad-based model to make the FT-NIR techniques more robust and practical for industrial applications.

Acknowledgements

This research was supported in part by the USDA-DOE Biomass Research and Development Initiative DEPA36-04GO94002. The authors would also like to thank Tennessee Agricultural Experiment Station for financial support (HATCH project No. TEN00325).

References

AACC. (1999). *Approved methods of the American Association of Cereal Chemists, Method 39-00. Near-infrared methods—Guidelines for model development and maintenance*. St. Paul, MN: AACC Press.

Albanell, E., Plaixats, J., Ferret, A., Bosch, L., & Casanas, F. (1995). Evaluation of near-infrared reflectance spectroscopy for predicting stover quality trait in semi-exotic populations of maize. *Journal of the Science of Food and Agriculture*, *69*, 269–273.

ASTM. (1996). *Standard test methods for direct moisture content measurements of wood and wood-base materials. ASTM D4442-92*. Philadelphia, PA: American Society for Testing Materials, ASTM., pp. 485–489

ASTM. (2002). *Standard Test Method for Ash in Biomass, ASTM E1755-95*. Philadelphia, PA: American Society for Testing Materials, ASTM.

Ben-Dor, A., Bruhn, L., Friedman, N., Nachman, I., Schummer, M., & Yakhini, Z. (2000). Tissue classification with gene expression profiles. *Journal of Computational Biology*, *7*(3–4), 559–583.

Boateng, A. A., Hicks, K. B., & Vogel, K. P. (2006). Pyrolysis of switchgrass (*Panicum virgatum*) harvested at several stages of maturity. *Journal of Analytical and Applied Pyrolysis*, *75*, 55–64.

Bothast, R. J., Saha, B. C., Flosenzier, A. V., & Ingram, L. O. (1994). Fermentation of L-arabinose, D-xylose and D-glucose by ethanologenic recombinant *Klebsiella oxytoca* strain P2. *Biotechnology Letters*, *16*, 401–406.

Christian, D. G., Riche, A. B., & Yates, N. E. (2002). The yield and composition of switchgrass and coastal panic grass grown as a biofuel in Southern England. *Bioresource Technology*, *83*, 115–124.

Curda, L., & Kukackova, O. (2004). NIR spectroscopy: A useful tool for rapid monitoring of processed cheeses manufacture. *Journal of Food Engineering*, *61*(4), 557–560.

Demirbas, A. (2001). Biomass resource facilities and biomass conversion processing for fuels and chemicals. *Energy Conversion and Management*, *42*, 1357–1378.

Duguid, K. B., Montross, M. D., Radtke, C. W., Crofcheck, C. L., Shearer, S. A., & Hoskinson, R. L. (2007). Screening for sugar and ethanol processing characteristics from anatomical fractions of wheat stover. *Biomass and Bioenergy*, *31*, 585–592.

Esbensen, K. H. (2002). *Multivariate data analysis—In practice* (5th ed.). Oslo: Camo ASA.

Fuentes, R. G., & Taliaferro, C. M. (2002). Biomass yield stability of switchgrass cultivars. In J. Janick, & A. Whipkey (Eds.), *Trends in new crops and new uses* (pp. 276–282). Alexandria, VA: ASHS Press.

Goodchild, A. V., El Haramein, F. J., El Moneim, A., Makka, H., & Williams, S. P. (1998). Prediction of phenolics and tannins in forage legumes by near infrared reflectance. *Journal of Near Infrared Spectroscopy*, *6*(1), 175–181.

Griffiths, P. R., De Haseth, & James, A. (2007). *Fourier transform infrared spectrometry* (2nd ed.). New York: Wiley-Interscience.

Grover, P. D. (1997). Biomass feed processing for energy conversion. In *Proceedings of the International Conference Biomass Energy Systems* New Delhi, (pp. 177–195).

Hames, B. R., Thomas, S. R., Sluiter, A. D., Roth, C. J., & Templeton, D. W. (2003). Rapid biomass analysis—New tools for compositional analysis of corn stover feedstocks and process intermediates from ethanol production. *Applied Biochemistry and Biotechnology*, *105*, 5–16.

Hildrum, K. I., Nilsen, N., Mielnik, M., & Næs, T. (1996). Near infrared reflectance spectroscopy in the prediction of sensory properties of beef. *Journal of Near Infrared Spectroscopy*, *3*(*2*), 81–87.

Hong, J.-H., Ikeda, K., Kreft, I., & Yasumoto, K. (1996). Near-infrared diffuse reflectance spectroscopic analysis of the amounts of moisture, protein, starch, amylose, and tannin in buckwheat flours. *Journal of Nutritional Science and Vitaminology*, *42*, 359–366.

Hopkins, A. A., Vogel, K. P., Moore, K. J., Johnson, K. D., & Carlson, I. T. (1995). Genotypic variability and genotype × environment interactions among switchgrass accessions from the midwestern USA. *Crop Science*, *35*, 565–571.

Kelley, S. S., Rials, T. G., Snell, R., Groom, L. H., & Sluiter, A. (2004). Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood. *Wood Science and Technology*, *38*, 257–276.

Kelley, S. S., Rowell, R. M., Davis, M., Jurich, C. K., & Ibach, R. (2004). Rapid analysis of the chemical composition of agricultural fibers using near infrared spectroscopy and pyrolysis molecular beam mass spectrometry. *Biomass and Bioenergy*, *27*(1), 77–88.

Kim, Y., & Kays, S. E. (2009). Near-infrared (NIR) prediction of trans-fatty acids in ground cereal foods. *Journal of Agricultural and Food Chemistry*, *57*(18), 8187–8193.

Klasek, S. E. (2006). *Terminal velocity determinations for component separation of biomass*. Knoxville: Dept. of Biosystems Engineering and Soil Science, University of Tennessee.

Kracl, N.P. (1986). *Corn harvester combine assembly with corn removal attachment*. US Patent Office, Pat. No. 4,583,354.

Lemus, R., Brummer, E. C., Moore, K. J., Molstad, N. E., Burras, C. L., & Barker, M. F. (2002). Biomass yield and quality of 20 switchgrass populations in southern Iowa, USA. *Biomass and Bioenergy*, *23*, 433–442.

Lestander, T. A., & Rhen, C. (2005). Multivariate NIR spectroscopy models for moisture, ash and calorific content in biofuels using bi-orthogonal partial least squares regression. *Analyst*, *130*, 1182–1189.

Liu, L., & Chen, H. (2007). Prediction of maize stover components with near infrared reflectance spectroscopy. *Guang Pu Xue Yu Guang Pu Fen Xi*, *27*, 275–278.

Manley, M., Van Zyl, L., & Osborne, B. G. (2002). Using Fourier transform near infrared spectroscopy in determining kernel hardness, protein, and moisture content of whole wheat flour. *Journal of Near Infrared Spectroscopy*, *10*, 71–76.

Marten, G. C., Brink, G. E., Buxton, D. R., Halgerson, J. L., & Hornstein, J. S. (1984). Near infrared reflectance spectroscopy analysis of forage quality in four legume species. *Crop Science*, *24*, 1179–1182.

Martens, H. A., & Dardenne, P. (1998). Validation and verification of regression in small data sets. *Chemometrics and Intelligent Laboratory Systems*, *44*, 99–121.

McCarthy, W. J., & Kemeny, G. J. (2001). Fourier transform spectraphotometers in near-infrared. In E. W. Burns, & D. A. Ciurczak (Eds.), *Handbook of near-infrared analysis* (p. 814). CRC Press.

McLaughlin, S. B., & Kszos, L. A. (2005). Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. *Biomass and Bioenergy*, *28*, 515–535.

Meuret, M., Dardenne, P., Biston, R., & Poty, O. (1993). The use of NIR in predicting nutritive value of Mediterranean tree and shrub foliage. *Journal of Near Infrared Spectroscopy*, *1*, 45–54.

Michell, A. J., & Schimleck, L. R. (1996). NIR spectroscopy of woods from *Eucalyptus globulus*. *Appita Journal*, *49*(1), 23–26.

Montross, M. D., & Crofcheck, C. L. (2004). Effect of stover fraction and storage method on glucose production during enzymatic hydrolysis. *Bioresource Technology*, *92*, 269–274.

Porter, C. L. (1966). An analysis of variation between upland and lowland switchgrass, *Panicum virgatum* L., in central Oklahoma. *Ecology*, *47*, 980–992.

Sanderson, M. A., Agblevor, F., Collins, M., & Johnson, D. K. (1996). Compositional analysis of biomass feedstocks by near infrared reflectance spectroscopy. In *Proceedings of the Liquid Fuel Conference on Liquid Fuel and Industrial Products from Renewable Resources* Nashville, TN, (pp. 37–42).

Schimleck, L. R., Raymond, C. A., Beadle, C. L., Downes, G. M., Kube, P. D., & French, J. (2000). Applications of NIR spectroscopy to forest research. *Appita Journal*, *53*, 458–464.

Shao, J. (1993). Linear model selection by cross-validation. *Journal of American Statistical Association*, *88*, 486–494.

Sivakesava, S., & Irudayaraj, J. (2001). Rapid determination of tetracycline in milk by FT-MIR and FT-NIR spectroscopy. *Journal of Dairy Science*, *85*(*3*), 487–493.

Sladden, S. E., Bransby, D. I., & Aiken, G. E. (1991). Biomass yield, composition and production costs for eight switchgrass varieties in Alabama. *Biomass and Bioenergy*, *1*, 119–122.

Smith, K. F., & Flinn, P. C. (1991). Monitoring the performance of a broad-based calibration for measuring the nutritive value of two independent populations of pasture using near infrared reflectance (NIR) spectroscopy. *Australian Journal of Experimental Agriculture*, *31*, 205–210.

Sokhansanj, S., Turhollow, A., Cushman, J., & Cundiff, J. (2002). Engineering aspects of collecting corn stover for bioenergy. *Biomass and Bioenergy*, *23*, 347–355.

Togersen, G., Isaksson, T., Nilsen, B. N., Bakker, E. A., & Hildrum, K. I. (1999). On-line NIR analysis of fat, water and protein in industrial scale ground meat batches. *Meat Science*, *51*, 97–102.

Towne, G., & Owensby, C. (1984). Long-term effects of annual burning at different dates in ungrazed kansas tallgrass prairie. *Journal of Range Management*, *37*, 392–397.

Van Zyl, C., Prior, B. A., & Du Preez, J. C. (1988). Production of ethanol from sugar cane bagasse hemicellulose hydrolyzate by *Pichia stipitis*. *Applied Biochemistry and Biotechnology*, *17*, 357–369.

Versari, A., Parpinello, G. P., Mattioli, A. U., & Galassi, S. (2008). Determination of grape quality at harvest using Fourier-transform mid-infrared spectroscopy and multivariate analysis. *American Journal of Enology and Viticulture*, *5*, 317–322.

Wang, J., Wang, J. Q., Bu, D. P., Guo, W. J., Shen, J. S., Wei, H. Y., et al. (2009). Application of near-infrared spectroscopy to quality detection of milk and its products. *Spectroscopy and Spectral Analysis*, *29*(*5*), 1281–1285.

Westa, F., & Martens, H. (2000). Variable selection in near infrared spectroscopy based on significance testing in partial leastsquares regression. *Journal of Near Infrared Spectroscopy*, *8*, 117–124.

Wold, S. (1978). Cross-validatroy estimation of the number of components in factor and principle components models. *Technometrics*, *20*, 397–405.

Ye, X. P., Liu, L., Hayes, D., Womac, A., Hong, K. L., & Sokhansanj, S. (2008). Fast classification and compositional analysis of corn stover fractions using Fourier transform near-infrared techniques. *Bioresource Technology*, *99*, 7323–7332.