



Research article

Investigating the impact of biomass quality on near-infrared models for switchgrass feedstocks

Lindsey M. Kline^{1,*}, Nicole Labbé^{1,*}, Christopher Boyer², T. Edward Yu²,
Burton C. English², and James A. Larson²

¹ Center for Renewable Carbon, University of Tennessee, 2506 Jacob Drive, Knoxville, TN 37996 USA

² Department of Agricultural and Resource Economics, University of Tennessee, Knoxville, TN USA

* **Correspondence:** Email: lmccull4@utk.edu, nlabbe@utk.edu; Tel: 865-946-1126; Fax: 865-946-1109

Abstract: The aim of this study was to determine the impact of incorporating switchgrass samples that have been in long term storage on the development of near-infrared (NIR) multivariate calibration models and their predictive capabilities. Stored material contains more variation in their respective spectral signatures due to chemical changes in the bales with storage time. Partial least squares (PLS) regression models constructed using NIR spectra of stored switchgrass possessed an instability that interfered with the correlation between the spectral data and measured chemical composition. The models were improved using calibration sample sets of equal parts stored and fresh switchgrass to more accurately predict the chemical composition of stored switchgrass. Acceptable correlation values ($r_{\text{calibration}}$) were obtained using a calibration sample set composed of 25 stored samples and 25 samples of fresh switchgrass for cellulose (0.91), hemicellulose (0.74), total carbohydrates (0.76), lignin (0.98), extractives (0.92), and ash (0.87). Increasing the calibration sample set to 100 samples of equal parts stored to senesced material resulted in statistically increased ($p = 0.05$) correlations for total carbohydrates (0.89) and ash (0.96). When these models were applied to a separate validation set (equal to 10% of the calibration sample set), high correlation coefficients (r) for predicted versus measured constituent content were observed for cellulose (0.94), total carbohydrates (0.98), lignin (0.91), extractives (0.97), and ash (0.90). For optimization of processing economics, the impact of feedstock storage must be investigated for implementation in conversion processes. While NIR is a well-known high-throughput technique for characterization of senesced

switchgrass, the selection of appropriate calibration samples and consequent multivariate models must be taken into careful consideration for NIR application in a biomass storage facility for rapid chemical compositional determination.

Keywords: switchgrass; feedstock quality; chemical composition; storage; near-infrared spectroscopy; partial least squares regression

Abbreviations

NIR	near-infrared (spectroscopy)
BT3	BaleTech3
HDPE	high density polyethylene
LLPE	low linear polyethylene
PS	particle size
PC	principal component
PCA	principal component analysis
PLS	partial least squares (regression)
$r_{\text{calibration}}$	calibration correlation
$r_{\text{validation}}$	validation correlation
RMSEC	root mean square error of calibration
RMSEV	root mean square error of validation
R	component range (%)
SEP	standard error of prediction
r	correlation coefficient

1. Introduction

Near-infrared (NIR) spectroscopy is an effective, non-destructive, and inexpensive high-throughput method, widely used to characterize biomass composition [1–5]. Recent work has focused on application of NIR in the analysis of perennial herbaceous species for rapid analysis of biomass for use in various bioenergy conversion processes, including classification of plants grown under various environments [6] and predicting compositional properties of switchgrass and projected performance interests such as ethanol yield [7]. Although rapid and requiring a small biomass sample, NIR spectroscopy presents unique challenges. Unlike ultraviolet-visible (UV-Vis) and mid-infrared (MIR) spectroscopy, NIR must be coupled with multivariate statistics such as principal component analysis (PCA) and partial least squares regression (PLS) to extract valuable information from the large and complex spectral datasets. A NIR spectrum provides a unique chemical fingerprint profile for a given sample with the majority of bands observed in this range being overtone and combination peaks of molecular vibrations from O-H, C-H, S-H, and N-H stretching modes [8]. Much work has been completed on addressing issues present during the development of NIR multivariate calibration models, such as spectral data analysis and sample selection [9–11], as well as error associated with the standard techniques used to characterize the calibration set [12].

Switchgrass is one of the leading feedstocks for potential use in bioenergy and thermochemical conversion processes, particularly in the Southeastern United States [13]. Maintenance of quantity, quality, and performance of the switchgrass feedstock through efficient storage is a major challenge for large-scale commercial conversion facilities [14,15]. In addition to reducing biomass dry matter loss, another effect of feedstock storage is the reduction of biomass recalcitrance, which can aid in conversion processes [16]. However, a large change in compositional properties for stored switchgrass is attributed to degradation and consumption of carbohydrates during storage. Dry matter loss is not uniform across all constituents, instead preferring water soluble components (extractives and nonstructural ash) and selective degradation of hemicellulose, leaving the biomass enriched in cellulose and lignin [17–19]. The two most important factors for prevention of dry matter loss during storage are high temperatures and microbial dry matter oxidation [16]. As a result, many storage techniques rely upon conditions of low pH (<4.5), fermentative microbial activity to produce organic acids that inhibit the growth of microorganisms that consume or degrade cellulose, and low oxygen concentration created by densification of the material to avoid biological degradation [20]. Heat and water are produced as a result of the exothermic aerobic respiration process of microorganisms, leading to a self-heating of the biomass feedstock during storage. The internal temperature of the bale will remain elevated with the presence of available water, oxygen, and carbohydrates. At the time one of these factors is limited, the biomass temperature will decrease and stabilize at ambient conditions [21]. The porosity of the stack is influential as moisture can infiltrate the bale at exposed surfaces or retain water at the base of the bale, the biomass becomes unstable and can support fungal growth and lower the energy value of the product [20].

To further the progress on biomass collection, preprocessing and storage practices, past work [22] evaluated the profitability of using a new commercial stretch-wrap baler, BaleTech 3 (BT3), to compact chopped switchgrass into a wrapped large round bale by film and net for outdoor storage. The wrapped bale from BT3 is similar to the dimensions of an agricultural round bale but has the potential to reduce dry matter loss by increasing densification. BT3 was developed to reduce particle size and increase densification to improve the economics of storage. In a recent study, Larson et al. [22] concluded that BT3 may be an economically feasible alternative to traditional agricultural bales for chopped switchgrass; however, dry matter loss and chemical composition data were not reported to accurately measure the profitability of BT3. Yu et al. [15] analyzed the dry matter loss from using BT3 and found that bales wrapped in both film and net experienced less dry matter loss than bales wrapped with net only [15]. Larson et al. [23] extended the study by determining the impacts of particle size, wrap material, and storage time on the chemical composition of switchgrass bales preprocessed with this storage technology within the same dataset. The study described here is meant to be a companion analysis to the other works surrounding the BT3 baling technology [15,23], as NIR modeling has proven valuable in the analysis and evaluation of these storage experiments by dramatically increasing the number of samples that can be analyzed during the course of the investigation. Improvement of NIR compositional models utilized at a biomass storage facility is necessary, as the treatment will have an inevitable impact on the estimated chemistry of the biomass and corresponding infrared spectra.

2. Materials and Method

2.1. Biomass materials

Table 1. Summary of storage conditions for switchgrass samples utilized in the development of NIR compositional models.

Storage type	Particle size (cm)	Storage time (days)
BT3 wrapped, multiple layers of LLPE green film + HDPE mesh net	(Small) 1.27–1.91	0
	(Medium) 7.62	75
	(Large) Full stock, ~243.84	150
		225
BT3 wrapped, multiple layers of LLPE clear film + HDPE mesh net	(Medium) 7.62	0
		75
		150
		225
BT3 wrapped, HDPE mesh net	(Small) 1.27–1.91	0
	(Medium) 7.62	75
	(Large) Full stock, ~243.84	150
		225
Square bales with twine	(Large) Full stock, ~243.84	0
		75
		150
		225

LLPE = linear low polyethylene

HDPE = high-density polyethylene

A summary of the experimental design for harvest, baling, and storage of switchgrass used in this study can be found in past work [15]. In brief, Alamo switchgrass located on approximately 200 acres around Vonore, TN was harvested in early February 2012 using a New Holland (New Holland Agriculture, New Holland, PA) BB9080 large rectangular baler (1.2 × 0.9 × 2.4 m) without a cutter under contract with the Tennessee BioFuels Initiative. The square bales were transferred to the Biomass Innovation Park in Vonore and stored under cover while switchgrass bales were broken open, processed through a Vermeer (Vermeer Corporation, Pella, IA) TG5000 tub grinder, and sampled for moisture content. The ground material was then conveyed to a BT3 (TLA Bale Tech LLC, South Orange, NJ) to be formed into large round bales of 1.2 m in diameter and 1.5 m in width. The BT3 equipment was originally developed for storage and transport of domestic waste, and past work has focused on feasibility of utilizing the technology to compact chopped or shredded biomass in bales comparable to an agricultural round bale in a biorefinery setting [15,22]. A summary of all bales produced for the study is shown in Table 1. Samples included traditional square bales, round bales wrapped in a mesh net to encompass the outside circumference of the bale excluding the ends, or round bales with the mesh net plus a polyethylene film to induce an anaerobic storage environment. A subset of the bales was weighed, destroyed and material randomly sampled to quantify the initial composition of fresh switchgrass (time = 0 days). The remaining bales were

stored for 75, 150, or 225 days. A composite of 180 samples was collected in total, with 130 of the samples (60 stored bales and 70 bales with no storage) randomly selected following methods presented in the ASTM protocol E1655-5 [24] and analyzed for chemical composition.

2.2. Chemical compositional analysis

The 130 calibration samples underwent analysis for chemical composition following standard National Renewable Energy Laboratory (NREL) biomass analytical procedures. Samples were initially dried at 40 °C to < 10% moisture content then ground using a Thomas Scientific (Swedesboro, NJ) Model 4 Wiley mill using a 40 mesh (0.425 mm) screen. As the moisture content of ground biomass can rapidly change when exposed to air, the ground samples were equilibrated for three weeks at ambient conditions (approximately 23 °C, 63% relative humidity) and total solids measurements were completed using a sub-sample dried in a 105 °C convection oven for a minimum of 4 h to determine the percent of total solid prior to compositional analyses. Each sample was first combusted at 575 °C for 24 h and weighed for measurement of total ash content. Prior to quantification of the other structural constituents, the switchgrass samples were extracted to remove non-structural components in a Dionex (Sunnyvale, CA) Accelerated Solvent Extractor 350, following the methods described previously [25,26]. In this process, 5 g of raw biomass (40 mesh) containing < 10% moisture were added to a 33 mL extraction cell and sequentially extracted by pressurized water then ethanol. In this extraction process, the cell is heated over 5 min to 100 °C, filled with solvent until reaching a pressure of 1500 psi, and held at these conditions for a 7 min static cycle. This process was repeated three times with each solvent. The material was then allowed to air dry to less than 10% moisture content by weight and a change in weight < 1% in 24 h, determined using a dried sub-sample for total solids determination. The extractives-free material was stored in polyethylene bags at ambient temperature until further analyses were performed. The quantification of cellulose, hemicellulose, lignin, and ash in the switchgrass was performed following standard methods using three replicates [27]. The procedure is suitable for samples that do not contain extractives. A two-stage acid-catalyzed hydrolysis was performed to fractionate the sample into soluble and insoluble matter, and the two fractions were separated through vacuum filtration and ceramic fine porosity filtering crucibles. The insoluble solid fraction consisted of acid-insoluble lignin and ash. The acid-insoluble lignin was quantified gravimetrically after combustion of the residue at 575 °C for 24 h. The monomeric units of polysaccharides within the soluble liquid fraction were quantified via a Flexar high-pressure liquid chromatography (Perkin Elmer, Shelton, CT) with a refractive index detector. The system was equipped with an Aminex HPX-87P carbohydrate column (300 × 7.8 mmID, 9 μm particle size) and deashing guard column (125-0118) from Bio-Rad (Hercules, CA), using deionized water at 0.25 mL/min at 85 °C. The acid-soluble lignin content was measured using a dual beam Thermo Scientific (Waltham, MA) Genesys 10S spectrophotometer, and this value combined with the gravimetric value for acid-insoluble lignin was added as the total lignin content. A total of eight primary components were quantified as a mass percentage of the oven dried biomass (on % dry basis): extractives, cellulose, hemicellulose (combined values for xylan, galactan, arabinan, and mannan), lignin, and total ash. A separate validation sample set of 10 samples selected to encompass all storage conditions was simultaneously analyzed using the same methods.

To ensure that these methodologies were applicable to samples with potential degradation due to moisture and microbial growth from storage, the analyses were carefully monitored for certain red flags and inaccuracies, such as mass closure, and the presence of sugar degradation products and other byproducts (i.e. HMF and furfural) [28] from the hydrolysis of degraded carbohydrates, that could result in an underestimation of cellulose and hemicellulose content. However, no such problems were detected with the samples described here. All analyses were also carried out simultaneously with a National Institute of Standards Technology (NIST) QA standard (Wheat Straw 8494), and analyses were repeated if total compositional mass closure values were outside of the acceptable range (95–105%).

2.3. Near-infrared collection

Near-infrared spectral data were collected for all samples ground to 40 mesh at the BioEnergy Science and Technology Unit at the Center for Renewable Carbon, University of Tennessee, Knoxville. Samples were randomly scanned using an Analytical Spectral Devices (ASD) Field Spectrometer. In the scanning chamber, a high-intensity light source was positioned at a right angle, with a fiber optic oriented at 60 degrees to the sample surface. Five reflectance spectra were collected for each sample at wavelengths ranging between 350–2500 nanometers using a 10.2 cm dish spinning at approximately 50 rpm, with 40 scans collected and averaged into each spectrum, aiding in encompassing variation in the inherently heterogeneous samples. Spectra were transferred from the ASD to the Unscrambler® v9.0 software (CAMO, Woodbridge, NJ). The spectral datasets were converted to absorbance and replicates averaged to reduce the size of the dataset and time required for the statistical analyses. The spectral resolution was also reduced from the 1 nm collection interval to 4 nm to reduce uncertainty and noise [29–31]. Final pretreatments included application of a mean normalization and multiplicative scatter correction to remove light scatter effects due to the differences in the physical nature of the biomass particles.

2.4. Multivariate analyses

2.4.1. Distinguishing differences in spectral data with principal component analysis (PCA)

To detect differences in the NIR spectra collected for each switchgrass sample, multivariate analysis via PCA was performed on the spectral data using the Unscrambler statistical software. As each spectrum is a unique chemical fingerprint for the biomass sample, PCA can be used to detect differences between intensities of peaks and the overall spectral signature between samples, and therefore distinguish variations in their respective chemical composition. PCA is a descriptive method that allows for visualization of variability within a large data set. This technique transforms the multivariate data set into a different data set that is dependent on new variables, called principal components (PC). The first principal component accounts for as much of the variability in the data as possible and is associated with a set of loadings, which are directly related to contributing wavelengths, and each successive PC is orthogonal to the preceding components and is associated with a decreasing proportion of the variability. Each spectrum has an associated score on each PC. Plotting the scores of different PCs against one another reveals spectra with similar score values,

showing that they possess similar spectral features, and thus share similar chemical composition [32,33].

2.4.2. Partial least squares (PLS) regression models

PLS multivariate calibration models were built using the Unscrambler software to correlate the NIR spectra (dependent variables) to the wet chemistry compositional data (independent variables), allowing for prediction of compositional constituents such as structural carbohydrates, lignin, and ash in switchgrass samples outside of the calibration sample set. The models were developed using the spectral region of 1100–2300 nm. PLS models were also developed for the spectral region of 1000–2500 nm but did not significantly improve the model and are therefore not presented here. To improve model development, samples were chosen at random while the sample set as a whole was made to contain all combinations of storage treatment. This was done in an attempt to provide the largest range of variation in the concentration of the chemical constituents to be analyzed, and that these concentrations were uniformly distributed over their total range of variation [24]. In the PLS regression analysis, all of the calibration samples are used to create a model that is then used to predict the composition of each sample in the calibration set. The correlation between the predicted and measured content is given as $r_{\text{calibration}}$. Models were then generated using a full cross-validation procedure for model estimation and testing, in which one sample was left out from the selected calibration set and the model was calculated based on the remaining samples. The value of the left-out data point was used for prediction, and the process repeated until every sample has been left out one time. Therefore, the validation correlation ($r_{\text{validation}}$) is checking how well a model will perform for future samples taken from the same population as the calibration samples. This validation tests for predictive significance, i.e., a well-fit model with little to no predictive power or “over-fitting”, and allows for estimation of the prediction error in application of the model to future samples. More detailed descriptions of the technique can be found elsewhere [32–35].

In addition to the cross-validation, a separate validation sample set of 10 samples was used to test the accuracy of the prediction by each model. In this way, the NIR-predicted composition was compared to wet chemistry methodologies to evaluate the predictive capabilities of each model for rapid characterization of the feedstock material.

3. Results and Discussion

3.1. Compositional differences in bale types due to aging and storage condition

Switchgrass calibration samples were analyzed for quantification of compositional data including, cellulose, hemicellulose, lignin, extractives, and ash following the NREL standard biomass analytical procedures. More detailed analyses on the impact of densification using the BT3 technology and storage time on these switchgrass bales for various particle sizes can be found elsewhere [15,23]. Of value to this study is the composition distribution plot of the calibration sample set for switchgrass material with and without storage (Figure 1). For baseline comparison, the average composition of the fresh (without storage, 0 days) switchgrass material can be found in Table 2, and compositional means were compared with Fisher’s least significant difference method ($p = 0.05$). The average value of cellulose in the fresh switchgrass bales was $36.3 \pm 1.5\%$. When

comparing the chemical composition of the bales over time, dry matter loss as a result of degradation or consumption of certain constituents results in a bale that appears enriched in the chemical constituents with less degradation. For example as shown in Table 2, the percentage of cellulose in switchgrass under storage increased with longer storage durations, with a statistical increase in cellulose at 75 ($40.1 \pm 1.7\%$) and 150 ($41.0 \pm 1.9\%$). With limited storage, some bales contained over 43.9% cellulose (Figure 1a). However, after remaining in storage for 225 days, the cellulose content statistically decreased when comparing to 150 days (Table 2). This shows the relative stability of cellulose within the bale with shorter storage durations (less than 225 days). There was less of an impact of storage on hemicellulose content in the bales when compared to the fresh material except for a significant loss in hemicellulose sugars within the first 75 days (Table 2). For longer storage durations, the bales were not statistically different from the fresh feedstock, and there was little change in the overall hemicellulose distribution, ranging from approximately 24.5% to more than 27.6% (Figure 1b). The total carbohydrate content in the stored samples (Figure 1c) was higher than the fresh material (63.9 ± 3.7), with an average carbohydrate content of $66.7 \pm 2.7\%$ for all storage durations. Biomass material undergoing storage resulted in higher lignin content (Figure 1d) than the switchgrass with no storage ($21.1 \pm 1.5\%$). The extractives content (Figure 1e) in the bales was shown to decrease with storage time, with the distribution profiles shifting to lower values than the average value of $9.4 \pm 2.4\%$. The ash content was lower in all aged switchgrass, with inorganic content less than $3.1 \pm 0.6\%$ (Figure 1f). Therefore, it can only be concluded that the quality of switchgrass bales was greatly impacted by storage time as their chemical composition was statistically different ($p = 0.05$) from material with no storage, and an expanded component range should be represented in the calibration sample set for accurate quantification of any storage feedstock samples by NIR modeling.

Table 2. Average chemical composition for switchgrass bales undergoing no storage (time = 0 days). Standard deviation ().

	Average composition ¹ (% of dry matter)					
	Component range	No Storage	75 Days	150 Days	225 Days	All Storage
Total Carbohydrate	55.2 - 71.6	63.6 ^C (3.7)	66.3 ^{BC} (2.6)	68.4 ^A (2.6)	65.6 ^C (2.1)	66.7 ^B (2.7)
Cellulose	30.0 - 43.9	36.3 ^D (1.5)	40.1 ^B (1.7)	41.0 ^A (0.9)	38.5 ^C (1.4)	39.9 ^B (2.0)
Hemicellulose	24.4 - 30.5	27.3 ^{AB} (1.5)	26.0 ^C (0.1)	27.3 ^{AB} (1.2)	27.5 ^A (1.3)	26.9 ^B (1.4)
Lignin	18.6 - 25.4	21.2 ^B (1.5)	22.6 ^A (1.0)	22.6 ^A (0.8)	22.3 ^A (0.9)	22.5 ^A (0.9)
Extractives	4.7 - 15.6	9.4 ^A (2.4)	5.9 ^B (1.5)	4.8 ^C (1.5)	5.5 ^{BC} (1.4)	5.4 ^{BC} (1.5)
Ash	1.8 - 5.0	3.1 ^A (0.6)	2.1 ^B (0.6)	2.1 ^B (0.5)	2.2 ^B (0.6)	2.1 ^B (0.6)

1. Letter assignments correspond to mean separation using Fisher's protected LSD at the 5% significance level.

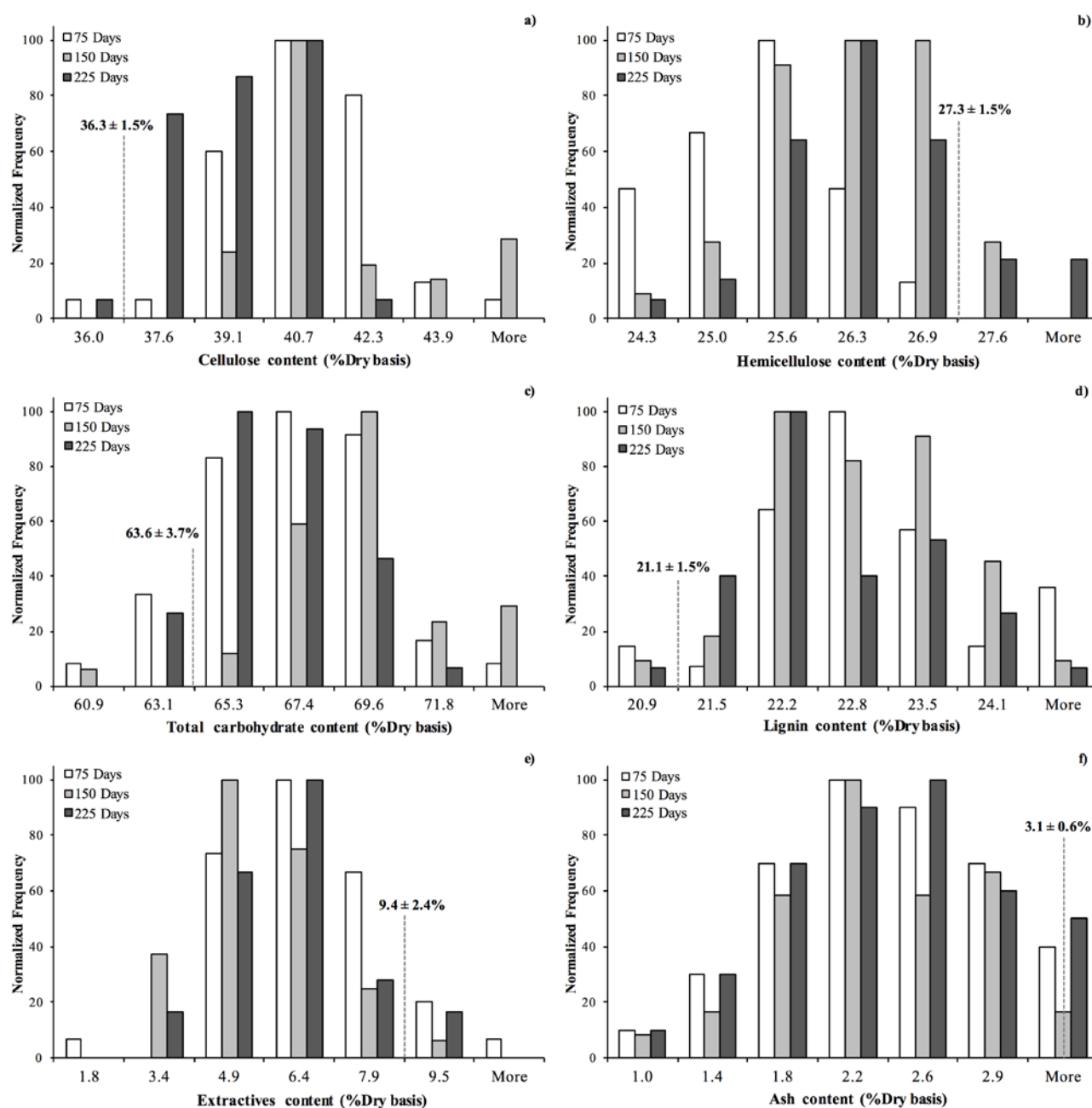
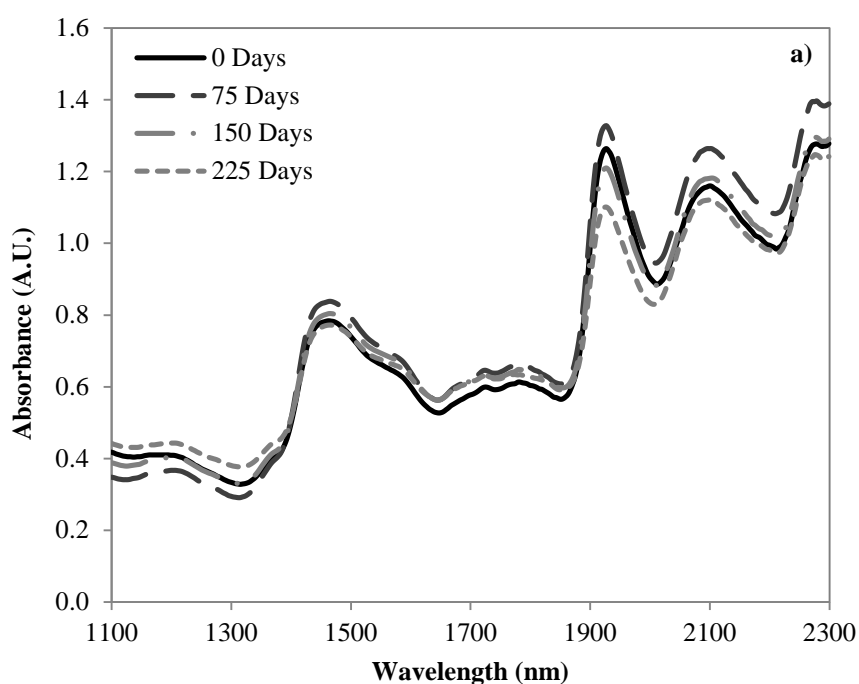


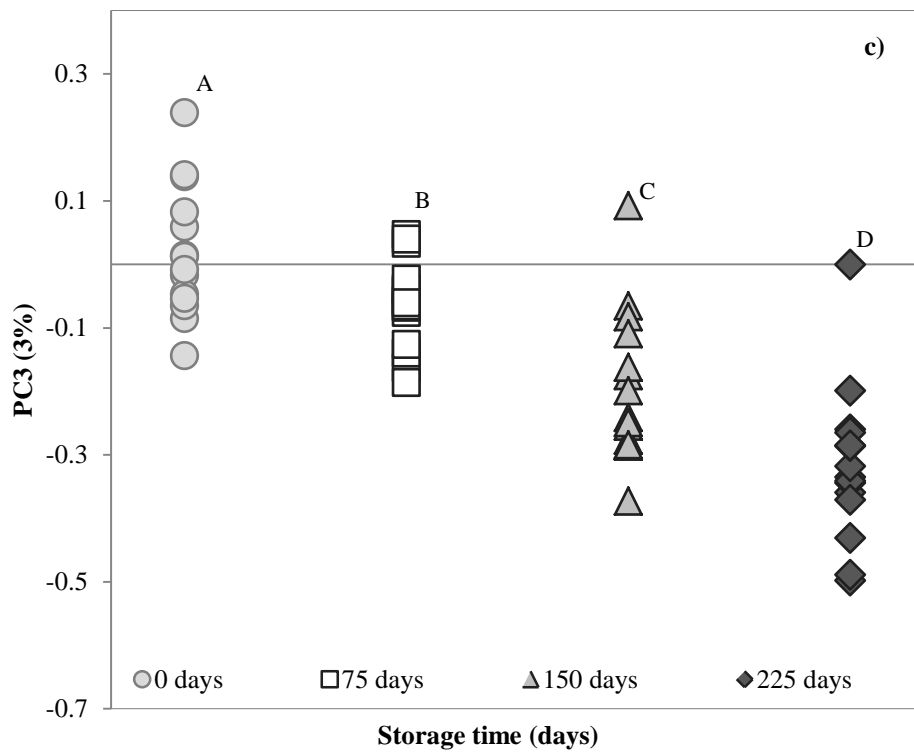
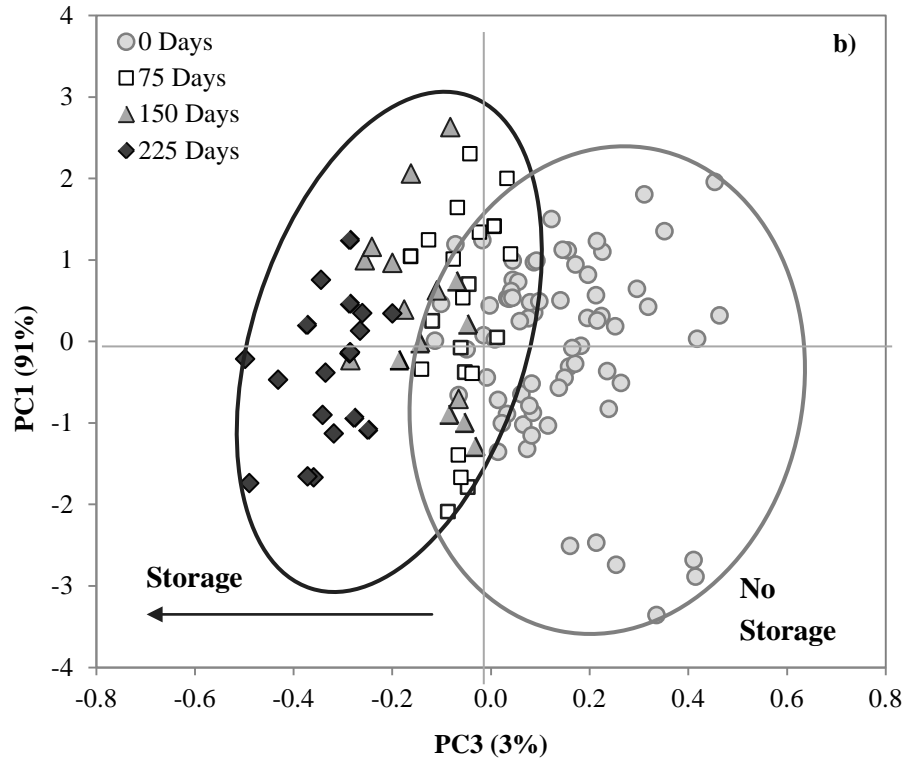
Figure 1. Distribution of chemical constituents within the calibration sample set of either bales stored with various treatments for 75–225 days (bars) or material with no storage (time = 0 days, average marked with dotted line) for (a) cellulose, (b) hemicellulose, (c) total carbohydrates, (d) lignin, (e) extractives, and (f) ash content.

3.2. Classification of spectral datasets due to innate chemical differences

Principal component analysis (PCA), a multivariate statistical analysis method used to detect differences, similarities, and trends in large datasets, was used to analyze the NIR spectra collected on all the switchgrass samples (Figure 2a). Despite the evident chemical differences found from the compositional analysis of the calibration sample set, examination of the NIR spectra detected very small differences between the feedstock materials (Figure 2b). No discernible clustering was

observed using PC1 and PC2; however, the scores plot of PC1 versus PC3 revealed two overlapping clusters for stored material and switchgrass with no storage. Fresh switchgrass was positive along the axis for PC3. Moving negatively along the horizontal axis, there was a progression of spectra for storage bales with time, as the 75 days samples were followed by spectra for 150 and 225 days. The scores plot of PC3, accounting for only 3% of the total variance in the spectra (Figure 2c), versus storage time revealed an indirect relationship between the significant spectral bands for this PC and increased storage time. The significant bands of PC3 were 2103, 1923, 1747, 1655, 1375, and 1279 nm, suggesting these portions of the spectrum may have been the most affected by the storage conditions (Figure 2d). The bands at 1923 and 2103 nm were assigned to hydroxyl groups, while the band at 1747 nm was attributed to C-H functional groups present in biomass components [8].





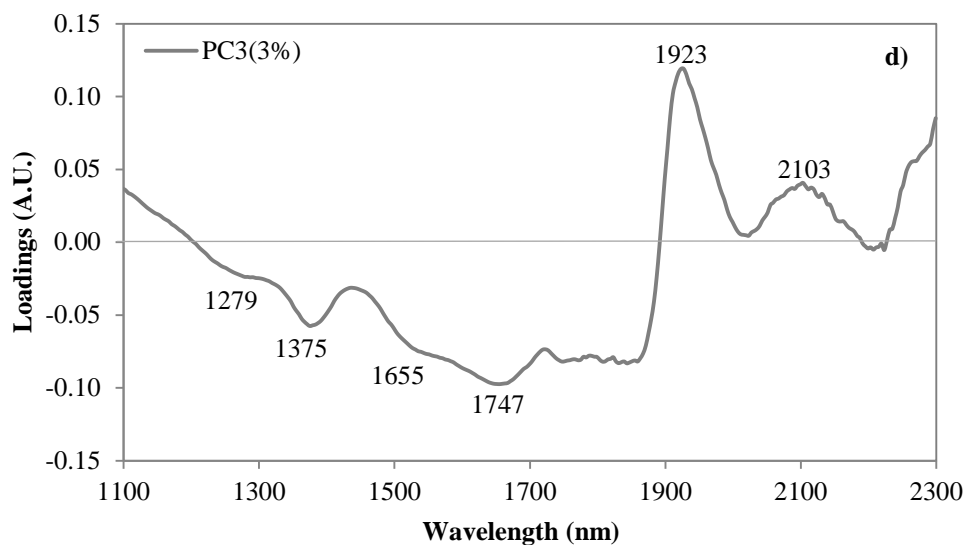


Figure 2. (a) Representative NIR spectra for senesced switchgrass undergoing limited or extended storage; (b) scores plot; (c) PC3 (principal component) derived from the NIR absorbance spectra with storage time; and (d) loadings plot for PC3 of the switchgrass samples using Fisher's test at a significance of 5%.

Specifically, the 1923 nm band is related to the intramolecular hydrogen bond between water and OH of switchgrass components [6,8,36]. While the differences between spectra based on storage time were minimal, this analysis demonstrates that the chemical fingerprint of the storage samples was unique and distinguishable from the spectra of switchgrass samples that were not subjected to any storage condition. Therefore, careful consideration should be taken in the selection of the calibration set to include samples with storage treatment to encompass these spectral variations.

3.3. PLS calibration models

PLS multivariate calibration models were constructed using the Unscrambler software and NIR spectra of switchgrass samples (Table 3). In the development of a multivariate model, the calibration sample set can be enhanced with the inclusion of a wider range in composition. Therefore, four model variations were built to determine the impact of spectra collected on aged samples using calibration sample sets consisting of (1) no storage, (2) exclusively storage, (3 and 4) equal parts of fresh and stored samples ($n = 50$, and $n = 100$). Comparisons of the regression correlations ($r_{\text{calibration}}$ and $r_{\text{validation}}$) and root mean square errors of calibration and validation (RMSEC and RMSEV) are commonly used to evaluate the robustness of NIR calibration models. To statistically compare correlation values between models, the Fisher z-transform ($z = \text{atanh}(R)$) was applied and confidence intervals were calculated around the z-transformed variable [37]. In general, correlations above ≥ 0.80 are considered robust and good for quality assurance [8]. Another statistical parameter uses the range of the constituent in the calibration sample set (R) divided by the standard error of prediction (SEP) to validate the models. Calculation of $R/\text{SEP} \geq 4$ should be considered fair and acceptable for screening, and $R/\text{SEP} \geq 10$ signifies a good and acceptable model for quality control [11]. Finally, the residual predictive deviation (RPD) is defined as the ratio between the standard deviation of the

population and the SEP. Values less than 1.5 can be considered insufficient for most applications, while NIR models with $RPD \geq 2$ are considered excellent [38].

When comparing the three models with ($n = 50$), the best correlations were observed when using samples without storage (Model 1), as all correlations were above 0.8 and RMSEC/RMSEV values were slightly higher than those typically observed with wet chemical uncertainties. In particular, strong correlations were observed with cellulose, lignin, and total ash with $r_{\text{calibration}}$ values of 0.93. The models built with stored materials only (Model 2) showed weaker correlations, especially for the carbohydrate constituents with $r_{\text{calibration}}$ values of 0.52, 0.65, and 0.60 for cellulose, hemicellulose, and total carbohydrates, respectively.

The impact of storage on the stability of the models was implied here as all correlations except extractives statistically decreased ($p = 0.05$) compared to the base model with no storage, while RMSEC/RMSEV values also increased for most constituents. Ideally, RMSEV values should be lower to indicate strong calibrations and therefore instill confidence in the prediction of compositional values. The third model was constructed with samples collected from both new bales and those stored for up to 225 days (Model 3). The calibration correlation values for lignin (0.98) and extractives (0.92) were the highest in this combination model, while cellulose (0.91), hemicellulose (0.74), total carbohydrates (0.76), and ash (0.87) were higher than those found in the storage only model (Model 2) but lower than Model 1. Validation correlations ($r_{\text{validation}}$) were slightly lower than the calibration correlations for all of Model 3, showing the stability of the model's predictive capability using the cross-validation procedure.

While only fifty samples were initially utilized for the calibration sample set following recommendations described in ASTM protocols for the development of Models 1–3, a fourth model was built by doubling the calibration sample set to 100 (Model 4) consisting of equal parts of switchgrass with no storage to aged switchgrass (75–225 days) in an attempt to alleviate the variation effects introduced to the model with inclusion of storage samples. In Model 4, the calibration correlations for cellulose and lignin slightly decreased compared to the smaller calibration model (Model 3) from 0.91 to 0.90, and 0.98 to 0.95, respectively, but none were found to be significant. Correlations for all other constituents increased, however, only correlations for ash and total carbohydrates were shown to be statistically improved over Model 3. The RMSEV for Model 4 indicated that the models were comparable to switchgrass NIR/PLS models that have been previously described, as the RMSEV for cellulose (1.4%) was somewhat higher than the reported value (glucose, 0.78%) using a model composed of approximately 110 switchgrass samples, while lignin (0.7%) was slightly lower than the values in the previously reported models (Klason lignin, 2.06%) [7]. The R/SEP values for the cellulose (10.7) and lignin (11.6) components in Model 4 showed that these PLS models are appropriate for quality control screening, with models for hemicellulose (9.8) and total carbohydrates (9.8) also applicable with R/SEP values just outside this range. The values for extractives (8.9) and ash (6.8) were in the lower category for fair and acceptable screening ($R/SEP \geq 4$), but the selection of wavelengths used in this study could be a factor in the reduction of the R/SEP statistic. For example, removal of wavelengths from 1100–1400 nm has been shown to reduce calibration errors for extractives, lignin, and hemicellulose [1]. Additionally, the RPD values for the cellulose (2.1), lignin (2.8), extractives (2.5), and ash (3.5) components in Model 4 were found to be excellent, while models for hemicellulose (1.6) and total carbohydrates (1.8) were sufficient for most applications.

In addition to cross-validation of the models, it is important to test their predictive capabilities using an independent validation sample set. Ten validation samples, equal to a total of 10% of the calibration sample set, were used to monitor the performance of the models. These samples were evenly distributed along the range of the model, both by constituent value, storage type, and time (Table 4). To better understand the accuracy of the model predictions, the correlation coefficient (r) for predicted versus measured (wet chemistry) content was calculated for each model.

Quantification of total carbohydrates is essential for application of switchgrass for bioconversion and production of biofuels. As seen in Table 4, the non-storage model (Model 1) underestimated by an average of 5% when compared to measured sugars originating from cellulose and hemicellulose. The model composed of only stored samples (Model 2) also averaged a -5% bias for sugars content, and an error of approximately -8% for those samples with the highest total carbohydrate content (<72% dry basis). Both models showed very weak correlation between predicted values and those determined in the laboratory. As expected, a better prediction came from the combined model (Model 3), consisting of equal parts of stored samples and fresh material.

Table 3. Comparison of PLS models developed with NIR spectra for cellulose, hemicellulose, lignin, extractives, and ash for switchgrass bales with (75–225 days) and without storage (0 days).

(Model 1) No Storage n = 50	<i>Cellulose</i>	<i>Hemicellulose</i>	<i>Total Carbohydrate</i>	<i>Lignin</i>	<i>Extractives</i>	<i>Ash</i>
$r_{\text{calibration}}$	0.93 ^A	0.80 ^A	0.91 ^A	0.93 ^A	0.81 ^B	0.93 ^A
RMSEC _{calibration}	0.7	0.9	1.1	0.3	1.3	0.2
$r_{\text{validation}}$	0.87 ^a	0.7 ^a	0.78 ^a	0.85 ^a	0.66 ^c	0.8 ^a
RMSEV _{validation}	0.9	1.0	1.7	0.4	1.7	0.3
R/SEP	8.0	5.8	8.0	10.3	5.5	7.3
# PCs	6	5	6	6	6	7
Component range (R, %)	30.3 - 37.8	24.4 - 30.2	55.2 - 68.0	18.2 - 25.4	5.9 - 15.6	1.8 - 5.0
Component mean (%)	36.0	27.4	61.6	20.5	10.8	2.8
RPD	2.0	1.5	1.8	2.3	1.3	1.7
Most Significant Wavelengths	2223, 2191, 2159, 2039, 1999, 1955, 1891, 1779, 1735, 1639, 1547, 1479, 1415, 1339, 1127	2271, 2231, 2211, 2155, 2095, 1975, 1911, 1779, 1763, 1695, 1611, 1515, 1423, 1331, 1255, 1139	2271, 2223, 2159, 2131, 2035, 1955, 1891, 1779, 1735, 1707, 1643, 1543, 1419, 1331	2263, 2195, 2123, 2095, 2051, 1899, 1767, 1671, 1611, 1575, 1507, 1449, 1395, 1187	2287, 2235, 2191, 2159, 2027, 1999, 1959, 1903, 1779, 1735, 1639, 1583, 1555, 1423, 1339, 1263, 1127	2291, 2231, 2251, 2203, 2163, 2123, 2091, 2047, 1791, 1763, 1727, 1619, 1423, 1307

(Model 2) Storage n = 50	<i>Cellulose</i>	<i>Hemicellulose</i>	<i>Total Carbohydrate</i>	<i>Lignin</i>	<i>Extractives</i>	<i>Ash</i>
$r_{\text{calibration}}$	0.52 ^C	0.65 ^B	0.60 ^C	0.87 ^B	0.75 ^B	0.74 ^C
RMSEC _{calibration}	2.0	1.2	2.3	0.5	1.2	0.5
$r_{\text{validation}}$	0.41 ^b	0.30 ^c	0.36 ^b	0.77 ^a	0.67 ^c	0.68 ^b
RMSEV _{validation}	2.1	1.5	2.8	0.7	1.4	0.5
R/SEP	4.3	4.2	4.8	6.4	6.5	6.7
# PCs	2	6	5	5	3	2

Component range (R, %)	36.0 - 45.3	24.3 - 30.8	60.4 - 74.0	20.8 - 25.1	1.8 - 11.0	1.3 - 4.9
Component mean (%)	39.9	27.1	67.0	22.3	5.8	2.5
RPD	1.1	1.0	1.1	1.6	1.5	1.3
Most Significant Wavelengths	2255, 1935, 1787, 1671, 1567, 1439, 1369, 1231	2271, 2215, 2135, 2091, 2055, 1947, 1899, 1871, 1835, 1811, 1783, 1667, 1631, 1451, 1411, 1363, 1299, 1227, 1155	2271, 2255, 2219, 2167, 2075, 2055, 1967, 1947, 1919, 1907, 1871, 1839, 1803, 1735, 1675, 1631, 1567, 1507, 1467, 1415, 1375, 1299	2255, 2083, 1999, 1939, 1895, 1839, 1827, 1799, 1779, 1735, 1675, 1451, 1411, 1375, 1307, 1159	2271, 2215, 2103, 2059, 1919, 1851, 1827, 1783, 1751, 1651, 1423, 1303	2255, 2095, 1935, 1823, 1675, 1591, 1503, 1439, 1371, 1159

(Model 3) Combined n = 50	<i>Cellulose</i>	<i>Hemicellulose</i>	<i>Total Carbohydrate</i>	<i>Lignin</i>	<i>Extractives</i>	<i>Ash</i>
r _{calibration}	0.91 ^A	0.74 ^{AB}	0.76 ^B	0.98 ^A	0.92 ^A	0.87 ^B
RMSEC _{calibration}	1.3	1.0	2.3	0.4	1.4	0.3
r _{validation}	0.84 ^a	0.63 ^b	0.61 ^a	0.94 ^a	0.85 ^b	0.66 ^b
RMSEV _{validation}	1.8	1.1	2.9	0.6	1.8	0.4
R/SEP	8.4	4.9	6.5	10.5	8.3	6.3
# PCs	6	4	5	7	6	7
Component range (R, %)	30.3 - 45.3	24.3-29.9	55.2 - 74.0	18.3 - 24.4	1.8 - 15.6	1.6 - 4.0
Component mean (%)	36.8	27.0	64.3	20.8	8.1	2.7
RPD	2.1	1.3	1.4	2.9	1.9	1.5
Most Significant Wavelengths	2299, 2251, 2235, 2203, 2179, 2039, 1915, 1779, 1719, 1575, 1491, 1331, 1099	2291, 2279, 2207, 2171, 2203, 2159, 2075, 1995, 2095, 2051, 1963, 1895, 1779, 1719, 1643, 1779, 1771, 1543, 1427	2299, 2259, 2235, 2211, 2203, 2159, 2075, 1995, 1895, 1779, 1719, 1643, 1483, 1431, 1331, 1195, 1099	2291, 2255, 2151, 2095, 2051, 1963, 1899, 1779, 1727, 1675, 1587, 1503, 1399, 1331, 1099	2235, 2299, 2151, 2031, 1927, 1879, 1799, 1779, 1727, 1639, 1575, 1479, 1383, 1267, 1155	2255, 2159, 2011, 1907, 1779, 1723, 1671, 1587, 1487, 1399, 1259, 1171, 1099

(Model 4) Combined n = 100	<i>Cellulose</i>	<i>Hemicellulose</i>	<i>Total Carbohydrate</i>	<i>Lignin</i>	<i>Extractives</i>	<i>Ash</i>
$r_{\text{calibration}}$	0.90 ^A	0.86 ^A	0.89 ^A	0.95 ^A	0.94 ^A	0.96 ^A
RMSEC _{calibration}	1.2	0.8	1.8	0.6	1.5	0.5
$r_{\text{validation}}$	0.80 ^a	0.79 ^a	0.84 ^b	0.93 ^a	0.92 ^a	0.96 ^a
RMSEV _{validation}	1.4	0.9	2.1	0.7	1.7	0.7
R/SEP	10.7	9.8	9.8	11.6	8.9	6.8
# PCs	5	8	7	7	7	4
Component range (R, %)	30.9 - 45.5	21.8 - 30.8	53.5 - 74.0	16.6 - 25.1	1.8 - 16.9	1.3 - 4.1
Component mean (%)	36.9	26.6	63.2	20.4	9.4	2.7
RPD	2.1	1.6	1.8	2.8	2.5	3.5
Most Significant Wavelengths	2207, 2179, 2095, 1995, 1915, 1779, 1731, 1671, 1583, 1479, 1343, 1175	2231, 2171, 2091, 2033, 1887, 1771, 1691, 1635, 1511, 1439, 1363, 1327, 1191	2231, 2207, 2179, 2087, 1887, 1779, 1719, 1603, 1515, 1363, 1311, 1163	2259, 2175, 2055, 1867, 1779, 1735, 1671, 1587, 1427, 1275	2255, 2179, 2095, 1887, 1779, 1731, 1667, 1583, 1475, 1387, 1279	2259, 2179, 2095, 1995, 1907, 1779, 1731, 1667, 1583, 1467, 1279

- Letter assignments correspond to confidence intervals calculated around the Fisher z-transform ($z = \text{atanh}(R)$) of each correlation coefficient ($p = 0.05$). Model correlations compared by column for each biomass constituent for calibration (capital) and validation (lowercase).

Table 4. Predicted versus measured content biomass constituents using four versions of PLS modeling: (1) no storage bales, (2) storage bales only, (3 and 4) combined model of equal parts fresh and stored switchgrass, (n = 50 or n = 100) with correlation coefficient (r) between the NIR-predicted compositional values versus the standard method (wet chemistry). Standard deviation ().

Storage Time (days)	Cellulose					Hemicellulose					Total Carbohydrates				
	Model 1	Model 2	Model 3	Model 4	Wet Chemistry	Model 1	Model 2	Model 3	Model 4	Wet Chemistry	Model 1	Model 2	Model 3	Model 4	Wet Chemistry
0	32.0 (1.7)	40.5 (2.4)	38.6 (1.3)	40.8 (1.5)	39.7 (0.1)	27.8 (1.7)	26.2 (1.5)	26.4 (0.6)	27.0 (1.0)	26.6 (0.1)	59.9 (2.9)	66.6 (2.8)	66.4 (2.2)	66.4 (2.3)	66.3 (0.1)
0	35.1 (1.2)	39.3 (2.0)	41.3 (1.4)	40.6 (1.1)	41.2 (0.1)	27.8 (1.3)	27.8 (1.2)	26.5 (1.0)	26.1 (0.9)	26.5 (0.1)	64.6 (2.2)	67.5 (3.2)	66.2 (2.2)	67.7 (2.4)	67.7 (0.0)
75	37.0 (1.2)	40.7 (2.0)	40.4 (1.5)	39.8 (1.3)	39.4 (0.5)	28.8 (1.3)	27.2 (1.3)	26.1 (0.7)	26.3 (1.3)	25.6 (0.1)	66.6 (2.0)	68.3 (1.6)	66.5 (2.2)	64.4 (2.9)	65.0 (0.4)
75	31.9 (2.1)	40.3 (2.0)	44.7 (1.4)	44.9 (1.3)	45.5 (0.2)	26.9 (2.3)	26.6 (1.3)	27.0 (1.5)	27.9 (1.3)	27.3 (0.1)	58.6 (3.7)	66.7 (2.2)	71.7 (1.9)	71.8 (2.1)	72.7 (0.1)
75	32.2 (2.2)	40.5 (1.0)	37.4 (1.2)	37.9 (1.5)	38.0 (0.3)	26.7 (2.3)	26.6 (0.7)	25.8 (0.8)	25.5 (0.9)	24.9 (0.1)	60.7 (4.1)	66.6 (2.1)	63.4 (1.7)	63.7 (2.6)	62.9 (0.2)
150	35.2 (1.7)	38.5 (1.9)	42.8 (1.1)	43.1 (1.4)	45.0 (0.5)	28.7 (1.7)	27.8 (1.4)	28.6 (1.0)	27.9 (1.1)	28.6 (0.2)	63.3 (2.7)	67.2 (2.7)	71.6 (1.6)	72.0 (1.9)	73.7 (0.4)
150	42.5 (1.5)	40.1 (1.8)	42.1 (1.9)	42.3 (2.1)	45.0 (0.3)	29.5 (1.9)	27.0 (1.3)	27.8 (1.6)	27.4 (1.5)	27.1 (0.1)	71.4 (2.1)	69.6 (2.6)	71.6 (2.5)	70.2 (2.1)	72.1 (0.2)
225	32.7 (1.5)	37.4 (1.4)	34.3 (1.7)	36.4 (2.6)	36.4 (0.1)	28.5 (1.6)	28.0 (3.0)	27.5 (1.0)	26.9 (1.1)	26.0 (0.1)	61.8 (2.5)	66.7 (4.7)	63.4 (2.5)	63.7 (2.6)	62.4 (0.1)
225	30.4 (1.6)	38.6 (1.1)	36.6 (1.2)	37.8 (1.5)	35.6 (0.2)	26.1 (1.8)	26.3 (1.4)	27.3 (1.6)	27.1 (1.2)	27.9 (0.1)	57.4 (2.9)	66.1 (2.2)	63.9 (2.1)	64.3 (1.5)	63.5 (0.1)
225	41.3 (1.3)	39.7 (1.9)	38.2 (1.6)	37.6 (2.8)	38.5 (0.2)	29.0 (1.7)	27.1 (2.1)	26.2 (1.5)	26.5 (2.6)	26.6 (0.1)	68.7 (2.2)	69.7 (3.1)	64.2 (4.1)	66.8 (2.8)	65.1 (0.1)
r (Pred vs Meas)	0.34	0.28	0.93	0.94	-	0.01	0.02	0.79	0.84	-	0.23	0.22	0.98	0.98	-

Storage Time (days)	Lignin					Extractives					Ash				
	Model 1	Model 2	Model 3	Model 4	Wet Chemistry	Model 1	Model 2	Model 3	Model 4	Wet Chemistry	Model 1	Model 2	Model 3	Model 4	Wet Chemistry
0	18.2 (0.5)	21.3 (0.6)	21.4 (0.5)	21.4 (0.9)	20.8 (0.4)	10.8 (3.1)	9.1 (1.1)	7.7 (1.4)	7.9 (1.2)	8.1 (0.6)	3.0 (0.5)	3.0 (0.4)	3.1 (0.4)	2.9 (0.7)	3.0 (0.2)
0	19.6 (0.4)	21.5 (0.5)	21.9 (0.4)	21.9 (0.6)	21.1 (0.3)	9.1 (2.4)	6.7 (1.3)	7.8 (0.6)	7.2 (2.6)	7.8 (0.1)	3.0 (0.4)	2.7 (0.4)	2.8 (0.3)	2.6 (0.3)	2.9 (0.1)
75	19.3 (0.4)	21.8 (0.5)	22.1 (0.5)	21.8 (0.8)	22.0 (0.1)	8.1 (2.5)	6.8 (1.4)	6.2 (1.0)	6.5 (2.1)	6.4 (0.7)	2.6 (0.4)	2.1 (0.3)	2.3 (0.4)	1.9 (0.4)	1.9 (0.1)
75	18.4 (0.7)	22.3 (0.5)	22.0 (0.4)	21.8 (0.5)	21.6 (0.2)	9.4 (4.1)	6.1 (1.0)	6.3 (1.5)	6.5 (0.7)	6.8 (0.1)	3.2 (0.5)	2.8 (0.5)	2.6 (0.4)	2.8 (0.3)	2.6 (0.3)
75	18.9 (0.7)	22.9 (0.5)	23.1 (0.3)	22.8 (0.5)	22.9 (0.2)	9.8 (4.2)	5.2 (1.2)	5.7 (0.6)	5.3 (0.7)	6.0 (0.6)	3.1 (0.6)	2.9 (0.3)	2.6 (0.3)	2.4 (0.3)	2.0 (0.0)
150	18.9 (0.5)	22.2 (0.6)	22.9 (0.3)	23.1 (0.7)	23.4 (0.1)	6.8 (3.1)	4.4 (1.2)	4.2 (0.8)	4.1 (0.9)	4.2 (0.4)	3.2 (0.4)	2.1 (0.4)	2.2 (0.3)	1.8 (0.3)	1.9 (0.1)
150	20.0 (0.6)	21.2 (0.5)	20.9 (0.6)	21.0 (0.7)	20.9 (0.2)	3.7 (3.2)	5.9 (2.1)	6.9 (0.9)	7.1 (2.1)	8.4 (0.1)	1.5 (0.4)	1.2 (0.6)	1.6 (0.5)	1.6 (0.3)	1.7 (0.1)
225	19.0 (0.5)	21.7 (0.7)	21.3 (0.5)	21.3 (0.7)	20.6 (0.1)	9.8 (2.8)	8.7 (1.8)	7.9 (3.6)	7.9 (2.8)	8.2 (0.3)	3.6 (0.4)	2.2 (0.9)	2.1 (0.4)	2.2 (0.3)	1.7 (0.1)
225	19.2 (0.5)	22.7 (0.6)	22.2 (0.4)	22.0 (0.7)	22.7 (0.2)	9.2 (2.9)	5.7 (1.4)	4.2 (1.4)	4.4 (1.8)	4.5 (0.3)	4.1 (0.4)	2.7 (0.5)	3.3 (0.4)	3.2 (0.3)	2.9 (0.1)
225	20.2 (0.5)	21.1 (1.4)	20.4 (0.9)	21.4 (1.1)	20.3 (0.4)	5.8 (2.4)	4.6 (2.1)	5.2 (1.7)	5.4 (0.8)	6.0 (0.3)	3.9 (0.4)	1.1 (0.6)	1.6 (0.7)	1.5 (0.3)	1.6 (0.2)
r (Pred vs Meas)	0.32	0.83	0.91	0.89	-	0.05	0.71	0.96	0.97	-	0.21	0.77	0.90	0.89	-

Model calibration with (1) senesced field switchgrass with no storage, n = 50; (2) BT3 switchgrass bales with storage, n = 50; (3) both senesced field switchgrass and BT3 switchgrass bales under storage, n = 50 and (4) n = 100.

A strong relationship was found between the predicted values and those measured in the laboratory for this model with a correlation coefficient of 0.98. Interestingly, the correlations were not significantly improved by increasing the calibration sample set size in Model 4 for either total carbohydrates (0.98) or cellulose (0.94 versus 0.93). When considering quantification of lignin for the production of bio-based products, the non-storage model (Model 1) under predicted from 0.5–20% of the lignin content and revealed no correlation between predicted and measured values for the validation set. Incorporation of storage data into the analysis improved the prediction of lignin content to $r = 0.83$ for Model 2, but the correlation coefficient slightly decreased from 0.91 to 0.89 when doubling the size of the combined calibration set, showing no improvement between Models 3 and 4. Model 1 revealed no correlation between the predicted and measured extractives content. This was most likely due to the range of extractives in the stored switchgrass falling below the component range included in the calibration sample set. The combination models (Models 3 and 4) predicted extractives content more accurately with values offset no more than 1.3% from the standard measured chemical data. Without the inclusion of storage samples, the predicted ash content in the validation samples was overestimated by up to 2.3% of the sample's total composition, or approximately 150% of the total ash content (Model 1), which could prove detrimental to and severely impact thermochemical processes that require specific ranges in ash content. Only Models 3 and 4 demonstrated acceptable correlations for the predicted ash concentrations with $r = 0.90$ and 0.89 , respectively. However, the ash model has a bias for low concentrations ($< 2.0\%$) as samples stored for 225 days were still overestimated by up to 30% of the measured ash content. While these differences comprise a small percent of the total composition ($+0.5 \text{ wt}\%$), this error could have a large effect on thermochemical conversion processes by allowing for possible catalytic side reactions [39,40].

The independent validation samples were generally well predicted with errors equal to or slightly higher than those obtained with standard wet chemistry. There were little differences in predicted values between Models 3 and 4, showing that 50 samples efficiently characterized the switchgrass, especially for screening of feedstock potential. Only quantification of the hemicellulose component was a challenge, with correlation coefficients between predicted and measured values only reaching 0.84 (Model 4). This was most likely due to the selective degradation of the hemicellulose during storage contributing to the instability of the bale and decreasing the correlation between the NIR spectra and wet chemistry analysis. Overall, these data indicate that NIR-PLS modeling can be used to predict the chemical composition of switchgrass as part of a long-term storage study only when the models are developed with both fresh and stored samples.

There are many sources of variability in the construction of a biomass compositional model. In particular, the natural variation in the concentration of chemical constituents in the plant is a function of variety, anatomy, age, and environmental conditions. While some uncontrollable factors (harvesting variations and mechanical damage, soil types, temperature, precipitation, etc.) may contribute to differences observed in the feedstock material, the impact of anaerobic storage and exposure time most likely had a greater influence on the composition of the feedstock material. Biomass conversion economics are driven by the quality of the biomass used in the process; therefore accurate characterization of stored feedstocks is critical to reduce variability. Alternatively, high-throughput evaluation techniques such as NIR can be used to rapidly determine the potential of specific bales based on suitability for a specific conversion process. The process has potential for

even increased high-throughput capabilities with possible integration into an online processing management system without extensive sample preparation [41–44].

Through analysis with NIR/PLS, it was observed that each spectrum represented a unique chemical fingerprint of the biomass; therefore, any changes to the switchgrass chemistry with storage degradation and aging will be reflected in the NIR profiles. All storage treatments included in this study resulted in higher carbohydrates and lignin, with lower extractives and ash in the feedstock material when compared to switchgrass with no storage, making these bales ideally suited for both biochemical and thermochemical conversion processes.

4. Conclusion

The appropriate and adequate number of samples to build a successful PLS calibration model should be carefully investigated. As demonstrated in this study, few statistical differences were observed after doubling the number of samples in the calibration sample set (50 versus 100) for a model with equal parts material with no storage and aged switchgrass. However, the error associated with the predictions decreased to values closer to those found using traditional wet chemistry methods, allowing for rapid analysis that is comparable to standard but time-consuming methods performed in a laboratory. Continued evaluation of novel baling, transport, and storage techniques requires analysis of large sample volumes possessing chemical signatures that differ from traditional bales. Stored switchgrass bales will most likely have ongoing microbial degradation at the time of sampling. Overall, this apparent instability caused by degradation or aging resulted in lower correlations between the NIR spectra and the measured chemical composition values with time, demonstrating that stored switchgrass bales are not characterized well by NIR techniques when the models are constructed with only fresh materials. Therefore, to overcome the challenges described here, continued evaluation of high-throughput spectral analysis techniques is required for implementation of these technologies into modern feedstock supply facilities.

Acknowledgements

This project was partially funded by the University of Tennessee, Institute of Agriculture Innovation Grant and Southeastern SunGrant Center, and the US Department of Agriculture-funded Integrated Biomass Supply Systems center (Grant number 2011-68005-30410). The authors also acknowledge the professional assistance by Dr. Samuel Jackson, the VP of Business Development in Genera Energy, Inc. The technical assistance by Mr. Wes Whitehead at Schuster Engineering UK Limited and Mr. Jose Falconi at Schuster Engineering USA are also greatly appreciated. In addition, we thank research assistance by Dr. Vivian Zhou and graduate students in the Department of Agricultural and Resource Economics.

Conflict of Interest

All authors declare no conflict of interest in this paper.

References

1. Sanderson MA, Agblevor F, Collins M, et al. (1996) Compositional analysis of biomass feedstocks by near infrared reflectance spectroscopy. *Biomass Bioenerg* 11: 365–370.
2. Moore KA, Owen NL (2001) Infrared spectroscopy studies of solid wood. *Appl Spectros Rev* 36: 65–86.
3. Kelly SS, Jellison J, Goodell B (2002) Use of NIR and MBMS coupled with multivariate analysis for detecting the chemical changes associated with brown-rot biodegradation of spruce wood. *FEMS Microbiol Lett* 209: 107–111.
4. Hames BR, Thomas SR, Sluiter AD, et al. (2003) Rapid biomass analysis : new tools for compositional analysis of corn stover feedstocks and process intermediates from ethanol production. *Appl Biochem Biotech* 105–108: 5–16.
5. Starks JP, Zuli D, Phillips WA, et al. (2006) Development of canopy reflectance algorithms for real-time prediction of bermudagrass pasture biomass and nutritive values. *Crop Sci* 46: 927–934.
6. Labbé N, Ye XP, Franklin JA, et al. (2008) Analysis of switchgrass characteristic using near-infrared spectroscopy. *BioResources* 3: 1329–1348.
7. Vogel KP, Dien BS, Jung HG, et al. (2011) Quantifying actual and theoretical ethanol yields for switchgrass strains using NIRS analyses. *Bioenerg Res* 4: 96–110.
8. Shenk JS, Workman JJ, Westerhaus MO (2001) Application of NIR spectroscopy to agricultural products, In *Handbook of Near-infrared Analysis*, 2 eds., DA Burns, and EW Ciurczak, New York : Marcel Dekker, Inc., 383–431.
9. Wolfrum EJ, Sluiter AD (2009) Improved multivariate calibration models for corn stover feedstock and dilute-acid pretreated corn stover. *Cellulose* 16: 567–576.
10. Nkansah K, Dawson-Andoh B, Slahor J (2010) Rapid characterization of biomass using near infrared spectroscopy coupled with multivariate data analysis: Part 1 yellow-poplar (*Liriodendron tulipifera* L.). *Bioresource Tech* 101: 4570–4576.
11. Monono EM, Haagensohn DM, Pryor SW (2012) Developing and evaluating NIR calibration models for multi-species herbaceous perennials. *Industrial Biotech* 8: 285–292.
12. Templeton DW, Scarlata CJ, Sluiter JB, et al (2010) Compositional analysis of lignocellulosic feedstocks. 2. Method uncertainties. *J Agric Food Chem* 58: 9054–9062.
13. English BC, De La Torre Ugarte DG, Walsh ME, et al. (2006) Economic competitiveness of bioenergy production and effects on agriculture of the southern region. *J Agric Appl Econ* 38: 389–402.
14. Mooney DF, Larson JA, English BC, et al. (2012) Effect of dry matter loss of profitability of outdoor storage of switchgrass. *Biomass Bioenerg* 44: 33–41.
15. Yu TE, English BC, Larson JA, et al. (2015) Influence of particle size and packaging on storage dry matter losses for switchgrass. *Biomass Bioenerg* 73: 735–144.
16. Chaoui H, Eckhoff SR (2014) Biomass feedstock storage for quantity and quality preservation, In: *Engineering and science of biomass feedstock production and provision*, 1ed. New York: Springer Science+Business Media, 165–193.
17. Kenney KL, Smith WA, Gresham GL, et al. (2013) Understanding biomass feedstock variability. *Biofuels* 4: 111–127.
18. Shinnors KJ, Boettcher GC, Much RE, et al. (2010) Harvest and storage of two perennial grasses as biomass feedstocks. *Trans ASABE* 53: 359–370.

19. Shah A, Darr MJ, Webster K, et al. (2011) Outdoor storage characteristics of single-pass large square corn stover bales in Iowa. *Energies* 4: 1687–1695.
20. Smith WA, Bonner IJ, Kenney KL, et al. (2013) Practical considerations of moisture in baled biomass feedstocks. *Biofuels* 4: 95–110.
21. Darr MJ, Shah A (2012) Biomass storage: An update on industrial solutions for baled biomass feedstocks. *Biofuels* 3: 321–332.
22. Larson JA, Yu TE, Boyer CN, et al. (2010) Cost evaluation of alternative switchgrass producing, harvesting, storing, and transporting systems and their logistics in the southeastern USA. *Agric Fin Rev* 70: 184–200.
23. Larson JA, Yu TE, English BC, et al. (2015) Effect of particle size and bale wrap on storage losses and quality of switchgrass. SunGrant Conference, Auburn, AL.
24. American Society of Testing Materials (ASTM), ASTM in West Conshohocken, PA: From Standard practices for infrared multivariate quantitative analysis. Available from: http://enterprise.astm.org/filtrexx40.cgi?+REDLINE_PAGES/E1655.htm. Accessed 15 December 2014.
25. National Renewable Energy Laboratory (NREL). (2008) Determination of extractives in biomass. National Renewable Energy Lab. NREL/TP-510-42619.
26. National Renewable Energy Laboratory (NREL). (2008) Determination of total solids in biomass and total dissolved solids in liquid process samples. National Renewable Energy Lab. NREL/TP-510-42621.
27. National Renewable Energy Laboratory (NREL). (2008) Determination of structural carbohydrates and lignin in biomass, National Renewable Energy Lab. NREL/TP-510-42618.
28. National Renewable Energy Laboratory (NREL). (2008) Determination of sugars, byproducts, and degradation products in liquid fraction process samples. National Renewable Energy Lab. NREL/TP-510-42623.
29. Labbé N, Ye XP, Franklin JA, et al. (2008) Analysis of switchgrass characteristics using near infrared spectroscopy. *Bioresources* 3: 1329–1348.
30. Çelen İ, Harper D, Labbé N (2008) A multivariate approach to the acetylated poplar wood samples by near infrared spectroscopy. *Holzforschung* 62: 189–196.
31. Taylor A, Labbé N, Noehmer A (2011) NIR-based prediction of extractives in American white oak hardwood. *Holzforschung* 65: 185–190.
32. Martens H, Naes T (1989) *Multivariate calibration*. Chichester: Wiley.
33. Esbensen KH (2002) *Multivariate data analysis in practice: An introduction to multivariate data analysis and experimental design*, 5 eds. Oslo, Norway: CAMO Process AS.
34. Rials TG, Kelley SS, So CL (2002) Use of advanced spectroscopic techniques for predicting the mechanical properties of wood composites. *Wood Fiber Sci* 34: 398–404.
35. Wold S, Sjöström M, Eriksson L (2001) PLS-regression: A basic tool of chemometrics. *Chemometr Intell Lab.* 58: 109–130.
36. Tsuchikawa S, Siesler HW (2003) Near-infrared spectroscopic monitoring of the diffusion process of deuterium-labeled molecules in wood. Part I: softwood. *Appl Spectrosc* 57: 667–674.
37. Goulden CH (1952) *Methods of statistical analysis*, 2 eds. New York: John Wiley.
38. Garde-Cerdán T, Lorenzo C, Alonso GL, et al. (2010) Employment of near infrared spectroscopy to determine oak volatile compounds and ethylphenols in aged red wines. *Food Chem* 119: 823–828.

39. Arvelakis S, Koukios EG (2013) Critical factors for high temperature processing of biomass from agriculture and energy crops to biofuels and bioenergy. *WIREs Energy Environ* 2: 441–455.
40. Evans RJ, Milne TA (1987) Molecular characterization of the pyrolysis of biomass. 1. Fundamentals. *Energ Fuel* 1: 123–137.
41. Everard CD, Fagan CC, McDonnell KP (2012) Visible-near infrared spectral sensing coupled with chemometric analysis as a method for on-line prediction of milled biomass composition pre-pelletising. *J Near Infrared Spectrosc* 20: 361–369.
42. Lestander TA, Johnsson B, Grothage M (2009) NIR techniques create added values for the pellet and biofuel industry. *Bioresource Tech* 100: 1589–1594.
43. Hayes DJM (2012) Development of near infrared spectroscopy models for the quantitative prediction of the lignocellulosic components of wet *Miscanthus* samples. *Bioresource Tech* 119:393–405.
44. Sluiter A, Wolfrum E (2013) Near infrared calibration models for pretreated corn stover slurry solids, isolated and *in situ*. *J Near Infrared Spectrosc* 21: 249–257.



AIMS Press

© 2015 Lindsey M. Kline, Nicole Labbé, et al. licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)