COMPOSITIONAL ANALYSIS AND CLASSIFICATION OF MISCANTHUS USING FOURIER TRANSFORM NEAR INFRARED SPECTROSCOPY

BY

DANIEL A. WILLIAMS

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Agricultural and Biological Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Master's Committee:

Assistant Professor Mary-Grace C. Danao, Chair and Director of Research
Associate Professor Kent D. Rausch
Professor Emeritus Marvin R. Paulsen

ABSTRACT

Miscanthus × *giganteus* is a woody rhizomatous C4 grass species that is a high yielding lignocellulosic material for energy and fiber production. The cellulose and hemicellulose fractions of Miscanthus can be converted into energy and chemicals through biological conversion. Since only a fraction of the biomass can be converted into chemical energy, bioethanol yields per unit mass of biomass are directly proportional to the composition of the biomass, which can vary due to age, stage of growth, growth conditions, and other factors. It is advantageous to know these variations prior to conversion so that enzyme mixtures, yeast strains, and process control parameters can be adjusted accordingly to maximize yields. Knowing the composition at earlier stages of the supply chain can also help in the development of quality-based valuations which incentivize farmers and suppliers to implement best management practices to ensure a uniform and consistent supply system.

Therefore, in this study, the variability of composition of Miscanthus bales stored under a variety of conditions for a period of 3 to 24 months was described, along with the compositional variability of its botanical fractions. High throughput assays based on Fourier transform near infrared (FT-NIR) spectroscopy, partial least squares regression (PLSR), and linear discriminant analyses (LDA) to provide quantitative and qualitative measures of Miscanthus composition were developed. Results showed large variations (mean ± S.D.) in glucan (40.4 ± 2.70%), xylan (20.7 ± 1.50%), arabinan (1.90 ± 0.40%), acetyl (2.84 ± 0.28%), lignin (20.5 ± 1.40%), ash (2.60 ± 1.80%), and extractives (5.60 ± 0.86%) - contents were observed for samples that were collected from Miscanthus bales stored indoors, under roof, outdoors with tarp cover, and outdoors without tarp cover for

3 to 24 months after harvest and baling. There was also a wide variability for all components: glucan, 32.2 to 46.1%; xylan, 20.9 to 25.3%; arabinan, 0.0 to 6.1%; lignin, 18.7 to 25.5%; and ash, 0.4 to 8.9%, observed in botanical fractions of Miscanthus. The ranges in composition were comparable to corn stover botanical fractions. While the sum of glucan, xylan, and arabinan contents for the rind, pith, and sheath fractions were not different from each other, the variations across some botanical fractions were significant with the blade having lowest glucan, lowest lignin, and highest ash contents.

PLSR models were developed to predict glucan, xylan, lignin, and ash contents in Miscanthus bale samples with RPD values of 4.86, 4.08, 3.74, and 1.71, respectively. The geometric mean particle size ranged from 0.36 to 0.49 mm, with the smallest size observed with samples from bales stored outdoors for 17 months and the largest size observed with samples from bales stored outdoors with a tarp cover for 5 months. On average, PLSR predictions of glucan, arabinan, and lignin content were not sensitive to the particle size of ground Miscanthus, but predictions of xylan and ash content were. The predicted xylan content using the non-sieved samples was lower than those for sieved samples and ash levels increased with decreasing particle size.

When the PLSR models were coupled with LDA to classify the Miscanthus samples based on their glucan, lignin, and ash contents, the best classification results were found with the PLS-DA lignin model. While the PLSR and PLS-DA models developed in this study were based on a small sample size, the approaches presented in this study demonstrated FT-NIR spectroscopy is a practical tool for screening biomass at different stages of the supply chain, making the delivery of consistent feedstock to conversion facilities year round a realistic possibility.

ACKNOWLEDGMENTS

This work was partially funded by the Energy Biosciences Institute through the program titled, “Engineering Solutions for Biomass Feedstock Production.” I thank Camo Software, Inc. for providing Unscrambler® X and technical support, and the following individuals for their technical assistance: Stefan Bauer, Shih-Fang Chen, Xiangwei Chen, Joshua Jochem, Gary Letterly, Tim Mies, and my thesis committee members. I would also like to thank my parents for their constant love and support.

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION

It has been estimated that to replace 30 percent of our current energy demand with fuels from an agricultural resource the United States will need to produce one billion tons of material annually. Of the one billion tons, 377 million tons need to be from a dedicated lignocellulosic feedstock such as *Miscanthus* × *giganteus* (Perlack and Stokes, 2005). In order to make a technology of this magnitude feasible, many obstacles need to be overcome such as the ability to supply the processing facilities with a steady stream of material with a known composition to convert throughout the year. Currently this presents many challenges as the material will need to be moved multiple times as it passes through the feedstock supply chain (Figure 1.1).

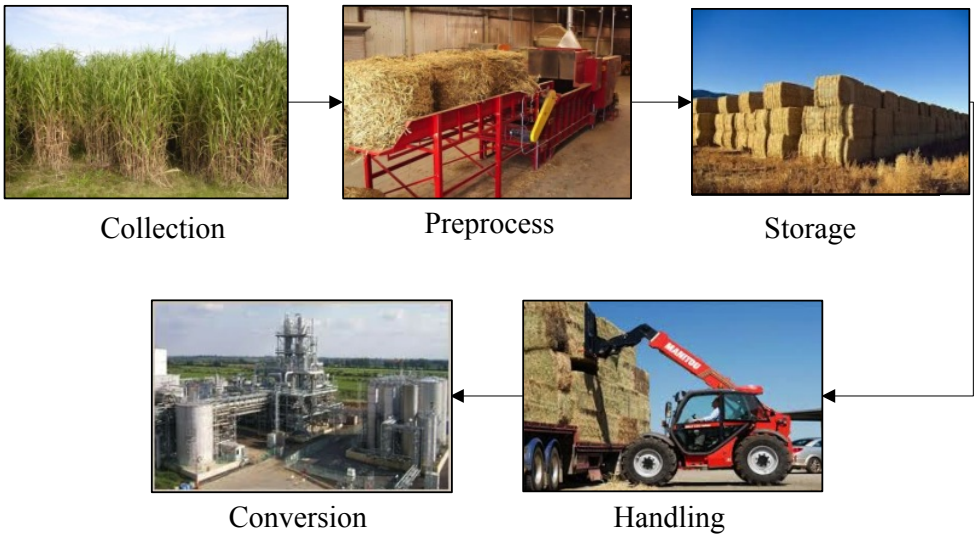


Figure 1.1. Biomass supply chain. Adapted from Aden et al. (2002). Images are from http://www.biogreentech.com; http://www.rotochopper.com; http://www.123rf.com; http://www.feedcentral.com.au; www.praj.ne

Since harvesting takes place during a short period of the year while processing facilities are operational year round, storage of biomass is imminent. The quality will not improve from the last day of growth and the biomass is expected to undergo dry matter loss and quality losses during storage. The main components of lignocellulosic materials

(cellulose, hemicellulose, and lignin) will degrade through multiple pathways. The theoretical yield of bioenergy is directly proportional to composition, so knowledge of the composition at any stage of the supply chain is desirable. Biomass feedstock composition affects efficiency and optimization of conversion processes. For example, high lignin is preferred for thermochemical conversion because it has a higher heating value compared to the structural carbohydrates (Hodgson et al., 2010); however, low lignin and high structural carbohydrates are desired for biochemical conversion because only the structural carbohydrates can be converted and lignin interferes with pretreatment (Claassen et al., 1999). High ash content is not desired in thermochemical and biochemical conversions processes since ash cannot be used and, more importantly, it can inhibit catalysis and cause slagging in pyrolysis (Kenney et al., 2013).

Considering these challenges, it would be advantageous (1) to know the variation in biomass composition and what factors cause these variations and (2) to have the ability to determine chemical composition of the biomass that is being produced, purchased, and processed, and be able to classify and utilize variations in optimizing conversion processes. While compositional variations can be determined with conventional wet chemistry methods, these methods are not readily available or practical in the field as they are time consuming, destructive, and usually require extensive sample preparation, expensive laboratory equipment, and well trained personnel. One alternative to current wet quantification methods is to utilize near infrared (NIR) spectroscopy coupled with multivariate analysis. NIR has been used in the agricultural and food industries for years, from analysis of moisture and protein content in wheat (Manley et al., 2002), to the compositional determination of biomass, such as cornstover, switchgrass, and Miscanthus

in plant breeding studies (Ye et al., 2008; Templeton et al., 2009; Liu et al., 2010; Hayes, 2012; Haffner et al., 2013). NIR spectroscopy has also been used to provide near real time assessment of moisture content and the amount of active ingredient in the final product for quality control in the pharmaceutical industry (Blanco et al., 1998).

In this study, Fourier transform near infrared (FT-NIR) spectroscopy was used as the basis for developing a high throughput assay for quantifying and classifying *Miscanthus* × *giganteus* based on its chemical composition after storage. The specific objectives were to:

Objective 1. Describe variability in composition (glucan, xylan, arabinan, lignin, ash, acetyl, and extractives content) of Miscanthus samples from bales that were stored under a variety of conditions for a period of 3 to 24 months.

Objective 2. Determine variability in composition of different botanical fractions (rind, node, pith, sheath, and blade) of Miscanthus.

Objective 3. Develop partial least squares regression (PLSR) models to predict composition of Miscanthus based on FT-NIR spectra of bale core samples.

Objective 4. Determine the effects of particle size on FT-NIR spectra of the sample and resulting predicted composition using PLSR models from Objective 3.

Objective 5. Classify Miscanthus bale core samples using the PLSR models from Objective 3 and linear discriminant analysis (LDA).

CHAPTER 2. LITERATURE REVIEW

2.1. U.S. bioenergy demand

In 1970 the Clean Air Act was implemented, "… to foster the growth of a strong American economy and industry while improving human health and the environment (Public Law 88-206)." While this Act covered a wide range of technologies to combat environmental and health concerns, a portion of the act was to produce energy from renewable sources. In doing so, it would allow America to become less dependent on foreign fossil fuels while balancing the carbon cycle.

To achieve this goal, 209 bioethanol plants have been constructed since 1999 to produce ethanol from glucose, which has been derived mainly from cornstarch, a food-based feedstock. In 2011, the 209 bioethanol plants produced 13.9 billion gallons of ethanol (RFA, 2012) and, by 2022, the U.S. has a goal of producing 36 billion gallons of biofuel per year according to the Clean Air Act (Public Law 88-206) and the Energy Independence and Security Act of 2007 (Public Law 110-140).

As demands for bioenergy continue to increase, it is essential to develop technologies for a diverse set of feedstocks and not rely solely on food-based materials. An alternative to food based feedstocks are lignocellulosic materials. Lignocellulosic materials account for 50 percent of the world's biomass and are composed of three main components: cellulose, lignin, and hemicellulose (Claassen et al., 1999). In lignocellulosic biofuel production, the cellulose and hemicellulose can be converted to a biofuel while the lignin and smaller constituents are typically waste byproducts of the process (Limayem and Ricke, 2012). Berndes et al. (2001) studied both food-based and lignocellulosic feedstocks and concluded that biofuel production from food-based

feedstocks would not be feasible on a large enough scale that is being asked for by the US government, but production from lignocellulosic feedstocks would be.

The U.S. Department of Energy (DOE) released a study in 2005 titled, "Biomass as a Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasibility of a Billion-Ton Annual Supply", which is also referred to as "the billion ton study" or 2005 BTS (Perlack et al., 2005). The 2005 BTS was a strategic analysis to determine if U.S. agriculture (e.g., agricultural waste, crop residue, and dedicated herbaceous perennial grasses) and forest resources have the capability to produce at least one billion dry tons of biomass annually in a sustainable manner, which was the amount of biomass needed for bioconversion to meet more than 30% of the U.S. oil consumption. In the analysis, 55 million acres were assessed to produce 377 million dry tons of perennial biomass feedstock a year.

In 2011, DOE released an update to the 2005 BTS. Modifications to the projected biomass supply in the 2005 BTS were reflected in the 2011 BTS, which included the use of fewer forest residues due to a decrease in the paper industry; fewer agricultural residues would be used to preserved soil carbon; and more dedicated bioenergy crops would be used to counteract the loss in the other categories. The new baseline projections for energy crops were set at 400 million dry tons per year, an increase of 23 million tons from the 2005 BTS (Perlack and Stokes, 2011).

2.2. *Miscanthus* × *giganteus*

Since the 2005 BTS, the U.S. government has encouraged research using herbaceous feedstocks, such as *Miscanthus* × *giganteus* and prairie cordgrass, for conversion to ethanol for use as transportation fuel. *Miscanthus* × *giganteus* is a highly

productive, sterile, rhizomatous, C4 perennial grass that is regarded as an ideal feedstock for bioenergy production because of its potential to produce large quantities of biomass with minimal inputs (Jones and Walsh, 2001). In a study conducted at the University of Illinois, Heaton et al. (2008) reported average yields over a three year period of 20.9, 33.4, and 34.6 dry ton/ha, in the north, central and southern parts of the state. Gauder et al. (2012) studied the cultivation of four types of Miscanthus in Germany from 1997 to 2010 and reported plots where *Miscanthus* × *giganteus* had the highest average yields ranging from 12.6 to 14.1 dry ton/ha over the 14 year period, excluding the first two establishment years. By the end of the study, *Miscanthus* × *giganteus* yields had increased to more than 20 dry ton/ha. In addition to high yields and minimal inputs, Miscanthus is also a C4 plant like maize, which are estimated to be 40% more efficient in the photosynthetic process than C3 plants like wheat (Monteith, 1978).

Miscanthus is composed of three main components – cellulose, hemicellulose, and lignin – which account for nearly 40, 20, and 20 % (w/w), respectively, while the balance is composed of organic acids, ash, and extractives (Sanderson et al., 1996). In addition to producing high quantities of materials with desired compositional characteristics, Miscanthus is a perennial, has low nutrient input needs, can utilize existing harvesting equipment for collection, and can be managed using commercially available herbicides if the ground needs to be utilized for food production (Heaton et al., 2004). Compared to other C4 plants, Miscanthus can tolerate lower temperatures, which allows it to utilize the spring and fall months for growth when the temperatures are lower (Naidu and Long, 2004).

2.3. Botanical fractions of Miscanthus

Miscanthus is a type of solid body monocot, having similar structural features to that of corn and sugarcane (Figure 2.1; Głowacka, 2011; Evert, 2006). The plant can be separated into three main parts – the root system, the stalk, and the leaves. The root system makes rhizomes, which travel perpendicularly to the force of gravity and can send out roots and shoots from its nodes (Evert, 2006). Since *Miscanthus* × *giganteus* is a sterile plant, it does not produce seeds and the propagation relies on the rhizomes (Heaton et al., 2008). In Miscanthus production for bioenergy applications, most of what is harvested and utilized are the stalks and leaf structures. The stalk can be further divided into nodes and internodes, while the sheath and the blade make up the leaf structures.

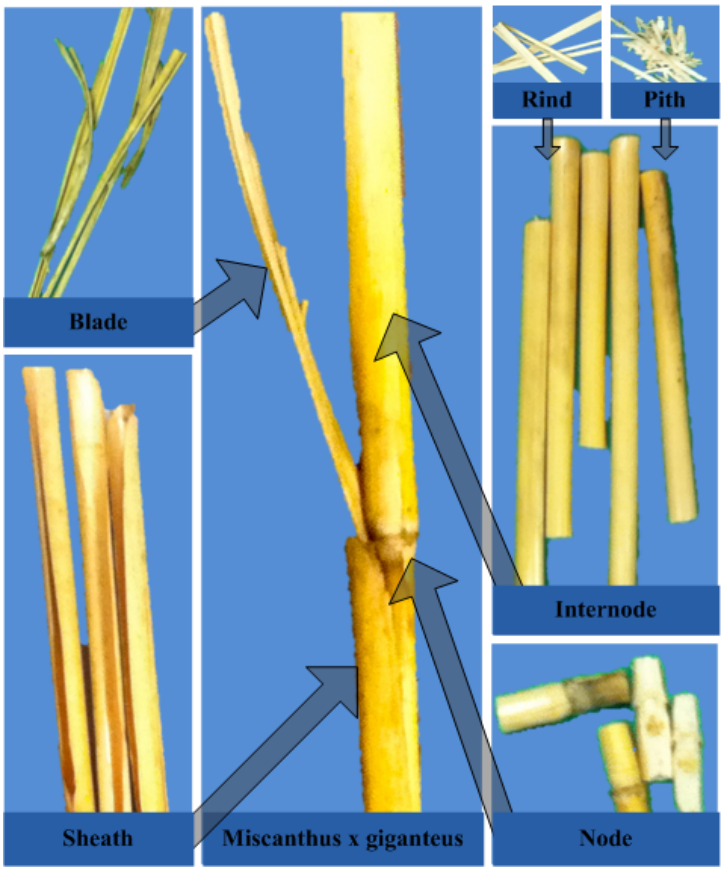


Figure 2.1. Diagram of the stalk and leaf of a *Miscanthus* × *giganteus* stem.

The stalk of the plant can then be broken down into three main tissue systems – the dermal, the vascular, and the fundamental (Figure 2.2; Evert, 2006). The dermal tissue makes up the epidermis, which is the primary outer protective covering of the plant body. The vascular tissue includes the phloem, which conducts food, and the xylem, which conducts water. The dermal and vascular tissues are complex and are composed of many different types of cells. The fundamental, or ground, tissue is the simplest tissue type, usually consisting of only one cell type. Parenchyma cells make up most of the fundamental tissues in non-woody plants and often have highly specialized structures with thicker, harder, lignified walls. However, collenchyma cells have been observed in the fundamental tissue, which are similar to parenchyma calls, but have thicker cell walls.

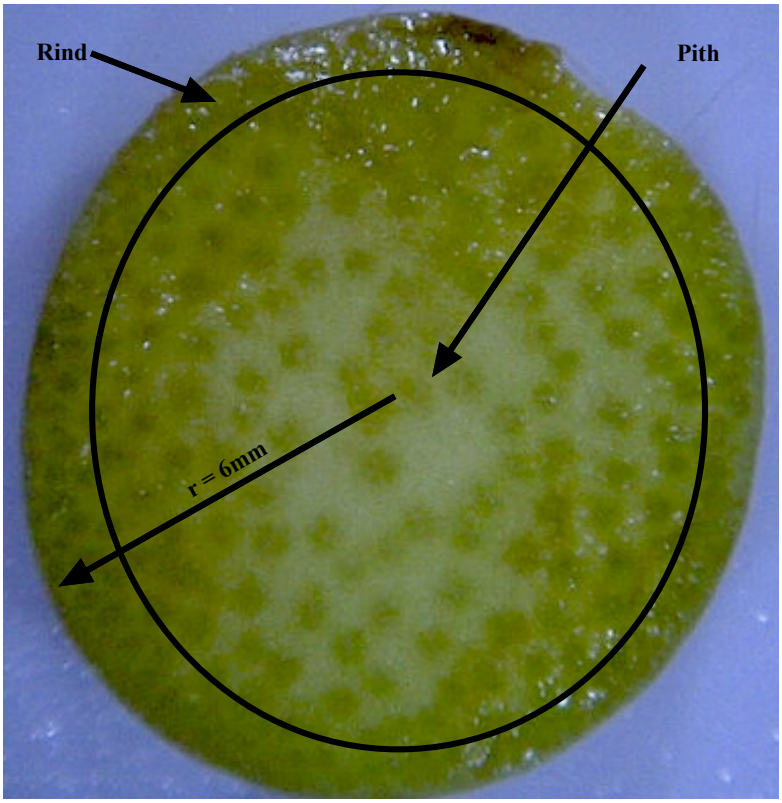


Figure 2.2. Cross section of a *Miscanthus* × *giganteus* stalk, a solid monocot stem. The epidermis, or rind, is on the exterior while the vascular tissue is scattered throughout the middle surrounded by the ground or fundamental tissue.

2.4. Chemical composition of Miscanthus

2.4.1. Cellulose, hemicellulose, and lignin

The principal component of the plant cell walls of Miscanthus and other herbaceous perennial grasses is cellulose, a polysaccharide composed of a chain of (1-4)-β-linked-D-glucan molecules. The cellulose chains tend to form hydrogen bonds in an antiparallel fashion to form microfibrils that range from 4 to 10 nm in diameter, forming a crystalline structure. The microfibrils wind together to form fine threads that coil around each other to form cables called macrofibrils, which are approximately 0.5 μm in diameter (Evert, 2006; Sun, 2010). Due to the stacking and the β-linkage, the cellulose compound is very resistant to chemical and biological degradation (Sun, 2010).

The cellulosic macrofibrils are embedded in a cross-linked matrix of polysaccharides called hemicellulose. Hemicellulose is a general term for a group of noncrystalline polysaccharides that are tightly bound in the cell wall. There are two main types of hemicelluloses: xyloglucans, which have a glucan backbone with xylose, galactose, and fructose branches, and glucuronoarabinoxylans, which have a (1-4)-β-D-xylose backbone. While the composition of the hemicellulose will vary in a plant depending on the tissue, overall, one type of hemicellulose tends to dominate in a specific plant. In Miscanthus, the dominant type is the glucuronoarabinoxylan with the xylan backbone. The xylan backbone has the ability to form hydrogen bonds with the cellulose backbone, covalent bonds with the lignin, and form ester linkages with acetyl units (Sun, 2010).

Surrounding the cellulose hemicellulose matrix are lignin structures. Lignins are phenolic polymers formed from the polymerization of three main monomeric units – the

monolignols *p*-coumaryl, coniferyl, and sinapyl alcohols. Generally, lignins are classified as guaiacyl, which are formed mostly from coniferul and sinapyl alcohols; guaiacyl-syringyl, which are mostly copolymers of coniferyl and sinapyl alcohols; or guaiacyl-syringyl-*p*-hydroxyphenyl lignins, which are formed from all three monomeric units. The exact lignin structure will vary greatly from one species to another, organ tissue, along with the part of the cell wall that is being observed. The final lignin structure ends up covalently linked to the cell wall polysaccharides (Sun, 2010).

The plant cell wall can then be broken down into three main components – the primary wall, the secondary wall, and the middle lamella – all varying in composition; proportions of cellulose, hemicellulose, and lignin; and geometry with respect to the vertical axes of the plant. The final cell wall structure has linear bundles of cellulose, with hemicellulose winding throughout, and lignin filling in the void space (Sun, 2010).

2.4.2. Other components

There are two other main components in lignocellulosic materials, extractives and ash. The National Renewable Energy Laboratory (NREL) has defined extractives as non-structural organic compounds in the biomass that are soluble in water or ethanol (Sluiter, 2005b). Some of the compounds that have been identified in the extractives from wood and straw are resin acids, triglycerides, sterol esters, fatty acids, sterols, fatty alcohols and various phenolic compounds (Sun, 2010).

Ash, on the other hand, is the inorganic material that resides after the sample has been heated to 575°C (Sluiter, 2005a). Kenney et al. (2013) discusses two forms of ash: structural ash, which is from the plant material and was utilized for physiological functions, and introduced ash, which is from materials like soil.

2.4.3. Sources of variation in composition and effects on conversion

The composition of the biomass can vary due to age, stage of growth, growth conditions, and other factors (Perez et al., 2002; Hames et al., 2003). The compositional variations in corn stover fractions were studied as the plant matured by Pordesimo et al. (2005). In their study, soluble solids, lignin, structural glucan, xylan, and protein in the stalk, husk, and leaves were measured with NIR spectroscopy. They found stalk increased in lignin content from 15 to 20% and xylan content from 13 to 23% during the growing season as the soluble solids content (i.e., extractives) decreased from 17 to 2%. The leaves exhibited the highest increases in glucan content, from 17 to 32%; xylan content, from 0 to 23%; and lignin content, from 2 to 19%; while their soluble solid content, too, decreased with time, from 35 to 6%. Similarly, in a study of switchgrass round bales harvested in October 1991 and stored outdoors for 26 weeks, unprotected from the weather, Wiselogel et al. (1996) found significant compositional changes in cellulose, lignin, ash, and extractives contents. These bales were exposed to high rainfall (65 cm) and weathering on the outer layers of the bales was observed. When the test was repeated with switchgrass harvested in August 1992, smaller and less significant compositional changes were observed. Shinners et al. (2007) studied the dry matter loss (DML) of corn stover baled at different moisture contents and stored under various conditions. For wet corn stover (moistures above 35% w.b.), DML after nine months of storage was 2.4% when baled and wrapped but could be as high as 5.4% when chopped and stored in plastic silo bags. For dry corn stover (moistures from 14.6 to 23.0% w.b.), average DML after eight months of storage was 3.3% for round bales stored indoors and 18.8% for round bales stored outdoors. For outdoor storage, DML could be reduced to

10.0% when bales were wrapped with net wrapping but could be as high as 30.4% when bales were left uncovered. More recently, Shah and Darr (2011) studied the dry matter loss (DML) of corn stover bales stored under three conditions (covered with tarp, covered with a breathable film, and indoors) and two initial moisture contents (15-20% w.b. and 30-35% w.b.) for three and nine months. They observed that tarp covered bales fared better than those covered with breathable film. DML for the tarp covered bales were 6 and 11% for three and nine months of storage, respectively, while DML were 14 and 17% for the three and nine months of storage, respectively, for those covered with a breathable film. In all the storage conditions tested, more than half of DML occurred within the first three months of storage and compositional changes (i.e., neutral detergent fiber, acid detergent fiber, and acid detergent lignin contents) were within a narrow range, from 1.4 to 3.9% of each other across the different storage conditions.

With this much variation in composition due to storage conditions alone, decision-making in all steps of the supply chain, from plant breeding, crop management, harvest, transportation, preprocessing (e.g., size reduction, densification) and storage needs to be guided by biomass conversion requirements (Vidal et al., 2011). The cost, quality, and volume of lignocellulosic feedstocks are essential in inventory management and determine the viability of commercial scale bioenergy production.

Conversion facilities would like to receive feedstocks that are consistent, or uniform, in quality, in moisture content, ash content, and convertible carbohydrates so they can operate their chemical pretreatment and conversion processes efficiently (Kenney and Ovard, 2013). For example, in combustion and pyrolysis, the lower the oxygen level and lower the moisture of the material, the higher the heating value of the

material. Since carbohydrates have a high amount of oxygen (500g kg^{-1}) compared to lignin (300g kg^{-1}), higher lignin contents will increases the heating value (Lewandowski and Kicherer, 1997; Hodgson et al. 2010)**.** For biofuel production, the theoretical yield is directly proportional to the cellulose and hemicellulose concentration of the material to be converted (Perez et al., 2002)**.** Ye et al. (2008) suggested that corn stover should be divided into its botanical fractions so that fractions with high lignin content can be used in co-firing while fractions with higher cellulose and hemicellulose are used in fermentation processes.

Tao et al. (2013) demonstrated that the variability in corn stover composition strongly impacted the variability of the minimum ethanol selling price (MSEP) due to the variability in ethanol yields. The corn stover used in their analysis ranged in total carbohydrates from 53 to 64%, which corresponded in decreasing MESP values of $2.50 to $2.05 per gallon. Therefore, it is advantageous to know the composition prior to conversion so that enzyme mixtures, yeast strains, and process control parameters can be adjusted accordingly to maximize yields and lower final product costs.

Knowing the composition at earlier stages of the supply chain can also help in the development of quality-based valuations which incentivize farmers and suppliers to implement best management practices to ensure a uniform and consistent supply system (Kenney et al., 2013). For example, biochemical conversion processes are sensitive to carbohydrate content as the ratio of C5 to C6 sugars and accessibility of these sugars are important in optimizing pretreatment and fermentation conditions (Öhgren et al., 2007; Berlin et al., 2007). In addition to accessibility, C6 sugars are more desirable since most commercial yeast strains that are used for fermentation can only ferment glucose, a C6

sugar, while they cannot ferment C5 sugars such as xylose (Rudolf et al., 2008). Lignin content in biomass represents the recalcitrance of cell walls to saccharification, particularly during enzymatic hydrolysis (Öhgren et al., 2007; Chen and Dixon, 2007). Such a quality-based valuation is needed for biorefineries to enforce best management practices and for biomass to be treated as a traded commodity, like grains and oilseeds, with consistent quality standards or "grades" (Kenney et al., 2013).

2.5. Current methods to determine composition

The National Renewable Energy Laboratory has developed a set of protocols to determine the chemical composition of various types of biomass. The biomass undergoes an extraction process to remove extractives (or nonstructural carbon components), which interfere with the downstream compositional analysis. The samples are first subjected to water extraction followed by an ethanol extraction taking a minimum 22 hours, extraction time (Sluiter et al., 2005b). Following extraction, the structural carbohydrates and lignin content are determined. Samples are hydrolyzed with 72% sulfuric acid at a high temperature (121°C). Liquid and solids are separated by filtration; the solids are used to determine the acid insoluble lignin and ash content while the liquid portion is used to determine the structural carbohydrates using a high performance liquid chromatography (HPLC) and the acid soluble lignin (ASL) is determined with a spectrophotometer. (Sluiter et al., 2005a,b; Sluiter et al., 2011).

2.6. Near infrared (NIR) spectroscopy

Current wet chemistry methods for chemical characterization of biomass feedstock are not applicable for field or inline monitoring because they are expensive, labor-intensive, and cannot provide compositional information in real time for process

control (Ye et al., 2008). Hames et al. (2003) estimated that quantification using NREL standards can cost from $800 to $2,000 per sample and can take up to one week to get results, making the technology not feasible for rapid and cost efficient analysis. One approach to reducing the time and cost of compositional analysis is the development of a high throughput assay based on NIR spectroscopy and a good calibration obtained from multivariate analyses to provide either a quantitative or qualitative measure of biomass composition.

NIR spectroscopy is a type of vibrational spectroscopy that utilizes the optical region ranging from 4,000 to 12,500 cm^{-1} (2,500 to 800 nm). The energy absorbed in this region by a biomass sample corresponds to combinations of the fundamental vibrational transitions along with overtones associated with each chemical bond present in the sample (Blanco and Villarroya, 2002). Depending on which atoms are interacting, different anharmonicities give each compound a unique fingerprint (Theander and Aman, 1984). The NIR region has weaker bands compared to the mid-infrared (MIR) region due to the lower number of excitations to the higher states. While the weaker bands are less informative in the NIR region than the parent bands in the MIR region, it does allow for the sample thickness to vary. Therefore, NIR spectroscopy's advantage over MIR spectroscopy is the ability to use larger sample thickness, which typically translates to less sample preparation, allowing for rapid and less costly analyses (Siesler et al., 2002).

2.6.1. Types of NIR spectrophotometers

There are two main types of spectrophotometers used in NIR spectroscopy: a diffraction grating NIR spectrophotometer and a Fourier transform NIR (FT-NIR) spectrophotometer. In a diffraction grating NIR spectrophotometer, light enters a slit and

is collimated using a collimating mirror (Tkachenko, 2006). The collimating mirror reflects the light towards a diffraction grating. The diffraction grating is used to split the light into its respective wavelength. The diffracted light is then reflected to another collimating mirror and sent towards a point or array detector. The resolution of the diffraction grating spectrophotometer is proportional the spread of the light to the width of the detector window.

FT-NIR spectrophotometers, on the other hand, utilize a Michelson interferometer (Figure 2.3). Light enters the interferometer and is sent straight to a beam splitter where 50% of the light is allowed to pass and the balance is reflected (Tkachenko, 2006). Both the reflected and passed light hit their respective mirror and are reflected back to the beam splitter, with one of the mirrors being fixed and the other being mobile. The light reflected by the mirrors is recombined at the beam splitter. Depending on the difference between pathlengths and wavelength, the light either combines to form the initial signal or exhibits some degree of destructive interference. The final signal is a maximum when the difference between the mirrors and the beam splitter is:

$$d = m\lambda + \frac{\lambda}{2} \qquad \text{[Equation 2.1]}$$

where *d* is the difference in the two mirrors distance from the beam splitter, *m* is the peak number from the central wavelength , and λ is the wavelength of the light.

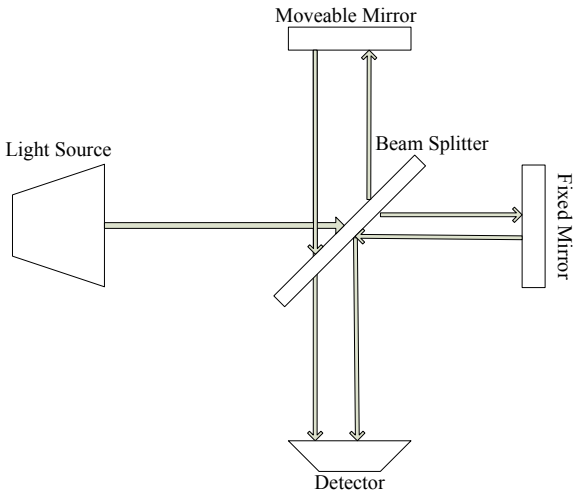


Figure 2.3. Michelson interferometer in a FT-NIR spectrophotometer.

Since the detector sees a sinusoidal pattern, the spectral resolution of the interferometer is defined as the full width of the peak at half the maximum intensity, which is inversely proportional to the path length of the mirrors. Therefore, to obtain a high resolution spectrum, an interferometer with a long mirror travel is needed.

Although both technologies are currently used today, FT-NIR spectrophotometers offer several advantages over diffraction grating spectrophotometers. They offer higher resolution spectral data; their detectors are able to collect large amounts of light at a single time point, which is referred to as the Jacquinot's advantage or the advantage of high throughput; and, all wavelengths can be collected at a single time point, which is referred to as Fellgett's advantage or the advantage of multiplexing (Siesler et al., 2002).

2.6.2. Spectral data collection and preprocessing

The main objective in spectroscopy is to measure the amount of energy that has been absorbed by the sample. Depending on the sample type this can be done two different ways. The first way is by transmittance, in which light passes through a sample

and the difference in incident and transmitted light intensity is proportional to the energy absorbed by the material. The second option is to measure the amount of light that is reflected off the surface of the material. The light absorbed can then be correlated to the amount of light before hitting the sample to the light that is reflected. For transparent samples, transmittance is typically utilized while, for opaque samples such as plant materials, reflectance is used (Blanco and Villarroya, 2002).

Once the spectra are collected, they need to be preprocessed. It is important to note that while preprocessing is an important step before multivariate calibration can be done, preprocessing techniques do not increase the resolution of the spectra. Preprocessing is merely conducted to abate noise and nonchemical effects contained in spectra (Siesler et al., 2002).

The first type of preprocessing involves scatter correction, which corrects for variations in light scattering properties of the samples like particle size. These methods correct for additive and multiplicative effects (Helland et al., 1995). One technique, multiplicative scatter correction (MSC), is based on the fundamental principle that depending on the physical properties of a sample, the sample will reflect light differently. In doing so, it can have an additive effect where the spectra are merely shifted up or down from the average spectra and/or it can have a multiplicative effect where the intensities of the bands are heightened or damped. To correct for both of these physical phenomena, the MSC algorithm first calculates the mean spectra (Figure 2.4; Næs et al., 1990). Next the spectra are plotted on the x-axis against the mean spectrum on the y-axis. A linear regression is then conducted for each of the spectra:

$$y = ax + b \qquad \text{[Equation 2.2]}$$

Keeping in mind that this algorithm assumes that the additive and multiplicative effects are non-chemical, the next step is to abate them or to normalize them for all the spectra. This is done by correcting the spectra, so the regression models are the same for all,

$$x_{k,new} = \frac{x_{k,old} - a}{b} \qquad \text{[Equation 2.3]}$$

where the offset (a) is subtracted from every wavenumber recorded (k) and then is divided by the slope of the regression model (b). This corrects all the spectra so they have a slope of one and a null offset (Figure 2.4).

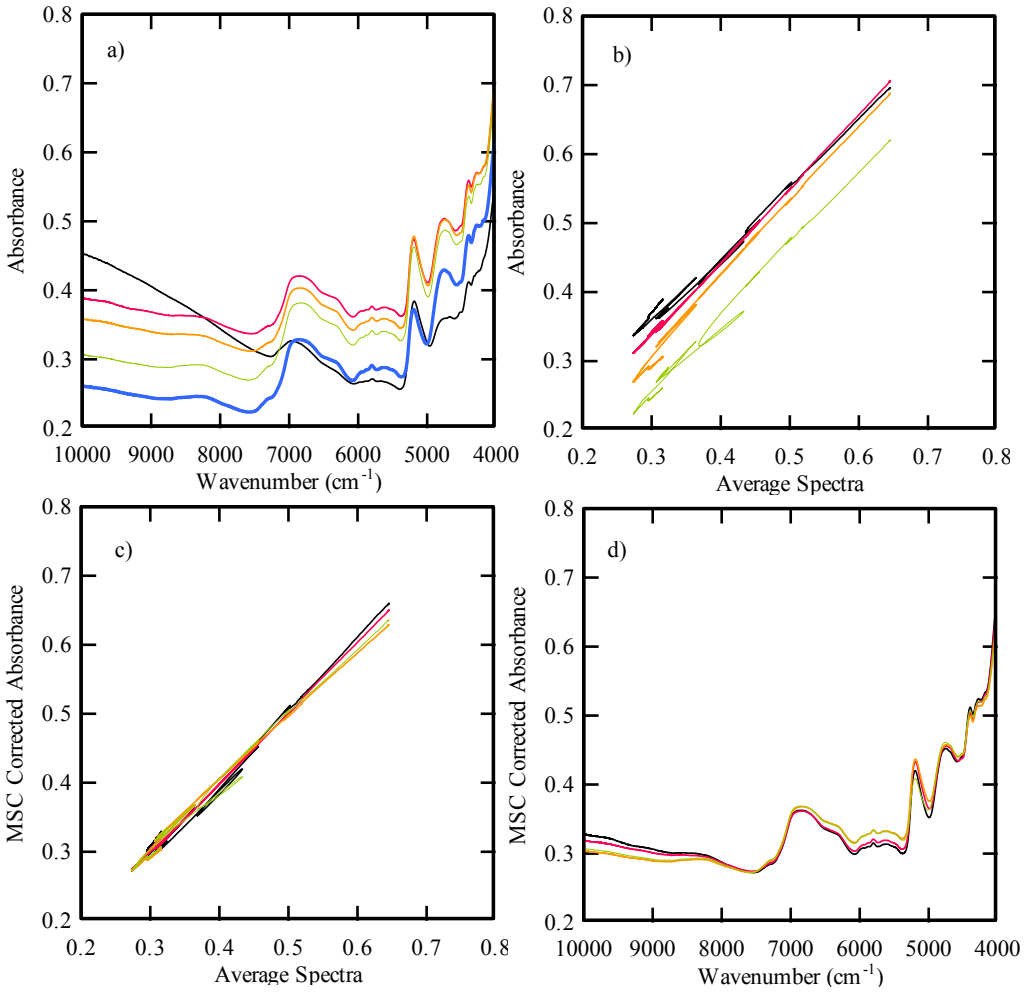


Figure 2.4. Demonstration of MSC on FT-NIR spectra with (a) the raw spectra; (b) comparison between the raw spectral values to the averaged spectral values; (c) application of the MSC algorithm to the comparison; and (d) resulting MSC corrected spectra.

The second type of preprocessing involves smoothing the spectra and/or conducting a derivative estimation to smoothen the spectral data. The estimated derivative curves are used to remove baseline offset and to see if the slopes of the spectra contain information by increasing the visual resolution of the spectra (Savitzky and Golay, 1964; Siesler et al., 2002). While derivatives abate baseline shifts, they can, however, add noise to the spectra. Hence derivatives are usually accompanied by smoothing filters to remove some of the noise that was added during the derivation (Siesler et al., 2002).

One derivative-smoothing algorithm is called the Savitzky-Golay (SG) derivative. It works by taking the derivative then smoothing the derivative by fitting a polynomial to the data set with a specific window width (Savitzky and Golay, 1964). When implementing the SG derivative, the derivative order, the polynomial order, and the window width or the region used for the calculation need to be set. Each variable can affect the resulting derivative curve by either under- or over-smoothing the spectra (Figure 2.5). Typically, higher order polynomials can fit the steeper peaks better than lower order polynomials when using the same window width (Siesler et al., 2002).

In addition to utilizing the MSC and SG derivative separately they can also be utilized together in series. Chen et al. (2013) studied the utilization of the MSC and SG filters together and determined that using the SG filter followed by the MSC pretreatment yielded better results than conducting vice-versa. However, it has been customary in NIR spectral analysis to apply the MSC pretreatment first, followed by the SG filter (Hayes, 2012; Shetty et al., 2012).

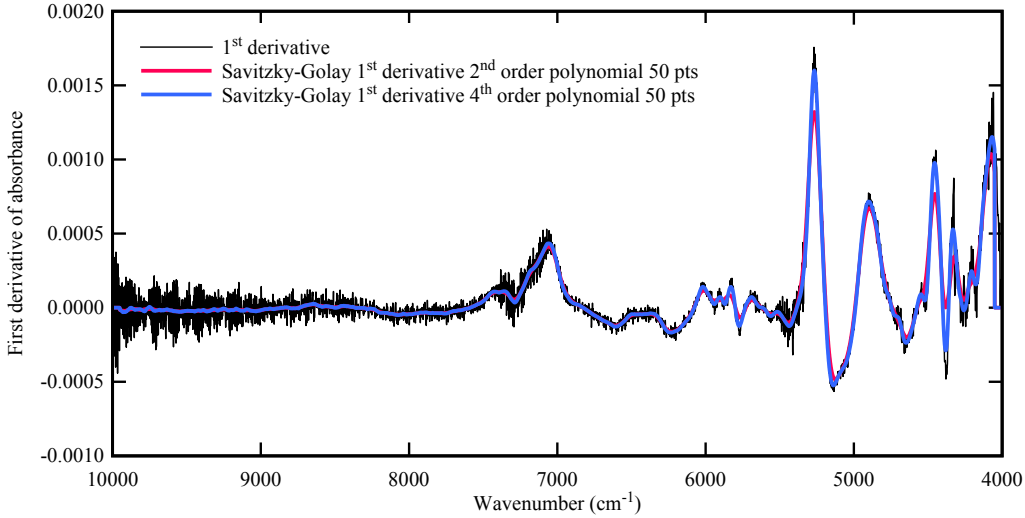


Figure 2.5. Demonstration of a first order Savitzky Golay (SG) derivative. The three spectra include a first derivative with no smoothing, the second is a first order SG with 50 left and right smoothing points and a second order polynomial, and the third is a first order SG with 50 left and right smoothing points and a fourth order polynomial.

2.7. Multivariate analysis of spectral data

2.7.1. Calibration and validation data sets

To construct robust models the data set that is used should have a few key characteristics such as a wide range that is evenly distributed having a low kurtosis. The sample set should also include samples over the range that could be encountered when using the model. The sample set size needs to be greater than 100 samples in the calibration set and 30 to 50 in the validation set. (AACCI Method 39-00, 1999)

2.7.2. Development of calibration models

Multivariate calibration is based on the fundamental principles of taking many variables *X* and projecting them onto a few variables *T*. This projection compresses the data into a more reliable model leaving out much of the noise and collinearity that can accompany large data sets such as NIR spectra (Lattin et al., 2003), allowing for a single constituent (e.g., lignin content) to be determined in complex samples. In the case of

spectral data sets, X are the absorbance, reflectance, or transmittance values at each wavenumber while T represents resulting principal components' scores of the analysis. Two of the main multivariate calibration methods are principal component regression (PCR) and partial least squares regression (PLSR) (Martens and Næs, 1989).

PCR is based on using principal components from principal component analysis (PCA) and regressing y (e.g., reference data, such as glucan, xylan, or lignin content of Miscanthus) onto the principal components' scores using multiple linear regression (MLR) (Lattin et al., 2003; Martens and Næs, 1989). However PCR utilizes eigenvectors to determine the axis that will bring the most variability to the data set X without taking into consideration y. This can be problematic where there are large variations in the data set that are not caused by y but, instead, include other effects such as light scattering, in the case of spectral data. PLSR solves those problems by determining the variability of X and y simultaneously.

Similarly, in PLSR, the spectral data matrix, $\boldsymbol{X}$, are compressed into a few predicted factors, $\hat{T}$ while considering the **y** reference data vector. This lessens the effect that noise and non-correlated variables have on the final calibration (Martens and Næs, 1989). This is done by determining the covariance matrix between $\boldsymbol{X}$ and the reference data, $\boldsymbol{y}$, and maximizing the covariance with the loading weight vector $\boldsymbol{w}$.

First, the spectral data are mean-centered by subtracting the mean from each measurement. For the spectra data, the mean absorbance $\overline{\boldsymbol{x}}$ is subtracted from the absorbance measurement $\boldsymbol{X}$ at each wavenumber (or wavelength) while, for the reference data set, the mean of the reference data $\bar{y}$ is subtracted from each measurement y:

$$\boldsymbol{X_0} = \boldsymbol{X} - \overline{\boldsymbol{x}} \qquad \text{[Equation 2.4]}$$

$$\boldsymbol{y_0 = y - \overline{y}} \qquad \text{[Equation 2.5]}$$

The weight vector w is calculated by maximizing the covariance between the $\boldsymbol{X}$ and $\boldsymbol{y}$ variables, which will expose regions of the spectral data X that have a good correlation to the reference y values. This is done by maximizing Equation 2.6 where y is least squares fitted to X.

$$\boldsymbol{w_a^{'} X_{a-1}^{'} y_{a-1}} \qquad \text{[Equation 2.6]}$$

$$\boldsymbol{X_{a-1} = y_{a-1} w_a + E} \qquad \text{[Equation 2.7]}$$

Once the weight vector has been determined, it is used to create the new variables $\hat{\boldsymbol{t}}_{\boldsymbol{a}}$, called factor scores:

$$\boldsymbol{\hat{t}_a = X_{a-1} \hat{w}_a} \qquad \text{[Equation 2.8]}$$

With the factor scores determined, the $\boldsymbol{X}$ and $\boldsymbol{y}$ loadings can then be calculated by regressing the $\boldsymbol{X}$ matrix on $\hat{\boldsymbol{t}}_{\boldsymbol{a}}$ for the X loadings and regressing $\boldsymbol{y}$ on $\hat{\boldsymbol{t}}_{\boldsymbol{a}}$ for the y loadings. The regressions are done in a least squares fashion:

$$\boldsymbol{\hat{p}_a = \frac{X_{a-1}^{'} \hat{t}_a}{\hat{t}_a^{'} \hat{t}_a}} \qquad \text{[Equation 2.9]}$$

$$\boldsymbol{\hat{q}_a = \frac{y_{a-1}^{'} \hat{t}_a}{\hat{t}_a^{'} \hat{t}_a}} \qquad \text{[Equation 2.10]}$$

where $\hat{\boldsymbol{p}}_{\boldsymbol{a}}$ and $\hat{\boldsymbol{q}}_{\boldsymbol{a}}$ represent the $\boldsymbol{X}$ and $\boldsymbol{y}$ loadings, respectively.

With the scores and loading that were determined for this factor, the new **X** and **y** data sets can be created for the next factor:

$$\boldsymbol{X_a = X_{a-1} - \hat{t}\hat{p}_a^{'}} \qquad \text{[Equation 2.11]}$$

$$\boldsymbol{y_a = y_{a-1} - \hat{t}\hat{q}_a^{'}} \qquad \text{[Equation 2.12]}$$

The process is repeated several times, until there are enough factors to explain the variance in the spectral data; hence, the variable *y* can be predicted:

$$\hat{\boldsymbol{y}} = \hat{\boldsymbol{b}}_{\boldsymbol{0}} + \boldsymbol{X}\hat{\boldsymbol{b}} \qquad \text{[Equation 2.13]}$$

where, $\hat{\boldsymbol{b}}_{\boldsymbol{0}} = \overline{\boldsymbol{y}} - \overline{\boldsymbol{x}}' \, \hat{\boldsymbol{b}}$ and the regression coefficient matrix is

$\hat{\boldsymbol{b}} = \widehat{\boldsymbol{W}}\left(\hat{\boldsymbol{P}}' \, \widehat{\boldsymbol{W}}\right)^{-1} \hat{\boldsymbol{q}}.\, \widehat{\boldsymbol{W}}$ is the matrix containing all the weight vectors for the factors, $\hat{\boldsymbol{P}}$ is the matrix containing all the *X* loadings for the factors, and $\hat{\boldsymbol{q}}$ is the vector containing all the scalars from the *y* loadings.

2.7.3. Development of classification models

In certain applications, the properties or composition of samples may not need to be predicted but, instead, samples need to be classified into certain groups or categories. Supervised pattern recognition techniques use the information about the class membership of the samples to a certain group or category in order to classify new unknown samples based on the pattern of measurements; in the case of near infrared spectroscopy, the classification is based on the absorbance spectra of the samples. The general procedure of supervised pattern recognition techniques include the selection of calibration and validation sets, in which the class memberships of the samples are known; selection of a variables or spectral data; model development using the calibration set only; and validation of the model with an independent set of samples. Several kinds of pattern recognition methods have been applied to agricultural, food, and pharmaceutical products. Two methods, linear discriminant analysis (LDA) and soft independent modeling of class analogy (SIMCA) are often applied.

There are three main applications of LDA: profiling, differentiation, and classification (Lattin et al., 2003). Profiling looks at how groups differ with respect to set

variables; differentiation tests variables to see whether they are similar or not; and classification is used to determine in which group an unknown variable will lie. LDA classification works by taking dependent *y* variables (class) and independent *x* variables and determining scores for the *x* variables that maximize the sum of squares between groups. However for the LDA to be executed the matrix for each group must have more samples than variables. Therefore the spectrum needs to be reduced by PLSR to a few *t* variables. Once the spectrum has been reduced to a few *t* variables, the new *x* variables (*t* variables from PLSR) can be imported and the model can be built (Lattin et al., 2003).

LDA is considered a hard classification method because the algorithm looks at the difference between groups, which always classifies a sample into a group; no matter how different a sample is from the closest group (Berrueta et al., 2007). SIMCA, on the other hand, is a soft classification method that looks at similarities within groups and the algorithm may classify a sample into a group or not (Esbensen et al., 2002). This is done by grouping the variables into their respective groups, followed by building a PLSR for each group. The PLSR models for all the groups are then utilized in the SIMCA. To run SIMCA, a sample set is imported and tested in each of the PLSR model. If a sample is similar to the samples that were used to create the PLSR model for a specific group, then the sample is classified into that group. If it is not similar to any of the groups, the sample is not classified. Therefore a sample could be classified into any number of the groups or none at all (Esbensen et al., 2002; Berrueta et al., 2007).

2.7.4. Validation

Quantification and classification models need to be validated before they are used. Validation can be conducted two ways, either by designating a set of samples to be used

for the validation or, in quantification, doing a cross validation (Martens and Næs, 1989; Esbensen et al., 2002). In the case where a validation set is designated, the calibration model is run using an independent validation and the efficacy of the model is determined. In cross validation, during calibration, several models are constructed and, each time a new model is constructed, a few samples are left out. In a "leave one out" cross validation procedure, one sample is left out for validation and the rest are used to construct calibration sub models; the sample that was left out is then used to validate the sub-model. The process is repeated until each sample has been left out once. With large sample sets, this process can become time consuming, so a group of samples, instead of one sample, may be used for cross validation. This process of building several submodels and validating with a different designated set each time is done until all the samples have been in the validation set. The final model is then the average of all the submodels created (Martens and Næs, 1989).

The uncertainty of the submodels during cross validation can be determined using the Martens uncertainty test (Esbensen et al., 2002). For each sub-model, a set of model parameters – β (regression) coefficients, scores, loadings, and loading weights – have been calculated. The Martens uncertainty test utilizes the mean and standard deviation to determine if a specific regression coefficient is different from zero. If the regression coefficient is not different from zero, it can then be left out in the next model built. The test allows for identification of possible wavelengths that are measuring non-chemical data and adding noise to the system since those wavelengths will have null regression coefficients. The whole process of calibration can then be repeated using a reduced

spectrum, i.e., excluding wavelengths with null regression coefficients, and validated with an independent sample set.

The reduced spectrum can then also be confirmed or compared against similar materials. While NIR is highly selective and the models can only be used for a single material type, similar materials should have similar significant wavenumbers.

2.7.5. Evaluation of calibration and classification models

To evaluate each model, a set of parameters will be considered. In the first level of evaluation, the correlation coefficient (R^2), mean square error (MSE), standard error (SE), and bias will be defined. The correlation coefficient is a measure of the linear dependence between and reference, x, and predicted, y, data sets; therefore an $R^2 = 1$ is desired showing x and y variables are equal to each other. Since errors are present in the data set they need to be measured, one way to do so is with the MSE. The MSE is the average square difference between the actual value and the predicted value of a sample. However, since the MSC will end up with squared units of measure, the square root of the MSE is often taken to get the root mean square error (RMSE). The MSE can be broken down into two forms of error – standard error (SE) and bias. The standard error is the difference between the deviation from the reference data and the average deviation from the reference data. Bias, on the other hand, is a measure of the average deviation from the reference value. A bias of zero is desired, showing that the model predicts the validation set the same as the calibration set, or that there is no additive offset.

The second level of evaluation involves comparing errors that are present in the model to the data set that was used to validate it. Three measurements are often used - the ratio of performance to deviation (RPD), the range to standard error of prediction

(R/SEP), and the relative ability of prediction (RAP). The RPD is a comparison of the standard deviation of the validation set to the standard error of prediction, the R/SEP is a comparison of the range of the validation data set to the standard error of prediction, while the RAP is a comparison of the difference between the variance of the validation set and the mean square error predicted, to the difference between the variance of the validation set and the mean variance of the reference data results (Martens and Næs, 1989).

$$RPD = \frac{\sigma_{validation}}{SEP} \quad \text{[Equation 2.14]}$$

$$\frac{R}{SEP} = \frac{Range_{Validation}}{SEP} \quad \text{[Equation 2.15]}$$

$$RAP = \frac{\sigma^2_{Validation} - MSEP}{\sigma^2_{Validation} - \sigma^2_{wet\ Chem}} \quad \text{[Equation 2.16]}$$

The American Association of Cereal Chemists International (AACCI) has defined Guidelines for RPD and R/SEP values for model development and maintenance (AACCI Method 39-00, 1999). For an RPD value greater than 2.5, the model is deemed good for screening in breeding programs; greater than 5, the model is acceptable for quality control; greater than 8, the model is useful for process control, development, and applied research. For an R/SEP value greater than 4; the model is good for screening; greater than 10, the model is good for quality control; and greater than 15, the model may be used for quantification purposes.

A good PLSR model also uses a low number of factors; the more factors in a model the higher risk of over fitting and explaining noise. The explained variance plot should be close for both calibration and validation, where the explained variance is how much each factor contributes to the correlation coefficient (Martens and Næs, 1989).

2.8. Application of NIR spectroscopy in biomass compositional analysis

Several studies have shown NIR spectroscopy as a promising technique to assessing biomass composition. Aenugu et al. (2011) conducted a review and specified where specific organic bonds should have absorption bands. Schwanninger et al. (2011) conducted a review of wavenumbers that should correspond to chemicals in wood. The review highlighted several regions specific to cellulose, hemicellulose, and lignin.

Sanderson et al. (1996) demonstrated individual carbohydrates can be estimated in woody and herbaceous feedstocks such as straw, corn stover, poplar, etc. using standard normal variate-detrend (SNV-D) preprocessing to correct the scatter in the NIR spectra collected and regression by PLS. Hames et al. (2003) reported NIR calibration models for corn stover feedstock and dilute acid pretreated corn stover. Pordesimo et al. (2005) later used the corn stover feedstock model to investigate the variability of stover composition with crop maturity at harvest. They took samples from corn plants from approximately two weeks before the corn grain reached physiological maturity to approximately one month after the grain was at a moisture content suitable for harvesting. Their results showed large decreases in the extractives content of the samples, with increases in both xylan and lignin content. The corn stover feedstock model was also used by Hoskinson et al. (2007) to provide compositional data for a study investigating the variation in quality and quantity of corn stover available under different harvesting scenarios. PLSR models of NIR spectra were used to evaluate compositional variation and sources of variability in 508 commercial hybrid corn stover samples collected from 47 sites in eight Corn Belt states after the 2001, 2002, and 2003 harvests (Templeton et al., 2009). Similarly, Haffner et al. (2013) demonstrated the use of PLS regression models

of NIR spectra of 241 *Miscanthus* × *giganteus* samples harvested from seven sites in Illinois for fast monitoring of Miscanthus in plant breeding studies.

Besides utilizing the spectral data to construct models that can quantify components, NIR spectra have also been utilized to classify materials. Ye et al. (2008) fractionated corn stover into botanical fractions (node, leaf, rind, pith, sheath, and husk) and scanned them with an FT-NIR spectrophotometer. For each fraction the spectral data was taken and a PCA was conducted followed by SIMCA. Results showed that the SIMCA model developed could classify 60 additional botanical fractions correctly. Similarly, Yang et al. (2007) utilized NIR spectra and SIMCA to classify rotted wood. Utilizing PCA models for non-degraded wood, white rot, and brown rot, the SIMCA model was able to predict the test set for non-decay, white-rot, and brown-rot, at 100%, 85%, and 100%, where the misclassified samples were white-rot samples placed into the brown rot model. To determine the type of feed that was being fed to ruminant animals, Cozzolino et al. (2008) utilized PLSR, PCA, and LDA to classify feed types. Using grain silage, grass and legume silage, and sunflower silage, they constructed a PLS-DA model and could predict the silage type more than 90% of the time. These studies show NIR with either SIMCA or LDA analyses can be a powerful tool in screening or classifying samples against a set of quality standards.

2.9. Development of biomass specifications

Current feedstock production is driven by cost over quality where the price of the material is determined on a dry ton basis and not on the amount of material available for conversion. However as biorefineries start to optimize their processes, the focus on the

quality of the feedstock will be more important since many components in lignocellulosic feedstocks cannot be converted or can inhibit the conversion process.

Since variations can be large in agricultural samples and can be contributed to many factors such as genetics, growing conditions, plant age, multiple plant fractions and tissue types, handling, and storage conditions (Hames et al., 2009), specifications would be advantageous to have, to ensure that materials that could hurt the conversion process are kept to a minimum or are left out entirely.

While the use of biomass for biochemical conversion is still a developing technology and industry, the use of biomass for combustion has been around for many years. The European Committee for Standardization (CEN) has come up with a set of standards that deal with specifications, classifications, and quality assurance of solid biofuels, essentially classifying various solid fuels into categories. With these categories in place, material can be combined to form blends and mixtures and designed to meet certain specifications. CEN defines the terms *blends* as intentionally mixed biofuels with a known composition correlated to a specific heating value while *mixtures* are unintentionally mixed (Alakangas et al., 2006).

Bringing that ideology to the lignocellulosic biofuel industry, clean, consistent feedstocks that meet quality specifications may be possible by blending different grades of material or different feedstock sources. Researchers at DOE's Idaho National Laboratory have proposed the concept of an advanced uniform system for a commodity-based biomass industry (Hess et al., 2009). In this system, various types of biomass (i.e., corn stover, switchgrass, Miscanthus, etc.) and physical characteristics (i.e., bulk densities, moisture contents, etc.) are converted into a standardized format (e.g., pellets)

early in the supply chain. As a standard format, the biomass needs to adhere to a set of specifications where it can be classified, bought and sold in the market, allow farmers to contract directly with biorefineries, and enable large-scale conversion facilities to operate with a continuous, consistent, and economic feedstock supply.

Because little is known about the variability in all biomass feedstocks and how they can be economically mitigated, the development of specifications has been slow. For ash content, current conversion process analyses rely upon an average modeled value of approximately 5% dry basis (Aden and Foust, 2002) and, in pyrolysis, ash levels must be kept below 1% (Kenney et al., 2013). Biorefineries have not yet specified minimum carbohydrate contents in biomass, but Humbird et al. (2011) chose to establish a total structural carbohydrate specification for corn stover of 59% (w/w) for the technoeconomic modeling of cellulosic ethanol production. This specification is comparable to the mean structural carbohydrates (i.e., sum of glucan and xylan contents) that have been reported for corn stover (53%), corn cobs (59%), Miscanthus (61%), and wheat (50%) (Kenney et al., 2013). A specification for moisture content is crucial as DML rates in aerobic storage increase with moisture content (Emery and Mosier, 2012). The threshold moisture content for safe storage varies among biomass types and conditions (e.g., storage indoors, storage outdoors with a tarp, enclosed in a silo bag, etc.) but a moisture content of 20% (w.b.) is a generally recognized rule of thumb for limiting DML (Darr and Shah, 2012).

CHAPTER 3. COMPOSITION OF MISCANTHUS FROM STORED BALES

3.1. Introduction

Miscanthus, a lignocellulosic bioenergy feedstock, is mainly composed of cellulose (40%), hemicellulose (20%), and lignin (20%) with the balance consisting of extractives, ash, and other constituents. Variations in composition arise due to age, stage of growth, growth conditions, and other factors (Perez et al., 2002; Hames et al., 2003). Large variations in composition can be problematic since, with all biochemical conversions, the quality of the material entering the conversion process can impact the efficiency of the process. Several researchers have stressed the importance of supplying conversion processes with a "clean" and consistent feedstock since product yields are not proportional to the total mass of the input, but rather the mass of certain components, or composition, of the input (Kenney et al., 2013; Liu et al., 2010; Hames et al., 2003).

As mentioned in Chapter 2, how biomass is stored can affect its composition over time. Wiselogel et al. (1996) studied the compositional changes in round switchgrass bales and found glucan contents ranging from 35.6 to 40.8%; xylan contents ranging 23.4 to 26.1%; arabinan contents ranging from 2.9 to 3.4%; lignin contents ranging from 20.1 to 23%; and ash contents ranging from 4.8 to 6.1% after nine months of storage. These data show biomass degrade over time, especially when exposed to the high moistures, when not stored properly.

In this study, in order to develop robust calibration and classification models for Miscanthus, samples with a large variability in composition were needed. To achieve this, samples from Miscanthus bales stored under a variety of conditions – indoors, under roof, outdoors with tarp cover, and outdoors without tarp cover - for different time

periods and different years were used. The variability in glucan, xylan, arabinan, acetyl, lignin, ash, and extractives content across this wide range of storage conditions and periods are described in this chapter.

3.2. Materials and methods

Bale core samples were collected using a hay probe bale sampler (Part No. BHP550C, Best Harvest, St. Petersburg, FL) from stacked bales that were stored in Urbana, Griggsville, and Taylorville, IL for a period of 3 to 24 mo. (Figure 3.1). All bales measured 0.91 x 1.21 x 2.43 m. The bales stored in Urbana, IL were harvested at the senescent stage (December to January) from the Energy Biosciences Institute (EBI) farm at the University of Illinois in Urbana-Champaign in 2008 to 2011. The bales stored in Griggsville and Taylorville were harvested at the senescent stage in Pana, IL in December 2008 to January 2009. The bales were stored under different conditions: indoors (Taylorville); under roof (Urbana); outdoors with a tarp (Urbana and Griggsville); and outdoors without a tarp (Urbana and Griggsville).

Each bale sample, approximately 40 g, was a collection of multiple core samples from the bale (Table 3.1). After collection, the samples were dried at 60°C for 72 h according to ASABE Standard S358.2 (1998). The dried samples were milled using a cutting mill (SM 2000, Retsch, Inc., Haan, Germany) fitted with a 2 mm sieve. The dry ground samples were bagged and stored at room temperature prior to sending to the EBI Analytical Chemistry Laboratory at the University of California in Berkeley campus for compositional analysis (glucan, xylan, arabinan, acetyl, lignin, ash, and extractives content). Compositional analyses were conducted in duplicates following standard

procedures developed by the National Renewable Energy Laboratory (NREL) and discussed in Haffner et al. (2013).

(a) Bale stack in Griggsville, IL. Bales were covered with a tarp for 12 months and ripped tarp for the next 12 months.

(b) Bale stack in Taylorville, IL. Bales were stored indoors for a period of 24 months.

(c) Bale stacks in the Energy Bioscience Institute (EBI) Farm in Urbana, IL. From left to right, bales were stored under roof, outdoors without tarp cover; and outdoors with tarp cover.

Figure 3.1. Sources of *Miscanthus* × *giganteus* core samples used in this study.

Table 3.1 Description of Miscanthus bale samples

Sample Group	Storage			Sample Names
	Location	Conditions	Period (mos.)	
1	Taylorville	Indoors[a]	24	Ty32, Ty33, Ty34
2	Griggsville	Outdoors, with tarp[b]	24	Gr1, Gr2, Gr3, Gr4, Gr10, Gr14, Gr16, Gr21
3	Urbana	Under roof[c]	24	E185, E186, E187, E188, E189, E190
4	Urbana	Under roof	12	E37, E38, E39, E40, E52, E53, E55, E56, E59
5	Urbana	Outdoors, without a tarp	12	E43, E44, E45, E46, E73, E74, E75, E76, E79, E272
6[d]	Urbana	Under roof	6	E89, E90, E91, E92, E93
7[d]	Urbana	Outdoors, without a tarp	6	E180, E181, E182, E183, E184

Table 3.1 Continued

Sample Group	Storage			Sample Names
	Location	Conditions	Period (mos.)	
8[d]	Urbana	Outdoors, without a tarp	6	Z108, Z120, Z132, Z135, Z144, Z153, Z198, Z213, Z231
9	Urbana	Outdoors, with tarp	6	Y114, Y123, Y126, Y141, Y150, Y159, Y200, Y216, Y234
10[d]	Urbana	Under roof	6	X111, X117, X129, X138, 147, X156, X196, X219, X237
11	Urbana	Outdoors, without a tarp	3	O206, O207, O208, O221, O222, O223, O239, O240, O241
12	Urbana	Outdoors, with tarp	3	T209, T210, T211, T224, T225, T226, T242, T243, T244
13	Urbana	Under roof	3	I203, I204, I205, I227, I228, I229, I245, I246, I247

[a]Bales stored at the Taylorville site where inside a locked storage building, completely protected from weather elements for 2 yr.
[b]Bales stored at the Griggsville site were covered on top with a tarp. After the first 12 mo. of storage, parts of the tarp had worn out and blown away. In the second year of storage, therefore, the bale stack was only partially covered.
[c]Bales stored under roof at the Urbana site were placed on a concrete floor but were still exposed to ambient temperature, moisture, and wind on all sides.
[d]Bales in Groups 6 and 10 were harvested from the same field in the same year but stacked separately. Likewise, Groups 7 and 8 represent separate stacks of bales harvested from the same field in the same year.

The composition means of each sample group were determined, compared, and tested for significance ($p < 0.05$) using Tukey's test (Table 3.2). Statistical analyses were conducted using *R* (Version 2.15.2, 2012).

3.3. Results and discussion

Compositions of Miscantus bale samples ranged from 25.8 to 44.1% glucan, 16.6 to 25.1% xylan, 1.0 to 3.1% arabinan, 1.7 to 3.4% acetyl, 17.5 to 26.5% lignin, 0.5 to 14.0% ash, and 3.6 to 9.1% extractives. Group 2 (outdoors with tarp, 24 mo.) had the lowest mean glucan content and the highest standard deviation of all the groups that were tested (Table 3.2). Groups 1 (indoors, 24 mo.) and Groups 4 (under roof, 12 mo.), 5

(outdoors without tarp, 12 mo.), and 6 (under roof, 6 mo.) had comparable glucan contents to Group 2 even though these bales were stored for less time and under different storage conditions as Group 2.

Table 3.2. Composition of Miscanthus bale samples from different storage conditions and time periods.

Sample Group	Mean[a] ± S.D.[b] (%)						
	Glucan	Xylan	Arabinan	Acetyl	Lignin	Ash	Extractives
1	40.8ab ± 2.9	20.5ab ± 0.7	1.9abcd ± 0.4	3.2a ± 0.3	19.8abc ± 1.6	2.7a ± 1.4	5.2a ± 0.9
2	35.8b ± 5.8	20.0b ± 3.2	1.7bcd ± 0.4	2.6ab ± 0.3	22.2a ± 3.1	4.1a ± 4.0	6.1a ± 1.9
3	43.2a ± 0.6	19.0b ± 0.4	1.3d ± 0.2	2.9ab ± 0.1	21.7ab ± 0.6	1.4a ± 0.4	6.1a ± 0.2
4	40.3ab ± 1.5	22.5a ± 1.1	2.2abc ± 0.3	2.4b ± 0.1	19.2bc ± 0.9	2.3a ± 1.3	5.4a ± 0.4
5	38.7ab ± 2.1	22.4a ± 1.6	2.2ab ± 0.3	2.6ab ± 0.5	19.0c ± 0.6	3.7a ± 3.4	5.6a ± 0.5
6	38.4ab ± 2.5	21.7ab ± 1.0	2.6a ± 0.5	2.6ab ± 0.2	19.7abc ± 1.4	0.9a ± 0.2[a]	6.2a ± 0.7
7	41.2a ± 2.1	20.3ab ± 0.6	1.7bcd ± 0.3	2.9ab ± 0.1	21.0abc ± 0.8	3.0a ± 1.3	5.0a ± 0.3
8	40.9a ± 1.0	20.5ab ± 0.4	1.6cd ± 0.2	3.0a ± 0.1	20.9abc ± 0.7	2.7a ± 0.5	5.7a ± 1.0
9	41.4a ± 0.7	20.5ab ± 0.3	1.7bcd ± 0.1	3.0a ± 0.2	20.8abc ± 0.4	2.4a ± 0.5	5.3a ± 0.6
10	41.1a ± 0.9	20.3ab ± 0.4	1.7bcd ± 0.1	2.9a ± 0.2	20.7abc ± 0.2	2.7a ± 0.6	5.9a ± 0.5
11	41.5a ± 0.4	20.4ab ± 0.2	1.9bcd ± 0.1	2.4ab ± 0.1	20.6abc ± 0.2	2.4a ± 0.2	5.3a ± 0.6
12	41.2a ± 0.4	20.3ab ± 0.2	1.9bcd ± 0.1	2.8ab ± 0.3	20.6abc ± 0.3	2.5a ± 0.2	5.5a ± 0.9
13	41.3a ± 0.6	20.1ab ± 0.2	1.8bcd ± 0.2	2.9ab ± 0.1	20.7abc ± 0.8	2.5a ± 0.2	5.5a ± 1.1
All Samples	40.4 ± 2.7	20.7 ± 1.5	1.9 ± 0.4	2.8 ±0.3	20.5 ± 1.4	2.6 ± 1.8	5.6 ± 0.9

[a]Component values followed by the same lowercase letter in the same column, are not different ($p > 0.05$).
[b]S.D. = one standard deviation.

Group 2, along with Group 3 (under roof, 24 mo.), had the lowest mean xylan contents. The highest standard deviation in xylan content was also observed with samples from Group 2. Groups 4 and 5, which contained Miscanthus samples from the same plot, harvested in the same year, and stored for 12 mo., exhibited the highest xylan contents. There were no differences in xylan content observed for bales stored for 3 to 6 mo., irrespective of storage condition.

The low glucan and xylan contents in Group 2 translated into high lignin contents in this group, with a mean of 22.2% and a standard deviation of 3.1%. Older bale samples (stored for 12 mo. or longer) exhibited the highest variations in lignin content, as well, and no differences were observed for bales stored for 6 mo. or less.

The mean arabinan and acetyl contents across all samples was 1.9 and 2.8%, respectively, which were the lowest constituents measured in the analysis. The highest arabinan content was observed in Group 6 (stored under roof, 6 mo.) while the highest acetyl content was observed in Group 3 (stored indoors, 24 mo.). Overall, however, there were no differences observed across all samples for these two components. Likewise, all groups had comparable ash contents, with a mean of 2.6% and a standard deviation 1.8%. Because of the high standard deviations across all sample groups, the honestly significant difference (HSD) was also high, at approximately 3%. While Group 2 (outdoors with tarp, 24 mo.) had the highest mean ash content, 4.1%, this was due to a wide range of 0.59 to 14.0%. The sample with 14.0% ash content was likely an outlier since the next highest ash content in Group 2 was 4.47%.

In terms of extractives content, there were no differences observed among groups. It should be noted that Groups 2 and 3 (stored outdoors with tarp and under roof,

respectively, for 24 mo.) had visible fungal growth in all samples and these two groups had the highest extractives content. Since mold is very similar in composition to plant materials, the additional non-structural carbon compounds extracted from the biomass may be attributed to the mold contamination.

Group 2 (outdoors with tarp, 24 mo.) had a distinct structural carbohydrates content (i.e., sum of glucan, xylan, and arabinan contents) at 57% from all the other groups, which ranged from 62.8 to 65.0%. This result showed that, while the groups may have variation in single components, in terms of available carbohydrates for conversion, there were no variations across the conditions tested especially when the biomass was stored for less than 24 mo.

Overall, the composition values of Miscanthus from the stored bales were comparable to those reported by Haffner et al. (2013) and Hayes (2012), who used Miscanthus samples that were manually cut above ground at either pre-senescent or at senescent stages of growth. In contrast, samples from this study came from mechanically harvested Miscanthus, cut at the senescent stage, baled, and stored. Haffner et al. (2013) reported glucan contents ranging from 36.3 to 45.3% whereas glucan contents found in this study has a wider range, from 25.8 to 44.1%. The lower ranges found with stored bales was attributable to DML that occur during storage. Xylan, arabinan, acetyl, and ash contents reported from the two studies were comparable. Lignin contents in the stored bales was as high as 26.5% while the highest lignin content reported by Haffner et al. (2013) was 23.2%. This difference was likely due to Haffner et al. (2013) focusing on pre-senescent or senescent Miscanthus while this study focused on stored baled samples. Additionally, the extractives contents in the stored bales were only as high as 9.1% – as

extractives tended to be lost with storage time (Wiselogel et al., 1996) – whereas higher extractives, up to 12.2%, were found by Haffner et al. (2013).

Compared to compositional variability observed in corn stover (Templeton et al., 2009), Miscanthus exhibited a wider range and higher values in glucan, xylan, and lignin contents (Table 3.3). Ash levels in corn stover and Miscanthus were similar except for two samples from Groups 2 and 5 in this study. Compared to the compositional variability observed in switchgrass (Wiselogel et al., 1996), Miscanthus had a wider range in glucan content and lower ash levels. Their xylan and lignin contents were comparable.

Table 3.3. Glucan, xylan, lignin, and ash contents of Miscanthus, corn stover, and switchgrass.

Component	Range in composition (%)		
	Miscanthus* ×*giganteus **(this study)**	**Corn stover (Templeton et al., 2009)**	**Switchgrass (Wiselogel et al., 1996)**
Glucan	25.8 – 44.1	26.5 – 37.6	35.6 – 40.8
Xylan	16.7 – 25.1	14.8 – 22.7	23.4 – 26.1
Lignin	17.5 – 26.5	11.2 – 17.8	20.1 – 23.0
Ash	0.5 – 14.0[a]	0.8 – 6.6	4.7 – 6.2

[a]Two samples had ash contents of 12.8% (Group 5) and 14.0% (Group 2). Neglecting these samples, ash content ranged from 0.5 to 5.16%.

3.4. Conclusions

Large variations in Miscanthus compositions were observed in core samples from bales that were stored under a variety of conditions (indoors, under roof, outdoors with tarp, and outdoors without a tarp) for a period of 3 to 24 mo. Glucan and lignin contents ranged from 25.8 to 44.1% and 17.5 to 26.5%, respectively. In general, few trends were observed when looking that the sample sets were examined using a univariate approach. However, Group 2, which included bales stored outdoors with a tarp cover for 24 mo., exhibited lowest glucan and xylan levels and highest lignin levels compared to the rest of the bales. Mold growth was also observed in Group 2 samples. Compared to other

lignocellulosic feedstocks, overall, Miscanthus had higher carbohydrate and lower ash contents than corn stover while Miscanthus composition was similar to switchgrass.

CHAPTER 4. COMPOSITION OF BOTANICAL FRACTIONS OF MISCANTHUS

4.1. Introduction

Hames et al. (2003) noted that feedstocks can vary with plant genetics, growth environment, harvesting method, and storage conditions and the resulting variations are difficult to control. Compositional variation also exists among different botanical fractions, as each fraction is composed of different types of cells and serves different functions in the plant. Ye et al. (2008) found the sheath and pith fractions of corn stover had the highest glucan content, 39%; husks had the highest xylan content, 22%; and leaves having the highest lignin and ash contents, 23% and 7.4%, respectively. The variations can be attributed to the desired structural and physiological qualities of the plant fractions. Similarly, Liu et al. (2010) found the internodes of switchgrass had the highest glucan content, 40.9%; the node and internode had the highest xylan contents, 22.7% and 22.6%, respectively; and leaves had the highest lignin, 24.4%, and ash, 4.9%, contents. While the glucan, xylan, and lignin contents of corn stover and switchgrass were comparable, the difference in the mean ash contents of the leaves was as high as 2.5%.

The variation in composition of botanical fractions could also affect glucose yields. Crofcheck and Montross (2004) pretreated corn stover fractions with sodium hydroxide prior to hydrolysis and compared the results to a control (non-pretreated) sample set. Their results showed that the average glucose released for pretreated cobs, leaves and husk, stalks, and whole stover were 0.50, 0.35, 0.28, and 0.36 g/g while the control samples corresponded to glucose releases of 0.32 (cobs), 0.23 (leaves and husk), 0.17 (stalks), and 0.20 (whole stover) g/g respectively.

Kenney et al. (2013) also noted that biomass fractionation, i.e., selectively removing a particular anatomical fraction or tissue, takes feedstock selection a step further and is a solution to meeting specifications and ensuring a uniform format feedstock supply. For example, debarking of woody biomass is the current solution for meeting the aggressive ash specification, less than 1%, of the thermochemical conversion processes. Several studies have consistently shown that corn stover composition changes as the different proportions of the plant are harvested, with glucan content increasing with higher proportion of stalk and xylan increasing with higher proportions of cob and husk fractions (Pordesimo et al., 2005; Prewitt et al., 2007; Hoskinson et al., 2007; Karlen et al., 2011). Similar findings were observed with wheat fractions (Duguid et al., 2007).

In this study, the mass and composition of the botanical fractions of Miscanthus were determined and compared to the composition of corn stover and switchgrass fractions, as reported in the literature. Doing so provided a better understanding of the variability in core samples from bales, as each core sample could contain a mixture of botanical fractions different from the next. Furthermore, blending of different botanical fractions to demonstrate the delivery of Miscanthus with a wide range of structural carbohydrate and lignin contents was explored.

4.2. Materials and methods

4.2.1. Miscanthus stalks and plant fractionation

Miscanthus × *giganteus* samples (n = 6), each consisting of four stalks from a single stand, were harvested by hand in January 2013 from the EBI farm in Urbana, IL. Since the stalk samples were harvested during the senescent stage, the leaves (blades) had already fallen off and were not included in the mass fraction analysis, but were collected

just beneath the stand so the composition could be determined. The samples were taken to the laboratory where they were manually separated into stalks and leaf structures. The stalks were further separated into nodes and internodes by cutting (Figure 4.1). The internodes were then further separated by manually peeling off the sheath, splitting the stalk in half, longitudinally, and scraping the pith from the rind.

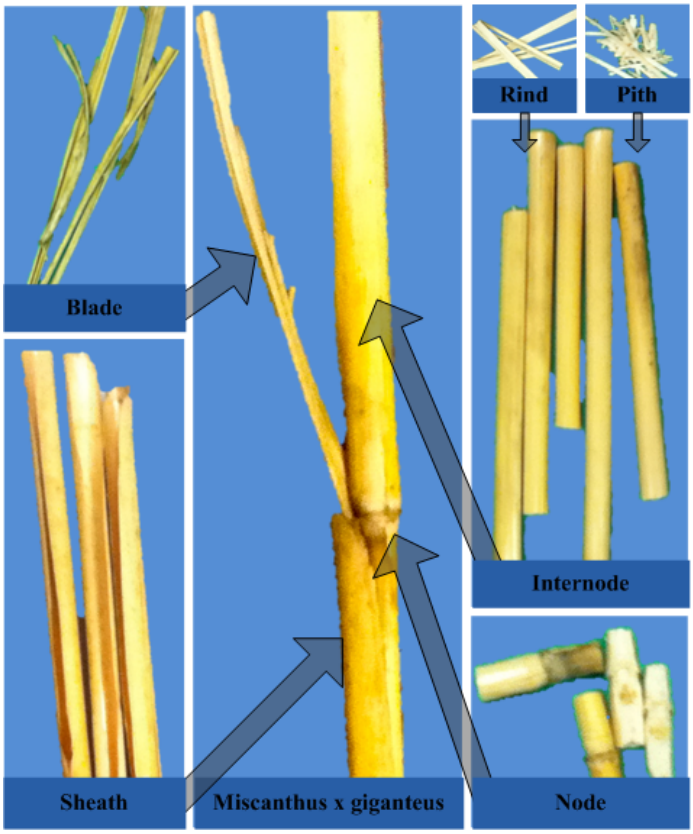


Figure 4.1. Miscanthus stalks were fractionated into pith, rind, sheath, leaves, and nodes.

Once all stalks per sample had been fractionated, the mass of each botanical fraction for each sample was determined. The moisture content of the samples was determined by drying a 1 g subsample at 103°C for 24 h according to ASABE Standard S358.2 (1998) for forage. The rest of the samples were dried at 60°C for 72 h for eventual compositional analysis according to the same ASABE Standard.

The samples were prepared for compositional analysis and FT-NIR spectra collection in a similar fashion as the core samples from the Miscanthus bales. Dried botanical fractions were ground using a cutting mill (SM 2000, Retsch, Inc., Haan, Germany) fitted with a 2 mm sieve, bagged, and stored at room temperature until compositional analysis and FT-NIR scanning. Compositional analysis was conducted in Dr. Vijay Singh's bioprocess engineering laboratory at the University of Illinois.

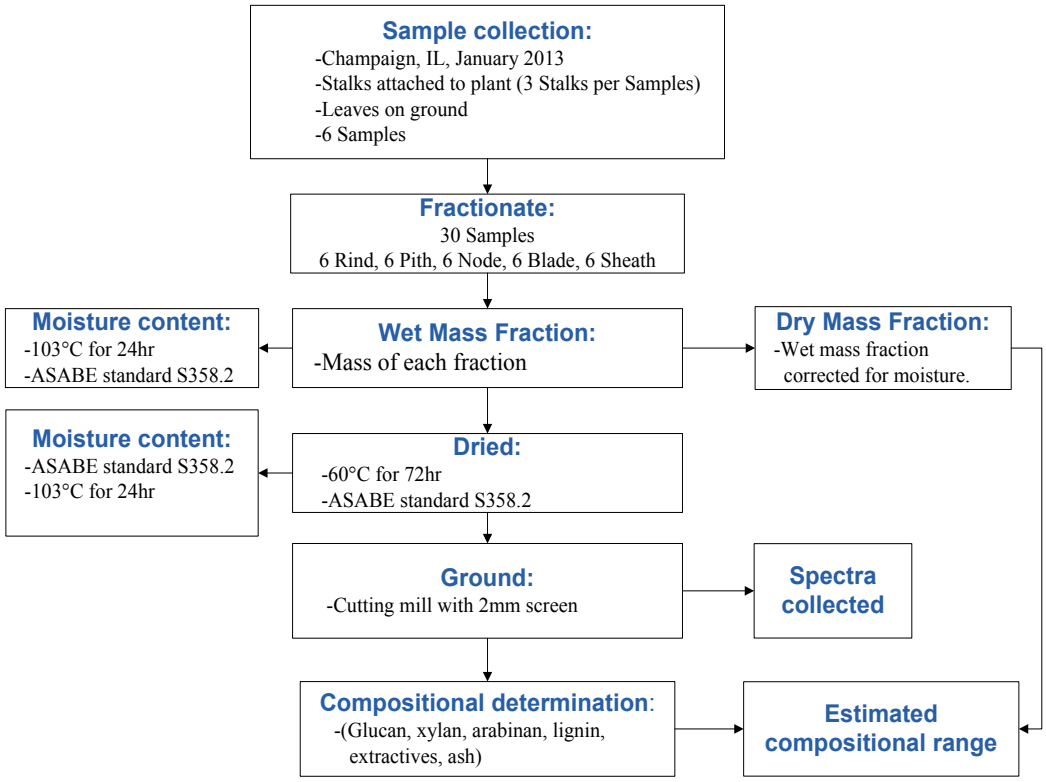


Figure 4.2. Flow chart of methods for the compositional determination of botanical fractions.

4.2.2. Chemical composition analysis

Chemical composition analyses were conducted in duplicates, except for moisture and extractive content, which were done in triplicates following standard procedures developed by NREL (Sluiter, 2005a,b; Sluiter, 2011).

Moisture and solids content. Prior to chemical analysis, the moisture and solids contents of the samples were determined in triplicates by weighing and placing 1 g subsamples in a dry disposable aluminum container. The mass of the empty container and mass of the container with subsample were recorded. The subsamples were dried in an oven at 105°C for 4 hr, cooled in a dessicator, and weighed for final moisture and solids content calculations:

$$Total\ solids\ (\%) = 100 - \frac{(m_{wet\ sample} - m_{dry\ sample}) \cdot 100}{m_{wet\ sample} - m_{container}} \quad \text{[Equation 4.1]}$$

$$ODW = \frac{m_{wet\ sample} \cdot Total\ solids\ (\%)}{100} \quad \text{[Equation 4.2]}$$

where *ODW* is the oven dry weight of the sample.

Extractives content. The extractives content of each sample was determined in triplicates using water and ethanol extractions in series. An extraction thimble was first weighed, filled with a 0.30 to 0.75 g subsample, and sealed using a heat sealer. The mass of the filled thimble was recorded. The subsamples were placed into a 300 ml Soxhlet extraction tube and extractives were removed with deionized water at 100°C for 6 h. The water was replaced with ethanol (95%) and further extraction was conducted for 16 h, for a combined extraction time of 22 h. The extraction tubes were drained and cleaned, and the thimbles were placed in an oven at 45°C oven for 24 h. The three replicates (dried thimbles) for each sample were weighed first and then combined for another moisture content determination.

The extractives content was calculated as follows:

$$m_{sample} = m_{thimble+sample} - m_{thimble} \quad \text{[Equation 4.3]}$$

$$m_{solids} = m_{sample} \cdot Total\ solids\ (\%) \quad \text{[Equation 4.4]}$$

Equation 4.4 was used to determine the solids content of the sample before ($m_{solids,i}$) and after ($m_{solids,f}$) extraction. The mass of the extractives and extractives content of the sample were then calculated as follows:

$$m_{extractives} = m_{solids,i} - m_{solids,f} \quad \text{[Equation 4.5]}$$

$$Extractives\ (\%) = \frac{m_{extractives}}{m_{solids,i}} \quad \text{[Equation 4.6]}$$

Acid hydrolysis and solids separation. In duplicates, 0.30 ± 0.01 g subsamples of the post-extraction samples were weighed and placed in 90 ml pressure tubes with 3 ml of 72% sulfuric acid. A set of sugar recovery standards (SRS), specifically, D-(+) glucose, D-(+)-xylose, D-(+)-galactose, L-(+)-arabinaose, and D-(+)-mannose, were also prepared as control samples. All samples were placed in an agitated water bath at 30 ± 3°C for 2 h. While in the water bath, every 15 min., the samples were manually agitated using a Teflon rod to break up the biomass as the acid hydrolyzed the sample. The sample and acid mixtures were then diluted with 84 ml of deionized water, Teflon caps were placed on the tubes and the tubes were autoclaved at 121°C for 1 h.

After autoclaving, the solids were separated from the dilute acid using a vacuum and placed in filtering crucibles, whose empty weights ($m_{crucible}$) had been recorded. The solids were used for acid insoluble lignin and ash determination while the recovered filtrates were used for sugars and acid soluble lignin quantification.

Acid insoluble lignin and ash determination. The filtering crucibles containing the solids from acid hydrolysis were weighed ($m_{crucible+solids}$) before being placed in an oven at 105°C for 24 h for drying. The dried solids were cooled in a dessicator for 20 min. and weighed again ($m_{crucible+solids,f,105℃}$). Afterwards, they were placed in a muffle furnace at 525°C for 4 h. where all organic material was incinerated leaving an inorganic

residue, or ash, behind. The filtering crucibles were weighed ($m_{crucible+solids,f,525℃}$) so that acid insoluble lignin (AIL) and ash contents could be determined as follows:

$$m_{AIL} = m_{\substack{crucible+solids,\\ f,105℃}} - m_{\substack{crucible+solids,\\ f,525℃}} \quad \text{[Equation 4.7]}$$

$$m_{ash} = m_{\substack{crucible+solids,\\ f,525℃}} - m_{crucible} \quad \text{[Equation 4.8]}$$

$$AIL\ (\%) = \frac{m_{AIL}}{ODW} \quad \text{[Equation 4.9]}$$

$$Ash\ (\%) = \frac{m_{ash}}{ODW} \quad \text{[Equation 4.10]}$$

Acid soluble lignin (ASL) determination. Within 6 h of acid hydrolysis, it was imperative to quantify the ASL content. The ASL content was determined colorimetrically using the filtrate samples and an Evolution 60s UV-VIS spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA). A 2 ml subsample was diluted 20x with 4% sulfuric acid solution so that the absorbance reading at 280 nm fell between 0.3 and 0.7, against a background of 4% (v/v) sulfuric acid solution. Absorbance measurements were then converted to ASL content using the following equation:

$$ASL\ (\%) = \frac{A_{280} \cdot V \cdot df}{\varepsilon \cdot ODW \cdot b} \quad \text{[Equation 4.11]}$$

where A_{280} is the average UV absorption of the sample at 280 nm; V is the volume of the filtrate sample, 0.087 L; df is the dilution factor, 20; ε is the absorptivity of the sample at 280 nm, 10 L/g/cm; ODW is the mass of the sample in grams,; and b is the path length of the cuvette used in the measurement, 1.1 cm.

Sugar content determination. The remaining filtrate samples were neutralized using calcium carbonate powder (Product No. 239216, Sigma Aldrich, St. Louis, MO) to a pH value between 5 and 6. Afterwards, the samples were centrifuged at 2000 rpm for 2 min.

to remove fine solids. The supernatant were removed and further filtered using a 0.2 μm filter and placed into sample vials for high performance liquid chromatography (HPLC) analysis. The samples were injected into an HPLC (Pump-Waters 1515, Autosampler-Waters 2707, Detector-Waters 2414, Waters, Milford, MA) with a Biorad Aminex HPX-87-P column (Part No. 125-0098, BIO-RAD, Hercules, CA) and a microguard de-ashing guard column (Part No. 125-0118, BIO-RAD, Hercules, CA) for glucan, xylan, and arabinan contents determination.

4.2.3. Statistical analyses

The composition means of each sample group were determined, compared, and tested for significance ($p < 0.05$) using Tukey's test. Statistical analyses were conducted using *R* (Version 2.15.2, 2012).

4.2.4. FT-NIR spectra collection, preprocessing, and analysis

An FT-NIR spectrophotometer (Spectrum™ One NTS, Perkin Elmer, Waltham, MA) was used to scan the dry ground botanical fractions (dry basis moisture contents were less than 2%). Approximately a 2-5 g subsample was poured in a near infrared reflectance accessory (NIRA) cup, leveled with a spatula making sure the spatula touched two points on the cup so the material was not packed, and scanned. The spectrophotometer was set to collect an average of 32 scans from 4,000 to 10,000 cm^{-1} at a spectral resolution of 4 cm^{-1}. Different blends (w/w) of the rind and blades were also scanned to demonstrate the range in composition achievable with blended fractions.

Unscrambler® (Version 10.1, Camo Software Inc., Woodbridge, NJ) was used to preprocess and analyze the spectral data. The data were preprocessed using multiplicative scatter correction (MSC) and mean centered. MSC was used to remove multiplicative

scatter or interferences resulting from baseline shifts and the sample's particle size distribution. Principal components analysis (PCA) was conducted on the preprocessed FT-NIR spectra.

4.2.5. Estimated Compositional Range (ECR)

To investigate how the composition of a bale may change as various blends of botanical fractions were incorporated into the bale, the estimated compositional range (ECR) was determined. The ECR was calculated by taking the compositions of the botanical fractions of a given sample, multiplying it by its respective mass fraction, and then dividing it by the sum of the mass fractions used:

$$ECR_{comp,i} = \frac{\sum_{i=1}^{N}(c_{fraction,i} \cdot m_{fraction,i})}{m_{total}} \qquad \text{[Equation 4.12]}$$

where $ECR_{,comp,i}$ is the ECR per component, $c_{fraction}$ is the component content for each fraction, $m_{fraction}$ is the mass of the fraction, m_{total} is the sum of the mass fractions utilized, and i is the i^{th} sample for a total of N = fraction.

Depending on harvest method and date, different botanical fractions may make up a bale so different scenarios were investigated. Early after the plant has dried down, the plant will be intact containing all five fractions: rind, node, pith, sheath, and blade. Therefore the ECR was calculated for whole Miscanthus containing all five fractions. Since the blade mass fraction data was not available when samples were collected in this study, the mass fraction of blade was assumed to be equal to the sheath's since there was at least one blade for every sheath while the plant was growing. In addition to the whole sample, samples with rind, node, pith, and sheath were made to show a late harvest variation, along with an all stalk (rind, node, pith) blend, and an all leaf (blade, sheath) blend.

The ECR means of each sample group were determined, compared, and tested for significance ($p < 0.05$) using Tukey's test. Statistical analyses were conducted using *R* (Version 2.15.2, 2012).

4.3. Results and discussion

4.3.1. Mass fraction of the botanical fractions

The majority of the Miscanthus stalks were composed of rind, the hard outer portion of the stalk, where the percent dry fraction (% w/w) ranged from 48 to 63% (Figure 4.3). All other botanical fractions, except the blade, had lower mass fractions ranging from 8 to 13% for the pith, 17 to 28% for the sheath, and 10 to 16% for the nodes. Since the stalk samples were harvested during the senescent stage, the leaves (blades) had already fallen off and were not included in the mass fraction analysis.

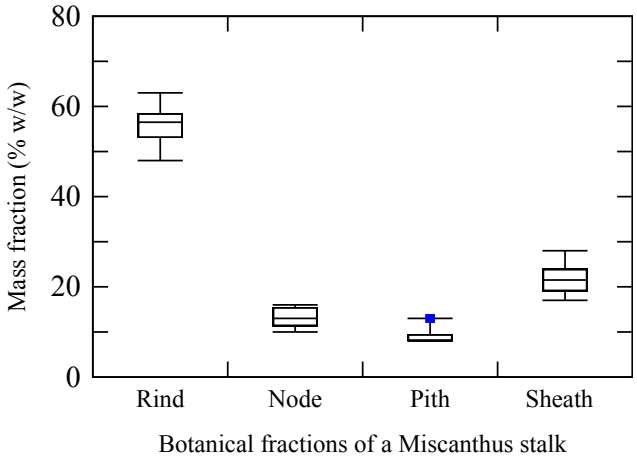


Figure 4.3. Mass fraction of botanical fractions of *Miscanthus* × *giganteus.* Median-based box plots represent the minimum, maximum, interquartile range (IQR), outliers (■, which are defined as data lying at 1.5·IQR distance from the median). The blades were excluded from the analysis as the samples were harvested at the senescent stage of plant growth.

4.3.2. Chemical composition

Across the botanical fractions, ranges for each component were as follows: glucan, 32.2 to 46.1%; xylan, 20.9 to 25.3%; arabinan, 0.0 to 6.1%; lignin, 18.7 to 25.5%; and ash, 0.4 to 8.9% (Table 4.1). While large variations were seen among individual sugars, the total structural carbohydrate levels for rind, pith, and sheath were not different from each other, showing that if glucan content (the main component in cellulose) decreased, xylan and arabinan contents (main components in hemicellulose) increased proportionally. However, nodes and leaves had the lowest total carbohydrate levels (62.0 and 59.8%, respectively) compared to rind, pith and sheath.

Table 4.1. Composition of botanical fractions of *Miscanthus* × *giganteus*

Botanical Fraction	Mean[a] ± S.D.[b] (%)						
	Glucan	**Xylan**	**Arabinan**	**Lignin**	**Ash**	**Extractives**	**Carbohydrates[c]**
Rind	46.1a ± 0.7	20.9a ± 0.3	---	24.2a ± 0.3	0.7a ± 0.1	5.2ab ± 0.5	67.0a ± 0.8
Pith	38.5c ± 0.6	25.3c ± 0.7	2.1a ± 0.3	21.1c ± 0.3	0.7a ± 0.1	6.3ad ± 1.1	64.9a ± 1.8
Sheath	40.4b ±1.1	23.5b ±0.5[b]	3.0b ± 1.6	20.5b ± 0.5	3.0c ± 0.5	4.1b ± 0.7	67.0a ± 1.5
Node	38.3c ± 0.3	23.8b ± 0.4	---	25.5d ±0.5	0.4a ± 0.1	7.6d ± 0.6	62.0b ± 0.5
Blade	32.2d ± 0.6	21.6a ± 0.5	6.1c ± 0.2	18.7e ± 0.4	8.9b ± 0.7	9.5c ± 0.8	59.8c ± 1.0

[a]Component values followed by the same lowercase letter in the same column, are not different ($p > 0.05$).
[b]S.D. = one standard deviation.
[c]Carbohydrates = glucan + xylan + arabinan contents.

Variations in composition within each botanical fraction of Miscanthus were, in general, low (Figure 4.4). The arabinan content of pith samples had the highest variability, but this was likely due to arabinan being measureable in only two of the six pith samples.

The compositions of the botanical fractions of Miscanthus were comparable to those of corn stover and switchgrass (Liu et al., 2010). Leaves had the lowest glucan

contents, respectively, in Miscanthus and corn stover and the rind had the highest glucan contents. In switchgrass, leaves had the lowest and internodes had the highest glucan contents. The glucan contents (mean ± S.D.) of the rind (37.83 ± 0.57%) and pith (39.03 ± 1.34%) of corn stover were comparable but, in the case of Miscanthus, their glucan contents were different from each other, with an absolute difference of 7.6%. The variations in glucan content of Miscanthus and corn stover were similar, 13.9% and 13.31%, respectively.

For the hemicellulose components, the pith fraction for Miscanthus, husk fraction for corn stover, and nodes fraction for switchgrass had the highest xylan contents. Liu et al. (2010) reported variations in xylan content of up to 13.48% in corn stover and 6.03% for switchgrass botanic fractions. Miscanthus was comparable to switchgrass having the largest absolute difference, 4.4%, in xylan content found in the rind and pith. Similarly, arabinan contents were low in the rind, pith and nodes of corn stover and Miscanthus and internodes of switchgrass; high levels were found in the leaves fraction of Miscanthus, husk fraction of corn stover, and nodes fraction of switchgrass. The variations in arabinan contents for all three feedstocks were below 5%.

Blades of Miscanthus had the lowest lignin content (18.7%) but, in the blades of corn stover and leaf of switchgrass, lignin contents were as high as 23.95% and 25.10%, respectively (Liu et al., 2010). In Miscanthus, the nodes had the highest lignin content (25.5%), which was comparable in lignin contents reported by Liu et al. (2010) to the nodes of corn stover (23.6%) and switchgrass (on average, 21.8%, depending on cultivars). The variations in lignin content of Miscanthus (6.8% difference between

nodes and leaves) and switchgrass (4.78% variation) were comparable, but was high in corn stover (14.26% variation).

Lastly, in terms of ash content, the three feedstocks were comparable in that the rind, node, and pith fractions had the lowest and the leaves had the highest levels. Again, the largest absolute difference (8.5%) in ash contents was between nodes and leaves of Miscanthus. This time, ash variation was comparable to corn stover (7.61%) than to switchgrass (3.79%).

These observations suggested that as a lignocellulosic feedstock, variations in Miscanthus composition were comparable to that of corn stover composition and not as consistent as switchgrass composition. This could be attributed to both Miscanthus and corn stover having more botanic fractions and, hence, specialized tissues, than switchgrass.

Overall, the carbohydrate levels in Miscanthus were higher than those in corn stover and switchgrass. Humbird et al. (2011) looked at the process design and economics for biochemical conversion of corn stover. The study set the total carbohydrate level to 59% for the analysis, which was the mean carbohydrates content of the corn stover samples used in their analysis. The mean carbohydrate levels for the fractions of Miscanthus ranged from 59.8% (blade) to 67.0% (rind and sheath), placing the lowest mean carbohydrate level fraction (blade) at the average composition of corn stover samples.

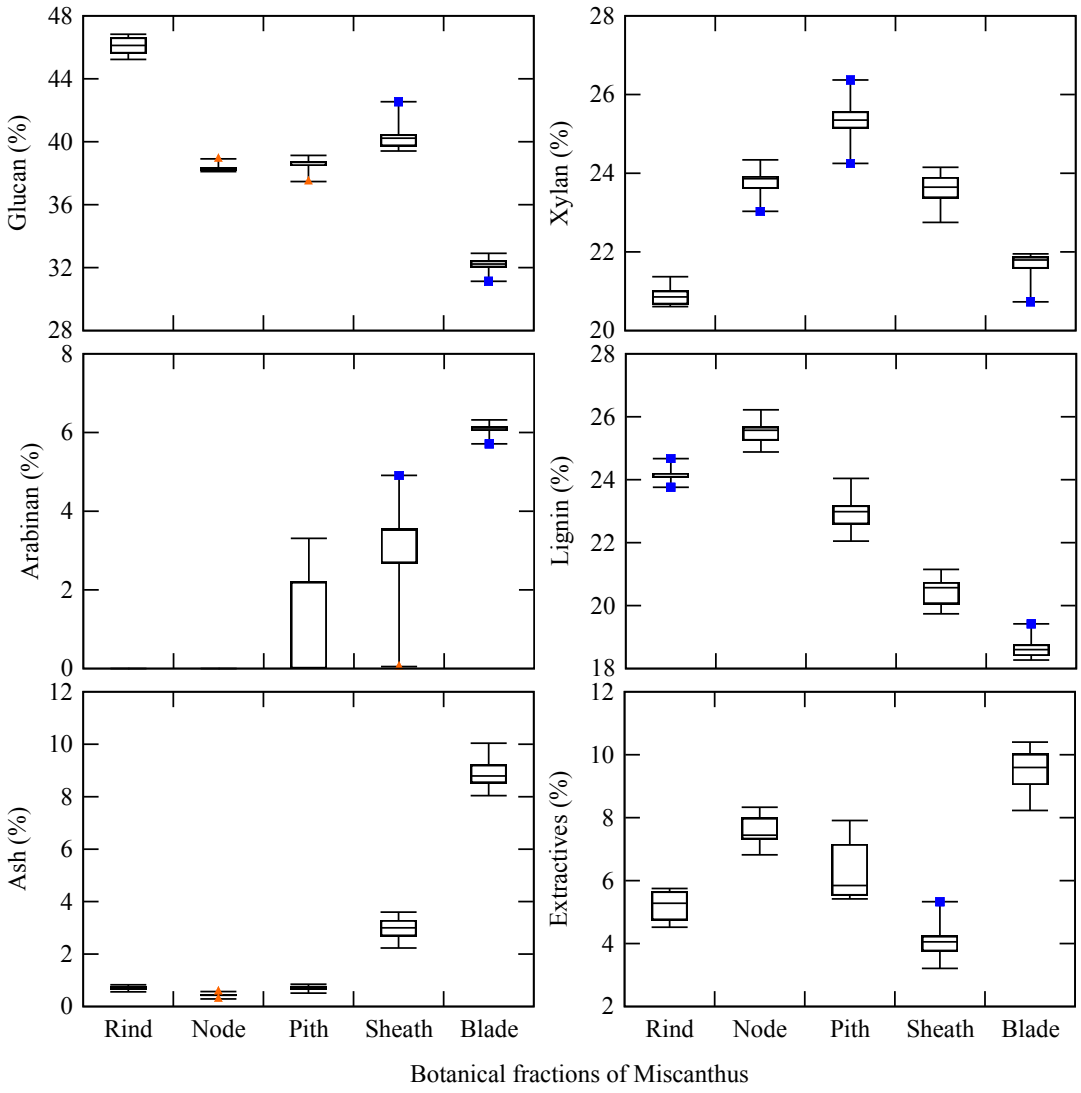


Figure 4.4. Composition of botanical fractions of *Miscanthus* × *giganteus.* Median-based box plots represent the minimum, maximum, interquartile range (IQR), outliers (■, which are defined as data lying at 1.5·IQR distance from the median), and extremes (▲, which are defined as data lying at 3·IQR distance from the median).

4.3.3. PCA of FT-NIR spectra of botanical fractions and their blends

The high variability in composition of the different botanical fractions was reflected in their FT-NIR spectra (Figure 4.5). Variations can be seen among fractions with different absorption intensities across all wavenumbers. The leaf structures – blade and sheath – exhibited additional peaks in the combination region from 5,000 to 4,000

cm^{-1}, which could be attributed to the differences in tissue types between the leaves and stalks and the lower lignin and higher arabinan and ash contents of the leaf structures.

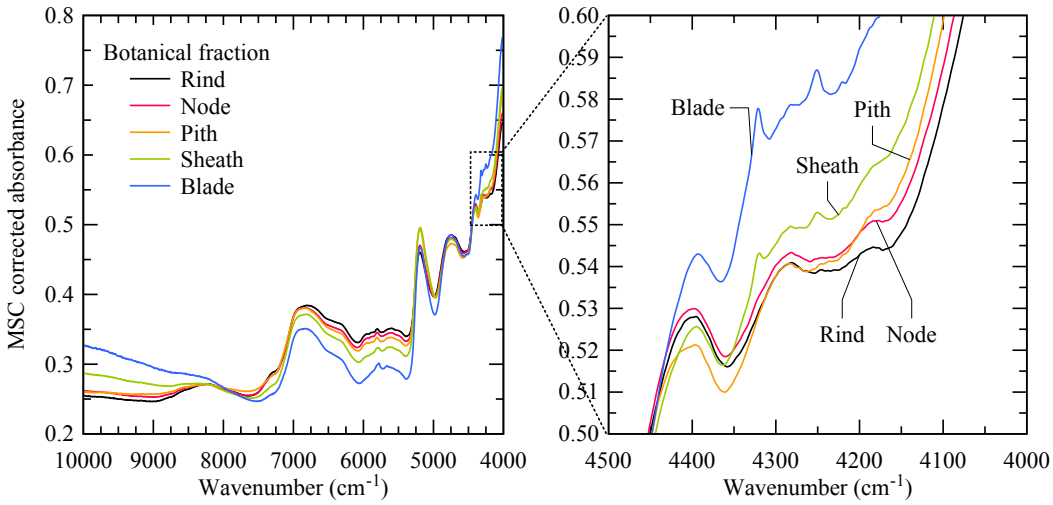


Figure 4.5. MSC spectra of each botanical fraction reflect variations in composition.

With blade and rind fractions having the largest differences in spectral response, blends with different mass fractions of blade and rind were scanned and analyzed using PCA. The FT-NIR spectra of the different blends showed the same trends as the pure botanical fractions with absorption variations across all wavenumbers and additional peaks forming in the combination region as the blade content was increased (Figure 4.6).

Recall that rind had the highest glucan and lignin contents and one of the lowest ash content; while the blade had the lowest glucan and lignin contents and the highest ash content among the botanical fractions. A plot of the first principal component (PC1) of the PCA against the third principal component (PC3) showed PC1 was sensitive to these two components (Figure 4.7), showing as the mass fraction of blade and, presumably, ash content of the blend increased, the samples had a positive PC1 value. As the rind portion of the blend increased and, presumably, its glucan and lignin content increased, the PC1 value was negative.

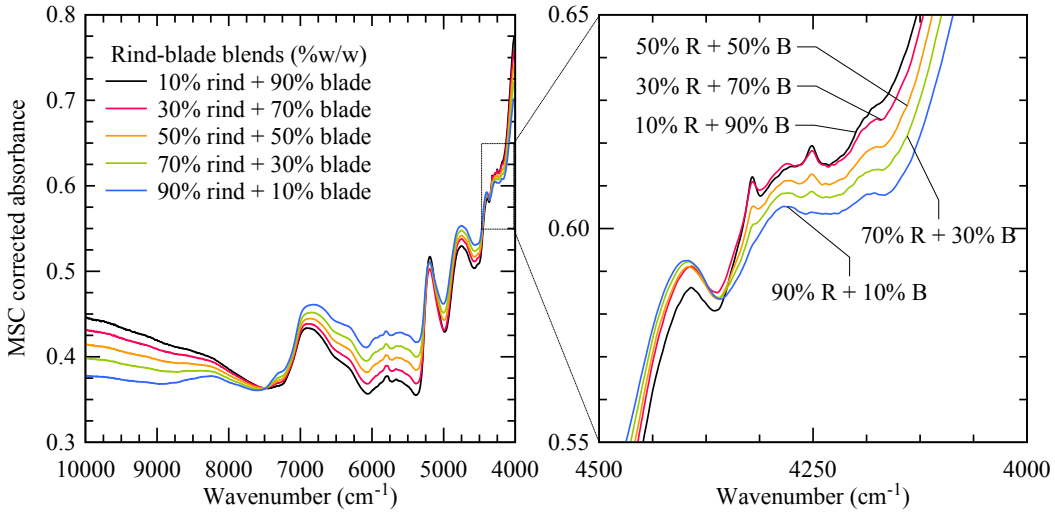


Figure 4.6. MSC corrected spectra for five rind (R) and blade (B) blends showed variations in composition and distinct absorption bands in the 4250 to 4350 cm^{-1} range as blade content increased.

Based on the mean values for glucan, lignin, and ash presented in Table 4.1 and the proportions of the blends, glucan and lignin were expected to increase from 32 to 46% and 18 to 24%, respectively, as PC1 value decreased. Likewise, ash content was expected to decrease from 9 to 1%. This result demonstrated the possibility of blending fractions to meet a specification and to use FT-NIR spectroscopy as a high throughput assay in monitoring the composition of the blends.

Combining the results from the dry mass fraction with the compositional variations, the ECR showed differences for each component across all the mixtures were 7.4% for glucan, 0.7% for xylan, 3.4% for arabinan, 4.8% for lignin, 5.3% for ash, 1.4% for extractives, and 2.5% for carbohydrates (Table 4.2). These large variations could also be leveraged in blending Miscanthus botanical fractions to meet quality specifications.

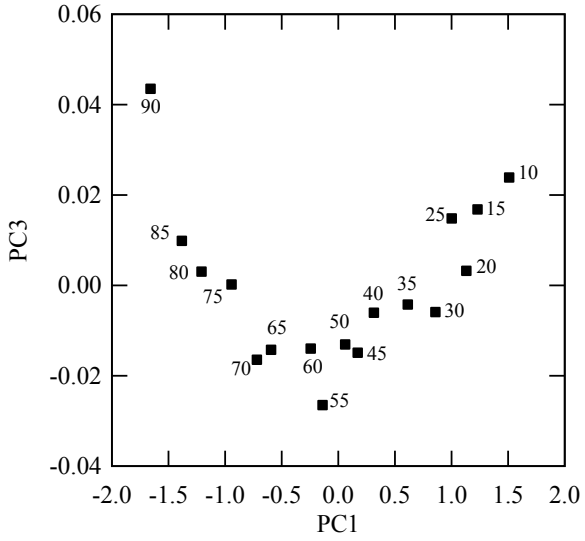


Figure 4.7. PCA showed blends of rind and blade fractions can be differentiated across the first principal component (PC1). As PC1 decreased, glucan and lignin content (associated with higher rind proportions) increased while ash content (associated with higher blade proportions) decreased. showed the different blends of rind and blade fractions. The labels represent the percentage of rind on a dry %w/w, present in the sample.

Table 4.2. Estimated compositional range (ECR) of different blends of botanical fractions of Miscanthus.

Component	ECR[a] ± S.D.[b] (%)			
	All[c]	Harvest[d]	Stalk[e]	Leaf structures[f]
Glucan	41.2a ± 0.8	43.1b ± 0.8	43.9b ± 0.7	36.3c ± 0.6
Xylan	22.1ab ± 0.3	22.3ab ± 0.3	21.9b ± 0.4	22.6a ± 0.2
Arabinan	1.75a ± 0.5	0.81a ± 0.5	0.15a ± 0.2	4.53b ± 0.8
Lignin	22.6a ± 0.3	23.4a ± 0.3	24.3a ± 0.3	19.6b ± 0.4
Ash	2.54a ± 0.4	1.17a ± 0.2	0.65a ± 0.7	5.94b ± 0.5
Extractives	6.11ab ± 0.4	5.39b ± 0.4	5.74b ± 0.5	6.79a ± 0.3
Carbohydrates[g]	65.1ab ± 0.7	66.2a ± 0.6	65.9a ± 0.6	63.4b ± 1.2

[a]Values followed by the same letter, per component (or row), are different ($p > 0.05$).
[b]S.D. = one standard deviation.
[c]All = rind (%), pith (%), node (%), sheath (%), blade (%)(w/w). All=$(m_r c_r+m_p c_p+m_n c_n+m_s c_s+m_b c_b)/m_{total}$
[d]Harvest = rind (%), pith (%), node (%), sheath (%) (w/w). Harvest=$(m_r c_r+m_p c_p+m_n c_n+m_s c_s)/m_{r+p+n+s}$
[e]Stalk = rind (%), pith (%), node (%) (w/w) Stalk=$(m_r c_r+m_p c_p+m_n c_n)/m_{r+p+n}$
[f]Leaf structures = sheath (50%), blade (50%) Leaf=$(m_s c_s+m_b c_b)/m_{s\_b}$
[g]Carbohydrates = glucan + xylan + arabinan contents (%)

4.4. Conclusions

When Miscanthus was manually harvested at the senescent stage, in terms of mass, the highest mass botanical fraction gathered was the rind, followed by node, sheath, and pith. Across the different botanical fractions, ranges for each component were as follows: glucan, 32.2 to 46.1%; xylan, 20.9 to 25.3%; arabinan, 0.0 to 6.1%; lignin, 18.7 to 25.5%; and ash, 0.4 % to 8.9%. Overall, however, the sum of glucan, xylan, and arabinan contents for rind, pith, and sheath fractions were not different from each other. While the compositional variation within each botanical fraction were not high, the variations across some botanical fractions were significant. The blade had the lowest glucan, lowest lignin, and highest ash contents making them different from the other botanical fractions. Variations in Miscanthus composition were comparable to that of corn stover composition and not as consistent as switchgrass composition. From the FT-NIR spectra, variations among botanical fractions can be detected through PCA, showing the possibility of constructing classification models such as linear discriminant analysis and soft independent modeling class analogy (SIMCA) based on different blends. With the variations that were observed it is now evident that the compositional variability within a bale core sample can be largely contributed to which fractions make up the core sample.

CHAPTER 5. PLSR MODELS OF MISCANTHUS COMPOSITION

5.1. Introduction

Since conversion facilities would like to receive feedstocks that are consistent, or uniform, in quality, moisture content, ash content, and convertible carbohydrates so they can operate their chemical pretreatment and conversion processes efficiently. It is advantageous to know the feedstock composition prior to conversion so that enzyme mixtures, yeast strains, and process control parameters can be adjusted accordingly to maximize yields. Knowing composition at earlier stages of the supply chain can also help in the development of quality based valuations which incentivize farmers and suppliers to implement best management practices to ensure a uniform and consistent supply system (Kenney et al., 2013). Current wet chemistry methods for chemical characterization of biomass feedstock are not applicable for field or inline monitoring because they are expensive, labor intensive, and cannot provide compositional information in real time for process control (Ye et al., 2008). One approach to reducing the time and cost of compositional analysis is the development of near infrared (NIR) spectroscopy and a good calibration based on multivariate analysis to provide either a quantitative or qualitative measure of composition.

Several studies have shown NIR spectroscopy as a promising technique to assessing biomass composition (Sanderson et al., 1996; Hames et al., 2003; Pordesimo et al., 2005; Hoskinson et al., 2007; Templeton et al., 2009; Hayes, 2012; and Haffner et al., 2013). In this study, calibration models based on partial least squares regression (PLSR) of FT-NIR spectra of Miscanthus to chemical composition (glucan, xylan, arabinan, lignin, ash, extractives, and acetyl contents) were developed. In order to get a broad range

in the composition of the samples in the calibration data, samples from Miscanthus bales stored under a variety of conditions – indoors, under roof, outdoors with tarp cover, and outdoors without tarp – for 3 to 24 mo. were used. To date, most calibration models have been based on freshly harvested biomass or using different cultivars of a specific biomass crop, e.g., upland vs. lowland cultivars of switchgrass.

5.2. Materials and methods

5.2.1. Sample collection, preparation, and compositional analysis

Bale core samples were collected and prepared for chemical compositional analysis and NIR scanning according to the procedure outlined in Section 3.2. The samples were sent to the Energy Biosciences Institute (EBI) Analytical Chemistry Laboratory at the University of California in Berkeley for compositional analysis (glucan, xylan, arabinan, lignin, ash, extractives and acetyl content). Compositional analyses were conducted in duplicates following standard procedures developed by the National Renewable Energy Laboratory (NREL) and details of which were discussed in Haffner et al. (2013).

5.2.2. Scanning and preprocessing of FT-NIR spectra

An FT-NIR spectrophotometer (Spectrum^{TM} One NTS, Perkin Elmer, Waltham, MA) was used to scan the dry ground samples (dry basis moisture contents were less than 2%). Approximately a 15 g subsample was poured in a near infrared reflectance accessory (NIRA) cup, leveled with a spatula, and placed in an automatic spinner. The spectrophotometer was set to collect an average of 32 scans from 4,000 to 10,000 cm^{-1} at a spectral resolution of 4 cm^{-1}.

Unscrambler® (Version 10.1, Camo Software Inc., Woodbridge, NJ) was used to preprocess and analyze the spectral data. The data were mean centered and either preprocessed using multiplicative scatter correction (MSC); Savitzy-Golay (SG) first derivative using a second, third, or fourth order polynomial; or combination of MSC and SG. MSC was used to remove multiplicative scatter or interferences resulting from baseline shifts and the sample's particle size distribution. The SG derivative algorithm does smoothing and differentiation of the data by simplified least squares, where each point (i.e., the absorbance at a specific wavenumber) became the weighted average derivative of the points surrounding it.

5.2.3. Absorption waveband selection and PLSR modeling

PLSR models for each component were first developed in Unscrambler® using 91 samples for calibration and validated using cross validation. With the cross validation results a Martens uncertainty test was performed. A separate test set, including 10 samples, was created by taking Miscanthus samples that were at the extremes of the range for glucan, xylan, arabinan, lignin, and ash. The test set remained fixed in the development of PLSR models for all components.

A series of cross validation models was used to detect null regression coefficients with the Martens uncertainty test option (Figure 5.1; Esbensen et al., 2002). Null regression coefficients were left out in the next model and the software highlighted wavebands that were found to be correlated to the specific component being modeled (e.g., glucan, xylan, lignin, etc.).

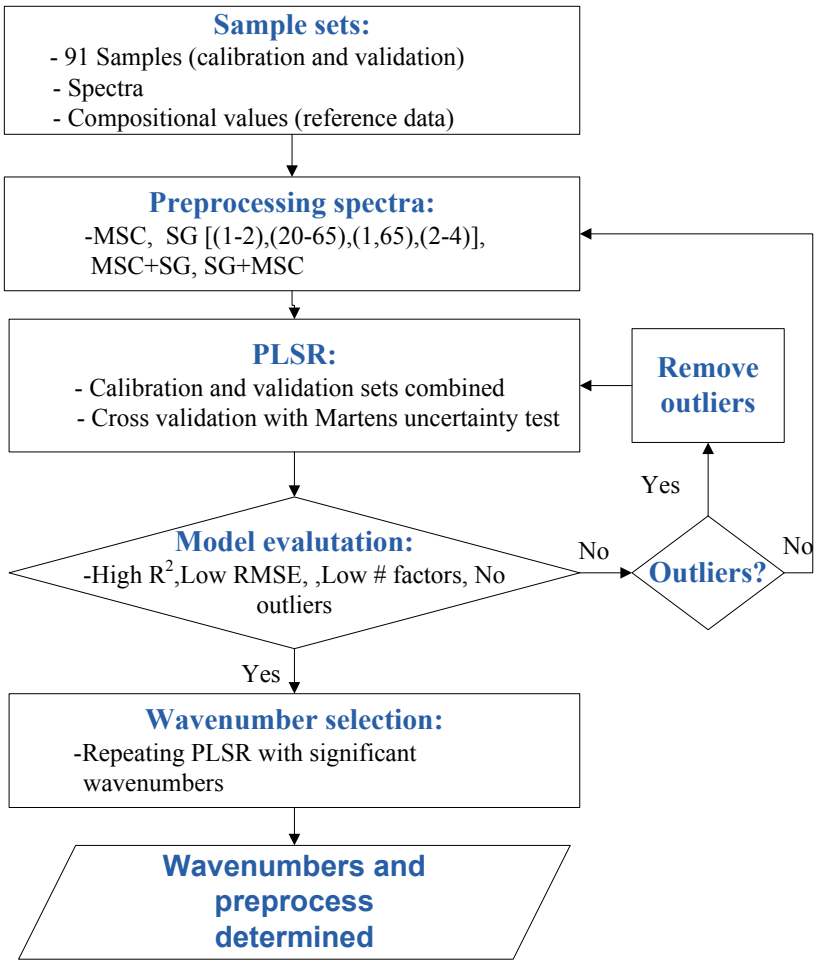


Figure 5.1. Wavenumber and preprocessing selection flow chart.

Once significant wavebands have been identified, PLSR models were re-calculated using only the significant wavebands and the 91 samples were divided into a calibration set (67 samples) and validation set (24 samples; Figure 5.2). In assigning calibration and validation sets for each component, the Miscanthus samples were arranged from highest to lowest values and every 4th or 5th sample in the range was used for validation. The remaining samples were used for calibration. This procedure ensured the calibration and validation sets for each component had a comparable range, mean, and standard deviation.

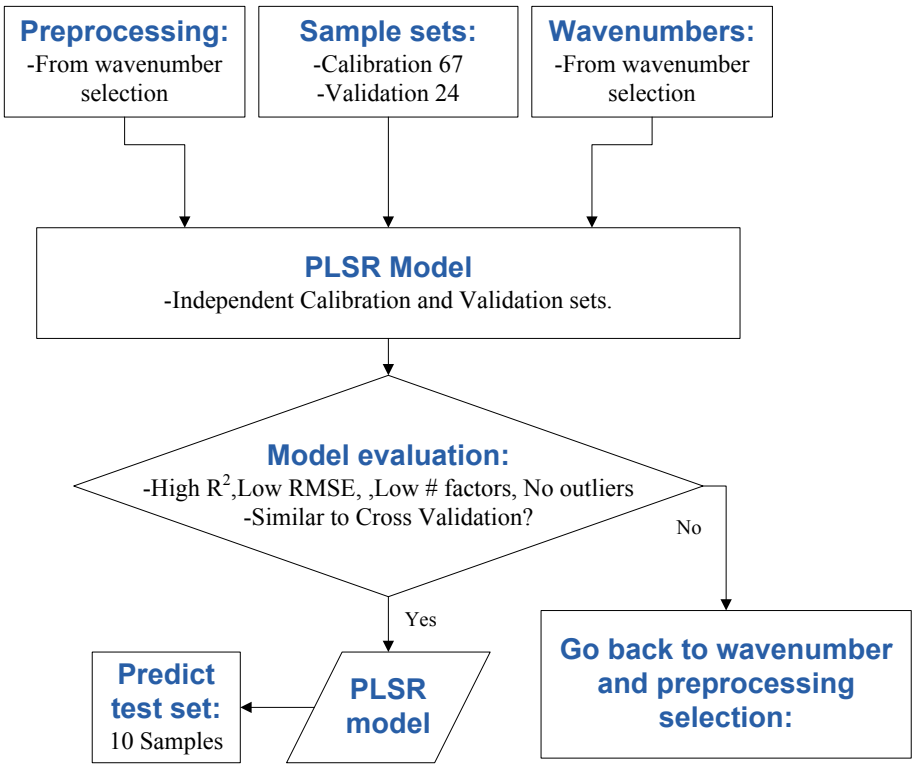


Figure 5.2. Flow chart of independent calibration and validation PLSR model building.

The resulting PLSR models were also compared to regression models for Miscanthus and other lignocellulosic feedstocks found in literature. Calibration models based on NIR spectra have a high specificity and selectivity and are rarely transferrable from one sample type to another; meaning, a model developed for corn stover is rarely applicable to Miscanthus or a model developed using 10 mm ground material may not be applicable to the same material that has been ground to a smaller or larger particle size. However, since lignocellulosic materials essentially have the same components – cellulose, hemicellulose, lignin, and ash – the PLSR models predicting composition should rely on similar NIR absorption wavebands.

The PLSR models were then compared and evaluated based on the number of factors used and corresponding explained variance, coefficient of determination (R^2), root

mean square error (RMSE), ratio of performance to deviation (RPD), range to standard error of prediction (R/SEP), and relative ability of prediction (RAP). Based on the waveband comparison and evaluation results, the *best* PLSR model for each component was chosen and used to predict the composition of an independent test data set (Figure 5.2).

5.3. Results and discussion

5.3.1. Descriptive statistics

For all components, ranges, means, and standard deviations of the samples chosen for the calibration, validation, and fixed test sets were comparable to each other (Table 5.1; kurtosis values for each component in each data set are available in Appendix, Table A.2). Differences were found in the kurtosis, K or degree of "peakedness", of the probability distribution of each data set for each component. A probability distribution with a $K = 0$ value represents a typical normal distribution curve. Data sets with a high kurtosis values ($K > 1$) refer to data sets centered around the mean while data sets with low kurtosis values ($K < -1$) have a flatter, more uniform distribution across the range probability distribution curves.

In the case of glucan PLSR modeling, the kurtoses of the calibration, validation, and test sets were different, with $K = 19.2$, 10.6, and -1.4, respectively. When high kurtosis is combined with a high range and low standard deviation, there is a tendency to overestimate the R/SEP values of the prediction models. This combination was found with the glucan, arabinan, and lignin calibration data sets where $K > 3$; for xylan, acetyl, ash, and extractives, $K > 1$. With high kurtosis the standard deviation of the data set will be narrower than the normally distributed data sets that are desired in NIR model building

sets. Therefore with the lower standard deviation the RPD will usually be under estimated since the standard deviation is in the numerator.

Table 5.1. Range, mean, and standard deviation (S.D.) of the composition of Miscanthus bale samples

Component	Calibration set (n = 67)		Validation set (n = 24)		Test set (n = 10)	
	Range (%)	Mean ± S.D. (%)	Range (%)	Mean ± S.D. (%)	Range (%)	Mean ± S.D. (%)
Glucan	26.5-44.0	40.7 ± 2.37	25.8-43.6	39.8 ± 3.70	36.9-43.7	40.5 ± 2.49
Xylan	17.7-24.2	20.6 ± 1.20	19.3-23.8	20.7 ± 1.02	18.1-24.2	21.6 ± 2.34
Arabinan	1.05-3.11	1.83 ± 0.36	1.2-2.71	1.85 ± 0.34	1.21-2.75	1.99 ± 0.50
Acetyl	2.06-3.40	2.74 ± 0.28	1.76-3.29	2.81 ± 0.38	2.09-3.01	2.61 ± 0.34
Klason lignin	17.7-26.5	20.7 ± 1.35	18.2-25.7	20.5 ± 1.61	17.5-22.2	20.1 ± 1.48
Ash after extraction	0.59-5.16	2.60 ± 1.64	0.59-4.47	2.53 ± 0.91	0.59-4.25	2.26 ± 1.46
Extractives	3.61-8.72	5.59 ± 0.86	3.59-9.10	5.66 ± 1.16	4.64-6.60	5.61 ± 0.57

5.3.2. Significant absorption wavebands and PLSR modeling

The cross validation with Martens uncertainty testing allowed for identification of significant absorption wavebands for each component of Miscanthus (Table 5.2). The glucan PLSR model utilized 35 wavenumbers; of the regions found useful for modeling, only the 8,250 to 8,260 cm^{-1} was similarly identified as cellulose by Schwanninger et al. (2011). Three other wavenumbers were recognized in Aenugu et al. (2011) as a ketone, methyl, and methylene. The xylan and arabinan PLSR models consisted of 351 and 57 wavenumbers, respectively, where within the regions utilized, 11 and 4 wavenumbers had been attributed to hemicellulose (Schwanninger et al., 2011). The lignin PLSR model utilized the entire 4,000 to 5,000 cm^{-1} region; therefore, the entire region was utilized. Schwanninger et al. (2011) reported five wavenumbers to be specific in that region. Finally, even though ash is an inorganic material, there was a P-OH bond that was reported by Aenugu et al. (2011) that matched up with one of the regions utilized for ash.

Table 5.2. Significant wavebands identified through cross validation with Martens uncertainty testing and their corresponding component or chemical bonds.

Wavebands (cm^{-1})		
This study Based on Miscanthus samples	Aenugu et al. (2011) Based on fundamental groups	Schwanninger et al. (2011) Based on wood samples
Glucan 8605-8602, 8255-8253, 5954-5941, 5922-5918, 5856-5848, 4166-4164	8605 C=O 4th overtone 8253 CH_2 2nd overtone 1708 CH_3 1st overtone	Cellulose 8250-8260
Xylan 7064-7022, 5310-5222, 5190-5172, 5155-5140, 5110-5100, 5047-5038, 4829-4813, 4714-4694, 4450-4415, 4406-4360, 4340-4314, 4186-4184, 4153-4133		Hemicellulose 5848, 5865, 5814, 5816, 5245, 4686, 4546, 4435, 4404, 4401, 4392
Arabinan 5860-5858, 5818-5817, 5780-5779, 5100-5099, 4684-4683, 4635-4631, 4585-4582, 4549-4543, 4486-4481, 4412-4407, 4392-4390, 4372-4367, 4340-4338, 4316-4314, 4239		Hemicellulose 5865, 5818, 4686, 4404
Lignin 5000-4030		Lignin 4686, 4546, 4411, 4280, 4195
Ash 5312-5223, 4779-4662, 4419-4355, 4277-4248, 4115-4178	5240 P-OH	

Multiple pretreatments and combinations of pretreatments, i.e., MSC, SG first or second derivatives, MSC followed by SG, and SG followed by MSC, were tested. While many papers have used the second derivative, the best results were obtained from the first derivative in this work. Hayes (2012) reported success with using derivative smoothing alone, without the use of MSC. However his models also utilized over eight factors for glucan, xylan, arabinan, lignin, ash, extractives, and acetyl.

Table 5.3. PLS regression models for glucan, xylan, arabinan, lignin, ash, extractives, and acetyl contents of Miscanthus.

	Glucan	Xylan	Arabinan	Lignin	Ash	Extractives	Acetyl
Preprocessing	MSC + SG (1,20,20,3)[a]	SG (1,30,30,3) + MSC	SG (1,50,1,4) + MSC	SG (1,30,30,3) +MSC	MSC + SG (1,65,65,4)	MSC	SG (1,40,40,4) + MSC
No. of factors	2	5	4	4	7	2	2
R^2							
Calibration	0.93	0.87	0.85	0.87	0.84	0.04	0.47
Validation	0.95	0.86	0.82	0.93	0.64	0.14	0.16
Test	0.52	0.60	0.69	0.70	0.85	N/A	N/A
RMSE[b] (%)							
Calibration	0.60	0.38	0.13	0.39	0.33	0.90	0.20
Validation	0.76	0.25	0.14	0.43	0.53	1.04	0.33
Test	1.62	1.39	0.26	0.75	0.53	N/A	N/A
Bias[c] (%)	0.10	-0.05	-0.04	0.06	-0.20	-0.002	0.06
RPD[d]	4.59	4.20	2.42	3.52	1.86	N/A[g]	0.80[g]
R/SEP[e]	22.13	18.75	11.61	16.38	8.05	N/A	3.55
RAP[f]	0.96	0.95	0.83	0.94	0.70	N/A	0.00

[a]Savitzky-Golay derivative parameters (derivative order, left points, right points, polynomial order)

[b]Root mean square error, $RMSE = \sqrt{\sum_{i=1}^{N}(y_i - \hat{y})^2/N}$, where y_i is the individual reference value (from wet chemistry assay), $\hat{y}$ is the NIR predicted value from the PLS regression model, and N is the total number of samples.

[c]$Bias = \sum_{i=1}^{N}(\hat{y} - y_i)/(N-1)$

[d]Ratio of performance to deviation, $RPD = \sigma_y/SEP$, where the standard deviation of the reference data is $\sigma_y = \sqrt{\sum_{i=1}^{N}(y_i - \bar{y})/(N-1)}$, $\bar{y}$ is the average reference value, and the standard error of prediction or prediction (using the validation set) is $\sqrt{\sum_{i=1}^{N}(y_i - \hat{y} - Bias)/N}$.

[e]Ratio of the range of the validation data set to the standard error of performance, SEP.

[f]RAP$= \frac{\sigma^2_{Validation} - MSEP}{\sigma^2_{Validation} - \sigma^2_{wet\ Chem}}$

[g]RPD values of models for extractives and acetyl contents were less than 1.0 and not used for prediction.

PLSR models were created for all the components but models for extractives and acetyl had *RPD* < 1. Moldy samples, which could provide an overestimate of extractives content, were removed from the calibration and validation data sets in the extractives PLSR modeling, but resulting models still had low *RPD* values. In the case of the acetyl model, the Miscanthus samples used had a very narrow range for acetyl content and a high mean standard deviation for the wet chemical values (Table 5.1).

The glucan model was the best of all the models that were constructed, with a calibration and validation R^2 = 0.93 and 0.96 and *RMSE* = 0.61 and 0.76%, respectively. The bias for calibration and validation were also low (< 0.1%). The *R/SEP* signified this model useful for quantification and, based on the *RPD,* it was also useful for sample screening in plant breeding programs. The difference between the two evaluation parameters is that the range is high compared to the standard deviation or in other terms the data set is not evenly distributed and has a high kurtosis. In addition, the RAP is at 0.96 in which a RAP of 1 means that the only variation in the model is due to the wet chemical analysis.

Similar to the glucan model, the lignin model was very good utilizing only 4 factors. The R^2 for the calibration and validation were 0.87 and 0.83, respectively. The RMSE was also low, being less than 0.43 for both the calibration and validation sets. The RPD and R/SEP was 3.52 and 16.38, showing that at a minimum the model was good for screening and at the best was good for quantification. The RAP also concluded a good model at 0.94.

Similarly, models developed for xylan and arabinan utilized five factors each, calibration and validation $R^2 > 0.80$, and a *RMSE* < 0.4%. Based on its *R/SEP* and *RPD* values, the PLSR model for xylan was at useful for quantification and sample screening while the PLSR model for arabinan was high enough for quality control applications.

The PLSR model for ash was a not as good as the other models due to a lower *RPD*, *R/SEP*, *RAP*, and R^2 for validation, but was still promising and likely to improve if more samples and a wider range of ash contents are included in the model. The R^2 was high with the calibration set (0.84) but low for the validation set (0.64). There was a large

difference between the *RMSE* for the calibration and validation models, 0.33% and 0.53%, respectively. The *R/SEP* value signified the model was good for screening purposes, while the *RPD* value was too low for this application. This discrepancy could be attributed to the low range in ash content and that ash, as an inorganic material, did not have enough organic compounds to produce vibrations in the NIR range.

Comparing the results to other PLSR models developed for biomass, Wolfrum and Sluiter (2009) created a model to predict the composition of corn stover. Their model included samples with glucan contents between 25.7 – 41.4% and a *RMSEP* = 1.96%; while the ranges were comparable, the *RMSEP* (*RMSE* of prediction, which refers to the validation set) was almost double than the one for this study. There was a wider range in xylan content (11.2 – 30.8%) and a *RMSEP* = 1.33%; again, double of the *RMSEP* found in this study. The lignin model was similar to the xylan model in that the range was wider and the RMSEP was three times the value determined in this study.

While models for ash, extractives and acetyl were attempted in this study, they were unreliable with $R^2 < 0.50$. The main difference between the two studies was that cross validation was used on a total of 77 samples in Wolfrum and Slutier (2009), where this study used an independent validation and calibration set of 67 and 24 samples. Also in this work insignificant wavenumbers were determined and not utilized which reduced some of the noise in the model.

Ye et al. (2008) constructed corn stover models, too, using NIR-PLSR and were more successful in terms of *R/SEP*. The results were similar to those in this study where the *R/SEP* for glucan was 14.31 vs. 22.13 (Miscanthus); xylan, 12.97 vs. 18.75

(Miscanthus); arabinan, 13.96 vs. 10.92 (Miscanthus); lignin, 13.71 vs. 16.38 (Miscanthus); and ash, 7.05 to 7.40 (Miscanthus).

Haffner et al. (2013) constructed NIR-PLSR models for Miscanthus from newly harvested samples. Their samples were ground to a particle size less than 2 mm using a ball mill and only a 0.5 g sample was scanned. Recall that, in this study, samples were ground using a knife mill fitted with a 2 mm sieve and a 15 g sample was scanned using a spinner accessory. The PLSR models for glucan and xylan in this study were comparable in *R/SEP* values to those developed by Haffner et al. (2013), but they reported higher *R/SEP* values for arabinan, lignin, and ash models. They were also able to construct models for extractives and acetyl content in Miscanthus, likely owing using clean samples from the field that were ball milled. Also in this study a 15 g sample was taken to get an average composition of the sample while in Haffner et al. (2013) they only scanned a 0.5g sample. Hayes' (2012) NIR-PLSR models for Miscanthus utilized independent calibration and validation sets with comparable *RPD* values for glucan and xylan models in this study, but he, too, reported better models for arabinan, lignin, ash, and extractives (*RPD* > 3). Like Haffner et al. (2013), Hayes' (2012) samples were newly harvested and the models included multiple factors (greater than 5 and as high as 17).

When the PLSR models were used to predict the composition of the test set, the R^2 values were 0.52 for glucan; 0.60 for xylan; 0.69 for arabinan; 0.70 for lignin; and 0.85 for ash. The lower R^2 values could be attributed to the low number of samples in the set, or that the samples were at the extremes of the glucan, xylan, arabinan, lignin, and ash data sets and could have been contaminated with foreign material like microbial growth.

5.4. Conclusions

PLSR models were created for glucan, xylan, arabinan, lignin, and ash with *RPD* values of 4.59, 4.20, 2.46, 3.52, and 1.86, respectively, with *R/SEP* > 8. These values indicated the models (except for ash) were useful, at minimum, for screening samples in a plant breeding program. PLSR models for extractives and acetyl were also developed but, since the mean deviation of the references data was high compared to the mean and range, the resulting models had a RPD < 1. These models could be improved by increasing the data sets and utilizing botanical fractions, which have a wider range of composition than bale core samples.

CHAPTER 6. EFFECTS OF PARTICLE SIZE ON PREDICTING MISCANTHUS COMPOSITION

6.1. Introduction

Most applications of NIR spectroscopy in agriculture depends on diffuse reflectance and scattering of infrared light. Light reflectance in itself is a complicated phenomenon and scattering due to size and shape variations in particulate samples can cause shifts in the reflected light (R). These shifts are often perceived as "noise" in the absorbance ($\log 1/R$) measurements and distort the information correlated to the sample's chemical composition (Isaksson and Næs, 1988). To correct spectral noise, pretreatments such as MSC, standard normal variate-detrend (SNV-D), and derivative smoothing functions are used (Geladi et al., 1985).

Another way to mitigate spectral noise is to get a better understanding of how particle size distributions affect light reflection and corresponding absorbance in the first place. Rantanen et al. (2000) measured the NIR absorbance of glass Ballotini of different particle sizes and noted smaller particles absorbed less light and tended to have lower baseline shifts. Furthermore, longer wavelengths were more affected by the particle size, which demonstrated particle size distribution effects were not necessarily linear, but these effects could be eliminated by the MSC pretreatment. Conversely, Chen and Thennadil (2012) demonstrated how MSC pretreated spectra could be used to predict particle size. They found that light scattering effects were also affected by particle concentration and their refractive indices.

In this study, the effects of the particle size distribution of the 2 mm ground Miscanthus bale core samples on the NIR spectra and predicted composition were investigated. These samples are not only a mixture of botanical fractions, but tend to have

a wide particle size distribution since the material can fall vertically through the sieve during milling, i.e., a material with a longitudinal axis longer than 2 mm can still fall through the sieve.

6.2. Materials and methods

6.2.1. Miscanthus samples and collection of FT-NIR spectra

Bale core samples (40 g) were collected and prepared according to procedures described in Section 3.2 from stacked bales in Groups 8-13 (Z284-86, Y287-89, X290-92, O293-95, T296-98, I299-301). Once the samples were dried and ground to 2 mm, they were placed in a near infrared reflectance accessory (NIRA) cup, leveled with a spatula, and scanned with an FT-NIR spectrophotometer (Spectrum^{TM} One NTS, Perkin Elmer, Waltham, MA). The spectrophotometer was set to collect an average of 32 scans from 4,000 to 10,000 cm^{-1} at a spectral resolution of 4 cm^{-1}.

A subsample (2 to 4 g) was ran in a sonic sifter separator (L3P, ATM Corporation of America, Coraopolis, PA) fitted with U.S standard sieve sizes of 18 (1.12 mm), 20 (0.850 mm), 40 (0.425 mm), 60 (0.250 mm), 120 (0.125 mm). A pan was also included in the separator to catch any remaining sample after sieving. The sonic sifter was run for 2 minutes and the particle mass in each sieve was weighed and recorded. The procedure was repeated for two additional replicates. Afterwards, the FT-NIR spectra of the sieved samples were collected.

6.2.2. Geometric mean diameter

The geometric mean diameters and geometric standard deviations were calculated according to ASABE Standard S319.3 (1997):

$$d_{gw} = \log^{-1}\left[\frac{\sum_{i=1}^{n}(W_i \log \overline{d_i})}{\sum_{i=1}^{n}(W_i)}\right]^{-1} \qquad \text{[Equation 6.1]}$$

$$S_{log} = \log^{-1}\left[\frac{\sum_{i=1}^{n} W_i(\log \overline{d_i} - \log d_{gw})^2}{\sum_{i=1}^{n}(W_i)}\right]^{1/2} \quad \text{[Equation 6.2]}$$

$$S_{ln} = 2.3 * S_{log} \quad \text{[Equation 6.3]}$$

$$S_{gw} = \frac{1}{2} d_{gw}[\log^{-1} S_{ln} - (\log^{-1} S_{ln})^{-1}] \quad \text{[Equation 6.4]}$$

where d_{gw} is the geometric mean; W represents the weight on the i^{th} sieve; $\overline{d_i}$ is the mean between the i^{th} sieve and the sieve above; S_{log} is the geometric standard deviation of the log-normal distribution by mass in $\log_{10}$; S_{ln} is the geometric standard deviation in natural log scale; and S_{gw} is the geometric standard deviation of particle diameter by mass.

6.2.3. Prediction of composition with PLSR models and statistical analysis

The spectra were then imported into Unscrambler® and the compositions were predicted using the PLSR models described in Chapter 5. The predictions for the non-sieved samples and for each particle size were compared using Students *t* test and Tukey's test at an alpla level of 0.05. Statistical analyses were conducted using *R* (Version 2.15.2, 2012).

6.3. Results and discussion

6.3.1. Geometric mean diameter

The mass fraction for each particle sized ranged from 19.6 to 25.2% for the #20 sieve, 32.7 to 36.1% for the #40 sieve, 20.2 to 22.3 for the #60 sieve, 12.3 to 15.8% for the #120 sieve, and 5.4 to 9.3% for the pan (Figure 6.1). There were no difference between the mass fractions for a given particle size. However, bale samples from Groups 11 to 14 had less material in sieve #120 and in the pan compared to bale samples from Groups 8 to 10). Within year, the bales stored under roof and with tarp cover had a higher mass percentage means than bales stored outdoors. The geometric means ranged

from 0.36 to 0.49 mm, where the smallest diameter came from the bale that was stored outside for 17 mos., and the largest diameter came from a bale that was covered with a tarp and was only 5 mos. old.

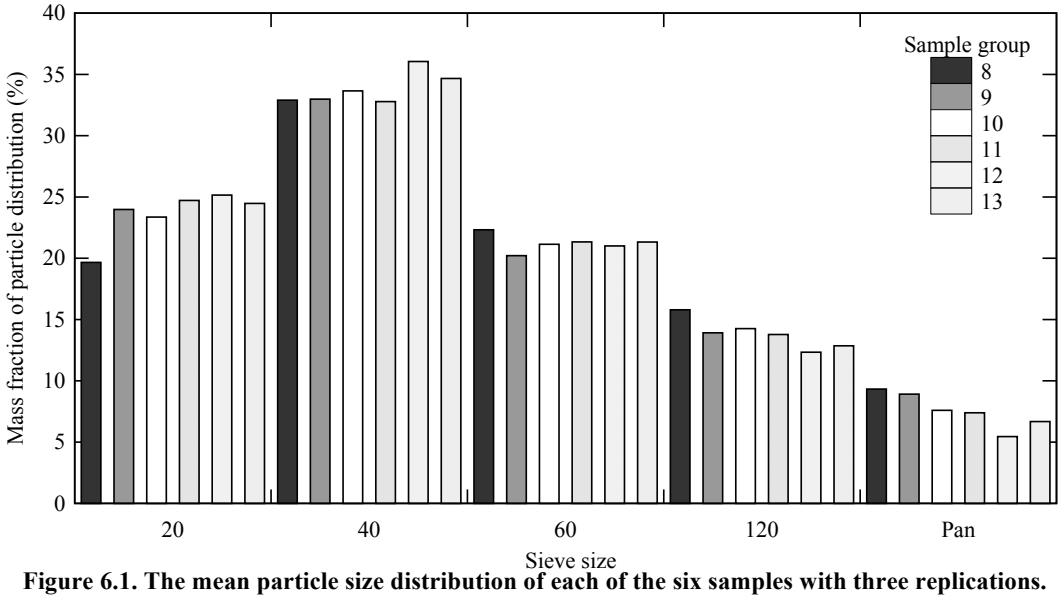


Figure 6.1. The mean particle size distribution of each of the six samples with three replications.

Table 6.1. Geometric mean diameter and geometric standard deviation of 2 mm ground Miscanthus.

Group	Age (mo.)	Storage conditions	d_{gw} (mm)				S_{gw} (mm)			
			1[a]	2[a]	3[a]	Mean ± S.D.	1[a]	2[a]	3[a]	Mean ± S.D.
8	17	Outside	0.42	0.36	0.42	0.40 ± 0.03	0.42	0.44	0.42	0.43 ± 0.01
9	17	Tarp	0.38	0.44	0.48	0.43 ± 0.05	0.44	0.41	0.40	0.42 ± 0.02
10	17	Under roof	0.44	0.46	0.41	0.43 ± 0.03	0.41	0.40	0.43	0.41 ± 0.02
11	5	Outside	0.49	0.43	0.41	0.44 ± 0.04	0.39	0.42	0.43	0.41 ± 0.02
12	5	Tarp	0.49	0.49	0.44	0.47 ± 0.03	0.39	0.39	0.41	0.40 ± 0.01
13	5	Under roof	0.46	0.45	0.45	0.45 ± 0.01	0.41	0.41	0.41	0.41 ± 0.00

[a]Replicate number.

A general trend for the older bales can be observed where the unprotected bale had a smaller particle size, which may indicate decay, while bales covered with a tarp or stored under roof had larger particle sizes. All the samples except the new bale stored under roof had a wide standard deviation of the average geometric mean diameters.

6.3.2. Prediction of composition with PLSR models

The FT-NIR spectra of the samples in each sieve showed the region from 4,500 to 10,000 cm^{-1} included additive and/or multiplicative scatter effects for which the MSC algorithm was not able to correct (Figure 6.2). In the region from 4,000 to 4,500 cm^{-1}, the non-sieved samples and those left in sieve sizes #20 and #40 showed increasing absorbance measurements as wavenumber decreased while the rest of the samples showed a negative or decreasing absorbance. This trend was also observed by Rantanen et al. (2000) – longer wavelengths tended to be more affected by particle size. Since the PLSR models developed for Miscanthus composition relied heavily on this combination region (4,000 to 5,000 cm^{-1}), the accuracy of their prediction is expected to be size-dependent.

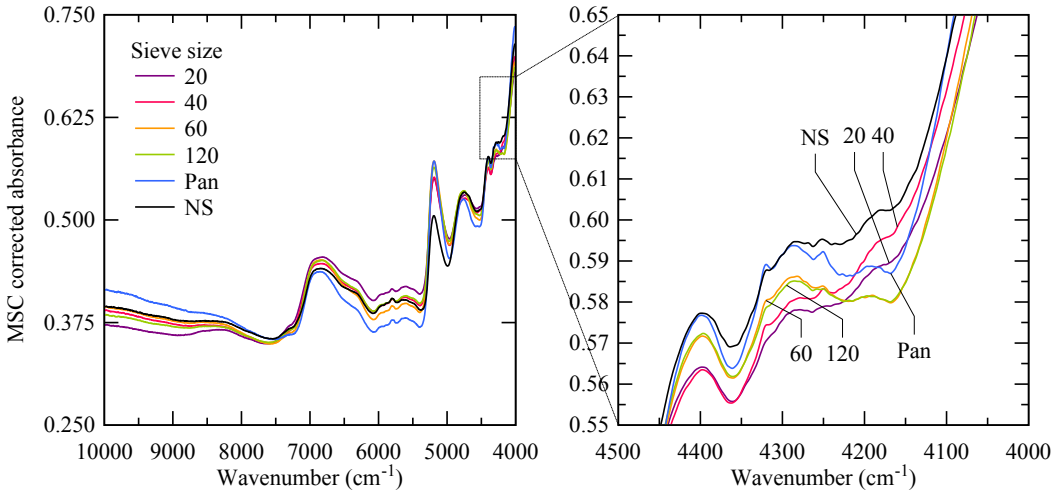


Figure 6.2. MSC corrected spectra of non-sieved (NS) samples and samples in each sieve size (#20, 40, 60, 120, and pan).

In general, as particle size decreased (i.e., sieve size number increased), the predicted composition also increased (Figure 6.3). The variability in predicted glucan content was comparable or all bale samples, while the variability in predicted xylan

content was larger with younger (Groups 11-13) than older bale (Groups 8-10) samples. Predicted arabinan content increased with decreasing particle size with the older bale samples but this trend was not apparent with the younger bale samples. As for lignin content, predictions based on the FT-NIR spectra of sieved younger bale samples were lower than the prediction based on the FT-NIR spectra of non-sieved older bale samples. The same trend was observed for older bale samples but the predicted lignin content was within 1% absolute difference of the predicted lignin content of the non-sieved samples. The predicted ash content of the bales, regardless of storage period, increased with decreasing particle size.

On average, across predictions of glucan, arabinan, and lignin content were not sensitive to particle size (Table 6.2). The predicted xylan content using non-sieved samples was different from sieved samples. Predicted ash content increased with decreasing particle size.

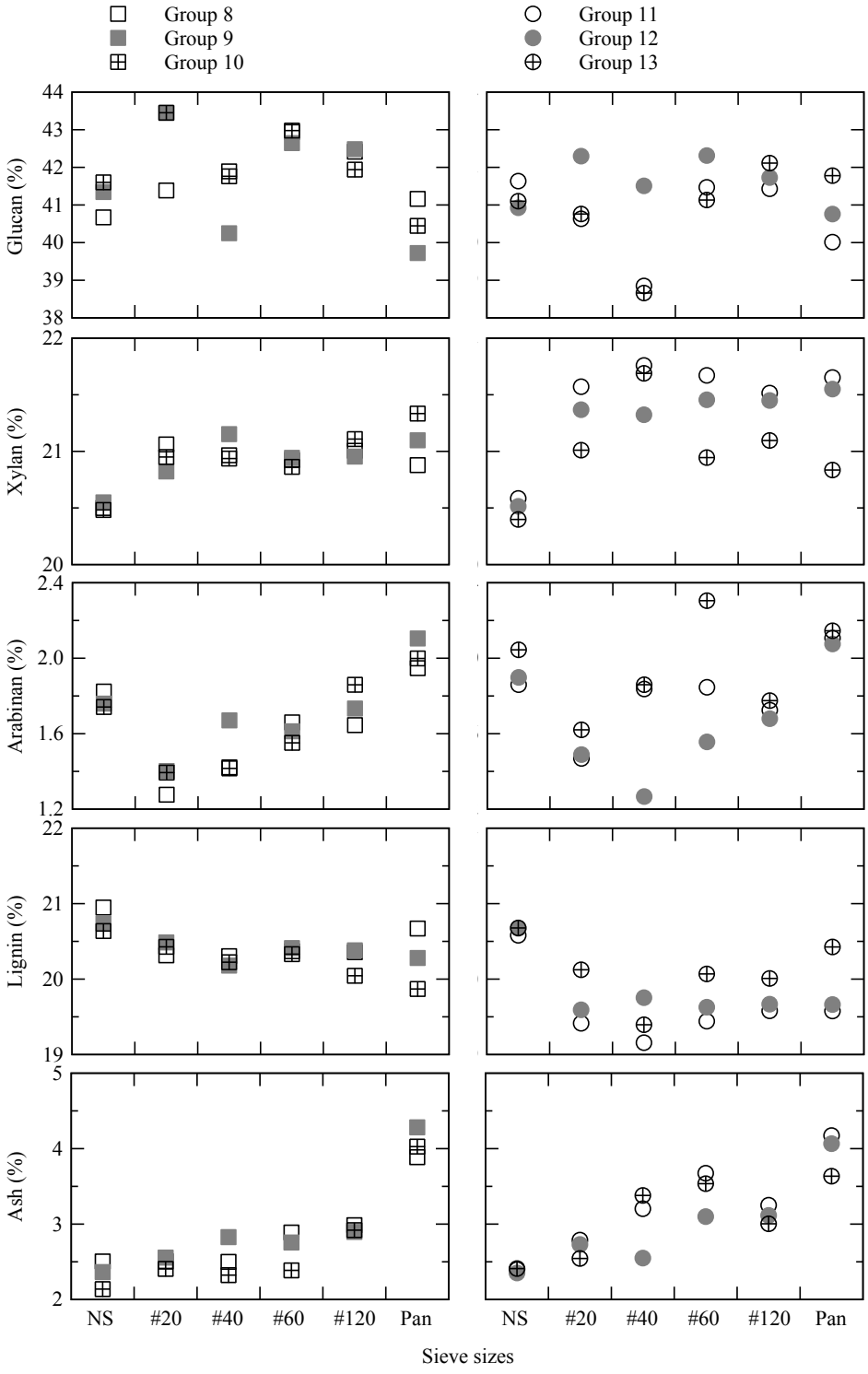


Figure 6.3. Predicted composition based on FT-NIR spectra of non-sieved samples (NS) and sieved samples.

Table 6.2. Predicted Miscanthus composition across particle size.

Sieve size (pore size)	Component Mean[a] ± S.D.[b] (%)				
	Glucan	Xylan	Arabinan	Lignin	Ash
Non-sieved (NS)	41.2a ± 0.4	20.5b ± 0.1	1.85ab ± 0.11	20.7a ± 0.1	2.36c ± 0.12
#20 (0.85 mm)	42.0a ± 1.3	21.1a ± 0.3	1.44b ± 0.11	20.1ab ± 0.5	2.59bc ± 0.14
#40 (0.425 mm)	40.5a ± 1.5	21.3a ± 0.4	1.58b ± 0.25	19.9b ± 0.5	2.80bc ± 0.42
#60 (0.25 mm)	42.3a ± 0.8	21.1a ± 0.3	1.76ab ± 0.29	20.0ab ± 0.4	3.06b ± 0.49
#120 (0.125 mm)	42.0a ± 0.4	21.2a ± 0.2	1.74ab ± 0.08	20.0ab ± 0.3	3.03b ± 0.13
Pan	40.7a ± 0.8	21.2a ± 0.3	2.06a ± 0.07	20.1ab ± 0.4	4.01a ± 0.23

[a]Per component, mean values followed by the same letter were not different ($p > 0.05$).
[b]S.D. = one standard deviation.

6.4. Conclusions

The effects of particle size on the predicted composition of ground Miscanthus was studied using core samples from 5 and 17 mo. bales. The geometric mean particle size ranged from 0.36 to 0.49 mm, with the smallest size observed with samples from bales stored outdoors for 17 mo. and the largest size observed with samples from bales stored outdoors with a tarp cover for 5 mo. All bale samples except those from the bale stored under roof for 5 mo. had a wide standard deviation in geometric mean diameter. On average, across predictions of glucan, arabinan, and lignin content were not sensitive to particle size, but predictions of xylan and ash content were. The predicted xylan content using the non-sieved samples was lower than those for sieved samples. In terms of ash content, predicted values increased with decreasing particle size.

CHAPTER 7. CLASSIFICATION OF MISCANTHUS BY PLS-DA

7.1. Introduction

Miscanthus is a potential dedicated energy crop with high cellulose and hemicellulose content (Heaton et al., 2004). However, like all agricultural materials, its composition varies greatly with age, stage of growth, storage condition, and genetics (Hames et al., 2003). These variations, especially in carbohydrate content, can affect process efficiencies during conversion, from pretreatment to fermentation (Öhgren et al., 2007; Berlin et al., 2007). Therefore, it would be advantageous for conversion facilities to receive lignocellulosic feedstocks that are consistent, or uniform in quality in moisture content, ash content, and convertible carbohydrates so they can operate their chemical pretreatment and conversion processes efficiently (Kenney and Ovard, 2013). Knowing the composition at earlier stages of the supply chain can also help in the development of quality-based valuations and incentivize farmers and suppliers to implement best management practices to ensure a uniform and consistent supply system (Kenney et al., 2013). To determine the composition of Miscanthus in near real time as it passes through the supply chain, high throughput assays based on spectroscopy, such as FT-NIR, has been explored (Haffner et al., 2013; Hayes, 2012). In some cases, a qualitative assessment, such as classification or "grade" of the Miscanthus material is preferable instead of a quantitative measure of its composition. Therefore, in addition to utilizing FT-NIR to predict compositional data, this study explored the use of FT-NIR PLSR models with linear discriminant analysis (LDA) to classify Miscanthus bale samples in a fixed number of groups or "grades".

7.2. Materials and methods

7.2.1. Classification of Miscanthus

The resulting PLSR regression models from Section 5.2 for glucan, lignin, and ash were further used to develop qualitative discriminant analysis models that may be used for screening Miscanthus samples based on these components. The method used in discriminant analysis was linear discriminant analysis (LDA), which explicitly attempts to model the difference between the classes of data.

To run the LDA the samples needed to be classified into their respective groups. The models were originally grouped into their respective classes by, first, looking for natural breaks in the histograms and, secondly, considering the range of the main portion of the data set for each component. Since there were few natural breaks, the groups were largely based on the range. Once the samples were placed in the groups the LDA was constructed with the groups and the scores from the PLSR models from Chapter 5 for glucan, lignin, and ash. The validation samples were then run along with the test set in the constructed LDA models.

7.2.2. Statistical analysis of groups

The resulting PLS-DA models were evaluated on the accuracy of the predicted classification and the uniformity of sample composition within each group. The means of each component across groups were compared using a Tukey's test to identify any difference between two means that is greater than the expected standard error. The Tukey's test was conducted using *R* (Version 2.15.2, 2012).

7.3. Results and discussion

7.3.1. Classification groups

Based on the histograms of the calibration sets, the maximum bin width was set at 2% and a minimum of three bins were used. For glucan, range was 8 percentage points, which led to 4 groupings of the samples (Figure 7.1) – Group G1 (less than 38%), Group G2 (38 to 40%), Group G3 (40 to 42%), and Group G4 (greater than 42%). Similarly, samples were split into three groups based on their lignin content (Group L1, less than 20%; Group L2, 20-22%; and Group L3, greater than 22%) and ash content (Group A1, less than 1.75%; Group A2, 1.75 to 3.25%; and Group A3, greater than 3.25%). Care was taken so the means of each group were different from each other (Table 7.1).

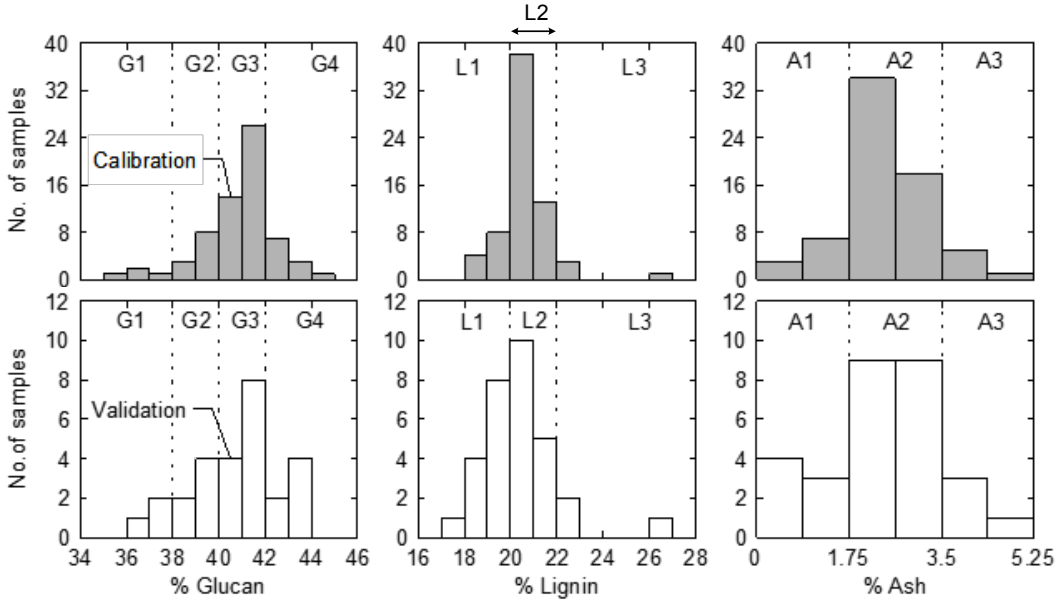


Figure 7.1 Classification of Miscanthus samples based on glucan, lignin, and ash contents determined by wet chemistry assays.

7.3.2. Accuracy of the PLS-DA classification models

During classification, three samples were identified as outliers – Gr1 (from a bale stored outdoors with tarp cover for 24 months), E44 (from a bale stored outdoors without

a tarp cover for 12 months), and E91 (from a bale stored under roof for 6 months). Using the PLS regression model for glucan for classification, 29 of 98 samples were misclassified into a neighboring group. The overall predictability of the model was 70%. A comparison of means showed that each predicted group, G1 to G4, was different from each other, whereas G3 and G4 were not. However each group was not different from the actual classification groups.

The classification and prediction results from the PLSR-DA lignin model showed that most of the samples were classified as Group L2. The model was able to classify 79 of 98 samples correctly. Samples with high lignin content tended to be classified correctly. The PLS-DA lignin model was able to classify 19 out of 25 samples correctly into Group L1; 54 out of 66 samples correctly into Group L2; and 6 out of 7 samples correctly in Group L3. When samples were misclassified, they fell into a neighboring group, i.e., no sample from Group L1 was misclassified into Group L3 and vice versa.

Similarly, the PLS-DA model for ash was able to classify 80 out of 98 samples correctly. Of the 18 samples that were misclassified they only fell one group away from their assigned group. This should be expected since the ash samples cover a continuous range but the samples are being classified into fixed groups. The PLS-DA ash model was able to classify 12 of 17 samples into Group A1; 56 of 65 samples into Group A2; and 11 of 15 into Group A3. A comparison of the means was run and all group means were different from each other.

Table 7.1 Classification of samples based on FT-NIR spectra and PLS-DA models for glucan, lignin, and ash contents.

		Actual classification by reference (wet chemistry) values[b]							
PLS-DA Glucan Model		Calibration				Validation			
Predicted classification[a]		G1	G2	G3	G4	G1	G2	G3	G4
	G1	*5*	1	0	0	*2*	2	0	0
	G2	0	*6*	*5*	0	2	*4*	3	0
	G3	0	4	*32*	3	0	1	*8*	2
	G4	0	1	7	*11*	0	1	3	*4*
PLS-DA Lignin Model		Calibration			Validation				
Predicted classification[a]		L1	L2	L3	L1	L2	L3		
	L1	*10*	0	0	*9*	3	0		
	L2	2	*45*	0	4	*8*	1		
	L3	0	6	*4*	0	3	*2*		
PLS-DA Ash Model		Calibration			Validation				
Predicted classification[a]		A1	A2	A3	A1	A2	A3		
	A1	*8*	0	0	*4*	1	0		
	A2	2	*44*	1	3	*12*	3		
	A3	0	6	*7*	0	2	*4*		

[a]Predicted classification based on the FT-NIR spectra.
[b]Number of samples correctly classified in their groups are presented in italics.

Table 7.2. Mean composition of all samples (calibration and validation) as classified by the PLS-DA models for glucan, lignin, and ash contents.

	Actual classification by reference (wet chemistry) values									
	Glucan Model				Lignin Model			Ash Model		
Means of actual classification groups (%)[a]	34.5 aA	39.2 bB	41.2 cC	43.0 dD	19.1 aA	20.8 bB	23.4 cC	0.97 aA	2.47 bB	3.90 cC
Means of predicted classification groups (%)[a]	34.9 aA	39.6 bB	41.2 cC	42.4 cD	19.4 aA	20.6 bB	22.2 cC	1.06 aA	2.40 bB	3.57 cC

[a]Values followed by the same lowercase letter in the same row, per PLS-DA model, are not different ($p > 0.05$). Means values of actual and predicted classification followed by the same uppercase letter in the same column are not different.

7.3.3. Uniformity of the samples after classification

Taking the predicted classes resulting from the PLS-DA modeling, the effect of this grouping on the other six components was evaluated. The box plots for each component showed the range and quartiles, with outliers designated by solid (laying 1.5

to 3 interquartile range distances from the mean) and hollow (laying more than 3 interquartile range distances from the mean) data points.

The PLS-DA glucan model would be most useful if it could also deliver a uniform feedstock within each grouping (Figure 7.2). In terms of xylan and arabinan content, the means for Groups G1 and G2 were not different from each other, similarly as the means for Groups G3 and G4. All four groups, however, did not differ in lignin and ash content. With the PLS-DA lignin model, while there were no difference in ash and extractives contents across Groups L1, L2, and L3, there were differences across the groups in terms of glucan and lignin contents and the variation within groups is small (Figure 7.3). The box plots for glucan, xylan, and lignin were tighter than the other groups for Group L2 which took 63 of the 98 samples representing a relatively consistent sample while Group L3 was wider, but was still relatively consistent. Even though the samples were classified into distinct groups when the PLS-DA for ash was applied, across groups, there was little variation in the other components (Figure 7.4). Group A2 had the least variation within its group. Based on these results, the best classification results were found with the PLS-DA lignin model, which classified the samples into three groups, with small variations with each group.

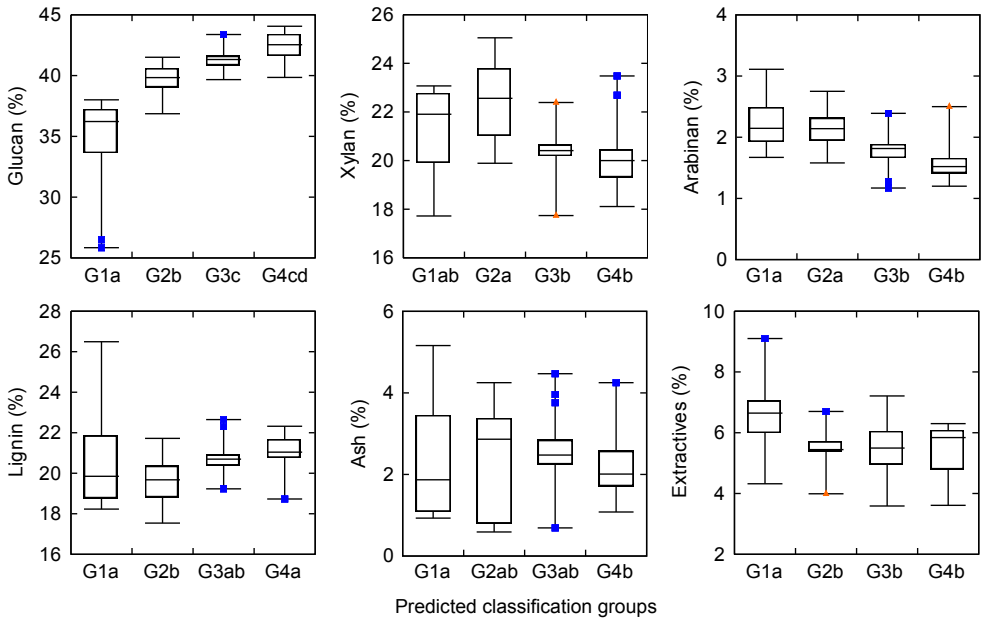


Figure 7.2. Variability in composition across groups as classified by the PLS-DA glucan model. Groups with the same letters, per component, are not different ($p > 0.05$). Median-based box plots represent the minimum, maximum, interquartile range (IQR), outliers (■, which are defined as data lying at 1.5·IQR distance from the median), and extremes (▲, which are defined as data lying at 3·IQR distance from the median).

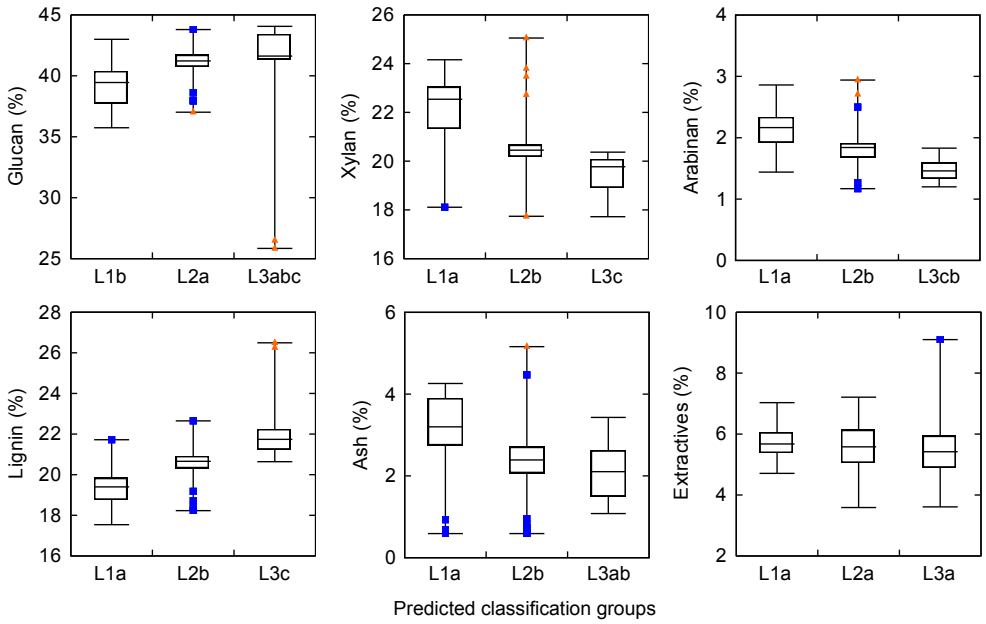


Figure 7.3. Variability in composition across groups as classified by the PLS-DA lignin model. Groups with the same letters, per component, are not different ($p > 0.05$). Median-based box plots represent the minimum, maximum, interquartile range (IQR), outliers (■, which are defined as data lying at 1.5·IQR distance from the median), and extremes (▲, which are defined as data lying at 3·IQR distance from the median).

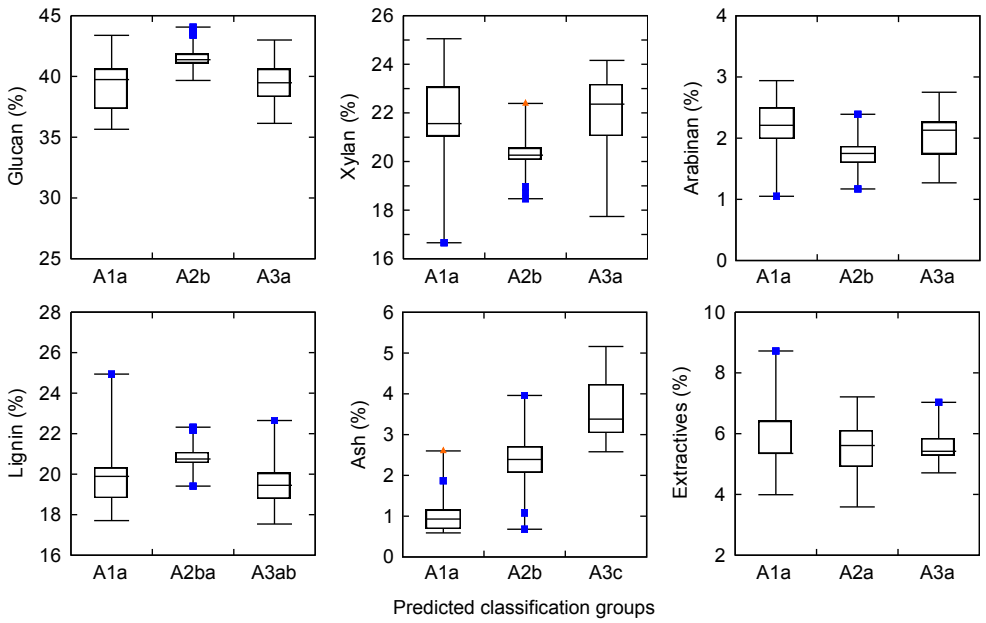


Figure 7.4. Variability in composition across groups as classified by the PLS-DA ash model. Groups with the same letters, per component, are not different ($p > 0.05$). Median-based box plots represent the minimum, maximum, interquartile range (IQR), outliers (■, which are defined as data lying at 1.5·IQR distance from the median), and extremes (▲, which are defined as data lying at 3·IQR distance from the median).

Tao et al. (2013) evaluated the minimum ethanol selling price (MESP) compared to corn stover composition and determined that the carybohydrates content (%) can have drastic effects on the conversion cost and final selling point of the ethanol. The corn stover used in their analysis ranged in total carbohydrates from 53 to 64%, resulting in an MESP range of $2.50 to $2.05 per gallon. The main carbohydrates in Miscanthus are cellulose (glucan) and the hemicellulose (arabino-xylan) and, when these components are summed up and plotted across groups, there were no differences across Groups G1 to G3 with G4 having the lowest polysaccharides content (Figure 7.5). Assuming that the conversion process is identical to that used by Tao et al. (2013) in their analysis, these data suggest Miscanthus as a promising alternative to corn stover since the majority of the samples had total carbohydrates contents greater than 60%. In addition to ash not

varying greatly between groups when the samples were classified using the PLS-DA models for glucan and lignin, overall, ash content after extraction were lower than those for corn stover samples (Templeton et al., 2009).

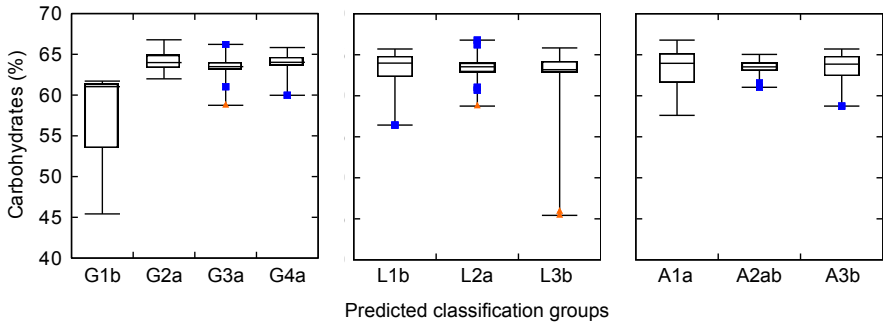


Figure 7.5 Carbohydrates (glucan + xylan + arabinan) content of the groups as classified by the PLS-DA models for glucan, lignin, and arabinan. Groups with the same letters, per PLS-DA model classification, are not different ($p > 0.05$). Median-based box plots represent the minimum, maximum, interquartile range (IQR), outliers (■, which are defined as data lying at 1.5·IQR distance from the median), and extremes (▲, which are defined as data lying at 3·IQR distance from the median).

7.4. Conclusions

These data indicate that the FT-NIR spectra and PLSR models from Chapter 5 can be used with LDA to classify the samples based on their glucan, lignin, and ash contents. The best classification results were based on the PLS-DA lignin model, which classified the samples into three groups, with small variations with each group. While the models developed in this study were based on a small sample size (less than 100 for calibration) and the small size contributed to some of the inaccuracy and imprecision in the predictions, the approach demonstrated that FT-NIR spectra and PLS-DA modeling can be used to rapidly screen Miscanthus samples at different stages of the supply chain, including after long-term storage.

CHAPTER 8. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

Overall, a large variation in Miscanthus compositions were observed in core samples from bales that have been stored under a variety of conditions (indoors, under roof, outdoors with tarp, and outdoors without a tarp) for a period of 3 to 24 mo. Glucan and lignin contents ranged from 25.8 to 44.1% and 17.5 to 26.5%, respectively.
In general, few trends were observed when looking that the sample sets in a univariate approach. However, Group 2, which included bales stored outdoors with a tarp cover for 24 mo., exhibited lowest glucan and xylan levels and highest lignin levels compared to the rest of the bales. Mold growth was also observed in Group 2 samples. Compared to other lignocellulosic feedstocks, overall, Miscanthus had higher carbohydrate and lower ash contents than corn stover while Miscanthus composition was similar to switchgrass.

When Miscanthus was manually harvested at the senescent stage, in terms of mass, the highest botanical fraction gathered was the rind, followed by the node, sheath, and pith. Across the different botanical fractions, ranges for each component were as follows: glucan, 32.2 to 46.1%; xylan, 20.9 to 25.3%; arabinan, 0.0 to 6.1%; lignin, 18.7 to 25.5%; and ash, 0.4 % to 8.9%. Overall, however, the sum of glucan, xylan, and arabinan contents for the rind, pith, and sheath fractions were not different from each other. While the compositional variation within each botanical fraction were not high, the variations across some botanical fractions were significant. The blade had lowest glucan, lowest lignin, and highest ash contents making them different from the other botanical fractions. The variations in Miscanthus composition were similar to that of corn stover composition and not as consistent as switchgrass composition. From the FT-NIR and PCA exploration, it has shown that PCA can pick up the variations among botanical

fractions, showing the possibility of constructing classification models such as linear discriminant analysis and soft independent modeling class analogy (SIMCA) based on different blends.

PLSR models were created for glucan, xylan, arabinan, lignin, and ash with *RPD* values of 4.59, 4.20, 2.46, 3.52, and 1.86, respectively, with *R/SEP* > 8. These values indicated the models (except for ash) were useful, at minimum, for screening samples in a plant breeding program. PLSR models for extractives and acetyl were also developed but, since the mean standard deviation of the reference data was high compared to the mean and range of the data set, the resulting models had RPD < 1.

The effects of particle size on the predicted composition of ground Miscanthus was studied using core samples from 5 mo. and 17 mo. old bales. The geometric mean particle size ranged from 0.36 to 0.49 mm, with the smallest size observed with samples from bales stored outdoors for 17 mo. and the largest size observed with samples from bales stored outdoors with a tarp cover for 5 mo. All bale samples except those from the bale stored under roof for 5 mo. had a wide standard deviation in geometric mean diameter. On average, across predictions of glucan, arabinan, and lignin content were not sensitive to particle size, but predictions of xylan and ash content were. The predicted xylan content using the non-sieved samples was lower than those for sieved samples. In terms of ash content, predicted values increased with decreasing particle size.

These data indicate that the FT-NIR spectra and PLSR models from Chapter 5 can be used with LDA to classify the samples based on their glucan, lignin, and ash contents. The best classification results were based on the PLS-DA lignin model, which classified the samples into three groups, with small variations with each group. While the models

developed in this study were based on a small sample size (less than 100 for calibration) and the small size contributed to some of the inaccuracy and imprecision in the predictions, the approach demonstrated that FT-NIR spectra and PLS-DA modeling can be used to rapidly screen Miscanthus samples at different stages of the supply chain, including after long-term storage.

The models and analyses presented here can be improved by (1) expanding the sample size to include variation in Miscanthus due to geographical location or collecting samples at different stages of the supply chain (e.g., after harvest, after transportation, handling of different forms (e.g., comminution, pelletizing, etc.); (2) blending of different fractions (rind, node, pith, sheath, and blade) of Miscanthus, verifying the composition of the blends with wet chemistry methods, and predicting the composition of the blends with FT-NIR spectroscopy; (3) expanding the classification models to include Miscanthus blends and comparing LDA models with SIMCA models.

REFERENCES

AACCI International. 1999. Guidelines for model development and maintenance. Method 39-00. The American Association of Cereal Chemists, 10th Edition. The Association: St. Paul, MN.

Aden A., M. Ruth, K. Ibsen, J. Jechura, K. Neeves, J. Sheehan, and B. Wallace. 2002. Lignocellulosic biomass to ethanol process design and economics utilizing co-current dilute acid prehydrolysis and enzymatic hydrolysis for corn stover. DOE/ Office of Biomass Programs Technical report NREL/TP-510–32438, National Renewable Energy Laboratory, http://www.nrel.gov.

Aden A. and T. Foust. 2002. Technoeconomic analysis of the dilute sulfuric acid and enzymatic hydrolysis process for the conversion of corn stover to ethanol. *Cellulose* 16(4): 535–545.

Aenugu, H., P. Reddy, D.S. Kumar, N.P. Srisudharson, S.S. Ghosh, and D. Banji. 2011. Near infrared spectroscopy – an overview. *International Journal of ChemTech Research* 3: 825-36.

Alakangas, E., J. Valtanen, and J. Levlin. 2006. CEN Technical specification for solid biofuels - fuel specification and classes. *Biomass and Bioenergy* 30: 908-14.

ASABE. 2003. Method of determining and expressing fineness of feed materials by sieving. Standard S319.3: ASABE: St. Joseph, MI.

ASABE. 2008. Moisture measurement - forages. Standard S358.2: ASABE: St. Joseph, MI.

Berlin, A., V. Maximenko, N. Gilkes, and J. Saddler. 2007. Optimization of enzyme complexes for lignocellulose hydrolysis. *Biotechnology and Bioengineering* 97: 287-96.

Berndes, G., C. Azar, T. Kaberger, and D. Abrahamson. 2001. The feasibility of large-scale lignocellulose-based bioenergy production. *Biomass and Bioenergy* 20: 371-83.

Berrueta, L.A., R.M. Alonso-Salces, and K. Heberger. 2007. Supervised pattern recognition in food analysis. *Journal of Chromatography A* 1158: 196-214.

Blanco, M., J. Coello, H. Iturriaga, S. Maspoch, and C. De La Pezuela. 1998. Near-infrared spectroscopy in the pharmaceutical industry. *Analyst* 123: 135R-50R.

Blanco, M., and I. Villarroya. 2002. NIR spectroscopy: a rapid-response analytical tool. *Trends in Analytical Chemistry* 21: 240-50.

Chen, H., Q. Song, G. Tang, Q. Feng, and L. Lin. 2013. The combined optimization of savitzky-golay smoothing and multiplicative scatter correction for FT-NIR PLS Models. *ISRN Spectroscopy* 2013: 1-9.

Chen, F. and R.A. Dixon. 2007. Lignin modification improves fermentable sugar yields for biofuel production. *Nature Biotechnology* 25: 759-61.

Chen, Y.C. and S.N. Thennadil. 2012. Insights into information contained in multiplicative scatter correction parameters and the potential for estimating particle size from these parameters. *Analytica Chimica Acta* 746: 37-46.

Claassen, P.A.M., J.B. van Lier, A.M. Lopez Contreras, E.W.J. van Niel, L. Sijtsma, A.J. M. Stams, S.S. de Vries, and R.A. Weusthuis. 1999. Utilization of biomass for the supply of energy carriers. *Applied Microbiology and Biotechnology* 52: 741-755.

Cozzolino, D., A. Fassio, E. Restaino, E. Fernandez, and A. La Manna. 2008. Verification of silage type using near-infrared spectroscopy combined with multivariate analysis. *Journal Agricultural Food Chemistry* 56: 79-83.

Crofcheck, C.L., and M.D. Montross. 2004. Effect of stover fraction on glucose production using enzymatic hydrolysis. *Transactions of the ASAE* 47: 841-844.

Duguid, K.B., M.D. Montross, C.W. Radtke, C.L. Crofcheck, S.A. Shearer, and R.L. Hoskinson. 2007. Screening for sugar and ethanol processing characteristics from anatomical fractions of wheat stover. *Biomass and Bioenergy* 31: 585-92.

Darr, M.J. and A. Shah. 2012. Biomass storage: an update on industrial solutions for baled biomass feedstocks. *Biofuels* 3(3): 321–332.

Emery, I.R., and N.S. Mosier. 2012. The impact of dry matter loss during herbaceous biomass storage on net greenhouse gas emissions from biofuels production. *Biomass and Bioenergy* 39: 237-246.

Esbensen, K. H., D. Guyot, F. Westad, and L. P. Houmøller. 2002. Multivariate Data Analysis-in Practice: An Introduction to Multivariate Data Analysis and Experimental Design. Woodbridge, NJ.

Evert, R. F. 2006. Esau's Plant Anatomy: Meristems, Cells, and Tissues of the Plant Body: Their Structure, Function, and Development. Hoboken, NJ.

Gauder, M., S. Graeff-Hönninger, I. Lewandowski, and W. Claupein. 2012. Long-term yield and performance of 15 different Miscanthus genotypes in southwest Germany. *Annals of Applied Biology* 160: 126-136.

Geladi, P., D. MacDougall, and H. Martens. 1985. Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Applied Spectroscopy* 39: 491-500.

Głowacka, K. 2011. A review of the genetic study of the energy crop miscanthus. *Biomass and Bioenergy* 35: 2445-2454.

Haffner, F. B., V. D. Mitchell, R. A. Arundale, and S. Bauer. 2013. Compositional analysis of Miscanthus giganteus by near infrared spectroscopy. *Cellulose* DOI 10.1007/s10570-013-9935-1.

Hames, B.R., S.R. Thomas, A.D. Sluiter, C.J. Roth, and D.W. Templeton. 2003. Rapid biomass analysis - new tools for compositional analysis of corn stover feedstocks and process intermediates from ethanol production. *Applied Biochemistry and Biotechnology* 105-108: 5-16.

Hames, B. 2009. Biomass compositional analysis for energy applications. *Methods Molecular Biology* 581: 145-67.

Hayes, D.J. 2012. Development of near infrared spectroscopy models for the quantitative prediction of the lignocellulosic components of wet Miscanthus samples. *Bioresource Technology* 119: 393-405.

Heaton, E. 2004. A quantitative review comparing the yields of two candidate C4 perennial biomass crops in relation to nitrogen, temperature and water. *Biomass and Bioenergy* 27: 21-30.

Heaton, E., J. Clifton-Brown, T. B. Voigt, M. B. Jones and S. P. Long. 2004. Miscanthus for renewable energy generation: European Union experience and projections for Illinois. *Mitigation and Adaptation Strategies for Global Change* 9: 433-451.

Heaton, E.A., F.G. Dohleman, and S.P. Long. 2008. Meeting U.S, biofuel goals with less land: the potential of Miscanthus. *Global Change Biology* 14: 2000-2014.

Helland, I.S., T. Næs, and T. Isaksson. 1995. Related versions of the multiplicative scatter correction method for preprocessing spectroscopic data. *Chemometrics and Intelligent Laboratory Systems* 29: 233-241.

Hess, J.R., K.L. Kenney, L.P. Ovard, E.M. Searcy, and C.T. Wright. 2009. Uniform-format solid feedstock supply system: A commodity-scale design to produce an infrastructure-compatible bulk solid from lignocellulosic biomass: Section 4. INL/EXT-08-14752 Revision 0 DRAFT. Idaho National Laboratory, Bioenergy Program, Idaho Falls, ID 83415 (April 2009).

Hodgson, E.M., R. Fahmi, N. Yates, T. Barraclough, I. Shield, G. Allison, A.V. Bridgwater, and I.S. Donnison. 2010. Miscanthus as a feedstock for fast-pyrolysis: does agronomic treatment affect quality? *Bioresource Technology* 101: 6185-91.

Hoskinson, R., D. Karlen, S. Birrell, C. Radtke, and W. Wilhelm. 2007. Engineering, nutrient removal, and feedstock conversion evaluations of four corn stover harvest scenarios. *Biomass and Bioenergy* 31: 126-136.

Humbird, D.A., R. Davis, L. Tao, C. Kinchin, D. Hsu, and A. Aden. (2011) Process design and economics for biochemical conversion of lignocellulosic biomass to ethanol. Dilute-acid pretreatment and enzymatic hydrolysis of corn stover. National Renewable Energy Laboratory, Golden, CO, USA.

Isaksson, T. and T. Næs 1988. The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. *Applied Spectroscopy* 42: 1273-1284.

James, A., R. Thring, S. Helle, and H. Ghuman. 2012. Ash management review—applications of biomass bottom ash. *Energies* 5: 3856-3873.

Jones, M.B. and M. Walsh. 2001. Miscanthus for energy and fiber. Reassessment of maximum growth rates for C3 and C4 Crops. *Experimental Agriculture* 14: 1-5.

Karlen, D.L., S.J. Birell, and J.R Hess. 2011. A five-year assessment of corn stover harvest in central Iowa, USA. *Soil and Tillage Research* 115-116: 47-55.

Kenney, K.L. and L.P. Ovard. 2013. Advanced feedstocks for advanced biofuels: transforming biomass to feedstocks. *Biofuels* 4: 1-3.

Kenney, K.L., W.A. Smith, G.L. Gresham, and T.L. Westover. 2013. Understanding biomass feedstock variability. *Biofuels* 4: 111-27.

Lattin, J., J. D. Carroll, P.E. Green. 2003. Analyzing Multivariate Data. Pacific Grove: Thomas Brooks/Cole, 2003.

Lewandowski, I. and A. Kicherer. 1997. Combustion quality of biomass: practical relevance and experiments to modify the biomass quality of Miscanthus x giganteus. *European Journal of Agronomy* 6: 163-77.

Limayem, A. and S.C. Ricke. 2012. Lignocellulosic biomass for bioethanol production: current perspectives, potential issues and future prospects. *Progress in Energy and Combustion Science* 38: 449-67.

Liu, L., P.X. Ye, A.R. Womac, and S. Sokhansanj. 2010. Variability of biomass chemical composition and rapid analysis using FT-NIR techniques. *Carbohydrate Polymers* 81: 820-29.

Manley, M., L. Van Zyl, and B.G. Osborne. 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.

Martens, H. and T. Næs. Multivariate Calibration. New York: John Wiley and Sons, 1989.

Monteith, J.L. 1978. Reassessment of maximum growth rates for C3 and C4 crops. *Experimental Agriculture* 14: 1-5.

Næs, T. and H. Martens. 1984. Multivariate calibration 2. Chemometric methods. *Trends in Analytical Chemistry* 3: 266-271.

Næs, T. C. Irgens, and H. Martens. 1986. Comparison of linear statistical methods for calibration of NIR instruments. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 35: 195-206.

Næs, T., T. Isaksson, and B. Kowalski. 1990. Locally weighted regression and scatter correction for near-infrared reflectance data. *Analytical Chemistry* 62: 664-73.

Naidu, S.L. and S.P. Long. 2004. Potential mechanisms of low-temperature tolerance of C4 photosynthesis in Miscanthus x giganteus: an in vivo analysis. *Planta* 220: 145-55.

Öhgren, K., R. Bura, J. Saddler, and G. Zacchi. 2007. Effect of hemicellulose and lignin removal on enzymatic hydrolysis of steam pretreated corn stover. *Bioresource Technology* 98: 2503-10.

Perez, J., J. Munoz-Dorado, T. de la Rubia, and J. Martinez. 2002. Biodegradation and biological treatments of cellulose, hemicellulose and lignin: An overview. *International Microbiology* 5: 53-63.

Pordesimo, L.O., B.R. Hames, S. Sokhansanj, and W.C. Edens. 2005. Variation in corn stover composition and energy content with crop maturity. *Biomass and Bioenergy* 28: 366-374.

R Development Core Team. 2012. *R*: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Rantanen, J., E. Räsänen, J. Tenhunen, M. Känsäkoski, J. Mannermaa, and J. Yliruusi. 2000. In-line moisture measurement during granulation with a four-wavelength near infrared sensor: an evaluation of particle size and binder effects. *European Journal of Pharmaceutics and Biopharmaceutics* 50: 271-276.

Rudolf, A., H. Baudel, G. Zacchi, B. Hahn-Hagerdal, and G. Liden. 2008. Simultaneous saccharification and fermentation of steam-pretreated bagasse using *Saccharomyces cerevisiae* Tmb3400 and *Pichia stipitis* Cbs6054. *Biotechnology and Bioenergy* 99: 783-790.

Renewable Fuels Association (RFA), "2012 Ethanol Industry Outlook," February 22, 2012, http://ethanolrfa.3cdn.net/d4ad995ffb7ae8fbfe_1vm62ypzd.pdf.

Sanderson, M.A., F. Agblevor, M. Collins, and D.K. Johnson. 1996. Compositional analysis of biomass feedstocks by near infrared reflectance spectroscopy. *Biomass and Bioenergy* 11: 365-370.

Savitzky, A. and M.J.E. Golay. 1964. Smoothing + differentiation of data by simplified least squares procedures. *Analytical Chemistry* 36: 1627-39.

Shah, A. and M. Darr. 2011. Outdoor storage characteristics of single-pass large square corn stover bales in Iowa. *Energies* 4: 1687-95.

Schwanninger, M., J. Rodrigues, and K. Fackler. 2011. A review of band assignments in near infrared spectra of wood and wood components. *Journal of Near Infrared Spectroscopy.* 19: 287-308.

Shinners, K., B. Binversie, R. Muck, and P. Weimer. 2007. Comparison of wet and dry corn stover harvest and storage. *Biomass and Bioenergy* 31: 211-221.

Siesler, H.W., Y. Ozaki, S. Kawata, and H.M. Heise. Near-Infrared Spectroscopy: Principals, Instruments, Applications. Weinheim (Germany): Wiley-Vch, 2002.

Sun, Run-Cang. Cereal Straw as a Resource for Sustainable Biomaterials and Biofuels Chemistry: Extractives, Structure, Ultrastructure, and Chemical Composition. Amsterdam: Elsevier, 2010.

Tao, L., A. Aden, R.T. Elander, V.R. Pallapolu, Y.Y. Lee, R.J. Garlock, V. Balan, B.E. Dale, Y. Kim, N.S. Mosier, M.R. Ladisch, M. Falls, M.T. Holtzapple, R. Sierra, J. Shi, M.A. Ebrik, T. Redmond, B. Yang, C.E. Wyman, B. Hames, S. Thomas, and R.E. Warner. 2011. Process and technoeconomic analysis of leading pretreatment technologies for lignocellulosic ethanol production using switchgrass. *Bioresource Technology* 102: 11105-11114.

Tao, L., D.W. Templeton, D. Humbird, and A. Aden. 2013. Effect of corn stover compositional variability on minimum ethanol selling price (MESP). *Bioresource Technology* 140: 426-430.

Templeton, D.W., A.D. Sluiter, T.K. Hayward, B.R. Hames, and S.R. Thomas. 2009. Assessing corn stover composition and sources of variability Via Nirs. *Cellulose* 16: 621-39.

Theander, O., and P. Aman. Anatomical and Chemical Characteristics. In Straw and Other Fiberous By-Products as Feed. Amsterdam: Elsevier, 1984.

Tkachenko, N.V. Optical Spectroscopy: Methods and Instrumentations. Oxford, UK: Elsevier, 2006.

Unscrambler. 2011. Camo Software, Inc., Woodbridge, NJ.

U.S. Department of Energy. Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The technical Feasibility of a Billion-Ton Annual Supply. R.D.Perlack, L.L. Wright, A.F. Turhollow, R.L. Graham, B.J. Stokes, and D.C. Erbach (Leads), ORNL/TM-2005/66. Oak Ridge National Laboratory, Oak Ridge, TN. 227p.

U.S. Department of Energy. U.S. Billion-Ton Update: Biomass Supply for a Bioenergy and Bioproducts Industry. R.D. Perlack and B.J. Stokes (Leads), ORNL/TM-2011/224. Oak Ridge National Laboratory, Oak Ridge, TN. 227p.

Vidal, B.C., B.S. Dien, K.C. Ting, and V. Singh. 2011. Influence of feedstock particle size on lignocellulose conversion--a review. *Applied Biochemistry and Biotechnology* 164: 1405-1421.

Wiselogel, A.E., F.A. Agblevor, D.K. Johnson, S. Deutch, J.A. Fennell, and M.A. Sanderson. 1996. Compositional changes during storage of large round switchgrass bales. *Bioresource Technology* 56: 103-109.

Wolfrum, E.J. and A.D. Sluiter. 2009. Improved multivariate calibration models for corn stover feedstock and dilute-acid pretreated corn stover. *Cellulose* 16: 567-76.

Yang, B. and C.E. Wyman. 2008. Pretreatment: The key to unlocking low-cost cellulosic Ethanol. *Biofuels, Bioproducts, & Biorefining* 2: 26-40.

Ye, P., L. Liu, D. Hayes, A. Womac, K. Hong, and S. Sokhansanj. 2008. Fast classification and compositional analysis of cornstover fractions using Fourier transform near-infrared techniques. *Bioresource Technology* 99: 7323-32.

APPENDIX A. CHEMICAL COMPOSITION OF MISCANTHUS

Table A.1. Composition of *Miscanthus* × *giganteus* from stored bales as determined by EBI[a].

Sample Group[b]	Sample Name	Mean ± S.D.[c] (%)						
		Extractives	Glucan	Xylan	Arabinnan	Acetyl	Lignin	Ash
1	Ty32	5.00 ± 0.10	40.60 ± 0.14	21.27 ± 0.15	2.05 ± 0.01	3.13 ± 0.02	19.47 ± 0.18	4.25 ± 0.17
	Ty33	6.14 ± 0.42	38.01 ± 0.13	20.53 ± 0.09	2.16 ± 0.01	2.63 ± 0.01	18.40 ± 4.21	1.87 ± 0.04
	Ty34	4.42 ± 0.42	43.80 ± 0.16	19.80 ± 0.07	1.44 ± 0.02	3.25 ± 0.26	21.61 ± 0.15	1.94 ± 0.01
2	Gr1	4.32 ± 0.21	25.84 ± 0.03	17.75 ± 0.07	1.83 ± 0.01	1.99 ± 0.03	26.49 ± 0.16	14.01 ± 0.05
	Gr10	3.99 ± 0.06	39.74 ± 0.10	25.05 ± 0.04	1.99 ± 0.00	2.74 ± 0.01	21.17 ± 0.66	0.59 ± 0.01
	Gr14	6.06 ± 0.06	40.42 ± 0.33	18.11 ± 0.16	1.44 ± 0.01	2.33 ± 0.00	20.75 ± 0.14	4.25 ± 0.02
	Gr16	7.03 ± 0.25	36.14 ± 0.12	22.93 ± 0.05	1.96 ± 0.00	2.85 ± 0.03	20.40 ± 0.02	4.26 ± 0.05
	Gr2	8.72 ± 0.25	35.65 ± 0.13	19.88 ± 0.10	2.08 ± 0.00	2.42 ± 0.08	17.71 ± 0.06	1.24 ± 0.02
	Gr21	5.39 ± 0.07	36.86 ± 0.08	24.15 ± 0.07	2.13 ± 0.05	2.66 ± 0.00	19.65 ± 0.04	4.22 ± 0.01
	Gr27	9.10 ± 0.16	26.50 ± 0.04	17.72 ± 0.07	1.67 ± 0.01	1.88 ± 0.00	26.26 ± 0.01	2.65 ± 0.18
	Gr3	6.31 ± 0.20	39.73 ± 0.12	17.74 ± 0.08	1.27 ± 0.00	2.30 ± 0.03	22.65 ± 0.03	4.47 ± 0.08
	Gr4	4.35 ± 0.04	41.77 ± 0.12	16.66 ± 0.05	1.05 ± 0.00	2.32 ± 0.00	24.94 ± 0.21	1.16 ± 0.02
3	E185	6.11 ± 0.01	42.67 ± 0.13	19.58 ± 0.09	1.56 ± 0.00	2.72 ± 0.05	20.94 ± 0.53	2.09 ± 0.04
	E186	6.30 ± 0.03	42.54 ± 0.60	19.20 ± 0.21	1.45 ± 0.03	2.73 ± 0.04	21.04 ± 0.02	1.62 ± 0.01
	E187	5.85 ± 0.12	43.63 ± 0.26	19.28 ± 0.17	1.29 ± 0.01	2.92 ± 0.22	21.58 ± 0.16	1.21 ± 0.01
	E188	6.00 ± 0.51	43.45 ± 0.08	18.72 ± 0.05	1.25 ± 0.00	2.67 ± 0.01	22.32 ± 0.04	1.46 ± 0.01
	E189	5.89 ± 0.00	44.06 ± 0.33	18.96 ± 0.05	1.20 ± 0.02	2.72 ± 0.00	21.97 ± 0.09	1.07 ± 0.03
	E190	6.24 ± 0.21	43.39 ± 0.23	18.47 ± 0.04	1.21 ± 0.02	2.80 ± 0.12	22.17 ± 0.18	1.18 ± 0.00
4	E37	5.93 ± 0.28	39.19 ± 0.18	22.76 ± 0.04	2.35 ± 0.01	1.92 ± 0.02	18.50 ± 0.11	3.38 ± 0.06
	E38	5.03 ± 0.22	41.40 ± 0.44	21.26 ± 0.21	2.18 ± 0.00	2.29 ± 0.16	19.23 ± 0.06	3.76 ± 0.15
	E39	5.35 ± 0.05	39.93 ± 0.05	22.99 ± 0.09	2.24 ± 0.01	1.98 ± 0.01	19.89 ± 0.01	0.59 ± 0.04
	E40	5.31 ± 0.24	39.67 ± 0.02	22.39 ± 0.08	2.39 ± 0.01	2.33 ± 0.00	19.41 ± 0.11	2.94 ± 0.03
	E52	6.05 ± 0.20	38.29 ± 0.15	23.79 ± 0.21	2.75 ± 0.05	2.66 ± 0.16	17.54 ± 0.14	3.20 ± 0.03
	E53	5.20 ± 0.26	43.38 ± 0.03	20.96 ± 0.02	1.88 ± 0.03	1.98 ± 0.02	19.96 ± 0.09	0.69 ± 0.04
	E55	5.34 ± 0.04	40.62 ± 0.40	21.44 ± 0.23	1.96 ± 0.03	2.30 ± 0.02	20.33 ± 0.43	0.83 ± 0.01
	E56	5.12 ± 0.04	38.98 ± 0.26	24.16 ± 0.05	2.22 ± 0.04	2.55 ± 0.11	18.65 ± 0.01	3.12 ± 0.18
	E59	4.76 ± 0.09	41.22 ± 0.05	22.69 ± 0.08	1.79 ± 0.02	2.56 ± 0.15	19.45 ± 0.38	2.58 ± 0.05
5	E43	5.65 ± 0.13	39.42 ± 0.29	23.00 ± 0.07	2.36 ± 0.03	2.17 ± 0.02	18.18 ± 0.18	2.75 ± 0.01
	E44	6.16 ± 0.01	35.74 ± 0.02	18.93 ± 0.03	1.76 ± 0.02	1.65 ± 0.05	18.42 ± 0.10	12.75 ± 0.03
	E45	5.44 ± 0.17	39.85 ± 0.59	23.48 ± 0.41	2.50 ± 0.09	2.21 ± 0.33	18.73 ± 0.11	2.60 ± 0.05
	E46	5.28 ± 0.02	38.64 ± 0.03	23.90 ± 0.04	2.71 ± 0.00	2.26 ± 0.02	19.18 ± 0.47	0.59 ± 0.01
	E72	5.70 ± 0.00	38.07 ± 0.44	22.36 ± 0.03	1.91 ± 0.01	2.58 ± 0.29	19.70 ± 0.15	3.64 ± 0.06
	E73	4.71 ± 0.14	43.00 ± 0.02	20.72 ± 0.04	1.69 ± 0.02	3.17 ± 1.65	19.39 ± 0.03	3.17 ± 0.00
	E74	5.37 ± 0.03	39.48 ± 0.01	23.37 ± 0.11	2.33 ± 0.03	2.94 ± 0.07	18.75 ± 0.04	2.87 ± 0.03
	E75	5.77 ± 0.16	39.95 ± 0.02	21.75 ± 0.10	2.15 ± 0.03	3.15 ± 1.82	18.30 ± 0.05	3.90 ± 0.06
	E76	6.68 ± 0.00	36.30 ± 0.16	23.07 ± 0.06	2.34 ± 0.03	2.25 ± 0.01	19.98 ± 0.22	0.93 ± 0.04
	E79	5.40 ± 0.10	37.43 ± 0.08	24.06 ± 0.00	2.31 ± 0.02	2.54 ± 0.04	19.16 ± 0.15	3.38 ± 0.01

Table A.1. Continued

6	E89	5.33 ± 0.07	40.66 ± 0.15	21.08 ± 0.09	2.21 ± 0.02	2.59 ± 0.01	20.47 ± 0.13	0.74 ± 0.03
	E90	6.61 ± 0.32	37.02 ± 0.25	22.73 ± 0.10	2.94 ± 0.00	2.45 ± 0.02	18.23 ± 0.18	0.96 ± 0.00
	E91	6.42 ± 0.08	37.20 ± 0.02	21.56 ± 0.05	2.86 ± 0.01	2.21 ± 0.02	19.30 ± 0.27	1.06 ± 0.01
	E92	7.14 ± 0.34	35.80 ± 0.20	22.58 ± 0.11	3.11 ± 0.01	2.14 ± 0.00	18.77 ± 0.78	1.22 ± 0.02
	E93	5.70 ± 0.30	41.51 ± 0.07	20.56 ± 0.03	1.97 ± 0.00	2.64 ± 0.01	21.72 ± 0.05	0.68 ± 0.03
7	E180	5.32 ± 0.07	37.93 ± 0.41	21.23 ± 0.25	2.13 ± 0.04	2.56 ± 0.00	19.72 ± 0.19	5.16 ± 0.02
	E181	4.64 ± 0.09	43.65 ± 0.18	19.81 ± 0.18	1.46 ± 0.00	2.87 ± 0.10	21.74 ± 0.12	1.82 ± 0.04
	E182	5.30 ± 0.05	41.59 ± 0.07	20.00 ± 0.09	1.57 ± 0.00	2.84 ± 0.22	20.84 ± 0.14	2.20 ± 0.03
	E183	5.01 ± 0.31	41.62 ± 0.27	20.37 ± 0.23	1.56 ± 0.02	2.73 ± 0.11	21.15 ± 0.03	2.57 ± 0.05
	E184	4.67 ± 0.04	41.36 ± 0.12	20.00 ± 0.13	1.63 ± 0.03	2.62 ± 0.01	21.45 ± 0.23	3.43 ± 0.02
8	Z108	5.60 ± 0.28	41.56 ± 0.05	20.66 ± 0.18	1.71 ± 0.03	3.01 ± 0.19	20.80 ± 0.02	2.71 ± 0.03
	Z120	6.91 ± 0.01	40.16 ± 0.75	20.53 ± 0.22	1.67 ± 0.02	2.86 ± 0.12	20.33 ± 0.28	3.19 ± 0.02
	Z132	6.00 ± 0.40	40.67 ± 0.22	20.93 ± 0.24	1.75 ± 0.03	2.70 ± 0.00	20.47 ± 0.09	2.86 ± 0.15
	Z135	6.70 ± 0.49	39.71 ± 0.00	20.95 ± 0.01	1.88 ± 0.01	2.74 ± 0.08	20.01 ± 0.16	3.14 ± 0.07
	Z144	5.42 ± 0.17	40.63 ± 0.24	20.25 ± 0.10	1.61 ± 0.01	2.73 ± 0.08	20.64 ± 0.08	2.96 ± 0.05
	Z153	3.61 ± 0.27	42.67 ± 0.40	20.36 ± 0.18	1.39 ± 0.01	2.93 ± 0.01	22.19 ± 0.07	2.01 ± 0.02
	Z198	5.84 ± 0.43	42.21 ± 0.12	20.53 ± 0.02	1.68 ± 0.00	2.82 ± 0.05	20.62 ± 0.01	2.02 ± 0.03
	Z213	4.71 ± 0.53	40.75 ± 0.08	21.00 ± 0.10	1.87 ± 0.01	2.94 ± 0.04	21.21 ± 0.12	2.36 ± 0.03
	Z231	6.41 ± 0.28	39.84 ± 0.01	19.57 ± 0.03	1.62 ± 0.05	2.81 ± 0.02	21.46 ± 0.07	2.89 ± 0.07
9	Y114	4.75 ± 0.32	41.83 ± 0.12	20.60 ± 0.04	1.63 ± 0.01	3.03 ± 0.12	21.44 ± 0.30	2.44 ± 0.02
	Y123	6.14 ± 0.25	42.15 ± 0.31	20.09 ± 0.32	1.52 ± 0.08	2.95 ± 0.23	20.59 ± 0.34	1.81 ± 0.03
	Y126	5.75 ± 0.36	41.37 ± 0.10	20.73 ± 0.05	1.72 ± 0.03	2.82 ± 0.06	20.51 ± 0.07	2.51 ± 0.00
	Y141	5.84 ± 0.26	40.41 ± 0.31	20.60 ± 0.08	1.74 ± 0.01	2.64 ± 0.02	20.36 ± 0.26	3.45 ± 0.15
	Y150	5.21 ± 0.19	41.44 ± 0.33	20.67 ± 0.15	1.63 ± 0.00	3.09 ± 0.07	20.98 ± 0.10	1.95 ± 0.01
	Y159	4.64 ± 0.03	42.52 ± 0.34	20.72 ± 0.15	1.60 ± 0.00	2.86 ± 0.15	20.71 ± 0.15	1.89 ± 0.00
	Y200	5.23 ± 0.19	41.72 ± 0.11	20.24 ± 0.05	1.58 ± 0.02	2.93 ± 0.27	21.06 ± 0.10	2.04 ± 0.01
	Y216	4.60 ± 0.27	41.22 ± 0.23	20.58 ± 0.08	1.77 ± 0.02	2.79 ± 0.04	21.25 ± 0.21	2.24 ± 0.03
	Y234	5.49 ± 0.18	40.32 ± 0.35	20.06 ± 0.24	1.77 ± 0.02	2.64 ± 0.02	20.22 ± 0.31	2.86 ± 0.05
10	X111	6.06 ± 0.44	40.33 ± 0.22	20.93 ± 0.03	1.89 ± 0.02	2.75 ± 0.06	20.19 ± 0.32	2.93 ± 0.03
	X117	6.16 ± 0.06	41.82 ± 0.30	20.56 ± 0.06	1.64 ± 0.00	3.09 ± 0.00	20.74 ± 0.15	1.86 ± 0.00
	X129	5.95 ± 0.06	41.86 ± 0.03	20.65 ± 0.03	1.58 ± 0.01	3.02 ± 0.01	20.68 ± 0.33	2.21 ± 0.02
	X138	6.20 ± 0.21	40.27 ± 0.03	20.23 ± 0.03	1.80 ± 0.02	2.60 ± 0.01	19.92 ± 0.06	3.96 ± 0.10
	X147	5.63 ± 0.14	41.36 ± 0.16	20.17 ± 0.10	1.58 ± 0.02	2.73 ± 0.13	21.14 ± 0.08	2.64 ± 0.12
	X156	6.56 ± 0.02	39.94 ± 0.32	19.86 ± 0.16	1.74 ± 0.01	2.66 ± 0.15	20.58 ± 0.31	3.28 ± 0.02
	X196	5.04 ± 0.08	42.51 ± 0.03	20.11 ± 0.04	1.61 ± 0.03	2.57 ± 0.04	21.09 ± 0.05	2.25 ± 0.06
	X219	5.45 ± 0.24	41.38 ± 0.07	19.94 ± 0.09	1.73 ± 0.02	2.75 ± 0.00	20.88 ± 0.08	2.24 ± 0.01
	X237	5.75 ± 0.24	40.35 ± 0.23	19.89 ± 0.05	1.75 ± 0.01	2.71 ± 0.03	20.85 ± 0.09	2.91 ± 0.04

Table A.1. Continued

11	O206	6.05 ± 0.12	40.87 ± 0.13	20.47 ± 0.14	1.84 ± 0.02	2.51 ± 0.06	20.22 ± 0.11	2.73 ± 0.02
	O207	6.14 ± 0.21	41.20 ± 0.47	20.12 ± 0.15	1.75 ± 0.01	2.69 ± 0.01	20.75 ± 0.06	2.25 ± 0.14
	O208	5.93 ± 0.21	41.66 ± 0.23	20.17 ± 0.13	1.81 ± 0.02	2.74 ± 0.06	20.36 ± 0.07	2.71 ± 0.12
	O221	5.03 ± 0.10	41.84 ± 0.16	20.45 ± 0.04	1.91 ± 0.02	2.73 ± 0.02	20.75 ± 0.12	2.27 ± 0.00
	O222	4.80 ± 0.27	42.03 ± 0.20	20.38 ± 0.10	1.86 ± 0.01	2.80 ± 0.00	20.80 ± 0.25	2.18 ± 0.01
	O223	5.61 ± 0.35	41.83 ± 0.25	20.35 ± 0.07	1.85 ± 0.01	2.88 ± 0.07	20.92 ± 0.05	2.24 ± 0.01
	O239	4.50 ± 0.10	41.91 ± 0.04	20.66 ± 0.15	1.94 ± 0.08	2.73 ± 0.00	20.77 ± 0.18	2.38 ± 0.04
	O240	4.85 ± 0.01	41.44 ± 0.06	20.76 ± 0.06	1.93 ± 0.02	2.87 ± 0.09	20.71 ± 0.02	2.46 ± 0.02
	O241	4.71 ± 0.09	41.17 ± 0.35	20.44 ± 0.17	1.91 ± 0.03	2.85 ± 0.01	20.60 ± 0.14	2.47 ± 0.00
12	T209	6.53 ± 0.08	40.60 ± 0.30	20.26 ± 0.08	1.88 ± 0.02	2.59 ± 0.00	20.02 ± 0.05	2.61 ± 0.02
	T210	6.70 ± 0.00	40.73 ± 0.03	20.22 ± 0.04	1.84 ± 0.00	2.06 ± 0.01	20.23 ± 0.03	2.60 ± 0.03
	T211	6.27 ± 0.26	40.94 ± 0.13	20.53 ± 0.16	1.88 ± 0.03	2.48 ± 0.12	20.28 ± 0.06	2.87 ± 0.07
	T224	5.25 ± 0.08	41.63 ± 0.04	20.15 ± 0.06	1.86 ± 0.00	2.85 ± 0.10	20.88 ± 0.09	2.30 ± 0.00
	T225	5.54 ± 0.17	41.10 ± 0.26	20.41 ± 0.03	1.94 ± 0.01	2.82 ± 0.06	20.65 ± 0.11	2.39 ± 0.08
	T226	5.58 ± 0.08	41.55 ± 0.03	20.02 ± 0.07	1.82 ± 0.02	2.83 ± 0.04	20.57 ± 0.39	2.06 ± 0.02
	T242	4.46 ± 0.16	41.33 ± 0.10	20.41 ± 0.09	1.93 ± 0.00	2.80 ± 0.21	20.92 ± 0.27	2.49 ± 0.03
	T243	4.32 ± 0.11	41.61 ± 0.22	20.54 ± 0.04	1.84 ± 0.01	2.95 ± 0.16	20.98 ± 0.03	2.48 ± 0.01
	T244	4.60 ± 0.01	41.35 ± 0.22	20.26 ± 0.08	1.91 ± 0.02	2.78 ± 0.06	20.96 ± 0.03	2.54 ± 0.01
13	I203	5.22 ± 0.11	41.04 ± 0.33	19.77 ± 0.18	1.38 ± 0.01	2.75 ± 0.00	22.30 ± 0.06	2.29 ± 0.02
	I204	7.21 ± 0.15	40.69 ± 0.01	20.10 ± 0.01	1.79 ± 0.00	2.93 ± 0.00	20.16 ± 0.01	2.62 ± 0.00
	I205	6.33 ± 0.25	41.13 ± 0.29	20.27 ± 0.12	1.81 ± 0.02	2.76 ± 0.00	19.88 ± 0.52	2.80 ± 0.08
	I227	5.83 ± 0.09	40.96 ± 0.14	20.03 ± 0.02	1.89 ± 0.01	2.71 ± 0.00	20.21 ± 0.01	2.44 ± 0.06
	I228	6.20 ± 0.35	41.32 ± 0.06	20.01 ± 0.03	1.87 ± 0.00	2.68 ± 0.03	20.73 ± 0.20	2.25 ± 0.01
	I229	5.88 ± 0.26	41.11 ± 0.31	19.87 ± 0.34	1.91 ± 0.02	2.69 ± 0.00	20.66 ± 0.21	2.66 ± 0.08
	I245	5.00 ± 0.34	41.35 ± 0.05	20.20 ± 0.05	1.85 ± 0.04	2.75 ± 0.01	20.73 ± 0.05	2.39 ± 0.03
	I246	3.59 ± 0.12	42.72 ± 0.18	19.76 ± 0.36	1.66 ± 0.02	2.77 ± 0.01	21.48 ± 0.08	2.07 ± 0.05
	I247	4.25 ± 0.03	41.37 ± 0.73	20.52 ± 0.30	1.90 ± 0.03	2.82 ± 0.11	20.10 ± 0.28	2.72 ± 0.00

[a]All measurements, except for extractives content, are reported as a percent of extracted biomass.
[b]Samples stored at different sites, conditions, and time. Group 1: Taylorville, indoors, 24 mos.; Group2: Griggsville, outdoors with tarp, 24 mos.; Group 3: Urbana, under roof, 24 mos.; Group 4: Urbana, under roof, 12 mos.; Group 5: Urbana, outdoors without a tarp, 12 mos.; Group 6: Urbana, under roof, 6 mos; Group 7: Urbana, outdoors without a tarp, 6 mos.; Group 8: Urbana, outdoors without a tarp; Group 9: Urbana, outdoors with tarp, 6 mos.; Group 10: Urbana, under roof, 6 mos.; Group 11: Urbana, outdoors without a tarp, 6 mos.; Group 12: Urbana, outdoors, with tarp, 3 mos.; and Group 13: Urbana, under roof, 3 mos.
[c]S.D. = one standard deviation

Table A.2. Range, mean, standard deviation (S.D.) and kurtosis of the composition of Miscanthus bale samples used in PLSR modeling.

Component	Calibration set (n = 67)			Validation set (n = 24)			Test set (n = 10)		
	Range (%)	Mean ± S.D. (%)	Kurtosis	Range (%)	Mean ± S.D. (%)	Kurtosis	Range (%)	Mean ± S.D. (%)	Kurtosis
Glucan	26.5-44.0	40.7 ± 2.37	19.2	25.8-43.6	39.8 ± 3.70	10.6	36.9-43.7	40.5 ± 2.49	-1.36
Xylan	17.7-24.2	20.6 ± 1.20	1.28	19.3-23.8	20.7 ± 1.02	3.88	18.1-24.2	21.6 ± 2.34	-1.60
Arabinan	1.05-3.11	1.83 ± 0.36	3.09	1.20-2.71	1.85 ± 0.34	1.46	1.21-2.75	1.99 ± 0.50	-0.93
Acetyl	2.06-3.40	2.74 ± 0.28	1.53	1.76-3.29	2.81 ± 0.38	1.50	2.09-3.01	2.61 ± 0.34	-1.15
Klason lignin	17.7-26.5	20.7 ± 1.35	8.30	18.2-25.7	20.5 ± 1.61	5.01	17.5-22.2	20.1 ± 1.48	-0.06
Ash after extraction	0.59-5.16	2.60 ± 1.64	1.18	0.59-4.47	2.53 ± 0.91	0.56	0.59-4.25	2.26 ± 1.46	-1.73
Extractives	3.61-8.72	5.59 ± 0.86	1.35	3.59-9.10	5.66 ± 1.16	3.48	4.64-6.60	5.61 ± 0.57	0.03

Table A.3. Composition of botanical fractions *Miscanthus* × *giganteus* as determined by bioprocess engineering laboratory

Sample ID	Composition (%)					
	Extractives	Glucan	Xylan	Arabinan	Lignin	Ash
Rind 1	5.73	46.55	20.94	-	23.76	0.56
Rind 2	5.45	45.57	20.63	-	24.67	0.69
Rind 3	4.52	46.83	21.37	-	24.10	0.67
Rind 4	5.75	46.66	20.77	-	24.25	0.65
Rind 5	5.11	45.23	21.04	-	24.05	0.83
Rind 6	4.62	45.69	20.61	-	24.09	0.77
Pith 1	7.55	38.72	25.60	2.94	22.05	0.76
Pith 2	5.42	39.13	24.25	-	22.92	0.59
Pith 3	5.70	38.39	25.48	-	24.04	0.85
Pith 4	7.91	38.76	25.22	-	22.47	0.51
Pith 5	5.46	37.47	26.37	3.31	23.23	0.76
Pith 6	5.99	38.68	25.11	-	23.05	0.75
Sheath 1	3.69	40.50	23.78	3.53	20.75	2.74
Sheath 2	3.21	39.60	23.94	3.56	21.15	2.66
Sheath 3	4.17	39.41	22.75	3.51	19.93	3.60
Sheath 4	3.94	42.54	23.32	2.40	20.73	2.23
Sheath 5	5.33	40.06	23.51	4.91	19.74	3.26
Sheath 6	4.27	40.39	24.15	0.05	20.41	3.31
Blade 1	10.40	32.91	21.76	5.71	19.42	8.04
Blade 2	10.08	32.01	21.51	6.03	18.27	8.66
Blade 3	8.23	32.43	21.95	6.16	18.67	8.47
Blade 4	9.87	32.03	21.83	6.08	18.53	8.92
Blade 5	8.95	32.46	21.89	6.32	18.36	9.32
Blade 6	9.32	31.13	20.73	6.15	18.80	10.04
Node 1	8.13	38.38	23.93	-	25.55	0.43
Node 2	7.31	38.22	23.54	-	26.22	0.29
Node 3	6.82	38.15	24.34	-	25.59	0.43
Node 4	8.33	38.24	23.03	-	25.14	0.41
Node 5	7.31	38.10	23.92	-	25.71	0.46
Node 6	7.58	38.91	23.80	-	24.88	0.57

APPENDIX B. PLS REGRESSION MODELS OF MISCANTHUS COMPOSITION

B.1. Glucan content

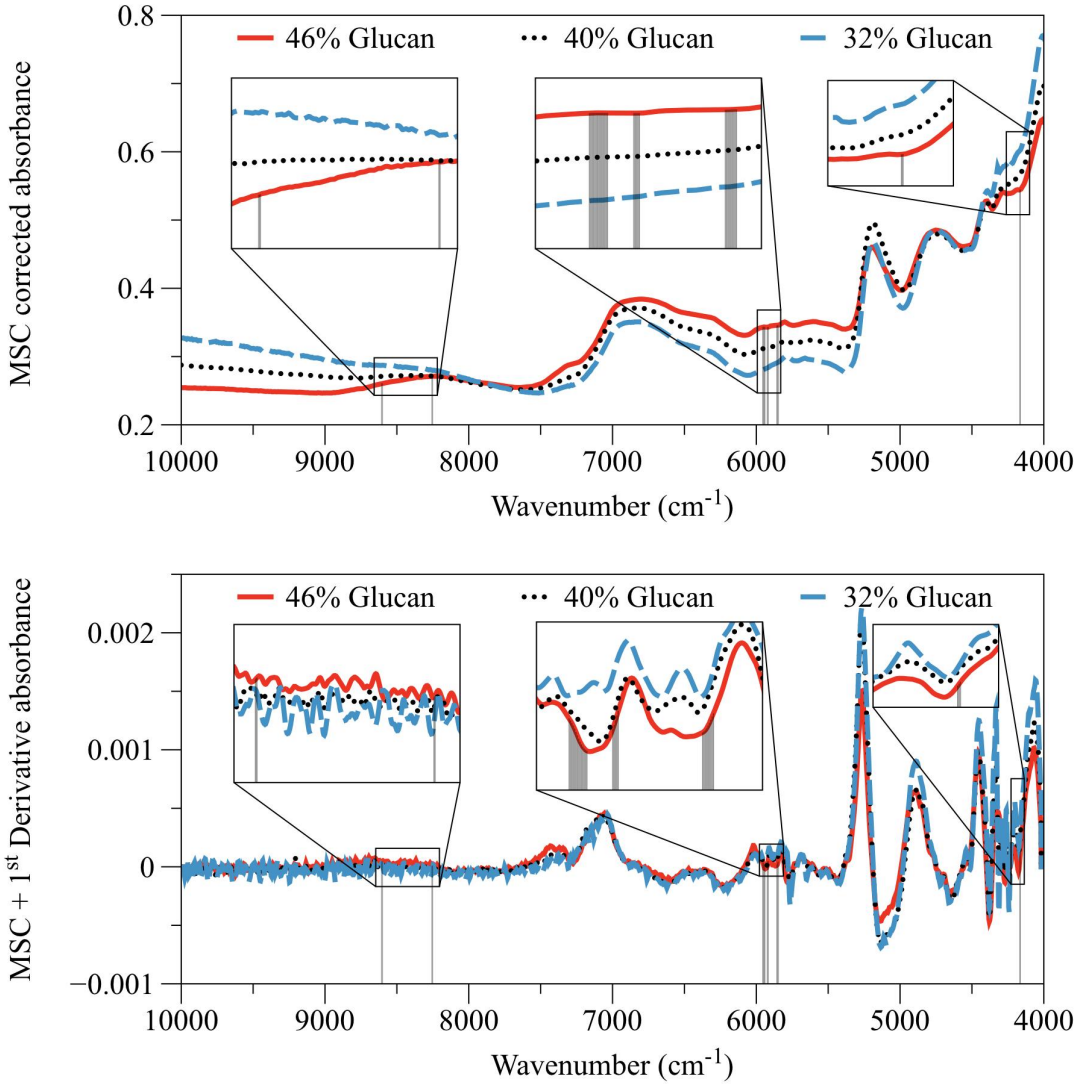


Figure B.1. MSC and MSC+SG 1st Derivative preprocessed spectra of select Miscanthus samples. The shaded regions represent wavenumbers that were utilized in the model to predict glucan content.

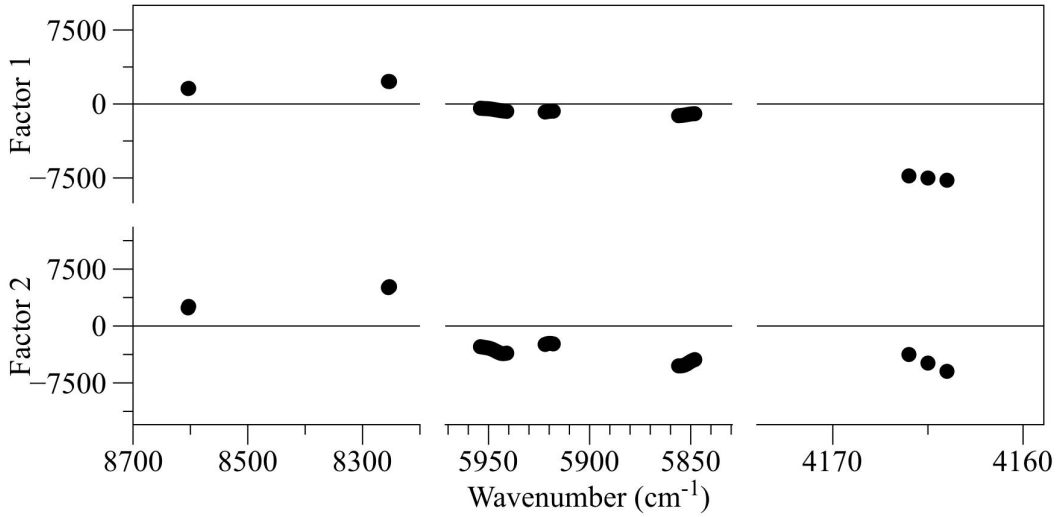


Figure B.2. Regression coefficients for the factors utilized in the glucan content model.

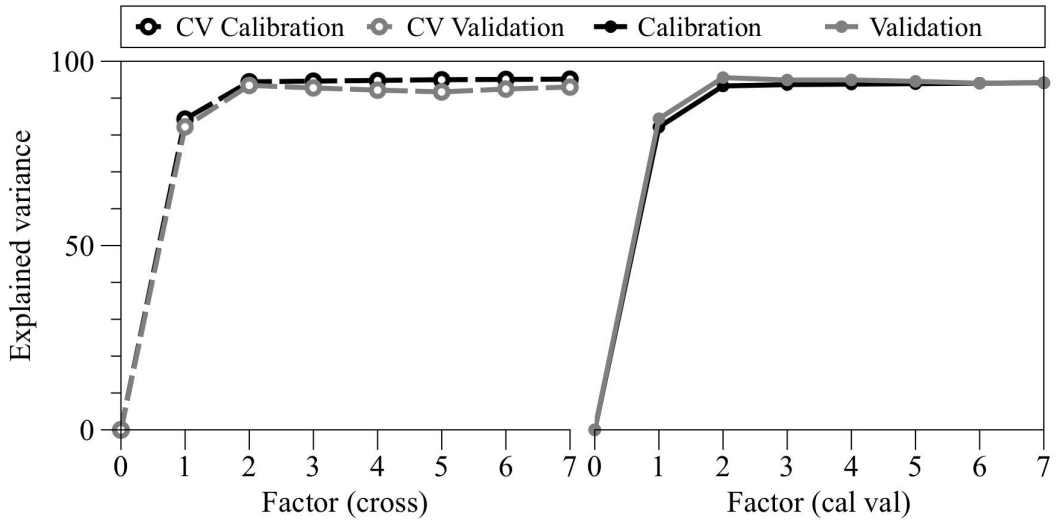


Figure B.3. Explained variance curves for the cross validation and independent calibration and validation models for glucan content.

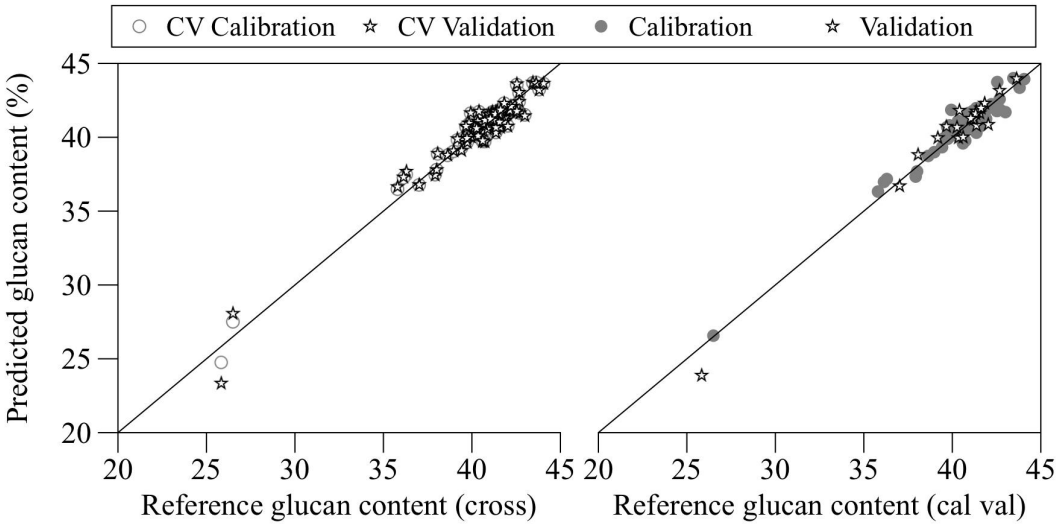


Figure B.4. Comparison of predicted to reference glucan contents after cross validation and independent validation.

B.2. Xylan content

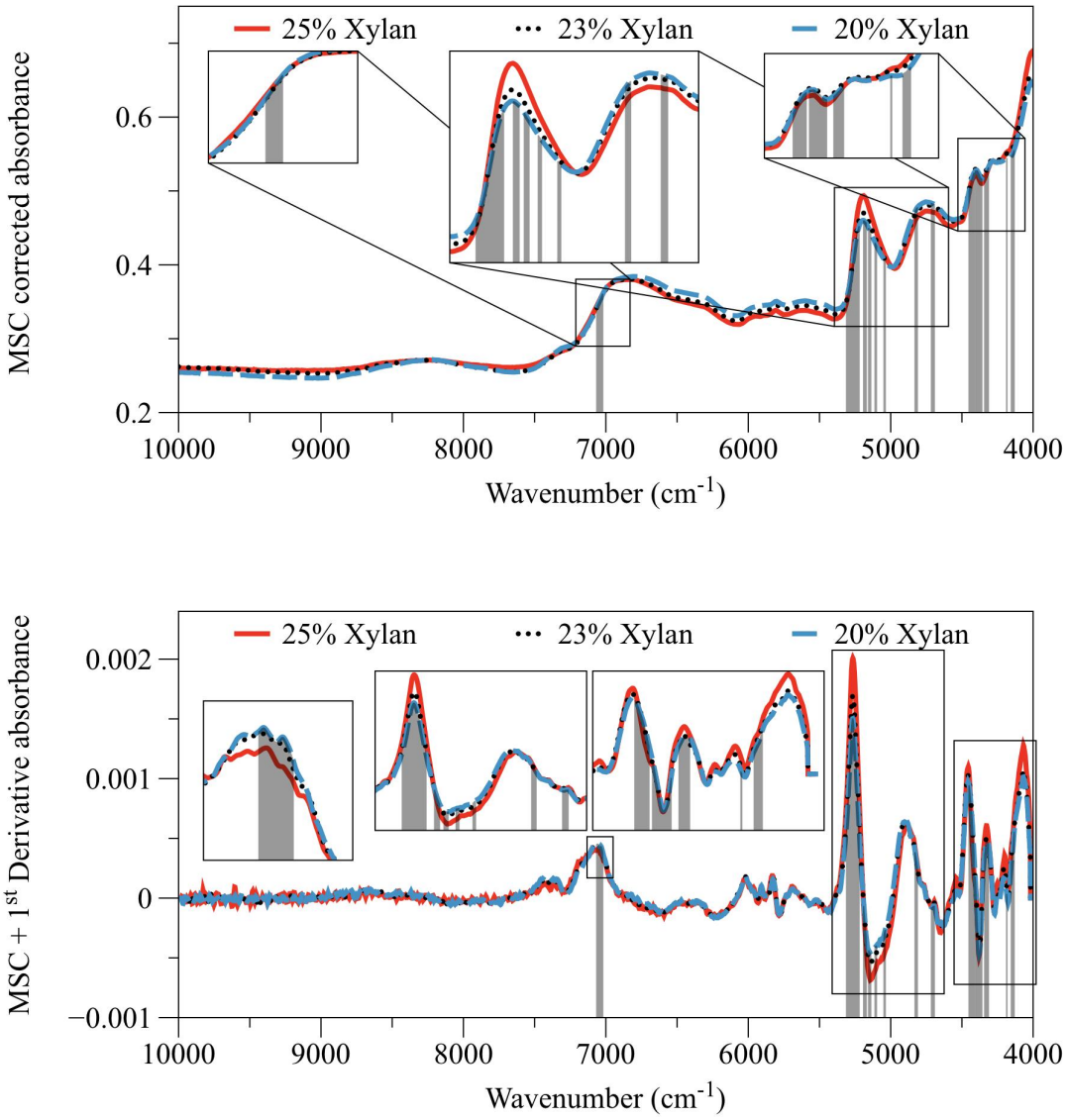


Figure B.5. MSC and MSC+SG 1st Derivative preprocessed spectra of select Miscanthus samples. The shaded regions represent wavenumbers that were utilized in the model to predict xylan content.

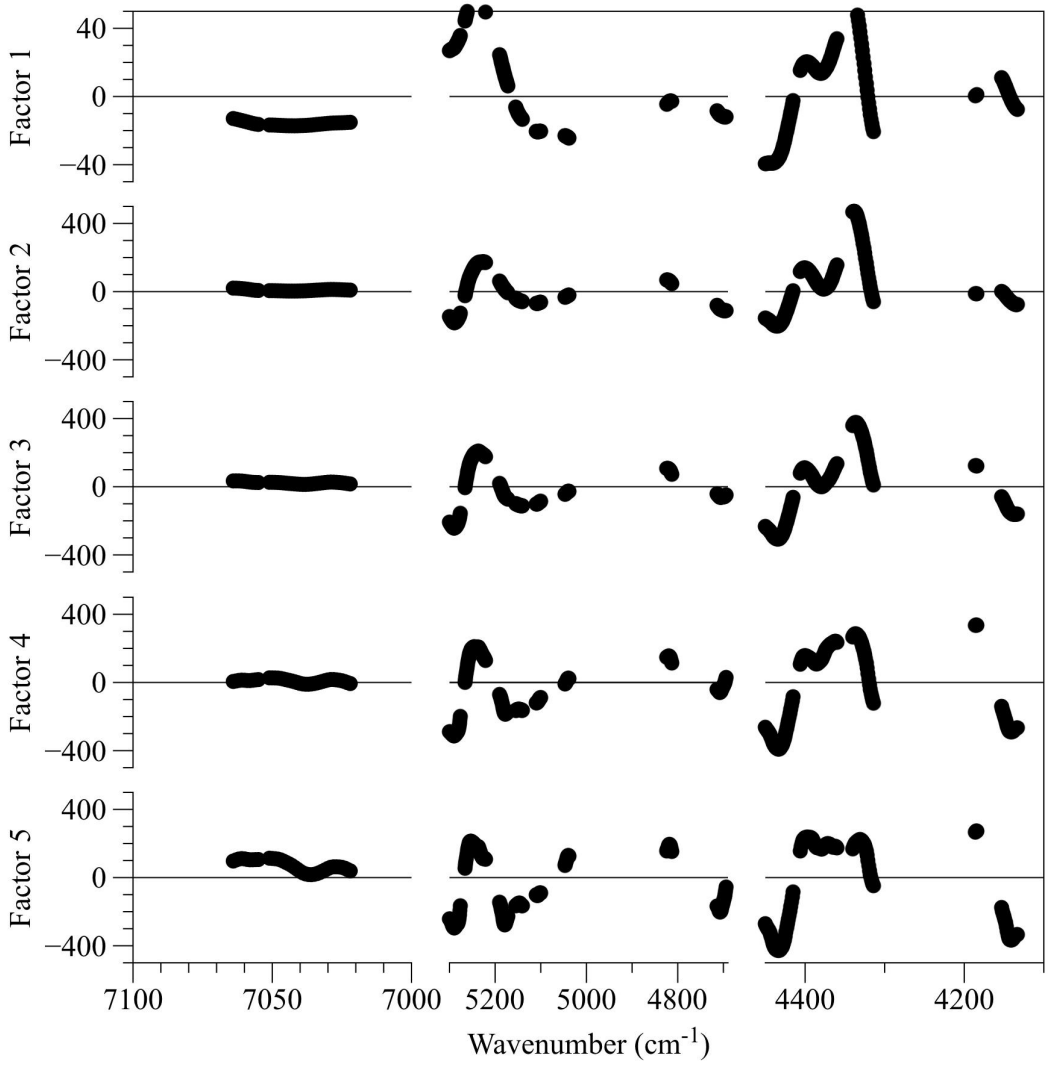


Figure B.6. Regression coefficients for the factors utilized in the xylan content model.

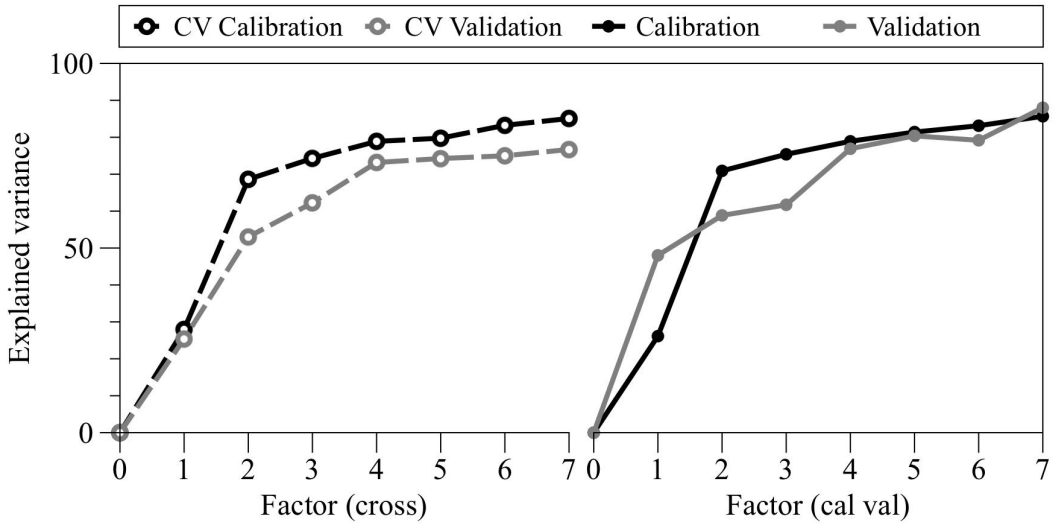


Figure B.7. Explained variance curves for the cross validation and independent calibration and validation models for xylan content.

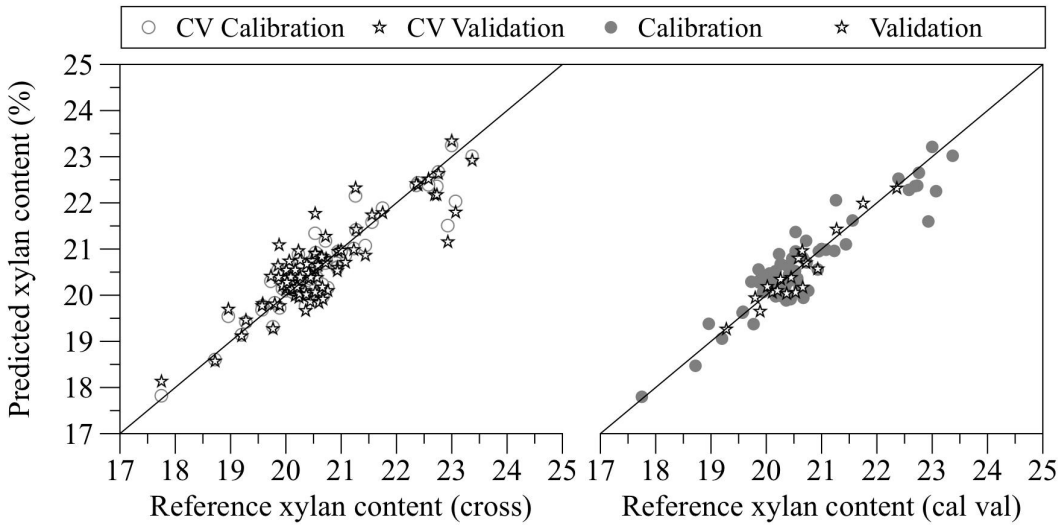


Figure B.8. Comparison of predicted to reference xylan contents after cross validation and independent validation.

B.3. Arabinan content

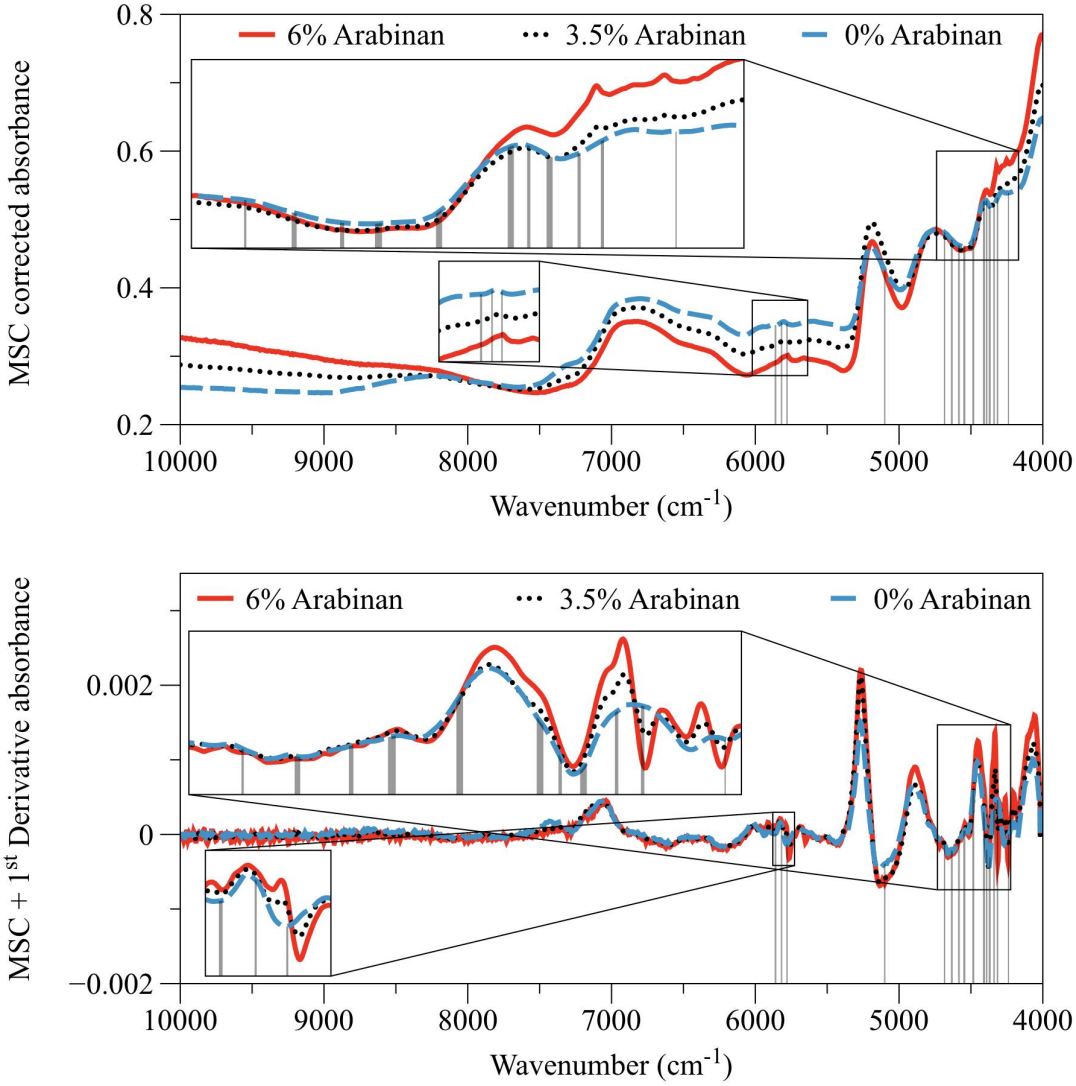


Figure B.9. MSC and MSC+SG 1st Derivative preprocessed spectra of select Miscanthus samples. The shaded regions represent wavenumbers that were utilized in the model to predict arabinan content.

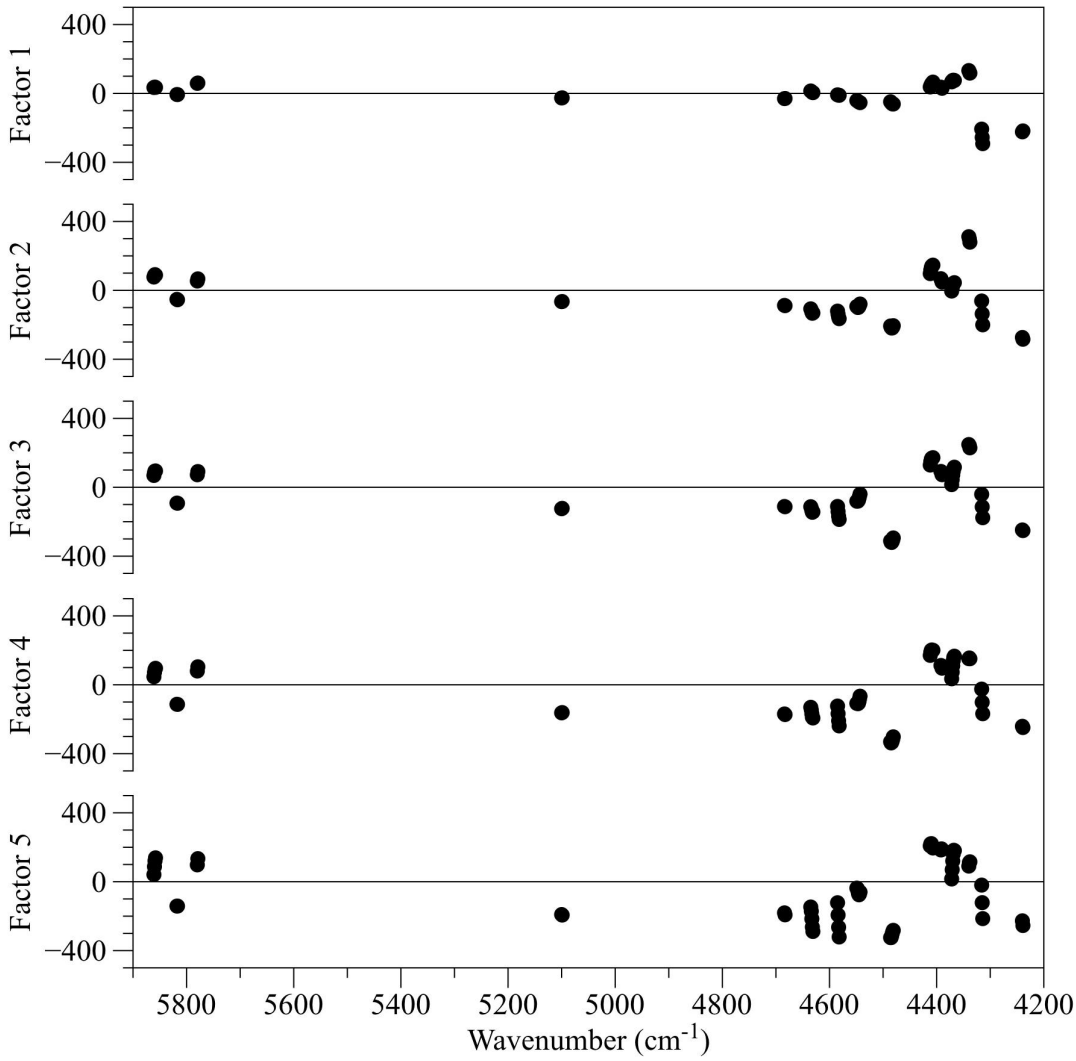


Figure B.10. Regression coefficients for the factors utilized in the arabinan model.

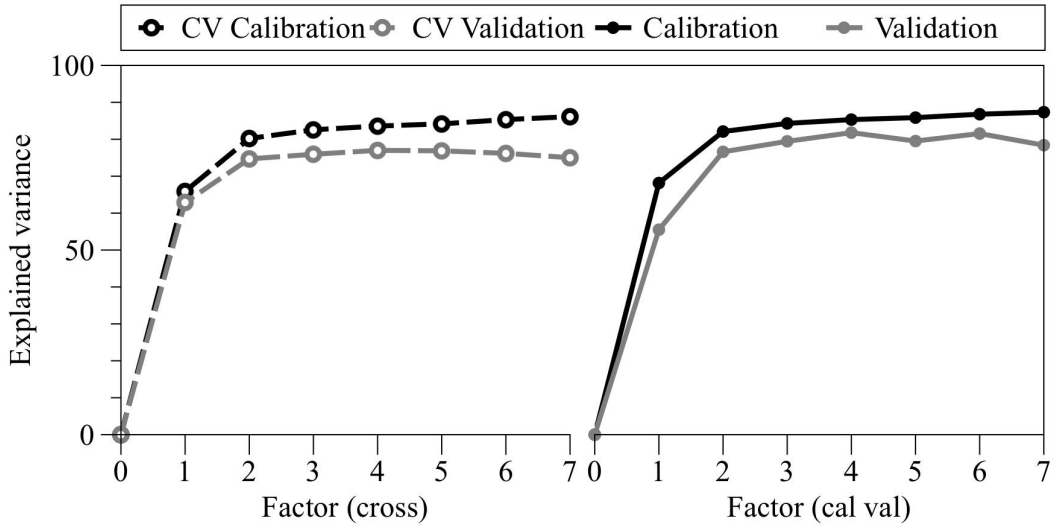


Figure B.11. Explained variance curves for the cross validation and independent calibration and validation models for arabinan content.

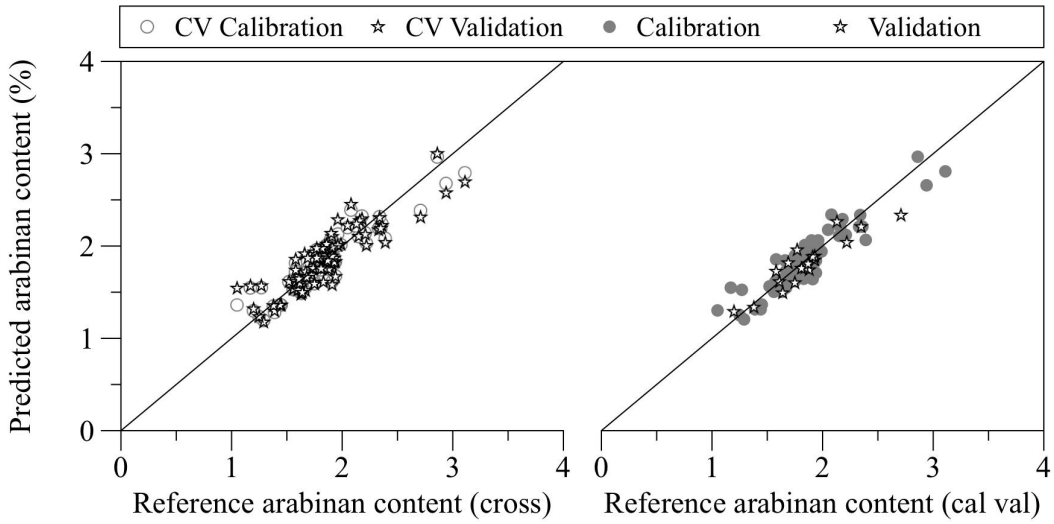
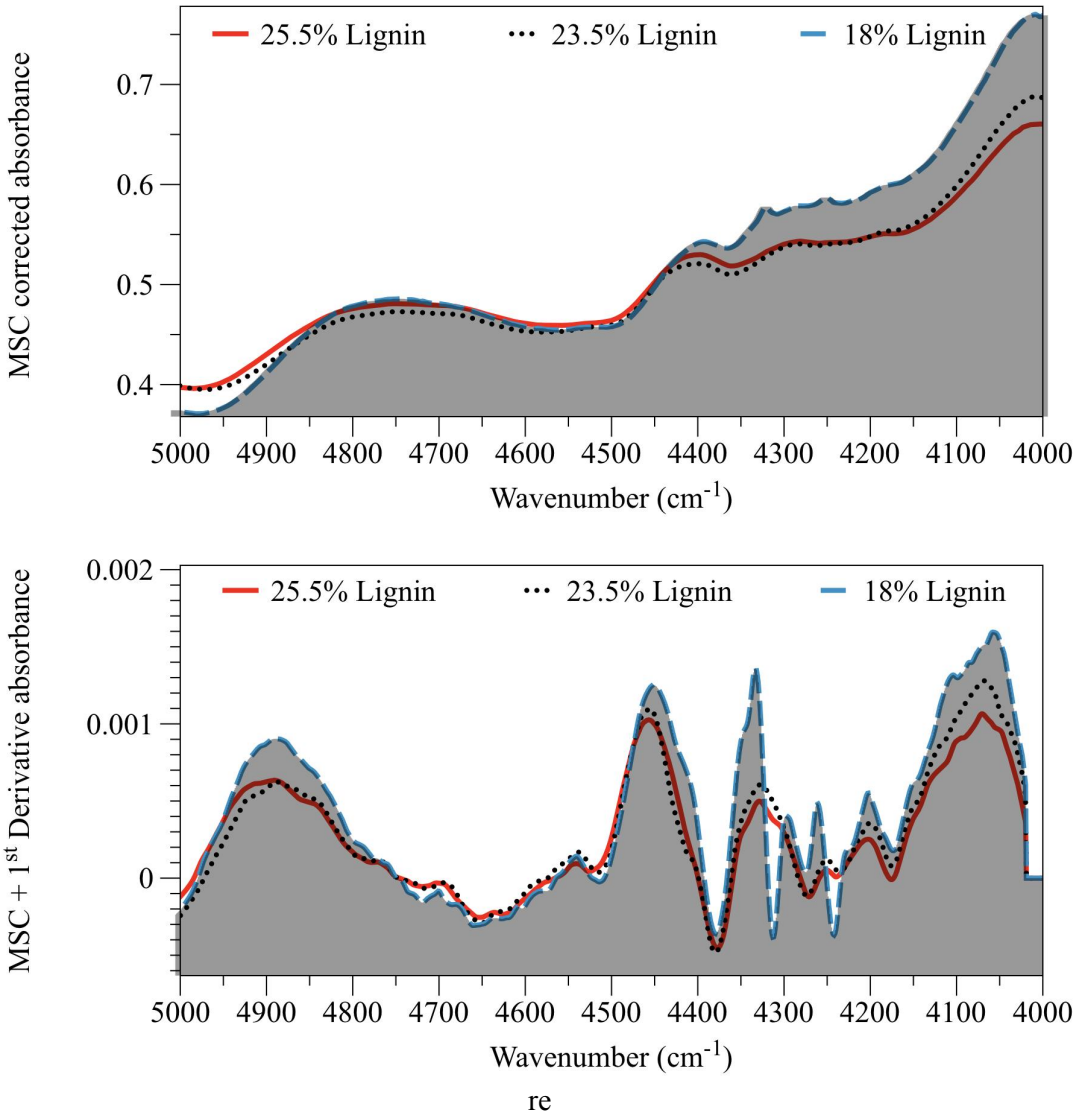


Figure B.12. Comparison of predicted to reference arabinan contents after cross validation and independent validation.

B.4. Lignin content



re

Figure B.13. MSC and MSC+SG 1st Derivative preprocessed spectra of select Miscanthus samples. The shaded regions represent wavenumbers that were utilized in the model to predict lignin content.

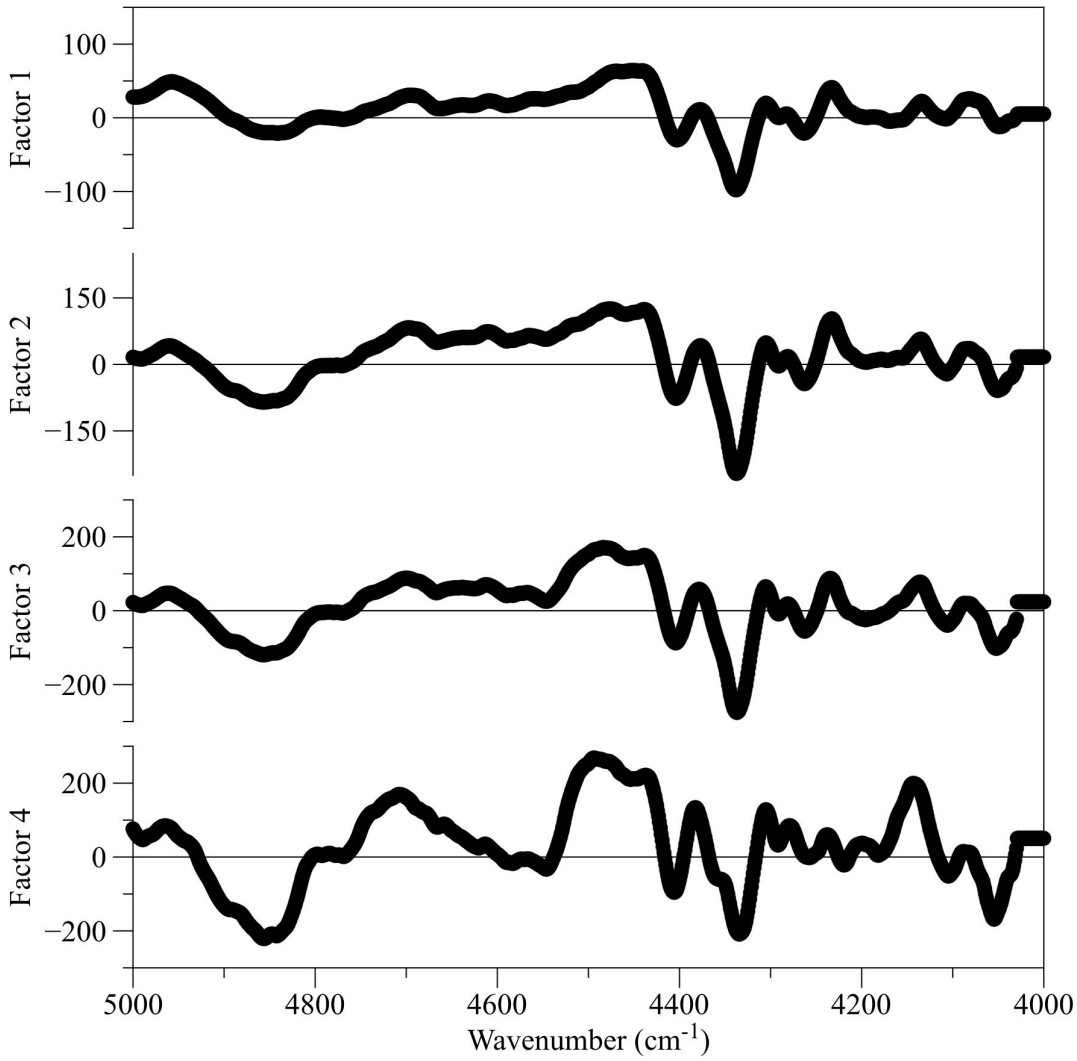


Figure B.14. Regression coefficients for the factors utilized in the lignin content model.

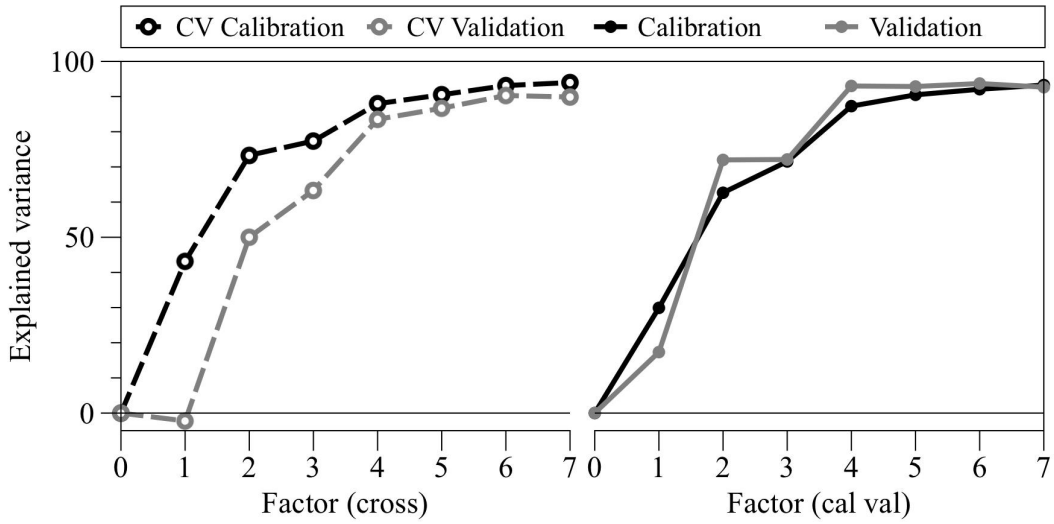


Figure B.15. Explained variance curves for the cross validation and independent calibration and validation models for lignin content.

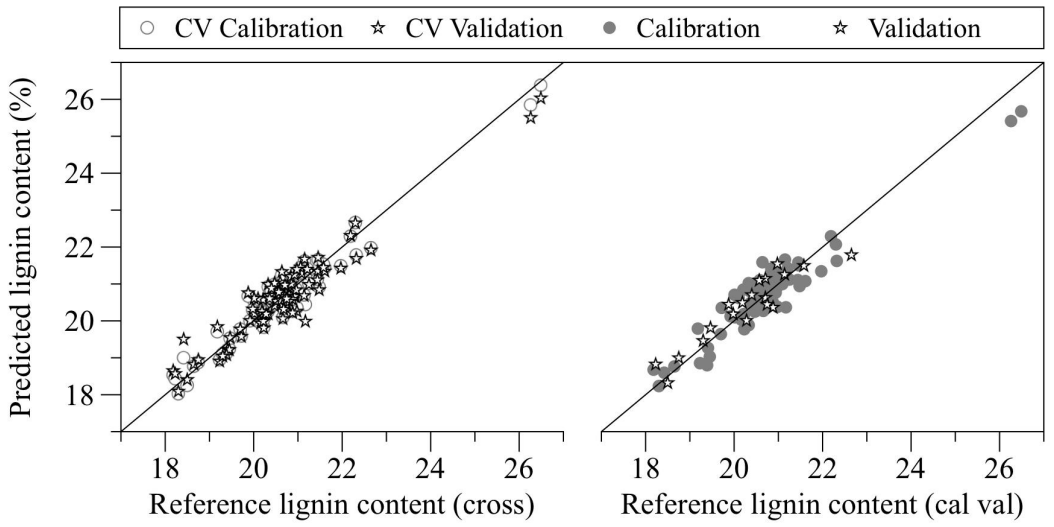


Figure B.16. Comparison of predicted to reference lignin contents after cross validation and independent validation.

B.5. Ash content

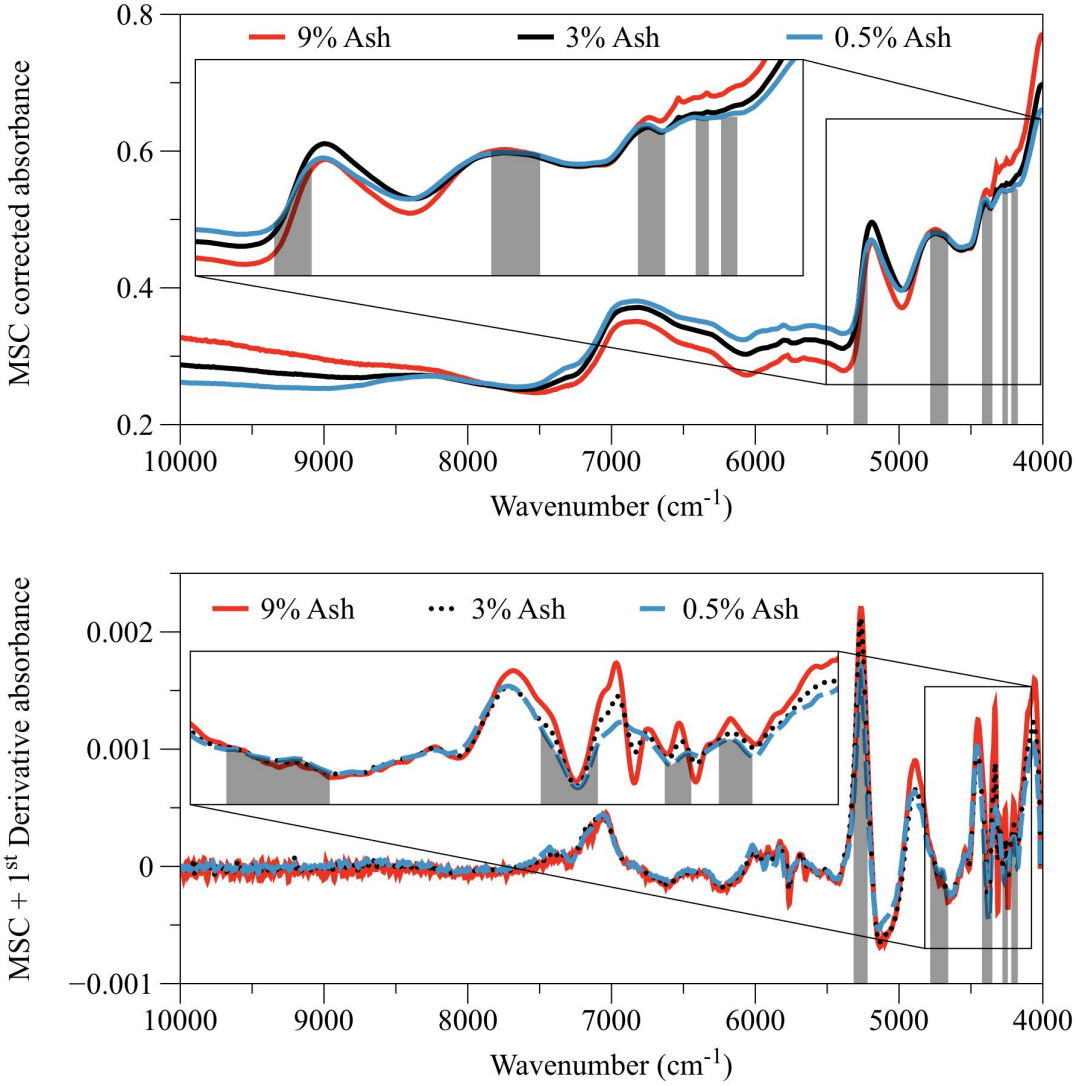


Figure B.17. MSC and MSC+SG 1st Derivative preprocessed spectra of select Miscanthus samples. The shaded regions represent wavenumbers that were utilized in the model to predict ash content.

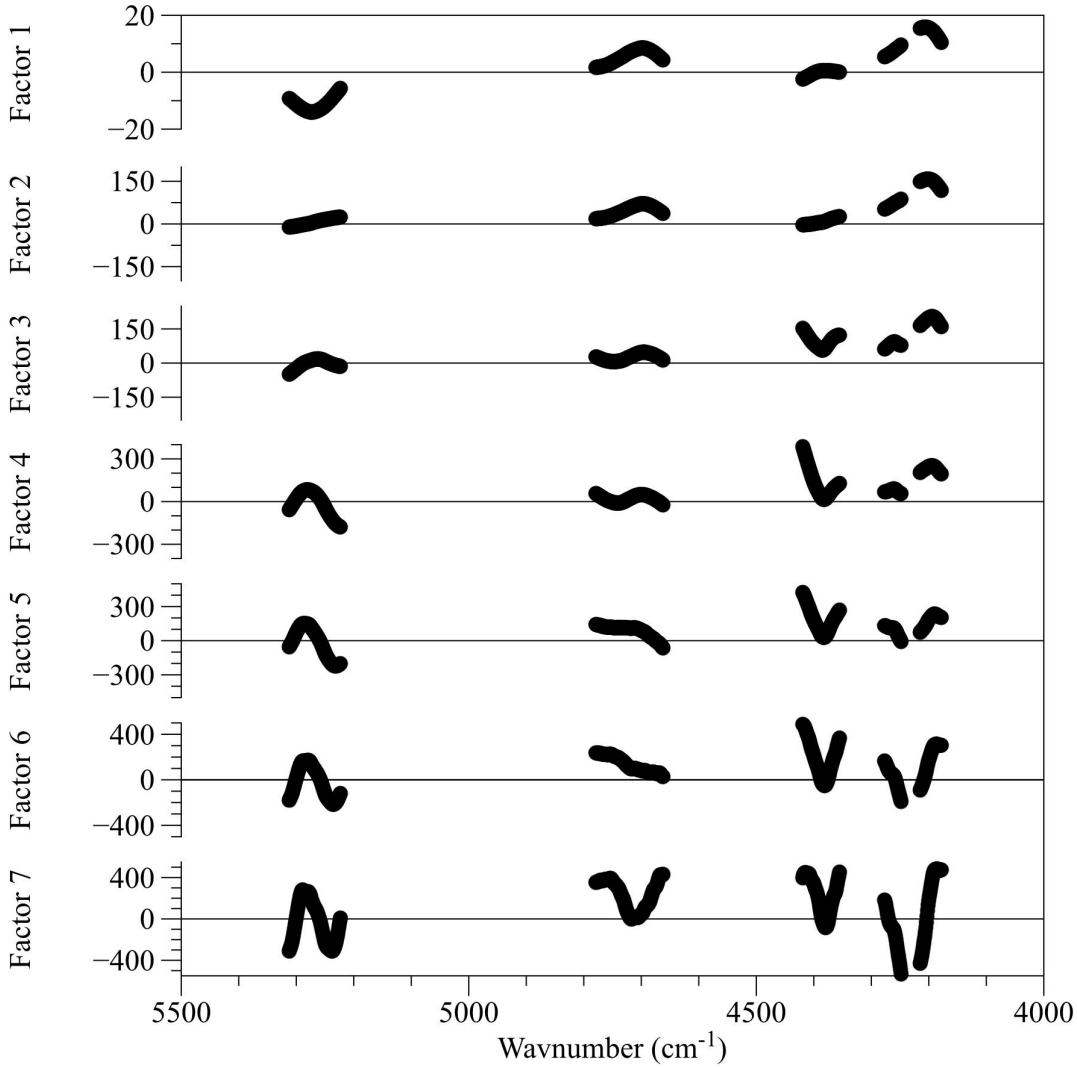


Figure B.18. Regression coefficients for the factors utilized in the ash content model.

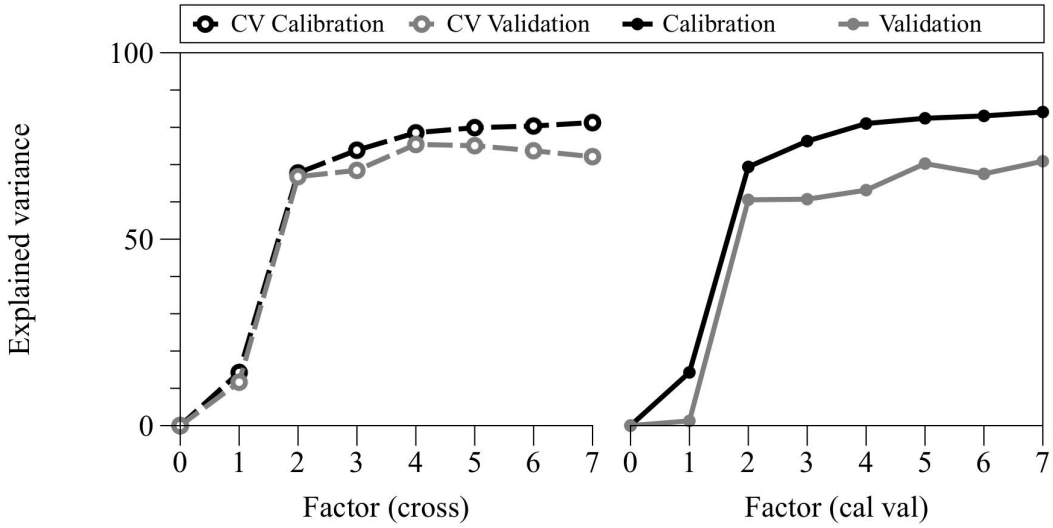


Figure B.19. Explained variance curves for the cross validation and independent calibration and validation models for ash content.

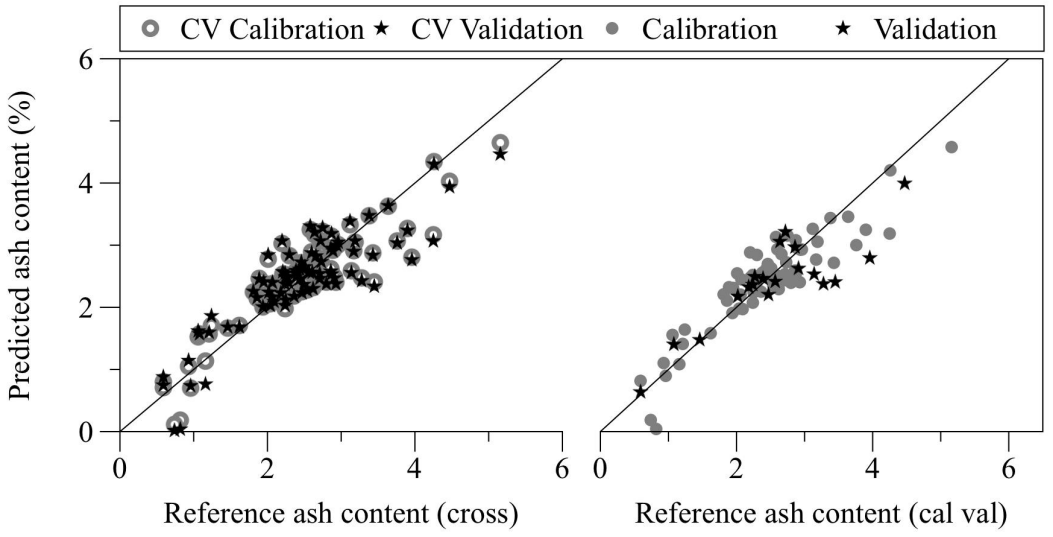


Figure B.20. Comparison of predicted to reference ash contents after cross validation and independent validation.