Chemometric Modeling of the Structural Integrity, Chemistry and Bioenergy Potential of Elite Loblolly Pine Families and Forest Biomass

by

Gifty Ewurama Acquah

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 10, 2016

Keywords: *Pinus taeda*, infrared spectroscopy, thermogravimetric analysis, wood quality, bioenergy potential, forest biomass

Approved by

Brian Via, Chair, Associate Professor of Forestry and Wildlife Sciences
Lori Eckhardt, Professor of Forestry and Wildlife Sciences
Oladiran Fasina, Professor of Biosystems Engineering
Nedret Billor, Professor of Mathematics and Statistics
Sushil Adhikari, Associate Professor of Biosystems Engineering

Dedication

This work is dedicated to my parents Samuel and Rebecca Acquah, my siblings Emmanuel Acquah, Samuella Acquah, Harriet Acquah and Samuel Acquah, and my husband Franklin Hewlett.

Abstract

Pinus taeda L. (loblolly pine) is the most economically important tree species in the USA. With 30 million acres in plantations in the southern United States alone, it accounts for over 50% of the standing pine volume of the region. Over the past sixty years however, Southern Pine Decline (SPD) has been causing the premature death of this species. As a management strategy, stakeholders are selecting elite loblolly pine families that are currently being screened for tolerance to SPD for deployment. However, before deploying these elite families, important wood traits that dictate the quality of this essentially new feedstock towards different end uses must be known.

This study examined the rapid screening of elite loblolly pine families for important wood properties and optimum utilization pathways using near infrared spectroscopy (NIR). In addition, the genetic variation, site and genotype by site interaction for the wood traits were investigated. Apart from its contribution to the conventional forest products industry, loblolly pine will play a role in the emerging bioeconomy. As such, this study also demonstrated the potential of NIR, as well as Fourier transform infrared spectroscopy (FTIR) and thermogravimetric analysis (TGA) in the rapid assessment of the heterogeneous loblolly pine biomass.

NIR was used to rapidly determine the density, strength properties, chemical composition and bioenergy potential of the elite loblolly pine families. Three to five latent variables were used in the development of NIR-based partial least squares (PLS) models that had R^2 values (cross-validation) of 0.58 to 0.88, and RPD values of 1.54 to 2.48. Validated models were employed in the screening of the families.

The effects of family, site and family by site interaction were tested for the properties. Genotype of the loblolly pine families affected all the studied traits. In addition, the interaction term was significant for all the properties except for MOE. As such, tree breeders should bear in mind that desired traits of the elite families might be unstable on different sites. Further studies with more sites would be useful in estimating the extent of the genotype by site interactions.

Nonetheless, desirable properties of some families remained high on the two forest sites. For example, A1, A26, A15, A2 and A9 which had consistently high cellulose contents on the two sites also had higher density, modulus of rupture (ultimate strength) and modulus of elasticity (stiffness). In addition, the amount of cellulose will affect the yield of pulp or ethanol. On the flip side, the strength-related properties of A33 and A21 remained low on both study sites. Apart from these two being undesirable for structural applications, their low strength properties could also make them more vulnerable to inclement weather on site.

For the studies on the loblolly pine biomass, NIR and FTIR were used to classify forest residue into the plant part components of wood, wood & bark and slash (i.e. tops and limbs). Linear discriminant models that were developed with raw NIR and FTIR spectra had classification accuracies of over 96% for both tools.

With respect to the quantitative assessment of biomass, full-cross-validated PLS and principal components regression (PCR) models were used. This study demonstrated that TGA coupled with chemometrics can be used for the compositional analysis of lignocellulosic biomass. The developed methodology enabled the simultaneous prediction of both the chemical and proximate properties from a single thermogram. According to the literature, this has not been attainable by the conventional deconvolution of TG data. In addition to its rapidity and simplicity, this alternative technique allowed the prediction of some monomeric sugars. Further studies will

however be necessary to validate the capability of chemometrics to model the thermal degradation and quantitative prediction of the individual monomeric sugars.

Comparing the predictive performance of the three analytical tools investigated in this study, the NIR models generally had better diagnostics relative to the FTIR and TGA models in predicting the chemical composition and thermal reactivity properties of the heterogeneous loblolly pine biomass.

Acknowledgements

First and foremost, I owe my deepest gratitude to God for His abundant grace, favor and strength throughout my stay and study in Auburn.

I sincerely thank Dr. Brian Via, my major advisor for his advice, guidance and support throughout my study. In addition, I am most grateful to my committee members, Dr. Lori Eckhardt, Dr. Oladiran Fasina, Dr. Nedret Billor and Dr. Sushil Adhikari for their steadfast support and insightful contributions to the success of this work. I also thank Dr. Gisela Buschle-Diller for serving as the outside reader for my dissertation.

I would also like to say a big thank you to the NSF-Auburn IGERT: Integrated Biorefining for Sustainable Production of Fuels and Chemicals, the Forest Health Cooperative, Rayonier, Plum Creek and West Fraser Inc. for their financial and material support of this study. Special thanks also go to the School of Forestry and Wildlife Sciences for giving me this great opportunity to begin with.

To all my colleagues at the Forest Products Development Center, Forest Health Dynamics Lab and the Center for Bioenergy and Bioproducts, my heartfelt appreciation to you all for the myriad ways you helped me with this research study. I am forever indebted to you.

Finally, to my family and friends who have supported me tirelessly with prayers, love and encouragement throughout my endeavors, I express my deepest gratitude and appreciation.

Table of Contents

List of Tables

List of Figures

Chapter 1 : Introduction and Literature Review

1.1 Timber Production in the United States

The United States has 304 million hectares of forested land, out of which two-thirds is classified as timberland. Of the remaining, 66 million hectares is classed as non-timberland and 30 million hectares is reserved. The forests of the southeastern region make up 27% of the total forest land and 30% of the unreserved forest area. *Pinus taeda* L. (loblolly pine) and *P. enchinata* Mill. forests cover almost a quarter (i.e. 20 million hectares) of southern forests and accounts for over 50% of the 38 million hectares of softwood forests in the eastern United States (Smith et al. 2009).

Seventy-seven percent of timberland is located in the eastern United States. The southeastern region, which makes up 55% of the total timberland area accounts for 64% of the total timber harvested in the country (Smith et al. 2001). This produces 60% of the wood consumed nationally (Wear and Greis 2002) and supplies 18% of the world's industrial timber, making the region the world's single largest producer of industrial wood (Prestemon and Abt 2002).

1.2 Loblolly Pine (*Pinus taeda* L.)

Loblolly pine is an important tree species in the United States. It dominates on approximately 13.4 million ha throughout the southeastern forests, and accounts for over 50% of the standing pine volume of this region (Schultz 1997). Loblolly pine is native to fourteen states extending from south New Jersey, down to central Florida and west to eastern Texas. Due to its adaptability, it has alsobeen successfully planted along the periphery of its natural range, and even on other continents (Baker and Langdon 1990). This tree species singularly provides some 110, 000

direct and indirect jobs and contributes approximately $30 billion to the economy of the region (Schultz 1999).

1.3 Problem Statement

Over the last six decades, Southern Pine Decline (SPD) (previously known as Loblolly Pine Decline or Pine Decline) has been causing reduced growth, decline and eventual premature deaths of loblolly pine especially in the southeastern United States. In a bid to control this disease complex, stakeholders would like to select and deploy elite loblolly pine families that are currently being screened for tolerance against fungi associated with SPD.

The southern pine industry is interested in knowing the wood properties of these elite loblolly pine families since they are essentially a new feedstock. However, current methods that are used to determine wood properties are laborious, costly and usually destructive and will not be feasible in tree breeding programs where a large number of trees have to be sampled. There is therefore the need for novel analytical tools that can rapidly and cost effectively characterize this resource in a non-destructive manner. This will aid in decision making concerning the suitability of this resource as a feedstock for specific applications such as paper, structural lumber or engineered wood products.

In addition to the conventional forest products industry, loblolly pine will play a key role as the nation seeks a sustainable bioeconomy (U. S. DOE 2016). The need to shift to a sustainable low-carbon economy is more pressing now than ever due to increasing negative anthropogenic impacts on the planet. One way of attaining such an economy is by using renewable resources to replace and/or supplement materials derived from non-renewable sources such as fossil fuels and metal ores (Scheffran 2010; White 2010; Perlack et al. 2005). Among the current sources of renewable energy, only biomass can be converted into liquid transportation fuels, chemicals, and products that

are currently obtained from crude oil (Klass 1998). In addition, it is the only one that can be transported to other locations or stored for future utilization. The use of lignocellulosic biomass to replace products derived from non-renewable sources will reduce net greenhouse gas emissions and persistent toxic materials resulting from the extraction and processing of non-renewable resources. This will enhance the health of both mankind and the environment (NRCan 2007), and at the same time decrease our dependence on these resources (Mayfield et al. 2007; Tampier et al. 2004).

In the United States, lignocellulosic biomass is primarily sourced from agricultural and forestry operations. Biomass from the forest consists of the higher end wood (for timber or pulpwood), as well as forest residues and mill residues that have been conventionally considered 'waste'.

Forest residues is a primary resource that consists mostly of logging residues (i.e. tree tops, limbs and stumps). It also may include excess biomass removed from fuel treatment and commercial thinning; rejected material that does not meet mill specifications such as oversized logs, cut off cankers and other defects, tree species that are currently not marketable as timber or pulpwood, as well as trees salvaged after being damaged by fire, wind and other types of forest disturbances (Bradley 2009; Ralevic et al. 2008). Globally, residues can be up to 60% of the volume of a forest harvest operation (FAO 1990). Forest residues generated in the United States and Canada are however lower at 13% to 24% and 9% to 14% respectively (U. S. DOE 2011; Smith et al. 2009). Mill residues on the other hand is a secondary resource generated during the manufacture of wood products such as lumber and oriented strand board.

Whereas mill residues have conventionally been 'reconstituted' to produce other wood products such as fiber boards and also combusted for heat and power, forest residues are usually

left on the forest floor because its small piece size makes it unsuitable and uneconomic for the manufacturing of forest products. Although this material provides important ecological functions such as nutrient cycling and coarse debris for wildlife habitation, it however also poses a fire risk and hazard (White 2010; Patton-Mallory 2008).

Apart from combustion, forest residues can be gasified in an oxygen-deprived environment (about a third of oxygen that would have been needed for conventional complete combustion) to produce syngas which can in turn be used in the production of heat, electricity and process steam, or further refined to serve as a feedstock for synthetic natural gas and high value chemicals (NRCan-CETC 2008). Alternatively, forest residues can be processed into bio-oils through the process of fast pyrolysis or liquefaction. Bio-oils can be utilized as an alternative fuel, or as a feedstock for the production of value-added products such as methane, ethanol, artificial flavors, wood resins and fertilizers. Phytochemicals also can be thermally extracted from forest biomass (NRCan 2007). In addition, due to the increasing cost of fiber (SCNR 2008), forest residues can be used as supplementary fiber in the manufacture of pellets (Bradley 2009) and composite board products.

As already stated, forest residues are currently an abundant resource that can become economically feasible to remove, most likely together with conventional harvest operations where the cost of extraction will be borne by the higher end timber and pulpwood as market for bioenergy and biofuel feedstocks develop (U.S. DOE 2016). As this emerging industry expands, it is most likely that it will compete especially with the pulp and paper industry for pulpwood. For the efficient and effective use of this resource as a raw material, its properties must be well understood. Just as in the case of the higher end solid wood, time-saving and cost effective analytical tools are needed for the rapid characterization of biomass.

Over the last two decades, infrared spectroscopy and thermogravimetric analysis have gained popularity as high throughput tools that can be used to accurately estimate the properties of wood and lignocellulosic biomass.

1.4 Infrared Spectroscopy

The infrared region is the wavelength range of 780 nm – 1 mm (wave number range 12820 – 10 cm^{-1}) that lies between the visible and microwave regions of the electromagnetic spectrum, Figure 1.1. It is subdivided into near infrared (NIR), mid infrared (MIR) and far infrared (FIR). The NIR region lies between the wavelength range of 780 to 2500 nm (wave number range 12820 – 4000 cm^{-1}) and the MIR region ranges from 2500 nm to 15380 nm (4000 – 650 cm^{-1}) (Adapa et al. 2009; Reich 2005; Sherman Hsu 1997).

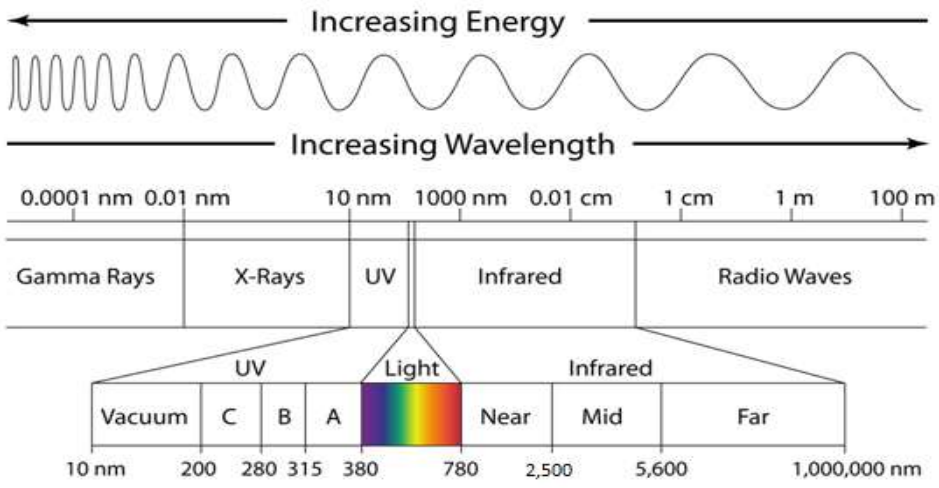

Figure 1.1: Infrared region in the electromagnetic spectrum (Adapted and modified from Google Images 2013).

Infrared spectroscopy is the measurement of the wavelength and intensities of the absorption and transmittance or reflectance of infrared light by a sample. NIR spectroscopy uses near infrared light to detect overtones and combinational vibrations of the molecular constituents

of the material under study; while Fourier Transform infrared spectroscopy (FTIR) uses mid infrared light to detect primarily functional and fundamental vibrations (Zhou et al. 2011).

In the NIR region, the measurement of reflectance instead of transmittance is possible because of the low molar absorptivity in the region; this low absorption coefficient however permits deeper depths of penetration. In transmittance, NIR radiation is detected by the instrument after it has passed through the sample, whereas in reflectance, the instrument detects radiation reflected by the sample (especially opaque sample) (Benito et al. 2008).

In the case of FTIR, an interferometer divides radiant beams and generates a difference in their optical paths. It then recombines them to produce repetitive interference signals (this contains infrared spectral information) which is measured by the detector as a function of the different paths (Sherman Hsu 1997).

Common absorption bands that occur in the NIR and FTIR regions are overtones and combinations of fundamental vibrations of C-H, O-H, C=O, N-H, -COOH, aromatic C-H groups and S-H functional groups in the mid infrared region; and can thus give the chemical and physical properties of a sample. In simpler terms, the chemical finger print (in the form of a spectrum) of a sample is taken by NIR or FTIR at a specific point in time (NIR Technologies Inc. 2012; Jorgensen and Goegebeur 2007; Reich 2005).

NIR and FTIR have several advantages that justify their widespread use. They can provide rapid and accurate analysis of a large number of heterogeneous samples in a non-destructive manner (i.e. as-is or with minimal preparation). Infrared spectroscopy also avoids the need of chemicals and their subsequent disposal. Furthermore, the cost-per-test is much lower than other analytical methods. Some NIR instruments operate on batteries, are lightweight and portable, thus making them suitable for field studies. Additional advantages for FTIR include increased optical

throughput, and elimination of stray light and emissions (Benito et al. 2008; Mark and Campbell 2008; Hames et al. 2003). Some common instruments used include photometers, spectrometers, Fourier transform spectrophotometers (FT-NIR), acousto-optically tuned filters (AOTF-NIR), attenuated total reflectance (ATR)-FTIR and diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS) among others (Heise and Winzen 2002; Pope 1995).

Infrared spectroscopy thus has a huge potential as a fast, accurate and non-destructive analytical tool to identify and characterize a myriad of products of instantaneous forms in such industries as the forestry and agricultural, petrochemicals, polymer and plastics, medical and pharmaceutical as well as military research and forensic science.

NIR and FTIR instruments use the complete spectra of several wavelengths to calculate the measurement of interest. Spectra therefore typically have broad ill-defined overlapping bands of all the chemical and physical information of an analyte that make it multivariate in nature, also, they can be non-linear due to scatter, stray light, path length variation, physical properties of sample such as size, inconsistency in instrument response and random noise (Reich 2005; Jorgensen 2000; Wold 1995). Therefore, mathematical, statistical and computer science methods known as chemometrics are needed in order to do a quantitative or qualitative analysis (i.e. relate spectral variables to the properties of a test sample.

1.5 Chemometrics and Multivariate Analysis (MVA)

Chemometrics extracts the relevant information and minimize irrelevant information from the chemical measurement data (Varmuza and Filzmoser 2009; Benito et al. 2008; Mark and Campbell 2008). It also enables the simultaneous analysis of all variables, and permits an overall evaluation of the significance of differences between groups and correlations. Results usually show unexpected patterns due to the interactive and synergetic effects of all the variables considered as

opposed to conventional analysis that typically consider one or several variables (Jorgensen 2000; Wold 1995).

In order to extract only the relevant information from chemical data, the NIR instrument must be 'calibrated' to do quantitative or qualitative analysis. Calibration involves the following steps: selection of a test set that is representative of the sample; acquisition of spectra and the determination of reference property; multivariate modeling to relate spectra to the reference; prediction and validation (Reich 2005; Blanco and Villarroya 2002). For quantitative calibration, sophisticated mathematical calculations relate measured spectra to known properties (including composition) of the samples determined by conventional methods; whereas in qualitative calibration (also known as classification), the instrument is trained to automatically identify components of an unknown sample by comparing its measurements to those taken on a known sample.

The role of chemometrics is to relate the spectra of all samples that are to be used in calibrating the instrument to the components of interest that have been determined by a conventional method so that when the spectra of similar samples are measured in the future the instrument can calculate their compositions (NIR Technologies Inc. 2012; Benito et al. 2008).

According to Varmuza and Filzmoser (2009), in spite of the broad definition of chemometrics, the most vital aspect of it is the application of multivariate analysis to chemistry-relevant data; multivariate analysis is powerful in analyzing and structuring IR spectra and conventionally acquired chemical data into empirical mathematical models that are capable of predicting properties of future measurements and even other properties that are not directly measurable.

IR spectra is usually pretreated before multivariate analysis due to scatter, stray light, path length variation, inconsistency in instrument response and random noise that cause such interferences as baseline shifts, vertical displacements, non-uniform slope of individual spectrum and curvilinear spectra. Pretreatment is a necessary step that minimizes, standardizes or even eliminates the impacts of these interferences on spectra and improves the robustness of calibration models to be developed. Commonly used pretreatment methods include standard normal variate (SNV) transformation, multiplicative scatter correction (MSC), derivatives and the newer orthogonal signal correction (OSC). SNV and MSC both minimize or remove baseline shifts, slope variation and intensity differences due to scatter by normalization algorithms. In addition to reducing baseline offsets, derivatives also improve the resolution of overlapping peaks. Derivatives however make spectra more complex and also amplify noise, thus they are done in conjunction with smoothing algorithms such as Savitzky-Golay algorithms. OSC removes latent variables that are orthogonal to the response variable (Lande et al. 2009; Blanco and Villarroya 2002; Jorgensen 2000).

Multivariate modeling means using many measured variables (X_1, X_2…..X_i) simultaneously to quantify a response or target variable (Y); X is NIR spectra and Y is the measured property (Martens and Naes 1990). Multivariate calibrations commonly used with IR spectroscopy are multiple linear regression (MLR), principal component regression (PCR) and partial least squares (PLS) also known as projection of latent squares.

Multiple linear regression: MLR is based on the inverse of Beer's law, that is, concentration is a function of absorbance (concentration is modeled linearly against combinations of absorbance). It assumes the error to be in the component concentrations and thus reduces squared errors in concentrations. A significant advantage of this model over the others is that, analysis based on it is invariant with respect to the number of chemical components. Even though it is a frequency-limited

method (because many frequencies cause collinearity and reduces precision), quantitative spectral analysis can be done even with the concentration of only one known component.

Principal component regression: PCR is the regression of principal components (i.e. PCA followed by a regression step). It is a factor analysis method that combines the many advantages of CLS and the ability to perform analysis one chemical component at a time, just like MLR. It also gives orthogonal intensities and thus avoids the collinearity seen in MLR. It presumes concentration to be a linear function of the intensities it generates in the new full-spectrum. PCR thus considers both spectra and concentration during modeling. It however factors the spectral data matrix without using information about the concentration, thus there is no assurance that the full-spectrum basis vectors associated with it will be relevant for predicting concentration.

Partial least squares: PLS is a generalization of ordinary least squares (OLS) regression that correlates subtle changes in spectra with the response variable (i.e. describes the underlying latent structure in the two) (Blanco and Villarroya 2002; Martens and Naes 1990). It has similar advantages to PCR. In addition, unlike PCR, PLS ensures a better predictive ability by sacrificing some fit of the spectral data so as to achieve better correlations to concentrations. It also gives coefficients of regression that can be used to relate chemical features in the spectra to the properties of the material. PLS also can predict more than one variable due to the non-iterative PLS (NIPALS) algorithm it uses to calculate model parameters. This has therefore made PLS the most extensively used with NIR applications in the biomass and bioprocessing industry (Lestander and Rhen 2005; Haaland and Thomas 1988).

The performances of models developed to predict the properties of wood and biomass are determined by such parameters as the standard error of calibration (SEC), standard error of cross-validation (SECV), coefficient of determination (R^2), standard error of prediction (SEP), bias and ratio

of preformance to deviation (RDP). The SEC evaluates how precisely the regression line fits the data. SECV is used to determine how well the models predicted the samples that are not used to create the calibration set. Bias detects any systematic difference between calibration set and the prediction set – thus the closer bias is to zero, the better the model both in fit and predicting ability. SEP measures the precision of a model's predicting ability corrected for bias during validation. R^2 mesures the total variance between measured and predicted that can be modeled linearly – the closer R^2 is to one, the better the model, but it not good as the ultimate measure of model performance. RDP is the ratio of SEP to standard deviation (SD); it evaluates SEP in terms of SD of the reference data (Kelley et al. 2004; Esbensen 2002).

1.6 Spectroscopic Studies of the Properties of Wood and Lignocellulosic Biomass

Cooper et al. (2010) partially differentiated the densities of earlywood and latewood with NIR and recommended that a large area should be illuminated especially when scanning is being done on the tangential surface so as to ensure a better representation of both earlywood and latewood. Nkansah et al. (2010) used diffuse reflectance NIR to predict the extractives, holocellulose and lignin content of *Liriodendron tulipifera* L. solid wood blocks (19 x 19 x 50 mm) whereas Sykes et al. (2005) used it to predict the properties of wafers (200 µm) sectioned from loblolly pine obtained from two sites. The former group noted that the models they developed with the full spectra (800 – 2500 nm) had slightly better predictive ability as compared to models they built with reduced spectra (1300 – 1800 nm). They had R^2 values of 0.84, 0.68 and 0.64 for extractives, holocellulose and lignin respectively. The latter group found out that even though models developed from samples of one site can be used to predict the cellulose content of wood from the other site, the R^2 value was lower than those for the individual site predictions and may be only sufficient for ranking and selection purposes. In both studies, the R^2 values for lignin content

were low. Sykes et al. (2005) attributed this result to the generally low variation of lignin content in wood and usually large errors associated with lignin determination in the lab.

Poke and Raymond (2006) also sought to increase the rapidity and cost-effectiveness of using NIR to predict the chemical composition of *Eucalyptus globulus* Labill using 20 x 20 mm test stripes. For comparison, they used both solid wood and ground wood for developing their models. It was realized that models built from ground wood samples did a poor job of predicting the chemical properties of solid wood samples. They however got better R^2 values when they used solid wood calibrations: extractives - 0.84, cellulose - 0.88 and lignin - 0.74.

Fardim et al. (2002) successfully utilized NIR to predict the chemical components of glucan, xylan, lignin and uronic acids; as well the the kappa number, viscosity, brightness and degree of polymerization of *Eucalyptus grandis Hill* ex Maiden kraft pulp obtained under different cooking parameters. Yet still, Derkyi et al. (2011) were able to determine the tannin content and Stiasny number of *Pinus caribaea* Morelet with the help of NIR and chemometric modeling. Other studies that investigated the ability of NIR to predict the chemical properties of lignocellulosic biomass include Li et al. (2015), Jones et al. (2006) and Hodge and Woodbridge (2004).

It is known that the extractives and lignin content have a positive linear relationship with the heating value of wood (Demirbas 2001; White 1987). When So and Eberhardt (2010) used reflectance NIR to predict the higher heating values of longleaf pine considering the effect of lignin content and extractives, they found out that the models predicted the HHV of unextracted wood samples better than they did extracted samples.

Not as much work has however been done on the use of NIR to predict the ash content of wood, and models generally give low coefficients of ash determination (Fagan et al. 2010; Allison et

al. 2008). Such poor results have been attributed to the fact that IR does not interact directly with the species that form ash, e.g. calcium, potassium and silica.

FTIR also has been used in quantitative and qualitative analysis of biomass. Via et al. (2013) built models for the HHV, volatile matter, fixed carbon and ash content of torrefied biomass using both NIR and ATR-FTIR spectra. They developed a novel 2D technique that made it possible to assign key NIR wavenumbers based on mid IR spectra. They were able to explain the depolymerization of key functional groups in hemicelluloses, cellulose and lignin by using a 2D correlation spectroscopy between NIR and FTIR. They noted however that NIR models performed better than FTIR models. Nuopponen et al. (2006) built calibration models of lignin, cellulose, extractives and density using fifty clones of *Picea sitchensis* (Bong.) Carr., twenty-four Ghanaian hardwoods and twenty *Pinus sylvestris* L. from DRIFT-MIR spectra of ground test samples. With RMSEP values of models constructed from only *Picea sitchensis* very similar to those computed from all the wood species, they concluded that a calibration model built from a particular species could be used to predict the similar properties of other wood species. Furthermore, they stated that when they reduced the number of wavelengths from 2386 to only the significant four or five (i.e. principal components), their RMSEP values were affected only slightly (lower for cellulose and density but higher for lignin); and thus, this advantage should be explored in the development of low-cost hand-held devices.

1.7 Thermogravimetric Analysis (TGA)

A typical thermogravimetric analyzer is basically made up of a balance, a furnace and its associated atmosphere controller, a data recorder and plotter (Price et al. 2000). TGA is a rapid type of thermal analysis that measures the change in mass as a function of temperature as a material is heated at a fixed rate under a set of conditions. The information provided by TG is inherently

quantitative in nature since the mass loss profile of a composite material or polymer usually equals the sum of the profiles of the individual components (Earnest 1988). The mass loss gives insight into a sample's chemical composition, thermal stability, number and sequence of reactions and kinetic parameters such as the order and activation energy of the chemical and physical reactions occurring (Klass 1998; Broido 1969). The activation energy is used to determine the amount of energy needed to bring about a chemical change, whereas the order of reaction and pre-exponential factor are needed in the calculation of the rate of reaction. Such knowledge is essential for the design and operation of thermochemical conversion units such as pyrolysis reactors and gasifiers (Parthasarathy and Narayanan 2013; Lee and Fasina 2009).

TGA has been a useful tool for determining the thermal decomposition behavior and kinetic parameters of lignocellulosic biomass. It has widely been used in the characterization of forestry residues (Lapuerta et al. 2004), softwoods and hardwoods (Grønli et al. 2002), corn stover (Kumar et al. 2008) and municipal solid waste (Becidan 2007). TGA also was utilized to study the degradation temperatures and kinetic parameters of several understory grasses found in a *Pinus palustris* Mill. ecosystem (Elder et al. 2011). Employing TGA together with differential scanning calorimeter (DSC), Owen et al. (2015) determined the rate, kinetics and energy involved in the thermal degradation of loblolly pine biomass in both air and nitrogen. In recent times, a couple of studies have explored the application of TGA in the quantitative (Carrier et al. 2011; Saldarriaga et al. 2015) and qualitative (Francisco-Fernández et al. 2012; Toscano et al. 2015) characterization of lignocellulosic biomass. Traditionally, researchers have determined the chemical composition of fuels by the deconvolution of derivative thermograms (DTGs), especially in quantitative analysis.

Systems integrating TGA with other analytical tools such as Fourier transform infrared spectroscopy (FTIR), gas chromatography (GC) and mass spectroscopy (MS) also have been used to

enable the identification and quantification of the composition and evolution rates of gaseous and liquid products during the pyrolysis and gasification of biomass feedstocks (Lee and Fasina 2009; Bassilakis et al. 2001; de Jong et al. 2003). Bodirlau and Teaca (2009) used FTIR to investigate the chemical reactions that occur during the esterification of wood and its polymers with organic anhydrides; and TGA to study the thermal stability of treated wood. FTIR spectra (of a reduced range of 1800 cm^{-1} – 1100 cm^{-1}) analysis showed that the O-H absorption band at 3456 cm^{-1} had reduced intensity, whereas the carbonyl absorption peak at 1735 cm^{-1} (C=O ester), C-H absorption at 1381 cm^{-1} (C-CH3) and C–O stretching band at 1242 cm^{-1} were enhanced. The latter results confirmed the formation of ester bonds. With the aid of a TGA, they were able to determine the initial temperature of decomposition, weight loss, activation energy and the reaction order to conclude that chemically modified wood was more thermally stable compared to untreated wood. Also, Kalisz et al. (2008) applied FTIR and TGA to characterize cotton, wood and ROFIRE (a wood/plastic composite) for pyrolysis and gasification. Using 3-dimensional plots of wavelength, absorbance and spectrum at a particular temperature and time in the reactor, they were able to follow the characteristic absorption at a specific wavelength of a gas evolving at a given time and temperature.

Furthermore, Dworzanski et al. (1991) developed a TG/FTIR/MS system interfaced with a GC/IRD/MS system to characterize *Liriodendron tulipifera* lignin and *Pseudotsuga menziesii* (Mirb.) Franco. They used a deactivated fused silica cap column to transfer TG effluents into the GC for gas separation and subsequent species identification by MS and/or FTIR. Their system enabled them to analyze selected spectra for univocal identification and chemical interpretation the thermal processes that occur during TG analysis.

Chapter 2 : Screening Elite Loblolly Pine Families for Structural Applications using Near Infrared Spectroscopy

2.1 Abstract

Near infrared (NIR)-based partial least squares (PLS) regression models were developed for the density, modulus of rupture (MOR) and modulus of elasticity (MOE) of elite loblolly pine families. These properties dictate wood quality for structural/dimensional applications. Moreover, the MOE and MOR are valuable traits in standing trees as they reduce mortalities caused by the failure of stems and uprooting of trees during inclement weather.

The mean densities of the families as predicted by NIR ranged from a low of 0.37 g/cm^3 (SD = 0.02) to a high of 0.5 g/cm^3 (SD = 0.07). Families did not necessarily perform the same on the two forest sites studied, with the interaction term being significant. The different densities noted between the individual families on a particular site were due to genetic variation (P-*value* < 0.0001). The MOR (i.e. strength) of the families studied was between 70 MPa to 116 MPa. MOR determined for the sixteen-year old elite families were comparable to what have been reported in the literature for older loblolly pine trees with a mean age of 22 years. A plot of density versus MOR showed that the former accounted for 41% of the variation noted in the MOR. With respect to the MOE (i.e. stiffness), ANOVA testing the effect of family, site and the interaction of family and site on the stiffness was only significant for family, with a *P-value* of < 0.0001. Stiffness values of the families ranged from 7782 MPa to 10946 MPa. Unlike for the strength, the relatively higher density of the elite families did not have as much impact on their stiffness.

As a means of further validating the robustness of NIR models developed in this study, the MOE of the loblolly pine families as estimated by NIR were compared to the MOE estimated from acoustics data. A One-way ANOVA gave a *P-value* of 0.45 at 0.05 significance level; an indication that there was no statistical difference between the MOE estimated for the families by the two nondestructive tools.

On the whole, four out of the fourteen screened families; A9, A1, A26 and A2 had high density, MOR and MOE values irrespective of site. As such, these should perform well in structural applications. On the flip side, using A33 and A21 in structural applications will not produce the optimum products due to their low density and strength properties. Moreover the low stiffness of these two families could impede their survival rate on site.

2.2 Introduction

As the wood basket of the nation, the southern United States accounts for 64% of the total timber harvested in the country (Smith et al. 2001). This produces 60% of the wood consumed nationally and contributes 18% to global wood supply (Wear and Greis 2002). *Pinus taeda* L. (loblolly pine) is an important timber species in the nation. The species dominates on approximately 13.4 million ha throughout the southeastern forests and accounts for over 50% of the standing pine volume of this region (Schultz 1997). Economically it generates some 110, 000 direct and indirect jobs and contributes approximately $30 billion to the economy of the region (Schultz 1999).

Unfortunately, Southern Pine Decline (SPD) (previously known as Loblolly Pine Decline or Pine Decline), which was first observed in the Talladega National Forest, Alabama back in 1959 has been associated with this species (Brown and McDowell 1968). Symptoms of SPD include a reduction of radial growth, thinning of foliage, deterioration of lateral roots and a heavy production of cones just before premature mortality. In a bid to control this disease complex, stakeholders would like to select and deploy elite loblolly pine families that are currently being screened for tolerance against root fungi associated with SPD.

Since the elite families are essentially a new feedstock, the southern pine industry is interested in knowing such important properties like the density, modulus of rupture (MOR) and modulus of elasticity (MOE) of this material. Furthermore, the stakeholders would like to incorporate this information back into tree breeding programs that aim to further improve wood quality.

The basic density of wood is defined as the ratio of its oven dry mass to its green volume. It is considered as the most important physical property of wood due to its effect on other wood attributes such as the strength properties, the yield of pulp per unit volume and the shrinkage and

swelling behavior of wood (Haygreen and Bowyer 1989). The porosity of wood (i.e. the ratio of cell wall material to lumen) is the main contributing factor to the basic density. Just like density, the strength properties of MOE and MOR dictate the quality of wood for structural applications. MOE (also known as the stiffness) is the resistance to deformation or distortion. It is the linear relationship between stress applied in bending and the resulting strain. The MOR (also known as ultimate strength) is the ultimate resistance of wood to applied loads. It is a measure of the maximum load carrying capacity of wood in bending (Green et al. 1999; Ritter 1990; Haygreen and Bowyer 1989). Apart from their importance in structural/dimensional timber, the MOE and MOR are valuable traits in standing trees. They reduce mortalities resulting from the failure of stems and uprooting of trees during inclement weather (Lachenbruach et al. 2011).

Currently, test methods used to determine the MOE, MOR and density require extensive sample preparation and destructive testing. With the large number of trees that have to be sampled in tree breeding programs, these conventional methods will not be practical and feasible. The industry would thus benefit from rapid, nondestructive and cost effective alternatives.

Near infrared spectroscopy (NIR) has evolved over the years as a reliable and rapid technique in the nondestructive assessment of wood and other forest products. A good number of studies have reported on the application of NIR to predict the density, MOE and MOR of wood. Kothiyal and Raturi (2011) related NIR spectra collected from the radial and tangential surfaces of five-year old *Eucalyptus tereticornis* Sm. to its specific gravity, MOE and MOR. They reported that PLS regression models from these two wood surfaces performed equally well in property prediction – with R^2 values ranging from 0.58 to 0.77. Cooper et al. (2011) have however pointed out that surface roughness of the material can affect the spectra. The researchers thus recommended that NIR spectra should be collected on similar surfaces that have ideally been planed. Furthermore,

they suggested that a large area should be illuminated especially when scanning is being done on the tangential surface so as to ensure a better representation of both earlywood and latewood. In an earlier study, Via et al. (2003) modeled the density, MOE and MOR of *Pinus palustris* mature wood, juvenile wood and pith wood. The authors attributed the poor performance of the developed models for predicting the strength properties of the pith wood to the narrow range of the MOE and MOR of the material used in their study. Another explanation they gave was that, the high concentration of extractives in the pith area although had a positive correlation with density, did not contribute to the strength. Several other researchers have employed NIR to estimate the density (Alves et al. 2011; Hein 2010; Gindl et al. 2001), as well as the MOE and MOR of solid wood and composite wood products (Adedipe and Dawson-Andoh 2008; Rails et al. 2002; Thumm and Meder 2001).

Even though a good number of studies have used NIR to estimate the density and strength properties of wood and wood-based products, the current study focuses on elite loblolly pine families, thus an essentially new resource. The objectives of this study were (i) to develop NIR-based PLS models to rapidly predict the density, MOE and MOR of loblolly pine families, and (ii) to screen out the elite loblolly pine families based on the understudied properties for the optimum structural applications.

2.3 Materials and Methods

2.3.1 Materials

Loblolly pine were acquired from two genetic research plantations established in 1998. Study Site 1 was located near Nahunta, Brantley County, Georgia (N 31° 12’16’’ W 81° 58’56’’) and Site 2 was located near Yulee, Nassau County, Florida (N 30° 63’ W 81° 57’). Each site was partitioned into fifteen blocks. Eighty trees representing eighty half-sib families were planted on

each block. Fifteen of these elite families were selected to be used; thus a total of 450 trees (i.e. fifteen families with one replication on each block per site) were earmarked for the study.

Firstly, cores were sampled at breast height from thirteen-year old trees during the spring and summer of 2011. Three hundred and fifty five tree cores were obtained because some trees were dead at time of sampling. The cores were stored in a walk in freezer (temperature: 4 ^{o}C) until time for further processing.

The second set of material comprised whole trees that were harvested from the selected families in 2014 and 2015. One tree per family was destructively sampled from each of the sites, thus a total of thirty trees. The DBH of trees ranged from 11.5 cm to 23.4 cm, with an average of 17.4 cm and standard deviation of 3.2 cm. Trees were crosscut into 1.5 m lengths along the bole, from which 50 cm bolts were cut from the basal end. Three disks representative of the butt, mid and top sections of each tree were selected to be further processed. For smaller diameter logs, two disks representing a section had to be processed. To hinder the rate of drying, the ends of boltswere coated with wax (Anchorseal Green Wood Sealer, U. C. Coatings Corp. Buffalo, NY, USA) in the field before they were transported to the laboratory. They were also stored in a walk in freezer (temperature: 4 ^{o}C) until time of processing.

The last set of material were nominal 2 x 4-in southern pine boards acquired from West Fraser Inc., a commercial sawmill located in Opelika AL. This is representative of material that is currently being processed and sold as lumber in the region.

Test samples that were processed from the harvested loblolly pine families and the commercial southern pine boards were used in NIR model calibration and validation.

2.3.2 Specimen Preparation and Determination of Wood Properties

Loblolly pine bolts and the commercial southern pine material were processed for three-point bending tests as specified in ASTM D143. Clear wood test specimens were cut to final dimensions of 2.5 x 2.5 x 41 cm and conditioned to an average moisture content (MC) of 9% in a control chamber (temp: 22 ^{o}C; relative humidity: 55%). Prior to destructive testing, the mass of a sample was measured and the dimensions obtained with calipers. This was used to compute the basic density as the ratio of the mass of a test specimen to its volume. Test samples were then loaded into a Zwick-Roell load frame equipped with 10KN load cell and a computer controlled screw-drive crosshead, and force applied at 1.3 mm/min on the tangential face. The span for the testing was 36 cm. The MOE (i.e. stiffness) was computed as the slope of the linear portion of the load-deflection curve. The MOR was calculated as

$$\text{MOR} = \frac{1.5\ \mathrm{P}*\mathrm{L}}{(\mathrm{b}*\mathrm{h}\text{^}2)}$$

Where: P is the ultimate load, L is the span, b is the width and h is the height of test specimen. Experiments were run in triplicates for density, MOE and MOR.

After destructive testing, each specimen was sawed into smaller blocks. Two of these cut pieces were used to determine the moisture content of the sample. The remaining materials were stored in airtight zip lock bags in a conditioning chamber until they were needed for further analysis.

2.3.3 Near Infrared Spectroscopy (NIR)

Cores obtained from the standing trees were dried in a conditioning chamber (temp: 22 ^{o}C; relative humidity: 55%) until the average moisture content (MC) was around 9%. The dried material was ground in a Wiley mill to pass an 80-mesh screen to be used for spectra collection. In addition, wood blocks saved after the destructive strength testing were first chipped using a chisel and

hammer, and then also ground to pass an 80-mesh screen. NIR spectra of test samples were collected on this material with a PerkinElmer Spectrum Model 400 NIR spectrometer (Waltham, MA, USA). The wavenumber range of the instrument is from 10000 cm^{-1} to 4000 cm^{-1}. A sample was scanned thirty-two times at a resolution of 4 cm^{-1} and averaged into one spectrum for analysis. Spectrum of a Spectralon standard was taken as the background reference sample every 20 minutes to correct for potential drifts with time.

2.3.4 Multivariate Data Analysis

PerkinElmer Spectrum Quant+ software (Waltham, MA, USA) was used to develop Partial Least Squares regression (PLS) models. The first derivatives of raw spectra were used for PLSmodel construction. The PLS procedure used the NIPALS algorithm to extract successive linear combinations of the predictors (i.e. NIR spectra) such that variations in both response and predictors were optimally explained. These linear combinations are known as factors, components or latent vectors. In extracting the first factor, If $X = X_0$, and $Y = Y_0$ are the centered matrix of the predictors and responses respectively, NIPALS starts with a linear combination $t = X_0w$ of the predictors where t is called a score vector and w is its associated weight vector. NIPALS predicts both X_0 and Y_0 by regressing them on t:

$X_0 = tp'$, where $p' = (t't)^{-1}t' X_0$; and $Y_0 = tc'$, where $c' = (t't)^{-1}t' Y_0$. The vectors p and c are known as the X- and Y-loadings respectively. The specific linear combination $t = X_0w$ is the one with maximum covariance t'u with some response linear combination $u = t = Y_0q$. Also, the X- and Y-weights w and q are proportional to the first left and right singular vectors of the covariance matrix $X'_0 Y_0$. The second factor is extracted in a similar way but X_0 and Y_0 are replaced with the X- and Y-residuals (called deflated X and Y blocks) from the first factor: $X_1 = X_0 - X_0$; $Y_1 = Y_0 - Y_0$. The process of extracting a score vector and deflating the data matrices is iterated for as many factors as are desired.

A total of two hundred and sixty samples were used in model calibration and validation. Samples were randomly assigned to either the calibration or test set. One hundred and ninety samples (i.e. 120 from the loblolly pine families and 70 from the southern pines stock) were used for model calibration and full cross-validation. The remaining 70 samples were used as an independent test set for external validation. The performance of validated models were evaluated using the standard error of calibration (SEC), standard error of cross-validation, standard error of prediction (SEP), coefficient of determination (R^2) and the residual predictive deviation / ratio of performance to deviation (RPD). Models that had the lowest error values were chosen and used to predict the basic density, MOE and MOR of the elite loblolly pine families.

In addition, the PROC GLM procedure in SAS (SAS Institute, Inc. Cary, NC, USA) was used in the analysis of variance (ANOVA) for the three properties of the different families. Tukey-HSD tests with alpha set to 0.05 were conducted when needed to further investigate pair-wise comparison between the treatments. All graphics and tables were produced with MS Excel (Microsoft Corp. Redmond, WA, USA).

2.4 Results and Discussion

2.4.1 NIR Model Calibration and Evaluation

Descriptive statistics of the density and strength properties of samples used in model training and validation are presented in Table 2.1. For the total sample set, the MOE ranged from 2380 MPa to 17300 MPa. The highest MOE measured for loblolly pine used in this study is however lower than what has been reported in the literature (So et al. 2002). Minimum and maximum values were respectively 25 MPa and 148 MPa for MOR and 0.37 g/cm^3 and 0.79 g/cm^3 for the density. Wide ranges observed in the dataset help to improve the robustness of models and in their predicting properties of future unknowns (Haartveit and Flæte 2006; Via 2003). Good overlaps were

noted between the means of the calibration and independent validation dataset for all three properties. The average MOE was 8923 MPa (SD = 2534 MPa) for the training set and 8643 MPa (SD = 3759 MPa) for the test set.

Table 2.1: Descriptive statistics of the basic density, MOE and MOR of southern pine wood.

	Property	Mean	SD	Min	Max	SE
	MOE (MPa)	8848	2909	2380	17300	180
Total set	MOR (MPa)	82	25	25	148	2
n = 260	Density (g/cm^3)	0.52	0.09	0.37	0.79	0.01
	MOE (MPa)	8923	2534	2380	15100	184
Training set	MOR (MPa)	82	21	26	132	2
n = 190	Density (g/cm^3)	0.52	0.08	0.37	0.75	0.01
	MOE (MPa)	8643	3759	2540	17300	449
Test set	MOR (MPa)	81	33	25	148	4
n = 70	Density (g/cm^3)	0.55	0.12	0.39	0.79	0.01
Loblloly pine	MOE (MPa)	8433	3128	2380	17300	180
families	MOR (MPa)	82	28	35	148	2
n = 180	Density (g/cm^3)	0.54	0.09	0.37	0.79	0.01
Commercial	MOE (MPa)	9782	2084	5780	14300	233
lumber	MOR (MPa)	81	15	41	112	2
n = 80	Density (g/cm^3)	0.49	0.06	0.39	0.63	0.01

Results obtained via the conventional methods were used as the response variables (i.e. Y-variables) whereas NIR spectra was used as the predictor variables (i.e. X-variables) in PLS modeling of properties. In Figure 2.1, a characteristic NIR spectra of pine wood is presented together with its 1st-derivatives.

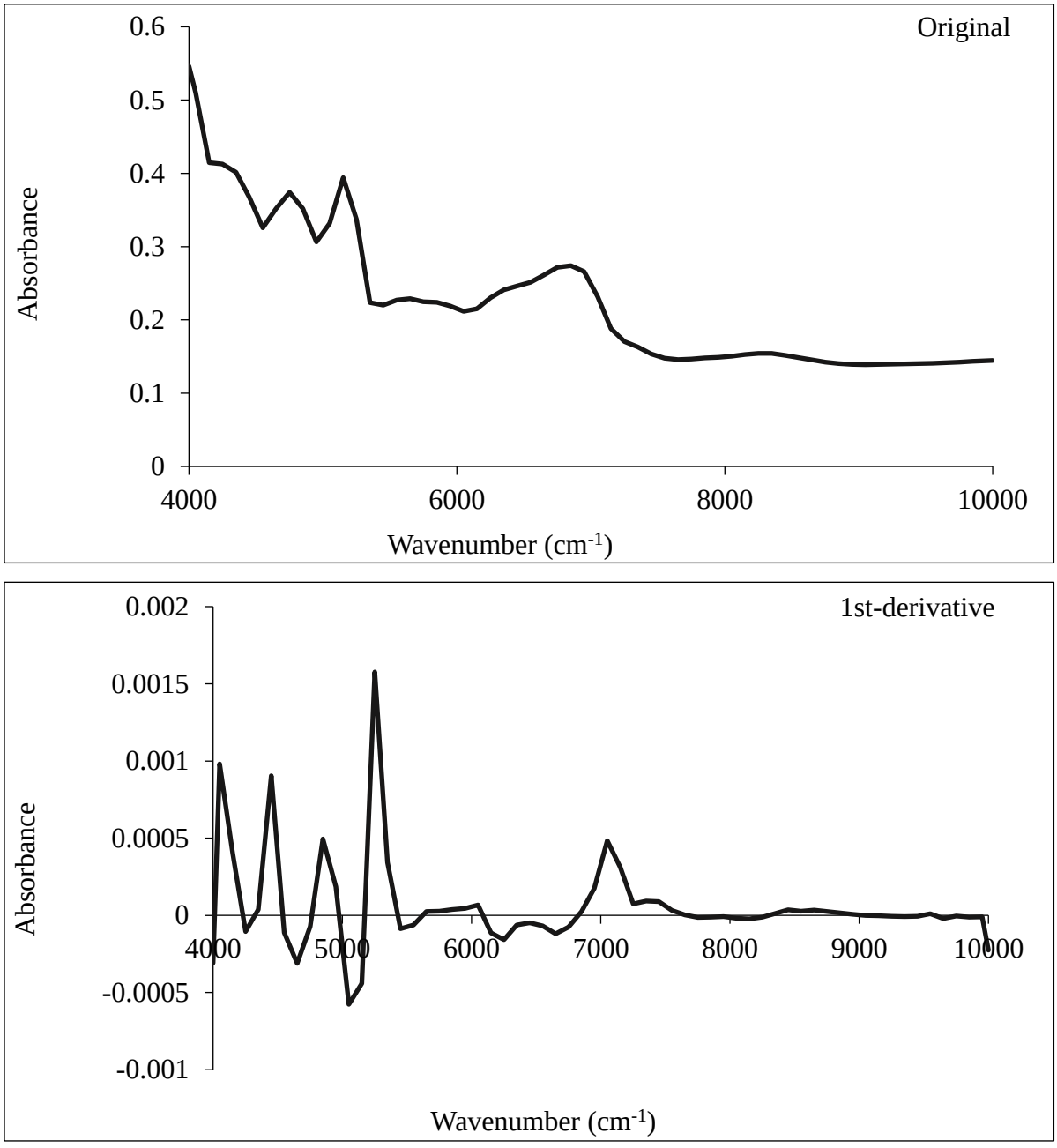

Figure 2.1: Characteristic NIR spectra of loblolly pine wood.

The absorption bands in NIR spectra arise when different organic functional groups absorb NIR light at different wavenumbers that equal the energy of their vibrations. As such, providing some insight about the chemical makeup of wood (or any other organic material). The most common absorption bands that occur in the near infrared region are overtones and combinations of fundamental vibrations of C-H, O-H, C=O, N-H, -COOH and aromatic C-H groups.

Table 2.2: Absorption bands associated with NIR spectra of wood.

Wavenumber (cm^{-1})	Bond Vibration	Chemistry Assignment
8368 - 8217	C-H str. 2^{nd} OT	Cellulose, methyl groups in lignin
7140 - 6250	O-H vib. 1^{st} OT	Cellulose, hemicelluloses
7092	O-H str	Phenolic hydroxyl groups in lignin and extractives
7027 - 6789	O-H str. 1^{st} OT	Amorphous or semi-crystalline regions of cellulose
6875	O-H str. 1^{st} OT	Phenolic groups in lignin
5596	C-H str. 1^{st} OT	Semi-crystalline or crystalline regions of cellulose
5291 - 4951	O-H str.	Interactions between hydroxyls in carbohydrates and water
5205	O-H asym. str. & def	Water
4765	O-H and C-H str. & def vib	Cellulose (and xylan)
4545	C-H and C=O str.	Lignin

Note: Str = stretching, OT = overtone, asym = asymmetric, def = deformation, vib = vibration.

For instance, absorbance peaks occur from 4000 cm^{-1} - 5000 cm^{-1} as combination bands due mostly to the interactions of C-H, O-H and N-H. Some peaks associated with specific wood components are presented in Table 2.2 (Alves et al. 2012; Kohan et al. 2012; Derkyi et al. 2011; Schwanninger et al. 2011; Kelley et al. 2004).

NIR is able to model non-chemistry secondary traits such as density and strength properties of wood because the chemical composition affects these traits. For instance, the fiber tracheid of wood, which is responsible for the strength of wood is made up of cellulose, hemicelluloses and lignin. The linear orientation and high degree of polymerization of cellulose makes it the primary contributor to strength; whereas lignin, in addition the binding the fibers together also serves as a stiffening agent. Hemicelluloses on the other hand acts as a matrix for cellulose as well as link the fibrous cellulose to the amorphous lignin. In the case of density, the ratio of cell wall material (i.e. cellulose, hemicelluloses and lignin) to lumen is a major determinant (Haygreen and Bowyer 1989; Winandy and Rowell 1984).

Results from PLS models that correlated 1^{st}–derivative NIR spectra to a single property are presented in Table 2.3. Four latent variables (LVs) were used in the construction of the optimum model for MOE and MOR, and three were used in developing the density model. The standard error of cross-validation (i.e. SECV), which estimates the errors that will be associated with a model's ability of predicting future unknown samples were 1267 MPa, 11 MPa and 0.04 g/cm^3 for MOE, MOE and density respectively. The coefficient of determination for the cross-validated models was least for density (i.e. $R^2 = 0.70$) and highest for MOE (i.e. $R^2 = 0.75$). The RPD, a measure of robustness was used to evaluate the predictive accuracy of cross-validated models. The relatively lower errors of PLS models compared to the standard deviation of the reference data produced models with good RPDs. For all the three properties, the RPD of models was greater than the 1.5

criteria required in order for a model to be considered as a preliminary screening tool (Hein et al. 2009). Schimleck et al. (2005) conducted a similar study using NIR to estimate the density, MOE and MOR of loblolly pine (mean age of 22) obtained from eighty-one plantations across the southern United States. The SEC and SECV obtained by the researchers were very similar to what was determined in this study (i.e. SEC = 0.03 g/cm^3, 1460 MPa and 8.6 MPa; and SECV = 0.03 g/cm^3, 1480 MPa and 9.5 MPa for density, MOE and MOR respectively). The authors however reported higher R^2 (0.82 – 0.89) and RPD values (2.3 – 2.6) for their cross-validated models.

Table 2.3: Calibration and prediction statistics of NIR-based PLS models.

	Density (g/cm^3)	MOR (MPa)	MOE (MPa)
Number of LVs	3	4	4
SEC	0.036	9.59	1100
SECV	0.042	11.33	1267
R^2_{cv}	0.7	0.71	0.75
RPD_{cv}	1.81	1.87	2
SEP_{iv}	0.065	19.4	2011
R^2_{iv}	0.19	0.41	0.45

Note: Subscript cv means cross-validation; iv means independent validation.

When the optimum cross-validated models were used in predicting the three properties for an independent test set (n = 70), the larger errors consequently reduced the coefficients of determination to 0.19 for density, 0.41 for MOR and 0.45 for MOE. The SEP values were 0.07 g/cm^3, 19 MPa and 2011 MPa for density, MOR and MOE respectively. Errors that are associated with the prediction of an independent dataset is usually larger since it factors in how much worse a model performs when applied to this test set not originally used in model training. In addition, this performance might have resulted from the diverse composition of the samples in the test set.

Regression plots relating the measured and NIR-predicted properties are shown in Figures 2.2, 2.3 and 2.4.

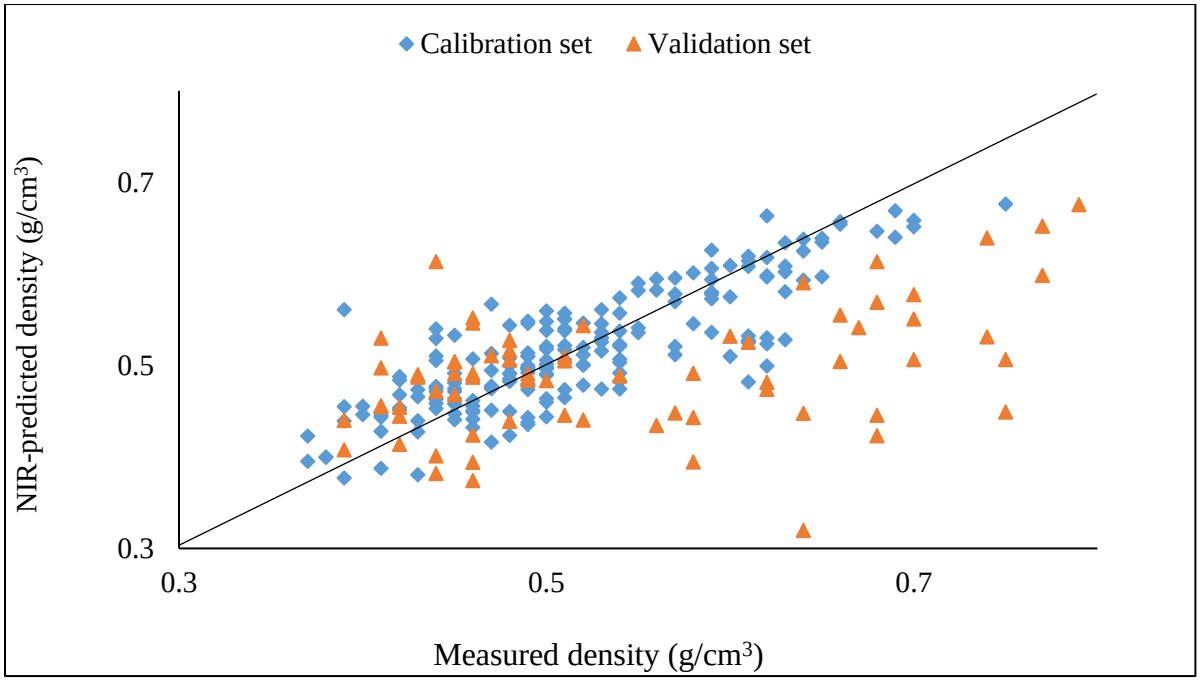

Figure 2.2: Relationship between measured and NIR–predicted basic density.

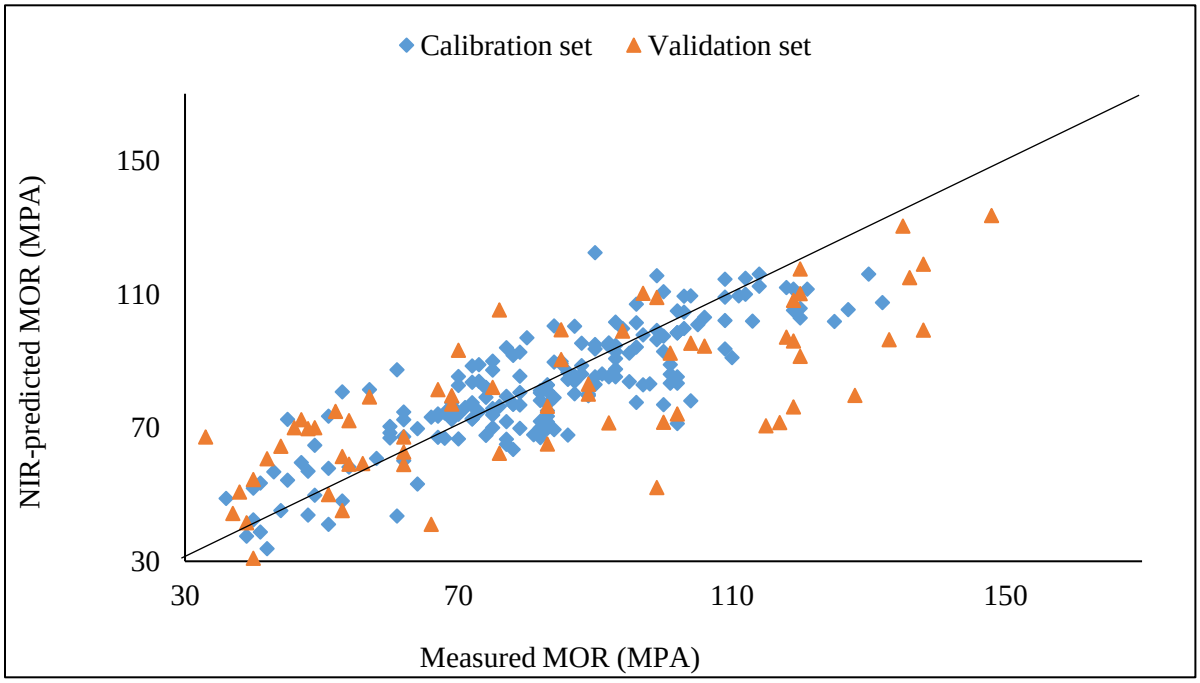

Figure 2.3: Relationship between measured and NIR–predicted MOR.

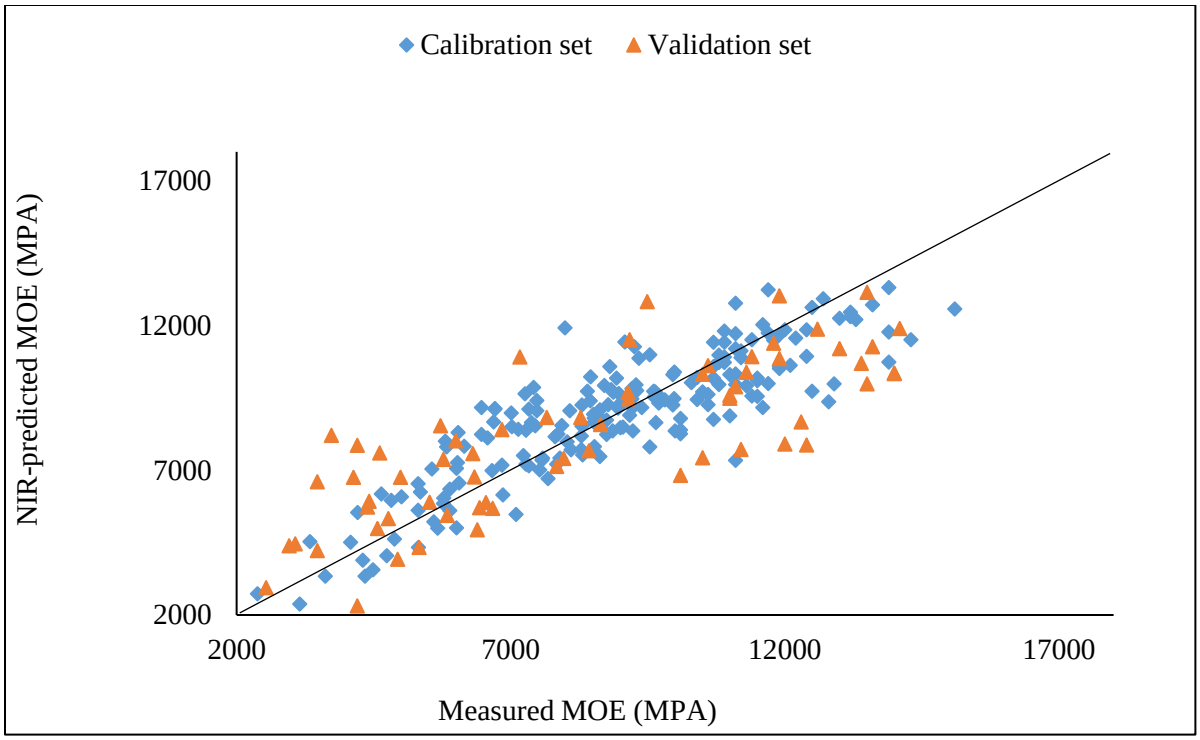

Figure 2.4: Relationship between measured and NIR–predicted MOE.

With the expected variations between the families, the predicted test samples were separated out into their respective families and a One-way ANOVA ($\alpha = 0.05$) conducted to test for equality of means between NIR-predicted and measured property. The results showed no statistically significant differences (i.e. *P-value* > 0.05) between NIR-predicted values and measured values for density, MOR and MOE for each of the fifteen elite families, Appendix A.

2.4.2 Prediction and Screening of Loblolly Pines Families

2.4.2.1 Density

Descriptive statistics for the density of the elite families on the two forest sites are presented in Table 2.4 and 2.5.

Table 2.4: NIR-estimated densities of loblolly pine families on Georgia site.

Family	Mean	SD	SE	COV	MIN	MAX	N
A1	0.46	0.06	0.02	12.85	0.36	0.61	15
A2	0.46	0.05	0.01	10.70	0.39	0.58	14
A5	0.40	0.05	0.01	11.67	0.36	0.51	13
A9	0.50	0.07	0.02	13.69	0.39	0.61	14
A10	0.44	0.07	0.02	15.48	0.36	0.56	14
A13	0.39	0.05	0.01	13.17	0.32	0.52	15
A15	0.45	0.07	0.02	15.54	0.37	0.59	12
A21	0.44	0.06	0.02	13.32	0.32	0.52	15
A26	0.49	0.07	0.02	13.38	0.39	0.62	15
A33	0.38	0.02	0.01	6.47	0.34	0.44	13
A34	0.44	0.08	0.02	18.47	0.34	0.58	14
A37	0.41	0.05	0.01	12.86	0.34	0.53	15
F17	0.43	0.05	0.01	11.86	0.36	0.54	13
F23	0.45	0.05	0.01	11.01	0.36	0.54	14

Table 2.5: NIR-estimated densities of loblolly pine families on Florida site.

Family	Mean	SD	SE	COV	MIN	MAX	N
A1	0.49	0.06	0.02	12.11	0.39	0.58	11
A2	0.46	0.04	0.01	8.40	0.40	0.52	10
A5	0.43	0.02	0.01	4.95	0.40	0.47	11
A9	0.47	0.03	0.01	6.87	0.42	0.51	10
A10	0.46	0.04	0.01	8.43	0.39	0.51	12
A13	0.37	0.03	0.01	7.19	0.31	0.40	7
A15	0.42	0.02	0.01	5.43	0.39	0.47	8
A21	0.39	0.03	0.01	6.87	0.36	0.43	12
A26	0.45	0.03	0.01	6.21	0.40	0.50	12
A33	0.40	0.03	0.01	7.04	0.35	0.47	12
A34	0.44	0.04	0.01	8.49	0.36	0.50	10
A37	0.41	0.01	0.00	3.25	0.39	0.43	12
F17	0.40	0.04	0.01	9.05	0.36	0.47	12
F23	0.39	0.03	0.01	8.38	0.34	0.44	11

The mean densities of the families ranged from a low of 0.38 g/cm^3 (SD = 0.02 g/cm^3) to a high of 0.50 g/cm^3 (SD = 0.07 g/cm^3) on the Georgia site. For the Florida site, the range was from 0.37 g/cm^3 (SD = 0.03 g/cm^3) to 0.49 g/cm^3 (SD = 0.06 g/cm^3). The range of density predicted by NIR is comparable to the 0.44 g/cm^3 to 0.51 g/cm^3 reported for 10-year old loblolly pine trees by Belonger et al. (1997) using X-ray densitometry; as well as the 0.39 g/cm^3 to 0.56 g/cm^3 determined for 12-year old loblolly pine trees by Jones et al. (2008). A Two-way ANOVA ($\alpha = 0.05$) showed that

the families differed significantly in their densities on the two sites. In addition, the interaction term had a *P-value* of less than 0.05, Table 2.6.

Table 2.6*:* ANOVA F-Test values and *P-values* of density by treatment.

Treatment	df	F-value	*P-value*
Family	13	9.24	< 0.0001
Site	1	4.79	0.0294
Family x Site	13	2.34	0.0055

From Figure 2.5, families A9 and A26 had the highest densities on the Georgia site. These were statistically higher than the densities of families A37, A5, A13 and A33. On the Florida site, families with the highest densities were A1 > A9> A10> A2 > A26, Figure 2.6. The densities of A1, A9, A10 and A2 were statistically higher than the densities predicted for A33, F17, A21, F23 and A13.

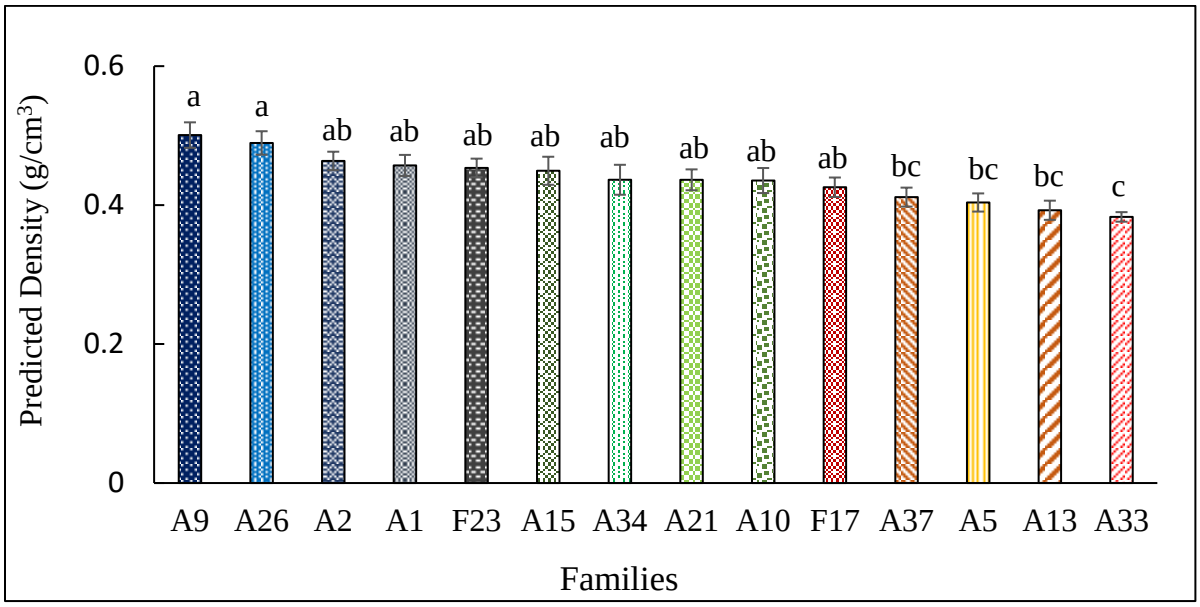

Figure 2.5: Rank of loblolly pine families for density on Georgia site. *Bars with different letters are significantly different at 95% confidence level (Tukey's HSD Test).

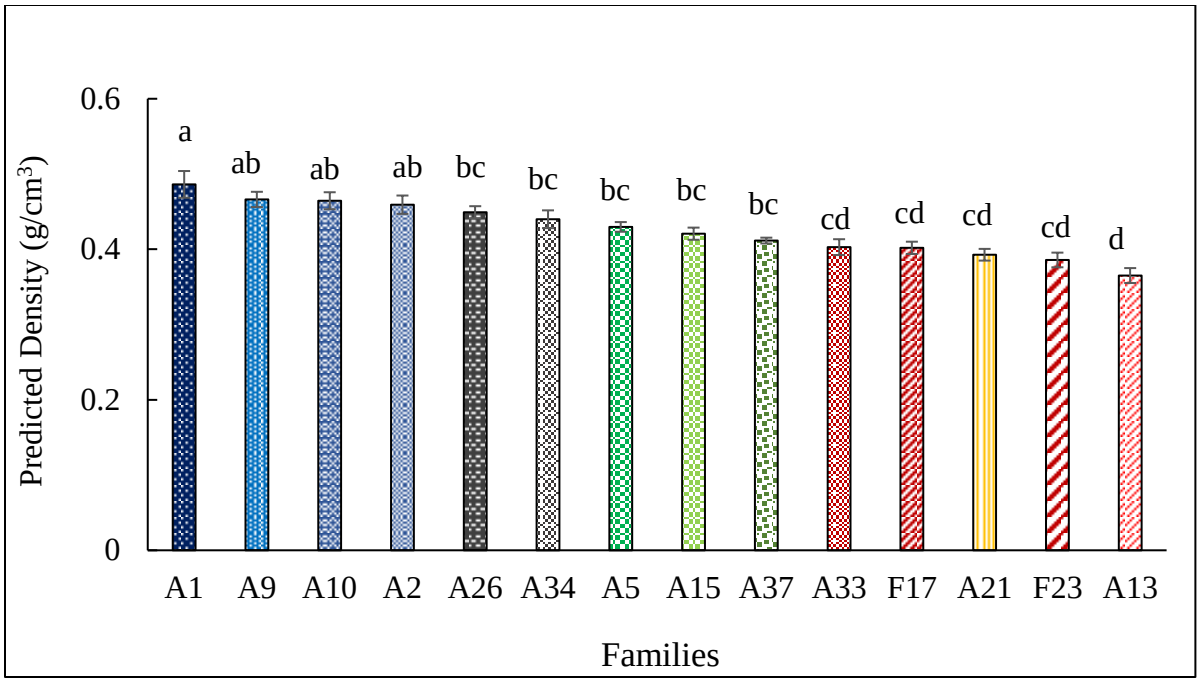

Figure 2.6: Rank of loblolly pine families for density on Florida site. *Bars with different letters are significantly different at 95% confidence level (Tukey's HSD Test).

The elite loblolly pine families did not necessarily rank the same on both sites as expected, with the interaction term being significant, Table 2.6. The most extreme case was for F23 whose density was among the highest on the Georgia site but among the lowest on the Florida site.

The density of wood has been shown to vary both within a single tree and among trees of a species. Within a tree, the ratio of cell wall material to lumen is known to be a major contributor to density. This overarching trait is due to several factors such as the ratio of earlywood to latewood, cell size and wall thickness, and the percentage of the various cell types such as fiber tracheids, ray cells and vessel elements. Furthermore, the occurrence of extractives can have a pronounced impact on density (Zobel and van Buijtenen 1989). Among softwoods, the density generally decreases further up the tree and increases outward from the pith towards the bark (Zobel and Sprague 1998). This is because wood produced by a young tree within the first five to twenty-five

years (i.e. juvenile wood or core wood) is usually inferior to wood that will be formed in later years (i.e. mature wood). For instance juvenile wood cells are shorter and are thin walled. Furthermore, juvenile wood has a lower proportion of latewood to earlywood. All these factors contribute to the lower density of wood found in the core as compared to the outerwood. Within a species, density has been reported to vary based on such factors as site conditions and genetic sources (Haygreen and Bowyer 1989). As Zobel and van Buijtenen (1989) pointed out, the properties of wood are a result of the interaction between the genetic potential of the tree with its growing environment.

The different densities noted between the individual families on a particular site are due to the genetic variation (*P-value* < 0.0001), Table 2.6. On the other hand, differences noted within an individual family on the two forest site are a function of the environment, as also observed by Jordan et al. (2008). According to the authors, the differences in wood properties is probably due to the duration of juvenile wood production which increases as one moves north-west from the South Atlantic region. Also, the occurrence of moisture, especially in the summer promotes the production of latewood, thus higher density. This was however not the case for all the families investigated in this study.

Even though there was a site effect on density and most families didn't rank the same on the two site, some families including A9, A1, A26 and A2 did consistently well on both sites. On the flip side, the densities of families A13, A33, F17 and A37 were low on the different sites.

2.4.2.2 Modulus of Rupture (Bending Strength)

The range of predicted MOR values for the elite families were narrower on the Georgia site (i.e. 47 MPa to 143 MPa), Table 2.7, compared to the Florida site (i.e. 34 MPa to 150 MPa), Table 2.8. A9 had the highest MOR (Mean = 109 MPa; SD = 28) on Georgia site, whereas A1 was the strongest on Florida site (Mean = 116 MPa; SD = 28 MPa). Conversely, the lowest MOR were recorded for A33 on the Georgia site, and for A21 on the Florida site.

Table 2.7: NIR-estimated MOR of loblolly pine families on Georgia site.

Family	Mean	SD	SE	COV	MIN	MAX	N
A1	102.7	20.6	5.3	20.1	54.7	134.1	15
A2	99.4	10.5	2.8	10.6	82.3	117.5	14
A5	91.9	32.4	9.0	35.3	48.0	143.3	13
A9	108.6	28.4	7.6	26.2	52.2	139.9	14
A10	96.0	30.0	8.0	31.3	47.7	143.4	14
A13	92.4	18.8	5.0	20.4	61.8	115.6	15
A15	95.9	16.4	4.7	17.1	73.7	128.1	12
A21	88.3	20.7	5.4	23.5	47.0	110.6	15
A26	98.2	18.7	4.8	19.0	59.7	117.7	15
A33	69.6	12.7	3.5	18.3	52.0	100.0	13
A34	83.7	21.7	5.8	25.9	59.0	124.5	14
A37	83.0	19.6	5.1	23.7	59.4	115.7	15
F17	94.7	17.3	4.8	18.2	70.8	128.1	13
F23	101.0	14.7	3.9	14.5	70.8	117.5	14

Table 2.8: NIR-estimated MOR of loblolly pine families on Florida site.

Family	Mean	SD	SE	COV	MIN	MAX	N
A1	116.1	27.7	8.4	23.9	68.4	159.3	11
A2	109.1	16.9	5.3	15.5	72.3	126.5	10
A5	86.9	10.2	3.1	11.8	71.2	100.3	11
A9	105.6	10.2	3.2	9.7	94.2	122.9	10
A10	103.1	19.1	5.5	18.5	72.3	136.9	12
A13	77.2	16.8	6.4	21.8	50.5	96.8	7
A15	92.6	27.4	9.7	29.5	43.5	122.7	8
A21	76.9	20.4	5.9	26.5	34.1	103.5	12
A26	100.7	21.2	6.1	21.1	58.5	123.3	12
A33	80.2	14.4	4.1	17.9	46.6	97.1	12
A34	104.4	20.9	6.6	20.0	64.2	124.7	10
A37	110.5	12.1	3.5	10.9	86.5	127.2	12
F17	92.1	18.4	5.3	20.0	66.7	120.5	12
F23	78.5	15.6	4.7	19.8	49.3	95.0	11

According to the ANOVA results in Table 2.9, there were significant differences in the MOR of families tested. Furthermore, the family x site interaction was significant. However, the effect of site was not statistically significant.

Table 2.9: ANOVA F-Test values and *P-values* of MOR by treatment.

Treatment	df	F-value	*P-value*
Family	13	5.7	< 0.0001
Site	1	0.79	0.3747
Family x Site	13	2.94	0.0005

On the Georgia site, A9, A1, F23, A2 and A26 had MOR values that were statistically higher than the MOR of A33. Meanwhile on the Florida site, A1, A37 and A2 are significantly stronger than A33, F23, A13 and A21. On the whole, the MOR of A1, A2, A9 and A26 ranked high on the two sites, whereas A33, A21 and A5 were among the least strong on both sites, Figure 2.7, Figure 2.8.

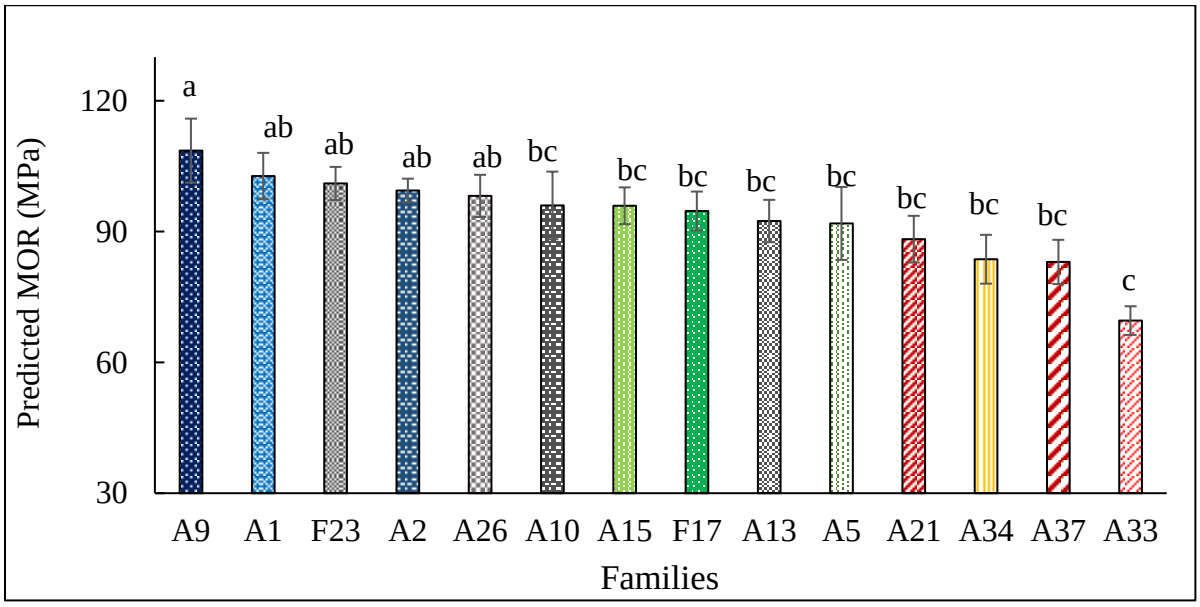

Figure 2.7: Rank of loblolly pine families for MOR on Georgia site. *Bars with different letters are significantly different at 95% confidence level (Tukey's HSD Test).

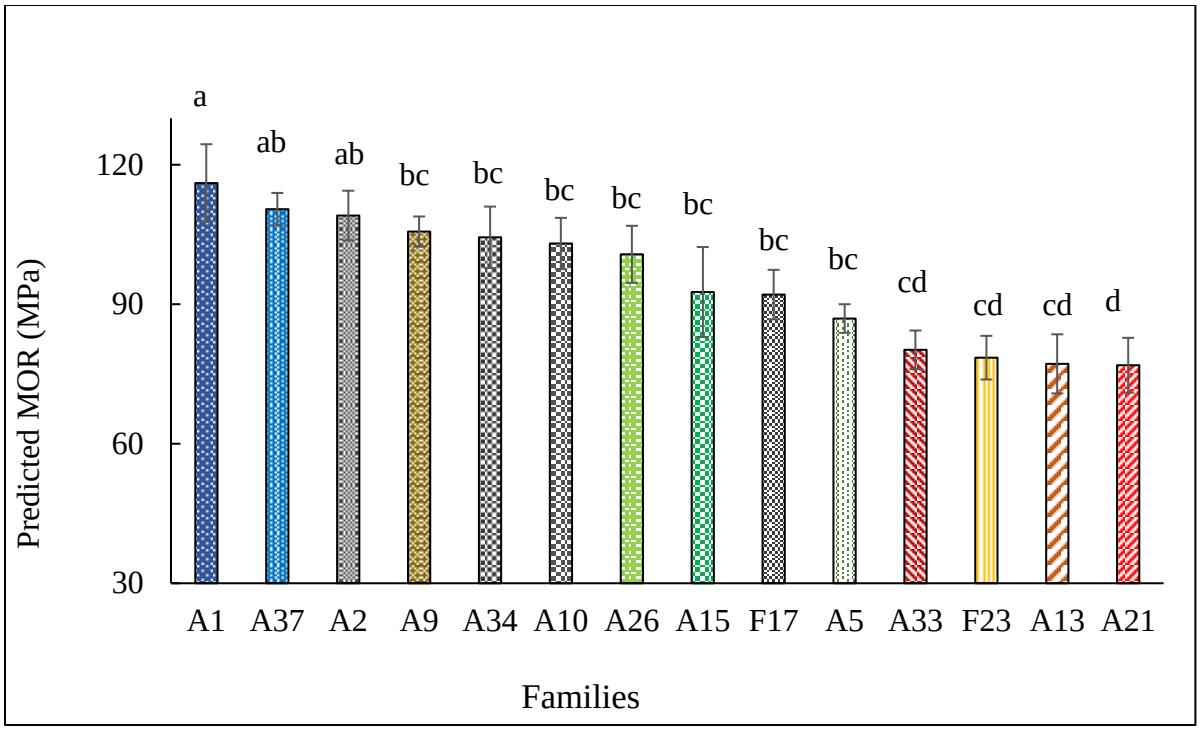

Figure 2.8: Rank of loblolly pine families for MOR on Florida site. *Bars with different letters are significantly different at 95% confidence level (Tukey's HSD Test).

Reviewing the results, rankings of families based on MOR seem to generally follow the ranks determined for the density. For instance, on the Georgia site, A9 had the highest density and MOR, with A33 having the lowest values for the two properties. A plot of density versus MOR showed that, the former was able to explain 41% of the variation noted in MOR, Figure 2.9. This result is similar to what has been reported in the literature about the correlation of density with MOR (Hein et al. 2013).

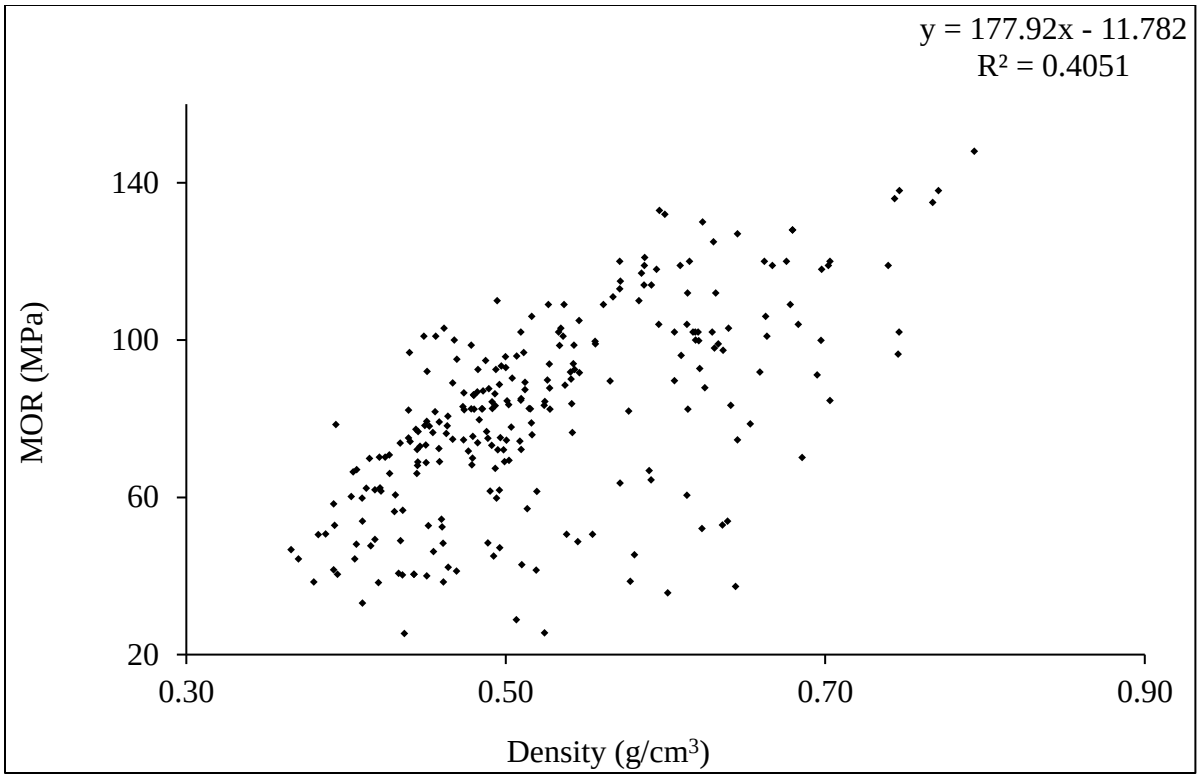

Figure 2.9: Correlation between the density and MOR of wood.

Mean MOR values determined for the elite loblolly pine families of this study were comparable to what has been reported by Schimleck et al. (2005) for older loblolly pine trees with a mean age of 22 years. An explanation for the higher MOR for these relatively younger trees might be due to their relatively higher density. MOR increases with increasing density because more stress is required to cause the failure of more material. In Table 2.1, it was noted that the mean density of the families was slightly higher (0.54 g/cm^3) than the mean density of the lumber (0.49 g/cm^3) acquired from the commercial sawmill. Pearson and Ross (1984) reported similar results whereby the density to 15-year old loblolly pine was slightly higher (0.52 g/cm^3) than that of 25-year old trees (0.5 g/cm^3). The saturation of wood near the pith with extractives has been known to increase the density of the juvenile wood without improving the strength (Via 2003; Larson et al. 2001). However, since higher density of the pine families corresponded to an increase in MOR, the

increased density may be attributed to the cell characteristics rather than extraneous depositions. Further studies of the anatomical features of the elite families will however be required in ascertaining this occurence.

2.4.2.3 Modulus of Elasticity (Stiffness)

ANOVA testing the effect of family, site and the interaction of family x site on the stiffness was only significant for family, Table 2.10. As such, replicates of families from the two sites were pooled together for further analysis.

Table 2.10: ANOVA F-Test values and P-*values* of MOE by treatment.

Treatment	df	F-value	*P-value*
Family	13	3.45	<0.0001
Site	1	0.34	0.5625
Family x site	13	0.93	0.5198

The descriptive statistics for the MOE of pine families are presented in Table 2.11. The range of MOE values predicted was from 2981 MPa to 15830 MPa. The MOE for the elite families was highest for A9 (Mean = 10946 MPa; SD = 2703) and lowest for A33 (Mean = 7782 MPa; SD = 2237).

Table 2.11: NIR-estimated MOE of loblolly pine families.

Family	Mean	SD	SE	COV	MIN	MAX	N
A1	10801	2424	475	22	4984	14200	26
A2	10453	1713	350	16	5999	14170	24
A5	9369	3092	631	33	4741	14890	24
A9	10946	2703	552	25	5329	15210	24
A10	10473	2894	567	28	5904	15830	26
A13	9561	2509	547	26	5321	13950	21
A15	10023	2396	536	24	4006	13200	20
A21	8909	2565	494	29	2981	12070	27
A26	10560	2247	441	21	5874	13570	27
A33	7782	2237	439	29	3791	12970	26
A34	9759	2688	549	28	5145	13460	24
A37	9285	2145	413	23	5259	12110	27
F17	10258	2217	435	22	6675	13640	26
F23	10805	1775	355	16	6813	13280	25

NIR-predicted MOE of loblolly pine in this study was comparable to the 2500 MPa - 15327 MPa reported by Kelley et al. (2004). However, the maximum MOE for the elite families were lower than the 23 000 MPa with a Silviscan for loblolly pine aged between twenty-one to twenty-six years (Jones et al. 2005). The upward shift of MOE with age is due to the fact that the proportion of juvenile wood (which has thin-walled shorter cells, as well as a higher proportion of earlywood to latewood) is inversely related to age. For instance, Zobel and Blair (1976) stated that 15-year old

loblolly pine has about 85% of its wood as juvenile wood but this proportion decreased drastically to 19% by age 40. Apart from the presence of juvenile wood adversely affecting the density, it also has large microfibril angles and excessive spiral grain.

Compared to the MOR, density accounted for less of the variation in MOE ($R^2 = 0.2$), Figure 2.10. Similar results were reported by Burdon et al. (2001) when they investigated the relation of density to the MOR and MOE of *Pinus radiata* D Don. Several studies have shown that the microfibril angle (MFA) has a greater influence on the MOE than density (Ivkovic et al. 2009; Evans and Ilic 2001; Tsehaye et al. 1998). Also, it has been reported that the MFA correlates to the MOR to a lesser extent than it does to MOE (Hein et al. 2013). As in the case of density, the MOE of loblolly pine increases from the pith towards the bark, but decrease as one moves further up the tree (Zobel and Sprague 1998).

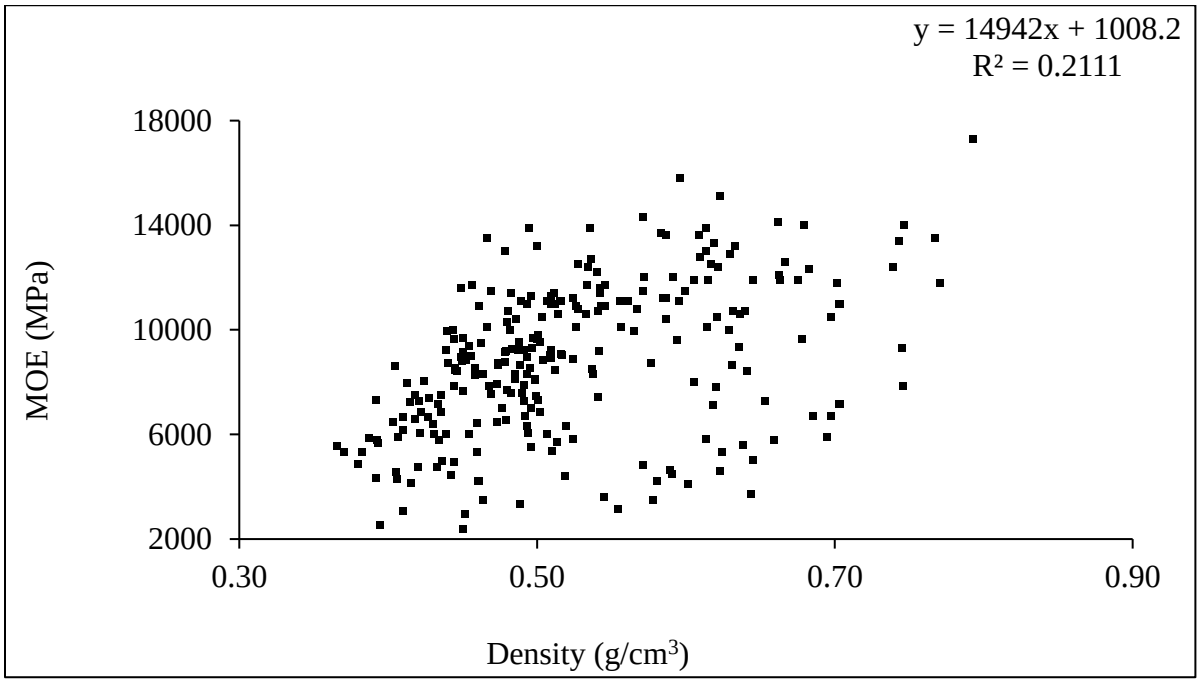

Figure 2.10: Correlation between the density and MOE of wood.

Thus, the MFA of the families, which is under moderate to strong genetic control, might have mainly contributed to the differences noted in the MOE of elite families. In terms of ranking, the MOE of families A9, F23, A1, A26, A10, A2 and F17 were all statistically different from the 7782 MPa predicted for A33. Figure 2.11.

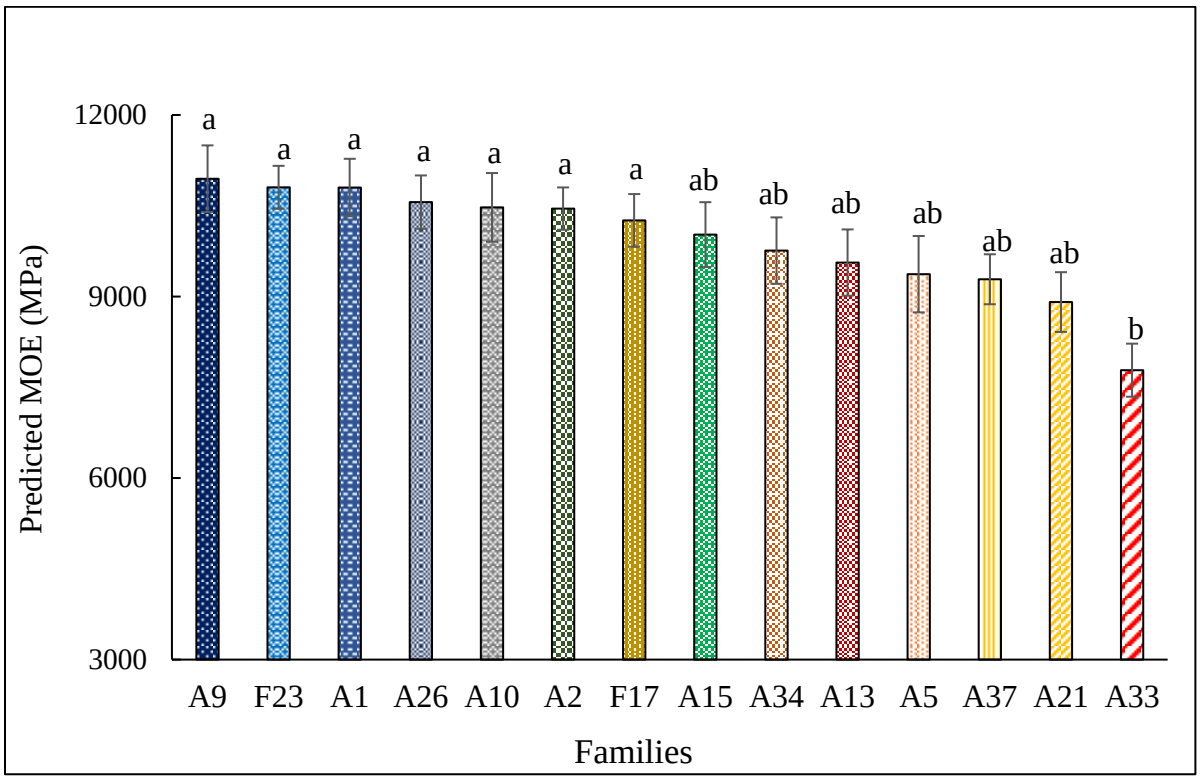

Figure 2.11: Rank of loblolly pine families for MOE. *Bars with different letters are significantly different at 95% confidence level (Tukey's HSD Test).

2.4.3 Validation of NIR- predicted MOE using Acoustics-predicted MOE

To further corroborate the robustness of the developed NIR-based models, the acoustic-MOE (MOE_{ac}) of the elite loblolly pine families were computed from their tree velocities and densities. Acoustic data was collected with a FAKOPP Microsecond Timer (Fakopp Enterprise, Agfalva, Hungary) which operates based on the Time-of-Flight (ToF) principle. For data collection, a transmitter and receiver probe were positioned 120 cm apart on the same side of a side of the tree,

ensuring that the midway of these probes is at breast height. A stress wave was then generated with the strike of a steel hammer at a steady force (Essien et al. 2016; Wang et al. 2007). The tree velocity (V_T) was computed as the ratio of the distance (m) between the probes (in m) and the time (s) it takes the stress wave to travel from the transmitter to the receiver. The MOE_{ac} was calculated as:

$MOE_{ac} = (V_T)^2$ x density.

With an F-value of 0.58 and *P-value* of 0.45 ($\alpha = 0.05$), results of the One-way ANOVA comparing the two tools showed no statistical difference between NIR-predicted MOE and acoustics-predicted MOE of loblolly pine families from the two forest sites. However, for better insight, the NIR-predicted versus acoustics-predicted was plotted for each of the families separately for the two sites. From Figures 2.12 and 2.13, the MOE of the families were estimated more or less equally by the two nondestructive techniques. However, NIR consistently underestimated the stiffness of A33. It has been determined earlier on that this family was among those with low densities. As such, the MFA might be a bigger contributor to the MOE (Ivkovic et al. 2009; Evans and Ilic 2001; Tsehaye et al. 1998) of this particular family than density. However, as already stated, further studies of the anatomical features of the elite families will be required to test this hypothesis. The underestimation of the MOE of this family on both study sites by NIR might be an indication of the limitation of the tool in being less sensitive to the MFA.

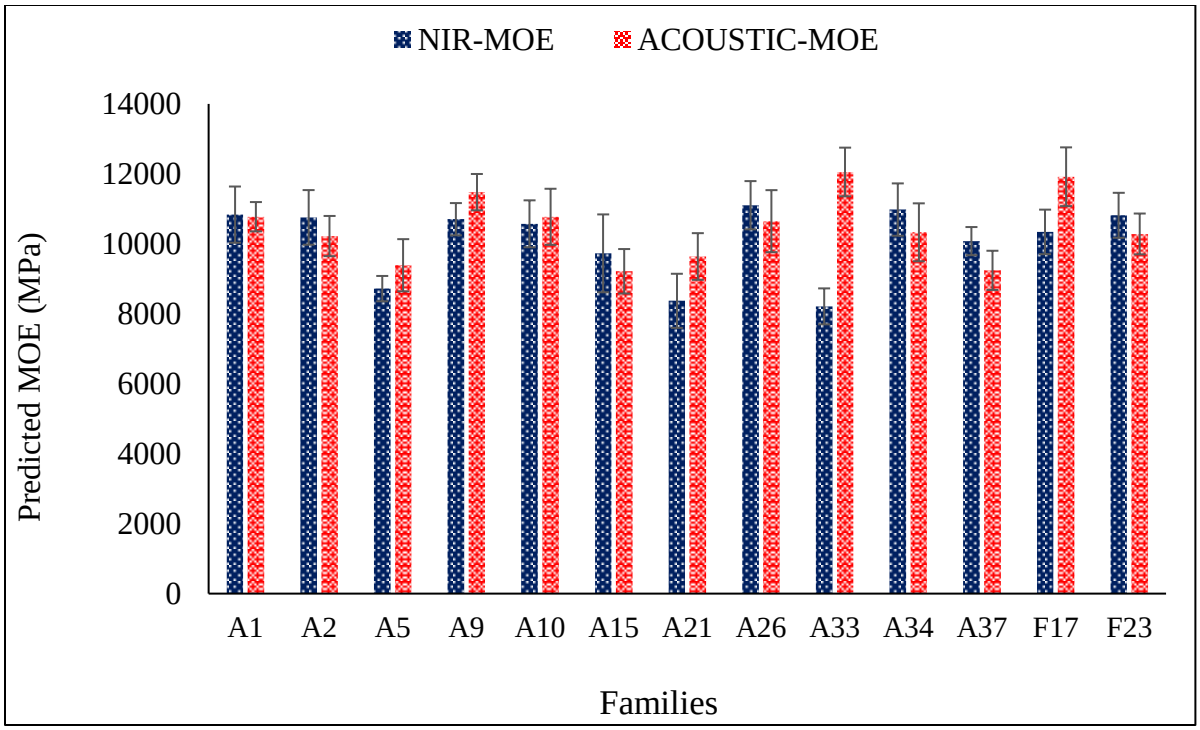

Figure 2.12: MOE of the loblolly pine families as predicted by NIR and Acoustics on the Florida site.

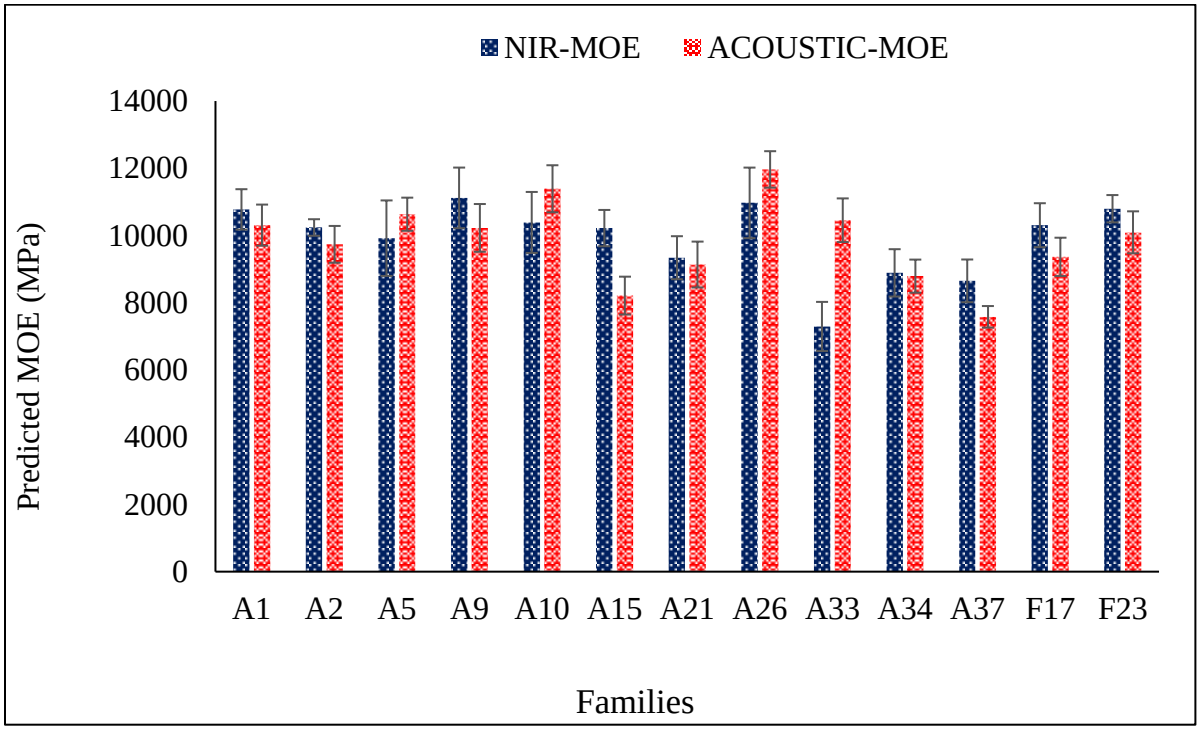

Figure 2.13: MOE of the loblolly pine families as predicted by NIR and Acoustics on the Georgia site.

2.4 Conclusions

NIR-based PLS-models were developed for the density, modulus of rupture (MOR) and modulus of elasticity (MOE). Models were then used to partition out fourteen elite loblolly pine families based on the aforementioned properties that dictate wood quality in structural applications.

Cross-validated models calibrated with 1^{st}-derivative treated NIR spectra had R^2 values ranging from a low of 0.7 for density to a high of 0.75 for MOE. Constructed models for the three properties all had their RPD values greater than 1.5; as such, they could be employed in the preliminary screening of the loblolly pine families.

The mean densities of the families as predicted by NIR was from 0.37 g/cm^3 (SD = 0.02) to 0.5 g/cm^3 (SD = 0.07). A Two-way ANOVA showed an effect of family, site and family x site interaction on density. Density variation in wood is known to be controlled by genetic factors such as the ratio of earlywood to latewood, cell size and wall thickness, and the length of juvenile wood production. However, these traits can be influenced by site conditions (i.e. the environment). In spite of the effect of site on density, families A9, A1, A26 and A2 had consistently high densities on the two study sites. Conversely, the densities of A13, A33, F17 and A37 were low on the two sites.

With respect to the MOR, NIR-predicted means for the families ranged from a low of 70 MPa for A33 on the Georgia site, to a high of 116 MPa for A1 on the Florida site. The treatments of family and family by site interaction had significant effects on MOR. MOR values of the loblolly pine families were noted to be comparable to what have been reported in the literature for older loblolly pine trees; suggesting that the elite families have improved strength. Within the scope of the current study, the relatively high MOR of the families could be attributed to the density. Further studies will however be needed at the anatomical level to determine other causes of this apparent

improvement in MOR since density accounted for only for 41 % of the variation in MOR. On the whole, the MOR of A1, A2, A9 and A26 ranked high on the two sites, while A33, A21 and A5 were among the least strong on both sites.

Analysis of variance testing the effect of family, site and the interaction of family x site on the stiffness was only significant for family, which had a *P-value* of < 0.0001. A9 had the highest stiffness (Mean = 10946 MPa; SD = 2703 MPa) while A33 had the lowest (Mean = 7782 MPa; SD = 2237). Unlike for the strength, the relatively higher density of the loblolly pine families did not impact the stiffness. In fact, regression analysis showed density could explain an even lesser proportion of the variation in the MOE (i.e. 20 %) compared to the MOR. As such, other factors such as the MFA and slope of grain would have to make up for the difference. A study of the anatomical features of the elite families will help clarify this. For the family ranking, the MOE of A9, F23, A1, A26, A10, A2 and F17 were all statistically different from that of A33.

In addition, NIR-predicted MOE was compared to acoustics-predicted MOE for the elite families. A One-way ANOVA gave a P-*value* of 0.45 at 0.05 significance level; an indication that there was no statistical difference between the MOE predicted for the loblolly pine families by the two nondestructive tools. This results further corroborated the robustness of NIR-based models developed in this study.

Finally, families A9 and A1, as well as A26 and A2 will perform better in structural application because they had higher density, MOR and MOE irrespective of site. On the other hand, using A33 and A21 in structural applications might not produce the optimum products due to their low density and strength properties. Apart from that, their low stiffness could impede their survival on site due to inclement weather such as storms.

Chapter 3 : Screening Elite Loblolly Pine Families for Chemical and Bioenergy Applications using Near Infrared Spectroscopy

3.1 Abstract

Near infrared (NIR) spectroscopy coupled with partial least squares (PLS) regression was used to rapidly estimate the chemical composition and bioenergy potential of select loblolly pine families. The chemical composition of wood impacts other wood properties, as well as the processing and quality of final products that will be derived; whether it be pulp and paper, bio-composites, lumber or biofuel. Also, knowledge about the bioenergy potential of this essentially new feedstock will be valuable in the emerging bioeconomy.

PLS models were developed using the first derivatives of NIR spectra as the independent variables (i.e. X-variables) and data obtained via conventional laboratory methods as the dependent variables (i.e. Y-variables). One hundred and ninety samples were used in model calibration and full cross-validation. An additional test set made up of 60 samples were used for external validation. Validated models that had the lowest error values were used to predict the extractives, lignin, cellulose, glucose, hemicelluloses, volatile matter, fixed carbon, ash and energy contents of elite loblolly pine families obtained from two forest sites.

R^2 values for cross-validated models ranged from 0.84 to 0.73 for the chemical traits; and from 0.88 to 0.58 for the bioenergy related properties. Models for all the properties studied had RPD values greater than 1.5, as such, they could be employed in the preliminary screening of the elite loblolly pine families. Property means of the families as predicted by NIR-based PLS models were as follows: extractives – 5.5% (SD = 1.1%), lignin – 30.7% (SD = 1.3%), glucose – 44.7%

(SD = 1.7%), hemicelluloses – 20.5% (SD = 0.4%), volatile matter – 84.2% (SD = 0.8%), fixed carbon – 14.9% (SD = 0.8%), ash – 0.2% (SD = 0.03%) and HHV 19.0 MJ/kg (SD = 0.2 MJ/kg).

The genotype of loblolly pine families affected the chemical, proximate and energy traits studied. The genetic variation detected for cellulose was the largest (*P-value* < 0.0001). Considering that the cellulose content has strong correlations with other properties, selecting and breeding for cellulose can generate gains. However, the family by site interaction was significant for all understudied properties, indicating the general instability of the elite families across different sites. Nonetheless, the cellulose contents of families A1, A26, A15, A2 and A9 were consistently higher on the two sites. High cellulose content implies that these families will be good choices for pulp and paper applications, or for cellulosic ethanol production. On the other hand, the relatively high volatile matter contents of F23, F17, A37 and A9 in spite of the significant interaction between genotype and environment make them good candidates for the production of bio-oil and syngas via thermochemical conversion processes.

Knowledge of the chemistry and bioenergy potential of these elite families will aid in decision making processes that optimize the suitability of this feedstock to support the conventional forest products industry as well as the emerging bioeconomy. Furthermore, making this information available to tree breeders will help in decisions about planting certain genotypes with desirable traits in different environments.

3.2 Introduction

Timber harvested in the southern United States accounts for 60% of the wood consumed nationally and 18% of global wood consumption (Wear and Greis 2002). The majority of this production is sourced from plantation grown loblolly pine. This widely grown tree species dominates on approximately 13.4 million ha throughout the southeastern forests and accounts for more than 50% of the standing pine volume in the region (Schultz 1997). As the most economically important tree species in the United States, loblolly pine provides some 110, 000 direct and indirect jobs and contributes approximately $30 billion to the economy of the southeastern United States (Schultz 1999).

For efficient utilization of loblolly pine, knowledge of its chemical properties will be valuable. Chemically, wood is principally composed of carbon (C), oxygen (O), and hydrogen (H). These three elements combine into the three structural components of cellulose, hemicelluloses and lignin. Cellulose is the most abundant and can make up 40% to 50% of the dry weight of wood. The percentage of hemicelluloses in the dry weight of wood ranges between 15% to 35%, and for lignin 18% to 35% (Haygreen and Bowyer 1989; Farmer 1967). Wood also has non-structural components known as extraneous compounds - examples of which include carotenoids, sterols, flavonoids, tannins, waxes essential oils and phenolics (Fengel and Wegener 1984). The fraction of extranoues compounds that can be extracted with polar and non-polar organic solvents are known as extractives. Extractives usually make up 1% to 5% of the dry weight of wood but can be as much as 14% in pines (Sjostrom 1981) and contribute to the natural durability of most tree species. Furthermore, several mineral elements also are present in wood. These mineral elements occur either as primary elements or trace elements (Ragland et al. 1991; Hakkila and Parikka 2002). Most of the trace elements occur in wood and plant biomass not because plants need them but because they are found in the soil. The

chemical composition of wood impacts other wood properties, as well as the quality of final products. For example, the lignin and extractives content that may be desirable to a degree because they impart durability against biological agents in lumber products are problematic in pulp and paper applications due to their adverse effects on the pulping process. Similarly, the chemical composition affects traits such as density and strength; two important properties that dictate the quality of wood for structural applications.

In addition to the conventional forest products industry, loblolly pine will form an important component of the renewable energy portfolio as the United States seeks national and energy security as well as environmental sustainability. In recent years, interest has been growing in the use of loblolly pine as a renewable energy feedstock due to existing knowledge of intensive southern pine plantation management, favorable production economics and high yields (U. S. DOE 2016; Jernigan et al. 2016).

Several loblolly pine energy plantations concepts have been proposed. One approach is through dual-cropping whereby a pine stand is established and managed to intentionally produce conventional sawtimber and pulpwood as well as bioenergy. The timber crop is planted in widely spaced rows; in between these, trees are planted in tightly spaced rows for bioenergy (Scott and Tiarks 2008). An alternative to this approach is to have more efficient fast growing dedicated pine plantations at higher densities and shorter rotations (Jernigan et al. 2016). However, fast-growing, densely planted short rotation pine plantations will be susceptible to endemic organisms such as fusiform rust, pine beetles, pine tip moths and seedling debarking weevils (Zalesny et al. 2011; Baker and Langdon 1990). As such disease tolerance should be of prime consideration in the establishment of pine energy plantations.

Just as for the conventional applications, the chemical composition will affect the yield and quality of bio energy/fuel products. For instance, in the production of ethanol, wood with a high concentration of cellulose is desirable for high yields. On the flip side, high amounts of extractives and lignin will inhibit the bioconversion process. In addition to the chemical properties, knowledge about the bioenergy potential (i.e. proximate composition and energy content) of loblolly pine is necessary in bio energy/fuel applications. The proximate composition of volatile matter, fixed carbon and ash give an indication of the thermal reactivity of a fuel (McKendry 2002). A fuel with high volatile matter content is easier to ignite and yield higher quantities of liquid products; whereas a higher fixed carbon content gives more solid products. For example, poplar, which has a 75% volatile matter content has an ignition temperature of 235 ^{o}C, whereas the volatile matter content and ignition temperature of eucalyptus has has been reported to be 64% and 285 ^{o}C respectively (Basu 2010). The mineral elements usually form particulates known as ash during combustion or gasification.

As such, knowing the chemical composition, proximate composition and energy content of loblolly pine could aid in the decision making process with regards to the suitability of this feedstock to support the conventional forest products industry as well as the emerging bioeconomy.

However, current methodologies employed for determining wood properties are laborious, costly and usually destructive. Alternative analytical tools that can be used to rapidly and cost effectively estimate these important properties will thus be invaluable to stakeholders.

Near infrared spectroscopy (NIR) has emerged over the years as a rapid and reliable tool for the nondestructive estimation of the properties of wood and other forest products. A good number of research studies have reported on the application of NIR to predict the chemical composition, proximate composition and energy content of wood and other lignocellulosic materials.

Sykes et al. (2005) used NIR to predict the properties of loblolly pine wafers obtained from two sites. The researchers determined that although models developed from samples of one site could be used to predict the cellulose content of wood from the other site, the R^2 value was lower than those for the individual site predictions and may thus be only applicable for ranking and selection purposes. Nkansah et al. (2010) also used NIR to predict the extractives, holocellulose and lignin content of *Liriodendron tulipifera* solid wood blocks employing the full (800 nm – 2500 nm) and reduced (1300 nm -1800 nm) NIR spectra. Their best models were those calibrated with the full NIR range; reporting R^2 values of 0.84 for extractives, 0.68 for holocellulose and 0.64 for lignin. In both studies, the models developed for lignin content had low predictive power. Sykes et al. (2005) attributed this result to the generally low variation of lignin content in wood and usually large errors associated with lignin determination in the laboratory.

In another study comparing NIR calibration models developed from solid and ground *Eucalyptus globulus* wood, Poke and Raymond (2006) noted that models built from ground wood samples did a poor job of estimating the chemical properties of solid wood samples. Jiang et al. (2012) however demonstrated that decreasing particle size improved the precision of NIR-based models when predicting the chemical composition of southern pines.

Several studies also have reported the proximate composition and energy content of wood as determined with NIR. According to the literature, extractives and lignin content have a proportional relationship with the heating value of wood (Demirbas 2001; White 1987). So and Eberhardt (2010) for instance used reflectance NIR to predict the higher heating value of *Pinus palustris* considering the effect of lignin content and extractives. They reported that the models predicted the HHV of unextracted wood samples better than they did acetone-extracted samples. In

addition, graphs of the regression coefficients showed similar plots for the HHV and extractives content; an implication that the two properties have similar molecular features.

With respect to the ash content of wood, NIR-based models generally give low coefficients of determination. For example, the model developed by Fagan et al. (2010) to determine the ash contents of two dedicated energy crops had an R^2 value of 0.58. These poor results have been attributed to the fact that NIR does not interact directly with the compounds that form ash, e.g. calcium, potassium and silica.

In this current study, NIR was used to rapidly determine the extractives, lignin, cellulose, glucose and hemicelluloses content (i.e. chemical composition) of elite loblolly pine families (i.e. an essentially new resource). In addition, the proximate analysis and energy content of these select families were also estimated to ascertain the bio energy/fuel potential of this feedstock. Developed NIR-based models were then used to partition out the pine families for optimum applications. Moreover, the stakeholders will be provided with this knowledge so that it can be incorporated back into tree breeding programs that aim to improve wood quality for different end users.

3.3 Materials and Methods

3.3.1 Material

Southern pine wood samples that were loaded until failure in structural testing were used (Chapter 2). Materials had been sawn into smaller blocks after destructive testing and stored in airtight zip lock bags in a conditioning chamber until they were needed for further analysis. Prior to subsequent analysis, wood blocks were first chipped using a chisel and hammer. For each sample, a Wiley mill was used to grind a portion to pass a 40-mesh screen, and the remaining was ground to pass an 80-mesh screen. The 40-mesh sample were used to determine the chemical and proximate composition as well as the energy content of pine wood. NIR spectra were collected from the materials that were ground to pass the 80-mesh screen.

3.3.2 Chemical Analysis

Laboratory experiments following conventional standards were used to determine the chemical composition of loblolly pine wood. Extractives content was determined following NREL/TP-510-42619 (Sluiter et al. 2008). Using a Soxhlet Apparatus, 150 ml of industrial grade acetone was used to extract 5 g of test sample for 6 hours. An additional 2 g of the sample was taken at this time for moisture content (MC) determination. Acetone was evaporated from the extract using a rotary evaporator. Extract was then dried at 40 ^{o}C for 24 hours in a vacuum oven and the final mass measured for extractives content determination.

The amount of total lignin and carbohydrates in samples were determined as described in NREL/TP-510-42618 (Sluiter et al. 2012). Air dried extractive-free test sample (0.5 g) was weighed into a dry pressure bottle and 5 ml of 72% sulfuric acid was added. Employing a two-step procedure to ensure complete hydrolysis of sugars, test sample was first placed in a water bath set at 30 ^{o}C +/-

3 ^{o}C and incubated for one hour. Afterwards, the mixture was diluted to a concentration of 4% with deionised water and placed in an autoclave set at 121 ^{o}C for another hour.

The total lignin was computed as the sum of acid soluble lignin (ASL) and acid insoluble lignin (AIL). ASL was determined with a UV – Visible spectrophotometer immediately after hydrolysis. Absorbance of a test sample was measured at the recommended wavelength of 240 nm, ensuring that the absorbance ranged between 0.7 – 1.0.

Monomeric sugars (i.e. glucose, xylose, galactose, arabinose and mannose) in test samples were determined via High Performance Liquid Chromatography (HPLC) using a Biorad Aminex HPX-87P column equipped with the appropriate guard column at a column temperature of 85 ^{o}C and run time of 35 minutes. Holocellulose was computed as the sum of all monomeric sugars; cellulose was computed as $\mathrm{glucose} - (\frac{1}{3} * \mathrm{mannose})$ and hemicelluloses computed as the difference between holocellulose and cellulose (Jiang et al. 2014).

3.3.3 Proximate Analysis and Energy Content Determination

Proximate analysis was conducted to determine the thermal reactivity of pine wood samples. The volatile matter of samples was determined as specified in CEN/TS 15148 (CEN 2005) using a furnace (VMF Carbolite, model 10/6/3216P, England). Empty crucibles with their lids were first heated to 900 ^{o}C ± 10 ^{o}C for 7 minutes. Crucibles were allowed to cool to room temperature in a desiccator before filling with 1 g ± 0.1 g of test material. These were then returned to the furnace for 7 minutes ± 5 seconds. They were removed from the furnace and allowed to cool in a desiccator. The volatile matter content was determined using the equation

% Volatile matter = {100 (M2 – M3) / (M2 – M1) – MC} X {100 / (100 – MC)}

M1 is the mass (g) of the empty crucible and lid;

M2 is the mass (g) of the crucible and lid and test portion before heating;

M3 is the mass (g) of the crucible and lid and contents after heating;

MC is the moisture content on wet basis.

Ash content was determined following NREL/TP-510-42622 (Sluiter et al. 2008). Crucibles were dried in a muffle furnace (Thermoscientific, model F6020C, Waltham, MA, USA) set at 575 ^{o}C ± 25 ^{o}C for four hours. Crucibles were then taken out and allowed to cool for one hour. They were filled with 0.5 g ± 0.1 g of test samples and placed back into the furnace set at 105 ^{o}C. This temperature was held isothermal for twelve minutes before ramping to 250 ^{o}C at 10 ^{o}C/min. After holding for thirty minutes at 250 ^{o}C, the temperature was increased to 575 ^{o}C at a rate of 20 ^{o}C/min and then held isothermal for 180 minutes. Furnace was allowed to cool down to 105 ^{o}C before samples were removed. The ash content of pine wood was determined using the equation:

$$\% \text{ Ash} = \{(M3 - M1) / (M2)\} \times \{100 / (100 - MC)\}$$

Where:

M3 is the final mass (g) of crucible with ash after completion of experiment

M1 is the mass (g) of empty crucible

M2 is the initial mass (g) of test sample

MC is the moisture content on wet basis

The fixed carbon content of wood was determined as the sum of the percentages of moisture, volatile matter and ash deducted from 100.

The energy of test samples was determined according to ASTM D5865 (ASTM 2013) using an IKA C-200 bomb calorimeter (IKA Works Inc., model C200, Wilmington, NC, USA). Approximately 0.5 g of test sample was pelletized and completely combusted in an oxidative environment.

3.3.4 Near Infrared Spectroscopy (NIR)

NIR spectra of test samples were collected using a PerkinElmer Spectrum Model 400 NIR spectrometer (Waltham, MA, USA). The wavenumber range of the instrument was from 10000 cm^{-1} to 4000 cm^{-1}. Each sample was scanned thirty-two times at a resolution of 4 cm^{-1} and averaged into one spectrum for analysis. Spectrum of a Spectralon standard was taken as the background reference sample every 20 minutes to correct for potential drifts with time.

3.3.5 Multivariate Data Analysis

PerkinElmer Spectrum Quant+ software (Waltham, MA, USA) was used to develop Partial Least Squares regression (PLS) models. The first derivatives of raw spectra were used for PLS1 model construction.

A total of two hundred and fifty samples were used in model calibration and validation. Samples were randomly assigned to the either the calibration or test set. One hundred and ninety samples were used for model calibration and full cross-validation. The remaining 60 samples were used as an independent test set for external validation. The performance of validated models were evaluated using the following five statistics: standard error of calibration (SEC), standard error of cross-validation (SECV), standard error of prediction (SEP), coefficient of determination (R^2) and the residual predictive deviation / ratio of performance to deviation (RPD). Models that had the lowest error values were selected and used to predict the chemical and thermochemical properties under study.

The PROC GLM procedure in SAS (SAS Institute, Inc. Cary, NC, USA) was used to conduct analysis of variance (ANOVA). Tukey-HSD tests with alpha set to 0.05 were conducted when needed to further investigate pair-wise comparison between the treatments. All diagrams and tables were produced with MS Excel (Microsoft Corp. Redmond, WA, USA).

3.4 Results and Discussion

3.4.1. Chemical Composition, Proximate Composition and Energy Content

The descriptive statistics of the chemical composition determined via conventional laboratory methods are presented in Table 3.1. The mean extractives content of southern pines used in this study was 3.06% (SD = 1.59%), with minimum and maximum values of 0.37% and 9.44% respectively. Percent extractives determined for the commercially acquired southern pine samples were slightly higher than that measured for the elite loblolly pine families (i.e. 3.8% versus 2.77%). The range of extractives content determined for pine samples in this study were low compared to some of what have been stated in the literature. Kelley et al. (2004) reported a range of 2.8% to 26.9% for loblolly pine whereas So and Eberhardt (2010) reported a range of 0.0% to 20.6% for *Pinus palustris*.

In the case of total lignin content, values ranged from a low of 26.72% to a high of 34.96%. The mean lignin content of the loblolly pine families were more similar to that of the southern pines -- 31.74% (SD = 1.52%) and 3.11% (SD = 1.64%) respectively. For the carbohydrates, wider ranges were recorded for the loblolly pine families than for the southern pines. For instance, cellulose ranged from 29.36% to 47.16% for the former but 30.42% to 42.76% for the latter. The mean cellulose, glucose and hemicelluloses of all samples were 37.2% (SD = 2.92%), 40.29% (SD = 3.07%) and 22.88% (SD = 1.57%) respectively.

Table 3.1: Descriptive statistics of the chemical composition of southern pine wood.

		% Extractives	% Lignin	% Cellulose	% Glucose	% Hemicelluloses
Calibration set (n = 190)	Mean	3.01	31.29	37.35	40.43	22.71
	SD	1.57	1.53	2.97	3.10	1.60
	Min	0.37	26.72	30.42	33.33	19.01
	Max	9.44	34.70	47.16	50.16	25.85
Test set (n = 60)	Mean	3.20	32.47	36.73	39.86	23.39
	SD	1.64	1.39	2.73	2.97	1.36
	Min	0.74	29.79	29.36	31.48	20.13
	Max	7.61	34.96	41.90	45.78	25.79
Total set (n = 250)	Mean	3.06	31.56	37.20	40.29	22.88
	SD	1.59	1.57	2.92	3.07	1.57
	Min	0.37	26.72	29.36	31.48	19.01
	Max	9.44	34.96	47.16	50.16	25.85
Loblolly pine families (n = 180)	Mean	2.77	31.74	37.26	40.39	23.13
	SD	1.27	1.52	3.13	3.26	1.50
	Min	0.37	27.11	29.36	31.48	19.43
	Max	7.61	34.96	47.16	50.16	25.85
Commercial lumber (n = 70)	Mean	3.80	31.11	37.04	40.03	22.21
	SD	2.03	1.64	2.32	2.51	1.56
	Min	0.38	26.72	30.42	33.33	19.01
	Max	9.44	34.33	42.76	46.30	24.73

In the literature, Sykes et al. (2005) reported the lignin and cellulose contents of 11-year old loblolly pine to range from 24.1% to 32.63%, and 35.6% to 55.2% respectively. Jones et al. (2006) also determined the mean chemical composition of loblolly pine aged between 21 to 25 years as follows: glucan – 39.2%, cellulose – 35.9%, hemicelluloses – 23.3% and lignin – 28.3%.

Results obtained from proximate analysis and bomb calorimetry are shown in Table 3.2. Percent volatile matter determined for all the southern pine wood samples ranged from a high of 87.05% to a low of 80.86%, with a mean of 83.55% and SD of 1.16. Also, the mean fixed carbon content of all test samples was 16.21% (SD = 1.17%); whereas % ash was between 0.14 – 0.59%. The mean percent volatile matter, fixed carbon and ash determined for southern pine wood in this study were similar to the respective 85.6%, 14.1% and 0.32% reported by Owen et al. (2016) for loblolly pine stemwood.

Although the range of calorific value determined for southern pine in this study was relatively wider (i.e. 17.0 – 22.65 MJ/kg) than what have been usually reported for clean wood, the overall mean value of 17.9 MJ/kg (SD = 0.6) was lower nonetheless. So and Eberhardt (2010) presented a narrower HHV range of 20.2 – 23.26 MJ/kg for unextracted *Pinus palustris*, which reduced further after extracting samples with acetone. Notwithstanding, wide calorific values ranges have been reported of lignocellulosic biomass comprising of needles, twigs and bark of conifers, broad leaved trees, shrubs and grasses (Gillon et al. 1997).

Table 3.2: Descriptive statistics of the proximate composition and energy content of southern pine wood.

		% Volatile matter	% Fixed carbon	% Ash	HHV (MJ/kg)
Calibration set (n = 190)	Mean	83.63	16.12	0.25	17.84
	SD	1.14	1.15	0.09	0.46
	Min	80.86	12.62	0.14	17.02
	Max	87.05	18.84	0.59	19.26
Test set (n = 60)	Mean	83.27	16.48	0.25	18.10
	SD	1.21	1.20	0.10	0.90
	Min	81.31	12.72	0.15	17.01
	Max	86.93	18.32	0.54	22.65
Total set (n = 250)	Mean	83.55	16.21	0.25	17.90
	SD	1.17	1.17	0.09	0.60
	Min	80.86	12.62	0.14	17.01
	Max	87.05	18.84	0.59	22.65
Loblolly pine families (n = 180)	Mean	83.43	16.32	0.26	17.88
	SD	1.17	1.18	0.10	0.66
	Min	80.86	12.72	0.14	17.01
	Max	86.93	18.84	0.59	22.65
Commercial lumber (n = 70)	Mean	83.85	15.92	0.22	17.96
	SD	1.10	1.10	0.07	0.42
	Min	81.48	12.62	0.14	17.22
	Max	87.05	18.32	0.43	19.26

Generally, wide ranges with good overlaps were noted among the training/cross-validation and independent datasets. These have been known to help improve the robustness of models when applied in predicting properties of future unknowns (Haartveit and Flæte 2006; Esteves and Pereira 2008).

3.4.2. NIR-Model Calibration and Evaluation for Chemical Composition

Partial least squares regression models were developed using the first derivatives of NIR spectra as the independent variables (i.e. X-variables) and the data obtained via the conventional laboratory methods were used as the dependent variables (i.e. Y-variables). One hundred and ninety samples were used in model calibration and full cross-validation. An additional test set made up of 60 samples were used for external validation. Calibrations obtained for the chemical components are summarized in Table 3.3. Using four or five latent variables, good calibration coefficients were obtained for all the properties with R^2_{cv} values greater than 0.7 (Williams 2004). Also, relatively small differences were observed between the errors of calibration (SEC) and cross-validation (SECV). This is an indication that the selected model did a good job of predicting the properties of the single-element test set at each iteration (Sanderson et al. 1996). The strong correlation (R = 0.99) between glucose and cellulose was reflected in their similar fit statistics. The relatively lower errors associated with NIR-based PLS models relative to the standard deviation of the laboratory determined reference data generated models with good RPDs. RPD values of cross-validated models ranged from a low of 1.58 for % hemicelluloses to a high of 2.48 for % glucose.

Table 3.3: Calibration and prediction statistics of NIR-based PLS models for chemistry.

	% Extractives	% Lignin	% Glucose	% Cellulose	% Hemicelluloses
LVs	4	4	5	5	4
SEC	0.66	0.78	1.15	1.10	0.87
SECV	0.76	0.89	1.25	1.20	1.01
R^2_{cv}	0.77	0.78	0.84	0.84	0.73
RPD	2.07	1.73	2.48	2.47	1.58
SEP	1.29	1.41	2.46	2.28	1.72
R^2_{iv}	0.43	0.23	0.47	0.46	0.04

Note: Subscript cv means cross-validation; iv means independent validation.

According to the literature, this qualified all developed models to be used as a preliminary screening tool (Via 2013; Hein et al. 2009).

The model statistics for the chemical components were comparable to what was determined by Jones et al. (2006) for lignin (SEC = 0.48%; SECV = 0.92%; R^2 = 0.85), cellulose (SEC = 1.03%; SECV = 1.86%; R^2 = 0.8), glucan (SEC = 1.09%; SECV = 1.96%; R^2 = 0.82) and hemicelluloses (SEC = 0.92%; SECV = 1.24%; R^2 = 0.59) of loblolly pine wood.

In a more similar study, Sykes et al. 2005 sought to use NIR to predict some chemical properties of 14 full-sib loblolly pine families obtained from two different sites. The study developed several models using the latewood, earlywood, growth rings three and eight, as well as separate models using materials from the two sites. For α-cellulose models, the researchers reported an R^2 value range of between 0.56 to 0.63 for the different wood types and rings, which increased to 0.75 (i.e. SEC = 2.4%; SECV = 2.0%) when the whole core was used in modeling. Their lignin models however had bad performances, with for instance, the site B model having an R^2 value of 0.16 and the model developed with the complete data set having an R^2 of 0.37.

When models developed in this study were applied in predicting the chemical properties of an independent test set, the errors (i.e. SEP) increased as expected and consequently affected the R^2_{iv} values adversely; especially for % hemicelluloses. This poor performance can be attributed to two reasons. Firstly, a model usually performs much worse on an independent test data not originally included in model training. Secondly, the test data set was made up of fifteen different loblolly pine families. In the previous study (i.e. Chapter 2), with similarly low R^2_{iv}, the predicted test samples were separated out into the individual elite families and a One-way ANOVA (α = 0.05) conducted to test for equality of means between NIR-predicted and lab-measured property. The

results showed no statistically significant differences (i.e. $p > 0.05$) between NIR-predicted values and measured values for all the three properties understudied for each of the fifteen families, Appendix A.

3.4.3. NIR-Model Calibration and Evaluation for Proximate Composition and Energy Content

Calibrations that were obtained for the bio energy/fuel related properties are presented in Table 3.4. Between three to five latent variables were used in the construction of PLS models. The best cross-validated models were for % volatile matter (SECV = 0.4%; R^2_{cv} = 0.88; RPD = 2.23) and HHV (SECV = 022%; R^2_{cv} = 0.83; RPD = 2.04). Three latent variables were used in building the model for ash which gave an R^2_{cv} value of 0.58 and RPD value of 1.54. Contrasting results have been reported in the literature about the capability of infrared spectroscopy to model the ash content of lignocellulosic biomass. For instance, whereas the models developed by Lestander and Rhen (2005) performed very well in predicting the ash content of Norway spruce wood, those by Fagan et al. (2010) had varying prediction success based on the wavenumber range and spectral pretreatment technique used.

Table 3.4: Calibration and prediction statistics of NIR-based PLS models for bioenergy.

	% Volatile matter	% Fixed carbon	% Ash	HHV (MJ/kg)
LVs	5	3	3	4
SEC	0.37	0.55	0.04	0.20
SECV	0.40	0.68	0.06	0.22
R^2_{cv}	0.88	0.65	0.58	0.83
RPD	2.23	1.68	1.54	2.04
SEP	0.95	0.65	0.08	0.32
R^2_{iv}	0.09	0.10	0.20	0.44

Note: Subscript cv means cross-validation; iv means independent validation

Among the four properties modeled, the prediction errors (i.e. SEP) for volatile matter was the highest – actually more than twice the SECV. In contrast, the SEP of the model for % fixed carbon was slightly lower than its SECV. Although errors associated with a model is generally expected to increase when used for predicting future unknown, this is not always the case (Workman Jr. 2011). As is apparent for the fixed carbon model in this study and others studies (Jones et al. 2006; Kelley et al. 2004). Except for the model for HHV, the others did not perform well when used to predict properties of the independent testset.

3.4.4 Prediction and Screening of the Elite Loblolly Pine Families for Chemistry

Validated models that had the lowest error values were used to predict the chemical composition of loblolly pine families obtained from two forest sites.

Results of Two-way ANOVA testing the effect of family, site and family x site interaction are summarized in Table 3.5. The chemical composition of the loblolly pines families were generally differernt due to the variation in their genetic makeup.

Table 3.5: ANOVA *P-values* per treatment for the chemical constituents.

Property	Family	Site	Family x Site
% Extractives	0.0077	< 0.0001	0.0435
% Lignin	0.0186	< 0.0001	< 0.0001
% Cellulose	< 0.0001	0.0001	0.0252
% Glucose	0.001	0.0327	< 0.0001
% Hemicelluloses	0.0773	< 0.0001	0.0356

In addition, the significant interaction term for all properties studied indicates that the chemical composition of a family could also vary dependng on the environment. As stated by Zobel and van Buijtenen (1989), the properties of wood are a consequent of the interaction between a tree's genetic potential and growing environment.

Descriptive statistics for the chemistry of the elite pine families on the two forest sites are presented in Table 3.6 and Table 3.7.

Table 3.6: Chemical composition of loblolly pine families on the Florida site.

Family	N	% Extractives	% Lignin	% Glucose	% Cellulose	% Hemicelluloses
A1	14	3.81 (1.2)	32.07 (0.92)	45.90 (2.79)	39.67 (2.31)	20.50 (0.66)
A2	10	4.69 (1.33)	32.36 (1.23)	44.60 (2.69)	38.89 (2.40)	20.71 (1.17)
A5	11	4.60 (1.27)	32.00 (0.90)	42.22 (1.60)	37.05 (1.51)	20.66 (0.81)
A9	10	4.30 (1.12)	31.61 (1.18)	42.38 (1.89)	39.70 (1.75)	20.99 (0.43)
A10	12	4.94 (1.98)	32.01 (1.23)	47.83 (5.24)	38.61 (3.96)	21.14 (0.60)
A13	7	4.18 (1.05)	32.22 (1.51)	45.86 (2.99)	39.97 (2.81)	21.12 (0.76)
A15	8	4.34 (1.45)	31.40 (0.61)	42.66 (2.05)	40.34 (2.16)	21.20 (1.20)
A21	12	4.03 (1.03)	31.63 (0.87)	43.23 (1.59)	39.90 (1.92)	21.01 (0.90)
A26	12	5.02 (1.84)	32.61 (0.95)	42.61 (1.37)	39.57 (2.26)	21.18 (0.67)
A33	13	4.33 (1.16)	32.04 (0.89)	43.77 (2.28)	38.27 (1.97)	20.62 (0.78)
A34	10	4.71 (1.22)	31.68 (0.70)	44.46 (2.48)	38.89 (2.11)	20.56 (0.71)
A37	12	5.25 (0.73)	30.51 (0.93)	45.84 (1.71)	39.88 (1.40)	20.40 (0.47)
F17	13	6.69 (0.87)	32.48 (0.65)	45.17 (2.38)	39.29 (2.10)	20.74 (0.80)
F23	12	4.70 (1.42)	31.75 (75)	43.43 (2.58)	37.32 (2.42)	21.10 (0.91)

Table 3.7: Chemical composition of loblolly pine families on the Georgia site.

Family	N	% Extractives	% Lignin	% Glucose	% Cellulose	% Hemicelluloses
A1	15	6.33 (1.76)	28.98 (1.27)	46.92 (3.58)	41.97 (3.13)	19.93 (0.25)
A2	14	4.41 (1.94)	29.21 (0.77)	46.79 (1.30)	41.86 (1.44)	20.09 (0.38)
A5	13	6.67 (2.12)	29.90 (1.45)	43.73 (3.86)	38.94 (3.09)	20.22 (0.26)
A9	14	6.65 (2.04)	29.14 (1.16)	45.64 (3.53)	40.65 (3.00)	20.24 (0.48)
A10	14	6.38 (2.63)	29.38 (1.35)	44.09 (3.42)	39.23 (2.79)	20.33 (0.43)
A13	14	6.86 (1.62)	29.55 (1.23)	45.42 (3.29)	40.49 (2.68)	20.13 (0.19)
A15	12	5.98 (2.35)	29.10 (1.07)	46.01 (3.91)	41.02 (3.52)	20.13 (0.41)
A21	15	4.34 (2.50)	29.60 (1.51)	45.03 (3.70)	40.29 (3.16)	20.15 (0.33)
A26	15	6.57 (2.24)	28.97 (1.13)	46.13 (4.13)	41.48 (3.97)	19.79 (0.54)
A33	13	6.94 (1.57)	30.82 (0.82)	40.80 (2.71)	36.69 (2.20)	20.10 (0.12)
A34	14	5.97 (2.24)	29.74 (1.48)	44.58 (4.42)	39.98 (3.99)	20.12 (0.48)
A37	15	7.36 (2.20)	30.16 (1.25)	43.57 (3.49)	38.96 (2.92)	20.17 (0.24)
F17	13	6.83 (1.76)	29.21 (1.64)	44.62 (4.07)	39.83 (3.39)	20.14 (0.40)
F23	14	6.24 (1.74)	28.69 (0.88)	46.82 (2.38)	41.82 (2.28)	19.93 (0.36)

The extractives content of the elite loblolly pine families ranged between 3.81 to 7.35%, Table 3.6 and Table 3.7. This is comparable to what was reported by Shupe et al. (1997). The researchers determined an extractives range of 2.5 to 6.98% for loblolly pine wood obtained from five stands with different silvicultural treatments. They noted the innerwood contained a relatively higher amount (i.e. 5.23 to 6.98%) of alcohol-benzene extractives compared to the outerwood (i.e. 2.5 to 4.53%). Via et al. (2007) also reported similar results whereby the extractives content of *Pinus*

palustris decreased as one moved from the pith towards the back and also from the butt higher up the stem.

On the Florida site, the lowest extractives content was determined for A1 (Mean = 3.81%; SD = 1.2%) and the highest estimate was for F17 (Mean = 6.69%; SD = 0.87%). Except for A37, A2 and A13, the extractives content of F17 was statistically higher than that predicted for the other families. For the Georgia site, the highest of 7.36% (SD =2.2) predicted for A37 was however statistically significant from only A21 (Mean = 4.35; SD = 2.5) and A2 (Mean = 4.41; SD = 1.94).

As can be seen in Table 3.5, the differences in the extractives content between the families on a site is a function of genotypic variation; whereas differences within a single family on the two sites is a to response to environmental conditions such as the presence of biological degraders or silvicultural treatments that aim to increase tree growth (Eckhardt et al. 2009; Pettersen 1984). In spite of the effect of environment on the concentration of extractives, some families including A2, A21 and F17 had similar values on the two sites.

Extractives in wood impart decay resistance. They also are however responsible for several issues related with the utilization of wood. For instance, extractives can contribute to the corrosion of metals in contact with wood, inhibit the setting of adhesives and finishes, as well as affect the swelling, shrinking, chemical treatability / permeability, light stability and flammability of wood (Larson et al. 2001). Furthermore, extractives cause various problems during papermaking (Sithole and Allen 2002).

Means for the lignin content of the loblolly pine families ranged from a low of 28.69% for F23 on the Georgia site to 32.61% for A26. The range of lignin content predicted by NIR for the elite families is consistent with what has been reported in the literature as the natural variation of lignin in juvenile loblolly pine (Sykes et al. 2003). Generally, trees growing on the Georgia site had a lower

lignin content (Mean = 29.46%; SD = 0.56%) than those from the Florida site (Mean = 31.88%; SD = 0.53%), Table 3.5.

The NIR-predicted lignin of 30.5% for A37 was no different than the estimated 31.8% for F23 on the Florida site. Similarly for the Georgia site, the 30.2% determined for A37 also was statistically the same as the 28.5% for F23. In addition, no significant variations were noted among F23, A34, A21, A9, A15 and A37 on both sites with respect to the lignin content.

The cellulose content estimated for the elite families ranged from 36.7% to 42.0%. Even though cellulose content of the families fall within what has been reported in the literature, the range determined is narrower (Jones et al. 2006; Sykes et al. 2005). It also was noted that, the NIR-estimated cellulose content of the juvenile loblolly pine families were higher than that measured for the commercially acquired southern pine samples; i.e. respective means of 39.3% (SD= 3.1%) and 37.0% (SD = 2.32%).

On the Florida site, the predicted cellulose content of A5 was similar to those of all the other families except for A1. Meanwhile on the Georgia site, the cellulose content estimated by NIR models for A1, A2, F23, A26, A15 and A9 all varied significantly from that of A33. Although the family by site effect was significant, some families ranked similarly high or low on the two sites. Examples of such include A1, A15, A26 and A9 on the high end; and A5, A33 and A34 on the low end. The cellulose content of wood is highly correlated pulp yield. Material with a higher percentage of cellulose would increase the efficiency of pulp and paper mills and reduce associated pulping costs (Kube and Raymond 2002). Furthermore, the amount of cellulose have close relationships with the density and strength of wood (Haygreen and Bowyer 1989; Megraw 1985; Winandy and Rowell 1984). Results from this study were consistent with what have been reported

in the literature. The loblolly pine families that had higher concentrations of cellulose also were determined to have higher density, modulus of rupture (ultimate strength) and modulus of elasticity (stiffness) values (Chapter 2).

For the hemicelluloses, the highest value of 21.2% was determined for A15 on the Florida site, and the lowest value of 19.8% was determined for A26 on the Georgia site. No statistical differences were determined among the hemicelluloses content of the elite loblolly pine families on the Florida site. Meanwhile on the Georgia site, only A10 differed significantly from A26.

Percent glucose estimated was 42.2% to 47.8% for the families on the Florida site and 40.8% to 46.9% for families on the Georgia site. The 47.8% of glucose determined for A13 was statistically higher than the 42.2% predicted for A5, as well as the 43.8% predicted for A33 on the Florida site. Similarly, on the Georgia site, the glucose content of 40.8% estimated for A33 differed statistically from the glucose contents of A1, F23, A2, A26, A15, A9 and A13.

As expected, there was a strong correlation ($r = 0.99$) between the glucose and cellulose content of the loblolly pine families. For some families such as A15, A37 and A10 on the Florida site and A1, A2, F23 and A26 on the Georgia site, high glucose corresponded with high cellulose; and the inverse was true for A33, A5, A37 and A10 on the Georgia site. In certain cases, a high glucose content didn't necessarily yield high cellulose, suggesting that more of the glucose was locked in the hemicelluloses such as for A13. On the other hand, certain families although having a low glucose content also had relatively high cellulose content, indicating that more of of the glucose was probably bonded in the cellulose, thus maybe less hemicelluloses; as noted in A1 and A2 on the Florida site.

Graphical representations of how the elite loblolly pine families ranked in their chemical composition on the two forest sites are presented in Figure 3.1 through Figure 3.5.

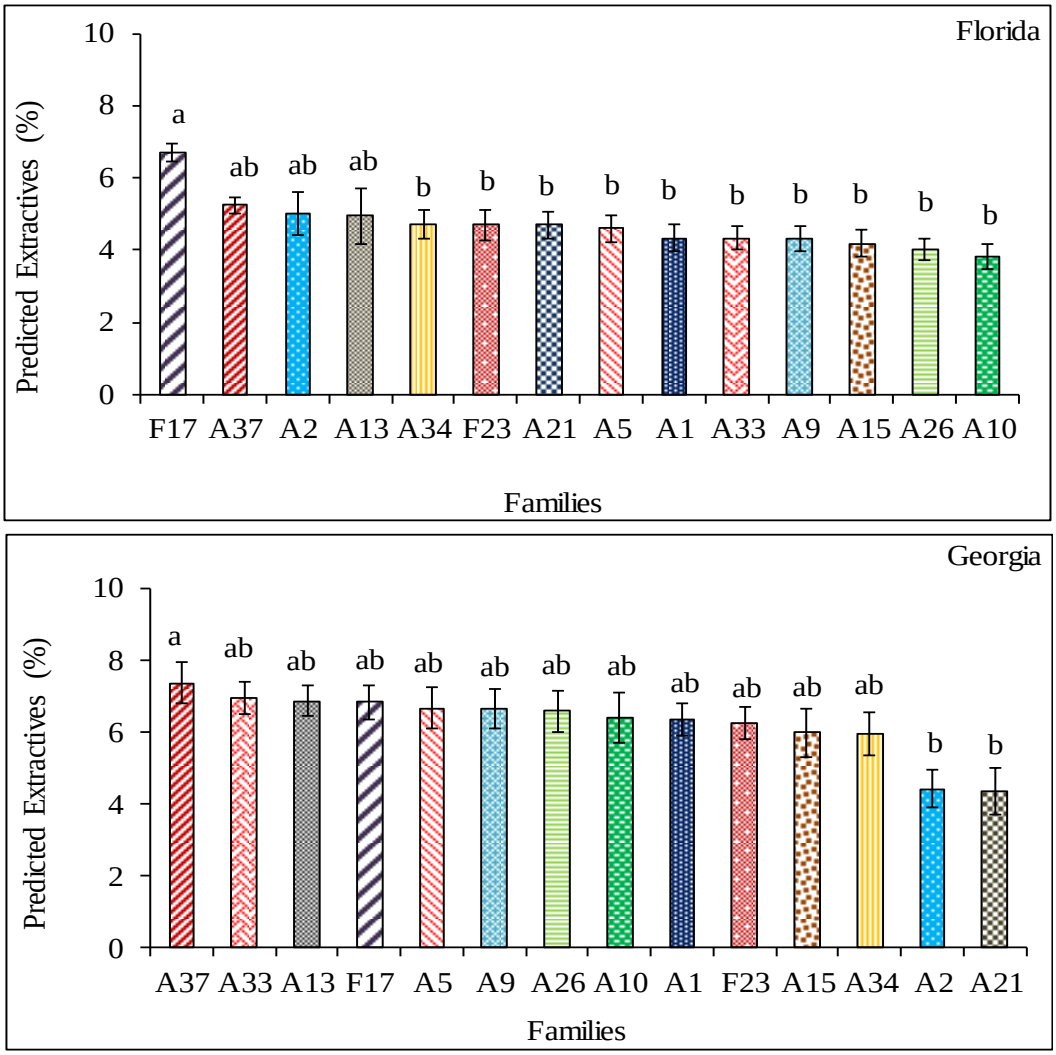

Figure 3.1: Rank of loblolly pine families for extractives content. *Bars with different letters are significantly different at 95% confidence level (Tukey's HSD Test).

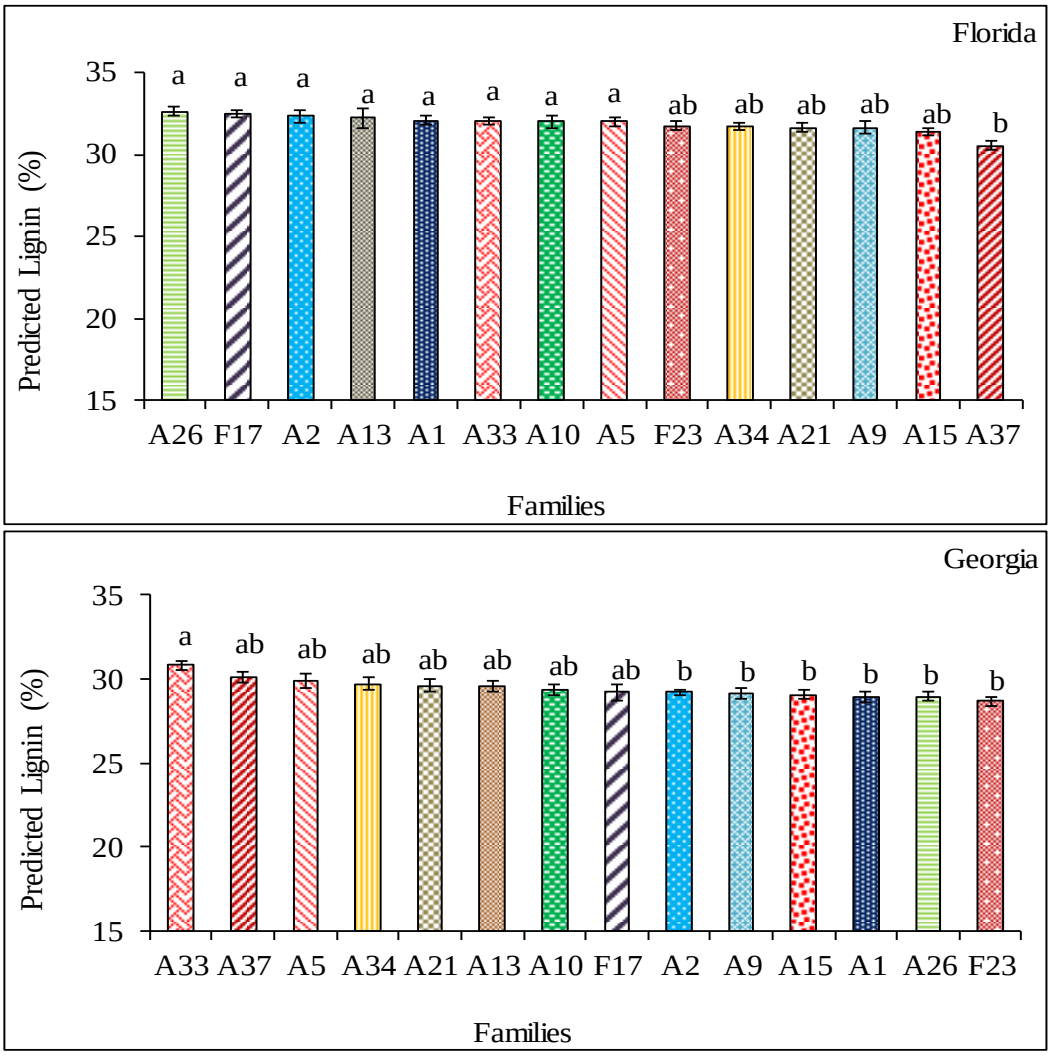

Figure 3.2: Rank of loblolly pine families for lignin content. *Bars with different letters are significantly different at 95% confidence level (Tukey's HSD Test).

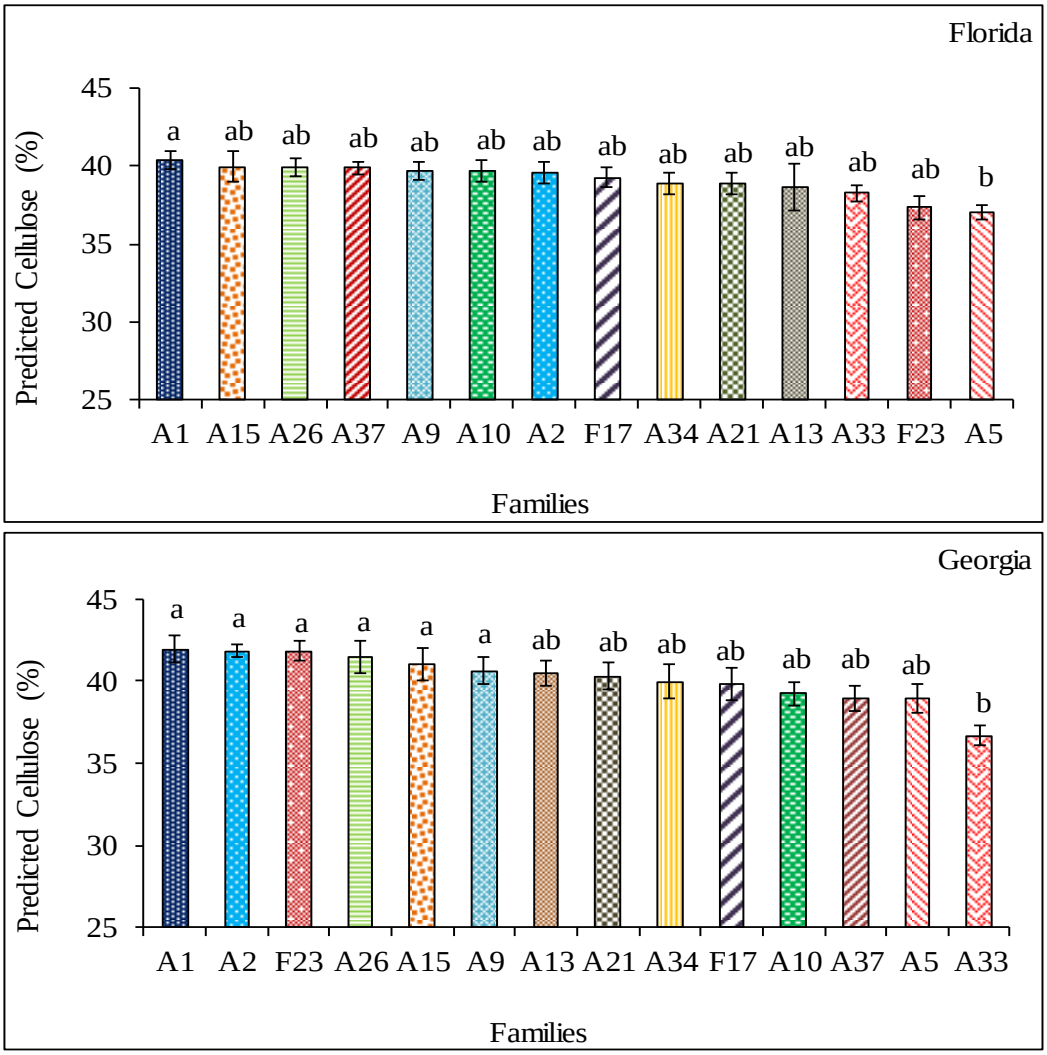

Figure 3.3: Rank of loblolly pine families for cellulose content. *Bars with different letters are significantly different at 95% confidence level (Tukey's HSD Test).

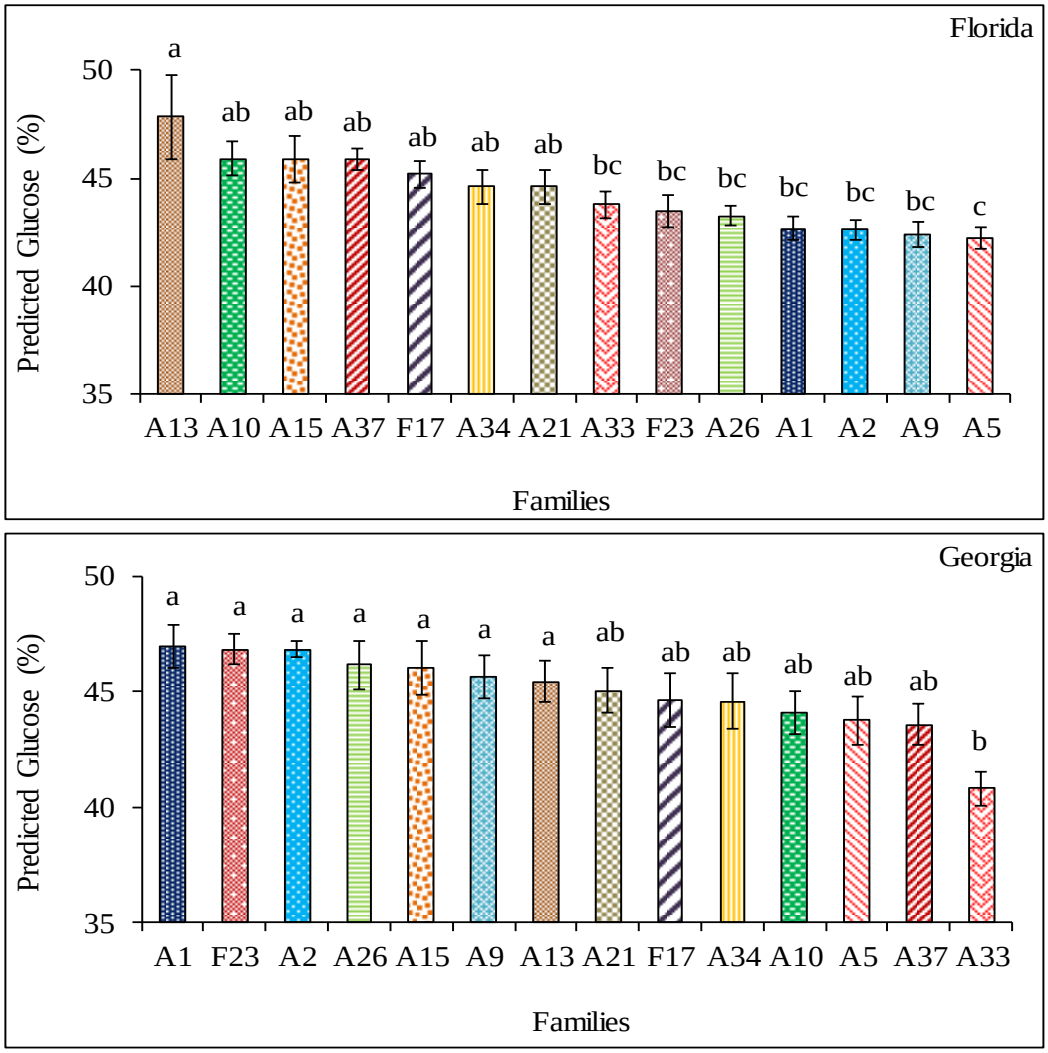

Figure 3.4: Rank of loblolly pine families for glucose content. *Bars with different letters are significantly different at 95% confidence level (Tukey's HSD Test).

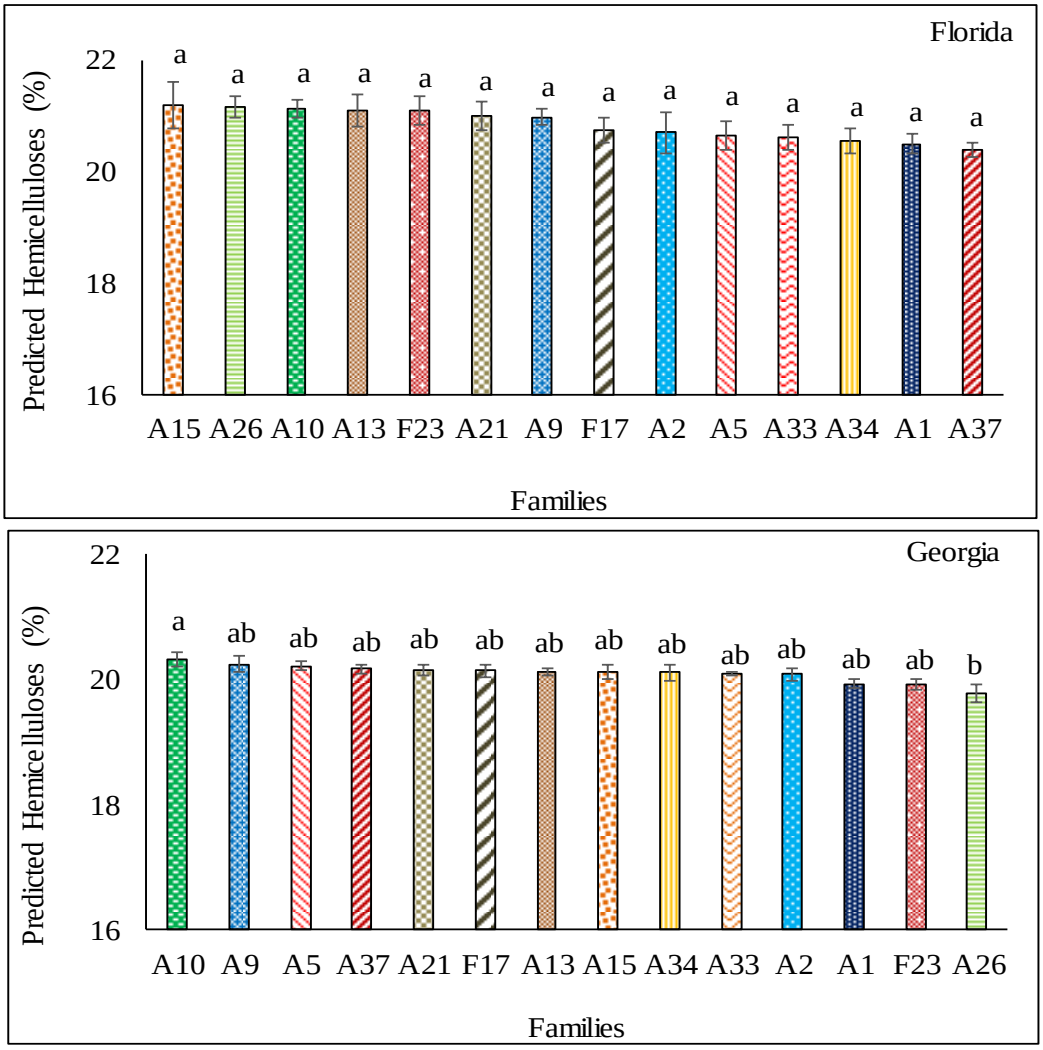

Figure 3.5: Rank of loblolly pine families for hemicelluloses content. *Bars with different letters are significantly different at 95% confidence level (Tukey's HSD Test).

Unlike for cellulose (*P-value* < 0.0001) and glucose (*P-value* = 0.001) contents, less variations were noted among the families with respect to the contents of lignin (*P-value* = 0.0186) and hemicelluloses content (*P-value* = 0.0773). Sykes et al. (2003) reported similar results for 11-year old full-sib loblolly pine families. The effect of site was generally more pronounced for the chemical composition of the loblolly pine families than genetics in this study. Table 3.5.

3.4.5 Prediction and Screening of the Elite Loblolly Pine Families for Bioenergy Potential

Two-way ANOVA results testing the effect of family, site and family x site interaction on proximate composition and energy content are summarized in Table 3.8. With the interaction term for all the understudied properties being significant, the loblolly pine families were separated based on the two sites for further analysis.

Table 3.8: ANOVA *P-values* per treatment for the proximate composition and energy content.

Property	Family	Site	Family x Site
% Volatile matter	<.0001	0.4749	<.0001
% Fixed carbon	0.0029	0.7237	<.0001
% Ash	0.1107	< 0.0001	0.0009
HHV (MJ/kg)	0.0002	0.0166	0.0007

Descriptive statistics for the bioenergy potential of the elite families on the two forest sites are presented in Table 3.9 and Table 3.10. The mean volatile matter content for the families as estimated by NIR was highest for A34 (Mean = 85.7%; SD = 0.3) on the Florida site and lowest for A33 (Mean = 82.6%; SD = 1.5%). Meanwhile, the highest fixed carbon content was determined for A33 to be 17.1% (SD = 1.1%) on the Georgia site, whereas A37 on the Florida site had the lowest of 13.6 (SD = 0.3%). Relatively smaller within-family variations in the volatile matter and fixed carbon contents of trees on the Florida site resulted to more significant differences between the families. On this site, A10, A15, A2 and A5 all had significantly lower volatile matter content compared to A34, F17, F23, A37, A26 and A13. The amount of volatile matter were consistently high on both sites for F17, F23, A37, A9 and A1; while that of A1 and A15 remained low on the two sites. During pyrolysis and gasification, the families with higher concentrations of volatile matter will yield bio-oil

and syngas and lesser amounts of char (Owen et al. 2015). The trade-off between % volatile matter and % fixed carbon was evident in this study. For instance on the Florida site, A34, F17 F23 and A37 all of which had high volatile matter contents also had consequently low fixed carbon contents. Similarly, when the volatile matter content was low, a high fixed carbon content resulted, as noted in A5, A2, A15, A10 and A21. A similar trend occurred among these proximate components for the loblolly pine families on the Georgia site.

Table 3.9: Proximate composition and energy content of loblolly pine families on the Florida site.

Family	N	% Volatile matter	% Fixed carbon	% Ash	HHV (MJ/kg)
A1	14	84.00 (1.64)	14.54 (2.05)	0.24 (0.04)	19.11 (0.42)
A2	10	83.02 (1.16)	15.98 (0.85)	0.24 (0.04)	18.69 (0.49)
A5	11	82.80 (1.42)	15.43 (1.07)	0.24 (0.03)	18.83 (0.46)
A9	10	84.22 (0.86)	15.18 (0.67)	0.24 (0.03)	18.80 (0.56)
A10	12	83.47 (1.07)	15.53 (0.70)	0.24 (0.03)	18.83 (0.49)
A13	7	84.36 (0.76)	14.84 (0.89)	0.27 (0.06)	18.93 (0.70)
A15	8	83.04 (0.49)	15.76 (0.32)	0.24 (0.03)	18.54 (0.83)
A21	12	83.84 (1.37)	15.40 (0.97)	0.24 (0.06)	19.22 (0.55)
A26	12	84.44 (0.60)	15.21 (0.39)	0.20 (0.04)	18.78 (0.29)
A33	13	84.16 (0.68)	14.04 (0.54)	0.21 (0.03)	19.06 (0.42)
A34	10	85.74 (0.29)	13.75 (0.51)	0.20 (0.02)	18.96 (0.44)
A37	12	84.53 (0.39)	13.57 (0.34)	0.21 (0.02)	19.10 (0.54)
F17	13	85.72 (0.81)	13.97 (0.53)	0.23 (0.03)	19.52 (0.46)
F23	12	85.06 (1.20)	14.44 (0.68)	0.24 (0.04)	19.17 (0.30)

Table 3.10: Proximate composition and energy content of loblolly pine families on the Georgia site.

Family	N	% Volatile matter	% Fixed carbon	% Ash	HHV (MJ/kg)
A1	15	83.95 (2.00)	14.47 (2.11)	0.19 (0.04)	19.09 (0.40)
A2	14	84.61 (2.37)	14.63 (1.85)	0.20 (0.05)	18.94 (0.39)
A5	13	84.04 (1.95)	14.16 (2.28)	0.18 (0.04)	19.20 (0.42)
A9	14	84.65 (2.20)	14.39 (2.28)	0.19 (0.05)	19.07 (0.43)
A10	14	85.33 (1.27)	14.24 (1.90)	0.18 (0.03)	19.00 (0.40)
A13	14	84.56 (1.65)	14.31 (1.90)	0.19 (0.02)	19.11 (0.32)
A15	12	83.90 (1.83)	15.89 (1.76)	0.19 (0.04)	19.04 (0.22)
A21	15	84.39 (1.64)	15.04 (1.67)	0.20 (0.03)	18.84 (0.35)
A26	15	83.52 (1.66)	15.09 (1.75)	0.19 (0.05)	19.10 (0.59)
A33	13	82.59 (1.45)	17.13 (1.14)	0.21 (0.02)	19.54 (0.19)
A34	14	84.03 (2.20)	15.40 (2.28)	0.20 (0.04)	18.97 (0.31)
A37	15	84.74 (1.46)	14.71 (1.62)	0.19 (0.04)	19.25 (0.43)
F17	13	84.76 (1.58)	14.53 (2.02)	0.19 (0.03)	19.11 (0.42)
F23	14	84.96 (1.47)	14.49 (1.90)	0.18 (0.04)	18.90 (0.39)

The mean ash content estimated for the families on the Florida site was highest for A13 (Mean = 0.27%; SD = 0.06%) and lowest for A26 (Mean = 0.20%; SD = 0.04%). For the Georgia site, A33 and A10 contained the most and least amounts of ash respectively, Table 3.10. The average ash content determined for families on the Florida site was significantly higher than what was determined for the families on the Georgia site (*P-value* < 0.0001) due to soil contamination on the former site. Even though the mean ash contents of the families were different on the Georgia site, none was statistically significant due to the relatively narrower range couple with the large within-family variations. On the Florida site, the ash contents of A37, A33, A34 and A26 were all

significantly lower than that of A13. In spite of the significant effect of the family by site interaction, similar ash contents were determined for some families on the two forest sites.

Inherently, wood has very little ash compared to other plant parts of tree species, energy grasses and agricultural residues (Owen et al. 2015; Acquah et al. 2015; Allison et al. 2009; Naik et al. 2010). From Table 3.8, the effect of family on ash was not significant (*P-value* = 0.1107) whereas that of site was significant (*P-value* < 0.0001). Environmental factors that have been known to increase the ash content of lignocellulosic biomass includes soil and fertilizer treatments (Allison et al. 2009), operational practices (Owen et al. 2015) and storage (White et al. 1986).

With respect to the calorific value, the range estimated for the elite families was from 19.5 MJ/kg to 18.5 MJ/kg, with a mean of 19.0 MJ/kg (SD = 0.21 MJ/kg). The determined mean energy content of the loblolly pine families is consistent with what has been reported in the literature for loblolly pine and hybrid poplar (Owen et al. 2015; Acquah et al. 2015; Maranan and Laborie 2007); but the range is narrower than the 20.2 to 23.6 MJ/kg reported for *Pinus palustris* by So et al. (2012). Such relatively small variations is to be expected as the energy content of wood from different tree species has been reported to vary by less than 15% (Senelwa and Sims 1999).

On the Florida site, the HHV of F17 was significantly higher than the HHVs of A26, A2 and A15; and on the Georgia site, the energy content of A33 varied significantly from what were estimated for A10, A34, A2, F23 and A21. Despite the positive interaction of family and site (*P-value* = 0.0007), the energy content determined for F17, A1, A37 and A33 were high on both sites, whereas that of A2 and A10 were low on the two site.

Although this did not hold in all cases, some families seemed to be consistent with the literature about how ash adversely affects (Acquah et al. 2016), but extractives and lignin boost

(Demirbas 2001; White 1987) the calorific value of lignocellulosic biomass. For example, on the Georgia site, Families A33, A37 and A5 which had relatively high percentages of extractives and lignin also had higher HHVs. Meanwhile on the Florida site, A9 which had lower extractives and lignin together with high ash content had the lowest energy content among the families. Furthermore, high amounts of ash in biomass hinder the production of liquid and organic yields during thermochemical conversion (Fahmi et al. 2008). It was also noted that families that had higher extractives content also mostly had a higher energy content, compared to families that had higher lignin content. Similar results were reported by So et al. (2012) about how the extractives content had a stronger correlation with the energy content compared to the lignin content.

Plots of how the elite loblolly pine families ranked on energy potential on the two forest sites are presented in Figure 3.6 through Figure 3.9.

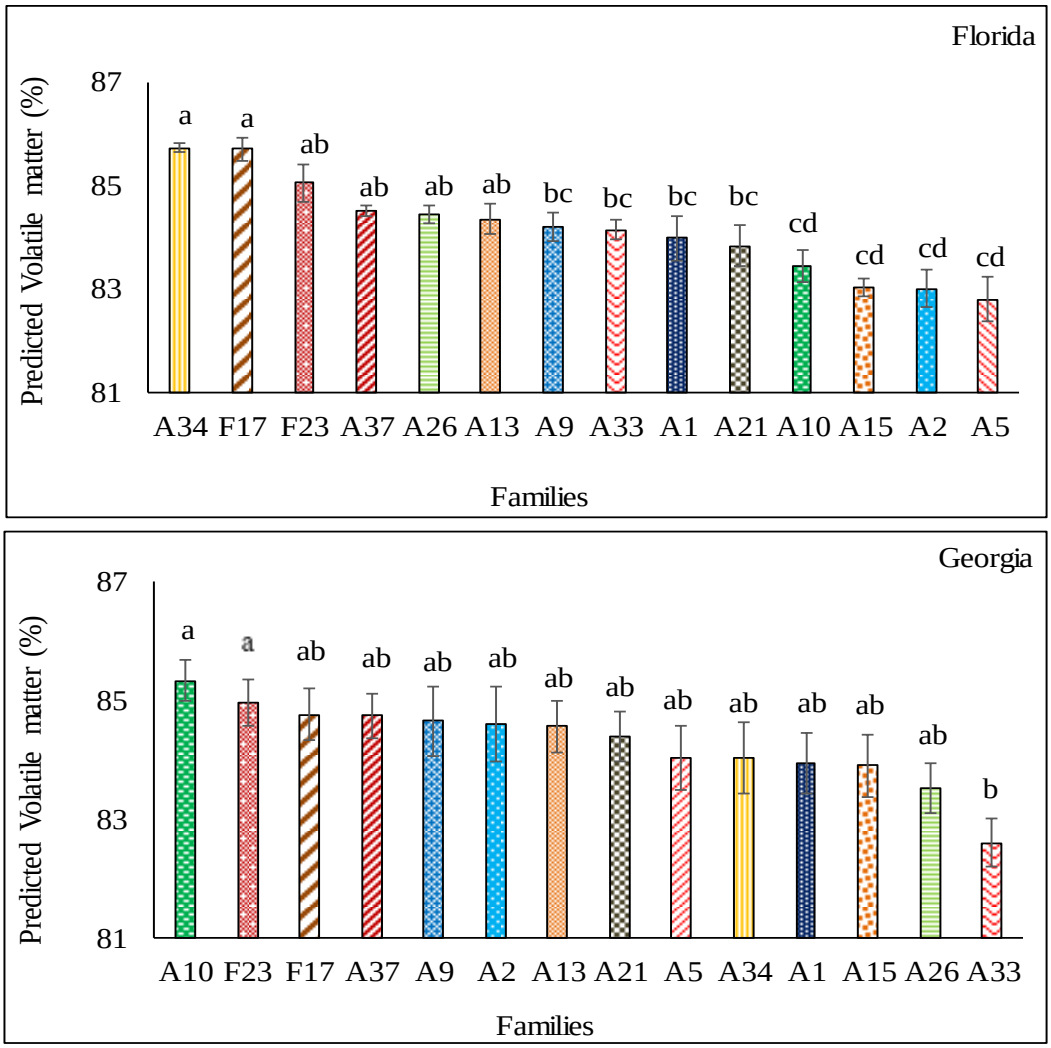

Figure 3.6: Rank of loblolly pine families for volatile matter content. *Bars with different letters are significantly different at 95% confidence level (Tukey's HSD Test).

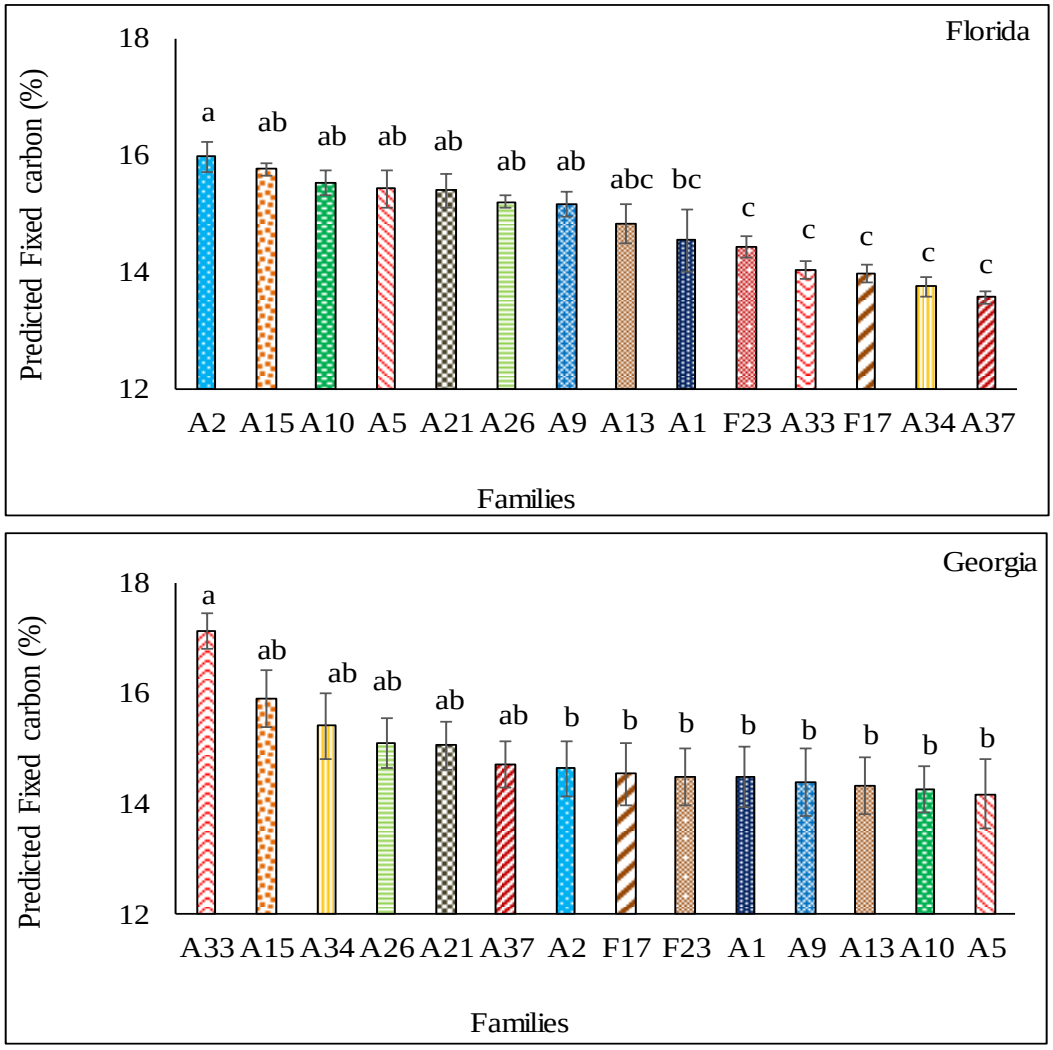

Figure 3.7: Rank of loblolly pine families for fixed carbon content. *Bars with different letters are significantly different at 95% confidence level (Tukey's HSD Test).

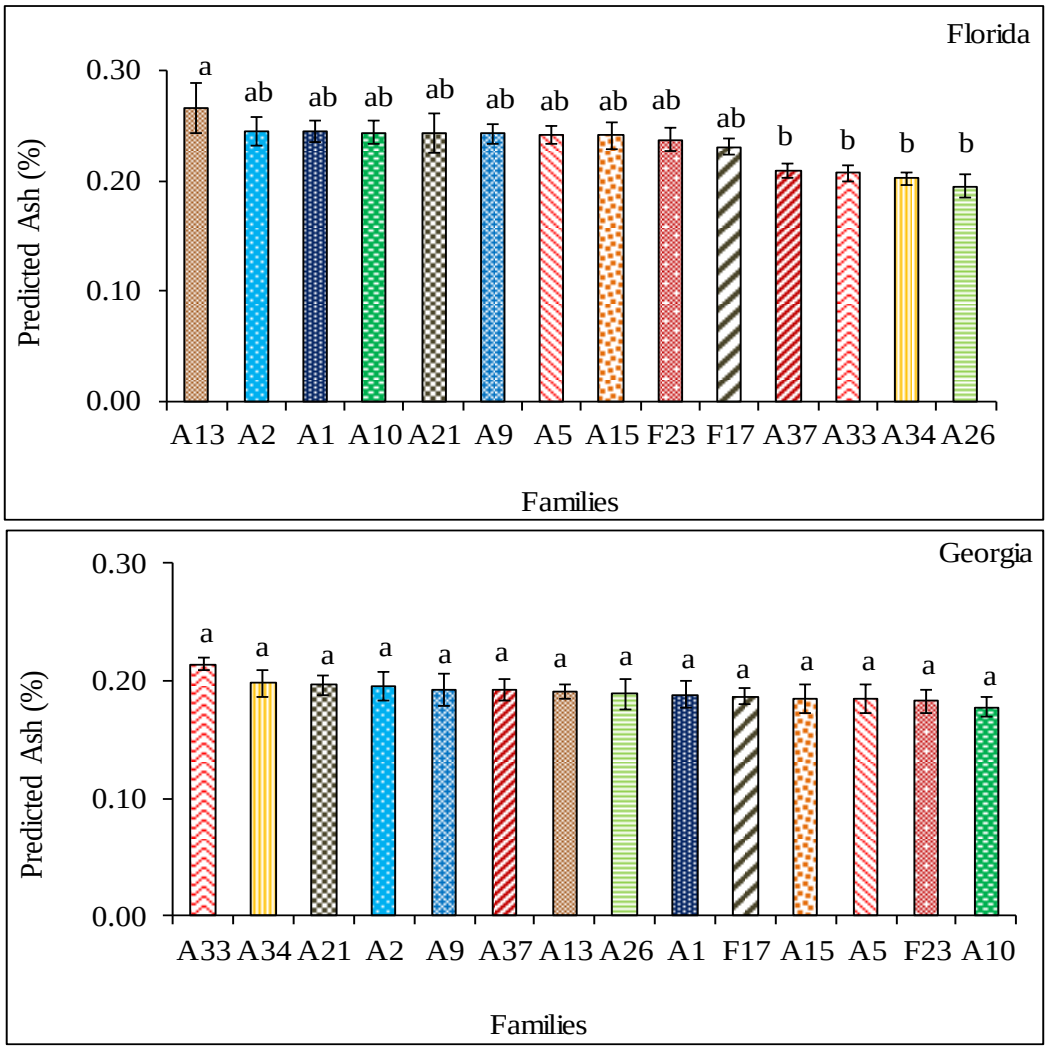

Figure 3.8: Rank of loblolly pine families for ash content. *Bars with different letters are significantly different at 95% confidence level (Tukey's HSD Test).

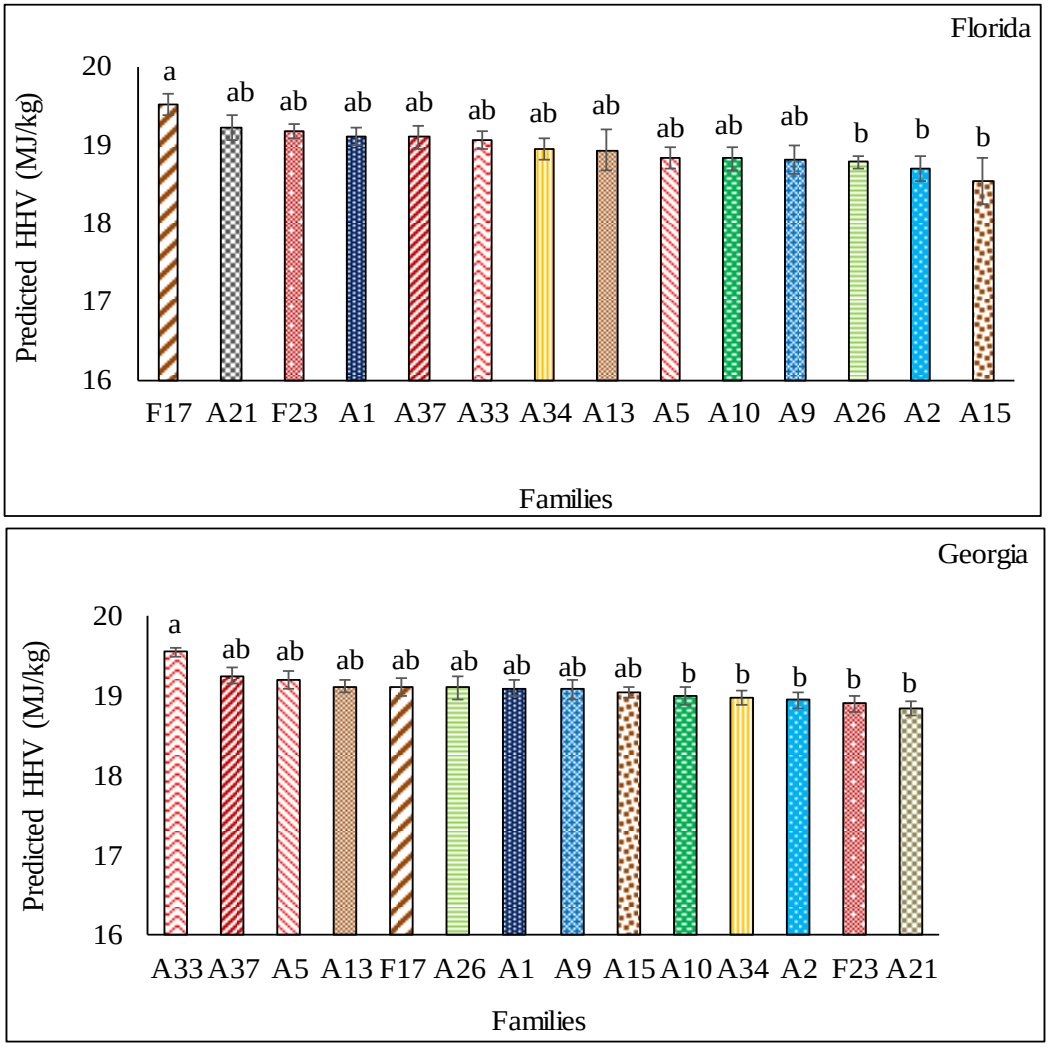

Figure 3.9: Rank of loblolly pine families for energy content. *Bars with different letters are significantly different at 95% confidence level (Tukey's HSD Test).

3.5 Conclusions

NIR-based PLS models were developed to rapidly predict the chemical composition of elite loblolly pine families. The chemistry of this essentially new feedstock will influence the processing and quality of final products that will be derived; whether it be pulp and paper, bio-composites, lumber or biofuel. In addition, the proximate composition and energy content were estimated to ascertain the bioenergy potential of these families for the emerging bioeconomy.

Models were calibrated with 1st-derivative treated NIR spectra using three to five latent variables. For the chemistry, R^2 values for cross-validated models ranged from 0.84 to 0.73. Meanwhile for the bioenergy related properties, the R^2 values of developed models ranged from a high of 0.88 to a low of 0.58. Models for all the nine properties studied had RPD values greater than 1.5, as such, they could be employed in the preliminary screening of the loblolly pine families.

NIR-predicted means of the loblolly pine families for the chemical and bioenergy traits studied are as follows: extractives – 5.5% (SD = 1.1%), lignin – 30.7% (SD = 1.3%), cellulose – 39.7% (SD = 1.4%), glucose – 44.7% (SD = 1.7%), hemicelluloses – 20.5% (SD = 0.4%), volatile matter – 84.2% (SD = 0.8%), fixed carbon – 14.9% (SD = 0.8%), ash – 0.2% (SD = 0.03%) and HHV 19.0 MJ/kg (SD = 0.2 MJ/kg).

The genotype of loblolly pine families affected the chemical, proximate and energy traits studied. The genetic variation detected for cellulose was the largest (*P-value* < 0.0001). Considering that the cellulose content has strong correlations with other properties, selecting and breeding for cellulose can generate some gains. However, the family by site interaction was significant for all understudied properties, indicating the general instability of the elite families across different sites. Further studies with more sites would be helpful in estimating the extent of the family by site

interaction. Such knowledge would be valuable for tree breeders in decisions to plant families with desired traits in certain environments.

Nonetheless, for this study, the cellulose contents of families A1, A26, A15, A2 and A9 were consistently higher on the two sites. High cellulose content implies that these families will be good choices for pulp and paper applications, or for cellulosic ethanol production. On the other hand, the relatively high volatile matter contents of F23, F17, A37 and A9 in spite of the significant interaction between genotype and environment make them good candidates for the production of bio-oil and syngas via thermochemical conversion processes.

Chapter 4 : Identifying Plant Part Composition of Forest Logging Residue using Infrared Spectral Data and Linear Discriminant Analysis[1]

4.1 Abstract

As new markets, technologies and economies evolve in the low carbon bioeconomy, forest logging residue, a largely untapped renewable resource will play a vital role. The feedstock can however be variable depending on plant species and plant part component. This heterogeneity can influence the physical, chemical and thermochemical properties of the material, and thus the final yield and quality of products. Although it is challenging to control compositional variability of a batch of feedstock, it is feasible to monitor this heterogeneity and make the necessary changes in process parameters. Such a system will be a first step towards optimization, quality assurance and cost-effectiveness of processes in the emerging biofuel/chemical industry.

The objective of this study was therefore to qualitatively classify forest logging residue made up of different plant parts using both near infrared spectroscopy (NIR) and Fourier transform infrared spectroscopy (FTIR) together with linear discriminant analysis (LDA). Forest logging residue harvested from several *Pinus taeda* (loblolly pine) plantations in Alabama, USA, were classified into three plant part components: clean wood, wood and bark and slash (i.e. limbs and foliage). Five-fold cross-validated linear discriminant functions had classification accuracies of over 96% for both NIR and FTIR based models. An extra factor/principal component (PC) was however needed to achieve this in FTIR modeling.

Analysis of factor loadings of both NIR and FTIR spectra showed that, the statistically different amount of cellulose in the three plant part components of logging residue contributed to their initial separation. This study demonstrated that NIR or FTIR spectroscopy coupled with PCA and LDA has the potential to be used as a high throughput tool in classifying the plant part makeup of a batch of forest logging residue feedstock. Thus, NIR/FTIR could be employed as a tool to rapidly probe/monitor the variability of forest biomass so that the appropriate online adjustments to parameters can be made in time to ensure process optimization and product quality.

4.2 Introduction

Lignocellulosic biomass is a renewable and largely untapped source of feedstock that can be converted into biopower, liquid and gas fuels, and other biobased products via thermochemical and biochemical conversion pathways. The development of economically and environmentally sustainable sources of biomass can help countries to reduce their dependence of imported fossil fuels and diversify their energy portfolios.

Bioenergy accounts for 4% of the total primary energy consumption in the USA. The country utilizes approximately 200 million dry tons of biomass that is mostly sourced from forestlands for energy. In addition to this volume, some 93 million dry tons of forest biomass that is made up of 68 million dry tons of logging residue and 25 million dry tons of other removal residue are however, currently left onsite annually. It is estimated that the USA has the potential to harvest an amount of 239 to 251 million dry tons of forest biomass on an environmentally and economically sustainable basis by 2030 as new markets and technologies emerge (U. S. DOE 2011; White, 2010).

Forest logging residue is mostly made up of tops, branches and limbs of merchantable trees, salvageable dead trees and small unmerchantable trees. Owing to the different plant part components, there is variability in the composition and quality of this resource. The heterogeneity of forest logging residue can influence important properties that dictate feedstock quality for specific applications. These physical, chemical and compositional characteristics can either setback or boost biomass conversion processes. A good understanding of these properties is crucial in the establishment of a successful biomass conversion facility.

Reprinted from Sensors. doi:10.3390/s16091375

Furthermore, an ability to monitor the properties of the feedstock entering the process and accordingly making the necessary online adjustments in process parameters will be essential in the optimization, quality assurance and cost-effectiveness of conversion technologies in the emerging biofuel/chemical industry. A first step in this direction will be a system that can probe the variability in the composition of a batch of feedstock. Such a system should work in processing/online conditions, provide rapid and accurate analysis of a large number of heterogeneous samples in a non-destructive manner, be easy to use and be cost-effective. Infrared spectroscopy has been shown to have the potential for use in process optimization applications.

Infrared spectroscopy is the measurement of the absorption, transmittance or reflectance of infrared light by a sample. The infrared region is the wavelength range of 780 nm–1 mm (i.e. wave number range of 12,820 cm^{-1} – 10 cm^{-1}) that lies between the visible and microwaves regions of the electromagnetic spectrum. The region is subdivided into near infrared (NIR), mid infrared (MIR) and far infrared (FIR). NIR region lies between the wavelength range of 780 nm to 2500 nm (i.e. 12,820 cm^{-1} to 4000 cm^{-1}); and MIR from 2500 nm to 15380 nm (4000 – 650 cm^{-1}) (i.e. 4000 cm^{-1} to 650 cm^{-1}) (Adapa et al. 2009; Benito et al. 2008; Hsu, 1997). Near infrared spectroscopy (NIR) uses near infrared light to detect overtones and combinational vibrations of the molecular constituents of the material under study; whereas Fourier transform infrared spectroscopy (FTIR) uses mid infrared light to detect primarily functional and fundamental vibrations. Absorption bands that commonly occur in the NIR region and MIR region are overtones and combinations of fundamental vibrations of O-H, C=O, N-H, -COOH, C-H, aromatic C-H groups and S-H functional groups; and can, thus, give the chemical and physical properties of a material (Stuart 2004; Reich 2015; Jorgensen and Goegebeur 2007).

NIR and FTIR have mostly been used for quantitative analysis in the forest products industry and emerging bioeconomy. In an earlier study, we employed NIR in the prediction of some important properties of forest logging residue for bioenergy, fuel and chemical applications (Acquah et al. 2015). It has also been used in the quantitative prediction of cellulose, cellulose crystallinity, hemicellulose, lignin and extractives in wood and lignocellulosic biomass (Poke and Raymond 2006; Jiang et al. 2007; Gierlinger et al. 2002; Sykes et al. 2005; Derkyi et al. 2011). In addition, NIR has successfully been used to model secondary properties that correlate well with wood chemistry including density (Via et al. 2003.), compression wood (McLean et al. 2014), tracheid morphology (Jones et al. 2005), mechanical properties (Acquah et al. 2016), kraft pulp yield (Downes 2009) and energy content (So and Eberhardt 2010). Although NIR does not interact directly with inorganic species, some recent studies have been able to predict the ash content of wood and biomass with varying degrees of success (Fagan et al. 2011; Allison et al. 2009).

Just like NIR, FTIR has been used to predict the density of loblolly pine (Via et al. 2011). Models were also built for the higher heat value (HHV), volatile matter, fixed carbon and ash content of torrefied biomass using ATR-FTIR spectra (Via et al. 2013). Nuopponen et al. (2006) used DRIFT-FTIR spectra to model lignin, cellulose, extractives and density using fifty clones of Sitka spruce, twenty-four Ghanaian hardwoods and twenty Scots pines. In an earlier study, Tucker et al. (2000) used partial least squares models developed with FTIR spectra to quantify the glucose, mannose, galactose, xylose, acetic acid and 5-hydroxymethyl-2-furfural (HMF) of dilute acid pretreated biomass.

With respect to the utilization of infrared spectroscopy for qualitative analysis, different wood species were separated using principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) on NIR spectra (Yang et al. 2015). The researchers were however not

as successful in their attempt to distinguish between wood samples from different locations. Other studies classified wood thermally treated under different conditions (Bächle et al. 2012), herbaceous biomass (Labbé et al. 2008), botanical fractions of cornstover (Ye et al. 2008) and wood-based materials (Tsuchikawa et al. 2003) using NIR coupled with chemometric methods such as soft independent modeling of class analogies (SIMCA), Mahalanobis' generalized distance, Kernel PLS and PCA among others. Similarly, FTIR has been used in the discrimination and classification of wood and wood-based materials (Carballo-Meilan et al. 2014; Chen et al. 2010; Rana et al. 2010; Hobro et al. 2010).

It is hypothesized that, since infrared light is sensitive to the chemical composition of a sample, NIR and/or FTIR spectra can be used to separate out materials that have different chemistry.

The objective of this study was, therefore, to use both near infrared spectroscopy (NIR) and Fourier transform infrared spectroscopy (FTIR) together with PCA and linear discriminant analysis (LDA) in the qualitative classification of *Pinus taeda* (loblolly pine) forest logging residue made up of different plant parts in a comparative study. As mentioned earlier, forest logging residue is a largely untapped resource that can play a key role in the bioeconomy as technologies advance in biomass supply chain logistics and new markets emerge for biofuel and other bioproducts.

4.3 Materials and Methods

4.3.1 Materials

Forest biomass was obtained during harvesting operations on loblolly pine plantations located on several forest tracts in Greenville, Alabama, (31°49'52.583'' N, 86°37'39.241'' W) and Georgiana, Alabama (31°38'24.313'' N, 86°44'21.991'' W). The stands were between 10 and 18 years old, and the diameter at breast height (DBH) of trees ranged from 10 to 20 cm. Biomass was made up of 'Clean wood', 'Slash' and 'Wood & bark'. Clean wood was sampled from either debarked disks that were removed at 5 feet interval along the main stem or from the whole debarked stems of loblolly pine trees. All disks from a tree were combined into a single sample. Slash material is the limbs and foliage of delimbed loblolly pine trees. For 'Wood and bark', material was sampled from the wood and bark of southern pines (mostly loblolly pine) whole stems. Except for the debarked disks that were transported and chipped at Auburn University, AL, all other materials were sampled onsite from chip streams at chipper discharge. A sampling pipe was raised into a chip stream 8–10 times per load. Final representative subsamples were obtained in the lab through coning and quartering. Harvesting, chipping and sampling of biomass spanned several months; from November 2010 to March 2012.

Material used in this study is representative of biomass feedstock that will most likely be used in a bioprocessing plant located in this region. It is typical of feedstock material a manufacturing facility will be acquiring either as pre-commercial thinnings, whole tree utilization of loblolly pine dedicated as an energy crop, or pulpwood chips.

4.3.2 Methods

4.3.2.1 Determination of Chemical Composition and Ash Content

The major chemical components of biomass, i.e. cellulose, hemicellulose, lignin and extractives, were measured via conventional wet chemistry and High Performance Liquid Chromatography (HPLC) (Shimadzu Corporation, Kyoto, Japan). Samples were prepared for analysis by grinding through a 40-mesh screen using a Wiley Mill (Thomas Scientific, model 3383-L10, Swedesboro, NJ, USA).

Extractive content of forest logging residue was determined following NREL/TP-510-42619 and TAPPI T-204. Test samples were extracted in 150 ml of industrial grade acetone for 6 hours in a Soxhlet Apparatus. The amounts of carbohydrates and total lignin were determined as described in NREL/TP-510-42618. After a two-step acid hydrolysis of extractive-free samples, HPLC was employed in the measurement of monomeric sugars (i.e. glucose, xylose, galactose, arabinose and mannose). The sum of all monomeric sugars gave the holocellulose content. Cellulose was computed as $\text{glucose} - (\frac{1}{3}\text{x mannose})$ and hemicelluloses computed as the difference between holocellulose and cellulose. The total lignin was calculated as the sum of acid soluble lignin (ASL) and acid insoluble lignin (AIL). ASL was determined with a UV-Visible spectrophotometer immediately following hydrolysis. Absorbance of a test sample was measured at the recommended wavelength of 240 nm, ensuring that it ranged between 0.7 and 1.0. The ash content of forest logging residue was determined as residue after dry oxidation of test samples at 575 °C, as specified in NREL/TP-510-42622.

For each plant part/group, the chemical composition and ash content was determined using ten of the seventeen samples (i.e. n = 10). Experiments were run in duplicate for each sample.

Knowledge of the chemical composition and ash content of the different plant parts will be useful in the interpretation and elucidation of PC and discriminant analysis.

4.3.2.2 Infrared Spectra Collection

Spectra of forest biomass were acquired with a PerkinElmer Spectrum 400 FT-IR/FT-NIR Spectrometer (Waltham, MA, USA). The FT-IR unit was equipped with a diamond crystal attenuated total reflectance device (i.e. ATR-FTIR) and a torque knob to ensure consistent application of pressure to samples during spectra collection. Samples were ground to pass an 80-mesh screen and oven dried for 4 hours before spectra acquisition. Spectra were collected at 1 cm^{-1} interval from 10,000 cm^{-1} to 4000 cm^{-1} for near infrared and from 4000 cm^{-1} to 650 cm^{-1} for mid infrared. This resulted in 6000 and 3500 data points/variables for NIR and FTIR respectively. A sample was scanned thirty-two times at a resolution of 4 cm^{-1} and averaged into one spectrum for analysis. For NIR, spectrum of a Spectralon standard was taken as the background reference sample every 10 min to correct for potential drifts with time. In the case of FTIR, the background was spectrum of a clear window. Due to the very high dimension of these data sets, spectra were compressed to 10 cm^{-1} interval before exporting to SAS (SAS Institute, Inc. Cary, NC, USA) for further analysis. An earlier study by Via et al. (2011) showed that such compressing/averaging allows the analysis of large data matrices without compromising the integrity of results.

4.3.2.3 Multivariate Data Analysis

Principal Component Analysis (PCA)

PCA is a widely used statistical technique which attempts to explain the covariance structure of data by using a small number of components. These components are linear combinations of the original variables, and often allow for an interpretation and a better understanding of the different sources of variation. PCA is concerned with data reduction. Therefore, it is commonly used for the analysis of high-dimensional data which arise frequently in chemometrics, computer vision,

engineering, genetics and other fields. PCA is, thus, used as a preliminary step of data analysis, followed by further multivariate statistical methods.

As an initial step, PCA was employed to reduce dimension of the data (p = 600 wavelengths for NIR spectra and p = 335 for FTIR spectra). PCA takes a set of correlated variables (as is the case in IR spectra) and transforms them into a smaller set of uncorrelated variables known as principal components (PCs) while maintaining as much of the information in the original data as possible. In other words, assuming that there are n observations X_{ij} on p correlated variables X_1, X_2, …, X_p, i = 1, …, n, j = 1, …, p, PCA finds new uncorrelated Z_1, Z_2,…,Z_p that are linear combinations of X_1, X_2, …,X_p as

$$Z_i = e_{i1}X_{1i} + e_{i2}X_{2i} + \ldots\ldots + e_{ip}X_{pi} \;\&\; Var(Z_i) = \lambda_i,\ i = 1, \ldots, p$$

where λ_is ($\lambda_1 > \lambda_2 > \ldots > \lambda_p$) and $\mathbf{e}_i$ are the eigenvalues and the corresponding eigenvectors of the covariance matrix of data matrix X (n by p). The coefficient, e_{ij} is a measure of the importance of the j^{th} original variable to the i^{th} PC irrespective of the other variables. The coefficients, known as component loadings or eigenvectors are proportional to the correlation between Zs and Xs and can be used in interpreting PCs. The values of the i^{th} principal component are called the PC scores.

The first PC (i.e. Z_1) corresponds to the direction in which the projected observations have the largest variance (i.e. $Var(Z_1) = \lambda_1$, which is the largest eigenvalue). The second component is then orthogonal to the first and again maximizes the variance of the data points projected on it. Continuing in this way produces all the principal components, which correspond to the eigenvectors of the covariance matrix of the data matrix X. In order to determine the number of components, the Proportion of Explained Variance (>99.5%) was used.

For model calibration and validation, a 5-fold cross validation was utilized due to the relatively small sample size (i.e. n = 51). The data set was randomly split into five blocks prior to PCA. Then, PCA was performed on standardized variables by using the correlation matrix of raw NIR and FTIR

spectra by employing the PRINCOMP procedure in SAS (SAS Institute, Inc. Cary, NC, USA). Four blocks were used together at a time as the training data for calibration and the remaining one block as the test data for validation. This was repeated until each of the five blocks was used as an independent test data (i.e. five total runs). As such, for each run, the data used for validation was independent/exclusive of the data used in developing the classification function. For each run, component loadings of the training data set were used to score the test data set.

Linear Discriminant Analysis (LDA)

Scores of retained PCs were used as input data for linear discriminant analysis. LDA is a supervised pattern recognition technique that seeks to find one or several linear functions or discriminants of the dependent variables that can be used to separate out classes/groups. Groups to which observations belong to are known and are defined by the multivariate data structure of its observations. LDA uses these structures to establish rules that allow new unknown samples to be assigned to one or another class (Varmuza and Filzmoser 2009; Johnson and Wichern 2007). Before classification, there is the natural probability (i.e. prior probabilities) that samples belong to one of the labeled groups and after classification there is also a probability (i.e. posterior probabilities) that samples belong to a group. The difference in prior and posterior probabilities enables the allocation of objects to one of the groups. Performances of discriminant functions were evaluated by their error rates or misclassification probabilities. The DISCRIM Procedure in SAS (SAS Institute, Inc. Cary, NC, USA) was used for LDA.

4.4 Results and Discussion

The major chemical composition and ash content as determined in the three plant part components of forest logging residue is presented in Figure 4.1. There were statistical differences (significance level of 0.05) between the plant parts for all properties measured. Clean wood and Slash were the most different, while the chemical make-up of Wood and bark was generally more like Clean wood. For instance, Slash had the highest amount of lignin (44%), with Wood and bark and Clean wood having 36% and 34% respectively. Additionally, the 2% ash in Slash was statistically higher than the 1.6% in Wood and bark and 0.4% in Clean wood.

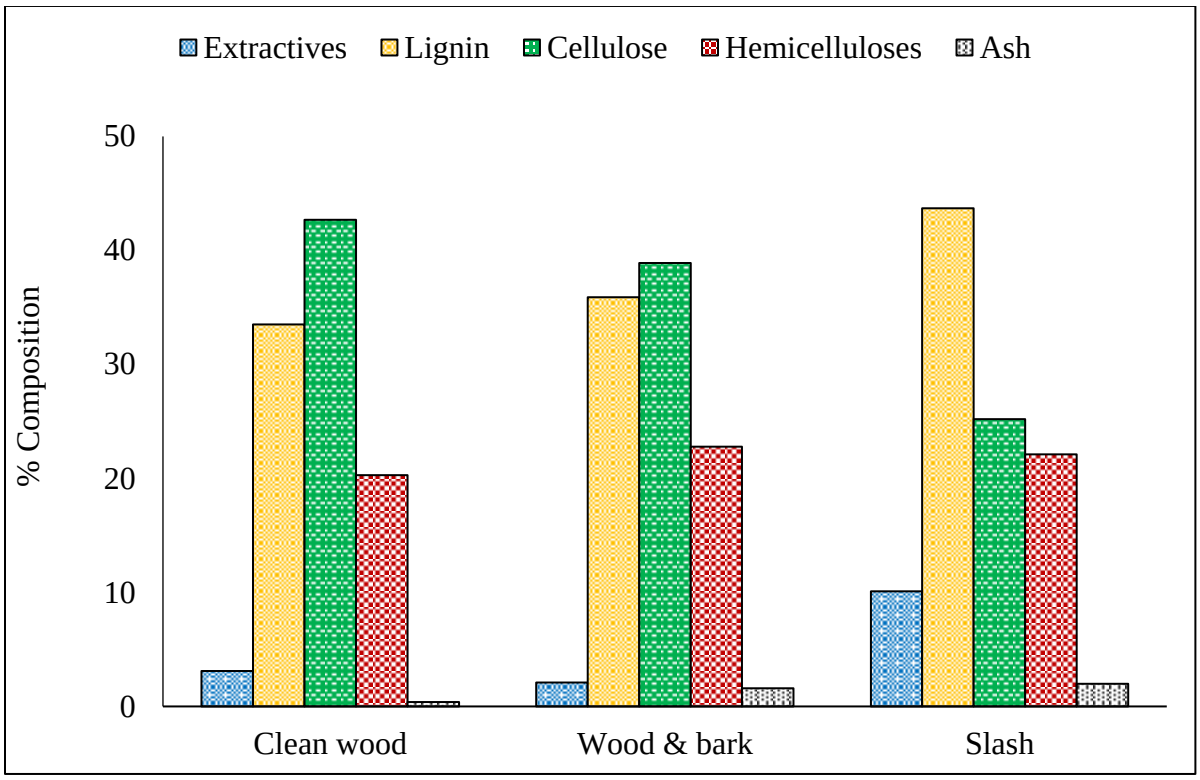

Figure 4.1: Chemical composition and ash content of forest logging residue.

4.4.1 Infrared Spectra

Averaged NIR and FTIR spectra of the three plant part components of forest logging residue used in this study are presented in Figure 4.2. There was a general trend in the absorbance of near infrared and mid infrared by the plant parts. There were however variations in the intensity of light

absorbed. For both NIR and FTIR, Slash absorbed the most. Clean wood absorbed the lowest amount of energy for a good portion of the near infrared region, but in the mid infrared region, its absorbance values were slightly higher or lower than the values for Wood and bark.

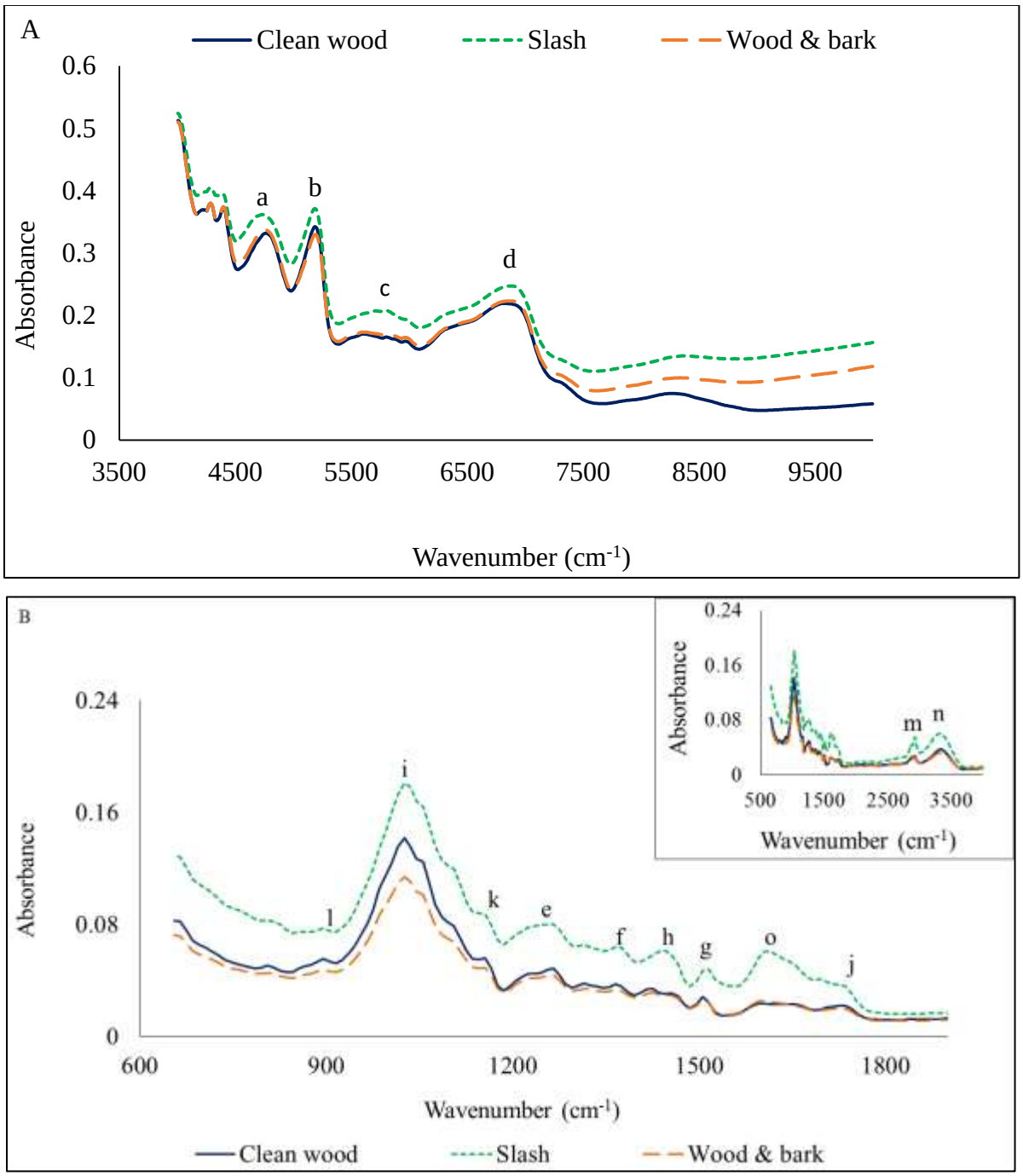

Figure 4.2: Raw NIR (A) and FTIR (B) spectra of the different types of forest logging residue.

Large baseline shifts noted in the 7100 to 10,000 cm^{-1} region might have resulted from the different ash contents of the three biomass types, Figure 4.1 (Sanderson et al. 1996). Although infrared light interacts directly with only organic compounds in materials, these interactions may be influenced by the presence of their associated inorganic species (Ono et al. 2003).

In the near infrared region (Figure 4.2A), the absorbance peaks occurring from 4000 cm^{-1} to 5000 cm^{-1} are as a result of the interactions of O-H, C-H and N-H functional groups interacting with one another (i.e. combination bands). Peaks have also been ascribed to specific chemical constituents of lignocellulosic biomass: (a) 4765 cm^{-1} results from O-H and C-H stretching and deformation vibration of cellulose (and xylan); (b) 5205 cm^{-1} is due to the asymmetric stretching and/or deformation of O-H in water; and (c) 5845 cm^{-1} credited to the first overtone stretching of C-H in hemicelluloses. In addition, the peak at (d) 6875 cm^{-1} has been attributed to the first overtone of O-H stretching of phenolic groups in lignin (Derkyi et al. 2011; Schwanninger et al. 2011).

As in the near infrared region, peaks arise in the mid infrared region due to the presence of functional groups in biomass. Although this region ranges from 4000 to 600 cm^{-1}, the fingerprint region (1800 to 600 cm^{-1}) is usually used for analysis because it contains the most spectral information pertaining to the molecular/chemical composition of a material (Figure 4.2B). According to the literature, bands at (e) 1270 cm^{-1}; (f) 1365 cm^{-1}; (g) 1505 cm^{-1} and (h) 1435 cm^{-1} have been associated with lignin; (e) and (f) result from guaiacyl ring breathing and syringyl ring breathing respectively, whereas (g) is due to the aromatic skeletal vibration with C=O stretch. For the carbohydrates, C=O stretch of unconjugated ketones mostly in hemicellulose generate bands at (i) 1025 cm^{-1} and (j) 1735 cm^{-1}; whereas the peaks at (k) 1154 cm^{-1} and (l) 895 cm^{-1} result respectively from C-O-C stretching and P-chains of cellulose. Furthermore, the peak at (m) 2935 cm^{-1} outside the fingerprint range (Figure 4.2B) have been associated with the bending and stretching of C-

H, as well as its aromatic ring vibration in lignin, while that occurring at (n) 3345 cm^{-1} is due to N-H stretching. Spectra of Slash had a very prominent peak at (o) 1635 cm^{-1} compared to Clean wood and Wood and bark. This has been attributed to C-O stretching of conjugated or aromatic ketones and/or C=O stretching vibration in flavones (Allison et al. 2009; Nuopponen et al. 2006; Chen et al. 2010; Hobro et al. 2010).

4.4.2 Principal Component Analysis

Partial results from PC analysis showing the first ten PCs are presented in Table 4.1. A preset criteria for the number of PCs to include in further analysis was that the eigenvalue of a PC should be more than 0.7 (i.e. PCA on the correlation matrix) and the cumulative variance should be greater or equal to 99.5%. In addition, the Scree Test Criterion was used. A Scree diagrams plots λ_i against i for i = 1,…,q; and λ is the eigenvalues. The point at which the curve begins to straighten out indicates a cut-off point. Based on Table 4.1 and the Scree plots, the first six PCs were tentatively retained for linear discriminant analysis.

The first six PCs out of the possible 600 for NIR and 335 for FTIR were able to account for over 99.5% of the total variation in the data. For NIR, PC 1 and PC 2 accounted for 76% and 16%, respectively, of the spectra data; in the case of FTIR, they were 81% and 15%, respectively.

Employing PCA as a preliminary classification tool, scores of the retained PCs were plotted against each other. In Figure 4.3, a graph of the scores of raw near infrared spectra for PC 1 and PC 2 is presented. Separation was better along PC 1; Clean wood clustered furthermost from Slash, with Wood and bark in between the two classes.

Table 4.1: Eigenvalues of the correlation matrix.

	NIR			FTIR		
PC	Eigenvalue	PVE (%)	CV (%)	Eigenvalue	PVE (%)	CV (%)
1	454.4 (4.8)	74.1	74.1	271.0 (5.2)	80.9	80.9
2	95.4 (3.7)	15.9	92.0	49.9 (4.0)	14.9	95.8
3	41.8 (4.5)	4.0	98.9	9.6 (3.6)	2.9	98.7
4	3.6 (1.3)	0.6	99.5	1.8 (0.3)	0.6	99.2
5	1.3 (0.1)	0.6	99.7	1.3 (0.2)	0.4	99.6
6	0.85 (0.08)	0.14	99.88	0.29 (0.04)	0.09	99.68
7	0.36 (0.03)	0.06	99.94	0.16 (0.03)	0.05	99.73
8	0.13 (0.01)	0.02	99.96	0.09 (0.01)	0.03	99.76
9	0.06 (0.01)	0.01	99.97	0.07 (0.01)	0.02	99.78
10	0.04 (0.01)	0.01	99.98	0.06 (0.01)	0.02	99.80

Note: Values are the means of the five folds used as training data sets; SD values in brackets; PVE is the Proportion of Variance Explained and CV is the Cumulative Variance.

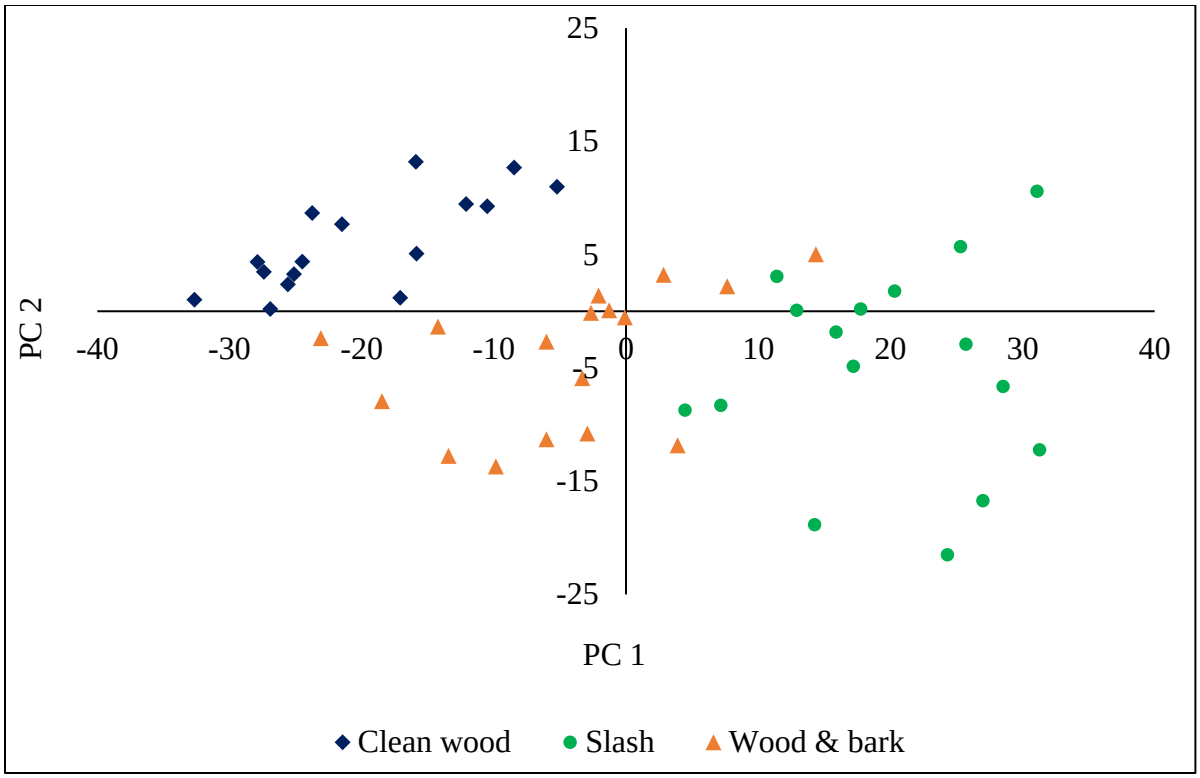

Figure 4.3: NIR scores plot of PC 1 versus PC 2.

According to the loadings plot of PC 1 (Figure 4.4), cellulose content (peaks at 4605 and 7325 cm^{-1}) was a good initial separator of the different plant parts as it had higher coefficient values. This could be backed by results from the conventional chemical analysis. Reviewing Figure 4.1, Clean wood had the highest percentage of cellulose (43%), followed by Wood and bark (39%), then Slash (25.2%). Thus, on the PC 1 axis, the three biomass types separated from left to right due to decreasing cellulose content. Other significant coefficients noted in NIR spectra loadings that contributed to the classification of the three groups of forest logging residue were at 7095 cm^{-1} in PC 3, which is attributed to the phenolic groups in lignin and/or extractives and a peak at 5835 cm^{-1} in PC 4 occurring due to C-H stretching in hemicelluloses. Again, results from PC analysis were buttressed and elucidated by chemical composition determined via conventional laboratory methods.

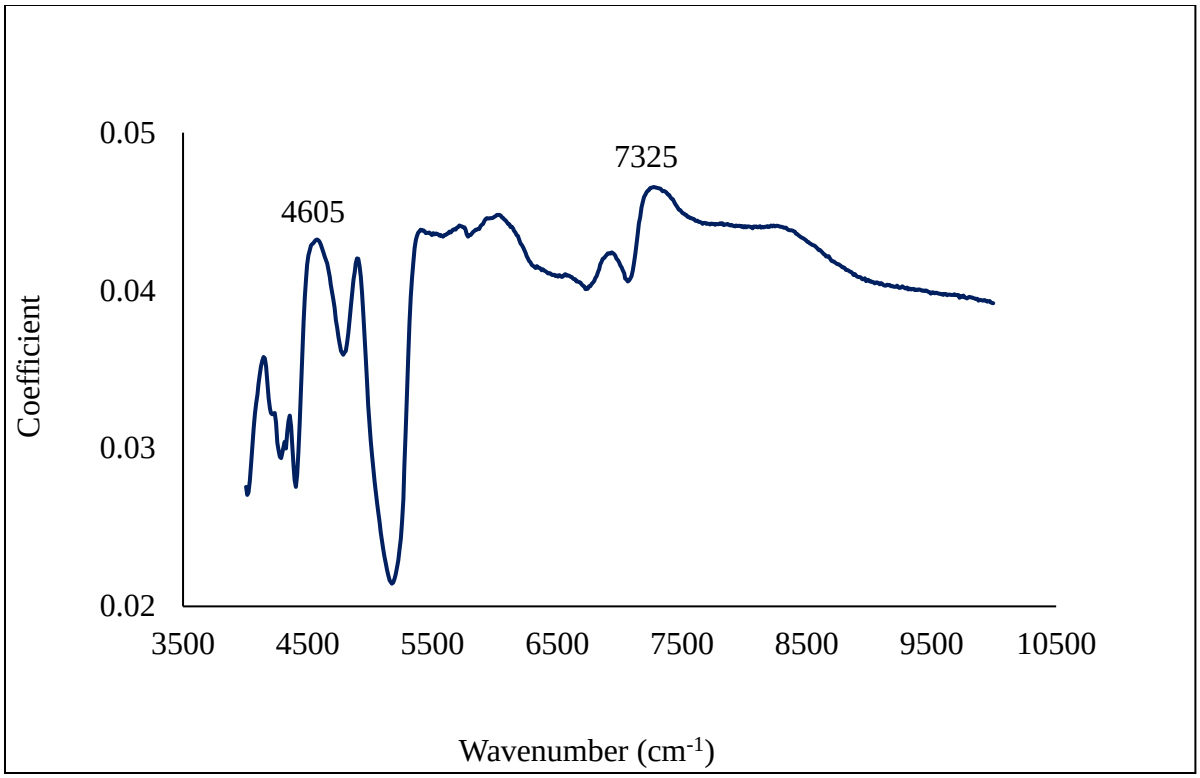

Figure 4.4: NIR loadings plot of PC 1 showing significant peaks.

In the case of FTIR, a plot of PC1 against PC5 (Figure 4.5) gave the best initial separation with better separation along PC 1. This was however not as distinct especially between Clean wood and Wood and bark as was seen in the scores plot of NIR spectra. A characteristic cellulose peak occurring at 1725 cm^{-1} was again observed in the loadings of PC 1, Figure 4.4. In addition, the large loadings coefficients of 1485 cm^{-1} suggests that vibrations attributed to both lignin and polysaccharides also contributed to the initial distinction of different biomass types in the mid infrared region.

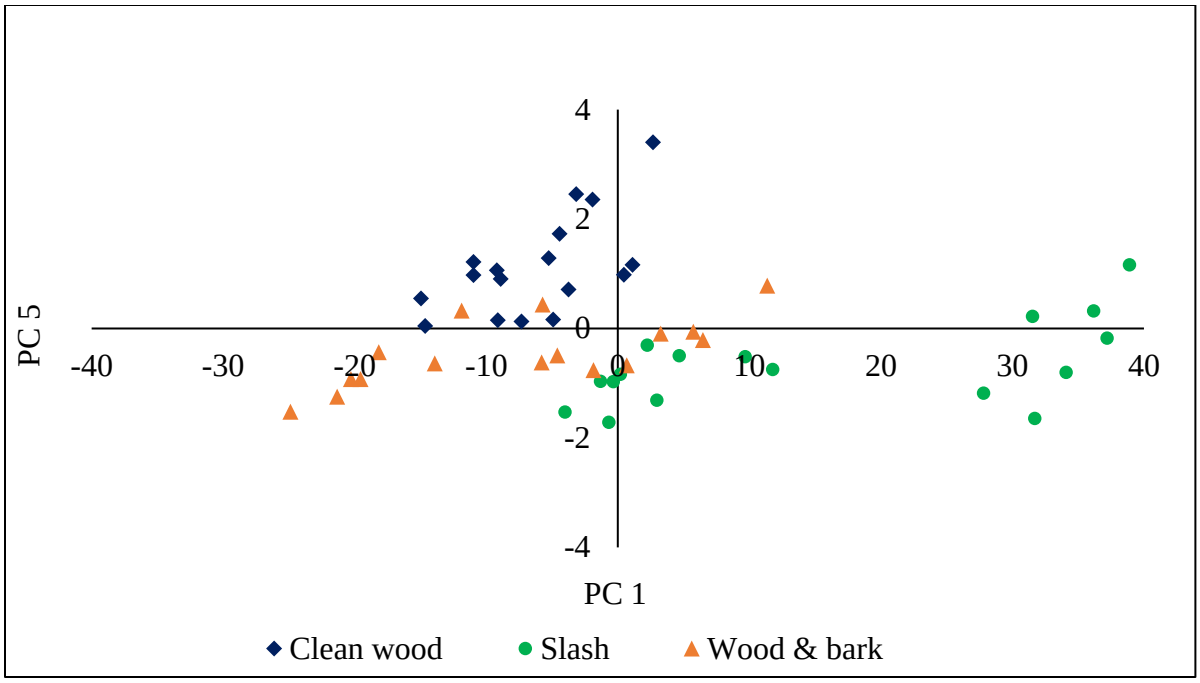

Figure 4.5: FTIR scores plot of PC 1 versus PC 5.

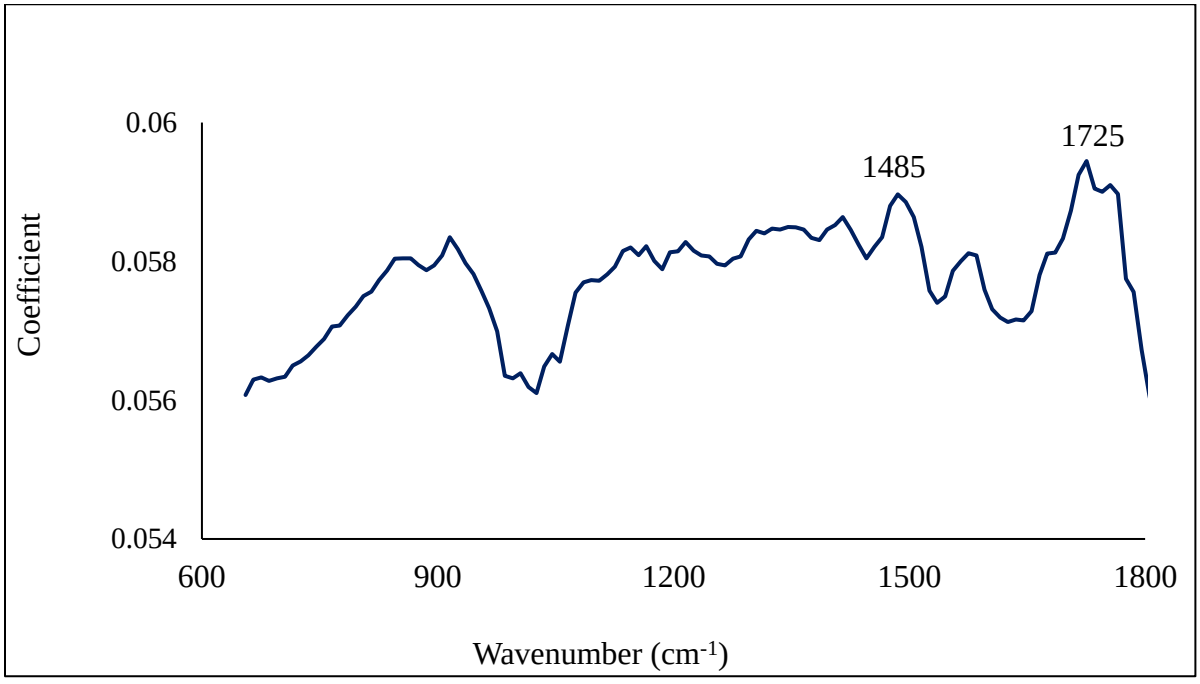

Figure 4.6: FTIR loadings plot of PC 1 showing significant peaks.

4.4.3 Linear Discriminant Analysis (LDA)

The first six PCs retained (chosen based on the eigenvalue, variance explained and Scree Test Criteria) were used in LDA. Examining the effect the inclusion of PCs had on errors associated with classification (Figure 4.7), the discriminant functions (Table 4.2) developed with four and five PCs were chosen as the optimum for NIR and FTIR, respectively. These selections were made because the difference in errors for the training data set and test data were the least. Furthermore, standard deviations of the five folds used in model calibration were smallest for the selected number of PCs.

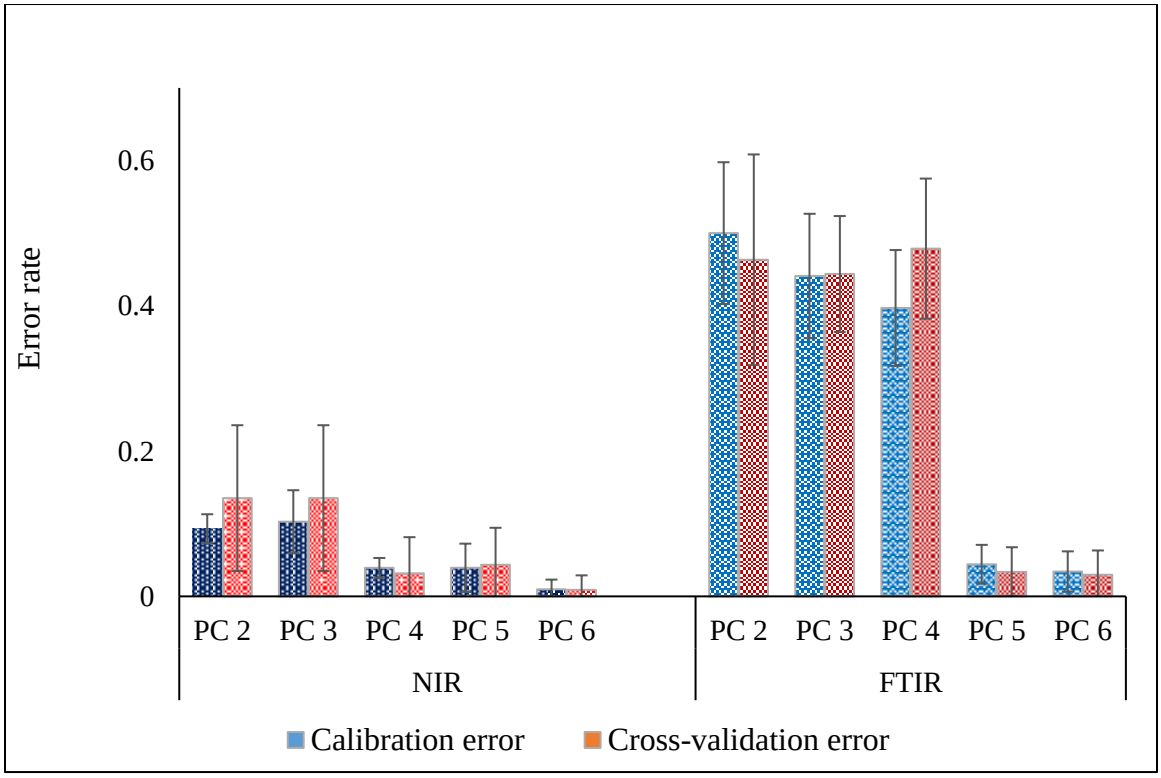

Figure 4.7: Effect of changing number of PCs on classification error.

Calibration errors were computed using the Lachenbruch's Holdout procedure, whereby all samples except the first sample were used in building the discrimination function to classify the first, then the second is held out and the process repeated until all samples have been used as single-element test sets in a fold (Brereton 2007). Unlike errors generated in five-fold cross-

validation of the test set, errors estimates by the Lachenbruch's Holdout procedure (i.e. a leave-one-out cross-validation technique) may be overoptimistic due to the exclusivity of the one-sample test data. Up until inclusion of the fifth PC, errors associated with FTIR-based discriminant functions were very high. For instance, while the model built from NIR spectra using four PCs had only 4% and 3% as misclassification errors of cross-validation for the respective training data and test data, errors for FTIR were 40% and 48% respectively. Studies that have been conducted to compare NIR and FTIR in the quantitative or qualitative analysis of lignocellulosic biomass and plant-based materials have reported differing results mostly in favor of the latter, albeit slightly (Via et al. 2013; Lupoi et al. 2014; Olale et al. 2013; Sankaran and Ehsani 2013; So et al. 2012).

Table 4.2: Linear discriminant functions used for classifying plant part components.

	NIR			FTIR		
Variable	CW	SL	WB	CW	SL	WB
Constant	-6.13 (0.71)	-9.29 (3.51)	-2.14 (0.36)	-6.89 (1.73)	-10.34 (3.96)	-2.02 (0.51)
PC1	-0.39 (0.08)	0.51 (0.17)	-0.11 (0.04)	-0.4 (0.08)	0.56 (0.16)	-0.14 (0.03)
PC2	0.42 (0.03)	-0.43 (0.04)	0.02 (0.06)	-0.16 (0.06)	0.2 (0.07)	-0.05 (0.04)
PC3	-0.04 (0.06)	0.07 (0.05)	-0.03 (0.01)	0.21 (0.23)	-0.41 (0.35)	0.22 (0.13)
PC4	1.23 90.35)	-2.35 (0.66)	1.07 (0.19)	-3.06 (1.74)	3.94 (1.52)	-1.12 (0.39)
PC5	-	-	-	6.56 (1.66)	-7.59 (2.84)	0.78 (0.41)

Note: CW means Clean wood; SL means Slash; WB means Wood and bark.

Better performance of FTIR relative to NIR have been attributed to the fundamental vibrations in the MIR region as opposed to overlapping and weaker overtone and combination bands observed in the NIR region. The abundance of absorption bands especially in the fingerprint region of the former make identification/ qualification of molecular structures easier. On the other hand,

chemometric techniques are usually required in order to extract relevant information in the latter (Stuart 2004).

The generalized squared distance (Table 4.3) gives an indication of the degree of separation between classes in space. A new/unknown sample is classified into a group if it is similar enough to the other members, otherwise it is rejected. According to Table 4.4, Slash was most distinct from Clean wood and less so from Wood and bark. These results are in agreement with that from PCA, as can be seen in Figures 4.3 and 4.5 when PC scores were plotted.

Table 4.3: Generalized squared distances of the three plant part components of forest logging residue.

	NIR			FTIR		
From/Into	CW	SL	WB	CW	SL	WB
CW	2.2 (0.3)	52.5 (12.9)	10.6 (1.6)	2.2 (0.3)	59.9 (14.8)	11.1 (1.7)
SL	52.5 (12.9)	2.2 (0.3)	28.2 (8.2)	59.9 (14.8)	2.2 (0.3)	30.2 (4.3)
WB	10.6 (1.6)	28.2 (8.2)	2.2 (0.3)	11.1 (1.7)	30.3 (4.3)	2.2 (0.3)

Note: CW means Clean wood; SL means Slash; WB means Wood and bark.

From the error count estimate in Table 4.4, the performance of developed functions in predicting the class of independent test samples were computed.

Table 4.4: Five-fold cross-validation summary of error count estimates (%) for plant part component.

	NIR				FTIR			
	CW	SL	WB	Total	CW	SL	WB	Total
Rate	0%	3.3%	6.7%	3.2%	0%	3.3%	8.3%	3.4%

Note: Values are averages of the five groups of test samples used in validation; CW means Clean wood; SL means Slash; WB means Wood and bark.

As seen in Table 4.5, linear discriminant functions developed with NIR and FTIR spectra were able to classify the plant part components of logging residue with over 96% overall accuracy. Clean wood was the easiest to identify, while Wood and bark generally had the highest misclassification

rate. This was to be expected considering the plant part makeup of the three materials studied. Moreover, from the chemical and ash content analysis, it was determined that the properties of Wood and bark were more similar to the other two plant part components.

Table 4.5: Classification rates for forest logging residue.

	NIR					FTIR				
	Plant part component									
	CW	SL	WB	Total	%CC	CW	SL	WB	Total	%CC
CW	17	0	0	17	100	17	0	0	17	100
SL	0	16	1	17	96.7	0	16	1	17	96.7
WB	1	0	16	17	93.3	2	0	15	17	91.7
% Total Accuracy					96.7 (3.3)					96.1 (4.2)

Note: Classification rates calculated based on error count estimates in Table 4.4. SD values in brackets. CW means Clean wood; SL means Slash; WB means Wood and bark; %CC is the % correct classification.

4.4.4. Remarks

Ideally, samples used in model validation should be independent of the training dataset. However, this cannot always be the case due to limited resources. When sample size is small, researchers have employed cross validation (CV) to test the performance of calibration models instead of splitting up the data into a single training set and test set. A commonly used technique is the leave-one-out CV method. In this procedure, n − 1 samples are used in training a model that is validated with the held out sample. This is repeated n times until each observation has been used as validation data. The advantage of this approach is that, it uses the maximum available data in both model training and validation. However, due to the exclusivity of the one-sample test data, the errors estimates may be overoptimistic. To overcome this potential problem, the current study

opted for a five-fold cross validation. This ensured that a test dataset comprised of observations with varying backgrounds—for instance, different age, DBH or site. Additionally, taking the average of five repetitions instead of just one experimentation gives a significantly better estimate of the errors.

Another strength of the developed classifier lies in the range of samples used. Materials used in this study are representative of biomass feedstock that will most likely be used in a bioprocessing plant located in this region. Loblolly pine (and southern pine on one site) that were 10 to 18 years old, with a DBH range of 10 – 20 cm from several forest sites were used. This is typical of feedstock material a manufacturing facility will be getting either from pre-commercial thinnings, loblolly pine dedicated as an energy crop or pulpwood chips. Thus, models constructed in this study are robust and will perform well in classifying similar feedstock in this region.

The aim of this study was to demonstrate that NIR and FTIR can be used to rapidly identify what a batch of feedstock is made up of, as this, as is known will influence the chemical composition. A traditional way to do this is probably by visual inspection. Compared to this, NIR/FTIR has a higher throughput, and will have fewer errors, especially for comminuted feedstock. With this information, on-time adjustments could be made in the process parameters so that product yield and quality can be optimized/assured. Such information could also be used in future feedstock acquisition.

Any processing plant employing NIR/FTIR as a classification tool will first have to calibrate their system with samples that is within the range of materials characteristic to their locality. Apart from this qualitative probing/monitoring, a facility's system could also be trained to provide quantitative information, such as the cellulose, lignin, ash or energy content of feedstock coming into the process.

4.5 Conclusions

This study demonstrated that NIR or FTIR spectroscopy coupled with PCA and LDA has the potential to be used as a high throughput tool in classifying the plant part makeup of a batch of forest logging residue feedstock. Peaks noted at 4605 and 7325 cm^{-1} (i.e. NIR) in the loading plot of PC 1 suggested that the significantly different amount of cellulose contributed to the initial separation of the different plant parts. In the mid infrared region (i.e. FTIR) preliminary separation was made possible due to the varying concentrations of lignin and polysaccharides. Both NIR and FTIR based linear discriminant functions had very good classification accuracies (i.e. 94%) even though an extra variable/PC was needed to achieve this with FTIR modeling.

Applications for this study include its use as a rapid tool to probe/monitor the variability of forest logging residue so that the appropriate online adjustments to parameters can be made in time to ensure process optimization and product quality.

Chapter 5 : Nondestructive Prediction of the Properties of Forest Biomass for Chemical and Bioenergy Applications using Near Infrared Spectroscopy[2]

5.1 Abstract

Forest biomass will play a key role as a feedstock for bioproducts as the bioeconomy develops. Rapid assessment of this heterogeneous resource will help determine its suitability as feedstock for specific applications, aid in feedstock improvement programs and enable better process control that will optimize the biorefinery process. In this study, near infrared spectroscopy coupled with partial least squares regression was used to predict important chemical and thermal reactivity properties of biomass made up of needles, twigs, branches, bark and wood of *Pinus taeda* (loblolly pine).

Models developed with the raw spectra for property prediction used between 3 and 8 factors to give R^2 values ranging from a low of 0.34 for higher heat value (HHV) to a high of 0.92 for volatile matter. Pre-treating the raw spectra with 1st derivatives improved the fit statistics for all properties (i.e. min – 0.57, max – 0.92; with 2 or 3 factors). The best performing models were for extractives, lignin, glucose, volatile matter and fixed carbon ($R^2 > 0.81$). This study provided the capacity to predict multiple chemical and thermal/energy traits from a single spectrum across an array of materials that differ considerably in chemistry type and distribution. Models developed should be able to rapidly predict the studied properties of similar biomass types. This will be useful in rapidly allocating feedstocks that optimize biomass conversion technologies.

5.2 Introduction

The use of biomass as an alternative source of energy, fuels and chemicals derived from fossil fuel will reduce our dependence on non-renewable resources and also minimize net greenhouse gas emissions. Biomass is mostly sourced from forestry and agricultural sources in the USA, and other countries. It is estimated that some 93 million dry tons of forest biomass is available per year; out of which 73% is logging residues comprising of tops, branches and limbs, salvageable dead trees and small trees (U.S. DOE 2011). Even though most of this resource is currently left on site, a significant portion of it will become economically feasible for removal as new markets for bioenergy and bioproducts emerge (Greene et al. 2011). In addition to the conventional combustion of forest biomass for heat and power, it can also be thermochemically converted into syngas or bio oils that can further be refined into high value fuels and chemicals; biochemically converted into ethanol, artificial flavors and phytochemicals, or used as supplementary fiber in the manufacture of pellets and composite board products (Acquah 2010).

Forest biomass is a heterogeneous feedstock that has its properties dependent on such factors including species and plant-part composition, harvesting time and procedures, region and climate (Acquah 2010). Like any other resource, the properties of forest biomass must be well understood. The aim of this research is to have a system that will give complete and systematic characterization of this potential resource to determine its suitability as a feedstock for specific applications, aid in feedstock improvement programs and enable better process control that will optimize the biorefinery process.

---

[2] Reprinted with permission from the Journal of Near Infrared Spectroscopy. doi: dx.doi.org/10.1255/jnirs.1153

Conventional methods used to measure properties of biomass have been time consuming, expensive and usually destructive; thus the need for rapid and cost-effective analytical tools is necessary. One such tool is near infrared reflectance (NIR) spectroscopy. The most common absorption bands that occur in the NIR region are overtones and combinations of fundamental vibrations of C-H, O-H, C=O, N-H, -COOH, aromatic C-H groups and S-H functional groups in the mid infrared region, and can thus give the chemical and physical properties of a sample. In simpler terms, the chemical finger print (in the form of a spectrum) of a sample is taken by NIR at a specific point in time. NIR spectra typically exhibit broad ill-defined overlapping bands of the chemical information of an analyte. Mathematical, statistical and computer science methods, often called chemometrics, are used to extract relevant information and minimize irrelevant signals from the chemical measurement data during quantitative or qualitative analysis (Mark and Campbell 2008; Reich 2005). According to Varmuza and Filzmoser (2009), in spite of the broad definition of chemometrics, the most vital aspect of it is the application of multivariate analysis (MVA) to chemistry-relevant data. MVA -- such as partial least squares (PLS) regression, principal components regression (PCR) and multiple linear regression (MLR)-- is useful in analyzing and structuring NIR spectra and conventionally acquired chemical data into empirical mathematical models that are capable of predicting properties.

In recent times, there has been a lot of work on the potential use of NIR to estimate the properties of biomass. Several researchers used NIR to monitor the moisture content (MC) of wood samples and distinguish between capillary condensed and adsorptive water (Defo et al. 2007; Lestander and Rhen 2005). Cooper et al. *(*2011) developed partial least squares (PLS) regression models that predicted the average MC within less than 5% total error for samples with a range between oven dry to 80%. They reported that NIR was less capable of detecting differences in the

quantity of free water in wood as compared to differences of MC below the fiber saturation point. They were also able to partially differentiate the difference in density between earlywood and latewood with NIR and recommended that a large area should be illuminated especially when scanning is being done on the tangential surface so as to ensure a better representation of both earlywood and latewood.

Others developed NIR models to predict the chemical composition including extractives, cellulose, hemicellulose and lignin content of wood (Derkyi et al. 2011; Nkansah et al. 2010; Sykes et al. 2005; Gierlinger et al. 2002). Nkansah et al. (2010) used diffuse reflectance NIR to predict the extractives, holocellulose and lignin content of yellow poplar solid wood blocks (19 x 19 x 50 mm) whereas Sykes et al. (2005) used it to predict the properties of wafers (200 µm) sectioned from loblolly pine from two sites. The former group noted that the models they developed with the full spectra (4000 to 12500 cm^{-1}) had slightly better predictive ability as compared to models they built with reduced spectra (5555 to 7695 cm^{-1}). They witnessed R^2 values of 0.84, 0.68 and 0.64 for extractives, holocellulose and lignin respectively. Sykes et al. (2005) on the other reported low R^2 of 0.37 for lignin content. The authors attributed this to the lower variation of lignin content in wood coupled with larger errors during wet chemistry determination. They also found out that models developed from samples at one site could be used to predict the cellulose content of wood from another site, even though the R^2 values were lower than those for the individual site predictions. Poke and Raymond (2006) sought to increase the rapidity and cost-effectiveness of using NIR to predict the chemical composition of *Eucalyptus* using 20 x 20 mm test stripes. For comparison, they used both solid wood and ground wood for developing their models. The authors reported that the models built from ground wood samples did not do as well as models developed from solid wood. For instance, ground wood models gave R^2 of 0.27 and 0.63 for Klason lignin and cellulose

respectively. However, for the solid wood models, the R^2 value was 0.78 for Klason lignin and 0.88 cellulose. Jiang et al. (2014) investigated the effect of particle size on NIR spectra quality and model precision using lumber, 3-mm chips and wood flour ground to pass an 80-mesh screen. They concluded that size reduction improved the stability of spectra, and consequently, the predictive diagnostics of models developed. Hein et al. (2010) also had a similar conclusion, but they added that the difference between lumber and milled wood was more significant than the difference between particle sizes of milled wood.

NIR has also been utilized in the prediction of secondary traits that correlate to wood chemistry such as density and mechanical properties (Via et al. 2003), microfibril angle (Hein et al. 2010), compression wood (McLean et al. 2014), pulp yield (Downes et al. 2009) and energy content (So and Eberhardt 2010). When So and Eberhardt (2010) used reflectance NIR to predict the higher heat value (HHV) of *Pinus palustris* considering the effect of lignin content and extractives, models they developed to predict the HHV of unextracted wood samples performed better than those developed using extracted samples. Their plots of the regression coefficients showed similar peaks for the HHV and extractives content, implying that the two properties have similar molecular features. Not as much work has been done on the use of NIR to predict the ash content of wood, and models generally give low coefficients of ash determination. These poor results have been attributed to the fact that NIR does not interact directly with the species that form ash, e.g. calcium, potassium and silica.

Most of the literature did work on clean wood, an expensive resource that is currently used for higher end products. However as already stated, biomass from the forest that will be feasibly available as feedstock for the emerging bioenergy and bioproducts market is the 96 million dry tons of logging residues and other removals. Thus, the objective of this study was to use NIR coupled

with PLS to rapidly estimate the monomeric sugars, cellulose, hemicelluloses, holocellulose, lignin, extractives, HHV, ash, volatile matter and fixed carbon of *Pinus taeda* (loblolly pine) forest biomass. Loblolly pine was used because it is the most important tree species in the USA accounting for some 30 million acres in plantations in the southeastern US alone (Conner and Hartsell 2002).

5.2 Materials and Methods

5.3.1 Material

Logging residues were acquired from several loblolly pine plantations in the Greenville area in the State of Alabama. Residues were made up of whole tree, wood and bark, slash (i.e. limbs and foliage) and clean wood chips, Figure 5.1. A total of forty biomass sets (ten from each biomass type) were sampled.

Figure 5.1: Loblolly pine logging residues: (a) whole, (b) wood & bark, (c) slash and (d) wood.

5.3.2 Methods

5.3.2.1 Conventional Laboratory Methods

Laboratory experiments following conventional standards were used to determine the chemical composition, energy content and thermal reactivity of biomass. Test samples ground with a Wiley Mill to pass through a 40-mesh screen were used in all experiments which were ran in duplicates (ASTM D-1106).

Wet laboratory chemistry was used to determine the chemical composition of logging residues. Extractives content was determined following NREL/TP-510-42619 and TAPPI T- 204. Using a Soxhlet Apparatus, 150 ml of industrial grade acetone was used to extract 5 g of test sample for 6 hours. An additional 2 g of the sample was taken at this time for moisture content (MC) determination. Acetone was evaporated from the extract using a rotary evaporator. Extract was then dried at 40 ^{o}C for 24 hours in a vacuum oven and the final mass measured for extractives content determination.

The amount of total lignin and carbohydrates in biomass samples were determined as described in NREL/TP-510-42618. Air dried extractive-free test sample (0.5 g) was weighed into a dry beaker and 5 ml of 72% sulfuric acid was added. Employing a two-step procedure to ensure complete hydrolysis of sugars, test sample was first placed in a water bath set at 30 ^{o}C +/- 3 ^{o}C and incubated for one hour. Afterwards, the mixture was diluted to a concentration of 4% with deionised water and placed in an autoclave set at 121 ^{o}C for another hour.

The total lignin was computed as the sum of acid soluble lignin (ASL) and acid insoluble lignin (AIL). ASL was determined with a UV – Visible spectrophotometer immediately after hydrolysis. Absorbance of a test sample was measured at the recommended wavelength of 240 nm, ensuring that the absorbance ranged between 0.7 – 1.0.

Monomeric sugars (i.e. glucose, xylose, galactose, arabinose and mannose) in biomass samples were determined via High Performance Liquid Chromatography (HPLC) using a Biorad Aminex HPX-87P column equipped with the appropriate guard column at a column temperature of 85^{o}C and run time of 35 minutes. Holocellulose was computed as the sum of all monomeric sugars; cellulose was computed as $\mathrm{glucose} - (\frac{1}{3} * \mathrm{mannose})$; and hemicelluloses computed as the difference between holocellulose and cellulose (Jiang et al. 2014).

The HHV of biomass samples was determined according to ASTM D5865 using an IKA C-200 bomb calorimeter.

Proximate analysis was conducted to determine the thermal reactivity of biomass samples. Ash content was determined following NREL/TP-510-42622. Test sample weighing 1.0 g was put into a dry crucible and heated at 105 ^{o}C for 12 minutes. The temperature was increased to 250 ^{o}C and held constant for 30 minutes (in order to avoid flaming). Afterwards, temperature was increased to 575 ^{o}C and held isothermal for an additional 180 minutes. The volatile matter of biomass was determined as specified in CEN 15148/ ISO 562 using a furnace. The fixed carbon of biomass is a calculated value. It is the summation of the percentage of moisture, volatile matter and ash deducted from 100. The R software was used to perform Tukey pairwise comparison tests between the four biomass types ($\alpha = 0.05$).

5.3.2.2 NIR Spectra Acquisition

Spectra were collected over a wavenumber range of 10000 to 4000 cm^{-1} with a Perkin Elmer Spectrum model 400. Sample preparation for NIR spectra acquisition included grinding to pass an 80-mesh screen (Jiang et al. 2014) and oven drying for 4 hours. Each sample was scanned 32 times at a resolution of 4 cm^{-1} and averaged into one spectrum for analysis (Via et al. 2014). A background

spectrum (Spectralon reference sample) was taken every 10 minutes to correct for any potential drifts with time.

5.3.2.3 Multivariate Data Analysis

Partial least squares (PLS) regression models were developed with raw or pre-treated NIR spectra as regressors and conventional lab results as responses using Perkin Elmer Spectrum Quant+ software. The emphasis of models was on predicting one response at a time (i.e. PLS1) and not necessarily on trying to understand the underlying relationships between the variables since the regressors are many and highly collinear. Both regressors and responses were centered so as to enable equal comparison of loadings across different wavenumbers and also ensure that the criterion for choosing successive factors is based on how much variation they explain, in either the regressors or the responses or both. In addition, centering reduces the effect of multicollinearity in the highly dependent spectra. PLS, using the iterative NIPALS algorithm was used to extract successive linear combinations of the regressors (called factors, components, latent vectors or latent variables) such that variations in both response and regressors were optimally explained. In extracting the first factor, If $X = X_0$, and $Y = Y_0$ are the centered matrix of the regressors and responses respectively, NIPALS starts with a linear combination $t = X_0w$ of the regressors where t is called a score vector and w is its associated weight vector. NIPALS predicts both X_0 and Y_0 by regressing them on t:

$X_0 = tp'$, where $p' = (t't)^{-1}t' X_0$, and $Y_0 = tc'$, where $c' = (t't)^{-1}t' Y_0$. The vectors p and c are known as the X- and Y-loadings respectively. The specific linear combination $t = X_0w$ is the one with maximum covariance $t'u$ with some response linear combination $u = t = Y_0q$. Also, the X- and Y-weights w and q are proportional to the first left and right singular vectors of the covariance matrix $X'_0 Y_0$. The second factor is extracted in a similar way but X_0 and Y_0 are replaced with the X- and Y-residuals

(called deflated X and Y blocks) from the first factor: $X_1 = X_0 - X_0$; $Y_1 = Y_0 - Y_0$. The process of extracting a score vector and deflating the data matrices is iterated for as many factors as are desired.

A leave-one-out cross validation was used to validate the models (Mevik and Cederkvist 2004). In this technique, all samples except the first sample were used in building a model to predict the first, then the second is held out and the process repeated until all samples have been used as single-element test sets.

The performances of models developed to predict the properties of biomass were assessed with the standard error of calibration (SEC), standard error of cross validation (SECV), bias, coefficient of cross validation (R^2) and and residual predictive deviation / ratio of performance to deviation (RPD). SEC evaluates how precisely the regression line fits the data, bias detects any systematic difference between calibration set and the prediction set, SECV measures the precision of a model's predicting ability corrected for bias during validation, and R^2 measures the total variance between measured and predicted that can be modeled linearly (Kelley et al. 2004). SECV is synonymous to standard error of prediction (SEP) in this study because there was no validation with an independent test set. The RPD is a general indicator of the predictive adequacy of a model. It is used to evaluate SEP in terms of SD of the reference data.

5.4 Results and Discussion

5.4.1 NIR Spectra

Raw NIR spectra characteristic of the whole tree, wood and bark, slash and clean wood chips are shown in Figure 5.2. Even though a general pattern was observed in all spectra, there were variations in the amount of NIR light absorbed by the different biomass types. Slash generally absorbed the most, with wood and bark absorbing the least. Especially large baseline shifts were

noted in the 7100 to 10000 cm^{-1} region. Baseline shifts have primarily been attributed to density, particle size and ash content.

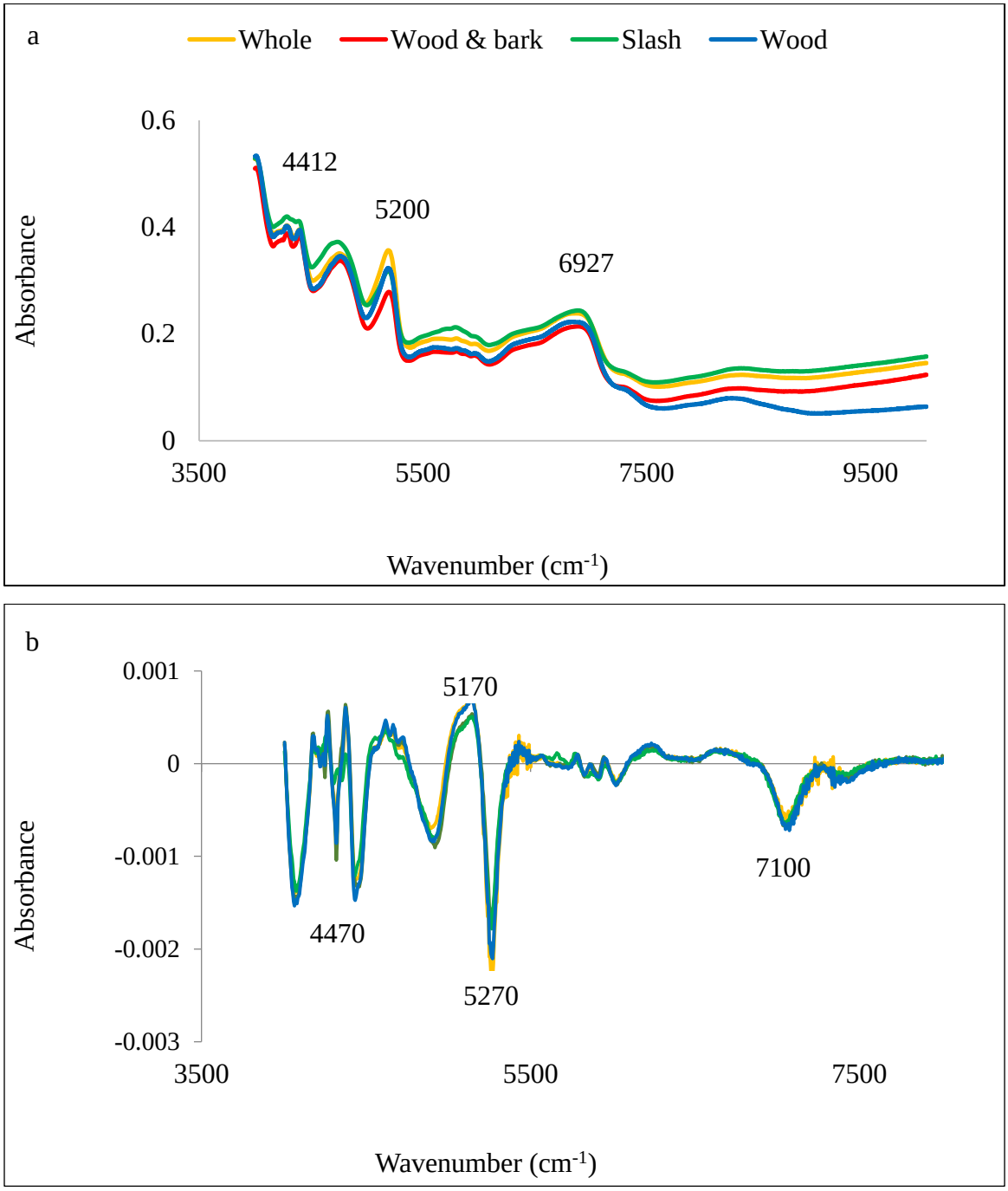

Figure 5.2: Characteristic NIR spectra of the different types of forest biomass for (a) raw and (b) 1st derivative spectra.

According to the Beer Lambert's Law, generally, a denser material or larger particle tends to absorb more infrared light than less dense or smaller particles (Via et al. 2013; Jones et al. 2006). Since there were no significant differences in the bulk densities of test samples ground to pass a 6-mesh screen -- whole tree {229 kgm^{-3}(11 kgm^{-3})}; wood and bark {241 kgm^{-3}(25 kgm^{-3})}; slash {234 kgm^{-3}(9 kgm^{-3})} and wood chips {227 kgm^{-3} (8 kgm^{-3})} -- baseline shifts may be attributed to the differences in ash contents of the biomass types (Sanderson et al. 1996). The absorbance peaks seen from 4000 cm^{-1} to 5000 cm^{-1} arise when C-H, N-H and O-H functional groups interact with one another (i.e. combination bands) (Derkyi et al. 2011; Jones et al. 2006; Schwanninger et al. 2011). Peaks have also been assigned to specific components of biomass: 4545 cm^{-1} is associated with the C-H and C=O stretching in lignin; 6250 to 7140 cm^{-1} has been assigned to the first overtone of O-H vibrations in cellulose and hemicellulose; and 4950 to 5291 cm^{-1} has been attributed to the interaction of O-H in carbohydrates and water.

To correct for baseline shifts, the raw spectra were pretreated with 1^{st}-derivatives (Via et al. 2013). Treating with 1^{st} derivatives is also known to reduce non-linearity and multicollinearity between factors. First derivative spectra for the different biomass types can be seen in Figure 5.2.

5.4.2 Chemical Properties

Descriptive statistics of the chemical components of forest biomass samples determined by conventional wet chemistry are presented in Table 5.1. The chemical composition of whole tree, wood and bark, slash and clean wood chips were different ($\alpha = 0.05$) except for galactose. As was expected, most of the significant differences were noted between the slash and clean wood. Lignin was highest in slash (44%) but lowest in clean wood chips (34%). On the other hand, wood contained the highest amount of cellulose at 43% whereas slash had only 25%. Slash also contained the most extractives at 10% with wood and bark having the least at 2%.

Table 5.1: Descriptive statistics of the chemical composition of whole tree, wood & bark, slash and wood. All values are expressed on percent oven-dry basis.

Property	Whole	Wood & Bark	Slash	Wood
Extractives	4.2 (0.7)	2.1 (1.2)	10.1 (1.9)	3.1 (0.3)
Lignin	37.3 (1.6)	35.9 (2.0)	43.7 (1.7)	33.5 (1.6)
Glucose	33.7 (2.4)	41.3 (3.6)	27.2 (2.4)	45.1 (2.5)
Mannose	8.2 (1.5)	7.2 (1.9)	6.1 (1.2)	7.3 (0.3)
Galactose	4.7 (2.1)	5.3 (1.7)	5.7 (2.1)	4.4 (0.5)
Arabinose	2.6 (0.4)	2.0 (0.4)	3.3 (0.5)	1.7 (0.1)
Xylose	5.7 (1.1)	6.0 (1.4)	4.9 (1.6)	4.5 (0.2)
Cellulose	31.0 (2.4)	38.9 (3.8)	25.2 (2.4)	42.7 (2.4)
Hemicelluloses	24.1 (2.2)	22.8 (2.8)	22.1 (5.2)	20.3 (0.9)
Holocellulose	55.0 (3.9)	61.8 (3.4)	47.2 (5.7)	63.0 (3.0)

Note: All values are expressed on percent oven-dry basis.

All 40 biomass samples were pooled together for NIR model calibration and validation using the leave-one-out cross validation technique. Models developed with the raw spectra performed worse than models built with 1^{st} derivative treated spectra, in that, they gave lower correlation coefficients using more latent variables (LVs). For raw spectra models, the model developed to predict extractives content performed best (SEC = 1.08%, SECV = 1.23%, R^2 = 0.91, LVs = 4), while the model for predicting the galactose content of forest biomass performed worst (SEC = 1.51%, SECV = 2.06%, R^2 = 0.40, LVs = 8), Table 5.2.

Table 5.2: PLS model statistics for the chemical properties of forest biomass using (i) raw spectra and (ii) first-derivative-treated spectra.

	Range	SEC		Bias		LVs		SECV		R^2		RPD	
		i	ii	i	ii	i	ii	i	ii	i	ii	i	ii
Ext.	1.1 - 13.1	1.08	0.99	-0.04	0.03	4	2	1.23	1.23	0.91	0.92	2.71	2.71
Lig.	30.8 - 45.9	1.55	1.77	0.07	0.01	4	2	1.75	2.09	0.87	0.83	2.38	2.0
Glu.	22.2 - 50.3	3.34	3.3	-0.1	-0.12	4	2	3.37	3.95	0.82	0.82	2.22	1.9
Man.	3.1 - 10.8	1	0.63	-0.02	-0.1	8	3	1.45	1.6	0.65	0.84	1.04	0.94
Gal.	1.6 - 9.2	1.52	0.97	-0.06	0	8	3	2.06	2.05	0.4	0.71	0.84	0.85
Ara.	1.5 - 4.1	0.37	0.36	-0.02	0	4	2	0.42	0.42	0.76	0.77	1.73	1.73
Xyl.	4.9 - 9.3	0.92	0.73	-0.01	0.05	5	3	1.19	1.23	0.57	0.71	1.1	1.06
Cel.	19.9 - 47.8	3.45	3.4	0.1	-0.14	4	2	3.87	4.09	0.81	0.8	1.92	1.81
Hem.	13.7 - 28.4	2.61	1.78	0.13	0	8	3	3.66	3.4	0.41	0.68	0.83	0.89
Hol.	40.0 - 68.1	4.04	3.73	-0.11	-0.15	4	2	4.53	4.44	0.73	0.75	1.62	1.65

Note: Ext. – Extractives; Lig. – Lignin; Glu. – Glucose; Man. – Mannose; Gal. – Galactose; Ara. – Arabinose, Xyl. – Xylose; Cel. – Cellulose; Hem. – Hemicelluloses. Hol. – Holocellulose.

SEC - Standard error of calibration; LVs - Latent variables (Factors); SECV - Standard error of cross validation; R^2 - Coefficient of cross validation.

Calibration models developed using 1st derivative spectra to predict extractives, lignin, cellulose, glucose and mannose had strong calibration coefficients Table 5.2. The highest R^2 value of 0.92 (SEC = 0.99, SECV = 1.23%) was obtained for extractives. Models for the other four components gave R^2 greater than 0.8. Model statistics for the remaining chemical constituents were also quite good, with the worst performance by hemicelluloses (R^2 = 0.68, SEC = 1.78%, SECV = 3.40). According to Williams (2004), models with R^2 from 0.92 to 0.96 can be used for most

application including quality assurance, while R^2 from 0.83 to 0.9 can be used for most applications but with some level of caution. He also stated that approximate calibration can be done with models with R^2 range of 0.66 to 0.81, rough screening done with 0.5 to 0.64, but models with R^2 less than 0.5 are non-usable in any NIR calibration.

Another parameter used to evaluate developed models was the RPD. Except for glucose, the RPDs of all the other carbohydrates were less than 2, with hemicelluloses having the lowest of 0.83, Table 5.2. A higher RPD is an indication of a robust model; a model with an RPD between 1.5 and 2.5 is expected to be successful in preliminary screening (Via 2013).

Error values noted in this study are comparable to the errors of models that were developed by So et al. (2012) to predict the extractives content of *Pinus palustris* (SEC = 1.78%, SECV = 3.33%). Sykes et al. (2005) also obtained an SEC of 2.42% and SECV of 2.13% for models they developed to predict the α-cellulose content of loblolly pine clean wood. Work done by Jones et al. (2006) to predict the monomeric sugars of loblolly pine wood gave similar SECs and SECVs as obtained in this study. For example, glucose: SEC = 1.12%, SECV = 1.78%, xylose: SEC = 0.16, SECV = 0.41%, and mannose: SEC = 0.30%, SECV = 0.64%. It should however be noted that, their models were developed with more factors (i.e. 4, 6 and 5 as against 2, 3, 3 respectively), and as observed by Kelley et al. (2004), this usually reduce the SEC, SECV and SEP. It is also worth mentioning that they used a more homogenous material comprising of only clean wood, as opposed to the relatively heterogeneous mixture of biomass types used in this study. A more comparable study in terms of type of biomass might be that by Ono et al. (2003). When they used select NIR wavelengths and multiple linear regression to predict the extractives, lignin and holocellulose of fresh leaves, litterfall and organic matter of softwoods and hardwoods from subtropical and boreal forests in Japan, the SEC and SECV of their models were generally higher than those for models in this study. Their

model for extractives had an SEC of 2.9%, SEP of 3.4% and an R^2 of 0.82. For lignin and holocellulose, the respective SECs and SEPs were 5.0% & 3.5%, and 5.0% & 3.5%. They reported 0.78 and 0.83 as the respective correlation coefficients between measured and NIR-predicted chemistry. Furthermore, the biomass they used had very wide chemical ranges. For instance, lignin in their samples ranged from 5.6 to 53.7% and holocellulose was from 20.4 to 68.4%.

Another indicator for good models is when SEC and SECV are comparable (Derkyi et al. 2011; Mevik and Cederkvist 2004). A small difference between calibration and validation errors is an indication that the test sets were well predicted by calibration models. According to Sanderson et al. (1996), a model having its SECV lesser or equal to 1.3 times its SEC has a good performance. Even though differences in SEC and SECV seemed large for models developed for chemical property predictions, only those for mannose, galactose, xylose and hemicelluloses exceeded this 1.3 limit. The performance of prediction models for these monomeric sugars and consequently the hemicelluloses may be due to overlapping signals caused by high concentrations of relatively similar sugars that vary only with respect to the location of H and OH groups on the cyclical 6 carbon structure. Other researchers have however hypothesized that, since the different sugars have specific ratios to one another after polymerization and result in different functional groups in the different types of hemicelluloses, this should not cause a problem but instead improve the diagnostics of predictive models (Ono et al. 2003). Plots relating laboratory determined properties to NIR-predicted properties are presented in Figure 5.3 and Figure 5.4.

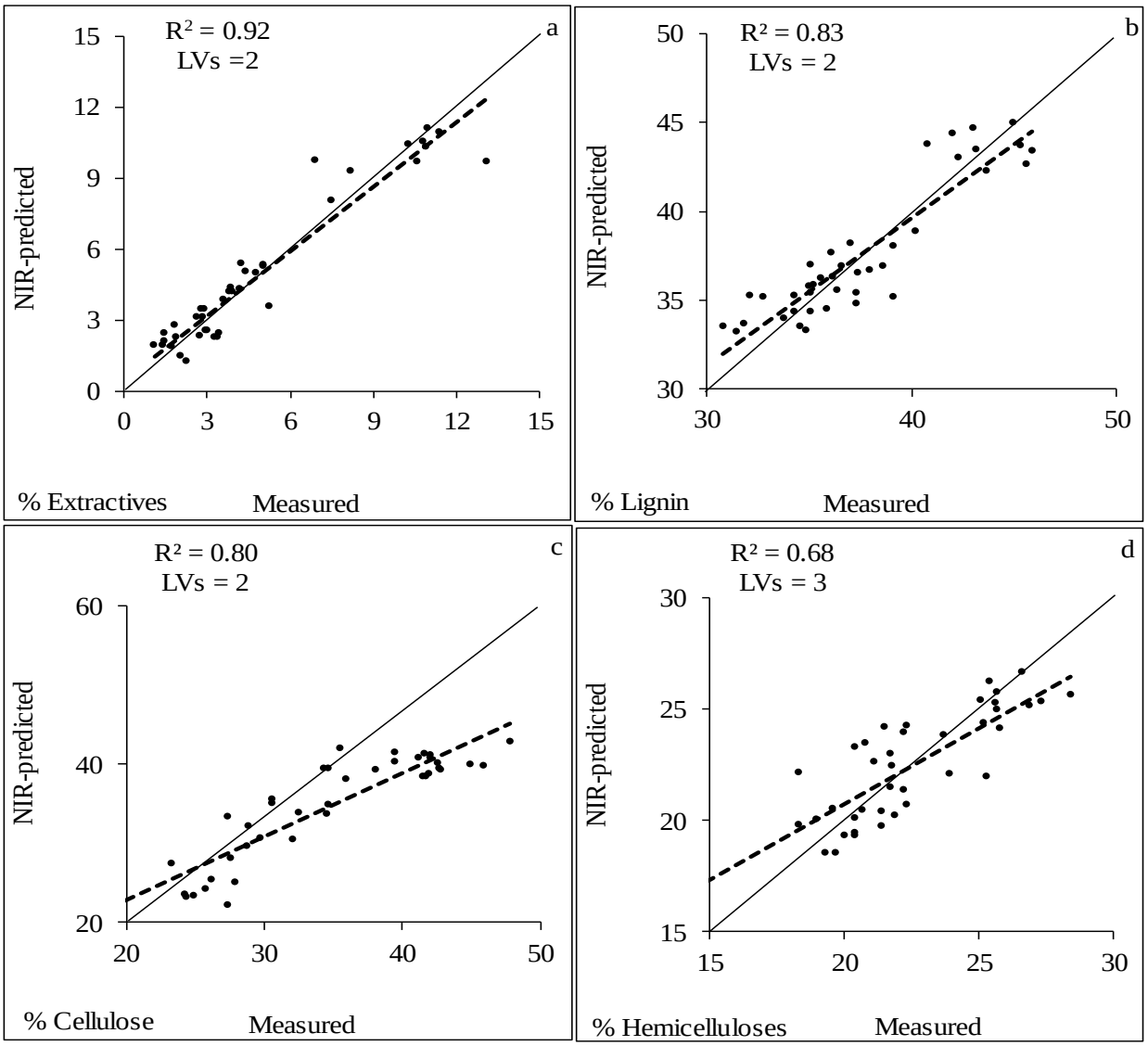

Figure 5.3: Regression plots of measured versus NIR-predicted values for (a) extractives, (b) lignin, (c) cellulose and (d) hemicelluloses of forest biomass.

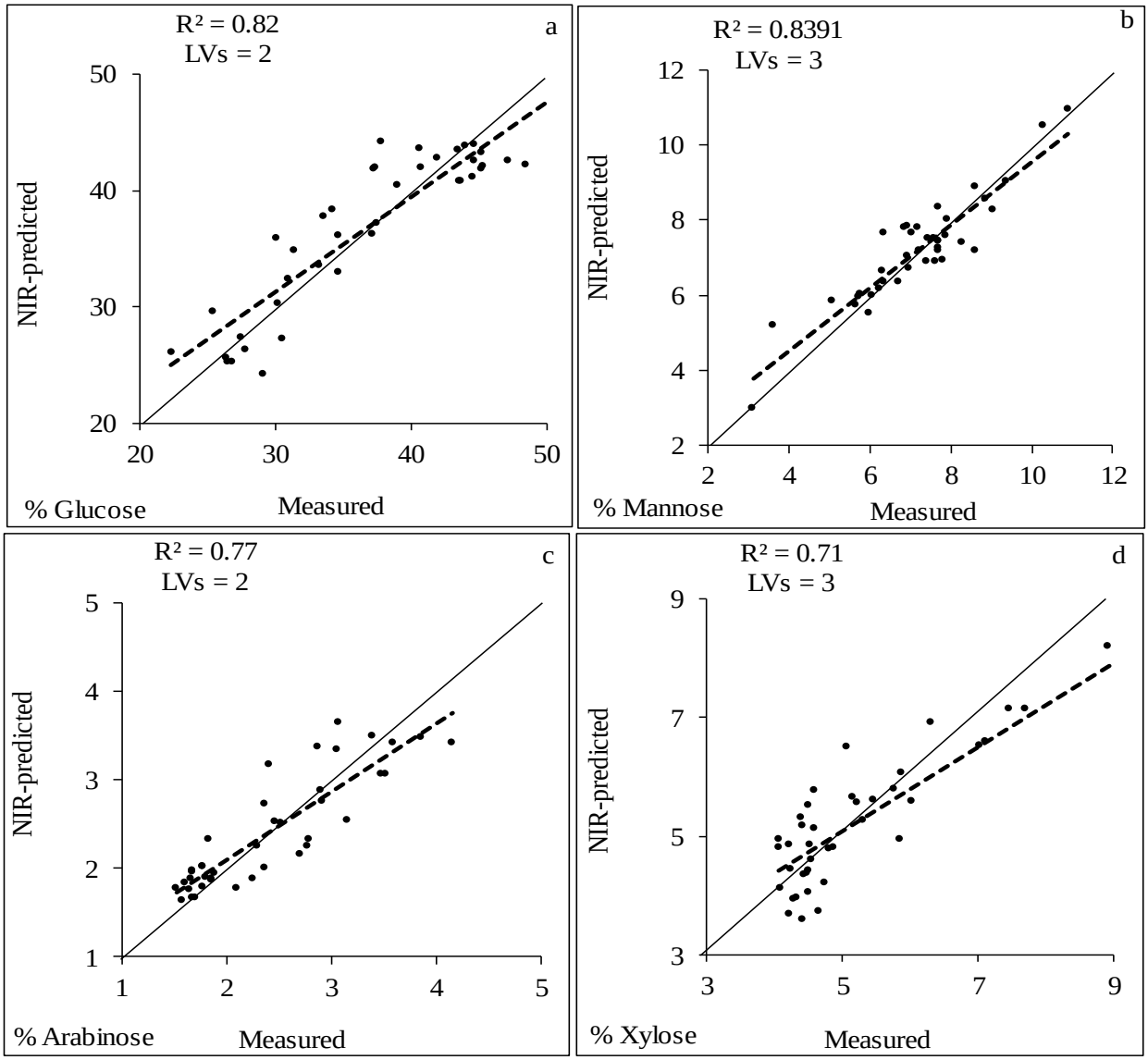

Figure 5.4: Regression plots of measured versus NIR-predicted values for (a) glucose, (b) mannose, (c) arabinose and (d) xylose of forest biomass.

5.4.3. Thermal Reactivity Properties

The thermal reactivity properties of volatile matter, fixed carbon and ash, as well as HHV were determined by proximate analysis and bomb calorimetry are shown in Table 5.3. As in the case of the chemical properties, some significant differences were noted in the thermal properties of the different types of forest biomass. For instance, the 1.9% of ash measured in slash was significantly higher than the 0.4% determined for wood. In reverse, the volatile matter content of wood (85.7%) was significantly higher than that of slash (77.3%). Also, HHV of wood and bark was significantly lower than that of slash; whereas the highest fixed carbon determined in slash (16.2%) was significantly different from that in the other biomass types.

Table 5.3: Regression plots of measured versus NIR-predicted values for (a) glucose, (b) mannose, (c) arabinose and (d) xylose of forest biomass.

	Whole	Wood & Bark	Slash	Wood
Volatile matter (%)	81.1 (1.4)	82.3 (0.8)	77.3 (0.6)	85.7 (0.5)
Fixed carbon (%)	9.8 (1.2)	9.7 (1.4)	16.2 (0.8)	8.9 (0.7)
Ash (%)	1.8 (0.7)	1.5 (1.6)	1.9 (0.2)	0.4 (0.1)
HHV (MJ/kg)	20.2 (0.4)	19.8 (0.5)	20.6 (0.7)	20.4 (0.3)

As in the case of chemical properties, thermal reactivity models developed using the 1st derivative spectra performed better than models developed with the raw spectra (Table 5.4). Calibration models developed using the PLS1 algorithm on 1st derivative treated spectra to predict volatile matter and fixed carbon used two latent variables in both models to give SECs of 1.0 and 1.27 respectively. When models were used to predict the single-element test sets, the SECVs were

comparable to the SECs (i.e. 1.15 and 1.51); the R^2 values were high, and their respective RPDs were 2.46 and 2.09.

Table 5.4: PLS model statistics for the thermal reactivity properties of forest biomass using (i) raw spectra and (ii) first-derivative-treated spectra.

		% Volatile matter	% Fixed carbon	% Ash	HHV (MJ/kg)
Range		76.3 - 85.5	6.6 - 17.2	0.2 - 5.8	18.9 - 21.5
SEC	i	0.82	1.45	0.87	0.48
	ii	1	1.27	0.61	0.37
Bias	i	0.06	0.05	-0.01	0
	ii	-0.02	0.01	0	0.01
LVs	i	4	4	3	5
	ii	2	2	3	2
SECV	i	0.93	1.65	0.92	0.54
	ii	1.15	1.51	0.97	0.48
R^2	i	0.92	0.81	0.36	0.34
	ii	0.88	0.85	0.68	0.57
RPD	i	3.04	1.91	1.14	1.02
	ii	2.46	2.09	1.08	1.14

Note: SEC - Standard error of calibration; LVs - Latent variables (Factors); SECV - Standard error of cross validation; R^2 - Coefficient of cross validation.

The HHV of lignocellulosic biomass has been related to its chemical composition, ultimate and proximate data. Thus, even though HHV is not a direct chemical property, it can be estimated and predicted with NIR y which is known to interact with overtones and combinations of fundamental vibrations of C-H, O-H, C=O, N-H, aromatic C-H groups and S-H functional groups (Schwanninger et al. 2011; Gillon et al. 1997). Similarly, NIR can be used to indirectly predict the

inorganic ash based on how they bound to organic moieties or how they affect some absorption bands (Lestander and Rhen 2005; Windham et al. 1991).

The HHV and ash content were however not predicted as well in this study as can be seen in Table 5.4. The range of HHV measured for test samples used in this study was relatively small (18.9 to 21.5 MJ/kg) compared to for instance, those measured by Gillon et al. (1997) (i.e. 17.1 to 25.6 MJ/kg) for biomass made up of leaves, needles, twigs and bark of conifers, broad leaved trees, shrubs and grasses. They had very good diagnostic statistics for their models; SEC = 0.42 MJ/kg, SECV = 0.50 MJ/kg, R^2 = 0.92. Even though prediction was not as strong in this study, the SEC and SECV were comparable to theirs. The models by So and Eberhardt (2010) to predict the HHV of unextracted and acetone-extracted *Pinus palustris* wood did not do as well either. They reported an SEC of 0.32 MJ/kg, SECV of 0.41 MJ/kg and R^2 of 0.72 for the unextracted biomass, and an SEC of 0.21 MJ/kg, SECV of 0.24 MJ/kg and a very poor R^2 of 0.05 for the extracted biomass.

Three prinicipal components were used in building the model that predicted ash with a R^2 of 0.68. Plots relating laboratory determined properties to NIR-predicted properties are presented in Figure 5.5. Models with better predictive capabilities were developed for stem and branch wood of Norway spruce by Lestander and Rhen (2005) using raw mean centered spectra or spectra pretreated with orthogonal signal correction (OSC) from select wavelengths. In all cases, the R^2 values were 0.98.

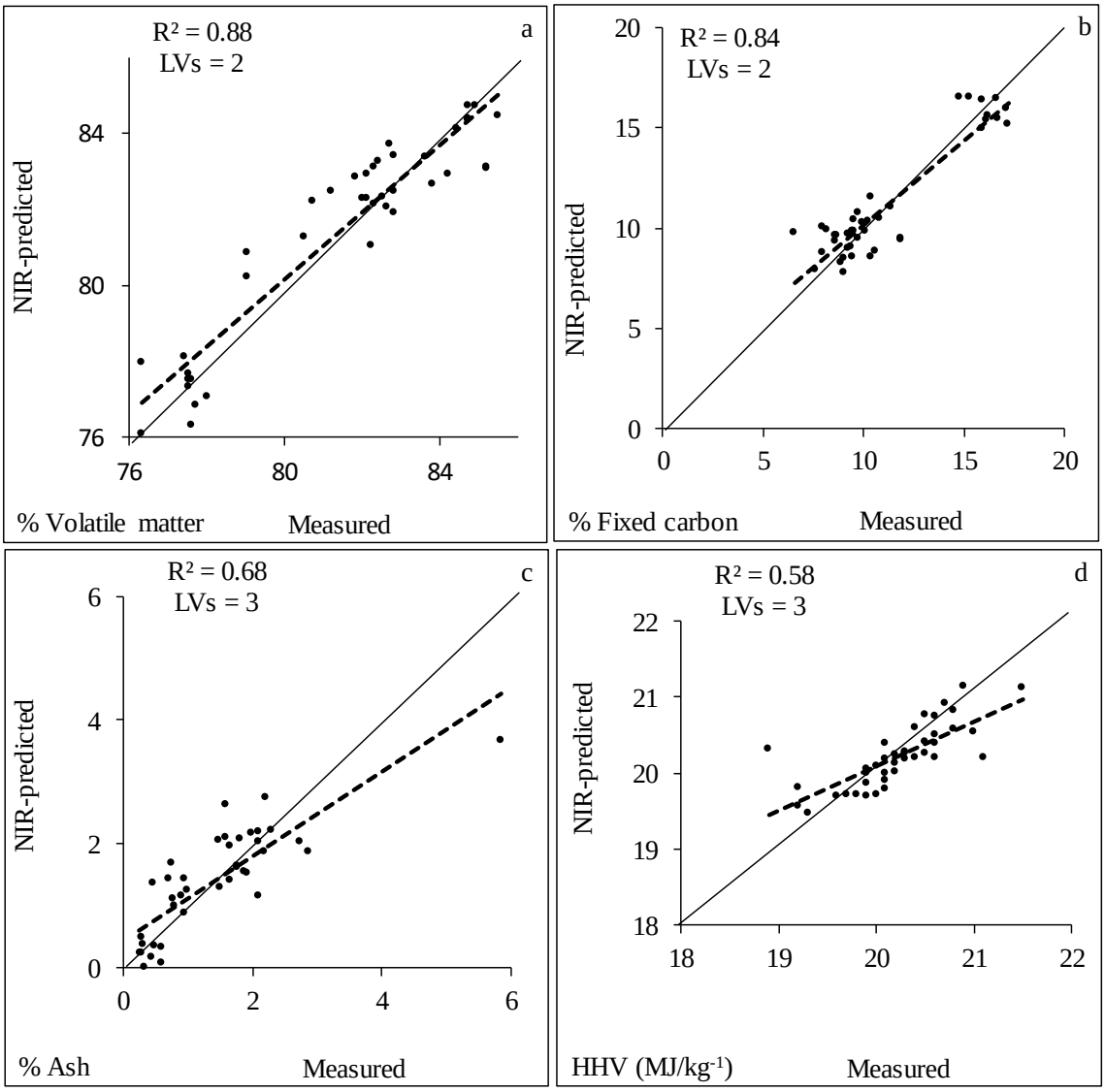

Figure 5.5: Regression plots of measured versus NIR-predicted values for (a) volatile matter, (b) fixed carbon, (c) ash and (d) HHV of forest biomass.

The performance of models developed in this study are promising considering the fact that, very different forest biomass types made up of needles, bark, twigs, stemwood and branchwood were used. Furthermore, errors associated with using models to predict the properties under study were mostly lesser than errors encountered when the various properties were measured in the laboratory using conventional methods.

5.5 Conclusions

NIR coupled with PLS was successfully used to predict important chemical and thermal reactivity properties of heterogeneous forest biomass comprising of whole tree, wood & bark, slash, and clean wood. Models developed to predict extractives, lignin, glucose, arabinose, cellulose, holocellulose, volatile matter and fixed carbon contents of forest biomass were strong with a minimum R^2 of 0.75. Their good performances were further supported by their RPD values which passed the 1.5 threshold for preliminary screening. Even though the coefficient of cross validation between lab measured and NIR-predicted mannose ($R^2 > 0.80$) was also high, its RPD made the adequacy of this model questionable. Further studies will be needed to improve the robustness and accuracy of this model, in addition to the models used in predicting galactose, xylose, hemicelluloses, HHV and ash.

.

Chapter 6 : Rapid Quantitative Analysis of Forest Biomass using Fourier Transform Infrared Spectroscopy (FTIR) and Partial Least Squares (PLS) Regression[3]

6.1 Abstract

Fourier transform infrared reflectance spectroscopy (FTIR) has been used to predict properties of forest logging residue, a very heterogeneous feedstock material. Properties studied included the chemical composition, thermal reactivity and energy content. The ability to rapidly determine these properties is vital in the optimization of conversion technologies for the successful commercialization of bio-based products.

Partial least squares regression of first derivative treated FTIR spectra had good correlations with the conventionally measured properties. For the chemical composition, constructed models generally did a better job of predicting the extractives and lignin content than the carbohydrates. In predicting the thermochemical properties, models for volatile matter and fixed carbon performed very well (i.e. $R^2 > 0.80$, RPD > 2.0). The effect of reducing the wavenumber range to the fingerprint region for PLS modeling and the relationship between the chemical composition and higher heating value of logging residue were also explored.

This study is new and different in that, it is the first to use FTIR to quantitatively analyze forest logging residue, an abundant resource that can be used as a feedstock in the emerging low carbon economy. Furthermore, it provides a complete and systematic characterization of this heterogeneous raw material.

6.2 Introduction

Lignocellulosic biomass is the only renewable resource that can be used in the production of biofuels and platform chemicals in addition to bioenergy. As the most abundant carbon neutral resource, using biomass instead of fossil fuels can help mitigate environmental pollution. However, many physical, chemical and structural factors can hinder the conversion of biomass into fuels and chemicals. A better understanding of the properties of biomass will be important in the allocation of feedstock to the appropriate end use. An ability to determine these properties in a timely manner is also necessary in the optimization of conversion technologies for the successful commercialization of biomass-based products. There is thus a need for high throughput methods and equipment in the monitoring and characterization of the raw feedstock as conventional methods have been laborious and destructive.

Fourier transform infrared reflectance spectroscopy (FTIR) has been used as a powerful analytical tool for the rapid characterization of lignocellulosic biomass. Since FTIR determines the presence of fundamental molecular vibrations that are characteristic of a chemical compound or class of compounds, it has widely been used in the qualitative elucidation of changes in biomass structure during and/or after treatment with processes. For instance, FTIR was used to study trembling aspen extracted with supercritical methanol (Grandmaison et al. 1987) and also to monitor the physical and chemical changes that result as corn stover undergoes ammonia fiber expansion (AFEX) and iconic liquid (IL) pretreatments (Li et al. 2011). With the advancement of multivariate data analysis, researchers are now coupling FTIR with chemometric methods for rapid quantitative of biomass feedstock.

---

[3] **Reprinted from the Journal of Analytical Methods in Chemistry. doi: dx.doi.org/10.1155/2016/1839598**

FTIR-based partial least squares (PLS) models were constructed to quantify the monomeric sugars, acetic acid and 5-hydroxymethyl-2-furfural (HMF) of dilute acid pretreated biomass (Tucker et al. 2013). The tool has also been used to model the ash, volatile matter, fixed carbon and higher heating value (HHV) of sweetgum, loblolly pine and switchgrass torrefied at different temperatures (Via et al. 2013). For some studies on the raw biomass, FTIR was employed in characterizing several agricultural residues and their extractives content (Naik et al. 2010), in the qualitative analysis of lignin from five timber species (Rana et al. 2010) and in predicting the chemical composition of hardwoods (Zhou et al. 2015). FTIR based models have also been employed for discriminant and classification analysis of biomass feedstocks and plant materials (Acquah et al. 2016; Carballo-Meilan et al. 2014; Chen et al. 2010; Hobro et al. 2010; Meng et al. 2015).

Most of the studies using FTIR were conducted on biomass that has been subjected to some kind of pretreatment; surprisingly, not very much was found on the raw resource.

In the USA, about 1.3 million dry tons of biomass can be sourced annually from forestry (27%) and agricultural (77%) operations, capable of replacing a third of the country's current fuel consumption (Perlack et al. 2005). Forest biomass includes logging residues, pre-commercial thinnings, fuel treatments, residues from primary and secondary mill processing and urban wood wastes. According to Smith et al. (2009), some 68 million dry tons of logging residues are currently produced in the USA, most of which is left onsite. Using logging residue as a raw material will ensure a more complete and sustainable utilization of trees. In addition, several studies have shown that the sustainable removal of logging residues can improve forest health, enhance replanting efforts and regeneration, and control forest fires (Spellman and Beiber 2011; Hacker 2005; Leinonen 2004).

Qualitative and quantitative analysis of biomass with FTIR can be quite precise when material vary considerably in chemical structure. For example, woody tissue could be easily differentiated from bark resulting in an easy calculation of bark content in aspen and birch (Brink et al. 2010) and beech could be differentiated from pine due to considerable differences in syringyl and guiaicyl moieties (Pandey and Pitman 2003). Such differentiation was less discriminative for the same tree species with tissue obtained from the juvenile and mature wood (Via et al. 2003). However when perturbations such as temperature (Via et al. 2013) and chemical treatments (Jiang et al. 2014) are introduced then model robustness for identification and/or concentration determination becomes more superior. The objective of this study was to employ FTIR coupled with partial least squares (PLS) regression to rapidly predict the chemical composition, thermal reactivity and energy content of logging residue of loblolly pine, the most economically important tree species in the USA. This study attempts to take advantage of the wide differences in bark, needle, and woody tissue chemistry that should allow for an easy discrimination and quantification. The accurate estimation of for instance the concentration of carbohydrates is important since it is directly proportional to the yield of biofuels; and having prior knowledge of the inorganic fraction will enable the anticipation of slagging or the extent to which the calorific value may be impacted.

6.3 Materials and Methods

6.3.1 Materials

Lignocellulosic biomass acquired during harvesting operations on several loblolly pine plantations in southern Alabama, USA were used for this study. Material comprised of Whole tree, Wood and bark, Slash (i.e. limbs and foliage) and clean Wood chips. Ten biomass sets were sampled for each of the four biomass types.

6.3.2 Methods

6.3.2.1 Conventional Laboratory Methods

Conventional standard methods were used to determine the chemical makeup, thermal reactivity and energy content of biomass.

The chemical composition of biomass samples was determined via wet chemistry. Extractive content was determined as specified in NREL/TP-510-42619 and TAPPI T- 204. 5 g of a test sample ground to pass a 40-mesh screen was extracted in 150 ml of acetone for 6 hours in a Soxhlet Apparatus. Afterwards, the acetone was allowed to evaporate before drying the extract in a vacuum oven at 40 ^{o}C for 24 hours. Air-dried extractive-free samples were used to determine lignin and carbohydrates following NREL/TP-510-42618. Test samples were first hydrolyzed with 72% sulfuric acid. This primary hydrolysis was carried out at 30 ^{o}C ± 3 °C for an hour. Then, the concentration of acid was diluted to 4% with deionized water and a secondary hydrolysis carried out in an autoclave at 121 °C for another hour. Hydrolyzed samples were allowed to cool before filtering through tared glass crucibles. An aliquot of this filtrate was collected to be used for determining the acid-soluble lignin (ASL) and monomeric sugars. The solid residue was thoroughly washed with distilled water, oven dried at 105 °C overnight, and the final weight used for calculating the acid-insoluble lignin (AIL) content of biomass. The ASL was determined with a UV/vis spectrophotometer immediately after hydrolysis. Total lignin was calculated as the sum of ASL and AIL. The monomeric sugars (i.e. arabinose, glucose, galactose, mannose and xylose) in test samples were measured using a Biorad Aminex HPX-87P column equipped HPLC. Holocellulose content was calculated as the sum of monomeric sugars; the cellulose content was computed as $\mathrm{glucose} - (\frac{1}{3} * \mathrm{mannose})$, and the difference between holocellulose and cellulose designated as hemicelluloses.

Bomb calorimetry, as specified in ASTM D5865 was used in calorific value determination; whereas proximate analysis was conducted following NREL/TP-510-42622 (for ash content) and CEN/TS 15148 (for volatile matter content).

Analysis of Variance (ANOVA) followed by Tukey pairwise comparison tests between the four biomass types ($\alpha = 0.05$) was performed using the R Stats Package. Duplicates experiments were ran for each test sample.

6.3.2.2 Collection of Spectra

Mid infrared spectra were collected over a wavenumber range of 4000 cm^{-1} to 650 cm^{-1} using a PerkinElmer Spectrum 400 FT-IR/FT-NIR spectrometer equipped with a diamond crystal attenuated total reflectance device (i.e. ATR-FTIR) and a torque knob to ensure that consistent pressure is applied to samples during spectra collection. Prior to spectra acquisition, samples were ground to pass an 80-mesh screen to improve model properties through light scatter reduction and oven dried at 40 ^{o}C for 4 hours. Each test sample was placed on the diamond plate, a pressure of 70 ± 2 psi applied using the torque knob, then scanned thirty-two times at a resolution of 4 cm^{-1}. The average of the thirty-two spectra is corrected for background absorbance by subtracting the spectrum of the empty ATR crystal and used for analysis.

6.3.3.3 Partial Least Squares (PLS) Regression

PLS regression is a statistical technique for developing predictive models of multivariate data that otherwise have high collinearity. The iterative NIPALS algorithm used extracted successive linear combinations of the predictors (called factors or latent vectors) such that variations in both predictors (i.e. MIR spectra) and responses (i.e. property under study) were optimally accounted for. A more detailed description of the procedure can be found elsewhere (Acquah et al. 2015).

PLS models were developed with PerkinElmer's Spectrum Quant^{+} software using spectra of the full mid infrared region (i.e. 4000 – 650 cm^{-1}) as well as the fingerprint region (1800 – 650 cm^{-1}). Spectra were pretreated with derivatives (1st order, 5-point) for baseline correction and also to help reduce non-linearity and multicollinearity among variables. 2nd derivatives were not used due to the generally lower signal to noise ratio. Both predictors and responses were mean centered prior to modeling. Due to the relatively small sample size (n = 40), the leave-one-out cross validation modeling method was employed. In this technique, for each run, 39 out of the 40 samples are used as training dataset for calibrating a model that is used to predict the 1-sample test dataset. This is iterated forty times until all samples are used as independent single-element test datasets.

Developed models were evaluated using such statistics as the standard error of calibration (SEC), standard error of cross-validation (SECV), coefficient of determination (R^2) and ratio of preformance to deviation (RPD).

6.4 Results and Discussion

6.4.1. FTIR Spectra

MIR spectra characteristic of the four biomass sets understudied are presented in Figure 1. Even though this region encompasses the 4000 to 650 cm^{-1} wavenumber range, the fingerprint region (1800 to 650 cm^{-1}) is usually of particular interest because it contains the most spectral information pertaining to the molecular/chemical composition of a material (Fig 6.1-I). In the literature, several bands have been linked to carbohydrates due to their associated functional groups. Within the fingerprint region, peaks that result due to the polysaccharides include: (P1) 897 cm^{-1} and (P2) 1030 cm^{-1} from the C-H deformation in cellulose and C-O stretch in polysaccharides respectively, (P3) 1157 cm^{-1} from C-O-C vibration, (P4) 1239 cm^{-1} from C-O stretch and O-H in plane in polysaccharides, (P5) 1465 cm^{-1} from C-H deformation and (P6) 1740 cm^{-1} from the C=O

stretching of unconjugated ketones mostly in the hemicelluloses. In the case of lignin, the peak at (L1) 1122 cm-1 occurs due to aromatic skeletal and C-O stretch. A guaiacyl ring breathing with C-O stretching causes a peak to arise at (L2) 1270 cm^{-1} and syringyl ring breathing creates the peak at (L3) 1365 cm^{-1}. The strong peak at (L4) 1505 cm^{-1} is attributed to the C=C stretch characteristic of aromatic skeletal compounds in lignin and extractives. Outside the fingerprint range (Figure 6.1-II), the peak occurring at (T1) 2935 cm^{-1} have been associated with the bending and stretching of C-H, as well as its aromatic ring vibration in lignin; whereas that at (T2) 3345 cm^{-1} has been assigned to bonded O-H (Grandmaison et al. 1987; Rodrigues et al. 1998; Faix 1992; Hergert 1971; Harrington et al. 1964).

The four biomass types followed a similar absorbance pattern in the mid infrared region. Slash generally had the highest absorbance values, followed by Wood, Whole then Wood & bark. The spectra of Slash had prominent peaks at (L5) 1635 cm^{-1} and (T1) 2935 cm^{-1} compared to the other biomass types. The former has been attributed to C-O stretching of conjugated or aromatic ketones and/or C=O stretching vibration in flavones, and the latter results from the aromatic ring vibration in lignin (Allison et al. 2009; Nuopponen et al. 2006; Padney 1998). These high peaks could thus be explained by the significantly high contents of extractives and lignin in Slash, Figure 6.1.

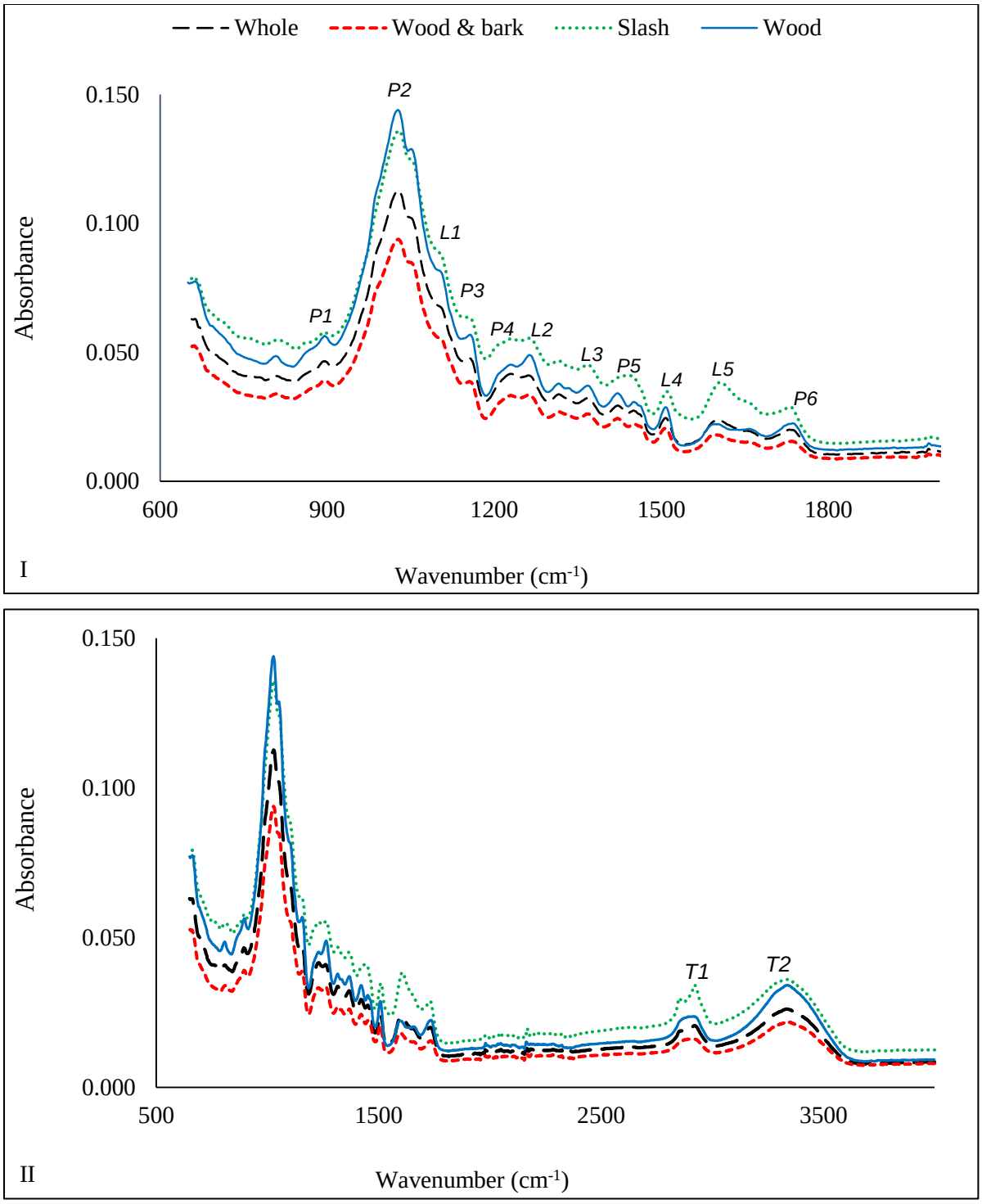

Figure 6.1: FTIR spectra of the different types of forest biomass. I – Fingerprint region; II – Full MIR range.

These assignments provide some insight into the chemical moieties present in the different biomass types. However, the overlapping peaks makes it challenging to tease out subtle difference by simple visual inspection sometimes (Stuart 2004). As such, the application of multivariate data analytical techniques to the spectra of lignocellulosic biomass help to extract relevant information and structure spectra and conventionally acquired chemical data into empirical mathematical models that are capable of predicting properties of future measurements and even other properties that are not directly measurable (Varmuza and Filmoser 2009). Scatter, stray light, path length variation, inconsistency in instrument response and random noise can cause interferences such as baseline shifts, vertical displacements, and non-uniform slope in infrared spectra. Pretreatment methods including standard normal variate (SNV) transformation, multiplicative scatter correction (MSC), derivatives and orthogonal signal correction (OSC) are therefore usually used to minimize, standardize or even eliminate the impacts of these interferences on IR spectra before multivariate data analysis to improve the robustness of calibration models. In this study, the first derivatives of spectra were used to reduce baseline offsets and improve the resolution of overlapping peaks (Heise and Winzen 2002).

6.4.2 PLS Modeling of the Chemical Composition of Forest Logging Residue

The chemical composition of forest logging residue determined via conventional methods are summarized in Figure 6.2. Some significant differences ($\alpha = 0.05$) were noted among the four biomass types. The mean concentration of extractives ranged from a low of 2% for Wood & bark to a high of 10% for Slash. The percentage of glucose was significantly lower in Slash (27%) and Whole (34%) compared to Wood & bark (41%) and Wood (45%). This pattern was unsurprisingly followed by the amount of cellulose in the four types of forest logging residue. Whole had the highest

amount of hemicelluloses, and this was statistically similar to the concentrations found in Slash and Wood & bark.

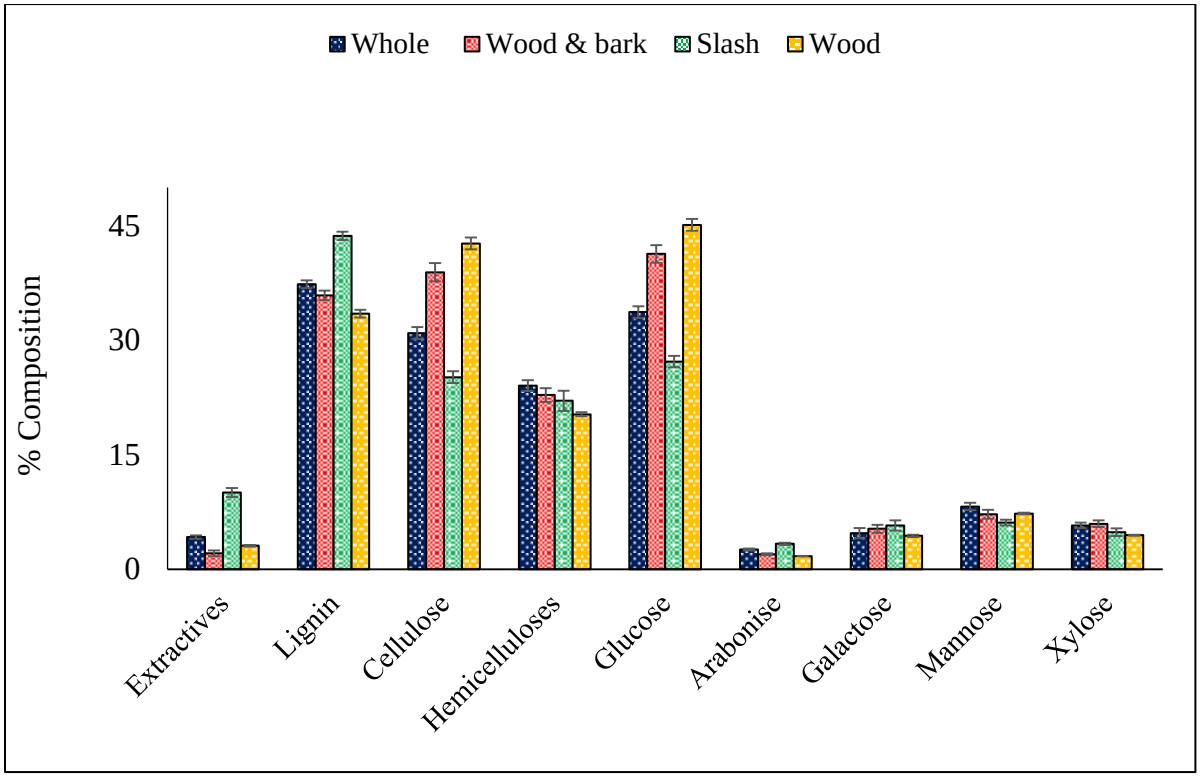

Figure 6.2: Descriptive statistics of the chemical composition of forest logging residue. *Bars represent ± standard error.

Using spectra as the independent variable and a measured property as the dependent variable, all forty biomass samples were employed in the calibration and cross validation of PLS predictive models. Models were developed using raw or 1st derivative spectra of first the entire MIR range, then the fingerprint region. As is generally the case, models built with 1st-derivative treated spectra have better predictive capabilities compared to those calibrated with untreated spectra, thus, only results of the former are presented in this paper (Table 6.1). Optimum models were chosen as those that used a lesser number of latent variables (LVs) to produce smaller error values. Another consideration in final model selection was a small difference between the SEC and SECV.

The SECV (which is a better measure of a model's predicting ability of future unknowns) is usually larger than the SEC (a statistic that evaluates how precisely the regression line fits the training data) because it also takes into consideration how much worse a model performs on an independent test data not originally included in model calibration. However, the SECV ideally should be no greater than 1.3 times the SEC (Sanderson et al. 1996). A big difference in SEC and SECV results when calibration models do a poor job of predicting the property under study for samples that were used in cross validation.

Table 6.1: Performance evaluation of PLS models developed using 1st derivative treated spectra of the full (i) and fingerprint (ii) regions for predicting chemical composition.

Constituent (%)	LVs		SEC		SECV		R^2		RPD	
	i	ii	i	ii	i	ii	i	ii	i	ii
Extractives	2	2	0.93	1.03	1.4	1.18	0.93	0.91	2.34	2.83
Lignin	2	2	1.58	1.77	2.02	2.04	0.86	0.83	2.06	2.04
Cellulose	2	2	3.89	4.04	6.1	4.58	0.74	0.72	1.46	1.61
Hemicelluloses	3	3	1.32	1.61	3.58	3.46	0.82	0.74	0.85	0.87
Glucose	2	2	3.71	3.88	4.6	4.4	0.77	0.75	1.63	1.7
Arabinose	2	2	0.36	0.4	0.46	0.46	0.77	0.72	1.57	1.6
Galactose	3	3	1.09	1.03	1.87	2.05	0.64	0.67	0.93	0.85
Mannose	3	3	0.85	0.82	1.87	1.84	0.71	0.73	0.81	0.84
Xylose	2	2	0.74	0.76	1.13	1.06	0.7	0.68	1.16	1.24
Holocellulose	2	2	3.92	4.25	6.05	4.79	0.73	0.68	1.45	1.53

Two or three LVs were used in the development of models that had R^2 values ranging from a low of 0.64 for galactose, to a high of 0.93 for extractives Table 1. Although the R^2 is an indicator of a good model (when greater than 0.5), it was not used as the sole assessor of models because it usually has a direct relationship with the number of LVs used in model development. When more

LVs are added in calibration, a model continues to fit random errors until every source of variation is accounted for in the training data (Bjorsvik and Martens 2011). The RPD, which is computed as the ratio of standard deviation of the validation set to the standard error of prediction (SEP) was used to compare the predictive ability of models. Except for galactose, xylose and hemicelluloses, the RPD values of models developed to predict the chemical components of forest logging residue fell within the preliminary screening criteria (i.e. 1.5 – 2.5) (Hein et al. 2009). The model for extractives was the most robust, having an RPD of 2.3.

PLS modeling of MIR spectra did a better job of predicting the extractives and lignin content of loblolly pine logging residue compared to the structural carbohydrates and their associated monomeric sugars in this study. The best performing models were for glucose and arabinose both of which had R^2 of 0.77 and RPD of 1.6; whereas the worst performing models were for galactose and mannose, the other two hexoses. Poor prediction of monomeric sugars have been attributed to similar conformation of sugars that only differ in the orientation of some hydroxyl groups. In a previous study by Acquah et al. (2015), galactose, mannose, xylose and consequently hemicelluloses were also poorly predicted by near infrared (NIR)-based PLS models. Similarly poor performing models were obtained by Jones et al. (2006) for galactose (R^2 = 0.11, RPD = 0.8) and hemicelluloses (R^2 = 0.30, RPD = 1.0). Since FTIR detects fundamental molecular vibrations as opposed to the overlapping and usually weaker combination bands in NIR, PLS models developed in this study were expected to do a better job of predicting the monomeric sugar content of forest logging residue, but this unfortunately was not the case.

The entire MIR range and fingerprint region were also used to model the lignin, cellulose and extractives of wood samples including Scots pine, Sitka spruce and tropical hardwoods from Ghana (Nuopponen et al. 2006). The authors reported the performance statistics of PLS models as

follows -- cellulose: R^2 = 0.65, SEC = 1.8, SEP = 3.3; lignin: R^2 = 0.65, SEC = 1.8, SEP = 3.3 and extractives: R^2 = 0.93, SEC = 0.3, SEP = 0.4. The seeming trend of infrared-based PLS models predicting the lignin and extractives of biomass relatively better than the polysaccharides was noted in this study also; and again in (Meder et al. 1999) when both diffuse reflectance (DRIFT) and transmission FTIR spectra were used in PLS modeling. A possible explanation of this trend could be the distinctive chemical structures of lignin and extractives, as opposed to the relative abundance of carbohydrates that have similar molecular makeup. Another study that quantitatively characterized the chemical composition of untreated wood was by (Tiovnen and Alen 2006). The researchers developed DRIFT-PLS models for lignin (R^2 = 0.66, SEP = 1), extractives (R^2 = 0.97, SEP = 0.9), arabinose (R^2 = 0.79, SEP = 0.1), galactose (R^2 = 0.80, SEP = 0.3), glucose (R^2 = 0.57, SEP = 1.7), mannose (R^2 = 0.63, SEP = 0.8) and xylose (R^2 = 0.73, SEP = 0.5). The standard deviation of the training data and prediction errors of monomeric sugars reported by the authors were low, even though the R^2 values are similar compared to what was obtained in current study.

For a fairer comparison of model performance, the same number of LVs that were retained as optimum for full spectra (4000 – 650 cm^{-1}) models were used in developing reduced spectra (1800 – 650 cm^{-1}) models. Reducing the wavenumber range to the fingerprint region did not adversely affect the performance of PLS models Table 6.1. In fact, this generally decreased the errors associated with cross-validation (employed as the SEP in current study) and improved RPD values for all models except that for lignin (Full: SECV = 2.02, RPD = 2.06; Reduced: SECV = 2.04, RPD = 2.04) and galactose (Full: SECV = 1.87, RPD = 0.93; Reduced: SECV = 2.05, RPD = 0.85) Table 6.1. Lowered SECV and improved RPD values is an indication that a model's predictive capability is reduced when irrelevant wavenumbers are included in model construction.

The relationship between laboratory reference data and FTIR-predicted chemical constituents are presented in Figure 6.3.

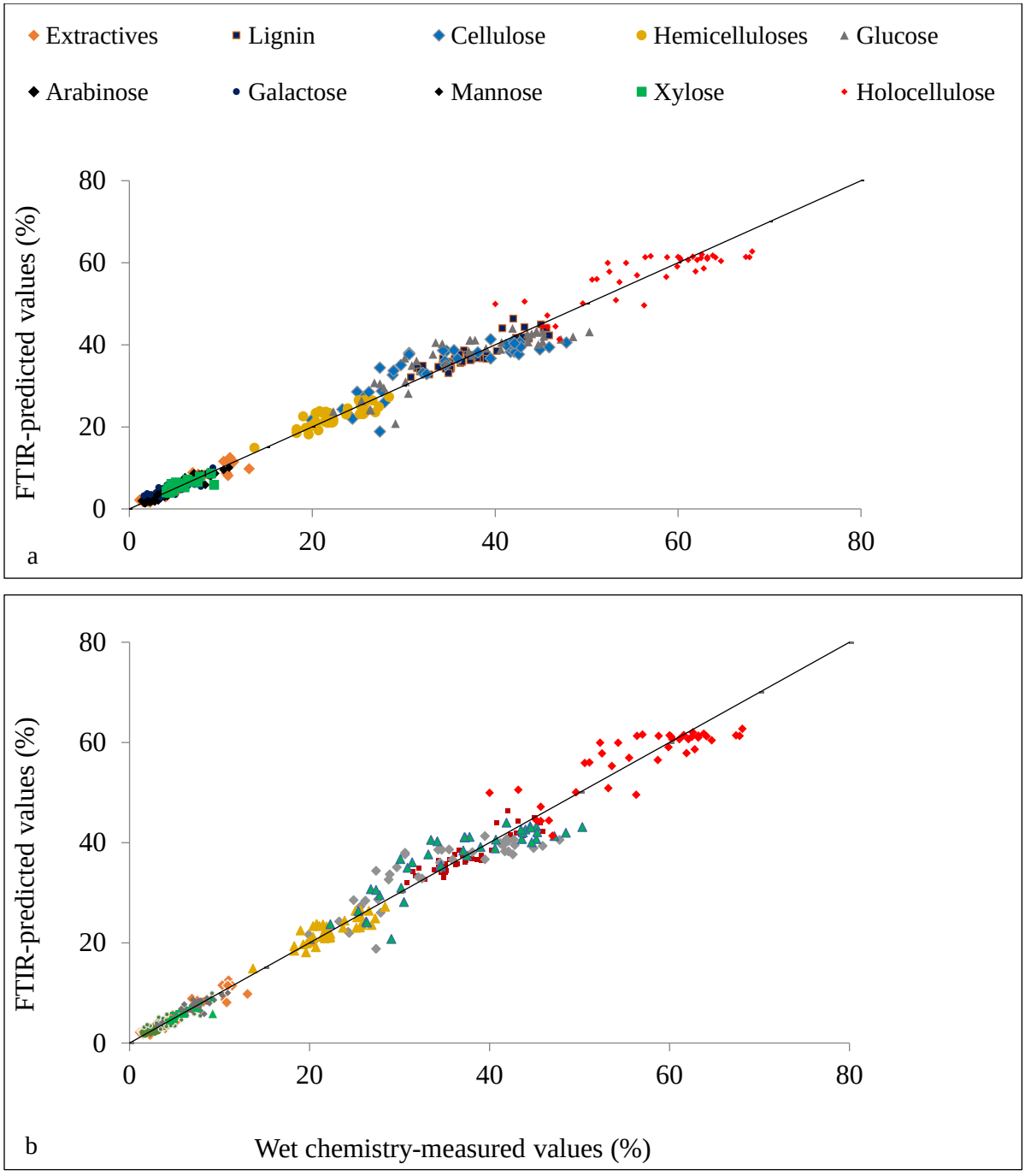

Figure 6.3: A regression plot of wet chemistry-measured versus FTIR-predicted values for chemical composition. (a) Modeled with full spectra; (b) - modeled with fingerprint region.

6.4.3 PLS Modeling of Thermal Reactivity and Energy Content of Forest Logging Residue

Summary statistics from proximate analysis and bomb calorimetry are presented in Figure 6.4. Ash content was significantly lower in Wood compared to the other three biomass types as expected. In contrast, Wood had the highest amount of volatile matter. Fixed carbon ranged from a low of 8.9% in Wood to a high of 16.2% in Slash. Among the four biomass types, Whole and Wood & bark samples were more similar in their thermal reactivity and energy content. The higher heating value (HHV), which is the maximum amount of energy that can be potentially recovered when fuel is completely combusted under adiabatic conditions ranged from 19.8 (MJ/kg) to 20.6 (MJ/kg) for loblolly pine logging residue.

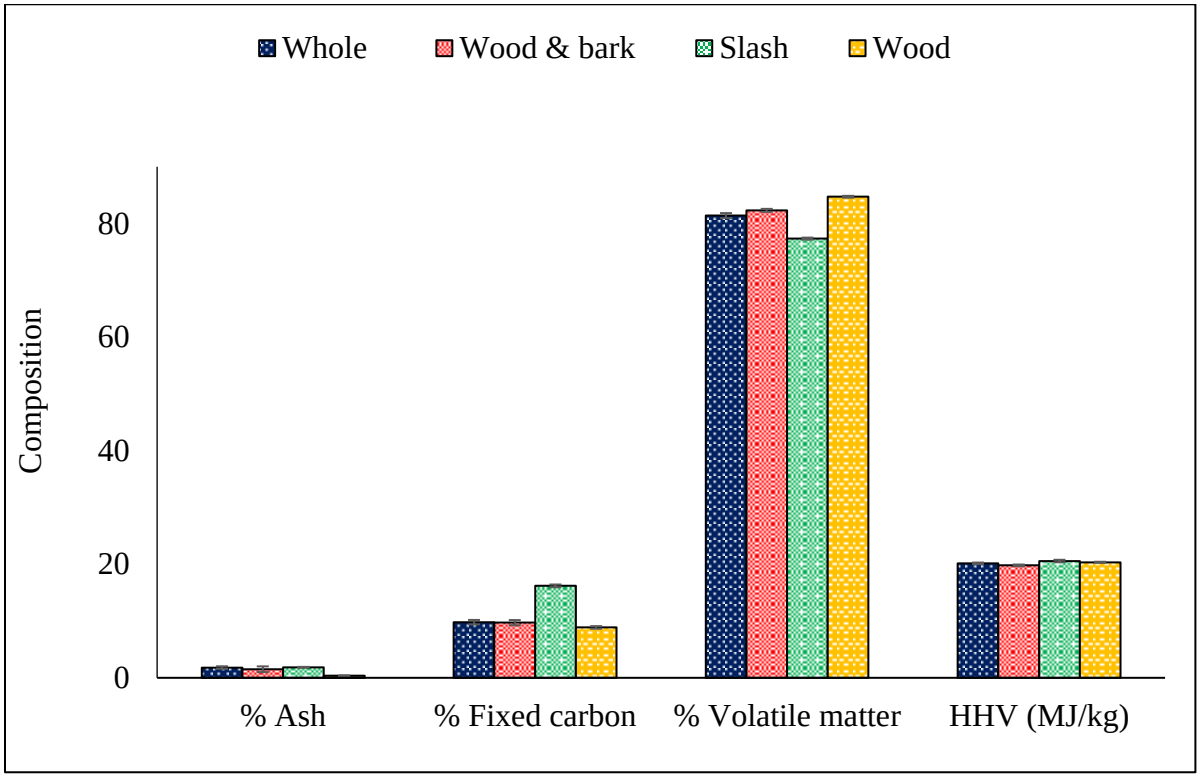

Figure 6.4: Descriptive statistics of the thermal reactivity and energy content of forest logging residue. *Bars represent ± standard error.

Two or three LVs were used in PLS modeling of the full or reduced MIR spectra. Fit statistics of cross-validated models calibrated with the 1st derivative of spectra are presented in Table 6.2.

Table 6.2: Performance evaluation of PLS models developed using 1st derivative treated spectra of the full (i) and fingerprint (ii) regions for predicting thermal reactivity and energy content.

Constituent	LVs		SEC		SECV		R^2		RPD	
	i	ii	i	ii	i	ii	i	ii	i	ii
Ash (%)	3	3	0.49	0.6	1.07	1.09	0.8	0.7	0.98	0.96
Fixed carbon (%)	2	2	1.26	1.35	1.6	1.54	0.85	0.83	1.96	2.04
Volatile matter (%)	2	2	1.07	1.03	1.31	1.15	0.87	0.88	2.17	2.31
HHV (MJ/kg)	2	2	0.34	0.38	0.53	0.44	0.64	0.54	1.03	1.23

For the reduced spectra of volatile matter content, two LVs gave the lowest values of SEC (1.03%) and SECV (1.15) falling within the ideal difference range. This optimized model had an R^2 of 0.88 and an RPD of 2.3. This was a 6% improvement over the RPD value of the model developed using the full MIR range. Similarly, utilizing the fingerprint region slightly improved the RPD value of the model for predicting percent fixed carbon. Correlation of ash content with spectra data were quiet high, although the RPD values were less than 1. Unlike for the organic components of forest biomass, developing models with the reduced spectra for the inorganic ash increased both the SEC and SECV, and reduced the R^2 and RPD values. Poor performance was also reported for full spectra FTIR-based PLS models constructed to predict the ash content of two energy crops (SEC - 1.02, SECV – 1.08, R^2 – 0.48) (Brink et al. 2010). However, Via et al. (2013) were able to better model the ash content of torrefied biomass. Simple monoatomic inorganic compounds do not produce any

vibrations in the mid (or near) infrared region. However, these form complexes with organic species to produce characteristic bands. As such, FTIR and NIR spectroscopy have been capable of quantitative and qualitative analysis of the ash in biomass, polymers, etc. (Stuart 2004).

Energy content is known to be influenced by the chemical composition of biomass. Lignin, (which can have as much as twice the calorific value of the carbohydrates) and extractives are mostly credited for this (FAO 1990). Consequently, FTIR which is sensitive to chemical signals has been used to model the HHV of biomass. PLS models constructed in this study to predict the HHV however did not perform very well. The R^2 values of models were 0.64 and 0.54; and RPD values were 1.03 and 1.23 using the full and reduced spectra respectively, Table 6.2.

Simple linear regression models were developed to explore the relationship between the chemical composition of forest logging residue and HHV. Only the extractive content had some meaningful linear correlation with HHV (R^2 = 0.31, *P-value* < 0.05); suggesting that the correlation between FTIR spectra and HHV is a secondary function of the correlation between the extractives and spectra. There have been conflicting reports in the literature about how especially lignin correlate with energy content (Demirbas 2002; Zhou et al. 2011). Comparing the regression spectra of extractives to that of HHV showed some common wavenumbers/peaks that made significant contributions in the prediction of the two properties; supporting results from the regression analysis, Figure 6.6. Peaks were noted at 1620 cm^{-1} (skeletal aromatic C=C in-plane vibration), 1440 cm^{-1} (C-O stretching, plus OH deformation of carboxylic acids or C-C stretching of aliphatic aldehydes) and 1190 cm^{-1} (C-O stretching of higher esters) (Parker 1983). However, unlike for the extractives, the peaks occurring in the regression spectrum of HHV could not account for as much of its variation, thus, the bad prediction performance of this model.

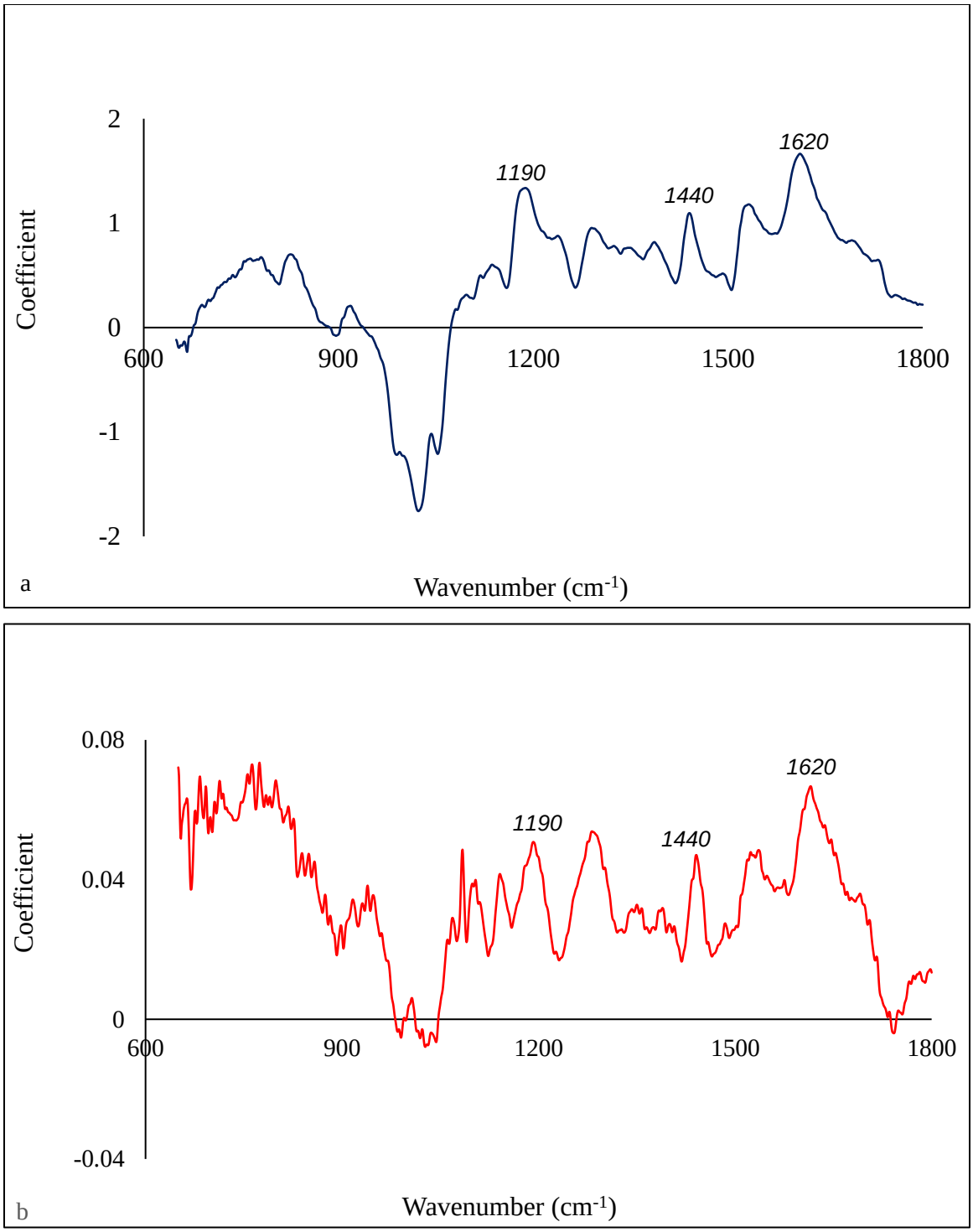

Figure 6.5: Regression spectra showing some common wavenumbers that made significant contribution to the modeling of extractives (%) (a) and HHV (MJ/kg) (b).

A scatter plot of how FTIR-based PLS models predicted the thermal reactivity and HHV as compared to results determined via proximate analysis and bomb calorimetry are presented in Figure 6.6.

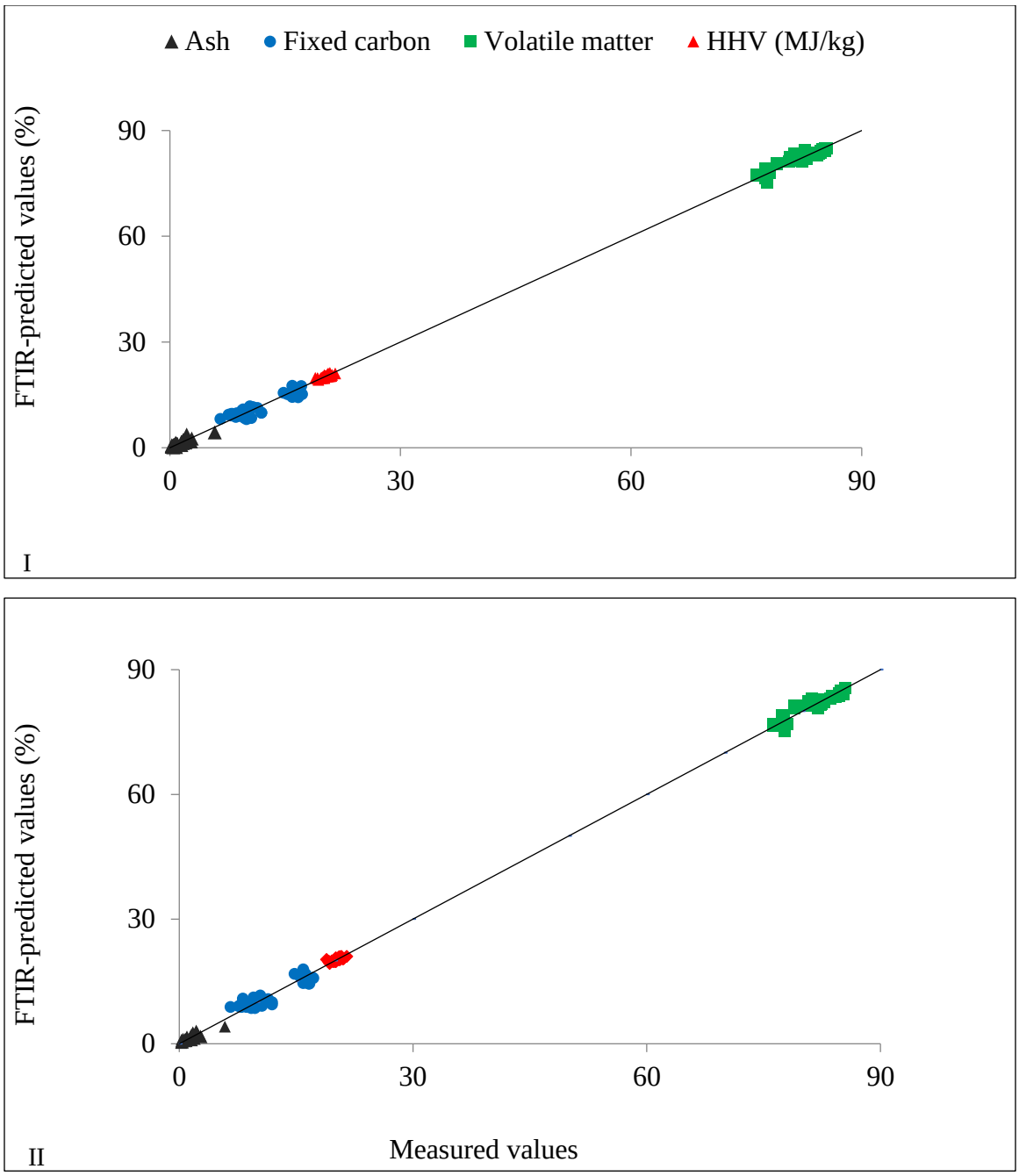

Figure 6.6: Regression plot of measured versus FTIR-predicted values for thermal reactivity and energy content. (a) Modeled with full spectra; (b) modeled with fingerprint region. Percent except for HHV.

6.5 Conclusions

FTIR spectra of forest logging residue made up of Whole tree, Wood & bark, Slash and Wood were acquired and related to the chemical, thermal reactivity and energy content of the biomass. PLS models were developed with the raw and 1st derivative of spectra spanning the entire MIR region or the fingerprint region. For chemical composition, developed models generally did a better job of predicting the extractives and lignin content than the carbohydrates; for the thermochemical properties, models for volatile matter and fixed carbon performed very well (i.e. R^2 > 0.80, RPD > 2.0). Reducing the wavenumber range to the fingerprint region for PLS modeling did not compromise the predictive ability of models. In fact, this mostly reduced the errors associated with prediction and improved the RPD values.

This study demonstrated that the chemical and thermochemical properties of forest logging residue can be predicted with FTIR coupled with PLS. The accuracies of prediction models constructed for this very heterogeneous biomass feedstock were comparable to that measured via lengthy and laborious conventional methods. The suite of important biomass properties understudied were predicted from a single FTIR spectrum without having to do any extra work for each of the properties. Thus FTIR can be employed as a high throughput tool for monitoring and characterizing this largely untapped resource to optimize processes in biorefineries that will depend on logging residues as new markets emerge and conversion technologies advance in the low carbon bioeconomy.

Chapter 7 : Chemometric Modeling of Thermogravimetric Data for the Compositional Analysis of Forest Biomass

7.1 Abstract

The objective of this study was to examine the use of chemometric modeling of thermogravimetric (TG) data as an alternative approach to estimate the chemical and proximate composition of lignocellulosic biomass. A capability to rapidly determine these properties is crucial in the optimization of bioconversion technologies.

The 38-minute long methodology developed in this study enabled the simultaneous prediction of both the chemical and proximate properties from the same TG data. Conventionally, two separate experiments had to be conducted to obtain such information. In addition, the chemometric models developed with normalized TG data outperformed models constructed via the traditional deconvolution of TG data. PLS and PCR models were especially robust in predicting the volatile matter (R^2 = 0.92, RPD = 3.58) and lignin (R^2 = 0.82, RPD = 2.40) contents of forest-derived biomass. The application of chemometrics to TG data also made it possible to predict some monomeric sugars. Moreover, the elucidation of PC loadings obtained from chemometric models provided some insights into the thermal decomposition behavior of the chemical constituents of lignocellulosic biomass. Further studies will however be necessary to validate the capability of chemometrics to model the thermal degradation and quantitative prediction of the individual monomeric sugars.

7.2 Introduction

The use of lignocellulosic biomass for energy and to replace other products derived from fossil fuel will reduce net greenhouse gas emissions and persistent toxic materials that result during the extraction and processing of fossil fuels. Furthermore, because of its widespread distribution, biomass utilization can present an opportunity for localities to develop new and innovative industries (NRCan 2007). Ligno-cellulosics may be herbaceous (annual) or non-herbaceous (perennial) and is made up of mainly cellulose, hemicellulose and lignin (Basu 2010).

The chemical composition of lignocellulosic biomass to a large extent determines the optimal conversion methodology and also affects the distribution and yield of products. The chemical compositional distribution can be complex due to presence of needles, bark or woody tissues (Acquah et al. 2015). Because of the recalcitrance of lignin during biochemical conversion processes, lignocellulosic biomass is usually converted via thermochemical conversion methods such as combustion, gasification and pyrolysis (Celikbag and Via 2016). When biomass is to be used as a source of energy or fuel, information about its proximate composition is necessary. Proximate analysis gives an indication of the thermal reactivity of biomass (McKendry 2002). It is used to measure the mass fraction of water, volatile matter, ash and fixed carbon (by difference) in lignocellulosic biomass. Biomass with high volatile matter content are easier to ignite and yield higher quantities of liquid products; whereas a higher fixed carbon gives more solid products. Ash is formed from incombustible minerals in biomass, and is increased when material is contaminated with soil during harvesting. Apart from decreasing the available energy, ash content also influences the choice of conversion pathway, the overall cost of processing and also creates pollution concerns. The chemical and proximate characteristics of a fuel feedstock impact the kinetics of degradation, thus, the efficiency and emission parameters of a processing plant.

Consequently, knowledge of the chemical and proximate composition of biomass is crucial in the optimization of conversion processes. Considering the heterogeneity of biomass, efficient operation of a bioconversion plant will require real time adaption of process parameters to the characteristics of the feedstock. In addition, the ability to determine these properties using rapid and cost-effective techniques will be necessary in the successful commercialization of bio-based products.

One technique with this potential is thermogravimetric analysis (TGA). TGA is a rapid type of thermal analysis that measures the change in mass as a function of temperature as a material is heated at a fixed rate under a set of conditions. The mass loss gives insight into a sample's chemical composition, thermal stability, number and sequence of reactions and kinetic parameters such as the order and activation energy of the chemical and physical reactions occurring (Klass 1998; Broido 1969). TGA has been a useful tool for determining the thermal decomposition behavior of biomass and the kinetic parameters required for the design and operation of thermochemical conversion equipment. It has widely been used in the characterization of forestry residues (Lapuerta et al. 2004), softwoods and hardwoods (Gronli et al. 2002), corn stover (Kumar et al. 2008) and municipal solid waste (Becidan 2007). TGA was also utilized to study the degradation temperatures and kinetic parameters of several understory grasses found in a *Pinus palustris* ecosystem (Elder et al. 2011). Employing TGA together with differential scanning calorimeter (DSC), Owens et al. (2015) determined the rate, kinetics and energy involved in the thermal degradation of loblolly pine biomass in both air and nitrogen. Systems integrating TGA with other analytical tools such as Fourier transform infrared spectroscopy (FTIR), gas chromatography (GC) and mass spectroscopy (MS) have also been used to enable the identification and quantification of the composition and

evolution rates of gaseous and liquid products during the pyrolysis and gasification of biomass feedstocks (Lee and Fasina 2009; Bassilakis et al. 2001; Dworzanski et al. 1991).

In recent times, a couple of studies have explored the application of TGA in the quantitative (Carrier et al. 2011; Saldarriaga et al. 2015) and qualitative (Francisco-Fernandez et al. 2012; Toscano et al. 2015) characterization of lignocellulosic biomass. Traditionally, researchers have determined the chemical composition of fuels by the deconvolution of derivative thermograms (DTGs), especially in quantitative analysis.

Chemometrics uses mathematical and statistical tools to extract pertinent information from chemical data (Varmuza and Filzmoser 2009). It has been used to determine the concentration of compounds in mixtures, identify substructures in unknown chemical compounds and predict their properties. The hypothesis is that, because the mass loss that occurs during TGA gives an indication of a material's chemical composition, chemometrics can be applied to TG data (also known as thermograms - TGs) to determine and then predict the thermo-chemical properties of lignocellulosic biomass. The objective of this study was to employ chemometric modeling of TG data as an alternative approach to estimate the chemical and proximate composition of biomass.

7.3 Materials and Methods

7.3.1 Sample Preparation

Biomass samples were obtained during harvesting operations on several *Pinus taeda* (loblolly pine) plantations in southern Alabama, USA. Materials acquired included the whole tree, wood and bark, slash (i.e. limbs and foliage) and clean wood chips of loblolly pine. Ten biomass sets were sampled for each of the four biomass types. In preparation for analysis, chipped materials were first ground to pass a 1/8 in. screen of a hammer mill, followed by further grinding in a Wiley

mill to pass a 40-mesh screen size. Test samples were stored in airtight plastic vials until time of analysis.

7.3.2 Methods

7.3.2.1 Compositional Analysis

The major chemical constituents of forest biomass (i.e. hemicellulose, cellulose and lignin) were determined via wet chemistry as specified in NREL LAP (2012). Air-dried extractive-free material was hydrolyzed with sulfuric acid in a two-step procedure.

The proximate composition was determined according to conventional standards. For ash determination, 1 g of test sample was weighed into a dry crucible and heated for 12 minutes at 105 ^{o}C. The temperature was raised to 250 ^{o}C, held isothermal for 30 minutes and then ramped up to 575 ^{o}C. The final temperature was maintained for another 180 minutes. The ash content was computed as given in NREL LAP (2008). Volatile matter content of biomass samples were determined as specified in CEN 15148 (2005) using a furnace (VMF Carbolite, model 10/6/3216P, England). Empty crucibles with their lids were first heated to 900 ^{o}C ± 10 ^{o}C for 7 minutes. After allowing to completely cool in a desiccator, they were filled with 1 g of test material, covered with lids and placed again in the furnace for 7 minutes. Unlike ash and volatile matter, fixed carbon of biomass is a calculated value. It is the summation of the percentage of moisture, volatile matter and ash deducted from 100.

Analysis of Variance (ANOVA) followed by Tukey pairwise comparison tests between the four biomass types ($\alpha = 0.05$) was performed using the R Stats Package. Duplicates experiments were ran for each test sample.

7.3.2.2 Thermal Analysis

Thermal decomposition of biomass samples was done in a Pyris 1 TGA thermogravimetric analyzer (PerkinElmer, Waltham, MA, USA) using different heating cycles. For the proximate analysis, adopting an earlier study by Acquah (2010) and ASTM E 872-82, 7 mg ± 2 mg of air dried test sample was heated from 30 ^{o}C to 105 ^{o}C at a rate 20 ^{o}C/min in an atmosphere of nitrogen. The temperature was held at 105 ^{o}C for 5 minutes, later ramped up to 800 ^{o}C at 50 ^{o}C/min and then held isothermal for 7 minutes. Next, air was introduced and maintained at 800 ^{o}C for an additional 10 minutes. The other method involved heating test samples from 30 ^{o}C to 800 ^{o}C at a rate of 20 ^{o}C/min in the presence of nitrogen. Both methods took approximately 38 minutes.

The TGs of samples were exported into excel for initial analysis. Several studies have been conducted in which the derivative of TG data have been used to show how the major chemical components thermally degrade in inert atmosphere at different temperature ranges. The amorphous, branched and lower molecular mass hemicelluloses is the first to decompose at mild temperatures (150 ^{o}C – 360 ^{o}C), followed by the linear, higher molecular mass cellulose (250 ^{o}C – 440 ^{o}C). The thermal degradation of lignin, a complex 3-dimensional polymer has been reported to occur over a wider range; from as low as 180 ^{o}C, or high as 300 ^{o}C, to 900 ^{o}C (Fasina and Lee 2009; Carrier et al. 2011; Naik et al. 2010; Yang et al. 2006; Garcia et al. 2013). The mass loss from room temperature to about 180 ^{o}C has been attributed to the loss of water and lower molecular mass volatiles. Several studies in recent times have also sought to determine the proximate composition (i.e. volatile matter, fixed carbon and ash contents) of biomass from thermogravimetric data (Acquah 2010; Garcia et al. 2013; Saldarriaga et al. 2015).

Based on the literature, two degradation regimes were adopted to be used for the quantitative computation of hemicellulose, cellulose and lignin. The following equations were employed: the chemical components were computed as follows:

$$\% \text{ Hemicellulose} = \{(B - C) / A\} * 100 \quad (1)$$

$$\% \text{ Cellulose} = \{(C - D) / A\} * 100 \quad (2)$$

$$\% \text{ Lignin} = \{D / A\} * 100 \quad (3)$$

For the first regime (Naik et al. 2010) which will be known as KIN-1 henceforth, A is the mass of test sample remaining after 130 ^{o}C, B is the mass after 250 ^{o}C, C is the mass after 350 ^{o}C, and D is the mass after 500 ^{o}C. For the second regime (Lee and Fasina 2009), dubbed KIN-2, A is the mass after 180 ^{o}C, B is the mass after 360 ^{o}C, C is the mass after 440 ^{o}C, and D is the sample mass after 440 ^{o}C.

Similarly, a degradation regime (i.e. KIN-3) was adopted (Acquah 2010) and used to calculate the amount of volatile matter, fixed carbon and ash using TG data. The computations were as follows:

$$\% \text{ Volatile matter} = \{(A - B)/A\} * 100$$

$$\% \text{ Ash} = \{C/A\} * 100$$

$$\% \text{ Fixed carbon} = 100 - \{\% \text{ Volatile matter} + \% \text{ Ash}\}$$

where A is the dry mass of test sample; B is mass at 800 °C after holding for 7 minutes in the presence of nitrogen and C is the residual mass after complete combustion in air.

Simple linear regression models were then developed to evaluate how these estimated properties compared to values experimentally obtained.

7.3.2.3 Chemometric Modeling

Chemometrics involve the application of multivariate analysis (MVA) to chemistry-relevant data. MVA uses many measured variables (X_1, X_2…..X_i) simultaneously to quantify a response or target variable (Y) (Martens and Naes 1990). In this case, X is thermogravimetric data and Y is the measured property. The PROC PLS package in SAS (SAS Institute, Inc. Cary, NC, USA) was used to develop both principal component regression (PCR) and partial least squares regression (PLS) models.

PCR is a two-step process involving principal component analysis (PCA) and multiple linear regression (MLR). PCA reduces the dimensionality of a dataset by taking a set of correlated X variables and transforming them into a smaller set of uncorrelated variables known as principal components (PC) scores. In other words, assuming that there are n observations X_{ij} on p correlated variables X_1, X_2,…,X_p, i = 1,…,n, j = 1,…, p, PCA finds new uncorrelated Z_1, Z_2,…,Zp that are linear combinations of X_1, X_2, …,X_p as

$$Z_i = e_{i1}X_{1i} + e_{i2}X_{2i} + \ldots\ldots + e_{ip}X_{pi} \quad \& \quad Var(Z_i) = \lambda_i\,,\ i=1,\ldots,p$$

where λ_is ($\lambda_1 > \lambda_2 > \ldots > \lambda_p$) and $\mathbf{e}_i$ are the eigenvalues and the corresponding eigenvectors of the covariance matrix of data matrix X (n by p). The coefficient, e_{ij} is a measure of the relevance of the j^{th} original variable to the i^{th} PC irrespective of the other variables. The coefficients, which are also known as eigenvectors or component loadings are proportional to the correlation between Zs and Xs; thus they can be used in the elucidation and interpretation of models. The values of the i^{th} principal component are called the PC scores (i.e. Zs). PCR then regress the PC scores against a response variable, Y.

The model structure for PLS is similar to that of PCR. However, unlike PCR, PLS takes the Y variable into consideration and generate latent variables scores (LVs, synonymous to PCs in PCA) in such a way that the covariance between X and Y is maximized.

Before exporting TG data into SAS (SAS Institute, Inc. Cary, NC, USA) for model construction, they were normalized to the dry mass (i.e. the temperature range associated with the loss of water: 30 °C to 105 °C were excluded). A total of forty samples (i.e. 10 of each biomass type) were used in the calibration and full-cross validation (also known as leave-one-out cross validation) of chemometric models. The PROC PLS procedure, using either the NIPALS (i.e. non-iterative PLS) or PCR algorithm gave the optimum models as those with the absolute minimum predicted residual sum of squares (PRESS). The final optimum models were chosen as those that used a lesser number of LVs or PCs to give a PRESS value that was not statistically different (Hotelling's T^2; $p > 0.1$) from the lowest PRESS value achieved with a higher number of LVs or PCs. The predictive performances of models were also assessed with the standard error of cross-validation (SECV), coefficient of determination (R^2) and ratio of preformance to deviation (RPD).

The chemical and proximate composition of forest residue used in this study is presented in Table 7.1.

Table 7.1: Properties of Loblolly Pine Biomass.

	Lignin	Cellulose	Hemi-celluloses	Ash	Volatile matter	Fixed carbon
Whole	37.3 a	31.0 a	24.1 a	1.8 a	81.4 a	9.8 a
Wood & bark	35.9 a	38.9 b	22.8 a b	1.5 a	82.3 a	9.7 a
Slash	43.7 b	25.2 c	22.1 a b	1.9 a	77.3 b	14.2 b
Wood	33.5 c	42.7 d	20.3 b	0.4 b	84.7 c	8.9 a
All	37.6 (4.2)	34.4 (7.4)	22.3 (3.0)	1.4 (1.0)	81.4 (2.8)	11.1 (3.1)

All values are expressed on % oven-dry basis. Statistically different [Tukey Test, $P<0.05$] properties noted with different letters. Mean values (SD in bracket). N = 10 for each group.

7.4 Results and Discussion

7.4.1 Thermogravimetric Analysis

Figure 7.1 shows representative TGs obtained for each biomass type when volatilized with the two temperature programs. The thermograms are an average (n = 10) for each of the four biomass types. In Figure 7.1(A), at about 150 ^{o}C, the devolatilization process starts, causing mass loss. The decomposition is steep until about 450 ^{o}C, after which the rate is more gradual. Maximum mass loss occurs approximately between 300 ^{o}C and 450 ^{o}C. The pyrolysis process removed 70% to 80% of the dry mass of forest biomass. The residual mass decreased drastically when air was introduced to facilitate combustion, Figure 7.1(B). Under both pyrolytic and combustive environments, wood had the least residual mass whereas slash had the highest. These results could be attributed to slash having a significantly higher percentage of the more thermally stable lignin and incombustible inorganics; while wood had the least, Table 7.1. The relatively more similar

chemical and proximate composition of whole and wood and bark were made evident in their overlapping thermograms.

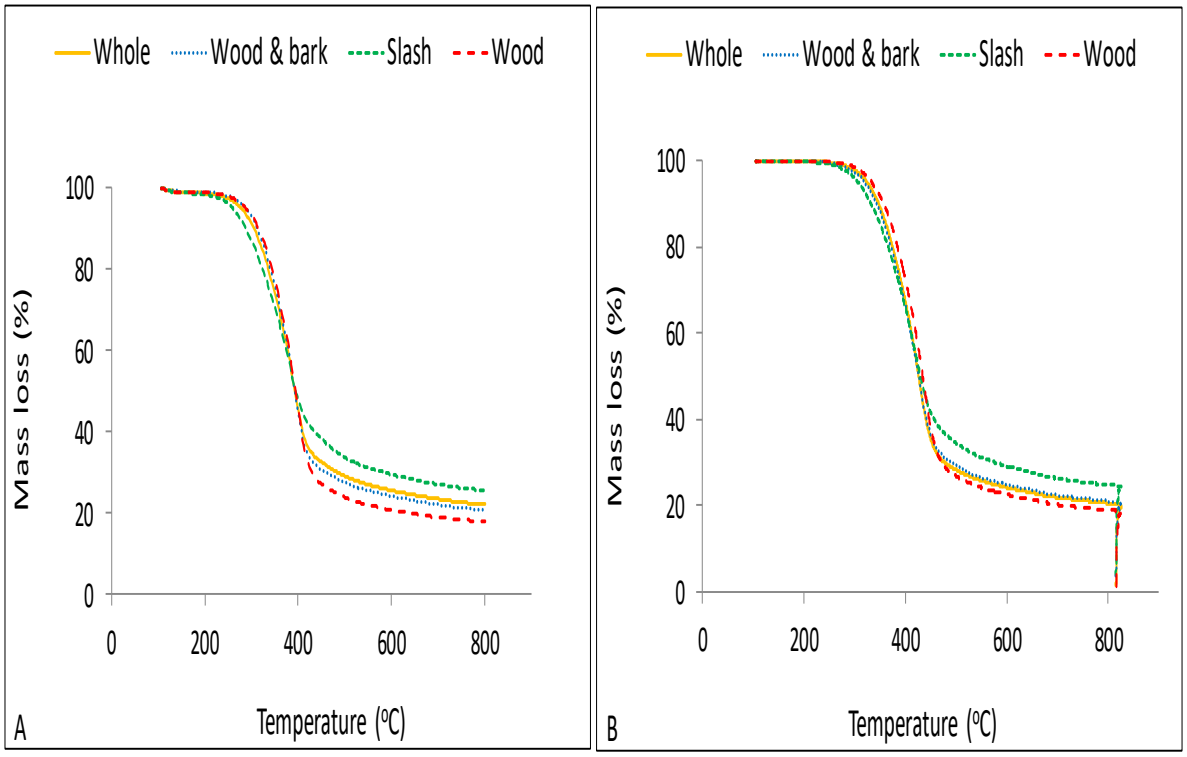

Figure 7.1: Mass loss from thermal decomposition of forest biomass in (A) nitrogen and (B) nitrogen plus air.

A DTG curve is a plot of the rate at which mass changes within a time range versus temperature (Bahadur and Sastry 2005). It can be used to determine the point at which mass loss is most apparent. From to the DTG curves in Figure 7.2, the highest rate of mass loss occurred at a slightly higher temperature for whole, wood & bark and wood (i.e. 420 °C) than it was for slash (400 °C). This behavior of slash could be as a result of the significantly higher concentration of easily volatilized amorphous extractives compared to the other biomass types (i.e. whole = 4%, wood & bark = 2%, slash = 10%, wood = 3%) (Chow and Pickles 1971). The bulk of mass loss happened from 150 °C to 550 °C as portrayed in the negative peak of the DTG.

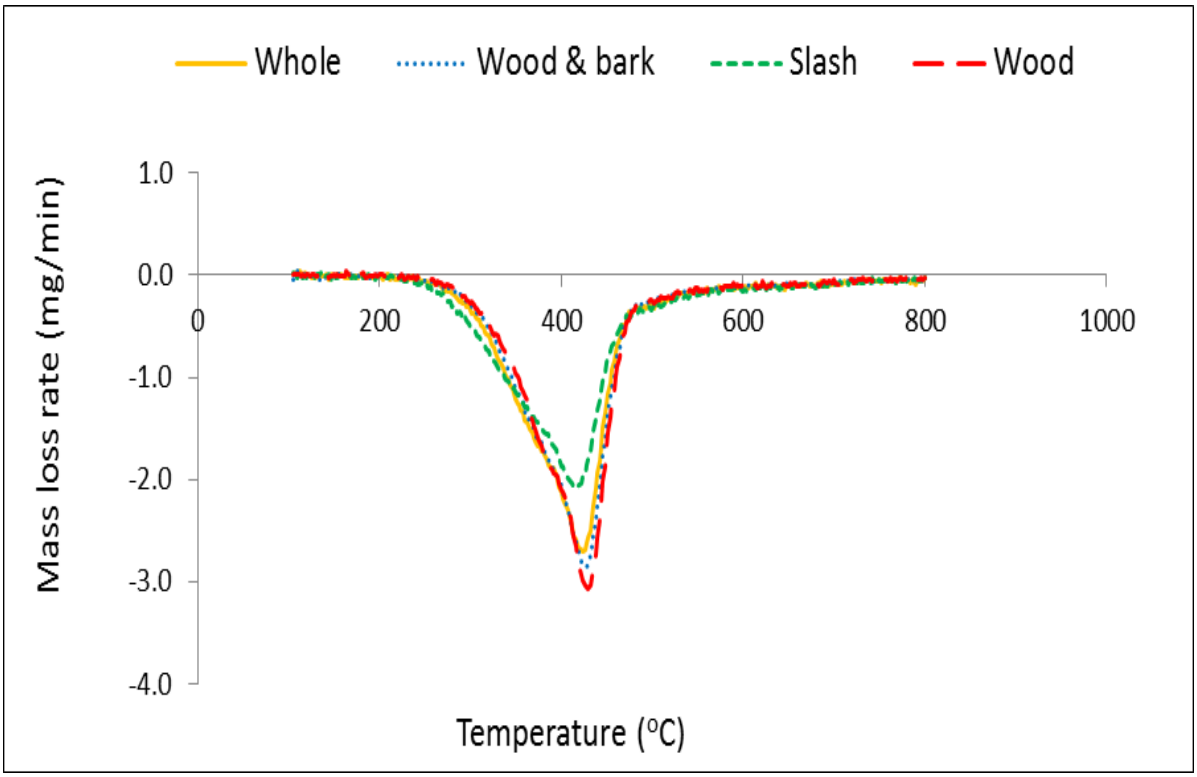

Figure 7.2: DTG curves of forest biomass in nitrogen plus air.

According to literature, this peak results from the overlap of the degradation of hemicellulose and celluloses. The presence of acetyl groups in the amorphous and branched hemicelluloses contribute to their relatively low thermal stability (Carrier et al. 2011; Yang et al. 2006; Saldarriaga et al. 2005). The smaller shoulder peak at (350 °C) attributed to the decomposition of hemicelluloses was not as prominent in whole, wood & bark and slash, as it was in wood, and other studies reported in the literature. When a similar occurrence was encountered by Owen et al. (2015) the researchers attributed it to higher ash content of the samples. Degradation of lignin has been reported to occur over a wider temperature range, with significant mass loss occurring after 550 °C. This gives the flat tail in the DTG of lignocellulosic biomass, Figure 7.2. There was a general upward shift in the temperature ranges that have been reported to

correspond to the degradation of the major chemical constituents of lignocellulosic biomass. This could be a consequence of the relatively high lignin content of samples used in this study, Table 7.1.

7.4.2 Chemometric Modeling for Property Prediction

Normalized TGs were used in the construction of PLS and PCR models. The PCR algorithm generally utilized more PCs to obtain the absolute minimum PRESS because the Y variables were not considered in the computations of PCs, Table 7.2. Considering parsimony, the final optimum models for prediction were chosen as those that used fewer LVs/PCs to give a PRESS value that was statistically the same as the absolute minimum PRESS; usually only two or three LVs/PCs could accomplish this. In the modeling of volatile matter, the minimizing number of factors and the smallest number of factors with $p > 0.1$ were the same. Once the final optimum LVs/PCs were chosen, both the PLS and PCR algorithms gave similar SEC and PRESS values, Table 7.2. This gave the assurance that the most stable and robust models had been selected.

Table 7.2: Calibration statistics of TG-based chemometric models.

	PLS				PCR			
	SEC	LVs (Opt)	LVs (Sig)	Press (Sig)	SEC	PCs (Opt)	PCs (Sig)	Press (Sig)
Lignin	1.63	6	2	0.48	1.63	10	2	0.48
Cellulose	3.44	3	2	0.58	3.45	4	2	0.58
Volatile matter	0.78	2	2	0.31	0.78	2	2	0.31
Fixed carbon	1.33	7	3	0.58	1.22	7	5	0.54
Ash	0.49	3	2	0.86	0.51	3	3	0.83

The predictive statistics of the constructed models are presented in Table 7.3. The PLS models performed slightly better in predicting the understudied properties than PCR models because LVs are extracted in a way that optimally explains the variation in both predictor and response variables (Varmuza and Filmoser 2009; Martens and Naes 1990). The R^2 -- which measures the degree of linear association between measured and predicted -- for cross-validated models ranged from a low of 0.32 (ash) to a high of 0.92 (volatile matter). Also, the RPD -- which gives an indication of the predictive adequacy of a model -- of models were from 0.59 (ash) to 3.58 (volatile matter). In the literature, a chemometric model with an R^2 greater than 0.5 could be used in several applications ranging in sensitivity; from rough screening to quality assurance (Nkansah et al. 2010). Likewise, a model that has an RPD of 1.5 or greater can be employed for preliminary prediction and screening (Hein et al. 2009).

For the major chemical constituent, TG-based models were especially able to predict the lignin content of forest biomass (R^2 – 0.82; RPD – 2.37 for both PLS and PCR), Table 7.3. The hemicelluloses were however poorly modeled when chemometrics was applied to TG data. Nonetheless, Carrier et al. (2011) were able to better estimate the hemicelluloses content of biomass by the deconvolution of DTGs. Their computations assumed a multi-component pyrolysis model in which hemicelluloses, cellulose and lignin have independent parallel reactions. However, as Cozzani et al. (1997) pointed out, these macro-components are intricately linked and the occurrence of their interactions during thermal degradation cannot be entirely overruled.

Table 7.3: Predictive performance of TG-based chemometric models.

	PLS			PCR		
	SECV	R^2	RPD	SECV	R^2	RPD
Lignin	1.76	0.82	2.37	1.76	0.82	2.37
Cellulose	4.01	0.70	1.85	4.02	0.70	1.85
Volatile matter	0.79	0.92	3.58	0.79	0.92	3.58
Fixed carbon	1.47	0.78	2.14	1.32	0.82	2.39
Ash	1.39	0.32	0.59	0.82	0.37	1.28

Note: Models developed with the smallest number of LVs/PCs that gave PRESS values statistically not different form the absolute minim PRESS.

Chemometric models constructed for the volatile matter content of biomass had the best predictive performance (R^2 – 0.92; RPD – 3.58). The predictive power for both PLS and PCR models for fixed carbon were also good. Unfortunately, models constructed for the ash content prediction didn't do as well. As TGs are a function of the mass loss of organics as samples are heated, the incombustibility of the inorganics might not have been adequately captured, as such, the poor performance of the ash models.

The performance of TG-based models developed in this study is similar to what have been reported in the literature for other widely utilized high throughput tools such as near infrared spectroscopy (NIR) (Acquah et al. 2015b; Ono et al. 2003) and Fourier transform infrared spectroscopy (FTIR) (Allison et al. 2009; Nuoponnen et al. 2006). TGA-based PLS models were constructed by Lande et al. (2010) to predict the furfuryl alcohol polymer content of *Pinus sylvestris* treated in a commercial plant. Depending on the pretreatment method used, as much as 94% of the variance could be accounted for. Comparing to NIR-based models, the authors concluded that both

tools had similar predictive powers; although more LVs were required in TGA models to attain this parity.

Regression coefficients of PCs (i.e. obtained from PCR) were investigated to identify temperatures that were important in modeling the various thermo-chemical constituents. PCs instead of LVs were employed for model interpretation because as mentioned earlier, the PCR algorithm only regards the X data when extracting its factors; as such most of the information in the original data is preserved. Also in an earlier study, we had determined that PC loadings worked better than LV loadings for model interpretation (Via et al. 2014). A plot of the regression coefficients is presented in Figure 7.3. Peaks were noted at about 320 ^{o}C and 410 ^{o}C for cellulose and lignin respectively. This is a suggestion that these temperatures had significant influence in the thermal degradation of the two chemical components. In the kinetic studies of TG/DTG of lignocellulosic biomass, thermal decomposition of cellulose have been determined to occur from 250 ^{o}C to 440 ^{o}C, while lignin degrades over a wider range (180 ^{o}C – 900 ^{o}C). Yang et al. (2006) reported that the maximum mass loss was obtained at 355 ^{o}C during the pyrolysis of pure cellulose. The authors could however not pinpoint the exact temperature at which the mass loss rate of lignin was highest. Instead, they noted that up until 700 ^{o}C, the rate of lignin degradation was slow (< 0.5 wt%/^{o}C) and this rate doubled at temperatures above 750 ^{o}C.

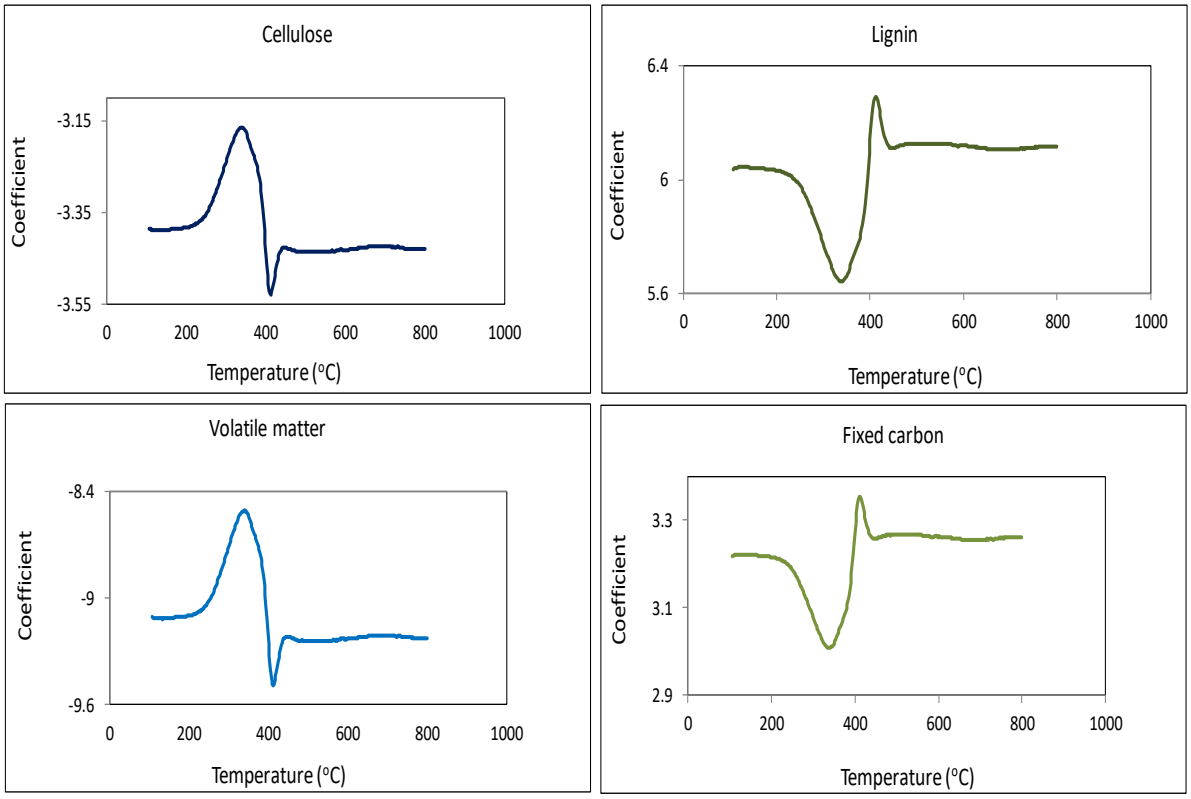

Figure 7.3: A plot of coefficients showing temperatures that had significant contribution in the prediction of thermochemical properties.

The loadings plot of volatile matter was identical to that of cellulose and those for fixed carbon and lignin also looked similar. This suggests common latent variables are shared between these chemical and thermal reactivity properties. These findings buttress what has been reported in the literature that the thermally less stable polysaccharides are responsible for the volatile matter while the lignin with its higher percentage of elementary carbon and low oxygen contributes to the yield of fixed carbon (Yang et al. 2006; Periera et al. 2013; Williams and Besler 2013). A simple regression of cellulose versus volatile matter and lignin versus fixed carbon provided strong linear correlations between these properties (cellulose vs. volatile matter: $R^2 = 0.7$, $p < 0.05$: lignin vs. fixed carbon: $R^2 = 0.77$, $p < 0.05$), reinforcing the literature.

7.4.3 Comparing Chemometric and Kinetic Models

Evaluation results of the performance of chemometric models compared to the kinetic models developed with temperature regimes adopted and modified from deconvolution studies reported in the literature are presented in Table 7.4.

Table 7.4: Predictive performance of TG-based chemometric models versus kinetic models.

		SEC	SECV	RPD	R^2
Lignin	PCR	1.63	1.76	2.37	0.82
	KIN-1	1.72	4.56	0.63	0.83
	KIN-2	1.72	2.65	1.57	0.81
Cellulose	PCR	3.45	4.02	1.85	0.7
	KIN-1	3.28	14.56	0.45	0.76
	KIN-2	3.22	5.95	1.25	0.71
Volatile matter	PCR	0.78	0.79	3.58	0.92
	KIN-3	0.74	1.45	1.57	0.90
Fixed carbon	PCR	1.22	1.32	2.39	0.82
	KIN-3	1.07	3.42	0.59	0.73
Ash	PCR	0.51	0.82	1.28	0.37
	KIN-3	0.40	0.91	0.58	0.46

Although the calibration errors associated with the two modeling approaches were similar, the cross validation errors were quiet high for the kinetic models; an indication that these models will poorly predict the understudied properties of future unknowns.

Between the kinetic models for chemistry, cellulose and lignin contents were better estimated with KIN-2 (i.e. cellulose as the mass loss occurring between 360 ^{o}C - 440 ^{o}C and lignin as the mass after 440 ^{o}C) than with KIN-1 (cellulose: 350 ^{o}C to 500 ^{o}C; lignin: after 500 ^{o}C). As can be seen from Table 7.4, the chemometric models outperformed the kinetic models in all instances. In addition to the superior predictive capability of TG-based chemometric models over the conventional deconvolution of DTG curves for quantitative analysis of biomass, chemometric modeling has several other advantages.

Usually when TGA has been used in proximate analysis, heating regimes include temperature ramps and isothermal conditions in both inert and reactive gaseous atmospheres (Acquah 2010; Garcia et al. 2013). On the other hand, when TGA is used in kinetic studies for chemical composition estimation, samples are typically heated at a constant rate in an inert environment (Yang et al. 2006; Carrier et al. 2011). As such, in order to estimate the chemical and proximate composition of biomass, two separate experiments have conventionally been conducted.

Using thermograms acquired under the inert conditions (total run time of 38 minutes), this study demonstrated that the chemical and thermal reactivity properties can be determined simultaneously when chemometrics is employed. In a recent study, Saldarriaga et al. (2015) were also able to determine several properties of lignocellulosic biomass from a single TG/DTG employing deconvolution and empirical modeling. The authors reported about a TGA procedure that occurred in both inert and oxidative environments. However, their methodology took over 240 minutes of run time per sample. Thus, the 38 minute procedure developed in this current study presents a big improvement in time saving.

Furthermore, the application of chemometrics to TG data enabled the quantitative modeling of some monomeric sugars. This has not been possible with the deconvolution of DTGs. The fit statistics of models for predicting hemicellulose and its associated monomeric sugars is presented in Table 7.5. The predictive performance of the glucose model was good, with R^2 of 0.77 and RPD of 2.11; thus meeting the criteria for preliminary screening. This was however not the case for mannose and galactose. Apart from these two hexoses having similar chemical structures to that of glucose (i.e. epimers), these also combine to form galactoglucomannans, the major hemicelluloses of softwoods. As such, analytical tools such as TGA might have some difficulty in distinguishing / predicting them. However, by employing chemometrics in this study, elucidation of the PC loadings gave some insight into their thermal degradation.

Table 7.5: Chemometric model statistics for monomeric sugars, hemicelluloses and holocellulose.

	PCs	SEC	SECV	RPD	R^2	R^2_{Adj}
Glucose	3	3.17	3.51	2.14	0.78	0.77
Galactose	2	0.69	1.55	1.12	0.19	0.17
Mannose	2	0.63	1.31	1.15	0.22	0.20
Xylose	2	0.66	0.97	1.34	0.43	0.42
Arabinose	3	0.33	0.38	1.92	0.72	0.72
Hemicelluloses	3	0.98	2.81	1.08	0.11	0.09
Holocellulose	2	3.43	4.09	1.76	0.67	0.66

A plot of the PC loadings is presented in Figure 7.4. PC 2 and PC 4 were essential in estimating the three sugars. The position of the peak in PC 2 suggests that significant devolatilization of these hexoses occurred at 338 ^{o}C. For mannose, PC 4 accounted for the most variation. PC 4 showed a more rounded peak ranging from 334 – 372 ^{o}C; with a maxima at 358 ^{o}C. This temperature shift could be due to the fact that mannose is the major component of the more stable backbone of galactoglucomannans (i.e. 0.1 - 1:1: 3 - 4) (Teleman 2009).

TGA performed better at predicting the five-ringed sugars of xylose and arabinose. The errors associated with the arabinose model were lower probably because this sugar is less similar to the abundant glucose, compared to xylose. PCs 2 and 4 once again made significant contributions in the modeling of this pentose. Xylose, which has been used in TGA kinetic studies to represent hemicelluloses was predicted with lesser success (R^2 = 0.43, RPD = 1.3). Unlike for the other monomeric sugars, PC 3 and PC 5 subsequently explained most of the variation in the thermal degradation of xylose, Figure 7.4. Decomposition of xylose started at a much lower temperature, with maximum mass loss occurring at 288 ^{o}C, Figure 7.4. Also, when Biagini et al. (2006) modeled commercial xylan as a model for hemicelluloses, they reported an onset degradation temperature of 253 ^{o}C, with bulk mass loss occurring at 299 ^{o}C. However, Werner et al. (2004) found out in their study that xylan decomposed in two stages, with significant mass loss at 243 ^{o}C - 292 ^{o}C. Results from the monomeric sugars chemometric models indicate that, thermal decomposition of hemicelluloses range from 282 ^{o}C to 372 ^{o}C. These are consistent with what have been reported in the literature using the deconvolution of thermograms (Fasina and lee 2009; Carrier et al. 2011; Naik et al. 2010; Yang et al. 2006).

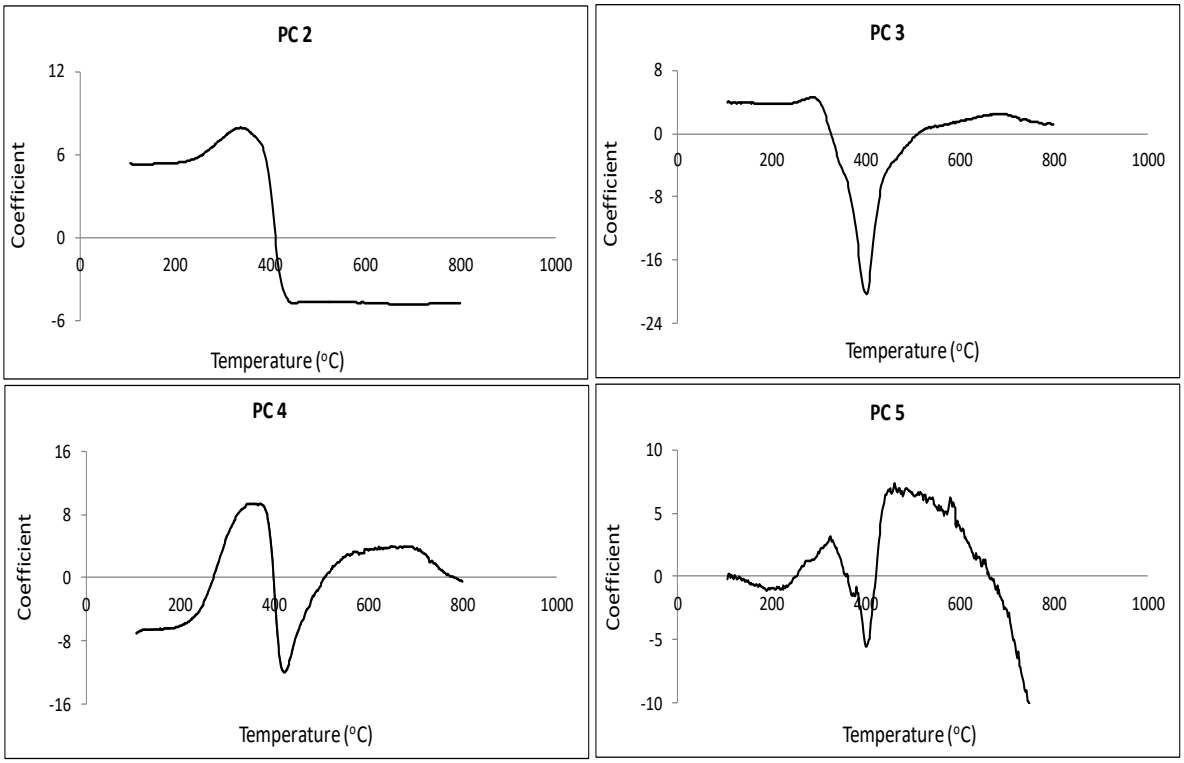

Figure 7.4: Loadings of PCs showing temperatures that had significant contribution in the prediction of chemical composition.

The poor prediction of the composite hemicelluloses could be a consequent of the difficulty in modeling mannose and galactose. However, TG-based chemometric models adequately predicted the holocellulose content of biomass. The holocellulose content could thus be used together with predicted cellulose content to provide a rough estimation of the hemicelluloses.

7.5 Conclusions

This study has demonstrated that chemometric modeling of thermogravimetric data can be used as an alternative approach to rapidly estimate the chemical and proximate composition of lignocellulosic biomass. Developed chemometric models had superior predictive capabilities than models constructed using the conventional deconvolution of TGs. PLS and PCR models calibrated with normalized TG data performed very well in predicting especially the lignin (R^2 - 0.82; RPD -

2.37) and volatile matter (R^2 - 0.92; RPD - 3.58) contents of forest-derived biomass. Examination of the loadings plots of PCR models suggested that significant degradation of cellulose and lignin occurred at around 320 ^oC and 410 ^oC respectively. Furthermore, these plots showed that common latent variables were shared between cellulose and volatile matter content; and between lignin and fixed carbon content.

The methodology developed in this study involved a 38-minutes procedure that allowed the simultaneous estimation of the chemical and proximate composition of lignocellulosic biomass from the same TG data. In addition to its rapidity and simplicity, this approach enabled the prediction of some monomeric sugars. Moreover, elucidation of PC loadings suggested that the thermal degradation of xylose started at a much lower temperature, with significant mass loss occurring at 288 ^oC, compared to the other monomeric sugars in lignocellulosic biomass. According to the literature, these have not been attainable by the conventional deconvolution of DTGs obtained from the composite lignocellulosic biomass. Further studies will be necessary to validate the capability of chemometrics to model the thermal degradation and quantitative prediction of the individual monomeric sugars.

Chapter 8 : Summary and Conclusions

8.1 Screening Elite Loblolly Pine Families for Structural, Chemical and Bioenergy Applications using NIR

Near infrared spectroscopy coupled with partial least squares was used to rapidly determine the density, strength properties, chemical composition and bioenergy potential of elite loblolly pine families. These properties will affect the processing of this essentially new feedstock, as well as the yield and quality of products to be derived. Furthermore, stakeholders would like to incorporate this information back into tree breeding programs that aim to further improve wood quality for different end users.

The loblolly pine families were obtained from two genetic research plantations established in 2000. Cross-validated models developed with 1st-derivative treated NIR spectra and conventionally-acquired laboratory data to predict the density, MOE and MOR of the elite families had R^2 values ranging from a low of 0.7 for density to a high of 0.75 for MOE; and the RPD values of all three models were higher than 1.5, the threshold that qualifies them as preliminary screening tools.

Genotype of the loblolly pine families affected density, MOR and MOE. The interaction of family and site was significant for density and MOR, but not for MOE. The range of density, MOR and MOE as predicted by NIR for the elite families were: 0.37 g/cm^3 (SD = 0.02) to 0.5 g/cm^3 (SD = 0.07); 34 MPa (SD = 10.2 MPa) to 150 MPa (SD = 14.4 MPa) and 7782 MPa (SD = 2237 MPa) to 10946 MPa (SD = 2703 MPa) respectively.

The MOR determined for the loblolly pine families were comparable to what has been reported in the literature for older loblolly pine trees, and the density higher than commercially sourced southern pines lumber. These results suggest that the elite families have improved the density and MOR. Further studies of the anatomical features of these families will however be required to ascertain these results.

The robustness of NIR-based models developed were buttressed when One-way ANOVA comparing NIR-predicted MOE and acoustics-predicted MOE of loblolly pine families from the two forest sites showed no statistical difference.

Finally, in spite of the significant family by site interaction, the structural-related properties of some families were consistently high or low irrespective of site. The density, MOR and MOE of A9, A1, A26 and A2 remained high whereas that of A33 and A21 remained low on the two study sites. Apart from the latter group being undesirable for structural applications, their low stiffness could also make them more vulnerable to inclement weather.

For the chemical and bioenergy potential of the elite loblolly pine families, NIR-predicted means for the studied properties were as follows: extractives – 5.5% (SD = 1.1%), lignin – 30.7% (SD = 1.3%), cellulose – 39.7% (SD = 1.4%), glucose – 44.7% (SD = 1.7%), hemicelluloses – 20.5% (SD = 0.4%), volatile matter – 84.2% (SD = 0.8%), fixed carbon – 14.9% (SD = 0.8%), ash – 0.2% (SD = 0.03%) and HHV 19.0 MJ/kg (SD = 0.2 MJ/kg).

Just as for the structural-related properties, genotype affected the chemical, proximate and energy traits of the elite loblolly pine families. Unlike for cellulose (*P-value* < 0.0001) and glucose (*P-value* = 0.001) contents, less variations were noted among the families with respect to the contents of lignin (*P-value* = 0.0186), hemicelluloses (*P-value* = 0.0773) and extractives (*P-value* = 0.0077). The genetic variation detected for cellulose was the the largest (*P-value* < 0.0001). Considering that

the cellulose content (which is under strong genetic control) has very good correlations with other properties, selecting and breeding for cellulose could generate some gains. For instance it was determined that, families that had higher concentrations of cellulose also had higher density, modulus of rupture (ultimate strength) and modulus of elasticity (strength). In addition, the amount of cellulose will affect the yield of pulp or ethanol. However, tree breeders must bear in mind that the desired traits of the elite families might be unstable on different sites. Further studies with more sites would thus help estimate the extent of this family by site interaction.

8.2 Rapid Assessment of Forest Biomass using NIR, FTIR and TGA

Apart from the contribution of loblolly pine wood to the conventional forest products industry, logging residues that will be available after harvest operations will play a key role in the emerging bioeconomy as new markets and technologies evolve.

NIR and FTIR were used to classify forest logging residue into the plant part components of wood, wood & bark, and slash. Being able to probe/monitor the variability of this feedstock means that the appropriate online adjustments to parameters could be made in time to ensure process optimization and product quality as biomass heterogeneity can influence its physical, chemical and thermochemical properties; and thus the final yield and quality of products. In addition, NIR and FTIR were used to determine the chemical composition, thermal reactivity and energy content of this resource. The ability to rapidly estimate these properties is vital in the optimization of conversion technologies for the successful commercialization of bio-based products.

Also, thermogravimetric analysis was used as another high throughput tool for the compositional analysis for forest-derived biomass. However, unlike the conventional kinetic modeling that is usually employed, this study explored the application of chemometric modeling of

thermogravimetric (TG) data as an alternative approach to estimate the chemical and proximate composition of lignocellulosic biomass.

For the qualitative analysis, five-fold cross-validated linear discriminant models that were developed with raw NIR and FTIR spectra had classification accuracies of over 96% in both cases. An extra factor/principal component (PC) was however needed to achieve this in FTIR modeling. Elucidation of factor loadings of both NIR and FTIR spectra showed that the statistically different amount of cellulose in the three plant part components of logging residue contributed to their initial separation.

For the quantitative assessment of biomass, partial least squares (PLS) regression models were used. In FTIR modeling, using only the fingerprint region (i.e. reduced spectra range) for model calibration did not compromise the predictive ability of models; instead, this mostly reduced the errors associated with prediction and improved the RPD values.

In predicting the chemical composition of the heterogeneous forest biomass, two or three latent variables were used in the development of models that had R^2 values ranging from 0.68 to 0.92, and RPD values of 0.87 to 2.83. The two tools had similar trends in their predictive capabilities. For instance in both cases, the models developed to predict the extractives content had better diagnostics whereas the models for predicting the hemicelluloses did not perform as well. In addition, the two tools did a better job of predicting the extractives and lignin content of forest biomass compared to the polysaccharides.

The energy content of biomass, a secondary trait which has been shown to have good correlation with the chemical components – especially extractives and lignin—where not predicted as well by both NIR and FTIR in this study. Likewise, PLS models developed for ash did not perform very well.

Comparing the FTIR regression spectra for the models for extractives and energy content, some common wavenumbers/peaks were noted to have made significant contributions in the prediction of the two properties, thus supporting the literature about the proportional relationship between these two properties.

With respect to using TGA for the compositional analysis of lignocellulosic biomass, the alternative approach of applying chemometrics to thermogravimetric data enabled the simultaneous prediction of both the chemical and proximate properties from a single thermogram. In addition, chemometrics made it possible to quantitatively model some monomeric sugars. These have not been achievable with the conventional deconvolution of TG data. PLS and principal components regression (PCR) models (i.e. the chemometric models) constructed with normalized TG data both outperformed models developed via the conventional deconvolution of TG data. PLS and PCR models were especially robust in predicting the volatile matter (R^2 = 0.92, RPD = 3.58) and lignin (R^2 = 0.82, RPD = 2.40) contents of forest-derived biomass. Between the two chemometric models, PLS performed slightly better in predicting some of the understudied properties than PCR.

Examination of the loadings plots of PCR models suggested that significant thermal degradation of cellulose and lignin occurred at around 320 ^oC and 410 ^oC respectively; whereas xylose started to decompose at a much lower temperature compared to the other sugars and lignin, with a significant mass loss occurring at 288 ^oC.

When the predictive performance of PLS models developed with TG data, NIR spectra and FTIR spectra were compared for some chemical components and the thermal reactivity properties, the NIR models generally had better diagnostics. For example, the coefficient of determination (R^2) between laboratory-measured and tool-predicted was higher for the NIR for % glucose, % cellulose

and % fixed carbon, Figure 8.1. Lastly, with the exception of ash content, the RPD values of models developed with all three analytical tools met the 1.5 preliminary screening criteria.

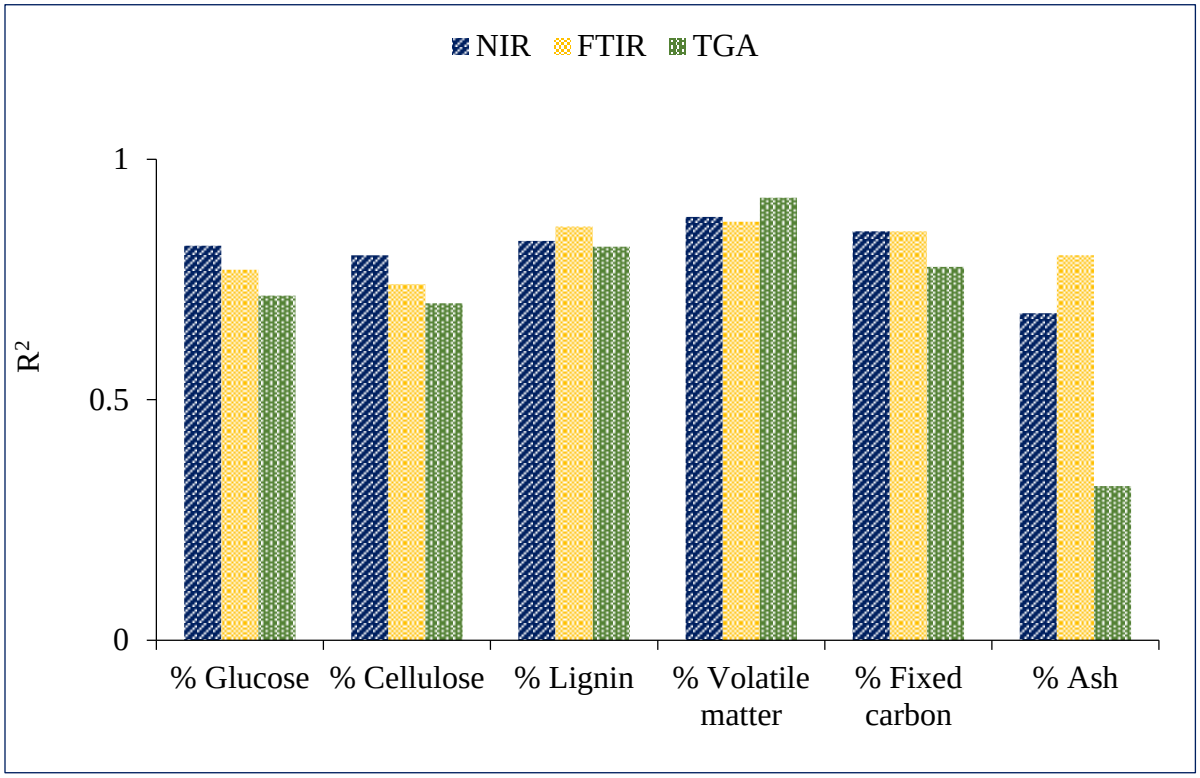

Figure 8.1: R^2 values of measured versus tool-predicted property.

8.3 Final Conclusions and Novelties

Near infrared spectroscopy, Fourier transform infrared spectroscopy and thermogravimetric analysis were used to rapidly and non-destructively determine several important properties that dictate the quality of wood and forest biomass as a feedstock for different utilization pathways.

The first focus of this study was on the screening of elite loblolly pine families, an essentially new resource for their structural integrity as well as for their chemical and bioenergy potential. This study is the first to determine an array of important wood traits from a single chemistry-sensitive data. With the developed methodology, it was possible to sample a large number of trees, a necessary requirement in tree breeding programs. Of the eleven wood traits studied, it was determined that genotype affected all. In addition, it was established that there were significant interactions between the families and the sites studied. This is valuable information for land owners and tree breeders as they keep in mind that a family with a desirable trait might not perform consistently when planted on different forest sites. Apart from the practical application of this information in the short term during the establishment of loblolly pine plantations, tree breeders can also use this information to further improve desirable traits in superior families in the long term.

The second emphasis of this dissertation was on forest biomass, an abundant but largely untapped resource. Prior to this current study, not much was known about the properties of this feedstock that has been projected to be a key contributor to the bioenergy portfolio as the United States strives for energy and national security. Moreover, apart from the more systematic characterization of this renewable resource, this study is the first to employ three high throughput analytical tools on the same materials with similar testing conditions and modeling techniques. As such, providing a more leveled comparison of the capabilities of these tools in the rapid assessment

of lignocellulosic biomass. Another novelty of this study is that, it demonstrated the potential of the chemometric modeling of thermogravimetric data. The developed methodology enabled the simultaneous prediction of both the chemical and proximate properties from data obtained in a single experimental run; this presents a significant gain in time savings.

This study made important first steps in the development of standard methodologies that can be leveraged towards the other hundreds of loblolly pine families in tree improvement programs, as well as other types of lignocellulosic biomass. Consequently, the right feedstock can rapidly be made available to support the conventional forest products industry, as well as new biofuels and chemical production initiatives.

References

- Acquah, G. E., Via, B. K., Fasina, O. O. and Eckhardt, L. G. 2015. Non-destructive prediction of the properties of forest biomass for chemical and bioenergy applications using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy.* 23: 93 – 102.
- Acquah, G. E., Via, B. K., Billor, N., Fasina, O. O. and Eckhardt, L. G. 2016. Identifying plant part composition of forest logging residue using infrared spectral data and linear discriminant analysis. *Sensors*. 16(9): Article ID 1376.
- Acquah, G. E., Krigstin, S., Wetzel. S., Cooper, P. and Cormier, D. 2016. Heterogeneity of forest harvest residue from eastern Ontario biomass harvests. *Forest Products Journal.* 66: 164 – 175.
- Acquah, G. E., Via, B. K., Eckhardt, L. G. and Fasina, O. O. 2016. Screening *Pinus taeda* (loblolly pine) families for physical and mechanical properties using vibrational spectroscopy. In: Schweitzer, C. J., Clatterbuck, W. K. and Oswalt, C. M. (eds.) *Proceedings of the 18th Biennial Southern Silvicultural Research Conference*. e–Gen. Tech. Rep**.**
- Acquah, G. E. 2010. Characterization of forest harvest residue from the Great Lakes-St Lawrence forests of South-eastern Ontario. MSc Thesis. Faculty of Forestry, University of Toronto, Canada.
- Adapa, P., Karunakaran, C., Tabil, L. and Schoenau, G. 2009. Potential applications of infrared and Rama spectromicroscopy for agricultural biomass. *Agricultural Engineering International*. 1081: XI.
- Adedipe O. E. and Dawson-Andoh B. 2008. Prediction of yellow poplar (*Liriodendron tulipifera*) veneer stiffness and bulk density using near infrared spectroscopy and multivariate calibration. Journal of Near Infrared Spectroscopy 16:487 – 496.
- Allison, G. G., Morris, C., Hodgson, E., Jones, J., Kubacki, M., Barraclough, T., Yates, N., Shield, I., Bridgwater, A. V. and Donnison, I. S. 2009. Measurement of key compositional parameters in two species of energy grass by Fourier transform infrared spectroscopy. *Bioresource Technology*. 100: 6428 – 6433.

- Alves, A. Santos, A., Rozenberg, P., Paque, L. E and Charpentier, J-P. 2012. A common near infrared – based partial least squares regression model for the prediction of wood density of *Pinus pinaster* and *Larix X eurolepis. Wood Science and Technology*. 46: 157-175.
- ASTM International. 2013. Standard test method for gross calorific value of coal and coke. ASTM Designation: D5865-13. West Conshohocken, PA.
- ASTM International. 2014. Standard test method for compositional analysis by thermogravimetric. Designation: E1131-08. West Conshohocken, PA.
- Bächle, H., Zimmer, B. and Wegener, G. 2012. Classification of thermally modified wood by FTNIR spectroscopy and SIMCA. *Wood Science and Technology.* 46: 1181 – 1192.
- Bahadur, P. and Sastry, N. V. 2005. Principles of Polymer Science. Second Edition. Alpha Science International Ltd. Harrow, Middlesex, U.K.
- Baker J. B. and Langdon O. G. 1990. *Pinus taeda* L., Loblolly Pine. In: Burns, R. M. and Honkala. B. H. (eds). *Silvics of North America, Vol. 1: Conifers.* Dept. Agric. Agric. Handb. 654, Washington, D.C. 497 – 512.
- Bassilakis, R., Carangelo, R. M. and Wojtowicz, M. A. 2001. TG-FTIR analysis of biomass pyrolysis. *Fuel*. 80: 1765 – 1786.
- Basu, P. 2010. Biomass gasification and pyrolysis. Practical design. Elsevier. Available at http://www.sciencedirect.com/science/book/9780123749888.
- Becidan, M. 2007. Experimental Studies on Municipal Solid Waste and Biomass Pyrolysis. PhD. Thesis, Norwegian University of Science and Technology.
- Belonger, P. J., McKeand, S. E. and Jett, J. B. 1997. Wood density assessment of diverse families of loblolly pine using x-ray densitometry. In: *Proceedings of the 24th Southern Forest Tree Improvement Conference*. pp. 133 – 142.
- Benito, M. T. J., Ojeda, B. C. and Rojas, F. S. 2008. Process analytical chemistry: Applications of near infrared spectroscopy in environmental and food analysis: An overview. *Applied Spectrosc*opy *Reviews.* 43: 452 – 482.
- Biagini, E., Barontini, F. and Togonotti, L. 2006. Devolatilization of biomass fuels and biomass components studied by TG/FTIR technique*. Industrial and Engineering Chemistry Res*earch. 45: 4486 – 4493.

- Bjorsvik, H-R. and Martens, H. 2011. Data analysis: calibration of NIR instruments by PLS regression. In: Burns, D. A. and Ciurczak, E. W. (eds), *Handbook of Near-Infrared Analysis*, 3rd edition. CRC Press.
- Blanco, M. and Villarroya, I. 2002. NIR spectroscopy: a rapid-response analytical tool. *Trends in Analytical Chemistry*. 21(4): 240 – 250.
- Bodirlau, R. and Teaca, C. A. 2009. Fourier transform infrared spectroscopy and thermal analysis of lignocellulose fillers treated with organic anhydrides. Romanian Journal of Physics. 54 (1–2): 93 – 104.
- Bradley, D. 2009. Canada report on bioenergy, 2009. Available at http://www.canbio.ca/ /publications/canadacountryreport2009.pdf
- Brereton, R.G. 2007. *Applied Chemometrics for Scientists*. John Wiley & Sons Ltd., West Sussex,
- Brink, M., Mandenius, C. F. and Skoglund, A. 2010. On-line predictions of the aspen fiber and birch bark content in unbleached hardwood pulp, using NIR spectroscopy and multivariate data analysis. *Chemometrics and Intelligent Laboratory Systems*. 103(1): 53 – 58.
- Broido, A. A. 1967. Simple, sensitive graphical method of treating thermogravimetric analysis data. *Journal of Polymer Science Part A-2-Polymer Physics.* 7: 1761 – 1773.
- Brown, H. D. and W. E. McDowell. 1968. Status of Loblolly Pine Die-off on the Oakmulgee District, Talladega National Forest, Alabama-1968. *US Department of Agriculture, Forest Service Report* no. 69 – 2: 28.
- Burdon, R. D., Britton, A. J. and Walford, G. B. 2001. Wood stiffness and bending strength in relation to density in four provenances of Pinus radiate. *New Zealand Journal of Forest Science*. 31(1): 130 – 146.
- Carballo-Meilan, A., Goodman, A. M., Baron, M. G. and Rodriguez, J. G. 2014. A specific case in the classification of woods by FTIR and chemometric: discrimination of Fagales from Malpighiales. *Cellulose*. 21: 261 – 273.
- Carrier, M., Loppinet-Serani, A., Denux, D., Lasnier, J-M., Ham-Pichavant, F., Cansell, F., et al. 2011. Thermogravimetric analysis as a new method to determine the lignocellulosic composition of biomass. *Biomass and Bioenergy*. 35: 298 – 307.

- Celikbag, Y. and Via, B. K. 2016. Characterization of residue and bio-oil produced by liquefaction of loblolly pine at different reaction times. *Forest Products Journal.* 66: 29 – 36.
- Chen, H., Ferrari, C., Angiuli, M., Yao, J., Raspi, C. and Bramanti, E. 2010. Qualitative and quantitative analysis of wood samples by Fourier transform infrared spectroscopy and multivariate analysis. *Carbohydrate Polymers*. 82: 772 – 778.
- Chow, S-Z. and Pickles, K. J. 1971. Thermal softening and degradation of wood and bark. *Wood* and *Fiber Science.* 3: 166 – 178.
- Conner, R. and Hartsell, A. 2002. Forest Area and Conditions. In: Wear, D. N. and Greis, J. G. (eds), *Southern forest resource assessment.* Gen. Tech. Rep. SRS-53. USD Forest Serv., Southern Res. Sta., NC-635.
- Cooper, P., Jeremic, D., Radivojevic, S., Ung, Y. T. and Leblon, F. 2011. Potential of near-infrared spectroscopy to characterize wood products. *Canadian Journal of Forest Research*. 41: 2150 – 2157.
- Cozzani, V., Lucchesi, A. and Stoppato, G. 1997. A new method to determine the composition of biomass by thermogravimetric analysis. *Canadian Journal of Chemical Engineering*. 75: 127 –133.
- de Jong, W., Pirone, A. and Wojtowicz, M. A. 2003. Pyrolysis of *Miscanthus giganteus* and wood pellets: TG-FTIR analysis and reaction kinetics. *Fuel.* 82(9): 1139 – 11.
- Defo, M., Taylor, A. M. and Bond, B. 2007. Determination of moisture and density of fresh sawn red oak lumber by near infrared spectroscopy. *Forest Products Journal.* 57: 68 – 72.
- Demirbas, A. 2001. Relationships between lignin contents and heating values of extractives. *Energy Conversion and Management*. 42 (2): 183 – 188.
- Demirbas, A. 2002. Relationships between heating value and lignin, moisture, ash and extractive contents of biomass fuels. *Energy Exploration & Exploitation*. 20(1): 105 – 111.
- Derkyi, N. S. A., Adu-Amankwa, B., Sekyere, D. and Darkwa, N. A. Rapid prediction of extractives and polyphenolic contents in *Pinus caribaea* bark using near infrared reflectance spectroscopy. *International Journal of Applied Science*. 2(1).
- Derkyi, S. A., Adu-Amankwa, B., Sekyere, D., and Darkwa, N. A. 2011. Application of near infrared spectroscopy in chemometric modeling of tannin content and stiasny number of

Pinus caribaea bark. *Journal of Emerging Trends in Engineering and Applied Sciences*. 2 (1): 132 – 136.

- Downes, G. M., Catela, F. and Meder, R. 2009. Developing and evaluating a multisite and multispecies NIR calibration for the prediction of Kraft pulp yield in eucalypts. *Southern Forests*. 71: 155 – 164.
- Dworzanski, J. P., Buchanan, R. M., Chapman, J. N. and Meuzelaar, H. L.C. 1991. Characterization of lignocellulosic materials and model compounds by combined TG/(GC)/FTIR/MS. *American Chemical Society, Division of Fuel Chemistry*. 36: 725 – 731.
- Earnest, C. M. (ed.), 1998. *Compositional Analysis by Thermogravimetry*. ASTM STP 997. American Society for Testing and Materials, Philadelphia, PA, USA.
- Eckhardt L. G., Menard R. D. and Gray E. D. 2009. Effects of oleoresins and monoterpenes on in vitro growth of fungi associated with pine decline in the Southern United States. *Forest Pathology*. 39 (3): 157 - 167.
- Elder, T., Kush, J. S. and Hermann, S. M. 2011. Thermogravimetric analysis of forest understory grasses. *Thermochimica Acta*. 512: 170 – 177.
- Esbensen, K. H. 2002. Multivariate data analysis in practice: An introduction to multivariate data analysis and experimental design, fifth ed. CAMO Process AS, Oslo, Norway, 598 p.
- Essien, C., Cheng, Q., Via, B. K, Loewenstein, E. F. and Wang, X. 2016. An acoustic operations study of loblolly pine standing saw timber with different thinning history. *BioResources.* 11(3): 7512 – 7521.
- Esteves, B. and Pereira, H. 2008. Quality assessment of heat-treated wood by NIR spectroscopy. *Holz Roh Werkst*. 66: 323.
- European Committee for Standardization (CEN). 2005. CEN/TS: 15148:2005: E. Solid biofuels – Method for the determination of the content of volatile matter.
- Evans, R. and Ilic, J. 2001. Rapid prediction of wood stiffness from microfibril angle and density. *Forest Products Journal.* 51(3): 53 - 57.
- Fagan, C. C., Everard, C. D. and McDonnell, K. 2011. Prediction of moisture, calorific value, ash and carbon content of two dedicated bioenergy crops using near-infrared spectroscopy. *Bioresource Technology.* 102: 5200 – 5206.

- Fahmi, R., Bridgwater, A. V., Donnison, I., Yates, N. and Jones, J. M. 2008. The effect of lignin and inorganic species in biomass on pyrolysis oil yields, quality and stability. Fuel 87: 1230 – 1240.
- Faix, O. 1992. Fourier transform infrared spectroscopy. In: Lin, S. Y. and Dence, C. W. (eds), *Methods in Lignin Chemistry*, Springer-Verlag.
- FAO Forestry Paper- 93. T0269/E 1990. Energy conservation in the mechanical forest industries. Available at http://www.fao.org/docrep/t0269e/t0269e08.htm#TopOfPage
- Fardim, P., Ferriera, M. M. C. and Duran, N. 2002. Multivariate calibration for quantitative analysis of eucalypt kraft pulp by NIR spectrometry. *Journal of Wood Chemistry and Technology*. 22(1): 67 – 81.
- Fengel, D. and Wegener, G. 1984. Wood: chemistry, ultrastructure and reaction. Walter de Gruyter, Berlin and New York, 613 pp.
- Francisco-Fernández, M., Tarrío-Saavedra, J., Mallik, A. and Naya, S. 2012. A comprehensive classification of wood from thermogravimetric curves. *Chemometrics and Intell*igent *Laboratory Systems.* 118: 159 – 172.
- Garcia, R., Pizarro, C., Lavin, A. G., Bueno, J. L. 2013. Biomass proximate analysis using thermogravimetry. *Bioresource Technology*. 139: 1 – 4.
- Gierlinger, N., Schwanninger, M., Hinterstoisser, B. and Wimmer, R. 2002. Rapid determination of the heartwood extractives in Larix sp. by means of Fourier transform near infrared spectroscopy. *Journal of Wood Science.* 10: 203 – 214.
- Gillon, D., Hernando, C., Valette, J-C. and Joffre, R. 1997. Fast estimation of the calorific values of forest fuels by near-infrared reflectance spectroscopy. Canadian Journal of Forest Research. 27: 760 – 765.
- Gindl, W., Teischinger, A. Schwanninger, M. and Hinterstoisser, B. 2001. The relationship between near infrared spectra of radial wood surfaces and wood mechanical properties. *Journal of Near Infrared Spectroscopy*. 9(4): 255-261.
- Grandmaison, J. L., Ahmed, A., Kaliaguine, S. and Chantal, P. D. 1987. Analysis of partially converted lignocellulosic materials. *Analytical Chemistry* 59(17): 2153 – 2157.

- Green, D. W., Winandy, J. E. and Kretschmann, D. E. 1999. Mechanical properties of Wood. In: Wood Handbook – Wood as an engineering material. Gen. Tech. Rep. FPL–GTR–113. Madison, WI.
- Greene, D., Baker, S., Mendell, B. and Lang, A. H. 2011. Integrating woody biomass into the U.S. South wood supply chain. 34th Council on Forest Engineering. Available @ https://www.cirrelt.ca/cofe2011 /proceedings/17-Greene_COFE2011_Final.pdf
- Grønli, M. G., Varhegyi, G. and Di Blasi C. 2002. Thermogravimetric analysis and devolatilization kinetics of wood. *Industrial and Engineering Chemistry Research.* 41: 4201 – 4208.
- Haaland, D. M. and Thomas, E. V. 1988. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Analytical Chemistry*. 60: 193 – 202.
- Haartveit, E.Y. and Flæte, P. O. 2008. Rapid prediction of basic wood properties by near infrared spectroscopy. *New Zealand Journal of Forestry Science*. 36(2/3): 393 – 407.
- Hacker, J. J. 2006. Effects of logging residues removal on forest sites – a literature review. Available at http://dnr.wi.gov/topic/forestbusinesses/documents/ loggingresiduereport.pdf
- Hakkila P., and Parikka M., 2002. Fuel resources from the forest. In: Richardson J., Bjorheden R., Hakkila P., Lowe A.T. and Smith C.T., (eds). Bioenergy from sustainable forestry: Guiding principles and practice. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Hames, B. R., Thomas, S. R., Sluiter, A. D., Roth, C. J. and Templeton, D. W. 2003. Rapid biomass analysis. New tools for compositional analysis of corn stover feedstocks and process intermediated from ethanol production. *Applied Biochemistry and Biotechnology.* 105 (108): 5 – 16.
- Harrington, K. J., Higgins, H. G. and Michell, A. J. 1964. Infrared spectra of *Eucalyptus regnans* F. Muell. and *Pinus radiata* D. Don. *Holzforschung*. 18(4): 108 – 113.
- Haygreen, J. G. and Bowyer, J. L. 1989. *Forest Products and Wood Science*, 2nd Ed. Iowa State University Press, Ames, Iowa – 500.
- Hein, P. R. G., Campos, A. C. M., Trugilho, P. F., Lima, J. T. and Chaix G. 2009. Infrared spectroscopy for estimating wood basic density in *Eucalyptus urophylla* and *Eucalyptus garndis*. *Cerne, Lavras.* 15(2): 133 – 144.

- Hein, P. R. G., Clair, B., Brancheriau, L. and Chaix, G. 2010. Predicting microfibril angle in Eucalyptus wood from different wood faces and surface qualities using near infrared spectra. *Journal of Near Infrared Spectroscopy.* 18(6): 455 – 465.
- Hein, P. R. G., Lima, J. T., and Chaix, G. 2010. Effects of sample preparation on NIR spectroscopic estimation of chemical properties of *Eucalyptus urophylla* S. T. Blake wood. *Holzforschung.* 64: 45 – 55.
- Hein, P. R. G., Silva, J. R. M. and Brancheriau, L. 2013. Correlations among microfibril angle, density, modulus of elasticity, modulus of rupture and shrinkage in 6-year-old *eucalyptus urophylla × E. grandis. Maderas. Ciencia y tecnología.* 15(2): 171 - 182.
- Heise, H. M. and Winzen, R. 2002. Fundamental chemometric methods In: Siesler, H. W., Ozaki, Y., Kawata, S., and Heise, H. M. (eds), *Near Infrared Spectroscopy: Principles, Instruments, Applications*. Weinheim: Wiley-VCH, 348 p.
- Hergert, H. L. 1971. Infrared spectra. In: Sarkanen, K. V. and Ludwig, C. H. (eds), *Lignins: Occurrence, Formation, Structure and Reactions*, Wiley Interscience.
- Hobro, A. J., Kuligowski, J., Doll, M. and Lendl, B. 2010. Differentiation of walnut wood species and steam treatment using ATR-FTIR and partial least squares discriminant analysis (PLSDA). *Analytical and Bioanalytical Chemistry.* 398: 2712 – 2722.
- Hodge, G. R. and Woodbridge, W. C. 2006. Use of near infrared spectroscopy to predict lignin content in tropical and sub-tropical pines. *Journal of Near Infrared Spectroscopy.* 12: 381 – 390.
- Ivkovic, M., Gapare, W. G., Abarquez, A., Ilic, J., Powell, M. B. and Wu, H. X. 2009. Prediction of wood stiffness, strength and shrinkage in juvenile wood of radiate pine. *Wood Science and Technology*. 43: 237 – 257.
- Jernigan, J., Gallagher, T., Mitchell, D., Smidt, M. and Teeter L. 2016. High Tonnage Harvesting and Skidding for Loblolly Pine Energy Plantations. *Forest Products Journal*. 66(3-4): 185 – 191.
- Jiang, W., Han, G., Via, B. K., Tu, M., Liu, W. and Fasina, O. 2014. Rapid assessment of coniferous biomass lignin-carbohydrates with near-infrared spectroscopy. *Wood Science and Technology*. 48 (1): 109 – 122.

- Jiang, W., Via, B. K., Han, G., Wang, Q. and Liu, S. 2014. Near infrared monitoring of untreated and chemically delignified wood. *Journal of Near Infrared Spectroscopy*. 21(6): 485 – 493.
- Jiang, Z. H., Yang, Z., So, C.-L. and Hse, C.-Y. 2007. Rapid prediction of wood crystallinity in *Pinus elliotii* plantation wood by near-infrared spectroscopy. *Journal of Wood Science.* 53: 449 – 453.
- Johnson, R. A. and Wichern, D. W. 2007. *Applied Multivariate Statistical Analysis*, 6th Ed. Pearson Education Inc, Upper Saddle River.
- Jones, P. D., Schimleck, L. R., Daniels R. F., Clark III, A. and Purnell, R. C. 2008. Comparison of *Pinus taeda* L. whole-tree wood property calibrations using diffuse reflectance near infrared spectra obtained using a variety of sampling options. *Wood Science and Technology*. 42: 385 – 400.
- Jones, P. D., Schimleck, L. R., Peter, G. F., Daniels, R. F. and Clark (III), A. 2006. Nondestructive estimation of wood chemical composition of sections of radial wood strips by diffuse reflectance near infrared spectroscopy, *Wood Science and Technology*. 40: 709 – 720.
- Jones, P. D., Schimleck, L. R., Peter, G. F., Daniels, R. F., Clark III, A. 2005. Nondestructive estimation of *Pinus taeda* L. wood properties for samples from a wide range of sites in Georgia. *Canadian Journal of Forest Research*. 5:85 – 92.
- Jones, P. D., Schimleck, L. R., Peter, G. F., Daniels, R.F. and Clark III, A. 2005. Non-destructive estimation of *Pinus taeda* L. tracheid morphological characteristics for samples from a wide range of sites in Georgia. *Wood Science and Technol*ogy. 39: 529 – 545.
- Jordan, L., Clark III, A., Schimleck, L. R., Hall, D. B and Daniels, R. F. 2008. Regional variation in wood specific gravity of planted loblolly pine in the United States. *Canadian Journal of Forest Research*. 38: 698 – 710.
- Jorgensen, A. 2000. Clustering excipient near infrared spectra using different chemometric methods. Seminar summary. Available at http://www.pharmtech.helsinki.fi/seminaarit/vanhat/annajorgensen.pdf
- Jorgensen, B. and Goegebeur, Y. 2007. Module 1: Chemometric and NIR spectroscopy. Available at http://statmaster.sdu.dk/courses/ST02/module01/module.pdf.

- Kalisz, S., Svoboda, K., Robak, Z., Baxter, D. and Andersen, L. K. 2008. Application of FTIR absorption spectroscopy to characterize waste and biofuels for pyrolysis and gasification. *Archives of Waste Management and Environmental Protection* 8: 51 – 62
- Kelley, S. S., Rials, T. G., Snell, R., Groom, L. H. and Sluiter, A. 2004. Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood. *Wood Science and Technology*. 38: 257 – 276.
- Klass, D. L. 1998. *Biomass for renewable energy, fuels and chemicals*. Academic Press.
- Kohan N. J., Via B. K. and Taylor S. E. 2012. Prediction of strand feedstock mechanical properties with near infrared spectroscopy. *BioResource* 7(3): 2996 – 3007.
- Kube, P. and Raymond, C. 2002. Prediction of whole-tree basic density and pulp yield using wood core samples in *Euclalyptus nitens*. *Appita Journal.* 55(1):43 – 48.
- Kumar, A., Wang L., Denis, Y. A., Jones, D. D. and Hanna, M. A. 2008. Thermogravimetric characterization of corn stover as gasification and pyrolysis feedstock. *Biomass and Bioenergy*. 32: 460 – 467.
- Labbé, N., Seung-Hwan, L., Hyun-Woo, C., Jeong, M. K. and André, N. 2008. Enhanced discrimination and calibration of biomass NIR spectral data using non-linear kernel methods. *Bioresource Technology.* 99: 8445 – 8452.
- Lachenbruch, B., Moore, J. R. and Evans, R. 2011. Radial variation in wood structure and function in woody plants, and hypothesis for its occurrence. In: Meinzer, F. C., Lachenbruch, B. and Dawson, T. E. (eds), *Size- and Age-Related Changes in Tree Structure and Function*. Springer. 514 p.
- Lande, S., van Riel, S., Høibø, O. A. and Schneider, M. H. 2010. Development of chemometric models based on near infrared spectroscopy and thermogravimetric analysis for predicting the treatment level of furfurylated Scots Pine. *Wood Science and Technology.* 44: 189 – 203.
- Lapuerta, M., Hernandez, J. J. and Rodriguez, J. 2004. Kinetics of devolatilisation of forestry wastes from thermogravimetric analysis. *Biomass and Bioenergy.* 27: 385 – 391.
- Larson, P. R., Kretschmann, D. E. Clark, A. III and Isebrands, J. G. 2001. Formation and properties of juvenile wood in southern pines: a synopsis. Gen. Tech. Rep. FPL-GTR-129. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 42 p.

- Lee, S. and Fasina, O. 2006. TG-FTIR analysis of switchgrass pyrolysis. *Journal of Analytical and Applied Pyrolysis.* 86: 39 – 43.
- Leinonen, A. 2004. Harvesting technology of forest residues for fuel in the USA and Finland, VTT Tiedotteita – Research Notes 2229. 132 p. + app. 10. p. Available at http://www.vtt.fi/inf/ pdf/tiedotteet/ 2004/T2229.pdf.
- Lestander, T. A. and Rhen, C. 2005. Multivariate NIR spectroscopy models for moisture, ash and calorific content in biofuels using bi-orthogonal partial least squares regression. *Analyst.* 130: 1182 – 1189.
- Li, C., Cheng, G. and Balan, V. 2011. Influence of physic-chemical changes on enzymatic digestibility of ionic liquid and AFEX pretreated corn stover. *Bioresource Technology.* 102(13): 6928 – 6936.
- Li, X., Sun, C., Zhou, B. and Yong, He. 2015. Determination of hemicellulose, cellulose and lignin in moso bamboo by near infrared spectroscopy. *Scientific Reports*. 5(17210).
- Lupoi, J. S., Singh, S., Davis, M., Lee, D. J., Shepherd, M., Simmons, B. A. and Henry, R. J. 2014. High-throughput prediction of eucalypt lignin syringyl/guaiacyl content using multivariate analysis: A comparison between mid-infrared, near-infrared, and Raman spectroscopies for model development. *Biotechnology for Biofuels*. 7(1): 93.
- Maranan, M. C. and Laborie, M. P. G. (2007). Analysis of energy traits of populous spp. Clones by near-infrared spectroscopy. *Journal of Biobased Materials and Bioenergy*. 1: 155-162.
- Mark, H and Campbell, B. 2008. An introduction to near infrared spectroscopy and associated chemometrics. Available @ http://www.idrc-chambersburg.org/cnirs/pdfs /nirandchemometrics.pdf
- Martens, H. and Naes, T. 1989. Multivariate Calibration, Wiley, New York, 419 p
- Mayfield, C. A., Foster, C. D., Smith C. T., Gan, J. and Fox, S. 2007. 2007. Opportunities, barriers, and strategies for forest bioenergy and bio-based product development in the Southern United States. *Biomass and Bioenergy.* 31: 631 - 637.
- Mckendry, P. 2002. Energy production from biomass (part 1): overview of biomass. *Bioresource Technology* 83: 37- 46.

- McLean, J. P., Jin, G., Brennan, M., Nieuwoudt, M. K. and Harris, P. J. 2014. Using NIR and ATR-FTIR spectroscopy to rapidly detect compression wood in *Pinus radiata. Canadian Journal of Forest Research*. 4: 820 – 830.
- Meder, R., Gallagher, S., Mackie, K. L. and Bohler, H. 1999. Rapid determination of the chemical composition and density of *Pinus radiata* by PLS modelling of transmission and diffuse reflectance spectra. *Holzforschung*. 53(3): 261 – 266.
- Megraw, R. A. 1985. Wood quality factors in loblolly pine-The influence of tree age, position in tree, and cultural practice on wood specific gravity, fiber length, and fibril angle. TAPPI Press, Atlanta, GA. 88pp.
- Meng, Y., Wang, S., Cai, R., Jiang, B. and Zhao, W. 2016. Discrimination and content analysis of fritillaria using near infrared spectroscopy. *Journal of Analytical Methods in Chemistry*. 2015: Article ID 752162, 8 pages.
- Mevik, B. and Cederkvist, H. R. 2005. Mean squared error of prediction (MSEP) estimates for principal component regression (PCR) and partial least squares regression (PLSR). *Journal of Chemometrics.* 18(9): 422 – 429.
- Naik, S., Goud, V. V., Rout, P. K., Jacobson, K. and Dalai, A. K. 2010. Characterization of Canadian biomass for alternative renewable biofuel. *Renewable Energy.* 35: 1624 – 1631.
- Natural Resources Canada - Canmet Energy Technology Centre (NRCan- CETC), 2008. Clean energy technologies. Bioenergy and Biomass. Available at http://canmetenergy-canmetenergie.nrcan-rncan.gc.ca/eng/bioenergy/publications/200843.html
- Natural Resources Canada, 2007. Forest bioproducts: toward a diversified Canadian forest bioeconomy.
- NIR Technological Inc. FT-NIR technical note. TN 001. Available at http://www.nirtechnologies.com/PDF/Technical%20Note%20001%20NIR%20Materials%20Identification.pdf
- Nkansah, K., Dawson-Andoh, B. and Slahor, J. 2010. Rapid characterization of biomass using near infrared spectroscopy coupled with multivariate data analysis: Part 1 yellow-poplar (*Liriodendron tulipifera* L.). *Bioresource Technology*. 101: 4570 – 4576.
- Nuopponen, M. H., Birch, G. M., Sykes, R. J., Lee, S. J. and Stewart, D. 2006. Estimation of wood density and chemical composition by means of diffuse reflectance mid-infrared

Fourier transform (DRIFT-MIR) spectroscopy. *Journal of Agriculture and Food Chem*istry. 54: 34 – 40.

- Olale, K., Yenesew, A., Jamnadass, R., Sila, A., Aynekulu, E., Kuyah, S. and Shepherd, K. 2013. Limitations to use of infrared spectroscopy for rapid determination of carbon-nitrogen and wood density for tropical species. *Advanced Analytical Chemistry.* 3: 21 – 28.
- Ono, K., Hiraide, M. and Amari, M. 2003. Determination of lignin, holocellulose and organic solvent extractives in fresh leaf, litterfall and organic material on forest floor using near-infrared reflectance spectroscopy*. Journal of Forest Research.* 8: 191 – 198.
- Owen, K., Fasina, O., Taylor, S. and Adhikari S. 2015. Thermal Decomposition Behavior of Loblolly Pine Stemwood, Bark, and Limbs/Foliage Using TGA and DSC Techniques. *Transactions of the ASABE*. 58: 509 - 518.
- Padney, K. K. 1998. A study of chemical structure of soft and hardwood and wood polymers by FTIR spectroscopy. *Journal of Applied Polymer Science*. 71(12): 1969 – 1976.
- Pandey, K. K. and Pitman, A. J. 2003. FTIR studies of the changes in wood chemistry following decay by brown-rot and white-rot fungi. *International Biodeterioration and Biodegradation*. 52(3): 151 – 160.
- Parker, P. S. 1983. *Application of Infrared, Raman and Resonance Raman Spectroscopy in Biochemistry.* Plenum Press.
- Parthasarathy, P. and Narayanan, S. K. 2013. Determination of kinetic parameters of biomass samples using thermogravimetric analysis. *Environmental Progress & Sustainable Energy.* 33 (1): 256 – 266.
- Patton-Mallory, Marcia, ed. 2008. Forest Service, U.S. Department of Agriculture woody biomass utilization strategy. Washington, DC: U.S. Department of Agriculture, Forest Service. 17 p.
- Pearson, R. G. and Ross, B. E. 1984. Growth rate and bending properties of selected loblolly pines. *Wood and Fiber Science*. 16: 37 – 47.
- Pereira, B. L. C., Carneiro, A. C. O., Carvalho, A. M. M. L., Colodette, J. L., Oliveira, A. C and Fontes, M. P. F. 2013. Influence of Chemical Composition of *Eucalyptus* Wood on Gravimetric Yield and Charcoal Properties. *BioResources*. 8: 4574 – 4592.

- Perlack, R. D., Stokes, B. J., and Erbach, D. C. 2005. Biomass as feedstock for a bioenergy and bioproducts industry: the technical feasibility of a billion-ton annual supply.
- Pettersen, R. C. 1984. The chemical composition of wood. In: Rowell, R. M. (ed), *The Chemistry of Solid Wood.* ACS Advances in Chemistry, Washington D.C. p 614.
- Poke, F. S. and Raymond, C. A. 2006. Predicting extractives, lignin and cellulose contents using near infrared spectroscopy on solid wood in Eucalyptus globules. *Journal of Wood Chemistry and Technology.* 26 (2): 187 - 199.
- Pope, J. M. 1995. Near infrared spectroscopy of wood products. In: Conners, T. E. and Banerjee, S. (eds), *Surface analysis of paper*. CRC Press LLC, Florida USA.
- Prestemon, J. P. and Abt, R. C. 2002. The Southern Timber Market to 2040. *Journal of Forestry*. 100: 16 – 22.
- Price, D. M., Hourston, D. J. and Dumont, F. 2000. Thermogravimetry of Polymers. In: Meyers, R. A. (ed.), *Encyclopedia of Analytical Chemistry*, pp. 8094 – 8105. John Wiley & Sons Ltd, Chichester.
- Ragland K. W., Aerts D. J. and Baker A. J. 1991. Properties of wood for combustion analysis. *Bioresource Technology* 37 (1991) 161-168.
- Ralevic, P., Karau, J., Smith, C. T. and Richardson, J. 2008. IEA Bioenergy Task 31 Country report: Canada.
- Rana, R., Langenfeld-Heyser, R., Finkeldey, R. and Polle, A. 2010. FTIR spectroscopy, chemical and histochemical characterisation of wood and lignin of five tropical timber wood species of the family of Dipterocarpaceae. *Wood Science and Technology.* 44: 225 – 242.
- Reich, G. 2005. Near infrared spectroscopy and imaging: basic principles and pharmaceutical applications. *Adv*anced *Drug Delivery Rev*iews. 57: 1109 – 1143.
- Rials, T. G., Kelley, S. S. and So, C. L. 2002. Use of advanced spectroscopic techniques for predicting the mechanical properties of wood composites. *Wood and Fiber Science* 34(3): 398 –407.
- Ritter, M. A. 1990. Timber Bridges: Design, Construction, Inspection, and Maintenance. Washington, DC: 944 pp.
- Rodrigues, J., Faix, O. and Pereira, H. 1998. Determination of lignin content *of Eucalyptus globulus* wood using FTIR spectroscopy. *Holzforschung*. 52: 46 – 50.

- Saldarriaga, J. F., Aguado, R., Pablos, A., Amutio, M., Olazar, M. and Bilbao, J. 2015. Fast characterization of biomass fuels by thermogravimetric analysis (TGA). *Fuel*. 140: 744 – 751.
- Sanderson, M. A., Agblevor, F., Collins, M. and Johnson, D. K. 1996. Compositional analysis of biomass feedstocks by near infrared reflectance spectroscopy. *Biomass and Bioenergy*. 11(5): 365 – 370.
- Sankaran, S. and Ehsani, R. 2013. Comparison of visible-near infrared and mid-infrared spectroscopy for classification of Huanglongbing and Citrus Canker infected leaves. *Agricultural Engineering International: CIGR Journal*. 15(3): 75 – 79.
- Scheffran, J. 2010. The global demand for biofuels: technologies, markets and policies. In: Vertes, A. A., Quereshi, N., Blaschek, H. P. and Yukawa, H. (eds), *Biomass to biofuels: strategies for global industries.* John Wiley & Sons, Ltd.
- Schimleck, L. R., Jones, P. D., Clark III, A., Daniels R. F. and Peter, G. F. 2005**.** Near infrared spectroscopy for the nondestructive estimation of clear wood properties of *Pinus taeda* L. from the southern United States. *Forest Products Journal*. 55 (12): 21 – 28.
- Schultz, R. P. 1997. Loblolly pine, the ecology and culture of loblolly pine (*Pinus taeda* L.). U.S. Dep. Agric., Forest Service, Agric. Handb -713 Washington, D.C. pp 1 – 16.
- Schultz, R. P. 1999. Loblolly – the pine for the 21st century. *New Forests* 17: 71 – 88.

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

- Schwanninger, M., Rogrigues, J. and Fackler, K. 2011. A review of band assignments in near infrared spectra of wood and wood components. *Journal of Near Infrared Spectroscopy.* 19: 287 – 308.
- Scott, A. D. and Tiarks, A. 2008. Dual-cropping loblolly pine for biomass energy and conventional wood products. *Southern Journal of Applied Forestry*. 32(1): 33 – 37.
- Senelwa, K. and R. E. Sims. 1999. Fuel characteristics of short rotation forest biomass. *Biomass and Bioenergy*. 17: 127 – 140.
- Sherman Hsu, C. P. 1997. Infrared spectroscopy. In: Settle F. A. (ed), *Handbook of Instrumental Techniques for Analytical Chemistry*. Prentice Hall PTR (ECS Professional), 1024 p.

- Shupe. T. F., Hse, C. H., Choong, E. T. and Groom, L. H. 1997. Differences in some chemical properties of innerwood and outerwood from five silviculturally different loblolly. *Wood and Fiber Science*. 29(1): 91 – 97.
- Sithole, B. B and Allen, L. 2002. The effects of extractives on system closure. Pulp & Paper Research Institute of Canada, Quebec, Canada.
- Sjostrom, E. 1981. Wood chemistry: fundamentals and applications. 1981. Academic Press, New York. 223 pp.
- Sluiter, A., Ruiz, R., Scarlata, C., Sluiter, J. and Templeton, D. 2005. Determination of Extractives in Biomass. Laboratory Analytical Procedure (LAP). Technical Report NREL/TP-510-42619.
- Sluiter, B. H., Ruiz R., Scarlata, C., Sluiter, J., Templeton, D. and Crocker D. 2012. Determination of Structural Carbohydrates and Lignin in Biomass. Laboratory Analytical Procedure (LAP). Technical Report NREL/TP-510-42618.
- Sluiter, B. H., Ruiz, R., Scarlata, C., Sluiter, J. and Templeton, D. 2008. Determination of Ash in Biomass. Laboratory Analytical Procedure (LAP). Technical Report NREL/TP-510-42622.
- Smith, W. B., Vissage, J. S., Darr, D. R. and Sheffield, R. M. 2001. Forest Resources of the United States, 1997. U.S. Dep. Agric. For. Serv., Gen. Tech. Rep. NC-219.
- Smith, W. R., Miles, P. D., Perry, C. H and Pugh, S. A. 2007. Forest resources of the United States. A technical document supporting the Forest service 2010 RPA assessment.
- Smith, W. R., Miles, P. D., Perry, C. H. and Pugh, S. A. 2009. Forest resources of the United States, 2007. A technical document supporting the Forest service 2010 RPA assessment.
- So C-L., Groom, L. H., Rials, T. G., Snell, R., Kelley, S. S. and Meglen, R. 2002. Rapid assessment of the fundamental property variation of wood. In: Outcalt, K. W. (ed.), Proceedings of the 11th biennial southern silvicultural research conference. Gen. Tech. Rep. SRS-48. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station: 176 – 180.
- So, C. and Eberhardt, T. L. 2010. Chemical and calorific characterization of longleaf pine using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*. 18: 417 – 423.

- So, C., Eberhardt, T. L., Groom, L. H. and Shupe, T. F. A. 2012. Spectroscopic study on the fuel value of softwoods in relation to chemical composition. In: *Proceedings of the 55th International Convention of Society of Wood Science and Technology*, Beijing, China.
- Spellman, F. and Bieber, R. 2011. *The Science of Renewable Energy*, CRC Press.
- Stuart, B. H. 2004. *Infrared spectroscopy: Fundamentals and applications*. John Wiley & Sons, Ltd., West Sussex, 224 p.
- Sykes, R., Li, B., Hodge, G., Goldfarb, B., Kadla, J. and Chang, H. M. 2005. Prediction of loblolly pine wood properties using transmittance near-infrared spectroscopy. *Canadian Journal of Forest Research*. 35: 2423 – 2431.
- Tampier, M., Smith, D., Bibeau, E. and Beauchemin, P. A., 2004. Identification of feedstock-to-product thread- Stage 1 Report. In: Identifying environmentally preferable uses for biomass resource. Available at http://www.cec.org/Storage/55/4777_Biomass-Stage-I-II_en.pdf
- Teleman, A. Hemicelluloses and pectins. In: Ek, M., Gellerstedt, G., Henriksson, G. (eds), *Wood Chemistry and Wood Biotechnology*. DeGruyter. Pp. 101 – 110.
- The Standing Committee of Natural Resources (2008). Third Report to Canada House of Commons. Canada's forest industry: Recognizing the challenges and opportunities. Available at http://www2.parl.gc.ca/content/hoc/Committee/392/RNNR/Reports/ RP3534643/ rnnrrp03/rnnrrp03-e.pdf
- Thumm, A. and Meder, R. 2001. Stiffness prediction of radiata pine clearwood test pieces using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy.* 9: 117 – 122.
- Toivanen, T. J. and Alen, R. 2006. Variations in the chemical composition within pine (*Pinus sylvestris*) trunks determined by diffuse reflectance infrared spectroscopy and chemometrics. *Cellulose.* 13(1): 53 – 61.
- Toscano, G., Duca, D., Rossini, G., Mengarelli, C and Pizzi, A. 2015. Identification of different woody biomass for energy purpose by means of Soft Independent Modeling of Class Analogy applied to thermogravimetric analysis. *Energy*. 83: 351 – 357.
- Tsehaye, A., Buchanan, A. H., Meder, R., Newman, R. H. and Walker, J. C. F. 1998. Microfibril angle: determining wood stiffness in radiate pine. In: Butterfield, B. G. (ed), *Microfibril Angle in Wood*. University of Cantebury, Christchurch, NZ. 410 p.

- Tsuchikawa, S., Yamato, K. and Inoue, K. 2003. Discriminant analysis of wood-based materials using near-infrared spectroscopy. *Wood Science and Technology.* 4: 275 – 280.
- Tucker, M. P., Mitri, R. K., Eddy, F. P., Nguyen, Q. A. and Gedvilas, L. M. 2013. Fourier transform infrared quantification of sugars in pretreated biomass liquors. *Applied Biochemistry and Biotechnology*. 84(86): 39 – 50.
- U.S. DOE. 2011. U.S. Billion-ton update: Biomass supply for a bioenergy and bioproducts industry. Perlack R. D. and Stokes B. J. (Leads), ORNL/TM-2011/224. Oak Ridge National Laboratory, Oak Ridge, TN, 227 p.
- U.S. DOE. 2016. 2016 Billion-Ton Report: Advancing domestic resources for a thriving bioeconomy, Volume 1: Economic availability of feedstocks. M. H. Langholtz, B. J. Stokes, and L. M. Eaton (Leads), ORNL/TM-2016/160. Oak Ridge National Laboratory, Oak Ridge, TN. 448 p.
- Varmuza, K. and Filzmoser, P. 2009. *Introduction to multivariate statistical analysis in Chemometrics*. CRC Press Taylor & Francis Group, Boca Raton.
- Via B. K. 2013. Characterization and evaluation of wood strand composite load capacity with near infrared spectroscopy. *Materials and Structures.* 46(11): 1801 – 1810.
- Via, B. K., Adhikari, S. and Taylor, S. 2013. Modeling for proximate analysis and heating value of torrefied biomass with vibration spectroscopy. *Bioresource Technology* 133: 1 – 8.
- Via, B. K., Oladiran, F. and Pan, H. 2011. Assessment of pine biomass density through mid-infrared spectroscopy and multivariate modeling. *BioResources*. 6: 807 – 822.
- Via, B. K., Shupe, T. F., Groom, L. H., Stine, M. and So, C. L. 2003. Multivariate modelling of density, strength and stiffness from near infrared spectra for mature, juvenile and pith wood of longleaf pine (*Pinus palustris*). *Journal of Near Infrared Spectroscopy*. 11(5): 365 – 378.
- Via, B. K., So, C. L., Groom, L. H., Shupe, T. F., Stine, M. and Wikaira, J. 2007. Within tree variation of lignin, extractives, and microfibril angle coupled with the theoretical and near infrared modeling of microfibril angle. *IAWA Journal*. 28(2): 189 – 209.
- Via, B. K., Zhou, C., Acquah, G. E., Jiang, W. and Eckhardt L. 2014. Near Infrared Spectroscopy Calibration for Wood Chemistry: Which chemometric Technique Is Best for Prediction and Interpretation? *Sensors*. 14: 13532 –13547.

- Via, B. K., Zhou, C., Acquah, G. E., Jiang, W. and Eckhardt L. 2014. Near Infrared Spectroscopy Calibration for Wood Chemistry: Which chemometric Technique Is Best for Prediction and Interpretation? *Sensors*. 14: 13532 –13547.
- Wang, X., Ross, R. J. and Carter, P. 2007. Acoustic evaluation of wood quality in standing tree. Part 1: Acoustic behavior in standing tree. *Wood and Fiber Science*. 39(1): 23 – 38.
- Wear, D. N. and Greis, J. G. (eds). 2002. Southern forest resource assessment. General Technical Report SRS-53, USDA Forest Service, Southern Research Station, Asheville, NC 635 pp.
- Werner, K., Pommer, L. and Brostrom, M. 2014. Thermal decomposition of hemicelluloses. *Journal of Analytical and Applied Pyrolysis.* 110: 130 – 137.
- White, E. M. 2010. Woody biomass for bioenergy and biofuels in the United States – A briefing paper.
- White, M. S., Argent, R. M. and Sarles, R. L. 1986. Effects of outside storage on the energy potential of hardwood particulate fuels: Part III. Specific gravity, ash content, and pH of water solubles*. Forest Products Journal.* 36(4): 69–73.
- White, R. H. 1987. Effect of lignin content and extractives on the higher heating value of wood. *Wood and Fiber Science.* 19(4): 446 - 452.
- Williams, P. 2004. Near infrared technology – Getting the best out of light. A short course in the practical implementation of near infrared spectroscopy for the user. PDK Grain, Nanaimo, British Columbia, Canada.
- Williams, P. T. and Besler, S. 2013. Thermogravimetric analysis of the components of biomass. In: Bridgewater, A. V. (ed), *Advances in thermochemical biomass conversion*. Science + Business Media; pp. 771 – 783.
- Winandy, J. E and Rowell, R.M. 1984. The Chemistry of solid wood. In: Advances in Chemistry. American Chemical Society, Washington DC.
- Windham, W. R., Hill, N. S. and Stuedemann, J. A. 1991. Ash in forage, esophageal, and fecal samples analyzed using near-infrared reflectance spectroscopy, *Crop Science.* 31: 1345 – 1349.
- Wold, S. 1995. Chemometrics; what do we mean with it, and what do we want from it. *Chemometrics and Intelligent Lab Systems*. 30: 109 –115.

- Workman, Jr. J. J. 2011. NIR spectroscopy calibration basics. Data analysis: calibration of NIR instruments by PLS regression. In: Burns, D. A. and Ciurczak, E. W. (eds), *Handbook of Near-Infrared Analysis*, 3rd edition. CRC Press.
- Yang, H., Yan, R., Chen, H., Zheng, C., Lee, D. H. and Liang, D. T. 2006. In-depth investigation of biomass pyrolysis based on three major components: hemicellulose, cellulose and lignin. *Energy and Fuel.* 20: 388 – 393.
- Yang, Z., Liu, Y., Pang, X. and Li, K. 2015. Preliminary investigation into the identification of wood species from different locations by near infrared spectroscopy. *BioResources.* 10: 8505 – 8517.
- Ye, X. P., Liu, L., Hayes, D. G., Womac, A. R., Hong, K. and Sokhansanj, S. 2008. Fast classification and compositional analysis of cornstover fractions using Fourier transform near-infrared techniques. *Bioresource Technology.* 99: 7323 – 7332.
- Zalesny, Jr. R. S., Cunningham, M. W., Hall, R. B., Mirck, J., Rockwood, D. L., Stanturf, J. A. Volk, T. A. 2011. Woody Biomass from Short Rotation Energy Crops. In: Zhu, J. Y., Zhang, X. and Pan, X. (eds), *Sustainable Production of Fuels, Chemicals, and Fibers from Forest Biomass.* ACS Symposium Series; American Chemical Society: Washington, DC.
- Zhou, C., Jiang, W., Cheng, Q. and Via, B. K. 2016. Multivariate calibration and model integrity for wood chemistry using Fourier transform infrared spectroscopy. *Journal of Analytical Methods in Chemistry*. 2015: Article ID 429846, 9 pages.
- Zhou, G., Taylor, G. and Polle, A. 2011. FTIR-ATR-based prediction and modelling of lignin and energy contents reveals independent intra-specific variation of these traits in bioenergy poplars. *Plant Methods*. 7(9).
- Zobel B. J. and van Buijtenen, J. P. 1989. Wood Variation: Its Causes and Control. Springer, New York, NY. 363 pp.
- Zobel, B. J. and Blair, R. 1976. Wood and pulp properties of juvenile wood and topwood of the southern pines. *Applied Polymer Symposia*. 28: 421 – 433.
- Zobel, B. J. and J. R. Sprague. 1998. Juvenile wood in forest trees. Springer-Verlag, Berlin. 300 pp.

Appendices

A: ANOVA results of NIR-predicted values and laboratory-measured values for density, MOR and MOE

Density

A1

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	0.028825	1	0.028825	3.95315	0.093943
Within Groups	0.04375	6	0.007292		
Total	0.072575	7			

A2

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	0.047204	1	0.047204	5.204026	0.062694
Within Groups	0.054424	6	0.009071		
Total	0.101629	7			

A5

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	0.044113	1	0.044113	4.585885	0.075996
Within Groups	0.057716	6	0.009619		
Total	0.10183	7			

A9

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	0.016409	1	0.016409	1.474253	0.270293
Within Groups	0.066781	6	0.01113		
Total	0.083189	7			

A10

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	0.01497	1	0.01497	0.897913	0.379927
Within Groups	0.10003	6	0.016672		
Total	0.115	7			

A15

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	0.004369	1	0.004369	0.4264	0.537963
Within Groups	0.061481	6	0.010247		
Total	0.06585	7			

A21

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	0.001499	1	0.001499	0.124985	0.73578
Within Groups	0.071937	6	0.011989		
Total	0.073435	7			

A26

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	0.000429	1	0.000429	0.043704	0.841323
Within Groups	0.05891	6	0.009818		
Total	0.059339	7			

A33

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	0.029819	1	0.029819	1.74127	0.235092
Within Groups	0.10275	6	0.017125		
Total	0.132569	7			

A34

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	0.000121	1	0.000121	0.014088	0.909394
Within Groups	0.051394	6	0.008566		
Total	0.051514	7			

A37

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	0.010682	1	0.010682	1.032314	0.348822
Within Groups	0.062086	6	0.010348		
Total	0.072768	7			

F3

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	0.000364	1	0.000364	0.048237	0.833441
Within Groups	0.045255	6	0.007542		
Total	0.045618	7			

F17

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	0.016059	1	0.016059	0.921044	0.374262
Within Groups	0.104614	6	0.017436		
Total	0.120673	7			

F18

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	0.001141	1	0.001141	0.062523	0.810893
Within Groups	0.109494	6	0.018249		
Total	0.110635	7			

F23

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	0.001753	1	0.001753	0.149634	0.712233
Within Groups	0.070276	6	0.011713		
Total	0.072029	7			

Lumber

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	0.003614	1	0.003614	4.606622	0.04573
Within Groups	0.01412	18	0.000784		
Total	0.017734	19			

Modulus of rupture

A1

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	445.4366	1	445.4366	0.402023	0.54944
Within Groups	6647.929	6	1107.988		
Total	7093.366	7			

A2

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	1139.029	1	1139.029	1.446991	0.27432
Within Groups	4723.023	6	787.1706		
Total	5862.052	7			

A5

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	1229.088	1	1229.088	1.154562	0.323896
Within Groups	6387.297	6	1064.55		
Total	7616.386	7			

A9

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	166.5404	1	166.5404	0.158628	0.704209
Within Groups	6299.29	6	1049.882		
Total	6465.83	7			

A10

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	166.586	1	166.586	0.082914	0.783074
Within Groups	12054.83	6	2009.139		
Total	12221.42	7			

A15

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	56.04758	1	56.04758	0.085869	0.779367
Within Groups	3916.273	6	652.7122		
Total	3972.321	7			

A21

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	3.499335	1	3.499335	0.00214	0.964605
Within Groups	9811.897	6	1635.316		
Total	9815.397	7			

A26

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	3.608641	1	3.608641	0.003042	0.957805
Within Groups	7117.345	6	1186.224		
Total	7120.953	7			

A33

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	286.3225	1	286.3225	0.263673	0.625962
Within Groups	6515.406	6	1085.901		
Total	6801.729	7			

A34

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	36.88546	1	36.88546	0.048685	0.832684
Within Groups	4545.802	6	757.6336		
Total	4582.687	7			

A37

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	555.7945	1	555.7945	0.577899	0.475937
Within Groups	5770.506	6	961.751		
Total	6326.301	7			

F3

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	194.8338	1	194.8338	0.229359	0.648953
Within Groups	5096.825	6	849.4709		
Total	5291.659	7			

F17

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	15.32088	1	15.32088	0.009217	0.926643
Within Groups	9973.401	6	1662.234		
Total	9988.722	7			

F18

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	292.6869	1	292.6869	0.143433	0.717931
Within Groups	12243.52	6	2040.587		
Total	12536.21	7			

F23

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	351.7348	1	351.7348	0.311472	0.596972
Within Groups	6775.594	6	1129.266		
Total	7127.328	7			

Lumber

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	3876137	1	3876137	1.517728	0.233815
Within Groups	45970333	18	2553907		
Total	49846470	19			

Modulus of elasticity

A1

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	2054060	1	2054060	0.131798	0.729022
Within Groups	93509762	6	15584960		
Total	95563822	7			

A2

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	2480433	1	2480433	0.320575	0.591793
Within Groups	46424660	6	7737443		
Total	48905092	7			

A5

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	2358249	1	2358249	0.294351	0.606994
Within Groups	48070106	6	8011684		
Total	50428355	7			

A9

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	6715196	1	6715196	0.687279	0.43882
Within Groups	58624199	6	9770700		
Total	65339395	7			

A10

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	1717082	1	1717082	0.076639	0.791195
Within Groups	1.34E+08	6	22404863		
Total	1.36E+08	7			

A15

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	13415.22	1	13415.22	0.001231	0.973147
Within Groups	65373203	6	10895534		
Total	65386618	7			

A21

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	679020.3	1	679020.3	0.030532	0.867036
Within Groups	1.33E+08	6	22239668		
Total	1.34E+08	7			

A26

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	12403.13	1	12403.13	0.000707	0.979644
Within Groups	1.05E+08	6	17533856		
Total	1.05E+08	7			

A33

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	347944.8	1	347944.8	0.033852	0.860083
Within Groups	61670343	6	10278390		
Total	62018288	7			

A34

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	131430.6	1	131430.6	0.009931	0.923866
Within Groups	79409397	6	13234900		
Total	79540828	7			

A37

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	26576.65	1	26576.65	0.003408	0.955343
Within Groups	46790173	6	7798362		
Total	46816749	7			

F3

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	292344.8	1	292344.8	0.019875	0.8925
Within Groups	88254050	6	14709008		
Total	88546395	7			

F17

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	270480.1	1	270480.1	0.011315	0.918755
Within Groups	1.43E+08	6	23905020		
Total	1.44E+08	7			

F18

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	509444.2	1	509444.2	0.028292	0.871951
Within Groups	1.08E+08	6	18006772		
Total	1.09E+08	7			

F23

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	10333058	1	10333058	0.664763	0.446033
Within Groups	93263825	6	15543971		
Total	1.04E+08	7			

Lumber

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	435.2445	1	435.2445	2.885759	0.106584
Within Groups	2714.849	18	150.8249		
Total	3150.094	19			