## Graduate Theses, Dissertations, and Problem Reports

2009

# Rapid characterization of biomass: The use of near infrared and fluorescence spectroscopy as process analytical technology (PAT) method

Kofi Nkansah
*West Virginia University*

Follow this and additional works at: https://researchrepository.wvu.edu/etd

# RAPID CHARACTERIZATION OF BIOMASS: THE USE OF NEAR INFRARED AND FLUORESCENCE SPECTROSCOPY AS PROCESS ANALYTICAL TECHNOLOGY (PAT) METHOD.

**Kofi Nkansah**

**Thesis submitted to the College of Agriculture, Forestry and Consumer Sciences at West Virginia University In partial fulfillment of the requirements for the degree of**

**Master of Science In Forestry**

**Ben Dawson-Andoh, Ph.D., Chair**
**James Armstrong, Ph.D.**
**John Renton, Ph.D.**
**Ben Spong, Ph.D.**

**Division of Forestry**

**Department of Wood Science and Technology**

**Morgantown, West Virginia**
**2009**

**Keywords: Near infrared (NIR), Fluorescence, Process analytical technology (PAT), Multivariate data analysis (MVDA)**

**ABSTRACT**


**RAPID CHARACTERIZATION OF BIOMASS: THE USE OF NEAR INFRARED AND FLUORESCENCE SPECTROSCOPY AS PROCESS ANALYTICAL TECHNOLOGY (PAT) METHOD.**

**Kofi Nkansah**


The heterogeneous property of biomass (wood) affects its potential of being converted into any form of fuel in different ways (both positive and negative effects). Therefore in other to efficiently utilized biomass as a raw material for conversion into any form of clean alternative fuel to displace some of the fossil fuel we consume in the United State on a commercial scale basis, a quick, robust, non destructive on/in/at-line method of characterizing the physical and chemical properties of biomass that are relevant to the bio-refinery industry is imperative.


This study discusses the potential of using near infrared spectroscopy (NIRS) and fluorescence spectroscopy (FS) coupled with multivariate data analysis (MVDA) as a robust and rapid process analytical technology (PAT) to characterize the physical and chemical properties of two potential biomass feedstock (yellow-poplar and northern red oak) in its solid state. This study is aimed at rapidly detecting the properties of potential biomass feedstock to be used in the bio-refinery online before any conversion process is begun. This will reduce cost of manufacturing bio-fuels, provide real time results of biomass characteristics reduce waste and produce a much consistent product. The potential utilization of fluorescence spectrometer which is much cheaper, rapid and sensitive spectrometer with equal model performance as the NIR spectrometer models will reduce the cost of PAT even further.


Generally, the results of this study showed that both NIR and FS can be used as rapid PAT method to characterize the physical and chemical properties of northern red oak and yellow-poplar with moderate to high prediction performance. The NIR prediction models generally exhibited slightly higher prediction model performance as compared to similar models of the same response variable developed with the fluorescence spectra data.

# DEDICATION

To my mum whose love is endless; thank you for your constant prayer and encouragement

To the rest of my family for their encouragement, prayer and support

&

To everyone who has contributed to my life positively towards my education especially my father who realized the essence of education and gave me one.

# ACKNOWLEDGEMENT

I can't compromise but first give thanks to the almighty God for his love, mercies and strength on my life. I will like to express my profound gratitude to my graduate study advisor, Dr. Ben Dawson-Andoh for his encouragement, guidance, patience and most importantly believing in me. I will also like to thank my graduate committee members for their time, valuable advice and comments on this thesis.

I am particularly grateful to Jeffrey Slahor of the Appalachian Hardwood center for his support and advice towards this study.

Finally to all my friends and colleagues at The West Virginia University, I cannot but single our Mr. Emmanuel Oluwatosin Adedipe for his support and advice during the course of my study. Thank you OLU!

Thank you all! Without you this thesis was far from reach!

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF SYMBOLS AND ABBREVIATIONS

| | |
|---|---|
| PAT | process analytical technologies |
| PAC | process analytical chemistry |
| NIR | near infrared |
| FT-NIR | Fourier transformed near infrared |
| MVDA | multivariate data analysis |
| PC | principal component |
| PCA | principal component analysis |
| PLS | partial least square regression |
| SIMCA | soft independent modeling of class analogies |
| CGC | Canadian Grain Commission |
| MSC | multiplicative scatter correction |
| PARAFAC | parallel factor analysis |
| R | coefficient of correlation |
| $R^2$ | coefficient of determination |
| SEP | standard error of prediction |
| SEC | standard error of calibration |
| RMSEP | root mean square error of prediction |
| RMSEC | root mean square error of calibration |
| SD | standard deviation |
| MAX | maximum |
| MIN | minimum |
| RDP | ratio of standard error of prediction to standard deviation |

| | |
|---|---|
| < | value greater than |
| < | Value less than |
| Φ | Quantum yield |

# LIST OF PAPERS TO BE PUBLISHED

I.      Nkansah, K.; Dawson-Andoh, B.; Slahor, J. Rapid characterization of biomass using near infrared spectroscopy coupled with multivariate data analysis: Part1. Yellow-poplar (*Liriodendron tulipifera L).* 2009: Unpublished.

II.     Nkansah, K.; Dawson-Andoh, B. Rapid characterization of biomass using near infrared spectroscopy coupled with multivariate data analysis: Part2. Northern red oak (*Quercus rubra*). 2009: Unpublished.

III.    Nkansah, K.; Dawson-Andoh, B. Rapid characterization of biomass using fluorescence spectroscopy coupled with multivariate data analysis: Part1. Yellow-poplar (*Liriodendron tulipifera* L). 2009: Unpublished

IV.    Nkansah, K.; Dawson-Andoh, B. Rapid characterization of biomass using fluorescence spectroscopy coupled with multivariate data analysis: Part2. Northern red oak (*Quercus rubra*). 2009: Unpublished

# CHAPTER ONE: Structure of thesis, introduction, objectives and literature review

## Structure of thesis

This thesis is a combination of four unpublished papers that emanated from one broad objective. This thesis will constitute six chapters. Chapter one presents a general introduction to the study and literature review on relevant concepts such as multivariate data analysis, process analytical technology, near infrared spectroscopy (NIRS) and fluorescence spectroscopy (FS).

Chapter two will constitute paper I, which investigate the use of near infrared (NIR) as a process analytical technology (PAT) tool to predict the physical and chemical properties of yellow-poplar (*Liriodendron tulipifera*). Chemical properties that were studied include ash content, extractive content, acid insoluble lignin, total lignin and holocellulose. Two types of bulk densities (aerated and tapped bulk density) were also measured and predicted.

Paper II will be presented in chapter three, which is a follow up study on the first paper. This study was necessitated based on the concept that biomass is heterogeneous and the prediction performance of yellow-poplar calibration models may not be taken as a universal prediction performance for all biomass types. This chapter evaluated the possibility of predicting the physical and chemical properties of northern red oak using NIR spectroscopy coupled with multivariate data analysis.

Chapter four of this thesis presents paper III which investigates the potential of using FS as a rapid PAT method for predicting the physical and chemical properties of yellow-poplar relevant to the bio-fuel industries that utilize biomass as a feedstock. This study also compares the performance of similar near infrared and fluorescence spectra-based models of each property of yellow-poplar which were considered in this study and predicted from the same population size.

Paper IV, the final paper of this study is presented in chapter five of this thesis. This paper is a follow up on paper III. This chapter presents a study on the potential of using FS as a PAT tool to predict the chemical and physical properties of yellow-poplar. This study also compared fluorescence spectra-based models to near infrared spectra-based models of similar chemical or physical property of northern red oak to assess their performance. This comparison

was reasonable because both models were developed with the sample population. Chapter six will present a general conclusion and recommendations of the whole study.

## Introduction

Recent increase in crude oil prices coupled with environmental concerns over the use of fossil fuels has lead to the search for alternative clean fuel. In this search for alternative clean fuels to complement or displace some of the fossil fuels that our economy depends on, transportation bio-fuels and bio-products has been on the fore front. The use of any biomass as feed stock for bio-fuel is affected by its physical and chemical properties. The heterogeneous nature of biomass requires it properties to be known in real time before it can be used efficiently as a feedstock for any form of bio-energy conversion process. Therefore there is the need to pursue very rapid, sensitive, non invasive, robust and simple technology that can be used online during biomass conversion to predict the physical and chemical properties. This will form part of process monitoring in the bio-fuel product processing stream.

Process analytical technology (PAT) has been successfully used for such process control activities in the petrochemical and pharmaceutical industry for more than two decades. In PAT, sensitive and robust analyzers are coupled with an appropriate multivariate data analysis (MVDA) to monitor the process stream of product manufacturing from the beginning to the end of the process.[1] In the forest product industry, near infrared spectroscopy (NIRS) has been successfully been used as a PAT method to predict and monitor a variety of properties of wood that affects the appropriate products being manufactured. These include physical and chemical properties of wood.[2,3] However, NIRS cannot be a catch-all PAT tool for all lignocellulosic properties, other efficient spectroscopic tools such as fluorescence which is 100-1000 times more sensitive need to be pursued. The efficient conversion of biomass into clean transportation fuel requires the development of appropriate cheap, robust, rapid and non invasive PAT method to characterize the physical and chemical properties of the feedstock relevant to the biochemical and chemo-thermal processing both online and offline. Fluorescence spectroscopy (FS) and NIRS meet the requirements to be used as a PAT method in the bio-fuel and forest product industry in general.

FS, a very sensitive spectroscopic tool is not widely used as a PAT analyzer in the biomass and forest products industry. Wood fluoresces by virtue of the presence of chromophores that gives wood its color. Even though Billa *et al*. (2000) demonstrated the potential of using FS coupled with MVDA to predict the ash content and monomeric lignin of eucalyptus wood, most of the use of FS in the forest product industry has predominantly been

limited to identification of wood.[5,6] Fluorescence spectrometers are much portable, sensitive and acquisition of spectrum is rapid as compared to other spectroscopic tools used in the forest products industry. FS if pursued as a possible PAT method in the wood industry will cut cost in process control activities and also time of analysis.

## Objective of study

The primary goal of this study was to investigate the use of NIR and fluorescence to characterize some physical and chemical properties of two potential bio-refinery raw material feed stocks: yellow-poplar (*Liriodendron tulipifera* L.) and northern red oak (*Quercus rubra*). Specific objectives of this study include:

1. Investigate the use of NIR and FS coupled with MVDA to predict important physical and chemical properties of northern red oak and yellow-poplar. These properties include ash content, extractives content, total lignin, acid insoluble lignin, holocellulose and bulk density.

2. Compare the performance of NIR and fluorescence based prediction models of each physical and chemical property of northern red oak and yellow-poplar.

# Literature review

## Process analytical chemistry (PAC) / Process analytical technology

Process Analytical Chemistry (PAC) a technique that combines analytical chemistry, process engineering, process chemistry and multivariate data analysis (MVDA) has been embraced by a lot of industries for decades. Process Analytics (PA) is broadly defines as the chemical and physical analysis of a material in the process stream through the use of an in-line or on-line analyzer. The principal focus of PAC is to gain a deep understanding of a process stream in other to achieve a more consistent end product, improve manufacturing efficiency, reduce waste, optimize the use of resources, enhance safety and reduce production cost that can be gathered from the mentioned advantages of PAC.

When the principal focus of PAC is directed towards process understanding, the technique is referred to as Process Analytical Technology. The past decade has seen approximately 5% annual increase in the use of PA instruments and this growth is expected to continue.[1] Early PAT utilized univariate tools such as oxygen sensors, pH meters and flow meters which are currently also used. Other tools that are been utilized along these univariate tools are on-line chromatographs and spectroscopic tool.[1] There has been extensive research in the potential use of PA instrument in the biomass based products industries. NIR is one of the most predominant PA tools that have been used over the years. NIR has been used as a PAT method to predict the physical and chemical properties of different types of biomass based products.[2,3] The potential used of other tools such as NMR and FS have also been investigated.[4,7,8]

## Near infrared spectroscopy (NIRS)

The use of NIR as a PAC technique has been applied to various process control variable measurements from a simple variable (moisture content) to very complex variables (full characterization of hydrocarbon composition and properties in refinery process stream).[1] NIRS is known to be simple, non-destructive and rapid analytical technique that enables multi constituent analysis of nearly any matrix with a precision and accuracy comparably similar to reference methods.[9] William Herschel, an astronomer, is largely credited with the discovery of NIR when he studied the distribution of heat in sunlight in 1800.[10] The first potential use of NIR in the

agricultural and food industry was reported by Karl Norris *et al*. in 1950. They exhibited the potential used of NIRS as a rapid analytical tool for studying the properties of grain.[11] The instrumentation of NIRS has lately been modified for rapid, non-destructive analysis of a variety of matrices; solids and liquids across agricultural and food industry.[9] The first commercial application of NIRS as a PAT tool in the agricultural products industry is credited to the Canadian Grain Commission (CGC). The CGC successfully started using NIRS to monitor the protein content of wheat grain from 1974. NIR coupled with chemometric tools has been successfully used for quantitative and qualitative analysis of organic compounds in an array of applications. Online and inline analysis of materials with NIRS permits real-time measurements of a process and minimize deviations from standard process specification.[1]

**Principles of NIRS**

NIRS is defined as the measurement of wavelength and intensity of absorption of the near infrared light by a material. The spectra region approximately between 750-2500nm is characterized as the NIR region of the electromagnetic spectrum.[12] (Figure 1)
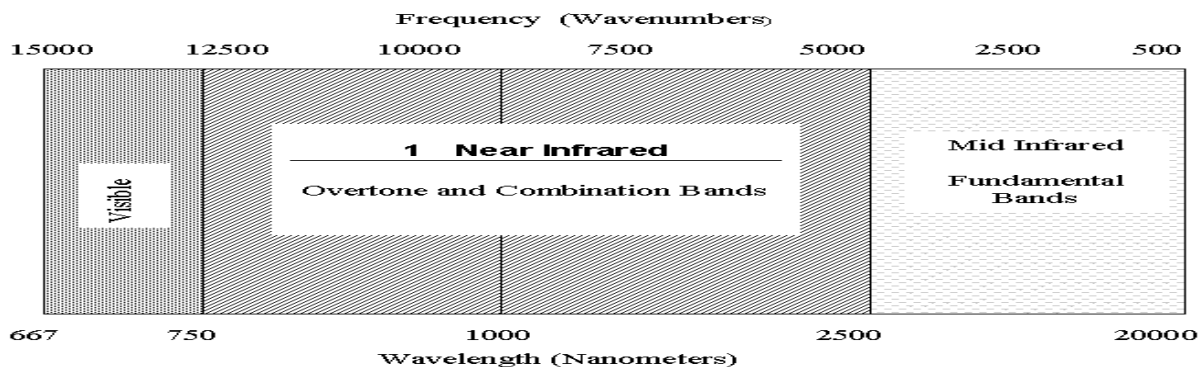


**Figure 1: The infra red region of the electromagnetic spectrum (667-20000nm)**

 **Source:** http://www.nirtech.net/technology.htm

6

The principle behind NIRS is the concept of vibration spectroscopy. NIRS is used to characterize the vibrational properties of a matrix or material.[13] The concept of vibrational spectroscopy is based on the principle that atom-to-atom bonds within molecules vibrate with different frequencies that can be quantified by laws of physics.[14] Most molecules at room temperature vibrates at their least energetic state allowed by quantum mechanics (rest state). NIR light has enough energy to excite molecular vibration of a material to higher energy levels from its rest state. When a material is radiated with EM radiation of NIR, the NIR light transfers some amount of energy unto the materials molecules and this causes transition from ground state of the molecules to higher energy levels. When this energy transfer matches the exact difference between the two energy levels (ground state-to-excited state) absorption of NIR radiation (energy) by the material occurs. Two major types of bond vibrational excitation occur during NIR radiation on a material, there are stretching (higher energy) and bending (lower energy) vibrations.[13-15]

Some groups of atoms characteristically absorb NIR at different wavelengths.[16] Spectrometers can be used to detect and measure these frequencies or energies being absorbed. The near infra red spectra region (700-2500nm) absorption bands are characterized by overtones and combination of fundamental stretching bands that are known to occur in the mid-IR region of the EM-spectrum. Certain bonds of fundamental group of atoms including C-H, O-H and N-H are known to be involved in NIR absorption of a material. This means most organic molecules and water molecules which possess theses functional groups does absorb radiations in the NIR spectra region. This provides unique advantage of using NIRs as a qualitative and quantitative tool in agricultural, food and organic materials industries since NIR absorption bands in organic materials will be due to these fundamental chemical bonds; C-H (Oils, fats and hydrocarbons), N-H (proteins) and O-H (water and alcohol). The overtones bands of the NIR may be absorbed by other chemical bonds of a material but these absorptions are generally weak to be potentially useful in any analysis of a heterogeneous material such as agricultural and pharmaceutical product.[2,14,16] Fundamentals absorption are known to be 10-100 times much stronger than overtones and combination bands. NIR radiation must cause a change in the dipole moment of a molecule before absorption occurs.[13] Since the NIR spectrum of a material is characterized by overlapping overtones and fundamental absorptions, calibration models does rely heavily on statistics (MVDA).[16]

## Near infrared spectra acquisition modes

T        wo modes of NIR spectra acquisition (transmission mode or reflection mode) are predominantly employed in the operation of NIR spectrometers.[17]

## Transmission spectroscopy

Transmission spectroscopy is known to be the most common form of NIRS measurement modes. It occurs in the region between 1200 to 1850nm which encompass the first and second overtones of the mid infra red region fundamental stretching bands. Transmission spectroscopy is generally based on the direct penetration of a material by an NIR radiation without any deflections and the energy absorbed by the chemical component of the material recorded by the spectrometer. Transmission spectroscopy measurement (1/log T) is usually used to analyze materials that are transparent such as liquids and film and other biological materials.[16] (Figure 2)
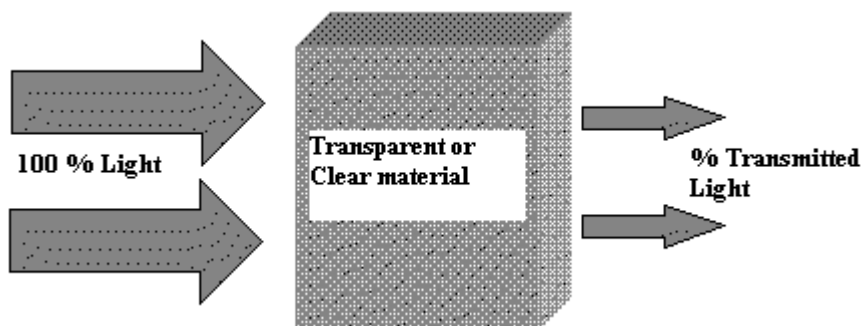
Figure 2: Schematics of transmission spectroscopy

## Diffuse reflectance

Diffuse reflectance spectroscopy measurements (1/R) is generally the most predominant reflection mode of spectra acquisition employed when analyzing opaque or light scattering matrices inkling solids, sullies and suspensions.[16,17] The basic concept of diffuse reflectance is depicted in Figure 3.  In diffuse reflectance, NIR radiation illuminates a material at zero degrees angle. The light interacts with the material and re-radiate the diffused energy that contains vital compositional and physical characteristics of the material back into the plane of illumination at an angle of 45 degrees. The deviation in angle gives an advantage of a reduction in specular reflection during spectra collection. Particle size and shape of the particles being tested affect

diffuse reflectance; therefore a potential sampling error ascribed to particle size and shape needs to be considered during sampling.[16]
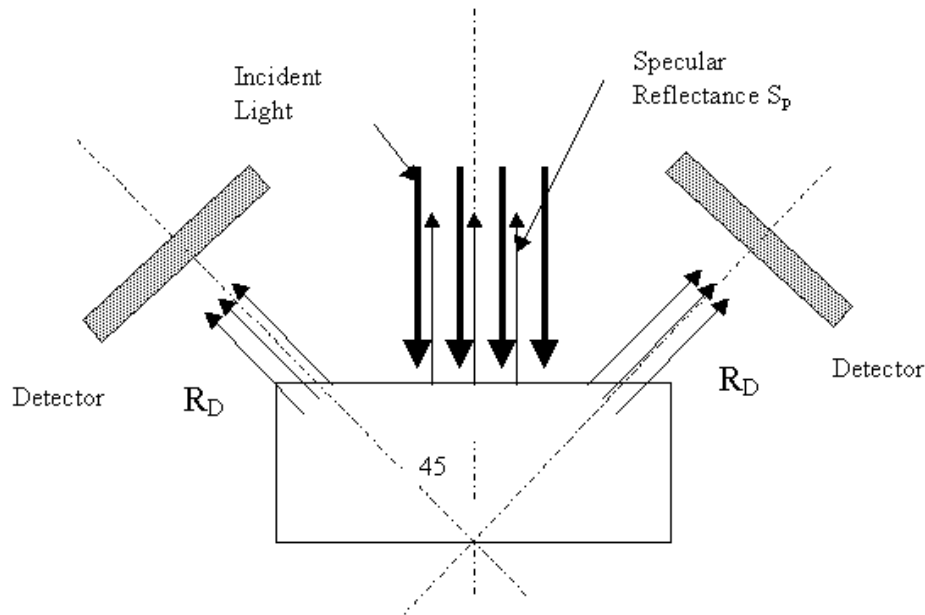


**Figure 3: Schematic of Reflectance spectroscopy**

**Source:** http://www.nirtech.net/technology.htm

**NIR Spectra analysis**

**Data (spectra) pre-processing**

The overtones and combination of fundamental vibrations that are exhibited by the three fundamental absorbers; C-H, O-H, and N-H produce similar, broad and overlapping bands that cannot be clearly seen.[16] This characteristic of the NIRS makes it rely greatly on statistics to decipher and extract information in the spectra produced. It is therefore common to apply some form of mathematical pre-treatment to spectra data during analysis. Three forms of spectra pre-processing are commonly used in NIR spectroscopic analysis of reflectance data. These include smoothing, derivation, and multiplicative scatter correction.[16]

Smoothing essentially used on spectra data where system noise is highest. Smoothing is used to improve multivariate calibration by reducing the spectra noise in the data. The loss of

spectra resolution can also occur when too much smoothing is employed therefore theoretical background of smoothing methods need to be known before being applied.[16]

Spectra derivation practically correct baseline shift and some variation that are encountered in using diffused reflectance spectroscopic measurements caused by particle characteristics. All spectra being analyzed are brought to a common base line before analysis is performed. The first derivative is used to offset baseline feature whiles the second derivate is used for linear base line feature (removing overlaps) that make absorption bands appear shaper.[16,18] Third and fourth derivative are possible and even produce much sharper bands but higher derivative are not encouraged since spectra noise increases as derivative increases from first to fourth.[16]

Multiplicative scatter correction (MSC) corrects the additive and multiplicative effects of spectra data. This method is used to correct the difference that exist between individual spectra data for all samples at a wavelength point and the average spectra data for all samples at that same wavelength.[16] MSC allows the variance between physical and chemical properties of a material to be more profound.[19]


**Potential use of NIR in the biomass products industry**

Biomass, an organic material has fundamental groups of chemical bonds (C-H, O-H and N-H) that form its primary constituents. These chemical bonds are known to have specific absorption bands in the NIR spectrum of the material. When NIR spectra are coupled with the appropriate MVDA inherent chemical and physical properties of the material can be extracted. NIRS has many advantages that make it the spectroscopic tool of choice in characterizing biomass. These advantages include; ease of operation and analysis, minimal sample preparation, non invasiveness and rapid spectra acquisition. NIR as a PAT tool has been used in the chemical and pharmaceutical industries for more than two decades.[1,9] Tremendous research on forestry products industry have successfully demonstrated the use of NIRS to characterize some physical and chemical properties of some wood species.[20-22] The mechanical properties of solid wood has also been successfully predicted with NIRS.[17,23-25] The paper industry also has extensively studied NIRS and has used it to characterize the chemical properties of pulp.[26-28]

## Fluorescence spectroscopy

### Fluorescence spectroscopy (FS)

Light can be absorbed and re-emitted by matter, this phenomenon is called luminescence. There are two types of luminescence where absorption of photons occurs: fluorescence and phosphorescence. The absorbing and re-emitting light in fluorescence occurs on an average timescale of about $10^{-5}$ to $10^{-8}$ seconds while phosphorescence processes are much slower, taking about $10^{-4}$ to several second to occur.[29,30]

Fluorescence spectroscopy (FS) is a widely used method within biochemistry and molecular biophysics. This is likely due to its extraordinarily high sensitivity in detecting changes in the structural properties as well as dynamic properties of biomolecules and biomolecular complexes.[31] In FS the radiation of electrons in certain molecules by a suitable light energy source usually ultraviolet light causes the molecule to emit light of a lower energy which is typically but not necessarily visible light. Sir G. G. Stokes in 1852, first documented the comprehensive mechanism of absorption and emission process that he named as flourecence.[32] The phenomenon of fluorescence could be defined as the spontaneous emission of light from an electronically excited state to different vibrational levels of the electronic ground state. The stages in this process involve electronic transfer after molecules are excited by photons.[33] Flourophores are characterized by important properties like the position of emission wavelength, intensity and life time. These properties are inherent in any flourophore and are modified by the environment.[34] The efficiency of fluorescence is expressed as quantum yield ($\Phi$) which is a measure of the ratio of emitted light to absorbed light. When a quantum yield is 0.10, only 10% of photons are absorbed by the system (material) resulting in fluorescence.[7] Quantum yield and wavelength of fluorescence spectra is affected by the structure of the material and the environment.[7]

### Principles of fluorescence spectroscopy

FS is primarily associated with excitation of molecules from the ground state through the absorption of photons. Materials at room temperature have its molecules in the ground state which is the lowest vibrational state of the molecule. In FS, the molecules absorb photons and are elevated to an excited state which results in the molecule reaching any of the vibrational sub-

levels with each electronic state. Collisions with other molecules cause the excited molecule to lose vibrational energy until it reaches the lowest vibrational state of the excited electronic state. The molecule then drops down to one of the various vibrational levels of the ground electronic state again, emitting a photon in the process.[29] Molecules may drop down into any of several vibrational levels in the ground state. The emitted photons will have different energies (emission), and thus frequencies. Therefore, by analyzing the different frequencies of light emitted in FS, along with their relative intensities, the structure of the different vibrational levels can be determined.[31] In a typical study, the different frequencies of fluorescent light emitted by a sample are measured, holding the excitation light at a constant wavelength. This is called an emission spectrum. An excitation spectrum is measured by fixing the emission wavelength and measuring the excitation wavelenghth that emerges from the monochromator.[34]

The Jabłoński diagram (Figure 4) summarizes the different relaxation pathways of an electronically excited chromphore. Light absorption usually starts from the first singlet state (S0) and its lowest vibrational level towards a vibrational level in S1. This is indicated by a vertical line in the diagram. In the S1 state the molecule rapidly relaxes to the lowest vibrational level (S0). There are several possible pathways of relaxation from the excited state and most of them are non-radiative. The most common pathways are internal conversion (IC) and fluorescence.[34]
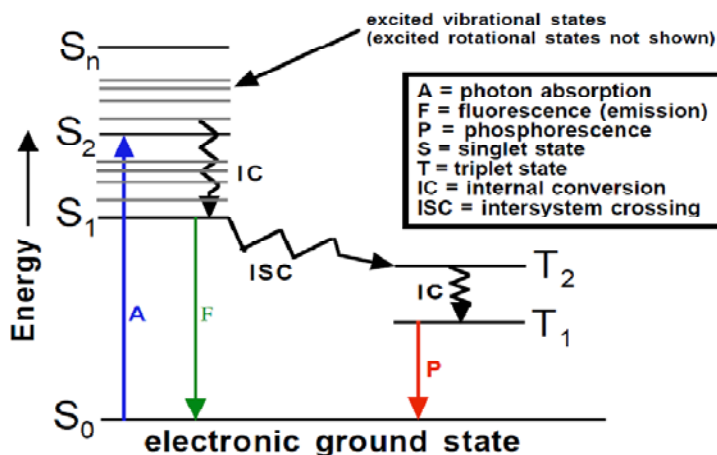


**Figure 4: The Jablonski's diagram**

**Source:** http://www.shsu.edu/~chm_tgc/chemilumdir/JABLONSKI.html

12

**Potential use of fluorescence spectroscopy in biomass products industry**

The presence of chromophores in wood (biomass) which gives its color is responsible for the fluorescence of wood. Fluorescence spectra coupled with the appropriate MVDA may be used to predict some important properties of wood. The use of FS spectroscopy in the forest products industry has been limited. Fluorescence has been successfully coupled with MVDA to characterize some chemical properties of wood.[4] The concept of different species of wood having different emission spectra (color) has also been demonstrated.[5,6,35] The paper industry has studied fluorescence extensively and has used it to characterize cellulose and lignin content of pulp.[7]

## Multivariate data analysis

In MVDA, mathematics and statistics are employed to extract relevant information from large data matrices with more than one variable. This approach is driven by data "mining" and statistics to get the desired results. Vital physical and chemical information of a material can be extracted from the broad spectrum of the matrix through MVDA. In contrast to traditional statistics, distributions assumptions are not met; the number of variables may be higher than the number of object and the variables are highly correlated. Hidden data structure of a material is fully investigated by the use of MVDA through the utilization of all data points in the data analysis.[10] Traditional statistics will not analyze multivariate data where the number of objects is not at least five times as large as the number of variables.[36] MVDA employs the correlation among the variables instead of looking at only one or few variables. The original variables are linearly combined to create new linear orthogonal variables.

Properties of materials, technological processes, and chemical measurement are multivariate in nature, i.e. many variables describe the observed phenomenon, i.e. properties, etc. Multivariate analysis unlike univariate analysis considers many variables simultaneously. In chemical analysis, traditional chemical approaches are univariate – absorbance of a solution is measured at only one wavelength.[37] In this approach, some important information about the solution is lost. In contrast, multivariate approach will record a spectrum of the solution over a wavelength region, i.e. – absorbance at several wavelengths are recorded. Since MVDA considers many variables simultaneously, it is able to obtain newer and higher quality of informationa from data.[36] The NIR and fluorescence spectra of most materials, including wood, contain valuable underlying chemical and physical information ("finger print") related to the

described material. However, without appropriate MVDA, the chemical, physical and biological information obtained from the spectra is limited.[16] Thus MVDA which analyzes data consisting of multiple samples (observation) and variables by decomposing and reducing these multiple data points (data dimensionality) into a few relevant latent variables that can explain the relationship between the observations and variables in the data and also capture them in new latent variables.[38,39] Multivariate data contains much more information about the phenomenon being studies than univariate data. There are many MVDA methods and the choice of a method in any analysis depends on the objectives of the study. These methods are categorized under three mains groups[39]:

(i)      Data description (explorative data structure modeling)

(ii)     Discrimination and classification

(iii)    Regression and prediction. (Partial least square and principal component regression)

This study employed two of these MVDA methods; they include data description method – principal component analysis (PCA) and principal least square regression method (PLS).

**Data description (Overview)**

The first steps in data analysis are descriptive statistics (mean, standard deviation, variance, etc) and PCA. PCA reduces the dimensionality of data by a linear combination of original data to generate new latent variables which are orthogonal and uncorrelated to each other. It distills the information in the original variables into a lower number of variables – principal components (latent variables).[40] The first principal component (PC) lies along the direction of maximum variance in the data set, whiles the second PC lies in the direction of the next greatest variance in the data matrice.[41,39] PCA, basically decomposes an X-data matrix into structure and noise parts. PCA thus minimizes errors in X-data matrix measurements by eliminating the noise part of the data matrix.[42] PCA is known to form the basis of MVDA by extracting and displaying systematic variations within a data matrix.[40,42-45] This method is used to characterize data and display its intrinsic structures visually by graph plots and summaries. Thus, it provides an overview of data and displays the relationship between variables in a data. PCA extracts and determines which variable is vital to variations in a data and also gives an overview of covariance variables that has the same effects on the x-matrix data.[38]

**Multivariate calibration**

      In any multivariate calibration process, the determination of X and Y matrices are essential. PLS and PCR relates the independent variables (X-matrix) to the response variables ($Y_i$-matrices). The model(s) developed is then used to predict new response variables from new independent population with similar sampling properties. The ability PLS and PCR to deal with noise, collinearity and interference have made them the two methods widely used for multivariate regression calibration modeling methods.

**Principal component regression (PCR)**

      The principal component regression (PCR) method is a two step procedure which involves PCA methods and multiple linear regression method (MLR). First a PCA of an X-matrix data is done and the transformed x-matrix data is later plugged into an MLR model.[39] The advantage of PCR over an MLR is that, PCR exhibits score plots, loadings and variance and can be interpreted with ease.[39]

**Partial least square regression (PLS)**

      PLS is simply considered to be the most used method for relating two data matrices, X and Y, to each other by a linear multivariate model. PLS considered to be useful in analyzing a data with many variable, noise, co-linearity, and incomplete variables in both X and Y matrices. Precision of PLS increase with increasing number of X variables. In PLS regression the variance in a data matrix X and a dependent matrix Y is decomposed by successively estimating PLS components that captures the variance and correlation between X and Y. PLS systematically extract (decompose) variation in the data matrix (X-matrix) while (PCR) is used to regress each response variable (Y-matrix) onto the decomposed spectra(X-matrix), and make a projection to latent structures. This process allows for simultaneous and independent decomposition of both X- and Y-matrices and then performs the regression of the Y-matrix onto the X-matrix. There are usually two forms of PLS algorithms calculated and used in chemometrics studies; PLS1 and PLS2. PLS1 analysis allows only one Y-variable to be projected against the x-matrix at a time, whiles PLS2 allows the projection of many Y-variables onto the X-matrix. Many studies have shown that PLS1 regression has a marginally better prediction ability of pertinent set of Y-variable than PLS2 modeling of multiple Y-variables set.[38,39,46]

The mathematical basis of these methods are reported elsewhere in literature.[38-39,47-48]

**Useful statistics used to evaluate NIRS and fluorescence calibration models**

A good PLS calibration model fit for accurate future predictions can be measured with several parameters. They include root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP), standard error of prediction (SEP), standard error of calibration (SEC), coefficient of determination ($R^2$), bias, RDP and residual variance. This study focused on future prediction ability of calibration models developed for each response variable (Y-reference value). Therefore the parameters used to measure the performance (prediction accuracy) of our models in this study included $R^2$, RMSEP, SEP and bias.

**Root mean square of prediction**

RMSEP, a measure of the model fit, is a direct estimate of the prediction error in the Y-variables (reference) of the model. It is expressed as the average error expected from future prediction of Y variables using our model and also expressed in the measurement units of the Y variables used in developing the model. Standard error of prediction (SEP) is a measure of precision of the models predicting ability corrected for the bias in the validation of the model.[38,39] SEP measures the average errors that are expected to be associated with future preditions.[39]

**Bias**

Bias is used to detect if there is a systematic difference between average values of the calibration set and the validation set. If such no difference exists, bias will be equal to zero. The closer the bias is to zero, the more accurate is the model developed, both in terms of model fit and predictability of the future sample set. Bias is computed as the average difference between predicted and measured Y-values of all samples in the validation set.[39]

**Coefficient of determination ($R^2$) between predicted and measured values.**

Coefficient of determination ($R^2$) is a measure of total variance between measured and predicted values that can be modeled by linear association. $R^2$ shows the proportion of X- data that can be explained by the variance in the Y-data. It varies between zero and one. The closer $R^2$ is to one, the stronger the correlation. $R^2$ can inflate and approach unity by increasing the complexity of the model. Therefore a high $R^2$, even though a good indicator of a good model

cannot be independently used as the ultimate measure of model performance. Interpretation of $R^2$ may be subjective.[16,38] Generally, an $R^2$ between 0.92-0.96 represents a model that can be used in most applications including quality assurance. An $R^2$ of 0.83-0.90 range also represent a model that can be used in most applications including research activities with some level of caution. Approximate calibration and screening purpose activities can be done with an $R^2$ range of 0.66-0.81. General rough screening activities can utilize a model with an $R^2$ range between (0.5-0.64). $R^2$ ($< 0.5$) may be deemed poor and not useable in any NIR calibration and the reasons should be investigated.[16,38]

# References

1. Baughman, E. 2005. Process Analytical Chemistry: Introduction and historical perspective. In "Process Analytical Technology. Spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries. Ed. K. A. Bakeev. Blackwell Publishing Ltd., Oxford, U.K, 2005; 1-12.

2. So, C.; Via, B. K.; Groom, L.H.; Schimleck, L.R.; Shupe T.F; Kelley, S.; and Rails, T. Near infrared spectroscopy in the forest product industry. Forest Product J. 2004, 54:4, 6-15.

3. Tsuchikawa, S. A review of recent infrared research for wood and paper. Applied spectroscopic reviews. 2007, 42:1, 43-71.

4. Billa, E.; Pastou, A.; Monties, B.; Romero, J and Koukios, E.G. Multivariate chemometric analysis of the fluorescence spectra of eucalyptus wood. Industrial crops and products. 2000, 187-196.

5. Pandey, K.K.; Upreti, N.K.; Srinivasan, V.V. A fluorescence spectroscopic study on wood. Wood Sci and technol. 1998, 32, 309-315

6. Dyer T.S., Wood Fluorescence of Indigenous South African Trees, IAWA Bulletin n.s., 9(1), 1988: 75-87.

7. Olmstead, J.A.; Gray, D.G. Fluorescence spectroscopy of cellulose, lignin and mechanical pulp: A review. J. of pulp and paper sci. 1997, 23:12, 571-581.

8. Obst, J.R.; Ralph, J. Characterization of hardwood Lignin: Investigation of Syringyl/Guaiacyl composition by 13C Nuclear Magnetic Resonance Spectroscopy. Holzforschung 37 (1983) 297-302

9. FOSS, A guide to Near Infrared Spectroscopy Analysis of Industrial Manufacturing processes. FOSS Press, 7703 Montpelliar Road, Suite1, Laurel, MD 20723 USA. www.foss.dk.

10. Hindle, P. "Historical development", In handbook of near infrared analysis, 2nd Edition, Revised and Expanded. Edited by D.A. Burns and E.W. Ciurczak. Practical spectroscopy series, Volume 27, Marcel Dekker , Inc., New York, NY, USA, pp1-6 (2001)

11. Norris, K.H., R.F. Barnes, J.E. Moore, and J.S. Shenk. Predicting forage quality by infrared reflectance spectroscopy. Journal of Animal. Science. 1976, 43, 889-897).

12. Stuart, B.H. 2004. Infrared spectroscopy: Fundamentals and application. John Wiley and sons. West Sussex, England.

13. Stordrange, L. Multivariate analysis of near infrared spectroscopic data from reaction monitoring. Ph.D. Thesis. Department of Chemistry, University of Bergen, Norway, 2003. p5.

14. Ciurczak, E. Principles of near infrared spectroscopy. In handbook of near infra-red analysis, 2nd Edition, Revised and Expanded. Edited by D.A. Burns and E.W. Ciurczak. Practical spectroscopy series, Volume 27, Marcel Dekker , Inc., New York, NY, USA, pp1-6 (2001)

15. Harwood, L.; Claridge, D.W. 2005. Introduction to Organic Spectroscopy. Oxford university press, Great Britain.

16. Williams, P. 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, 2004.

17. Adedipe, O.E. Application of process analytical technology to wood: A near infrared spectroscopy approach. Ph.D. Thesis. Department of Wood Science and Technology, west Virginia University, United states, 2008. pp 10-12

18. Savitzky, A.; Golay, M.J. Smoothing and differentiation of data by simplifies least squares procedures, Anal. Chem. 1964, 36(8). Pp1627-1639

19. Geladi, P.; MacDougall, D. and Martens, H. Linearization and Scatter –Correction for Near Infrared Reflectance Spectra of Meat. Appl. Spectrosc. 39(3), 491-500.(1985)

20. Poke, F.; Wright, J.; and Raymond, C. Predicting extractive and lignin contents in Eucalyptus globulus using near infrared reflectance analysis. J. Wood Chem. Technol. 2004, 24, 55-67.

21. Punsuvon, V.; Terdwongworakul, A.; Thanapase, W and Tsuchikawa, S. Rapid NIR analysis of chemical and mechanical properties for eucalyptus camaldulensis at plantation in Thailand. Near Infrared Spectrosopy: Proceedings of the 11th International conference, 781-785 (2004)

22. Hauksson, J.; Bergqvist, G.; Bergsten, U.; Sjostrom, M and Edlund , U. Prediction of Basic wood properties for Norway spruce: Interpretation of near infrared spectroscopy data using partial least squares regression. Wood Sci and Tech. 35, 475-485(2001)

23. Hoffmeyer, P. and Perderson, J. evaluation of density and strength of Norway spruce wood by near infrared spectroscopy. Holz Roh-werkstoff 53, 165-170 (1995)

24. Thumm, A and Meder, R. Stiffness prediction of radiate pine clearwood test pieces using NIR spectroscopy. J. Near Infrared Spectrosc. 19, 117-122 (2001)

25. Thumm, A.; Meder, R and Bier, H. Veneer stiffness predicted by NIR spectroscopy calibrated using mimi-LVL test panels. Holz als Roh-und Werkstoff, 60. 159-164 (2002)

26. Lindgren, T and Edlund, U. Prediction of lignin content and pulp yield from black liquor using NIR and partial least square regression. Nordic Pulp and Paper Res. J. 9, 76-80(1998)

27. Easty, D.; Berden, S.; De Thomas, F and Brimmer, P. Near Infrared spectroscopy for the analysis of wood pulp: quantifying hardwood-softwood mixtures and estimating lignin content. Tappi J 73, 257-261 (1990)

28. Wallbacks, L.; Edlund, U.; Noeden, B and Iverson, T. Multivariate characterization of pulp, Part1: Spectroscopy characterization of physical and chemical differences between pulp using 13C CP/ MAS NMR, FTIR NIR and multivariate data analysis. Nordic Pulp and Paper Res J. 26, 74-80 (1991)

29. An introduction to fluorescence spectroscopy. PerkinElmer Ltd, Buckingham, United Kingdom. (2000)

30. Wiberg, K. Multivariate spectroscopic methods for the analysis of solutions. Ph.D. Thesis. Department of Analytical Chemistry, Arrhenius Laboratory, Stockholm University, Sweden. 2004. Pp14-17.

31. Royer, C.A. Approaches to Teaching Fluorescence Spectroscopy. Biophysical Journal,68,1191-1195(1995)

32. Guilbault, G.G. Practical Fluorescence, 2nd ed., Marcel Dekkerr, Inc., NY(1990)

33. Christensen, J.; Norgaard, L.; Bro, R.; and Engelsen, S.B. Multivariate Autofluorescence of Intact Food Systems. Chemical Review, 2006, 106(6), pp 1981-1994.

34. Albani, J.R. Principles and application of fluorescence spectroscopy. Blackwell publishing, Oxford, United Kingdom (2007).

35. Sum, S.T.; Singleton, D.L.; Paraskevopoulos, G.; Irwin, R.S.; Barbour, R.J.; Sutcliffe, R. Wood sci and technol. 1991. 25, 405-413.

36. Varmuza K. and P. Filmoser. Introduction to multivariate statistical analysis in chemometrics. 2008, CRC Press, Boca Raton, FL, USA: 321pp).

37. Brereton, R.G. Chemometrics. Data analysis for the laboratory and chemical plant. 2002, John Wiley & Sons, Ltd.Chichester, England: 489pp).

38. Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; and Wold, S. Introduction to multi- and megavariate data analysis using projection methods (PCA, PLS). Umetrics AB, mea, Sweden, 1999, 490pp

39. K.H. Esbensen, Multivariate data analysis in practice: An introduction to multivariate data analysis and experimental design, 5th edition. CAMO Process AS, Oslo, Norway, 598pp (2002)

40. Wold, S.; Geladi, P.; Esbensen, K. Multiway principal components and PLS-analysis. Journal of Chemometrics. 1987:1, 41-56.

41. Analysis of color multiples in sugar processing using fluorescence spectroscopy and chemometrics. Ph.D. Thesis. Department of dairy and Food Science, The Royal Veterinary and Agricultural University, Denmark, 2000, p 18.

42. Gemperline, P.J. Principal Component Analysis. In: Practical guid to chemometrics. Edited by P. Gemperline. Taylor and Francis, Boca Ranton, FL, USA. Pp 69-104(2006)

43. Thielemans, A and Massart, D. The use of principal component analysis as a display method in the interpretation of analytical chemical, biological, environmental and epidemiological data. Chimia, 39, 236-242. (1985)

44. Jackson, J. a user's guide to principal component. John Wiley and Sons, Inc: New York, (1991)

45. Wold, S.; Albano, C.; Dunn, W.; Elmund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W and Sjostrom, M. Multivatriate data analysis in chemistry, In: Chemometrics: mathematics and statistics in chemistry. Kowalski, B.R. (ed). Reidel, D. Publishing Company, Dordrecht, Holland (1984).

46. Stephen S. Kelley, Timothy G. Rials, Rebecca Snell, Leslie H. Groom and Amie Sluiter "Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood", Wood science and technology 38,4(2004).

47. Martens, H., Naes, T.,(1991) Multivariate Calibration. Wiley, pp 419

48. Vandeginste B, Massart D, Buydens L, De Jong S, Lewi P, Smeyers-Verbeke J (1998) Handbook of Chemometics and qualimetrics: Part B. Elsevier, New York, pp 349-379

# CHAPTER TWO: Rapid characterization of biomass using near infrared spectroscopy coupled with multivariate data analysis: Part 1. Yellow-poplar (*Liriodendron tulipifera* L)

## Abstract

This paper is the first of a series of four papers on the use of near infrared (NIR) and fluorescence spectroscopy (FS) coupled with multivariate data analysis (MVDA) as a process analytical technology (PAT) tool for rapid characterization of important physical and chemical properties of two types of lignocellulosic biomass; Northern red oak (*Quercus rubra*) and yellow-poplar (*Liriodendron tulipifera* L) as a potential feed stock for the bio-refinery industry. This paper, Part 1, focuses on the use of NIR as a PAT tool in predicting the chemical and physical properties of a West Virginia hardwood (yellow-poplar). This study revealed that some form of mathematical preprocessing of the NIR spectra (first derivative) greatly improved all prediction models developed in this study. Two spectra regions were used in this study (800-2400nm and 1300-1800nm) and the full NIR spectra-based PLS1 models were comparably similar to models developed with reduced spectra that encompass parts of the first and second overtone of the NIR spectrum (1300-1800nm). Moderate to high correlation ($R^2$) between measured and predicted properties were observed in this study.

**Keywords**: Multivariate data analysis, Near Infrared, Fluorescence, Process analytical technology

## Introduction

For decades, the U.S. economy has depended largely on fossil fuels. More than half of the oil requirement of the U.S. is imported from politically unstable regions of the world. Additionally, predicted reserves of oil world-wide are grim. The energy crisis in 1970's clearly demonstrated that continued dependence on imported foreign petroleum oil by the US poses a serious security threat to the U.S. economy. The uncertainty about the future availability and prices of petroleum oil worldwide has necessitated an urgent need to find alternate U.S. energy sources that can displace imported transportation fuels to ensure the country's energy security.

The billion-ton study estimated that, approximately 3% of U.S. industrial energy needs are currently met by biomass.[1] In 2003, it contributed 2.9 quadrillion Btu of energy. A potential feedstock for the production of bio-fuels and bio-products is lignocellulosic biomass. Lignocellulosic biomass is both renewable and sustainable. It is a potential feedstock for the production of second generation transportation fuels and bio-products. Estimates for biomass contribution to total industrial and electric energy are put at 4% with annual increase of 3.9 quads in 2020.[1] Biomass was projected to contribute to about 10% of transportation fuel consumption in 2020 (4.0 quads) and 25% in 2030.[1] Further increase in production of chemicals and bio-products from biomass is expected. An increase in the production of chemicals from biomass is estimated to be 18% in 2020 and 25% in 2030. The Billion-ton study estimates that sustainable and renewable biomass from land resources can displace 30% or more of U.S. energy oil needs.[1] Biomass can therefore make significant contribution to the overall energy needs of the country.

Biomass is structurally heterogeneous both chemically and structurally. Its chemical, physical, and mechanical properties are source dependent. Sources of biomass include agriculture, logging residues, wood processing operations, and urban wood residues.[1] It is therefore important in the optimization of the bio-refinery process to match the different physical and chemical characteristics to the process parameters. This can be achieved by using rapid and non-destructive methods in place of the traditional slow and expensive wet chemical methods.

Recent developments in electronics and optics have led to the development of small portable spectrometers fitted with optical fibers. This configuration enables source of radiation to be brought to the specimen instead of the traditional setup where the specimen is brought to the radiation source. This reduces the time for analysis and unlike classical spectrometric analysis;

solid and liquid samples can now be analyzed in-situ with minimum sample preparation. These spectrometers are also capable of generating tremendous information about test specimens in a very short time periods (seconds). Such data contain "fingerprint" information about structure, physical and chemical properties of specimen under study. Using multivariate data analysis (MVDA), important and relevant information in the data can be "mined" and used to develop classification and prediction models. This constitutes the basis of Process Analytical technology (PAT).

PAT is defined as the application of analytical chemistry with multivariate data tools for process understanding and achieving control or optimization of manufacturing processes.[2] The immense capability of near infrared spectroscopy (NIRS) as a PAT tool was first demonstrated by the work of Karl Norris *et al.* in late 1950's at the U.S. Department of Agriculture in Maryland.[3] They demonstrated that NIR spectroscopy coupled with MVDA can be used to determine the physical and chemical properties of grains. NIR spectra represent overtones and combination bands. Most chemical and biochemical materials are known to have unique, well defined features at wavelengths that can be used for both qualitative and quantitative purposes. NIR spectra contain absorbance bands mainly due to three chemical bonds, i.e., C-H (fats, oil, and hydrocarbons), O-H (water, alcohol) and N-H (protein).[2] PAT has been successfully used in the petroleum and pharmaceutical industry. In addition to NIR, PAT has also been applied to other spectroscopic methods such as Raman, UV-Visible, Laser Induced Breakdown, Nuclear Magnetic Resonance and fluorescence spectroscopy (FS).

Considerable amount of wood residues is generated by various forestry operations in West Virginia. A recent survey by the Appalachian Hardwood Center (Division of Forestry and Natural Resources, West Virginia University) estimates weekly production of 12,077 tons of sawdust; 22,945 tons of chips and 17,748 tons of bark. Yellow-poplar and northern red oak are two of the most common wood species in West Virginia. Wood waste streams emanating from commercial forestry operations and wood processing plants provide potential lignocellulosic biomass feed stock for bio-refinery operations. Ability to rapidly characterize physical and chemical properties of these two wood species would enhance their potential as feed stock for the bio-refinery industry.

One of the overall objectives of this study was to evaluate FS as a potential PAT tool and compare its performance with near infrared (NIR). NIRS has demonstrated tremendous success

as a PAT tool for the rapid and nondestructive characterization of lignocellulosic biomass properties.[4,5] However, NIR spectroscopy cannot be a catch-all PAT tool for all lignocellulosic properties. FS has been reported as a rapid, nondestructive spectroscopic method for characterizing food products.[6] Fluorescence spectra can be acquired very quickly, thus lending itself as potential PAT tool.

In this study, NIR and Fluorescence were used to characterize some physical and chemical properties of two potential bio-refinery feedstock, yellow-poplar and northern red oak. Specifically, the objectives were to:

1. Use NIR and FS spectroscopy coupled with MVDA to predict important physical and chemical properties of northern red oak (*Quercus rubra*) and yellow-poplar (*Liriodendron tulipifera* L). These properties include ash content, extractives content, total lignin, acid insoluble lignin, holocellulose, and bulk density.

2. Compare the efficiency of two spectra region of NIR and fluorescence, in order to determine which region contains relevant spectroscopic information for accurate quantification of the chemical and physical properties of northern red oak and yellow-poplar.

3. Compare the performance of NIR based and Fluorescence based PLS1 prediction models of each physical and chemical property of northern red oak and yellow-poplar.

The results of this study will be useful to the bio-refinery and bio-products industry. It will encourage the industry to embrace PAT as a tool for process control. The results of this study are presented in four parts. Parts 1 and 2 of this study reports on predicted models developed for of yellow-poplar and northern red oak using NIR. In parts 3 and 4, prediction models for the same properties of the two species using FS are presented. This study allows us to compare the performance of the two potential PAT tools with respect to a common population of each wood species. These results are also presented in paper 3 and 4.

## Materials and methods

Materials and methods used in the overall study are described here in details. The remaining three papers will all make references to this section.

## Materials

A tree each of yellow-poplar and northern red oak were randomly selected in West Virginia University Research forest located about 13 km east of Morganton, WV. The height of each tree was approximately 14 meters. Five centimeter thick disks were cut at 3.5 meter intervals along the bole of the tree starting from the breast height (1.1m). A total of three disks were cut from each tree. Each disk was cut into blocks of size 19 X 19 X 50 mm. Only blocks from the heartwood of each disk were used in this study. Twenty specimens were randomly selected from each disk of both northern red oak and yellow-poplar disks. A total of 60 blocks (specimens) of both species were used in this study. Specimens were oven-dried at $103\pm2$ $^{o}$C for 24 hours and reconditioned at about $23^{o}$C in a vacuum dessicator for 24 hours prior to analysis.

## Methods

**Near infrared (NIR) measurements**

Measurement of NIR spectra of yellow-poplar and northern red oak are described here. The paper on northern red oak makes reference to this section. All wood specimens were scanned with Bruker Matrix-F FT-NIR spectrometer (Bruker Optics Inc., Billerica, MA, USA) with a fiber optic sampling probe for solids and liquids (IN263E) operating in the diffuse reflectance mode (range 833-2500 nm). Each wood specimen was scanned 10 times and averaged to obtain single spectrum. For each specimen, a total of four spectra from the tangential face were taken and converted to the JCAMP format using Bruker OPUS$^{TM}$ software (version 5.0, Bruker Optics. Inc, Billerica, MA, USA). The probe was aligned perpendicular to the collection face for each scan at a distance 5 mm from the sample.

**Determination of chemical properties**

Chemical properties measured included total lignin (insoluble and acid-soluble lignin), extractives, and ash contents. Specific chemical properties of yellow-poplar were determined by classical wet chemistry methods: ASTM D-1106-96, NREL LAP-004(1996), ASTM D1105-96(2007) and ASTM D-1102-84 respectively. Holocellulose was determined by the difference between initial weight of ground wood (oven-dried weight basis) and the sum of the above chemical components.

**Bulk density measurement**

Bulk density of biomass is a major contributing factor to the determination of "cost and logistic requirements for handling and moving biomass from point of production to biorefinery.[7] It is a function of the size, shape and individual particle density of the lignocellulose biomass. Two types of bulk densities were determined: aerated bulk density and tapped bulk density, according to the method described by Yu et *al* (1994)[8] and Adullah *et al. (*1998).[9] Volume and weight determinations for each method was replicated twice for each sample and the average computed for the sample. The chemical and physical properties determined were used as Y-response variable for PLS modeling.

**Development of NIR calibration**

All spectra were analyzed using Unscrambler multivariate data analysis software (version 9.8, Camo Inc., Woodbridge, New Jersey, US). The four spectra collected from each sample were averaged to one spectrum to give a total of sixty (60) averaged spectra for each species. (Sixty samples per species). All spectra data collected were converted from wave numbers to wavelength using OPUS[TM] software. The spectra data were divided into calibration and prediction sets. The calibration set consisted of two spectra out of every three consecutive spectra (40 spectra) and the prediction set comprised of one spectrum out of every three consecutive spectra (20 spectra).

As a first step in the development of a calibration model, a Principal Component Analysis (PCA) was applied to the data to obtain an overview. This process reduces the dimensionality of data by a linear combination of original data to generate new latent variables which are orthogonal and uncorrelated to each other. It distills the information in the original variables into

a lower number of variables – principal components (latent variables).[10] PCA lies along the direction of maximum variance in the data set. [11] The score plot and loading plots of the PCA were observed and this operation revealed a total of six outliers that were removed from the yellow-poplar data and excluded from subsequent model development. As a result of these outlier samples, the number of samples used for calibration (37 spectra) and testing (17spectra) of the model decreased. After removal of outliers, both X- and Y- data matrices were mean-centered. All NIR spectra (independent variables) were combined into a single data matrix [X-matrix] and the physical and chemical properties into separate response matrices [$Y_i$]. Calibration models were developed for two spectra region: (i) full spectra region (800-2500nm) and (ii) the second overtone region and first overtone region (1300-1800nm). The most important and distinct spectra information of NIR is known to be between 1000-2500nm (combination of first overtone, second overtone and combination band region).[12,13] Therefore this much reduced region was chosen to cover part of this important region of the NIR (1000-2500nm).

Calibration models were developed using partial least squares regression (PLS1). PLS1 models were developed for raw spectra data and Savitzky-Golay first derivative pre-processing of the raw data. (15 points of smoothing on both sides of the X-data matrix with a second polynomial order). This preprocessing was done to evaluate improvement in model quality, interpretation and prediction ability. Additionally, data pre-processed using second derivative and Multiple Scatter Correction (MSC) did not significantly improve the model developed and are therefore not reported here. Partial least squares calibration models were developed with X- and Y-matrices. The Y-matrices consisted of data from bulk densities (aerated bulk density and tapped bulk density), ash content, holocellulose content, total lignin, acid insoluble lignin and extractives contents. The response variable holocellulose content, total lignin content and acid insoluble lignin content were preprocessed using a cubed root transformation due to the large differences in the values measured.[14] Calibration models were developed using full-cross validation method and then used to predict the response of the prediction set.

The Unscrambler (version.9.8) software was allowed to determine the optimum number of principal components required with close supervision on the explained and residual variance plots of each model before it was accepted. This was done to prevent under fitting or over fitting of our model.

**Measure of calibration model performance**

The prediction performance of a PLS model can be evaluated using a number of parameters and they include root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP), standard error of prediction (SEP), standard error of calibration (SEC), coefficient of determination ($R^2$), bias, RDP and residual variance. This study focused on future prediction ability of calibration models developed for each response variable. Therefore, the parameters used to measure prediction performance of our models in this study included $R^2$, RMSEP, SEP and bias. Calibration models results are also presented for convenience purposes, even though they not discussed in the results and discussion section.

RMSEP, a measure of the model fit, is a direct estimate of the prediction error in the Y-variables of the model. It is expressed as the average error expected from future prediction of Y variables using our model and also expressed in the measurement units of the Y variables used in developing the model. Bias is used to detect if there is a systematic difference between average values of the calibration set and the validation set. If no such difference exists, bias will be equal to zero. The closer the bias is to zero, the more accurate is the model developed, both in terms of model fit and predictability of the future sample set. Bias is computed as the average difference between predicted and measured Y-values or all samples in the validation set. SEP is a measure of precision of the models predicting ability corrected for the bias in the validation of the model.[12,15,16]

Coefficient of determination ($R^2$) is a measure of total variance between measured and predicted values that can be modeled by linear association. $R^2$ shows the proportion of X- data that can be explained by the variance in the Y-data. It varies between zero and one. The closer $R^2$ is to one, the stronger the correlation. $R^2$ can inflate and approach unity by increasing the complexity of the model. Therefore, a high $R^2$, even though a good indicator of a good model cannot be independently used as the ultimate measure of model performance. Interpretation of $R^2$ may be subjective. Generally, an $R^2$ between 0.92-0.96 represents a model that can be used in most applications including quality assurance. An $R^2$ of 0.83-0.90 range also represent a model that can be used in most applications including research activities with some level of caution. Approximate calibration and screening purpose activities can be done with an $R^2$ range of 0.66-0.81. General rough screening activities can utilize a model with an $R^2$ range between (0.5-0.64). $R^2$ (< 0.5) may be deemed poor and not useable in any NIR calibration and the reasons should be

investigated.[16,17] The evaluation of possible end used of prediction models in this study will be based on this reference.

Result presented in Tables 2 and 3 are PLS1 models developed for the prediction of measured physical and chemical properties of yellow-poplar with raw spectra data matrix and first derivative spectra data matrix.

## Results and discussion

Chemical properties of yellow-poplar are reported on percentage basis (oven-dried weight) of original sample analyzed. (Table 1) Ash content exhibited a range of 0.70-0.79; extractive content had a range of 3.88-6.1, acid-insoluble lignin ranged from 16.59-21.30; acid-soluble lignin content had a range of 0.011-0.067; total lignin content ranged from 21.84-27.49 and holocellulose showed a range of 72.51-78.16. Chemical properties measured compared favorably to those reported previously by Peterson (1984). For the physical property measured, aerated bulk density and tapped bulk density exhibited a range of 0.16-0.20 and 0.21-0.27 g/cm$^3$ respectively. Like typical NIR spectra, all measured spectra of yellow-poplar looked alike (Figure 1).

## Evaluation of chemical and physical properties prediction using raw NIR spectra data

Results of calibration and prediction models of physical and chemical properties of yellow- poplar using the full spectra (800-2500nm) and reduced spectra regions (1300-1800nm) are presented in Table 2. Generally all the calibration and prediction models developed in this study required optimum PC's of 6-10. Almost all calibration models developed with the full spectra exhibited higher model performance as compared to similar prediction models developed with the reduced spectra regions.

For the prediction models developed with the full spectra region, aerated bulk density of yellow-poplar had a high R$^2$ (0.83). Both RMSEP and SEP were low (0.0053). Bias for the full spectra-based prediction model was also low (0.0014). There was only a small decrease in R$^2$ (0.82) when the reduced spectrum was used to predict aerated bulk density. Similar prediction model developed with the reduced spectra exhibited a slightly higher RMSEP (0.0057) and SEP (0.0058) as compared to the full spectra-based model. The reduced spectra region also predicted

aerated bulk density with a low bias (0.0010). Both models required 7 optimum PC's for model development.

Tapped bulk density prediction model also exhibited an $R^2$ of 0.78 and 0.77 for full spectra-based and reduced spectra-based models respectively. Both models had bias value of 0.0010. The full spectra-based model exhibited low RMSEP (0.0072) and SEP (0.0073). There was a slight increase in the RMSEP (0.0086) and SEP (0.0088) of prediction models developed with reduced spectra region as compared to full spectra-based prediction model. Few studies have predicted wood bulk density with NIR spectroscopy. Adedipe and Dawson-Andoh (2008) developed bulk density prediction model for solid yellow-poplar veneer with NIR spectra ($R^2$ of 0.79). Even though this study can not be compared to the results reported here due to differences in how the response variable (bulk density) were measured, the potential use of NIR spectroscopy to predict bulk density was exhibited.[14] Both models required moderate number of PC's for model development and could be used for approximate calibration and screening purposes.

Prediction models for ash content of yellow-poplar exhibited high $R^2$ of 0.87 and 0.83 using calibration models developed with the full NIR and reduced spectra regions respectively. Prediction model developed with the full spectra exhibited low RMSEP (0.0126) and SEP (0.0122) values whiles the prediction model developed with the reduced spectra had a slight increase in RMSEP (0.0144) and SEP (0.0148) values. The full spectra-based and the reduced spectra-based prediction models required slightly high number of PC's (10) and low bias values of 0.0041 and 0.0001 respectively. Both models could be used for applications where accuracy is imperative.

For extractives content, prediction models for yellow-poplar developed with the full spectra and reduced spectra region exhibited high $R^2$ of 0.84 and 0.78 respectively. Low RMSEP (0.2902) and SEP (0.2985) was exhibited by the prediction model developed with the full spectra. Concurrent with the slight decrease in $R^2$, the reduced spectra-based prediction model also exhibited a slight increase in RMSEP (0.3387) and SEP (0.3419) as compared to similar model developed with the full NIR region. The prediction models developed with the full spectra and the reduced spectra exhibited low bias values of 0.0090 and 0.0661 respectively. The development of both models required 8 PCs for model development. Based on the $R^2$ exhibited by both models, the prediction model developed with the full spectra could be used for most

applications where accuracy is important, and the reduced spectra-based prediction model could be used for approximate calibrations and screening purposes.

For acid-insoluble lignin, the prediction model developed with the full NIR region had a moderate $R^2$ (0.64) and a low bias (-0.0039). This model also exhibited low RMSEP (0.0895) and SEP (0.0922). Prediction models developed with the reduced spectra had a low $R^2$ (0.39) and bias (0.0099) values. RMSEP (0.1067) and SEP (0.1096) of the prediction model developed with the reduced spectra increased slightly as compared to the prediction model developed with the full spectra. The reduced spectra-based prediction model of acid insoluble lignin was poor and unusable because of low $R^2$ (0.39). The prediction model developed with the full spectra could be used for rough screening purposes ($R^2$ 0.64). The calibration and prediction models for both spectra regions required a high number of PC's (8-9) for model development.

Similarly, total lignin content of yellow-poplar was predicted with a moderate $R^2$ (0.64) using the calibration model developed with the full spectra whiles prediction model developed with the reduced spectra exhibited a low $R^2$ (0.38). Prediction model developed with the full spectra exhibited low RMSEP (0.0876), SEP (0.0905), and bias (-0.0024). Even though the reduced spectra-based prediction model of total lignin had low RMSEP and SEP values (< 0.12), the model was poor and unusable due to its low $R^2$ (0.38). The prediction model developed with the full spectra-based prediction model was adequate for rough screening activities based on its $R^2$. High number of PC's (8-9) was required to develop both prediction models using the full spectra and reduced spectra.

Holocellulose, a measure of total sugars in a biomass, was predicted with moderate $R^2$ values of 0.68 and 0.63 for the full spectra-based and reduced spectra-based prediction models respectively. The full spectra-based model exhibited low RMSEP (0.0474) and SEP (0.0489) value, with a bias (0.0003) close to zero. A slight increase in RMSEP(0.0507) and SEP(0.0518) of holocellulose prediction model developed with the reduced spectra region as compared to similar model developed with the full spectra region was observed. The increase in RMSEP and SEP can be attributed to the decrease in $R^2$ value; this invariably increases model error.[15,16] Similarly, holocellulose content prediction model developed with the reduced spectra region exhibited a low bias (-0.0067) close to zero. Both prediction model of holocellulose required 6-7 PC's for model development. Based on the $R^2$ exhibited by each model, the full spectra-based

prediction model could be used for approximate calibration activities whiles the reduced spectra-based prediction model could be used for rough screening purposes.

Comparatively, the prediction models developed with the full spectra was generally better than the reduced spectra-based prediction models; this observation was based on the slight improvement in $R^2$ and a small reduction in RMSEP/SEP values of the full spectra-based models as compared to the prediction models developed with the reduced spectra region.

## Evaluation of chemical and physical properties prediction using first derivative preprocessed NIR spectra data

Table 3 presents results of physical and chemical properties prediction models of yellow-poplar developed with first derivative preprocessed spectra. Generally, prediction models developed with the preprocessed full and reduced spectra regions improved greatly as compared to similar raw spectra-based prediction models developed for each response variable considered in this study. There was a great decrease in the number of PC's (3-6) required to develop the models and an increase or almost equal $R^2$ values. Comparatively, with the exception of acid-insoluble lignin and holocellulose content prediction models, the reduced spectra-based prediction models of the preprocessed data generally predicted all response variables with a much improved accuracy than the full spectra-based models. Acid-insoluble lignin and holocellulose content prediction models exhibited the reverse of this pattern.

The prediction model of aerated bulk density revealed a high $R^2$ for the full spectra-based ($R^2$ 0.83) and reduced spectra-based ($R^2$ 0.84) models respectively. The prediction model developed with the full spectra region predicted aerated bulk density with low RMSEP (0.0054), SEP (0.0055), and bias value close to zero (-0.0007). Equal RMSEP and SEP value (0.0051) was exhibited by the prediction model developed with the reduced spectra region. The latter model also exhibited a low bias (0.0012) value. Both models require low number of PC's for model development and could be used for most prediction activities with some caution based on the $R^2$ values exhibited.

Tapped bulk density, the second type of bulk density, exhibited good $R^2$ (0.79) when predicted with the calibration model developed with the full spectra region. The prediction model developed with reduced spectra exhibited a slightly higher $R^2$ (0.86). Both spectra regions predicted tapped bulk density with low RMSEP and SEP values (< 0.0073), with the reduced

33

spectra-based prediction model having a slightly lower RMSEP/SEP than the prediction model developed with the full spectra. Bias value of both prediction models of tapped bulk density were low, with the full spectra-based region and reduced spectra-based models having bias values of (-0.0031) and (0.0012) respectively. All the prediction models developed for both bulk densities required low number of PC's (4-5) for model development. With respect to the $R^2$ exhibited by both models, the full spectra-based prediction model was adequate for approximate calibration and screening purposes, whiles the reduced spectra-based prediction model was good for most prediction applications including research purposes.

Ash content was predicted with strong $R^2$ of 0.82 and 0.92 for the full and reduced spectra-based prediction models respectively. Both prediction models exhibited low bias values (< 0.0025). Prediction model developed with the full spectra region had low RMSEP (0.0132) and SEP (0.0133). The reduced spectra region predicted ash content with a slightly lower RMSEP and SEP (0.0099). Both prediction models required low number of PC's (5-6) for model development. Based on the $R^2$ exhibited by the models, the prediction model developed with the full spectra was judged to be adequate for approximate calibrations and screening activities whiles the reduced spectra-based prediction model was adequate for most application including quality assurance purposes.

For extractives content, prediction models developed with the full and reduced spectra regions exhibited good $R^2$ values of 0.78 and 0.83 respectively. The prediction model developed with the full spectra exhibited low RMSEP (0.3422) and SEP (0.3387). The reduced spectra-based prediction model had a slightly lower RMSEP (0.2945) and SEP (0.3029) as compared with the models developed with the full spectra. Low bias values were observed in both prediction models, with the full spectra-based and reduced spectra-based prediction models exhibiting bias values of (0.0954), and (-0.0057) respectively. The prediction models of both spectra regions for extractives content required low number of PC's (4) for model development. The prediction model developed with the full spectra was adequate for approximate calibration and screening purposes whiles the reduced spectra-based prediction models was good for most prediction applications including research activities. These observations were based on the $R^2$ exhibited by both models.

Acid-insoluble lignin content of yellow-poplar was predicted with good $R^2$ value of 0.79 and 0.75 for the full spectra-based and reduced spectra-based prediction models respectively.

These results were highly improved models as compared to the moderate $R^2$ (0.64) and low $R^2$ (0.39) values that was exhibited when the full raw and reduced raw spectra regions were used to predict acid-insoluble lignin. The full preprocessed spectra predicted acid-insoluble lignin with low bias (0.0037), RMSEP (0.0696) and SEP (0.0717) values. Similarly, low RMSEP (0.0792), SEP (0.0783) and bias (0.0224) values were exhibited by the prediction model developed with the reduced preprocessed spectra region. The number of PC's required to develop both prediction models were low and within the range of 3-5. With respect to the $R^2$ values exhibited by both models of the preprocessed spectra, the models were adequate for screening and approximate calibration purposes.

Prediction models of total lignin exhibited moderate $R^2$. The full spectra and reduced spectra-based models exhibited $R^2$ of 0.66 and 0.77 respectively. These $R^2$ values were significant improvement over respective similar models developed with the raw NIR spectra. Low RMSEP (0.0885) and SEP (0.0895) values were observed in the prediction models developed with the full spectra. The reduced spectra-based model also exhibited slightly lower RMSEP (0.0717) and SEP (0.0724). The slight reduction in RMSEP and SEP in the latter may be attributed to the slightly higher $R^2$ value over the full spectra-based prediction model. Low bias values of -0.0165 and 0.0143 were exhibited by the full spectra-based and the reduced spectra-based prediction models respectively. Low numbers of PC's (3-4) were required to develop both prediction models of total lignin.  Based on the $R^2$ values exhibited by the models, both models were judged adequate for approximate calibration and screening activities.

For holocellulose content, the prediction model developed with full spectra had a high $R^2$ (0.81), whiles the model developed with the reduced spectra region exhibited a moderate $R^2$ (0.77). Both prediction models developed with the full spectra and reduced spectra had low RMSEP and SEP (< 0.043). Very low bias values were exhibited by the prediction model developed with the full spectra (0.0012) and reduced spectra (0.0057) regions. Prediction model of each spectra region was developed with low PC's (4). This was a much lower number of PC's as compared to similar models developed with the raw spectra data. Both models were adequate for screening and approximate calibration purposes based on the $R^2$ range (0.77-0.81) exhibited by the models.

Improvement of prediction results of some physical and chemical properties of biomass using NIR after some form of  preprocessing of the raw spectra data have been reported, these

preprocessing include Savitzky- Golay first derivative, second derivative, Standard Normal Variate and Multiplicative Scatter Correction.[12-14]

## Conclusion

The results of this study have demonstrated that NIR can be used to predict some chemical and physical properties of yellow-poplar, a potential biomass feedstock for the production of bio-products and bio-fuels. Different response variables predicted showed varying degrees of model strength. With the exception of acid insoluble lignin and total lignin, which was predicted poorly with the reduced raw spectra ($R^2$ < 0.4), all the chemical and physical properties predicted with the full spectra and reduced spectra of the raw data exhibited $R^2$ values greater than 0.6. The findings of this study have also shown that first derivative preprocessing of raw spectra data generally improved the performance of the models. The improvement in model performance of preprocessed data is shown by the decrease in optimum number of PCs, and an increase in $R^2$ values.

This study has also shown that the reduced spectra region (1300-1800nm), which encompass part of the first and second overtone regions of the NIR spectrum, can be used to predict the physical and chemical properties of yellow-poplar with an $R^2$ greater than 0.75 after the preprocessing of the spectra. The $R^2$ values of the reduced spectra-based models of the preprocessed spectra were comparable to models developed with the preprocessed full spectra. The reduced spectra-based models predicted all the chemical and physical properties in this study with an $R^2$ greater than 0.77, with the exception of total lignin content which was predicted with a slightly lower $R^2$ of 0.66. This is an important finding for industries that are interested in pursuing cheaper and portable spectrometers with shorter wavelength range for rapid acquisitions and analysis. It must be noted that no attempt was made in this study to attribute wavelength regions to predicted response variables considered in this study.

**Table 1. Statistics of measured chemical and physical properties of yellow-poplar (60 samples)**

|  | Mean* | SD[a] | Min[b] | Max[c] |
|---|---|---|---|---|
| Aerated Bulk Density (g/cm$^{3)}$) | 0.18 | 0.014 | 0.16 | 0.20 |
| Tapped Bulk Density (g/cm$^{3)}$) | 0.24 | 0.017 | 0.210 | 0.266 |
| Ash (%) | 0.76 | 0.036 | 0.7 | 0.79 |
| Extractives (%) | 5.03 | 0.720 | 3.88 | 6.1 |
| Acid Insoluble Lignin (%) | 19.39 | 1.29 | 16.59 | 21.3 |
| Soluble Lignin (%) | 0.03 | 0.012 | 0.011 | 0.067 |
| Total Lignin (%) | 19.40 | 1.27 | 21.84 | 27.49 |
| Holocellulose (%) | 73.00 | 0.75 | 72.51 | 78.16 |

*Mean, [a]Standard deviation, [b]Minimum measured value, [c]Maximum maximum measured value

**Table 2. Results of PLS1 calibration and prediction models developed for the chemical and physical properties of yellow-poplar using the raw spectra at two NIR wavelength regions**

| | Wavelength (nm) | [a]PC | [b]$R^2$cal | [c]RMSEC | [d]SEC | [e]$R^2$val | [f]RMSEP | [g]SEP | BIAS |
|---|---|---|---|---|---|---|---|---|---|
| Bulk Density Aerated (g/cm$^3$) | | | | | | | | | |
| | 800-2500 (A) | 7 | 0.85 | 0.0054 | 0.0055 | 0.83 | 0.0053 | 0.0053 | 0.0014 |
| | 1300-1800 ( C ) | 7 | 0.86 | 0.0053 | 0.0054 | 0.82 | 0.0057 | 0.0058 | 0.0010 |
| Bulk Density Tapped (g/cm$^3$) | | | | | | | | | |
| | 800-2500 (A) | 7 | 0.80 | 0.0074 | 0.0075 | 0.78 | 0.0072 | 0.0073 | -0.0010 |
| | 1300-1800 ( C ) | 7 | 0.89 | 0.0055 | 0.0056 | 0.77 | 0.0086 | 0.0088 | 0.0010 |
| Ash (%) | | | | | | | | | |
| | 800-2500 (A) | 10 | 0.99 | 0.0043 | 0.0044 | 0.87 | 0.0126 | 0.0122 | 0.0041 |
| | 1300-1800 ( C ) | 10 | 0.98 | 0.0048 | 0.0490 | 0.83 | 0.0144 | 0.0148 | 0.0001 |
| Extractives (%) | | | | | | | | | |
| | 800-2500 (A) | 8 | 0.86 | 0.2640 | 0.2678 | 0.84 | 0.2902 | 0.2985 | 0.0090 |
| | 1300-1800 ( C ) | 8 | 0.83 | 0.2878 | 0.2917 | 0.78 | 0.3387 | 0.3419 | 0.0661 |
| Acid Insoluble Lignin (%) | | | | | | | | | |
| | 800-2500 (A) | 8 | 0.83 | 0.0599 | 0.0606 | 0.64 | 0.0895 | 0.0922 | -0.0039 |
| | 1300-1800 ( C ) | 9 | 0.86 | 0.0540 | 0.0550 | 0.39 | 0.1067 | 0.1096 | 0.0099 |
| Total Lignin (%) | | | | | | | | | |
| | 800-2500 (A) | 8 | 0.83 | 0.0592 | 0.0600 | 0.64 | 0.0878 | 0.0905 | -0.0024 |
| | 1300-1800 ( C ) | 9 | 0.85 | 0.0550 | 0.0550 | 0.38 | 0.1067 | 0.1097 | 0.0087 |
| Holocellulose (%) | | | | | | | | | |
| | 800-2500 (A) | 7 | 0.77 | 0.0289 | 0.0293 | 0.68 | 0.0474 | 0.0489 | 0.0003 |
| | 1300-1800 ( C ) | 6 | 0.78 | 0.0283 | 0.0288 | 0.63 | 0.0507 | 0.0518 | -0.0067 |

[a]No of principal components, [b]Calibration $R^2$, [c]Root mean square of calibration, [d]Standard error of calibration, [e]Validation $R^2$, [f]Root mean square of validation, [g]Standard error of prediction

**Table 3. Results of PLS1 calibration and prediction models developed for the chemical and physical properties of yellow-poplar using the first derivative preprocessed spectra at two NIR wavelength regions**

| | Wavelength (nm) | [a]PC | [b]$R^2$cal | [c]RMSEC | [d]SEC | [e]$R^2$val | [f]RMSEP | [g]SEP | BIAS |
|---|---|---|---|---|---|---|---|---|---|
| **Bulk Density Aerated (g/cm$^3$)** | | | | | | | | | |
| | 800-2500 (A) | 4 | 0.9 | 0.0044 | 0.0044 | 0.83 | 0.0054 | 0.0055 | -0.0007 |
| | 1300-1800 ( C ) | 4 | 0.89 | 0.0047 | 0.0047 | 0.84 | 0.0051 | 0.0051 | 0.0012 |
| **Bulk Density Tapped (g/cm$^3$)** | | | | | | | | | |
| | 800-2500 (A) | 4 | 0.86 | 0.0067 | 0.0068 | 0.79 | 0.0072 | 0.0067 | -0.0031 |
| | 1300-1800 ( C ) | 5 | 0.92 | 0.0051 | 0.0052 | 0.86 | 0.0063 | 0.0064 | 0.0012 |
| **Ash (%)** | | | | | | | | | |
| | 800-2500 (A) | 5 | 0.96 | 0.0074 | 0.0075 | 0.82 | 0.0132 | 0.0133 | 0.0024 |
| | 1300-1800 ( C ) | 6 | 0.98 | 0.0055 | 0.0056 | 0.92 | 0.0099 | 0.0099 | 0.0022 |
| **Extractives (%)** | | | | | | | | | |
| | 800-2500 (A) | 4 | 0.83 | 0.2846 | 0.2885 | 0.78 | 0.3422 | 0.3387 | 0.0954 |
| | 1300-1800 ( C ) | 4 | 0.81 | 0.3066 | 0.3108 | 0.83 | 0.2945 | 0.3029 | -0.0057 |
| **Acid Insoluble Lignin (%)** | | | | | | | | | |
| | 800-2500 (A) | 5 | 0.89 | 0.0488 | 0.0495 | 0.79 | 0.0696 | 0.0717 | 0.0037 |
| | 1300-1800 ( C ) | 3 | 0.7 | 0.0810 | 0.0828 | 0.75 | 0.0792 | 0.0783 | 0.0224 |
| **Total Lignin (%)** | | | | | | | | | |
| | 800-2500 (A) | 4 | 0.81 | 0.0633 | 0.0641 | 0.66 | 0.0885 | 0.0895 | -0.0165 |
| | 1300-1800 ( C ) | 3 | 0.68 | 0.0806 | 0.0817 | 0.77 | 0.0717 | 0.0724 | 0.0143 |
| **Holocellulose (%)** | | | | | | | | | |
| | 800-2500 (A) | 4 | 0.79 | 0.0370 | 0.0375 | 0.81 | 0.0416 | 0.0429 | 0.0012 |
| | 1300-1800 ( C ) | 4 | 0.64 | 0.0445 | 0.0451 | 0.77 | 0.0401 | 0.0409 | 0.0057 |

[a]No of principal components, [b]Calibration $R^2$, [c]Root mean square of calibration, [d]Standard error of calibration, [e]Validation $R^2$, [f]Root mean square of validation, [g]Standard error of prediction
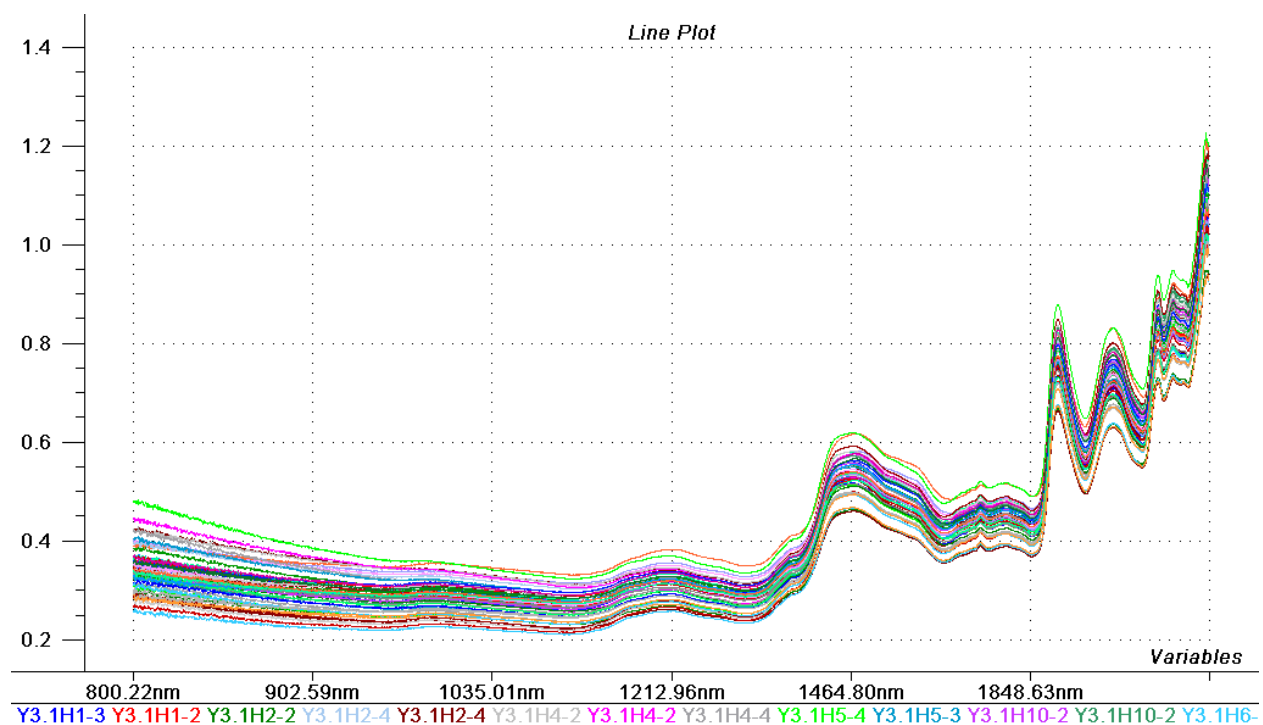
**Figure 1a. Representative spectra of northern red oak**
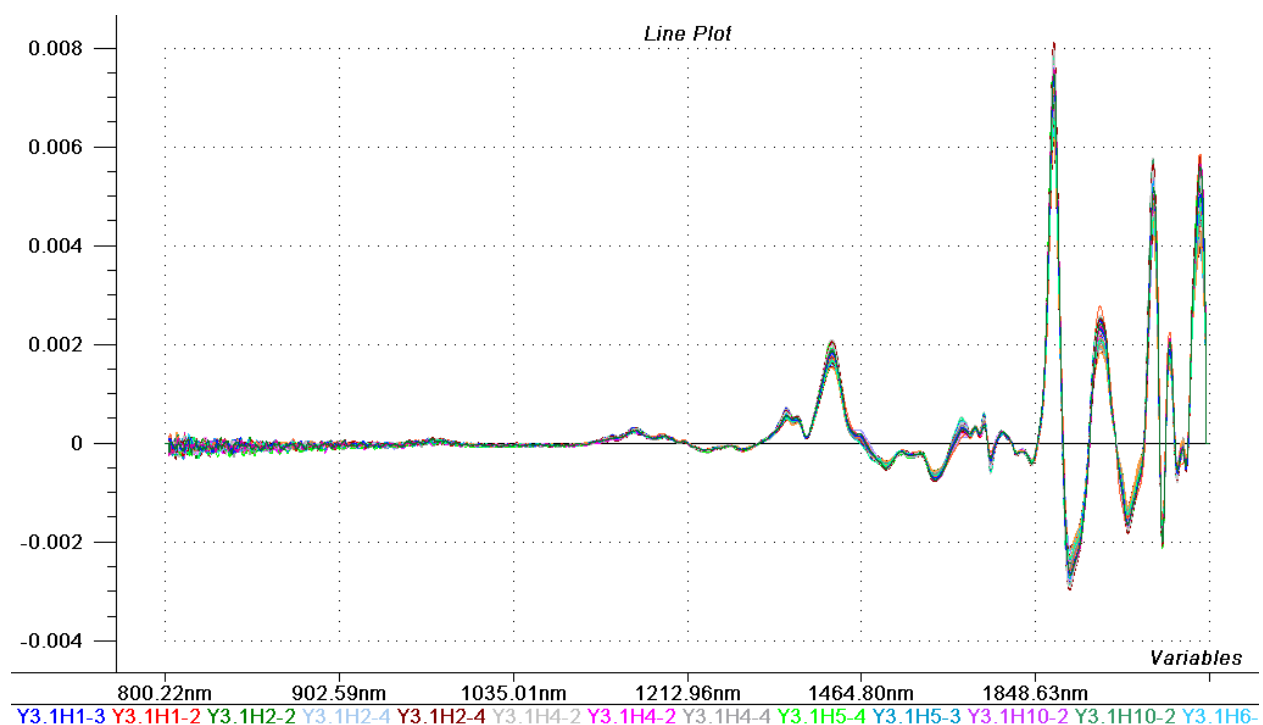


**Figure 1b. First derivative preprocessed spectra of yellow-poplar heart wood**
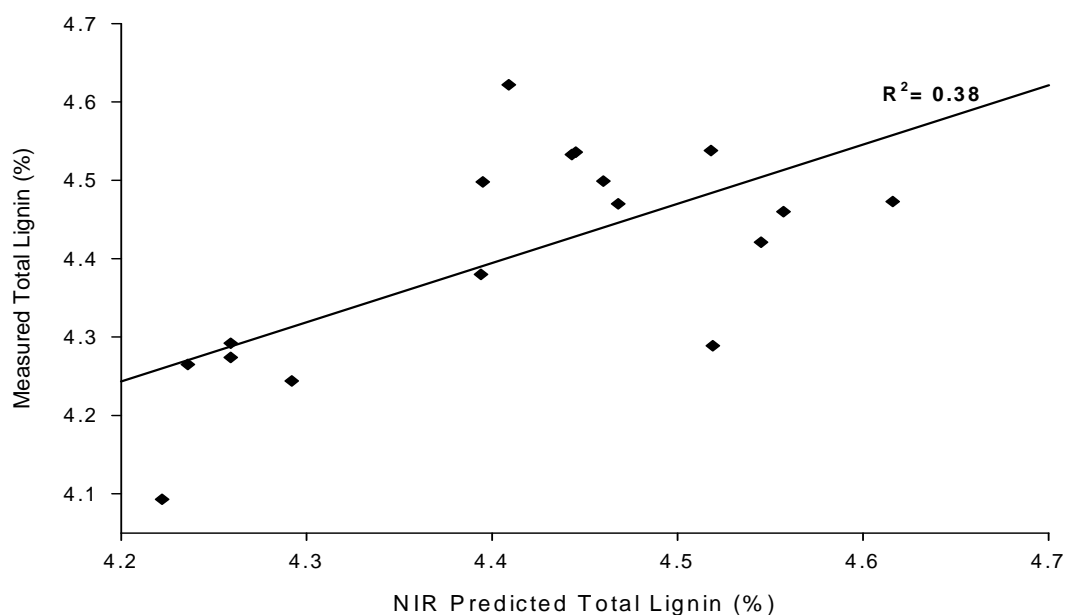
**Figure 2a. PLS1 plot of predicted against measured total lignin using raw NIR spectra region of 1300-1800nm**
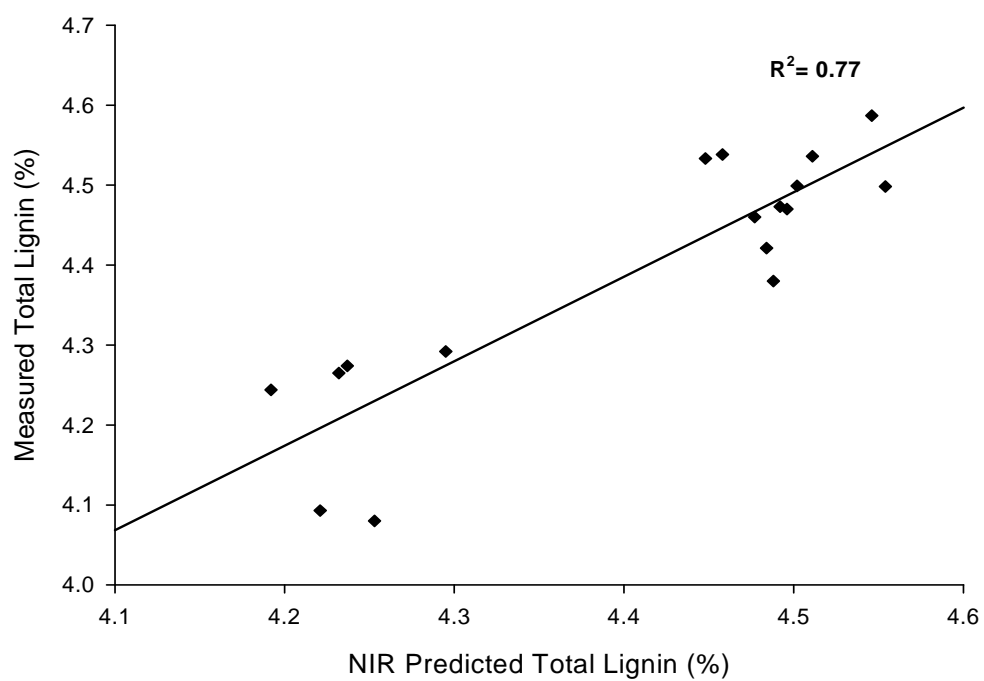


**Figure 2b. PLS1 plot of predicted against measured total lignin using preprocessed NIR spectra region of 1300-1800nm**

# References

1. Perlack R.D., Wright, L.L., Turnhollow, A.F., Graham, R.L., B.J. Stokes, and D.C. Erbach. Biomass as feed stock for a bioenergy and bioproducts industry: The technical feasibility of a billion-ton annual supply US Department of Energy. 2007, DOE/GO-102995-2135, http://www.ostl.gov/bridge)

2. Baughman, E. 2005. Process Analytical Chemistry: Introduction and and historical perspective. In "Process Analytical Technology. Spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries. Ed. K. A. Bakeev. Blackwell Publishing Ltd., Oxford, U.K,2005; 1-12.

3. Norris, K.H., R.F. Barnes, J.E. Moore, and J.S. Shenk. Predicting forage quality by infrared reflectance spectroscopy. Journal of Animal. Science. 1976, 43, 889-897).

4. So, C.; Via, B. K.; Groom, L.H.; Schimleck, L.R.; Shupe T.F; Kelley, S.; and Rails, T. Near Infrared spectroscopy in the forest product industry. Forest Product J. 2004, 54:4, 6-15.

5. Tsuchikawa, S. A review of recent near infrared research for wood and paper. Applied spectroscopic reviews. 2007, 42:1, 43-71.

6. Sadecka and J. Tothova. Fluorescence spectroscopy and chemometrics on the food classification – a review. Czech J. Food Sc. 2007, 25: 159-173.

7. Lam P.S.; Sokhansanj, S.; Bi, X.; Mani, S.; Lim, C.J.; Womac, A.R.; Hoque, M.; Peng, J.; JayaShankar, T.; Naimi, L.M.; and Nayaran, S. 2007. Physical characterization of wet and dry wheat straw and switchgrass – bulk and specific density. 2007 ASABE Annual International Meeting. Sponsored by ASABE, Minneapolis Convention Center, MN, 17-20 June 2007.

8. Yu, A.B.; Hall, J.S. Packing of fine powders subjected to tapping. Powder Technology. 1994, 78, 247–256.

9. Adullah E.C.; Geldart, D. The use of bulk density measurements as flowability indicators. Powder technology, 1999, 102, 151-165.

10. Wold, S.; Geladi, P.; Esbensen, K. Multiway principal components and PLS-analysis. Journal of Chemometrics. 1987:1, 41-56.

11. Analysis of color multiples in sugar processing using fluorescence spectroscopy and chemometrics. Ph.D. Thesis. Department of dairy and Food Science, The Royal Veterinary and Agricultural University, Denmark, 2000, p 18.

12. Sluiter, A.; Snell, R.; Kelley, S.; Rials, T.; and Groom, L. Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood. Wood Sci Technol, 2004, 38, 257-276.

13. Poke, F.; Wright, J.; and Raymond, C. Predicting extractive and lignin contents in Eucalyptus globulus using near infrared reflectance analysis. J. Wood Chem. Technol. 2004, 24, 55-67.

14. Adedipe, O.; Dawson-Andoh, B. Prediction of yellow-poplar (Liriodendron tulipifera) veneer stiffness and bulk density using near infrared spectroscopy and multivariate calibration. J. Near Infrared Spectrosc. 2008, 16, 487-496.

15. Esbensen, K.H. Multivariate data analysis in practice: An introduction to multivariate data analysis and experimental design, 5th edition. CAMO Process AS, Oslo, Norway, 2002, 598pp.

16. Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; and Wold, S. Introduction to multi- and megavariate data analysis using projection methods (PCA, PLS). Umetrics AB, mea, Sweden, 1999, 490pp.

17. Williams, P. 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., 2004.

18. Pettersen, C. The chemical composition of wood. In: Rowell, R.M., ed. The chemistry of solid wood. Advances in chemistry series. 207 Washington D.C: American Chemical Society; 1984: chapter 2.

# CHAPTER THREE: Rapid characterization of biomass using near infrared Spectroscopy coupled with multivariate data analysis: Part 2. Northern red oak (*Quercus rubra*)

## Abstract

Rapid characterization of biomass as a raw material is vital to the emerging bio-fuel and bio-products industry. In this study, NIR spectroscopy coupled with Multivariate data analysis was used to develop calibration models for the chemical and physical properties of northern red oak which was tested with separate samples not included in the calibration of the model. Two wavelength regions (800-2500nm and 1300-1800nm) were modeled. The performances of preprocessed PLS1 prediction models (first derivative) were superior to the raw spectra-based prediction models. Models based on the two spectra regions were similar in performance and exhibited moderate to strong coefficient of determination ($R^2$) between measured and predicted properties with the exception of Ash content that exhibited poor PLS1 prediction models.

**Keywords:** Multivariate data analysis, Near infrared spectroscopy,

## Introduction

Today, energy has become a global security and economic issue characterized by an earnest search for alternatives to fossil fuels. Approximately 97% of U.S. transportation industry energy needs come from petroleum-based sources.[1] The U.S. imports approximately 60% (approximately 190 billion gallons) of its petroleum oil needs. About 13% of this comes from politically unstable Persian Gulf region. U.S. energy consumption is projected to increase more than 30% by 2030 from its 2006 levels. Agriculture products and other finished goods, requires transportation from the source to the market. Therefore the cost of transportation fuel is an important contributor to prices of all services and goods in US economy. Petroleum-based fuels not only affect the economy, but also the environmental health of the country because they closely associated with the emission of green house gases. The latter exerts a significant negative impact on the environment. The deleterious environmental impacts of green house gas emissions precipitated the creation of the Kyoto Protocol.[2]

Lignocellulosic biomass conversion into clean forms of energy has been a subject of interest as a potential alternative to fossil fuel for some time.[3] It is estimated that the existing forestlands and agricultural lands in the U.S. could potentially produce 1.3 billion dry tons of biomass per year. This amount of biomass will meet more than one-third of the transportation needs of the U.S. Out of the 1.3 billion dry tons, forestlands will contribute approximately 368 million dry tons. This can be used to displace about 30% of fossil fuel need in the united state. Presently biomass fuels used in US industries, meet about 3% of the total energy needs annually and this is expected to increase in the near future.[3]

The conversion of biomass into any form of fuel presents a numbers of challenges because of its heterogeneous nature, both physically and chemically.[4] Physical and chemical properties of biomass impact its conversion into bio-fuels and bio products. Today, these properties are largely determined by standard wet chemical analyses which are destructive, slow, and expensive protocols. Also these standard methods of analysis do not lend themselves to on/at/in-line implementation. Development of robust rapid, non-destructive characterization methods will be a great boost to the emerging bio-refinery industry. Important physical and chemical properties of biomass that are relevant to the bio-refinery process include bulk density, particle size, moisture content, gross calorific value, lignin content, extractives content, ash content, cellulose and hemicelluloses contents.

Advancement in electronic and optics technologies have produced cheap, portable and simple spectrometers that brings the light source to the specimens and also generates tremendous multi-component information in very short periods (seconds). Additionally, these methods require no or minimum sample preparation.  These methods are the foundation of a new approach to measuring important properties of industrial processes - Process Analytical technology (PAT).  It has been successfully used with multivariate data analysis (MVDA) to develop prediction models for process parameter. PAT is explained in detail elsewhere.[5] Near infrared spectroscopy (NIRS) coupled with MVDA is one of the most predominant PAT tools available. This PAT tool has enjoyed considerable success in the characterization of physical, chemical and mechanical properties of lignocellulosic biomass. Heterogeneity in lignocellulosic biomass exists between, within species, and to some extent between geographic locations of growth.[6,7] Process analytical tools are also lignocellulosic biomass type dependent. Therefore prediction models developed for properties of yellow-poplar cannot be applied to other species or types. Consequently, new prediction models need to be developed for northern red oak. This paper reports the results of prediction models for the same physical and chemical properties for northern red oak using the same PAT tool – Near infrared (NIR) coupled with MVDA used in earlier study for yellow-poplar. Northern red oak used in this study was harvested from the same location as the yellow-poplar.

Specifically the objectives of this paper are to:

1. Use NIR coupled with MVDA as a PAT method to develop calibration and prediction tools for some physical and chemical properties of northern red oak (*Quercus rubra*). These properties include ash content, extractives content, total lignin, acid insoluble lignin, holocellulose, and bulk density.

2. Compare the efficiency of two spectra region of the NIR spectrum in order to determine which spectra region contains relevant spectroscopic information for accurate quantification of the chemical and physical properties of northern red oak.

## Materials and methods

### Materials

Northern red oak was randomly selected in West Virginia University Research forest located about 13 km east of Morganton, WV. Each tree was approximately 14 meters tall. Three 50 mm disks were removed from the bole and a total of 60 samples from the heartwood with size 19mm X 19mm X 50 mm each were randomly selected for this study. Detailed description of sampling methods and conditioning are described elsewhere.[5]

### Methods

#### Near infrared (NIR) measurements

Bruker Matrix-F FT-NIR spectrometer (Bruker Optics Inc., Billerica, MA, USA) coupled with a fiber optic sampling probe, for solids and liquids (IN263), operating in the diffuse reflectance mode (range 833-2500 nm), was used to collect NIR spectra of each sample. Measurement of NIR spectra of northern red oak specimens were done as per protocols described in details elsewhere s for yellow-poplar.[5]

#### Determination of chemical properties

Chemical properties measured include total lignin (insoluble and acid-soluble lignin), extractives content, holocellulose, and ash contents. These properties were determined by classical wet chemistry methods as per protocols of ASTM D-1106-96, NREL LAP-004(1996), ASTM D1105-96(2007) and ASTM D-1102-84 respectively. Holocellulose was given by the difference between initial weight of ground wood and the sum of the above chemical components.

#### Bulk density measurement

Bulk density of biomass is a major contributing factor to the determination of "cost and logistic requirements of handling and moving biomass from point of production to biorefinery.[8] It is a function of the size, shape and individual particle density of the lignocellulose biomass. Two types of bulk densities were determined; aerated bulk density and tapped bulk density according to the method described by Yu *et al.* (1994)[9] and Adullah *et al.* (1998).[10] Both volume and weight determination for each method was replicated twice for each sample and the average

computed for the sample. The measured physical and chemical properties of northern red oak were used as $Y_i$-response variables for PLS modeling.

## Development of NIR calibration

All spectra were analyzed using Unscrambler multivariate data analysis software (version 9.8, Camo Inc., Woodbridge, New Jersey, US). A total of 60 averaged spectra for the heartwood of northern red oak as described previously were used.[5] To obtain an overview of data, Principal Component Analysis (PCA) was performed and it revealed the presence of six outliers. These were eliminated from data in subsequent analysis. Spectra data were divided into calibration and validation sets. The calibration set consisted of two spectra out of every three consecutive spectra (n=37 spectra) and the prediction set (n=17 spectra) comprised of one spectrum out of every three consecutive spectra. For the development of calibration models, all NIR spectra were combined into a single data matrix (X-matrix, independent variables). Each measured physical or chemical property was also combined into separate data matrices (Y-matrices, dependent variables). The Y-matrices consisted of data from bulk densities (aerated bulk density and tapped bulk density), ash content, holocellulose content, total lignin, acid-insoluble lignin, and extractives contents.

Using the X-matrix, calibration models were developed for two spectra region: (i) full NIR Spectrum (800-2500nm) and (ii) part of the second overtone region and first overtone region (1300-1800nm). PLS1 calibration models were developed using the raw spectra data and Savitzky-Golay first derivative (15 smoothing points on both sides of the X- matrix data, with a second polynomial order) pre-processed data. PLS1 calibration models were developed with X- and Y-matrices. The response variables holocellulose content, total lignin content and acid-insoluble lignin content were preprocessed using a cubed root transformation due to the large differences in the values measured.[11] The Y-data matrix were mean –centered and calibration models were developed using full-cross validation method and then used to predict the response of the validation set. Detailed multivariate analyses (calibration and validation methods) of the data are described elsewhere.[5]

**Measure of calibration model performance**

Parameters used to evaluate the performance of our cprediction models include: (i) Coefficient of determination ($R^{2)}$, (ii) root mean square error of prediction (RMSEP), (iii) standard error of prediction (SEP), and (iv) bias. These parameters are discussed in details elsewhere.[5] Root mean square error of prediction is a direct estimate of the prediction error in the Y variables of the model; whiles SEP is a measure of precision of the models predicting ability corrected for the bias in the validation of the model. Bias is a measure of systematic difference between average values of the calibration set and the validation set of a model. If no such difference exists, the bias will be equal to zero. Coefficient of determination ($R^2$) is a measure of total variance between measured and predicted values that can be modeled by linear association. The closer $R^2$ value is to one, the stronger the correlation.[12-15] Result presented in Tables 2 and 3 are PLS1 models developed for prediction of measured physical and chemical properties with raw spectra data matrix and first derivative spectra data matrix of northern red oak.

## Results and discussion

Results of measured chemical and physical properties of northern red oak are presented in Table1. Chemical properties of northern red oak are expressed as a percentage of the original oven dry-weight of ground wood analyzed. Ash content had a range of 0.15-0.20; extractives content had a range of 4.48-7.74; acid insoluble lignin exhibited a range of 17.90-21.95 whiles total lignin content had a range of 18-21.97. Holocellulose content also had a range of 71.48-74.93. These results were comparable to earlier data reported by Petersen (1984). For the physical property measured; aerated bulk density and tapped bulk density exhibited a range of 0.23-0.28 and 0.31-0.38 g/cm$^3$ respectively. All collected NIR spectra were similar (Figure 1).

## Evaluation of chemical and physical properties prediction using raw NIR spectra data

Table 2 present results of PLS1 models developed with raw spectra data covering the full NIR spectrum (800-2500nm) and reduced spectra region (1300-1800nm). Generally, all prediction models developed for the physical and chemical properties required high number of principal component (PCs, 4-12).

Prediction models developed for aerated bulk density exhibited a high $R^2$ (0.80) for both full and reduced spectra-based models (Figure 2). The full spectra-based prediction model had a low RMSEP (0.0073) and SEP (0.0074). Slightly lower RMSEP (0.0069) and SEP (0.0070) were exhibited by the reduced spectra-based prediction model of aerated bulk density. Prediction models of aerated bulk density developed using the full and reduced spectra regions exhibited low bias values of -0.0008 and 0.0012 respectively. Both models required 9 PC's for model development. Both models will be adequate for screening and approximate calibration activities based on the $R^2$ (0.80) values exhibited.[15]

Tapped bulk density prediction models exhibited high $R^2$ (0.80) for the full spectra and reduced spectra-based prediction models. The prediction models developed with the full spectra region exhibited low RMSEP (0.0104), SEP (0.0103) and bias (-0.0026) value close to zero. Similarly, low RMSEP (0.0025), SEP (0.0099) and bias (-0.0035) were observed for the tapped bulk density prediction model developed with the reduced spectra region. Both prediction models of tapped bulk density required relatively high number of PC's (10-12) for model development. Adedipe and Dawson-Andoh (2008)[11] reported bulk density prediction model for yellow-poplar veneer with good $R^2$ (0.79). Even though these two studies used different wood species, different response variable measuring methods and probably different populations; it is a clear indication of the ability of NIR coupled with MVDA to predict bulk density of a biomass. Based on $R^2$ (0.80) values exhibited by the models, both prediction models were judged adequate for screening and adequate calibration purpose.[15]

For ash content, the prediction models developed for both each spectra regions were poor and exhibited very low $R^2$ (< 0.35). Additionally, they also exhibited low bias (< 0.015). This may be attributed to the low ash content values obtained and the subtle variance between the measured samples used in this study. Poke *et al.* (2004), similarly obtained moderate $R^2$ (0.62) for the calibration models for acid-soluble lignin of *Eucalyptus globlus* using NIR and attributed

the observation to low quantitative values of measured acid-soluble content of the species. The poor ash content model obtained in this study may be improved by increasing the number of samples in the study, which will increase the variance within the population and make it more profound and detectable by PLS modeling technique. The number of PC's required to develop both prediction models had a range within (4-10). Both models were poor and unusable based on the low $R^2$ values they exhibited.[15]

Extractives content of northern red oak was predicted with a very high $R^2$ (> 0.92) for both models developed with the raw full and the reduced spectra. The reduced spectra-based prediction model exhibited slightly lower RMSEP (0.2602) and SEP (0.2674) values as compared to the RMSEP (0.2660) and SEP (0.2717) of the model developed with the full spectra. Both prediction models required a higher number of PC's (10-11) for model development and exhibited lower bias values (0.0326-0.0126). Based on the high $R^2$ exhibited by both models they were adequate for most prediction purposes where accuracy is vital; including quality assurance activities.[15]

Prediction of acid-insoluble lignin exhibited a slightly low $R^2$ values of (0.56) and (0.58) for the full and reduced spectra-based prediction models respectively. These models even though had slightly low $R^2$ values, were suitable for rough screening purposes.[15] The prediction model developed with the full spectra of the raw data exhibited low RMSEP (0.0674) and SEP (0.0696). Similarly, low RMSEP (0.0793) and SEP (0.0817) were exhibited by acid-insoluble lignin prediction model developed with the reduced spectra region. Both prediction models required slightly higher number of PC's (5-9). Low bias values were exhibited by prediction models developed with the full spectra (0.0020) and the reduce spectra (0.0039) regions.

For total lignin, the prediction model developed with the full spectra region exhibited low $R^2$ (0.40). $R^2$ (0.56) was slightly improved (0.56) when total lignin was predicted with calibration model developed with the reduced spectra region. The reduced spectra-based prediction model exhibited slightly lower RMSEP (0.0654) and SEP (0.0669) values as compared to the RMSEP (0.0766) and SEP (0.0789) values exhibited by the full spectra-based model. The prediction models developed with full and reduced spectra regions also exhibited low bias values of -0.0028 and 0.0083 respectively. Higher numbers of PC's (7-12) were required to develop both models. The full spectra-based prediction model is unusable based on the low $R^2$ (0.4), whiles the reduced spectra-based model with an $R^2$ of 0.56 will be adequate for rough screening purposes.[15]

Holocellulose prediction model developed with the reduced spectra demonstrated slightly reduced RMSEP (0.0114) and SEP (0.0117) as compared to RMSEP (0.0137) and SEP (0.0141) for the full spectra-based model. This may be due to the slightly high number of PC's (9) and $R^2$ (0.89) associated with the reduced spectra-based model. The full spectra-based prediction model also exhibited high $R^2$ (0.89) and required high number PC's (8) for model developement. Models of both spectra regions had low bias (0.0009). Both models will be adequate for prediction purposes where accuracy is required including research activities. This observation was based on the range of $R^2$ (0.83-0.89) exhibited by both models developed.[15]

With the exception of prediction models developed for ash content, which were both poor, the reduced spectra-based prediction models of all response variables considered in this study exhibited slightly better prediction accuracy over the full spectra-based prediction models. This is based on improved $R^2$ and a decrease in RMSEP/SEP of the reduced spectra-based model as compared to the full spectra-based models. These results are shown in columns 6-8 of Table 2.

## Evaluation of chemical and physical properties prediction using first derivative preprocessed spectra

Prediction statistic of PLS1 models developed with the first derivate preprocessed data and their optimum number of PCs are shown in Table 3. Generally, preprocessing of the raw spectra using the first derivative method improved the PLS models of all the chemical and physical properties. This observation was based on a decrease in the number of optimum PCs and an increase in $R^2$ or almost equal $R^2$ values with respect to similar models developed with the raw spectra data (Table2).

For aerated bulk density, prediction models developed with the full spectra and reduced spectra exhibited high $R^2$ (0.84) values. Both models had low and equal RMSEP (0.0066) and SEP (0.0068). Prediction models developed with the full spectra and reduced spectra exhibited low bias values of 0.0005 and -0.0004 respectively. Both prediction models required lower number of PC's (6) for model development as compared to similar models (9 PC's) developed with the raw spectra data. Based on the high $R^2$ (0.84) exhibited by both models, they were judged adequate for most prediction purposes where accuracy is important; this includes research activities.[15]

High $R^2$ values were exhibited by prediction models of tapped bulk density developed with the full ($R^2$ 0.89) and reduced spectra ($R^2$ 0.88). The prediction model developed with the full spectra exhibited low RMSEP (0.0078) and SEP (0.0072) values. Similarly, the reduced spectra-based model also predicted tapped bulk density with a slightly higher RMSEP (0.0084) and SEP (0.0085) values as compared to the model developed with the full spectra region. Prediction models developed with the full and reduced spectra exhibited low bias values of -0.0035 and -0.0015 respectively. The number of PC's required to develop both models were much lower (5-6) than similar models developed with the raw spectra data (10-12 PC's). Both models are adequate for most applications including research purposes where accuracy is vital (Williams 2004). This observation was based on $R^2$ values (0.88-0.89) exhibited by both models.

For ash content, even though all the parameters used to evaluate models performance improved in the first derivate preprocessed spectra models in comparison with models developed with the raw spectra, the ash content prediction models developed with both spectra regions were poor and unusable based on low $R^2$ values (< 0.4). Similar results were obtained when the raw spectra data was used to predict ash content of northern red oak. The poor prediction ability of ash content model was attributed to its low content in northern red oak and hence the subtle variance between the measured samples. This can be remedied by increasing the sample population. Lower number of optimum PC's (4-5) was used to develop both models.

For extractives content, there was a strong correlation between the measured and NIR predicted values, with the full and reduced spectra-based prediction models exhibiting $R^2$ value of 0.93 and 0.94 respectively. Prediction models developed with the full spectra exhibited low RMSEP (0.2550) and SEP (0.2519) values. Likewise, the reduced spectra-based model also exhibited low RMSEP (0.2328) and SEP (0.2395), which were slightly lower than the full spectra-based prediction model. The number of PC's (6) required to develop both models were lower than similar models (10-11 PC's) developed with the raw spectra data. Extractives content prediction models developed with the full and reduced spectra exhibited low bias value of 0.0729 and 0.0059 respectively. Both models will be adequate for quality assurance activities. This observation was based on the high $R^2$ (0.03-0.94) exhibited by both models. [15]

The preprocessed full spectra region and preprocessed reduced spectra region was used to predict acid-insoluble lignin content with a much improved $R^2$ values (>0.7) as compared to low $R^2$ (<0.6) values of similar models developed with the raw spectra data. The full and reduced

spectra-based prediction models exhibited $R^2$ values of 0.73 and 0.79 respectively. Prediction models developed with the full spectra exhibited low RMSEP (0.0631) and SEP (0.0647) values. Slightly lower RMSEP (0.0567) and SEP (0.0583) were exhibited by prediction models developed with reduced spectra; this may be due to an increase in number of PC's (5) and $R^2$ values in the latter model. Models developed with the full and the reduced spectra predicted acid-insoluble lignin with low bias values of 0.0057 and 0.0047 respectively. Both prediction models were developed with few PC's (4-5) and based on their $R^2$ values both models were judged to be adequate for screening and approximate calibration purposes.[15]

Likewise, total lignin was predicted with a much improved $R^2$ (0.70) for the model developed with the full spectra-based model and an $R^2$ of (0.72) for reduced spectra-based models. Similar models developed with the raw spectra data had a much lower $R^2$ (<0.6) for both spectra regions. Prediction models developed with the full spectra region exhibited low RMSEP (0.0531) and SEP (0.0477). Similarly, the reduced spectra-based model predicted extractives content with a low RMSEP (0.0487) and SEP (0.0485). The full and reduced spectra-based prediction models exhibited low bias values of -0.0262 and -0.0127 respectively. Both prediction models required low number of PC's (5-6) for model development. With a moderate $R^2$ range exhibited by the two prediction models, both models were deemed adequate for approximate calibration and screening purposes.[15]

Holocellulose content was predicted with high $R^2$ (0.85-0.89) for both models developed with the full and reduced spectra regions (Figure3). The full spectra-based prediction models exhibited low RMSEP (0.0124) and SEP (0.0124) values. Prediction models developed with the reduced spectra region exhibited a slightly higher RMSEP (0.0129) and SEP (0.0133) in contrast to full spectra-based prediction models. Prediction models developed with the full and reduced spectra regions showed low bias values of -0.0028 and -0.0006 respectively. The number of PC's necessary to develop both prediction models was within the range of 6-7. Both models were deemed good for most prediction applications where accuracy is vital, including research activities.

Comparatively, the full spectra-based prediction models of tapped bulk density, extractives content and holocellulose content exhibited slightly higher prediction accuracy over the reduced spectra-based prediction models of each response variable. Prediction models developed for acid- insoluble lignin and total lignin exhibited the reverse of this pattern. Aerated

bulk density was predicted with the same accuracy using the full NIR and the reduced spectra regions of the preprocessed spectra.


## Conclusion:

NIRS coupled with MVDA can be used to predict the physical and chemical properties of northern red oak with a high degree of accuracy. However, ash content was poorly predicted with both the raw and preprocessed spectra. The raw spectra data could be used to predict aerated bulk density, tapped bulk density, extractive content, and holocellulose with high $R^2$ (> 0.8). Acid-insoluble lignin and total lignin was predicted moderately with $R^2$ (< 0.6). First derivative preprocessing of spectra data greatly improved the prediction models developed for all the physical and chemical properties of northern red oak considered in this study. These observations were based on a significant decrease in PC's required to develop prediction models and a general increase in $R^2$ values of the models with respect to similar models developed with the raw spectra data in this study.

Prediction models developed with the reduced spectra region (1300-1800nm), predicted both chemical and physical properties of northern red oak with comparable degree of accuracy as the full spectra region (800-2500nm). This result is significant because cheaper, rapid and more portable spectrometers with a much reduced spectra range can be used to predict the chemical and physical properties of northern red oak, a potential biomass feedstock, with a good level of accuracy. No attempt was made in this study to attribute wavelength regions to predicted response variables considered in this study.

**Table 1. Statistics of measured chemical and physical properties of northern red oak (60 samples)**

|  | Mean* | SD[a] | Min[b] | Max[c] |
|---|---|---|---|---|
| Aerated Bulk Density (g/cm$^3$) | 0.25 | 0.015 | 0.23 | 0.28 |
| Tapped Bulk Density (g/cm$^3$) | 0.35 | 0.023 | 0.31 | 0.38 |
| Ash (%) | 0.18 | 0.013 | 0.15 | 0.20 |
| Extractives (%) | 6.25 | 0.997 | 4.80 | 7.74 |
| Acid Insoluble Lignin (%) | 20.54 | 1.162 | 17.90 | 21.95 |
| Soluble Lignin (%) | 0.034 | 0.015 | 0.003 | 0.063 |
| Total Lignin (%) | 20.60 | 1.028 | 18.00 | 21.97 |
| Holocellulose (%) | 74.8 | 1.45 | 71.48 | 74.93 |

Mean, [a]Standard deviation, [b]Minimum measured value, [c]Maximum maximum measured value

**Table 2. Results of PLS1 calibration and prediction models developed for the chemical and physical properties of northern red oak using the raw spectra at two NIR wavelength regions**

| | Wavelength (nm) | [a]PC | [b]$R^2$cal | [c]RMSEC | [d]SEC | [e]$R^2$val | [f]RMSEP | [g]SEP | BIAS |
|---|---|---|---|---|---|---|---|---|---|
| **Bulk Density Aerated (g/cm$^3$)** | | | | | | | | | |
| | 800-2500 (A) | 9 | 0.94 | 0.0034 | 0.0034 | 0.80 | 0.0073 | 0.0074 | -0.0008 |
| | 1300-1800 ( C ) | 9 | 0.93 | 0.0039 | 0.0040 | 0.80 | 0.0069 | 0.0070 | 0.0012 |
| **BulkDensity Tapped (g/cm$^3$)** | | | | | | | | | |
| | 800-2500 (A) | 10 | 0.98 | 0.0031 | 0.0031 | 0.80 | 0.0104 | 0.0103 | -0.0026 |
| | 1300-1800 ( C ) | 12 | 0.98 | 0.0035 | 0.0035 | 0.80 | 0.0025 | 0.0099 | -0.0035 |
| **Ash (%)** | | | | | | | | | |
| | 800-2500 (A) | 10 | 0.95 | 0.0030 | 0.0030 | 0.33 | 0.0108 | 0.0108 | 0.0026 |
| | 1300-1800 ( C ) | 4 | 0.59 | 0.0086 | 0.0087 | 0.17 | 0.0140 | 0.0118 | 0.0081 |
| **Extractives (%)** | | | | | | | | | |
| | 800-2500 (A) | 10 | 0.98 | 0.1241 | 0.1256 | 0.93 | 0.2660 | 0.2717 | 0.0326 |
| | 1300-1800 ( C ) | 11 | 0.98 | 0.1524 | 0.1543 | 0.94 | 0.2602 | 0.2674 | 0.0126 |
| **Acid Insoluble Lignin (%)** | | | | | | | | | |
| | 800-2500 (A) | 9 | 0.91 | 0.0415 | 0.0420 | 0.56 | 0.0674 | 0.0696 | 0.0020 |
| | 1300-1800 ( C ) | 5 | 0.64 | 0.0820 | 0.0831 | 0.58 | 0.0793 | 0.0817 | 0.0039 |
| **Total Lignin (%)** | | | | | | | | | |
| | 800-2500 (A) | 7 | 0.89 | 0.0395 | 0.0400 | 0.40 | 0.0766 | 0.0789 | -0.0028 |
| | 1300-1800 ( C ) | 12 | 0.97 | 0.0201 | 0.0203 | 0.56 | 0.0654 | 0.0669 | 0.0083 |
| **Holocellulose (%)** | | | | | | | | | |
| | 800-2500 (A) | 8 | 0.81 | 0.0143 | 0.0145 | 0.83 | 0.0137 | 0.0141 | 0.0009 |
| | 1300-1800 ( C ) | 9 | 0.81 | 0.0136 | 0.0138 | 0.89 | 0.0114 | 0.0117 | 0.0009 |

[a]No of principal components, [b]Calibration $R^2$, [c]Root mean square of calibration, [d]Standard error of calibration, [e]Validation $R^2$, [f]Root mean square of validation, [g]Standard error of prediction

**Table 3. Results of PLS1 calibration and prediction models developed for the chemical and physical properties of northern red oak using the first derivative preprocessed spectra at two NIR wavelength regions**

| | Wavelength (nm) | [a]PC | [b]$R^2$cal | [c]RMSEC | [d]SEC | [e]$R^2$val | [f]RMSEP | [g]SEP | BIAS |
|---|---|---|---|---|---|---|---|---|---|
| Bulk Density Aerated (g/cm$^3$) | | | | | | | | | |
| | 800-2500 (A) | 6 | 0.96 | 0.0030 | 0.0031 | 0.84 | 0.0066 | 0.0068 | 0.0005 |
| | 1300-1800 ( C ) | 6 | 0.91 | 0.0044 | 0.0045 | 0.84 | 0.0066 | 0.0068 | -0.0004 |
| BulkDensity Tapped (g/cm$^3$) | | | | | | | | | |
| | 800-2500 (A) | 5 | 0.94 | 0.0055 | 0.0055 | 0.89 | 0.0078 | 0.0072 | -0.0035 |
| | 1300-1800 ( C ) | 6 | 0.94 | 0.0057 | 0.0057 | 0.88 | 0.0084 | 0.0085 | -0.0015 |
| Ash (%) | | | | | | | | | |
| | 800-2500 (A) | 4 | 0.62 | 0.0079 | 0.0080 | 0.36 | 0.0119 | 0.0098 | 0.0072 |
| | 1300-1800 ( C ) | 5 | 0.66 | 0.0076 | 0.0077 | 0.23 | 0.0159 | 0.0140 | 0.0082 |
| Extractives (%) | | | | | | | | | |
| | 800-2500 (A) | 6 | 0.98 | 0.1498 | 0.1517 | 0.93 | 0.2550 | 0.2519 | 0.0729 |
| | 1300-1800 ( C ) | 6 | 0.94 | 0.2267 | 0.2296 | 0.94 | 0.2328 | 0.2395 | 0.0059 |
| Acid Insoluble Lignin (%) | | | | | | | | | |
| | 800-2500 (A) | 4 | 0.82 | 0.0569 | 0.0576 | 0.73 | 0.0631 | 0.0647 | 0.0057 |
| | 1300-1800 ( C ) | 5 | 0.73 | 0.0693 | 0.0702 | 0.79 | 0.0567 | 0.0583 | 0.0047 |
| Total Lignin (%) | | | | | | | | | |
| | 800-2500 (A) | 5 | 0.92 | 0.0323 | 0.0327 | 0.70 | 0.0531 | 0.0477 | -0.0262 |
| | 1300-1800 ( C ) | 6 | 0.91 | 0.0344 | 0.0348 | 0.72 | 0.0487 | 0.0485 | -0.0127 |
| Holocellulose (%) | | | | | | | | | |
| | 800-2500 (A) | 7 | 0.99 | 0.0036 | 0.0037 | 0.89 | 0.0124 | 0.0124 | -0.0028 |
| | 1300-1800 ( C ) | 6 | 0.83 | 0.0129 | 0.0131 | 0.85 | 0.0129 | 0.0133 | -0.0006 |

[a]No of principal components, [b]Calibration $R^2$, [c]Root mean square of calibration, [d]Standard error of calibration, [e]Validation $R^2$, [f]Root mean square of validation, [g]Standard error of prediction
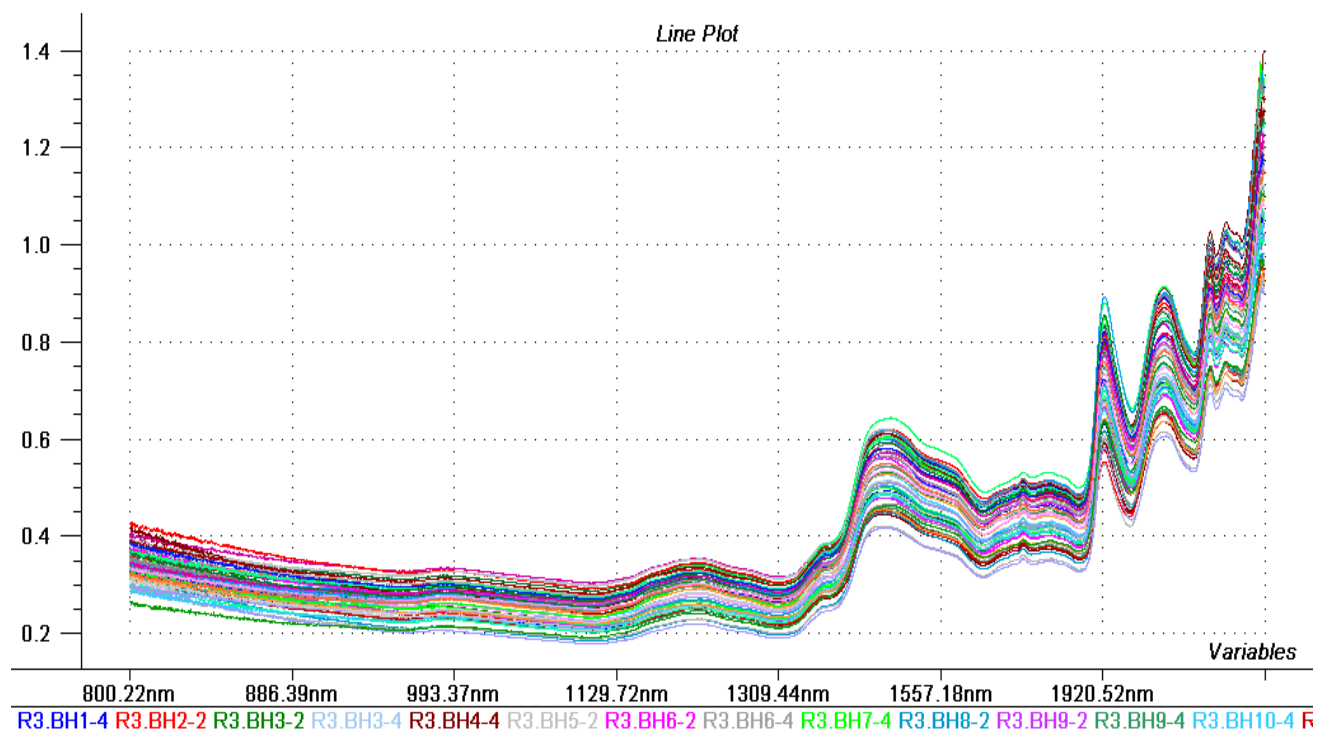
**Figure 1a: Representative spectra of northern red oak**
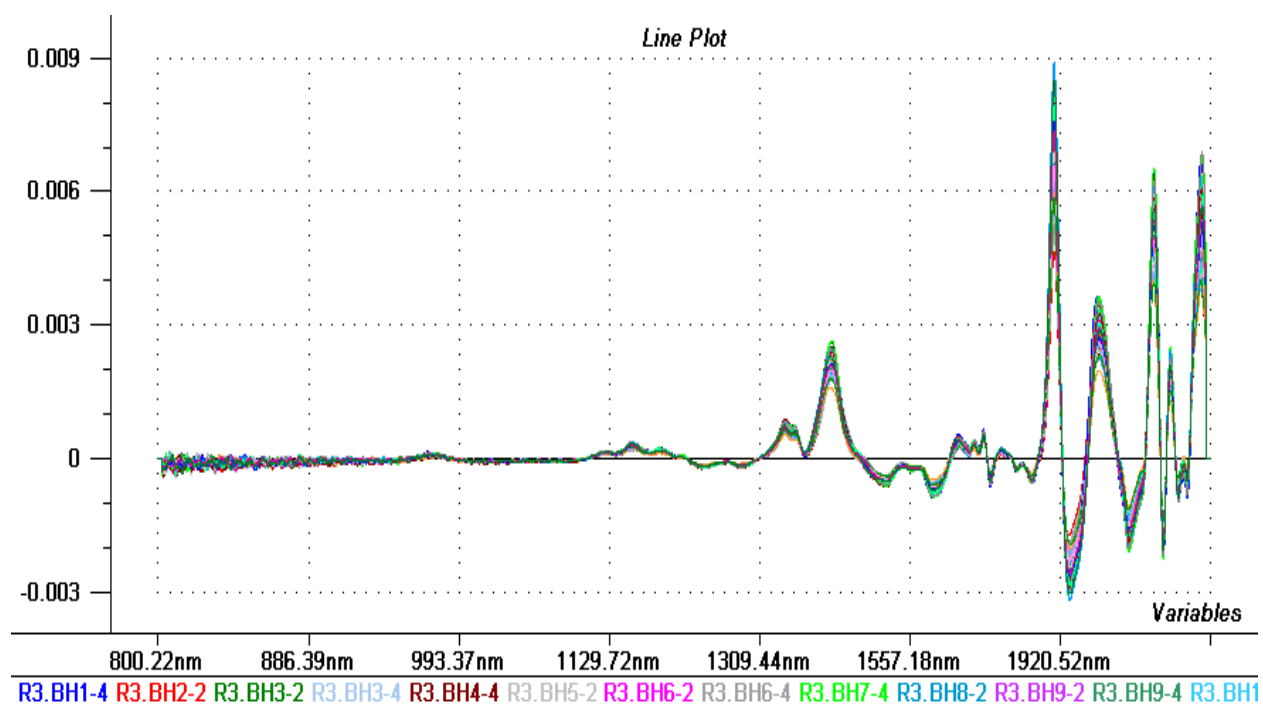


**Figure 1b. First derivative preprocessed spectra of northern red oak**
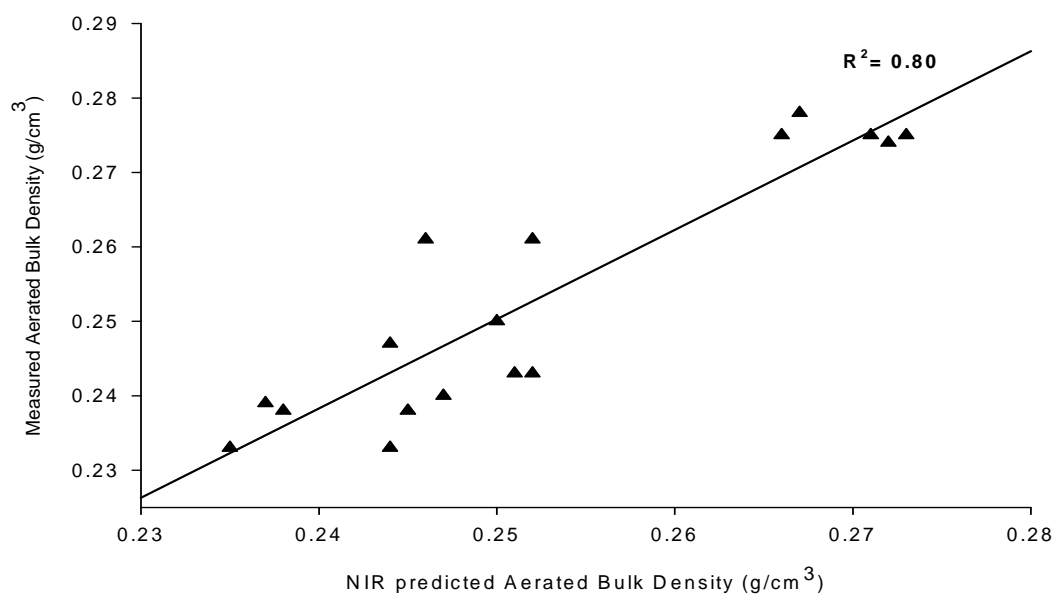
**Figure 2. PLS1 plot of predicted against measured Aerated Bulk Density using raw full NIR spectra region ( 800-2500nm)**
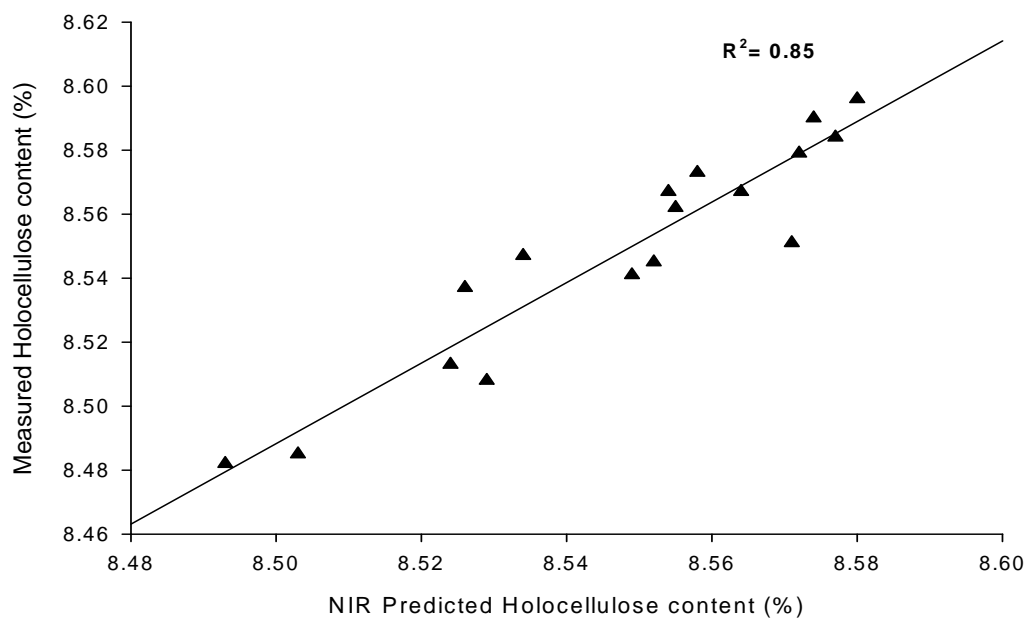


**Figure 3. PLS1 plot of predicted against measured Holocellulose content using preprocessed NIR spectra region of 1300-1800nm**

# References

1. Putche V, D. Sandor. 1996. Strategic, economic, and environmental issues for transportation fuels. In Handbook on Bioethanol: production and utilization. Editors: C. E. Wyman. C. E. Taylor and Francis, Washington, D.C. 21-35

2. The Kyoto Protocol to the Framework Convention on Climate Change. 1997. (http://unfcc.int/essential_background/Kyoto_protocol/background/items/1351.php).

3. Perlack R.D., Wright, L.L., Turnhollow, A.F., Graham, R.L., B.J. Stokes, and D.C. Erbach. 2007. Biomass as feed stock for a bioenergy and bioproducts industry: The technical feasibility of a billion-ton annual supply US Department of Energy. DOE/GO-102995-2135,( http://www.ostl.gov/bridge)

4. Labbe, N.; Lee, Seung-Hawn.; Cho, Hyun-woo.; Jeong, M.K.; Andre, N. Enhanced discrimination and calibration of biomass NIR spectra data using non-linear kernel methods. Bioresource Technology. 2008, 99, 8445-8452

5. Nkansah, K.; Dawson-Andoh, B.; Slahor, J. Rapid Characterization of Biomass: Part1. Near infrared spectroscopy characterization of the physical and chemical properties of yellow-poplar (*Liriodendron tulipifera* L). 2009: Unpublished.

6. So, C.; Groom, L.; Rails, T.; Snell, R.; Kelley, S and Meglen, R. Rapid assessment of the fundamental property variation of wood. Gen. Tech. Rep. SRS–48. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. pp 176-180.

7. Zobel BJ, Van Buijtenen JP. 1989 Wood variation - its causes and control. Springer Verlag: Berlin, etc. 363 pp. DM 278.00. 581.4241. Review by P.Baas in: IAWA Bull. (N.S.) 10 (1989) 439-40

8. Lam P.S., S. Sokhansanj, X. Bi, S. Mani, C.J. Lim, A.R. Womac, M. Hoque, J. Peng, T. JayaShankar, L.J. Naimi and S. Nayaran. 2007. Physical charatcterization of wet and dry wheat straw and switchgrass – bulk and specific. density 2007 ASABE Annual International Meeting. Sponsored by ASABE, Minneapolis Convention Center, MN, 17-20 June 2007)

9. Yu A.B.,and J.S. Hall. 1994. Packing of fine powders subjected to tapping. Powder Technology 78: 247–256.

10. Adullah E.C. and D. Geldart. 1999. The use of bulk density measurements as flowability indicators". Powdder technology 102:151-165.

11. Adedipe O.E. and B. Dawson-Andoh. 2008 Prediction of yellow-poplar (Liriodendron tulipifera) veneer stiffness and bulk density using near infrared spectroscopy and multivariate calibration, J. Near Infrared Spectrosc. 16: 487-496

12. Sluiter, R. Snell, S. Kelley, T. Rial, and l. Groom, "Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood. Wood Sci Technol (2004) 38: 257-276

13. Esbensen K.H. 2002.Multivariate data analysis in practice: An introduction to multivariate data analysis and experimental design, 5th edition. CAMO Process AS, Oslo, Norway:598pp

14. Eriksson L., E. Johansson, N. Kettaneh-Wold and S. Wold. 1999. Introduction to multi- and megavariate data analysis using projection methods(PCA  PLS). Umetrics AB, Mea, Sweden: 490pp

15. 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.

16. Pettersen, C. The chemical composition of wood. In: Rowell, R.M., ed. The chemistry of solid wood. Advances in chemistry series. 207 Washington D.C: American Chemical Society; 1984: chapter 2.

17. Poke F., J. Wright and C. Raymond. 2004. Predicting extractive and lignin contents in Eucalyptus globules using near infrared reflectance analysis.  J. Wood Chem. Technol. 24: 55-67

# CHAPTER FOUR:  Rapid characterization of biomass using fluorescence spectroscopy coupled with multivariate data analysis: Part 1.  Yellow-poplar (*Liriodendron tulipifera* L)

## Abstract

In this paper, fluorescence spectroscopy coupled with multivariate data analysis (MVDA) was used as a process analytical technology (PAT) tool to predict some physical and chemical properties of yellow-poplar (*Liriodendron tulipifera* L.) as a biomass feedstock relevant to the bio-chemical conversion of the material into bio-fuels and bio-products. With the exception of holocellulose content, all the properties considered in the study were predicted with moderate to strong correlation between measured and fluorescence predicted values. Fluorescence spectra-based prediction models of each property considered in this study were compared to Near Infrared (NIR) spectra-based prediction models of the same properties from a previous study using the same population. The NIR-based prediction models exhibited slightly superior model strength over the fluorescence spectra-based prediction models of similar properties.

**Keywords:** Process analytical technology, multivariate data analysis, fluorescence, near infrared

## Introduction

The emergence of cellulosic bio-fuel as an alternate clean fuel to displace fossil-based transportation fuel has gained momentum for the last couple of years. In the United States it is currently estimated that about 3% of our total industrial energy consumption is fulfilled by bio-energy derived from sustainable and renewable biomass.[1] It is projected that available renewable biomass in the U.S. can be used to displace about 30% of current fossil fuel needs.[1]

The potential of using biomass as feedstock for bio-refineries depends on linking both their physical and chemical properties to the bio-refinery process parameters. These properties include bulk density, particle size, moisture content, calorific values, ash content, lignin content, extractive content, and total sugars.[2] Both qualitative and quantitative information on these properties needs to be known at the beginning of the bio-refinery process. We should also be able to track these properties and that of intermediary to final products of the process for optimum process control and quality assurance. Currently, these properties are monitored by standard wet chemistry analysis which is slow, destructive, expensive, and can only be carried out primarily off-line. Consequently, rapid characterization of these properties before and during bio-refinery is critical to the success of the process. Standard wet chemistry methods of quantification are cumbersome, slow and expensive.

Recent advances in spectroscopic technologies such as near infrared (NIR), UV-VIS, fluorescence, Raman and Laser induced breakdown spectroscopy (LIBS) coupled with "data mining" methods such as multivariate data analysis (MVDA) have created new opportunities for industries such as the pharmaceutical to implement rapid on-/at/in-line analytical quantification and qualification methods (Process Analytical Technology, PAT) for monitoring manufacturing processes and quality assurance. Process analytical technology (PAT) is broadly defined as the application of analytical chemistry coupled with multivariate data (MVDA) to monitor the physical and chemical properties of a material through on-/at-/in-line for process understanding, control or optimization.[3] The "data mining" tool, MVDA, facilitates the extraction of important and relevant information in the large multivariable data (spectra) generated by spectrometers. This can be used to develop classification and prediction models.

The biomass and forest products industry have successfully utilized NIR as a PAT tool for quantitative and qualitative process control.[4-6] Extensive research has been done on the use of NIR as a PAT tool in the biomass and forest products industry to quantify the physical and

chemical properties of a number of wood species.[7,8] Fluorescence spectroscopy (FS), a vibrational spectroscopic tool known to be more sensitive, cheaper, and more rapid in spectra acquisition has been understudied in the biomass and forest products industry. Wood is known to fluoresce due to the presence of flourophores in its chemical structure.[9] The source of fluorescence is lignocellulose materials has been has been attributed to the presence of fluorophore molecules present in some of its chemical components; notably lignin, cellulose, hemicelluloses and extractives.[9-12] Some chemical properties of *Eucalyptus* wood species were successfully quantified by fluorescence coupled with chemometrics.[13] Most past fluorescence spectroscopic studies of wood were done on ground wood specimens. Sum *et al.* (1991), successfully used fluorescence spectra of solid wood to identify wood species on the basis of the characteristic spectra exhibited by different species.[14]

To the best of our literature review, limited studies of fluorescence spectra coupled with MVDA have been evaluated as a potential PAT tool for predicting physical and chemical properties of solid wood. This study therefore bridges the knowledge gap by evaluating the potential of using FS to predict comprehensive physical and chemical properties of solid biomass and also evaluate the relative performance of NIR and fluorescence to predict the physical and chemical properties of wood using the same population of wood samples. The specific objectives of this study were:

1. Use Fluorescence spectroscopy coupled with MVDA to develop calibration and prediction models for important physical and chemical properties of northern red oak (*Quercus rubra*) and yellow-poplar (*Liriodendron tulipifera* L.). These properties include ash content, extractives content, total lignin, acid insoluble lignin, holocellulose, and bulk density.

2. Compare the efficiency of two spectra region of the fluorescence spectrum in order to determine which region contains relevant spectroscopic information for accurate quantification of the chemical and physical properties of yellow-poplar.

3. Compare the performance of fluorescence and NIR spectra-based PLS1 prediction models of each physical and chemical property of northern red oak and yellow-poplar.

The above study results are presented in two parts based on the two wood species, yellow-poplar and northern red oak. Part 1 of the study, which is based on yellow-poplar is presented here. Part 2, which reports the results of the study on northern red oak, is reported elsewhere.

## Materials and methods

Details of materials and methods used in this study are described in this paper. Part 2 of this study makes extensive references to this section.

## Materials

A tree, each of yellow-poplar and northern red oak was randomly selected in West Virginia University Research forest located about 13 km east of Morganton, WV. The height of each tree was approximately 14 meters. Five centimeter thick disks were cut at 3.5 meter intervals along the bole of the tree starting from the breast height (1.1m). A total of three disks were cut from each tree. Each disk was cut into blocks of size 19 X 19 X 50 mm. Only blocks from the heartwood of each disk were used in this study. Twenty specimens were randomly selected from each disk of both northern red oak and yellow-poplar disks. A total of 60 blocks (specimens) of both species were used in this study. Specimens were oven-dried at $103 \pm 2$ $^{o}$C for 24 hours and reconditioned at about $23^{o}$C in a vacuum desiccator for 24 hours before analysis.

## Methods

**Fluorescence spectra measurement**

Fluorescence spectra measurements were acquired using USB2000-FLG fluorescence spectrometer (Ocean Optics Inc., Dunedin, FL, USA) with an application wavelength range of 380-1050nm. The spectrometer was fitted with a customized fiber optic probe for measuring the fluorescence of solid biomass materials. A UV light source (PX-2 Pulsed Xenon Lamp) of range 200-750 nm was used to excite the samples. The probe had an incident leg to direct light source and collection leg for collection of emitted light (fluorescence emission) after incidence on the material connected to the spectrometer. A linear variable filter (Ocean Optics Inc., Dunedin, FL USA) was fixed between the broad UV light source and the tip of the probe that directs light onto the material. The purpose of this filter was to select a band wavelength region for excitation of

the material and block out other wavelength regions not selected. The software of the spectrometer (Spectra Suite, Ocean Optics Inc., Dunedin, FL, USA) allows excitation period adjustments. Excitation was done with thirty milliseconds of exposure of incident light and the resulting emission recorded. The spectra of wood samples were collected in scope mode and excited with a band wavelength region of 300-400nm. Ten scans were collected and averaged to a single spectrum; a total of four spectra were taken on the tangential face along the length of each sample and averaged to one average spectrum. These spectra were collected with the tip of the probe perpendicular to the tangential face of the sample. Spectra were collected using Spectra Suite software in GRAMS SPC format. The Unscramble software version 9.8 (Camo Software Inc., Woodbridge, NJ, USA) was used for MVDA of the spectra data

**Determination of chemical properties**

Acid-insoluble lignin, total lignin (insoluble and acid-soluble lignin), extractives, and ash contents were the chemical properties determined in this study. Specific chemical properties of yellow-poplar were determined by traditional wet chemistry analytical methods: ASTM D-1106-96, NREL LAP-004(1996), ASTM D1105-96(2007) and ASTM D-1102-84 respectively. Holocellulose content is given by the difference between initial weight of ground wood (oven-dried weight basis) and the sum of the above chemical components. This data together with their physical property; bulk density was used as the Y-response variable for PLS modeling.

**Bulk density measurement**

Only one physical property of wood, bulk density was measured. Bulk density of biomass is a major contributing factor to the determination of "cost and logistic requirements for handling and moving biomass from point of production to biorefinery".[15] It is a function of the size, shape and individual particle density of the lignocellulose biomass. Two types of bulk densities were determined; aerated bulk density and tapped bulk density in accordance with the method described by [Yu et *al* (1994) and Adullah *et al. (*1998)]. Volume and weight determinations for each method were replicated twice for each sample and the average was computed for the sample.

**Development of fluorescence calibration**

All spectra were analyzed using Unscrambler multivariate data analysis software (version 9.8, Camo Inc., Woodbridge, New Jersey, US). The four emission spectra collected from each sample were averaged to one spectrum to give a total of sixty averaged spectra for each wood species. Spectra data of each species were divided into calibration and prediction sets. The calibration set consisted of two spectra out of every three consecutive spectra (40 spectra) and the prediction set comprised of one spectrum out of every three consecutive spectra (20 spectra).

Chemometrically, fluorescence data can be analyzed in several ways: (i) two-way structure – sample **X** emission spectra wavelengths using bilinear chemometric methods (principal component analysis, partial least squares) and (ii) multi-way notable three-way (samples **X** excitation wavelengths **X** emission wavelengths). The two-way method principal component analysis is used to find the principal directions of variation in the fluorescence data.[25-26] However, the three-way analysis produces more robust prediction models.[27-29] PARFAC method is one method used for multi-way model development.[30] It decomposed three way fluorescence data into spectra excitation and emission profiles of fluorophores in samples. Two-way models are less robust in prediction. However, a two-way analysis will be used in this paper and a three-way analysis will be presented in a subsequent paper.

As a first step in the development of a calibration model, a Principal Component Analysis (PCA) was applied to the data to obtain an overview. This process reduces the dimensionality of data by a linear combination of original data to generate new latent variables which are orthogonal and uncorrelated to each other. It distills the information in the original variables into a lower number of variables – principal components (latent variables).[18] PCA lies along the direction of maximum variance in the data set.[19] The score plot and loading plots of the PCA were observed and this operation revealed a total of six (6) outliers that were removed from the yellow-poplar data and excluded from subsequent model development. As a result of these outlier samples, the number of samples used for calibration (37 spectra) and testing (17 spectra) of the model decreased. After removal of outliers, both X- and Y- data matrices were mean-centered. All fluorescence spectra (independent variables) were combined into a single data matrix [X-matrix] and the physical and chemical properties into separate response matrices [$Y_i$]. Calibration models were developed for two spectra region: (i) full spectra region that contained

both excitation and emission regions (295-872nm) and (ii) the longer wavelength region (400-810nm) which encompassed only the emission region of the spectrum.

Calibration models were developed using partial least squares regression (PLS1). PLS1 models were developed using only the raw spectra data. It should be noted that two other forms of spectra preprocessing of the data including first derivative and multiple scatter correction (MSC) were evaluated but did not significantly improve prediction models developed and are therefore not reported here. Since models developed with the raw data were better than similar models developed with preprocessed data, only the raw data-based calibration and prediction models are reported in this study.

Partial least squares calibration models were developed with X- and Y-matrices. The Y-matrices consisted of data from bulk densities (aerated bulk density and tapped bulk density), ash content, holocellulose content, total lignin, acid-insoluble lignin and extractives contents. The response variables, holocellulose, total lignin and acid-insoluble lignin contents were each preprocessed using a cubed root transformation due to the large differences in the values measured.[6] Calibration models were developed using full-cross validation method and then used to predict the response of the prediction set. The Unscrambler MVDA software (version.9.8, Camo Software, Woodbridge, NJ, USA) was allowed to determine the optimum number of principal components required with close supervision on the explained and residual variance plots of each model before it was accepted. This was done to prevent under or over fitting of model.

**Measure of calibration model performance**

The prediction performance of a PLS1 model can be evaluated using a number of parameters. These include root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP), standard error of prediction (SEP), standard error of calibration (SEC), coefficient of determination ($R^2$), bias, RDP and residual variance.[22] Since this study focused on future prediction ability of models developed, parameters used to evaluate prediction performance of models were $R^2$, RMSEP, SEP and bias. Calibration models results are also presented even though they not discussed.

RMSEP, a measure of the model fit, is a direct estimate of the prediction error in the Y-variables of the model. It is expressed as the average error expected from future prediction of Y variables using our model and also expressed in the measurement units of the Y variables used in developing the model. Bias is used to detect if there is a systematic difference between average values of the calibration set and the validation set. If no such difference exists, bias will be equal to zero. The closer the bias is to zero, the more accurate the model developed both in terms of model fit and predictability of the future sample set. Bias is computed as the average difference between predicted and measured Y-values or all samples in the validation set. SEP is a measure of precision of the models predicting ability corrected for the bias in the validation of the model.[4,20,21]

Coefficient of determination ($R^2$) is a measure of total variance between measured and predicted values that can be modeled by linear association. $R^2$ shows the proportion of X- data that can be explained by the variance in the Y-data. It varies between zero and one. The closer $R^2$ is to one, the stronger the correlation. $R^2$ can inflate and approach unity by increasing the complexity of the model. Therefore, a high $R^2$, even though a good indicator of a good model cannot be independently used as the ultimate measure of model performance. Interpretation of $R^2$ may be subjective. Generally, an $R^2$ between 0.92-0.96 represents a model that can be used in most applications including quality assurance. An $R^2$ of 0.83-0.90 range also represent a model that can be used in most applications including research activities with some level of caution. Approximate calibration and screening purpose activities can be done with an $R^2$ range of 0.66-0.81. General rough screening activities can utilize a model with an $R^2$ range between (0.5-0.64). $R^2$ ($< 0.5$) may be deemed poor and not useable in any Fluorescence calibration and the reasons should be investigated.[21,22] The evaluation of possible end use of prediction models in this study are based on this reference.

Result presented in Tables 2 and 3 are PLS1 models developed for the prediction of measured physical and chemical properties of yellow-poplar with raw fluorescence spectra data and First derivative preprocessed NIR spectra (full spectra-based models).

## Results and discussion

Chemical properties of yellow-poplar are reported on percentage basis (oven-dried weight) of original sample analyzed (Table 1). Ash content exhibited a range of 0.70-0.79; extractives content had a range of 3.88-6.1, acid-insoluble lignin ranged from 16.59-21.30; acid-soluble lignin content had a range of 0.011-0.067; total lignin content ranged from 21.84-27.49 and holocellulose showed a range of 72.51-78.16. Chemical properties measured compared favorably to those reported previously by Peterson (1984). For the physical property measured, aerated bulk density and tapped bulk density exhibited a range of 0.16-0.20 and 0.21-0.27 g/cm$^3$ respectively. Typical representative fluorescence spectra of yellow-poplar are depicted in Figure1.

## Evaluation of chemical and physical properties prediction using fluorescence spectra data

PLS1 prediction models were developed to predict ash, extractives, acid-insoluble lignin, total lignin, holocellulose contents, aerated and tapped bulk densities of yellow-poplar using fluorescence spectroscopy coupled with MVDA and the results are presented Table 1.

The prediction of aerated bulk density with calibration models developed with the full spectra, and the emission spectra exhibited good correlation between the fluorescence predicted and measured aerated bulk density with an R$^2$ of (0.77) and (0.72) respectively (Figure 2). Low RMSEP (0.0068) and SEP (0.0068) were exhibited by the full spectra-based prediction model. The emission spectra-based prediction model had a slightly higher RMSEP and SEP (0.0072) values as compared to a full spectra-based model. Both PLS1 prediction models required slightly high number of PC's (7-8) for model development. Very low bias values were exhibited by the full spectra-based prediction model (-0.0001) and the emission spectra-based prediction model (-0.0016). Both models were adequate for screening and approximate calibration purposes based on the moderate R$^2$ exhibited by the models (0.72-0.77).[22]

Tapped bulk density, the second type of bulk density measured, was also predicted with a good R$^2$ (0.75) for the full spectra-based and the emission spectra-based (R$^2$ 0.69) prediction models. The full spectra-based prediction model exhibited low RMSEP and SEP (0.0086). Similarly, the emission spectra-based tapped bulk density prediction model also exhibited low RMSEP and SEP (0.0095) values. Low bias values were also exhibited by the full spectra-based

(-0.0021) and the emission spectra-based (-0.0023) prediction models. Slightly higher number of PC's (8) were required by both prediction models development. PLS1 prediction models developed for aerated bulk density and tapped bulk density exhibited slightly better model when predicted with calibration models developed with the full spectra than models developed with the emission spectra. This observation was based on slightly improved $R^2$, low RMSEP, SEP and bias values. Both models will be adequate for screening and adequate approximation activities. This observation was based on $R^2$ (0.69-0.75) values exhibited by the models.[22]

PLS1 calibration models developed for ash content with the full spectra and emission spectra exhibited high correlation between measured and fluorescence-predicted ash content values with $R^2$ values of (0.86) and (0.76) respectively. The full spectra-based prediction model exhibited low RMSEP (0.0130), SEP (0.0131) and bias (0.0021) values. Low RMSEP (0.0167), SEP (0.0170) and bias (-0.0014) values were also exhibited by the prediction models developed with the emission spectra. The full spectra-based prediction model was slightly a better model as compared to the emission spectra-based model; this was based on slightly lower RMSEP, SEP values and an improved $R^2$ value of the full spectra-based model (columns 6-9 of Table 2). Both prediction models required slightly high number of PC's (8) for development. Based on the $R^2$ of the model, the full spectra-based prediction models will be adequate for most prediction activities including research $R^2$ (0.86), whiles the emission spectra-based model will be good for screening and approximate calibration purposes $R^2$ (0.76).[22] Billa *et al.* (2000), similarly predicted the ash content of eucalyptus wood based on the fluorescence spectra of ground wood with a Pearson correlation coefficient (R) of 0.7.

For extractives content, prediction models developed with the full spectra and the emission spectra exhibited good $R^2$ of 0.75 and 0.68 respectively (Figure 3). Full spectra-based prediction model exhibited low RMSEP (0.3439), SEP (0.3529) and bias (-0.0333) values. Similarly the emission spectra-based prediction model also exhibited slightly higher RMSEP (0.3924) and SEP (0.4043) as compared to the full spectra-based model. Low bias (0.0113) was also exhibited by the emission spectra-based prediction model. The full spectra-based prediction model was consequently judged a better prediction model over the emission spectra-based model based on improved $R^2$, RMSEP and SEP values. Both prediction models required slightly high number of PC's (8) for development. Both prediction models will be adequate for screening and some approximate calibration purposes based on the $R^2$ of the models (0.68-0.75).[22]

Moderate correlation was exhibited between measured acid-insoluble lignin content and fluorescence predicted values based upon full spectra ($R^2$ 0.65) and emission spectra ($R^2$ 0.59) prediction models. The full spectra-based prediction model exhibited low RMSEP (0.0812), SEP (0.0838) and bias (0.0033). The emission spectra-based prediction model exhibited a slightly higher RMSEP (0.0912) and SEP (0.0905) value as compared to the full spectra-based model. Bias of the reduced spectra model was also low (-0.0245). Full spectra-based prediction model for acid-insoluble lignin will perform much better than models developed based with emission spectra. This observation was based on the much improved $R^2$, RMSEP and SEP values of the full spectra-based model. Both prediction models required a moderate number of PC's (7) for development. Based on the $R^2$ (0.59-0.65) values exhibited by each model, both models will be adequate for rough screening activities.[22]

Total lignin, the sum of acid-insoluble, and soluble lignin contents were also predicted with moderate $R^2$ values of 0.62 and 0.57, for the full spectra-based and the emission spectra-based prediction models respectively. The prediction model developed with the full spectra exhibited low RMSEP (0.0889), SEP (0.0918) and bias (-0.0015). Prediction model developed with the emission spectra also exhibited slightly higher RMSEP (0.0914), SEP (0.0943) and bias (0.0026). Based on the slight improvement in $R^2$, RMSEP, SEP and bias values of the full spectra-based prediction model; it was judged a better model for predicting total lignin content with an $R^2$ of 0.62. Both models required slightly high number of PC's (8-9) for development. Additionally, they will be adequate for rough screening purposes based on their $R^2$ range (0.57-0.62).[22]

For holocellulose content, prediction models for both spectra region exhibited very low $R^2$ values with a range of (0.1-0.3). Both models developed exhibited low RMSEP and SEP values (<0.068). Similarly, low number of PC's (6) were required for the development of both models and a low bias values were exhibited by both models as well (0.0012-0.0097). These models were judged poor and unusable for any purpose based on their low $R^2$ values.[22] The low prediction ability of holocellulose models may be as a result of insufficient sample size used to calibrate and test the model or as a result of the inability of fluorescence spectroscopy (FS) coupled with MVDA to predict holocellulose content of yellow-poplar accurately.

## Comparison of fluorescence and near infrared spectra-based models

The performance of first derivative based-NIR prediction models for the identical response variables for the same yellow-poplar population published elsewhere (Nkansah and Dawson-Andoh, 2009) are compared. Since each spectra-based model was developed with the sample population, models developed by the two spectroscopic methods can be compared. To make comparison simple, only the full spectra-based prediction models with three parameters ($R^2$, SEP, and number of PC's) of the two spectra-based models will be used.

The NIR-based prediction model predicted aerated bulk density with greater accuracy than the fluorescence-based model. The NIR-based model exhibited slightly higher $R^2$ (0.83); slightly lower SEP (0.0055) and PC's (4) in contrast to fluorescence-based model [$R^2$ 0.77, SEP (0.0068) and PC's (8)]. Similarly, NIR-based tapped bulk density model [$R^2$ (0.79); slightly lower SEP (0.0067); number of PC's (4)] performed better than fluorescence-based model [lower $R^2$ (0.72), much higher SEP (0.0086) and higher number of PC's (8)].

Even though the fluorescence-based prediction model of ash content exhibited higher $R^2$ (0.86) and a lower SEP (0.0131) as compared to the slightly lower $R^2$ (0.82) and a minimal increase in SEP (0.0133) for the NIR-based model, the NIR-based spectra was still comparable in performance to fluorescence-based model. This is due to the high number of PC's (8) required by the fluorescence-based model as compared to the NIR-based model (PC's 5). High number of PC's may inflate $R^2$ values and cause over fitting of the model, therefore a lower number of PC's is preferred.

For extractives content, the performance of NIR-based prediction model [$R^2$ (0.78), lower SEP (0.3387) and number of PC's (4)] was comparable to fluorescence-based models [$R^2$ (0.75), an increase in SEP (0.3529) and number of PC's (4)]. Performance of NIR-based acid-insoluble lignin content prediction model [$R^2$ (0.79), lower SEP (0.0717) and number of PC's (5)] was slightly better than fluorescence-based model [$R^2$ (0.65), an increase in SEP (0.0838) and number of PC's (7)]. Total lignin was predicted by NIR-based model [$R^2$ (0.66), reduction SEP (0.0895) and number of PC's (4)] with a greater accuracy than the NIR-based prediction model [$R^2$ (0.62), much higher SEP (0.0918) and number of PC's (8)]. Fluorescence-based prediction model failed to predict holocellulose content [low $R^2$ (0.37), high SEP (0.0608) and number of PC's (6)]. In contrast, NIR-based model predicted holocellulose content with greater accuracy [high $R^2$ (0.81), low SEP (0.0429) and number of PC's (4)].

Generally, the performance of the NIR-based prediction models developed with the first derivative preprocessed spectra for all response variables were higher than similar fluorescence based models. The fluorescence based models could still be used for screening, approximate calibration and research purposes ($R^2$ range: 0.59-0.86). The raw spectra NIR-based spectra models were not considered in this study because the preprocessed models performed much better.[24] Additionally, with the exception of holocellulose content, fluorescence spectra-based prediction models performed equally, and in some situations better than the raw NIR spectra-based prediction (data not shown here). Raw spectra NIR-based models are presented elsewhere.[24] Finally, two-way model analysis of fluorescence data is less robust and exhibits low prediction performance. A three-way analysis using PARAFAC may improve prediction models developed in this study and will be presented in a subsequent paper.

## Conclusion

Results from this study have shown that with the exception of holocellulose content, a two-way MVDA model analysis of full spectra fluorescence data (emission and excitation together) can predict ($R^2$ 0.62-0.86) ash, extractives, acid-insoluble lignin, total lignin contents, aerated and tapped bulk density of yellow-poplar with moderate to high accuracy than the emission spectra-based models ($R^2$ 0.55-0.76 ). The performance of full fluorescence spectra-based models was generally lower than preprocessed full NIR spectra-based models for similar response variables. Fluorescence data analyzed using two-way model principal component and partial least squares has low prediction accuracy and less robust. A three-way principal component analysis of this data will be presented in a subsequent paper.

This study further demonstrates the future potential of FS coupled with MVDA as a potential PAT tool. It has demonstrated the potential of using fluorescence spectrometer as an inexpensive, rapid, and portable spectrometer for characterizing some chemical properties of solid wood and the bulk density of ground wood based on fluorescence spectra of the solid wood block with fairly good results.

**Table 1. Statistics of measured chemical and physical properties of yellow-poplar (60 samples)**

|  | Mean* | SD[a] | Min[b] | Max[c] |
|---|---|---|---|---|
| Aerated Bulk Density (g/cm$^3$) | 0.18 | 0.014 | 0.16 | 0.20 |
| Tapped Bulk Density (g/cm$^3$) | 0.24 | 0.017 | 0.210 | 0.266 |
| Ash (%) | 0.76 | 0.036 | 0.7 | 0.79 |
| Extractives (%) | 5.03 | 0.720 | 3.88 | 6.1 |
| Acid Insoluble Lignin (%) | 19.39 | 1.29 | 16.59 | 21.3 |
| Soluble Lignin (%) | 0.03 | 0.012 | 0.011 | 0.067 |
| Total Lignin (%) | 19.40 | 1.27 | 21.84 | 27.49 |
| Holocellulose (%) | 73.00 | 0.75 | 72.51 | 78.16 |

*Mean, [a]Standard deviation, [b]Minimum measured value, [c]Maximum maximum measured value

**Table 2. Results of PLS1 calibration and prediction models developed for the chemical and physical properties of yellow-poplar using the raw fluorescence spectra at two wavelength regions**

| | Wavelength (nm) | [a]PC | [b]$R^2$cal | [c]RMSEC | [d]SEC | [e]$R^2$val | [f]RMSEP | [g]SEP | BIAS |
|---|---|---|---|---|---|---|---|---|---|
| **Bulk Density Aerated (g/cm$^3$)** | | | | | | | | | |
| | 295-872(A) | 8 | 0.88 | 0.0046 | 0.0047 | 0.77 | 0.0066 | 0.0068 | -0.0001 |
| | 400-810 (B) | 7 | 0.85 | 0.0052 | 0.0052 | 0.72 | 0.0072 | 0.0072 | -0.0016 |
| **Bulk Density Tapped (g/cm$^3$)** | | | | | | | | | |
| | 295-872(A) | 8 | 0.83 | 0.0072 | 0.0073 | 0.75 | 0.0086 | 0.0086 | -0.0021 |
| | 400-810 (B) | 8 | 0.89 | 0.0058 | 0.0059 | 0.69 | 0.0095 | 0.0095 | -0.0023 |
| **Ash (%)** | | | | | | | | | |
| | 295-872(A) | 8 | 0.94 | 0.0088 | 0.0089 | 0.86 | 0.0130 | 0.0131 | 0.0021 |
| | 400-810 (B) | 8 | 0.93 | 0.0091 | 0.0092 | 0.76 | 0.0167 | 0.0170 | -0.0014 |
| **Extractives (%)** | | | | | | | | | |
| | 295-872(A) | 8 | 0.89 | 0.2287 | 0.2319 | 0.75 | 0.3439 | 0.3529 | -0.0333 |
| | 400-810 (B) | 8 | 0.89 | 0.2185 | 0.2216 | 0.68 | 0.3924 | 0.4043 | 0.0113 |
| **Acid Insoluble Lignin (%)** | | | | | | | | | |
| | 295-872(A) | 7 | 0.75 | 0.0673 | 0.0683 | 0.65 | 0.0812 | 0.0838 | 0.0033 |
| | 400-810 (B) | 7 | 0.85 | 0.0528 | 0.0535 | 0.59 | 0.0912 | 0.0905 | -0.0245 |
| **Total Lignin (%)** | | | | | | | | | |
| | 295-872(A) | 8 | 0.81 | 0.0588 | 0.0597 | 0.62 | 0.0889 | 0.0918 | -0.0015 |
| | 400-810 (B) | 9 | 0.96 | 0.0266 | 0.0270 | 0.57 | 0.0914 | 0.0943 | 0.0026 |
| **Holocellulose (%)** | | | | | | | | | |
| | 295-872(A) | 6 | 0.51 | 0.0476 | 0.0482 | 0.37 | 0.0588 | 0.0608 | 0.0012 |
| | 400-810 (B) | 6 | 0.51 | 0.0547 | 0.0554 | 0.1 | 0.0652 | 0.0668 | 0.0097 |

[a]No of principal components, [b]Calibration $R^2$, [c]Root mean square of calibration, [d]Standard error of calibration, [e]Validation $R^2$, [f]Root mean square of validation, [g]Standard error of prediction

**Table 3. Results of PLS1 calibration and prediction models developed for the chemical and physical properties of yellow-poplar using the first derivative preprocessed full NIR spectra region.**

| | Wavelength (nm) | [a]PC | [b]$R^2$cal | [c]RMSEC | [d]SEC | [e]$R^2$val | [f]RMSEP | [g]SEP | BIAS |
|---|---|---|---|---|---|---|---|---|---|
| Bulk Density Aerated (g/cm$^3$) | | | | | | | | | |
| | 800-2500 (A) | 4 | 0.9 | 0.0044 | 0.0044 | 0.83 | 0.0054 | 0.0055 | -0.0007 |
| BulkDensity Tapped (g/cm$^3$) | | | | | | | | | |
| | 800-2500 (A) | 4 | 0.86 | 0.0067 | 0.0068 | 0.79 | 0.0072 | 0.0067 | -0.0031 |
| Ash (%) | | | | | | | | | |
| | 800-2500 (A) | 5 | 0.96 | 0.0074 | 0.0075 | 0.82 | 0.0132 | 0.0133 | 0.0024 |
| Extractives (%) | | | | | | | | | |
| | 800-2500 (A) | 4 | 0.83 | 0.2846 | 0.2885 | 0.78 | 0.3422 | 0.3387 | 0.0954 |
| Acid Insoluble Lignin (%) | | | | | | | | | |
| | 800-2500 (A) | 5 | 0.89 | 0.0488 | 0.0495 | 0.79 | 0.0696 | 0.0717 | 0.0037 |
| Total Lignin (%) | | | | | | | | | |
| | 800-2500 (A) | 4 | 0.81 | 0.0633 | 0.0641 | 0.66 | 0.0885 | 0.0895 | -0.0165 |
| Holocellulose (%) | | | | | | | | | |
| | 800-2500 (A) | 4 | 0.79 | 0.0370 | 0.0375 | 0.81 | 0.0416 | 0.0429 | 0.0012 |

[a]No of principal components, [b]Calibration $R^2$, [c]Root mean square of calibration, [d]Standard error of calibration, [e]Validation $R^2$, [f]Root mean square of validation, [g]Standard error of prediction
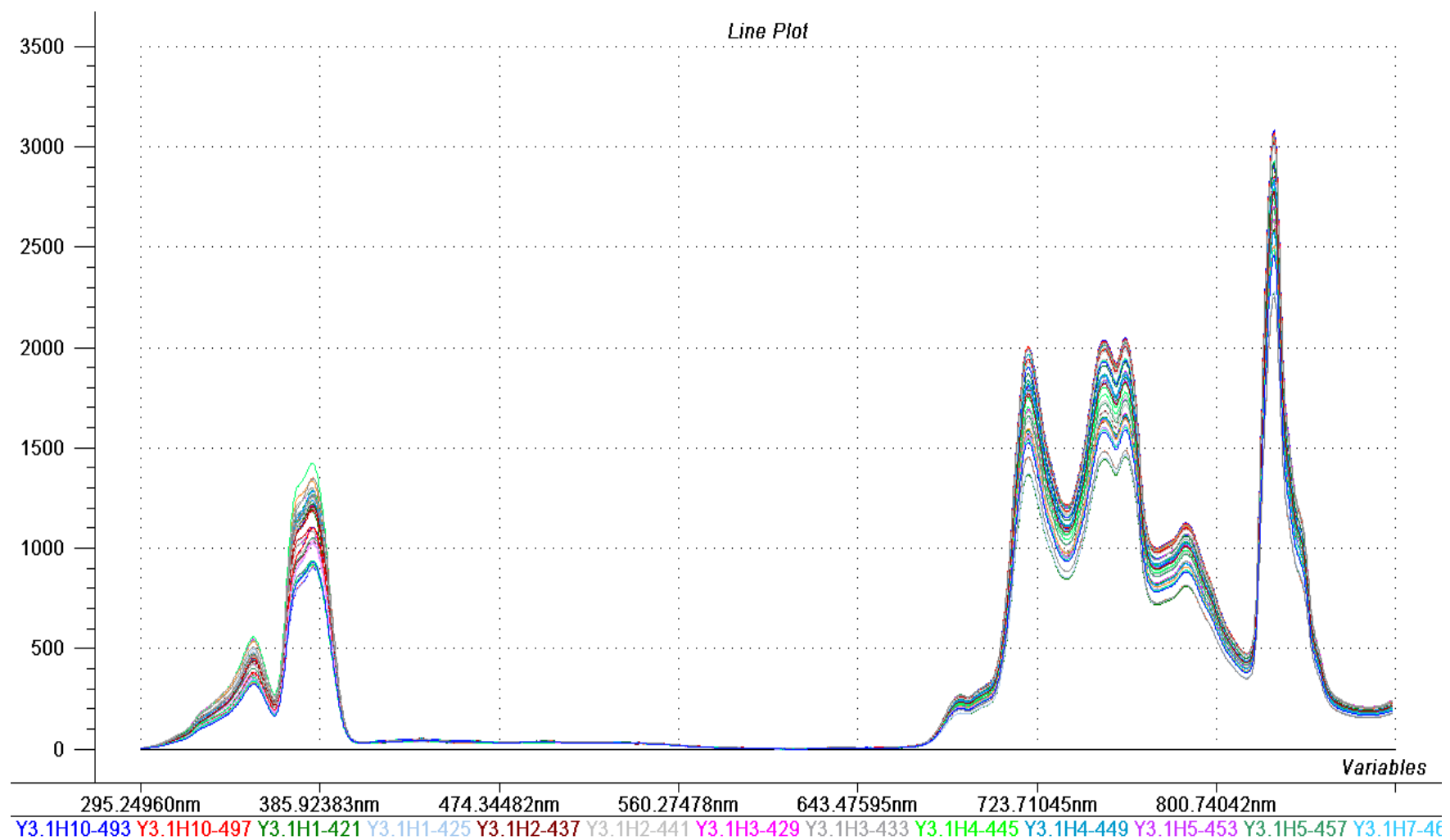
**Figure 1. Representative fluorescence spectra of yellow-poplar heartwood**
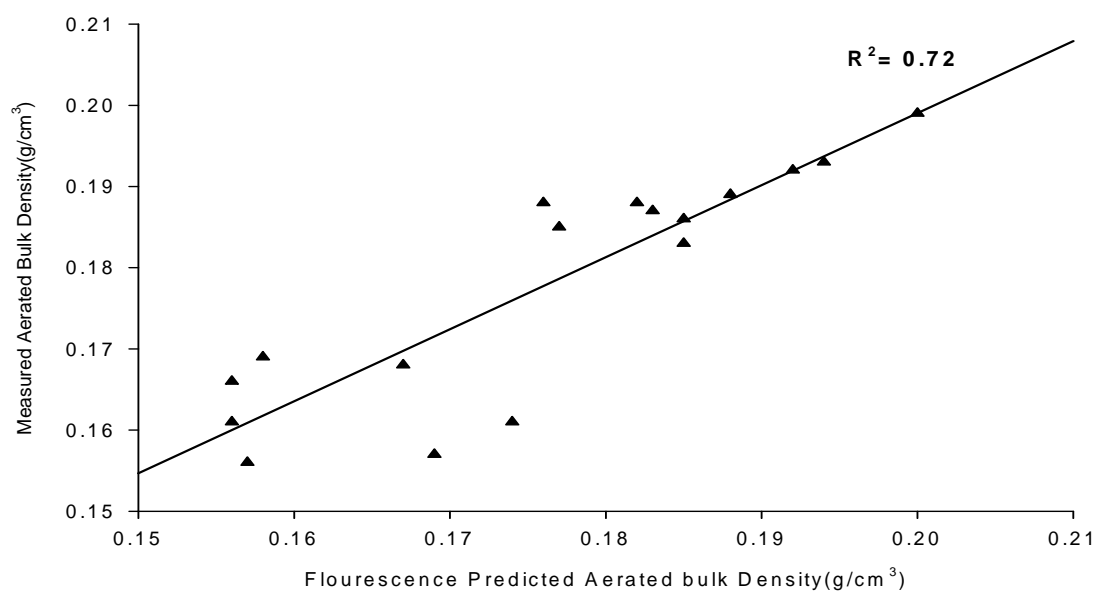
**Figure 2. PLS1 plot of predicted against measured aerated bulk density of yellow-poplar using the emission fluorescence spectra region (400-810nm)**
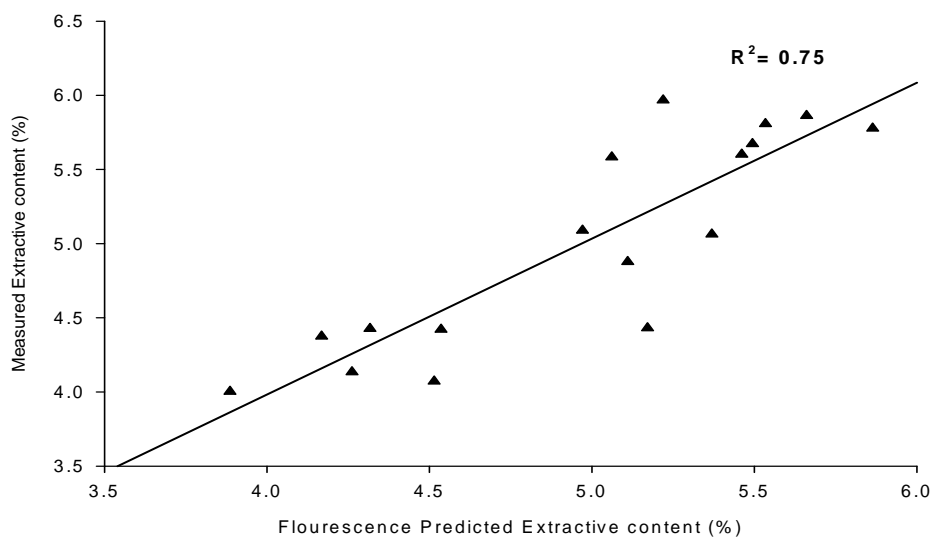


**Figure 3. PLS1 plot of predicted against measured extractive content using the full fluorescence spectra region (295-872nm)**

## References

1. Perlack R.D., Wright, L.L., Turnhollow, A.F., Graham, R.L., B.J. Stokes, and D.C. Erbach. Biomass as feed stock for a bioenergy and bioproducts industry: The technical feasibility of a billion-ton annual supply US Department of Energy. 2007, DOE/GO-102995-2135, http://www.ostl.gov/bridge)

2. Labbe, N.; Lee, Seung-Hawn.; Cho, Hyun-woo.; Jeong, M.K.; Andre, N. Enhanced discrimination and calibration of biomass NIR spectra data using non-linear kernel methods. Bioresource Technology. 2008, 99, 8445-8452

3. Baughman, E. 2005. Process Analytical Chemistry: Introduction and  and historical perspective. In "Process Analytical Technology. Spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries. Ed. K. A. Bakeev. Blackwell Publishing Ltd., Oxford, U.K,2005; 1-12.

4. Sluiter, A.; Snell, R.; Kelley, S.; Rials, T.; and Groom, L. Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood. Wood Sci Technol, 2004, 38, 257-276.

5. Poke, F.; Wright, J.; and Raymond, C. Predicting extractive and lignin contents in Eucalyptus globulus using near infrared reflectance analysis. J. Wood Chem. Technol. 2004, 24, 55-67.

6. Adedipe, O.;  Dawson-Andoh, B. Prediction of yellow-poplar (Liriodendron tulipifera) veneer stiffness and bulk density using near infrared spectroscopy and multivariate calibration. J. Near Infrared Spectrosc. 2008, 16, 487-496.

7. C. So, B. Via, L. Groom, Schimleck, L.R.; Shupe, T.; Kelley, S and Rials, T. Near Infrared Spectroscopy in the Forest Products Industry. Forest Product Journal. 2004, 54, 3

8. Tsuchikawa, S. A review of recent near infrared research for wood and paper. Applied spectroscopic reviews. 2007, 42:1, 43-71.

9. Castellan, A.; Ruggiero, R.; Frollini, E.; Ramos, L.A and Chirat, C. Studies on fluorescence of cellulosics. Holzforschung. 2007, 61, 504-508.

10. Olmstead, J.A.; Gray, D.G. Fluorescence spectroscopy of cellulose, lignin and mechanical pulp: A review. J. of pulp and paper sci. 1997,23:12, 571-581.

11. Pandey, K.K.; Upreti, N.K.; Srinivasan, V.V. A fluorescence spectroscopic study on wood. Wood Sci and technol. 1998, 32, 309-315

12. Albinsson, B.; Li, S.; Lundquist, K.; Stomberg, R. The Origin of Lignin Fluorescence, Journal of Molecular Structure. 1999, 508, 19-27.

13. Billa, E.; Pastou, A.; Monties, B.; Romero, J and Koukios, E.G. Multivariate chemometric analysis of the fluorescence spectra of eucalyptus wood. Industrial crops and products. 2000, 187-196.

14. Sum, S.T.; Singleton, D.L.; Paraskevopoulos, G.; Irwin, R.S.; Barbour, R.J.; Sutcliffe, R. Wood sci and technol. 1991. 25, 405-413.

15. Lam P.S.; Sokhansanj, S.; Bi, X.; Mani, S.; Lim, C.J.; Womac, A.R.; Hoque, M.; Peng, J.; JayaShankar, T.; Naimi, L.M.; and Nayaran, S. 2007. Physical characterization of wet and dry wheat straw and switchgrass – bulk and specific density. 2007 ASABE Annual International Meeting. Sponsored by ASABE, Minneapolis Convention Center, MN, 17-20 June 2007.

16. Yu, A.B.; Hall, J.S. Packing of fine powders subjected to tapping. Powder Technology. 1994, 78, 247–256.

17. Adullah E.C.; Geldart, D. The use of bulk density measurements as flowability indicators. Powder technology, 1999, 102, 151-165.

18. Wold, S.; Geladi, P.; Esbensen, K. Multiway principal components and PLS-analysis. Journal of Chemometrics. 1987:1, 41-56.

19. Analysis of color multiples in sugar processing using fluorescence spectroscopy and chemometrics. Ph.D. Thesis. Department of dairy and Food Science, The Royal Veterinary and Agricultural University, Denmark, 2000, p 18.

20. Esbensen, K.H. Multivariate data analysis in practice: An introduction to multivariate data analysis and experimental design, 5th edition. CAMO Process AS, Oslo, Norway, 2002, 598pp.

21. Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; and Wold, S. Introduction to multi- and megavariate data analysis using projection methods (PCA, PLS). Umetrics AB, mea, Sweden, 1999, 490pp.

22. Williams, P. 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., 2004.

23. Pettersen, C. The chemical composition of wood. In: Rowell, R.M., ed. The chemistry of solid wood. Advances in chemistry series. 207 Washington D.C: American Chemical Society; 1984: chapter 2.

24. Nkansah, K.; Dawson-Andoh, B.; Slahor, J. Rapid Characterization of Biomass: Part1. Near infrared spectroscopy characterization of the physical and chemical properties of yellow-poplar (*Liriodendron tulipifera L*). 2009: Unpublished.

25. Wold, S.K., K. Esbensen and P. Geraldi. 1987. Principal component analysis. Chemom. Intell. Lab. Systems. 2:37-52

26. Martens, H. and T. Naes. Multivariate calibration. Second Edition. John Wiley & Sons, N. York, USA. 1993.

27. Bro, R. Multiway calibration. Multi-Linear PLS. J. Chemometrics. 1997, 10:47-62.

28. Bro, R. PARAFAC: Tutorial and applications. Chemom. Intell. Lab. System. 1997, 38: 149-171.

29. Bro, R. Multi-way analysis in the food industry: models, algorithms, and applications. Ph.D. Thesis. University of Amsterdam, Netherlands. 1998.

30. Harshman, R.A. Foundations of PARAFAC procedure: Models and conditions for an "explanatory" multi-model factor analysis. UCLA Working papers in phonetics. 1970, 16:1-84

# CHAPTER FIVE: Rapid characterization of biomass using fluorescence spectroscopy coupled with multivariate data analysis: Part 2. Northern red oak (*Quercus rubra.)*

## Abstract

This paper is the final part of our study of the potential use of fluorescence spectrometer coupled with multivariate data analysis (MVDA) as a process analytical (PAT) tool for rapid characterization of a variety of biomass residues and this study focus on northern red oak (*Quercus rubra*) as a bio-energy feedstock. The results of the study showed that, with the exception of holocellulose content and ash content, fluorescence spectrometer can be used to predict the entire chemical and physical properties of biomass relevant to its bio-chemical conversion into bio-energy and bio-products. Similar near infrared (NIR) spectra-based prediction models of the same properties and population from a previous study were compared to this study. The performance of the NIR spectra-based prediction models of each property was slightly superior over the fluorescence spectra-based models even though both can be used for the same purpose. fluorescence spectra-based prediction models of holocellulose content and ash content of northern red oak were poorly predicted with a low $R^2$ (<0.5).

**Keywords:** Fluorescence, Multivariate data analysis, Process analytical technology, Near infrared

## Introduction

After the energy crisis that dominated the 1970's, interest in alternative sources of fuel has gained momentum. Bio-fuels have become the epitome of this search because its characteristics of being a clean source of energy. A comprehensive study of biomass resource (residues) as a potential feedstock for the bio-refinery industries put the contribution of biomass to total industrial and electric energy at 4% with annual increase of 3.9 quads in 2020 and 4.8 quads in 2030.[1] These estimates are based on the availability of approximately 1.3billion tons of dry biomass per year in the United States which is are sustainable resource and renewable.[1]

The heterogeneous nature of biomass presents a number of challenges for its biochemical conversion into energy.[2] Standard analytical methods of determining the properties of a biomass are slow and expensive. Therefore rapid characterization of the physical and chemical properties of any biomass is paramount prior to any conversion process.

The potential use of near infrared spectroscopy (NIRS) coupled with multivariate data analysis (MVDA) as a process analytical technology (PAT) tool in the forest products industry has been extensively studied and reviewed in a number of scientific reviews.[3-7] The used of much cheaper, sensitive, rapid, and portable spectroscopic tool such fluorescence have not gained much traction in the bio-fuel, forest products, and bio-products industry. Wood as a form of biomass is known to fluoresce due to the presence of flourophores in its chemical structure.[8] Olmstead and Gray (1997), extensively reviewed the source of fluorescence in cellulosic, lignin and mechanical pulp. A number of studies have also exhibited the possibility of classifying wood based on the fluorescence spectra from ground wood and solid wood.[10,11] The quantification of some wood properties based on fluorescence spectra from ground wood has also been reported.[12]

Dues to the heterogeneous nature of wood, the fluorescence spectra may vary from species to species.[11,13,14] This study is a follow up on our previous paper that employed fluorescence spectroscopy (FS) as a PAT method to predict the physical and chemical properties of biomass. Specific objective of this study were to:

1. Use Fluorescence spectroscopy coupled with MVDA to predict important physical and chemical properties of northern red oak (*Quercus rubra*) These properties include ash content, extractives content, total lignin, acid insoluble lignin, holocellulose and bulk density.

2.  Compare the efficiency of two spectra region of the fluorescence spectrum in order to determine which region contains relevant spectroscopic information for accurate quantification of the chemical and physical properties of northern red oak (*Quercus rubra*).

3.  Compare the performance of Fluorescence spectra-based and NIR spectra-based PLS1 prediction models of each physical and chemical property of northern red oak of the same population.

## Materials and Methods

Similar sampling methods used in our previous paper (Nkansah, K and Dawson-Andoh, B. 2009) will be applied in this study.

## Materials

Northern red oak (*Quercus rubra*) was randomly selected in West Virginia University Research forest located about 13 km east of Morganton, WV. Each tree was approximately 14 meters tall. Three 50 mm disk were removed from the bole and a total of 60 samples from the heartwood with size 19mm X 19mm X 50 mm each were randomly selected for this study. Detailed description of sampling methods and conditioning are described in our previous study.[15]

## Methods

**Fluorescence spectra measurement**

Ocean Optics USB2000-FLG fluorescence spectrometer (Ocean Optics, Florida) with an application wavelength range of 380-1050nm was used to capture the fluorescence spectra of the samples. The spectrometer was fitted with a customized fiber optic probe for measuring the fluorescence of solid biomass materials. A UV light source (PX-2 Pulsed Xenon Lamp) of range 200-750 nm was used to excite the samples at a band wavelength region of 200-400nm. Ten scans were collected and averaged to a single spectrum, a total of four spectra were taken on the tangential face along the length of each sample and averaged to one average spectrum. Detailed

description of the spectrometer, solid collection probe and technique of measurement used in this study are described in details elsewhere.[15] The Unscramble software v9.8 (Camo Smart, Woodbridge, NJ, USA) was used for MVDA of the spectra data

**Determination of chemical properties**

Chemical properties measured included total lignin (insoluble and acid-soluble lignin), extractives, holocellulose, and ash contents. These properties were determined by classical wet chemistry methods as per protocols of ASTM D-1106-96, NREL LAP-004(1996), ASTM D1105-96(2007) and ASTM D-1102-84 respectively. Holocellulose was given by the difference between initial weight of ground wood and the sum of the above chemical components. This data together with their physical property; bulk density was used as the Y-response variable for PLS modeling.

**Bulk density measurement**

Bulk density of biomass is a major contributing factor to the determination of "cost and logistic requirements of handling and moving biomass from point of production to biorefinery.[16] It is a function of the size, shape and individual particle density of the lignocellulose biomass. Two types of bulk densities were determined; aerated bulk density and tapped bulk density in accordance with the method described by Yu *et al.* (1994) and Adullah *et al.* (1998). Both volume and weight determination for each method was replicated twice for each sample and the average computed for the sample.

**Development of fluorescence calibration**

All spectra were analyzed using Unscrambler multivariate data analysis software (version 9.8, Camo Inc., Woodbridge, New Jersey, US). A total of 60 averaged spectra for the heartwood of northern red oak; one averaged spectrum for each specimen.

Several methods can be used in chemometrics to analyze fluorescence data. The two predominant methods are (i) two-way structure –samples X emission spectra wavelength using bilinear chemometric methods (PCA, PLS) and (ii) multi-way which is notable referred to as three-may (samples X excitation X emission wavelength). The three–way method is more robust when compared to the two-way method.[24-26] The two-way PCA is used to find the direction of

variation in the fluorescence data whiles the PLS decomposes the independent (emission spectra-X) and dependent (reference-Y) matrix successively and captures the variation and correlation between X and Y.[27-28] Detailed summary of these methods are presented in the first part of our study.[15] In this study the two–way method was used and a three-way analysis will be presented in a subsequent paper.

To obtain an overview of data, Principal Component Analysis (PCA) was performed and it revealed the presence of six outliers. These were eliminated from subsequent analysis. The spectra data were divided into calibration and prediction sets. Data was divided into calibration and validation sets. The calibration set consisted of two spectra out of every three consecutive spectra (n=36 spectra) and the prediction set (n=18 spectra) comprised of one spectrum out of every three consecutive spectra. All the fluorescence spectra were combined into a single data matrix [X-matrix] and the measured physical or chemical properties measured were also combined into separate response matrix [$Y_i$]. The Y-matrices consisted of data from bulk densities (aerated bulk density and tapped bulk density), ash content, holocellulose content, total lignin, acid insoluble lignin, and extractives contents measured values.

Calibration models were developed for two spectra region: (i) full fluorescence spectrum that encompassed the excitation and emission spectra (295-872nm) and (ii) emission spectra region (400-810nm) that encompassed only the emission spectrum. PLS1 calibration models were developed using the raw spectra data. In our study two other forms of spectra preprocessing of the data including first derivative and multiple scatter correction (MSC) were evaluated. These preprocessing techniques did not improve the prediction models developed to any degree than the raw spectra data models. Comparatively the raw spectra prediction models were much better than any of the preprocessing applied. Only the raw spectra data calibration and prediction models were evaluated and reported in this study. PLS1 calibration models were developed with X- and Y-matrices. The response variable holocellulose content, total lignin content, and acid-insoluble lignin content were preprocessed using a cubed root transformation due to the large differences in the values measured.[6] The Y-data matrix were mean –centered and calibration models were developed using full-cross validation method and then used to predict the response of the prediction/validation set. Detailed multivariate analyses (calibration and validation methods) of the data are described in our previous study.[15]

**Measure of calibration model performance**

Parameters used to evaluate the performance of our calibration models include: (i) coefficient of determination $R^2$, (ii) root mean square error of prediction (RMSEP), (iii) standard error of prediction (SEP), and (iv) bias. These parameters are discussed in details elsewhere.[15] Root mean square error of prediction is a direct estimate of the prediction error in the Y variables of the model; whiles SEP is a measure of precision of the models predicting ability corrected for the bias in the validation of the model. Bias is a measure of systematic difference between average values of the calibration set and the validation set of a model. If no such difference exists, the bias will be equal to zero. $R^2$ is a measure of total variance between measured and predicted values that can be modeled by linear association. The closer $R^2$ value is to one, the stronger the correlation.[4,19-21]

Result presented in Tables 2 and 3 are PLS1 models developed for prediction of measured physical and chemical properties with the raw fluorescence spectra and first derivative preprocessed full NIR spectra of northern red oak respectively.

## Results and discussion

Results of measured chemical and physical properties of northern red oak are presented in Table 1. Chemical properties of northern red oak are expressed as a percentage of the original oven dry weight of ground wood analyzed. Ash content had a range of 0.15-0.20; extractives content had a range of 4.48-7.74; acid insoluble lignin exhibited a range of 17.90-21.95 whiles total lignin content had a range of 18-21.97. Holocellulose content also had a range of 71.48-74.93. These results were comparable to a similar study reported by Petersen (1984). For the physical property measured, aerated bulk density and tapped bulk density exhibited a range of 0.23-0.28 and 0.31-0.38 g/cm$^3$ respectively. All collected fluorescence spectra were similar as depicted in Figure 1.

## Evaluation of chemical and physical properties prediction using fluorescence spectra data

Table 2 presents results of PLS1 calibration and prediction models of ash content, extractive content, acid insoluble lignin, total lignin, holocellulose content, aerated and tapped bulk density of northern red Oak (*Quercus rubra*). The number of optimum PCs required for developing calibration models for all properties in this study ranged between 1-7 PCs. Generally prediction models developed with the full spectra and the emission spectra for all the response variables considered in this study exhibited almost equal prediction accuracy with regards to $R^2$ values.

For aerated bulk density, prediction model developed with the full spectra and the emission spectra exhibited high $R^2$ values of (0.85) and (0.83) respectively. Prediction model developed with the full spectra had a low RMSEP (0.0055), SEP (0.0057) and bias (0.0001). Slightly higher RMSEP (0.0058), SEP (0.0059) and bias (0.0008) values were exhibited by the prediction model developed with the emission spectra as compared to the full spectra-based prediction model. Both models required moderate number of PC's (7) for model development. Prediction model developed with the full spectra exhibited a slightly better model over the emission spectra-based model. This observation was based on the slight improvement in $R^2$, RMSEP, SEP and bias values of the full spectra-based prediction model. Both models were also adequate for most activities that accuracy is important based on their high $R^2$ values. [21]

Tapped bulk density was predicted with a high $R^2$ using the full spectra-based ($R^2$ 0.87) and the emission spectra-based ($R^2$ 0.92) prediction models. The Prediction model developed with the full spectra exhibited low RMSEP (0.0074), SEP (0.0075) and bias (-0.0012). Slightly lower RMSEP, SEP (0.0057) and bias (0.0016) were exhibited by the prediction model developed with the reduce spectra region. Based on this observation the emission spectra region was judged a better model over the full spectra prediction model. Both models required a moderate number of PC's (7) for model development. The full spectra-based prediction model was adequate for most prediction activities whiles the emission spectra-based model was good for most applications including quality assurance purposes. These observations are based on high $R^2$ values both models exhibited (0.87-0.92).[21]

Even though the PLS1 prediction models developed for ash content with the full spectra and the emission spectra had low RMSEP, SEP values (<0.012) and a low bias within the range

(0.0011-0.0015), both models were judged to be poor models based on the low $R^2$ values exhibited (<0.24).[21] Similarly poor modeling of ash content was exhibited in our previous study using NIR as a spectroscopic tool in PLS1 prediction modeling on the same population used in this study. The low $R^2$ values in the previous study was ascribed to the very minute ash content values of red oak and the subtle variance within the population used in the calibration and prediction of the models.[23] Since this study used the same sample population as the previous study, the inability of fluorescence spectroscopy to predict ash content of northern red oak was also attributed to the minute ash content of northern red oak and the lack of variance within the population used in the study.

An increase in the sample space may increase the possibility of predicting ash content using (FS) coupled with MVDA by invariably increasing the variance within the population and making them more profound and detectable by the PLS1 modeling technique. Similar result was reported in a study of acid soluble lignin calibration model of *E.globulus* developed using the NIR spectrum. The poor $R^2$ of calibration was attributed to the minute quantitative measured values of acid soluble content of *E.globulus*. The study also suggested that the subtle variation within the measured values may contribute to the low $R^2$ of 0.62.[7]

Prediction models developed for extractive content exhibited high $R^2$ values of (0.92) and (0.91) for the full spectra-based and the emission spectra-based prediction models respectively. (Figure 2)  Low RMSEP (0.2693) and SEP (0.2767) were exhibited by full spectra-based prediction model. Slightly higher RMSEP (0.2894) and SEP (0.2935) were exhibited by the emission spectra-based prediction model. The prediction model developed with the full spectra and the emission spectra showed low bias values of (-0.0138) and (-0.0487) respectively. Both models required moderate number of PC's (6) for model development. The prediction accuracy of the full spectra-based model was generally stronger than the emission spectra-based model. This observation was based on the slightly lower RMSEP, SEP and bias of the full spectra-based model as compared to the emission spectra model. Based on high $R^2$ values exhibited by the models, both models were adequate for most applications where high accuracy is vital including quality assurance activities.[21]

Acid insoluble lignin, PLS1 prediction models, developed with the full spectra and the emission spectra exhibited moderate $R^2$ values of (0.71) as depicted in Figure 2. Low RMSEP (0.0676) and SEP (0.0691) exhibited by the full spectra-based prediction model. The RMSEP

(0.0693) and SEP (0.0713) values exhibited by the emission spectra-based prediction model were slightly higher than similar parameters of the full spectra-based model. Prediction models developed with the full spectra and the emission spectra exhibited low bias values (<0.0090). Even though the correlation between measure and fluorescence predicted acid insoluble lignin for both models were equal, the full spectra-based prediction model was a slightly better prediction model as compared to the emission spectra-based prediction model. This observation was based on the slightly lower RMSEP and SEP values exhibited by the full spectra-based prediction model as compared to the similar parameters of the emission spectra-based model. Both models required low number of PC's (5) for model development and they were both adequate for approximate calibration and screening purpose based on their $R^2$ values (0.71).[21]

Total lignin content was also predicted with high $R^2$ of (0.88) and (0.93) for models developed using the full spectra region and the emission spectra region respectively. The full spectra-based prediction model exhibited low RMSEP (0.0436) and SEP (0.0446). Much lower RMSEP (0.0355) and SEP (0.0346) was exhibited by the emission spectra-based prediction model. The emission spectra-based prediction model of total lignin exhibited a much higher model performance with respect to the decrease in RMSEP, SEP and an improved $R^2$ of the model as compared similar model parameters of the full spectra-based prediction model. Based on the $R^2$ of each model, the full spectra-based prediction model was judged to be adequate for most prediction purposes where accuracy is imperative, including research. The emission spectra-based prediction model was adequate for quality assurance activities where strict accuracy was required.[21]

For holocellulose content, prediction models developed with both spectra regions exhibited low RMSEP and SEP (<0.022). The full spectra-based and emission spectra-based prediction models exhibited low bias values of (0.0016) and (-0.0014) respectively. Both prediction models were judge poor and unusable for any purpose based on the low $R^2$ values they exhibited (0.23-0.48). The poor prediction modeling of holocellulose content of northern red oak may be as a result of insufficient population size for calibration and prediction of the model. An increase in the sample size may improve the models by increasing the variance within the measured sample population. It is also possible that the FS coupled with MVDA cannot be used to predict the holocellulose content of northern red oak with high accuracy. Similar results was

exhibited in our previous study; were holocellulose content of yellow-poplar was poorly predicted with low $R^2$ values (<0.40).[21]

## Comparison of fluorescence and near infrared spectra-based models

The first derivative preprocessed NIR spectra-based prediction models of northern red oak from our previous study (Nkansah, K and Dawson-Andoh. 2009) were compared with similar response variable prediction models developed with the fluorescence spectra to evaluate the relative performance of the models for each response variable. Both studies used the same sample population in developing prediction models and this allowed us to make a meaningful comparison between the two different spectra-based prediction models. Only the full spectra-based prediction models of the fluorescence and NIR study were compared with respect to their $R^2$, SEP values and the number of PC's of each prediction model.

For aerated bulk density; even though fluorescence spectra-based exhibited a slightly improved prediction model [$R^2$ (0.85) and SEP (0.0057)], as compared the NIR spectra-based prediction model [$R^2$ (0.84) and SEP (0.0068)]. The NIR spectra-based model was judged as better prediction model of aerated bulk density with regard to the lower number of PC's (6) required by the NIR spectra-based model as compared to the fluorescence spectra-based model (7 PC's). The $R^2$ of both models were comparably very close and could be used for similar activities including research purposes. The performance of the NIR spectra-based prediction model of tapped bulk density was judged a better prediction model over the fluorescence spectra-based model. This observation was based on the slightly higher $R^2$ (0.89), lower SEP (0.0072) and number of PC's (5) exhibited by the NIR spectra-based model as compared to the fluorescence spectra-based prediction model that exhibited a slightly lower $R^2$, an increase in SEP and number of PC's [$R^2$ (0.87), SEP (0.0075) and PC's (7) ].

Even though the ash content of northern red oak prediction models of both spectra-based models were poor and unusable, the NIR spectra-based model ($R^2$= 0.36) exhibited the potential of being a better model over the fluorescence based-spectra model ($R^2$ =0.23). Similarly, for extractive content, the NIR spectra-based prediction model performed slightly better than the fluorescence spectra-based model. The NIR spectra-based model exhibited a slightly higher $R^2$ (0.93) and a much lower SEP (0.2519) as compared to the fluorescence spectra-based model [$R^2$ (0.92) and SEP (0.2767)]. Both models had the same number of PC's (6).

Acid insoluble lignin of northern red oak was also predicted with slightly higher accuracy by the NIR spectra-based prediction model as compared to the corresponding fluorescence spectra-based model. The NIR spectra-based model exhibited a slightly higher $R^2$ (0.73), low SEP (0.0647) and low number of PC's (4) as compared the fluorescence spectra-based model [$R^2$ (0.71), SEP (0.0691) and PC's (5)]. Even though the NIR spectra-based prediction model of total lignin exhibited lower number of PC's (5) as compared to the fluorescence spectra-based (PC's: 7), the huge difference in $R^2$ between the NIR spectra-based model ($R^2$ 0.72) and fluorescence spectra-based models ($R^2$ 0.88) could not be over shadowed. The fluorescence spectra-based model also exhibited lower SEP (0.0446) as compared to the NIR spectra-based model [SEP (0.0477)]. Based on these observations the performance of the fluorescence spectra-based models in predicting total lignin was deemed superior over the NIR spectra-based model.

For holocellulose, the NIR spectra-based prediction model performed poorly and exhibited low $R^2$ (0.48), slightly higher SEP (0.0179), and number of PC's (5) as compared to the NIR spectra-based prediction model which exhibited a strong correlation between measured and predicted hollocelulose content ($R^2$ 0.89). This model also exhibited much lower SEP (0.0124) and required moderate number of PC's (7).

With the exception of total lignin content, NIR spectra-based prediction models for all response variables generally exhibited slightly higher model performance over the corresponding fluorescence spectra-based models. The performance of the latter spectra-based model was close in comparison to the performance of the NIR spectra-based models. The spectra-based models could be used to predict both bulk densities, extractive content, acid insoluble lignin content, and total lignin content with comparable accuracies. Some form of mathematical preprocessing, or a more robust MVDA regression technique including three-way analysis (PARAFAC) of fluorescence spectra may improve the performance of the fluorescence spectra-based models and should be investigated in future studies of FS as a PAT tool for biomass characterization. This will be presented in a subsequent paper.

The detailed comparison of the performance of the NIR-based prediction models developed with the raw spectra to similar fluorescence spectra-based prediction models were not conducted in this study because the preprocessed models performed much better than the raw NIR spectra-based models and these results are presented elsewhere.[23] It should be mentioned that with the exception of ash content and holocellulose content; a general overview of the

performance of the fluorescence spectra-based prediction model of this study were superior over similar response variable prediction models developed with the raw NIR spectra.

## Conclusion

This study has established that some chemical properties (extractive content, acid insoluble lignin and total lignin) and physical properties (Aerated bulk density and tapped bulk density) of northern red oak can successfully be predicted with moderate to high degrees of accuracy ($R^2$ between 0.71-0.93) using FS coupled with MVDA. Ash content and holocelluse content of red oak were poorly predicted with both spectra regions considered in this study ($R^2 < 0.5$).

This study has also shown the potential of using only the emission region of the fluorescence spectra coupled with MVDA to predict some chemical and physical properties of northern red oak that is much comparable to results that could be obtained using the full fluorescence spectrum. Model performances of fluorescence spectra-based prediction models were comparable to NIR spectra-based models of similar prediction models, even though the NIR spectra-based models were slightly superior over the fluorescence spectra-based models. A more robust MVDA analysis such as three-way PARAFAC technique may improve the fluorescence spectra-based models and need to be investigated. Tremendous number of research on rapid characterization of biomass properties have focused mainly on the use of NIR spectrometer which is a much expensive and bulk spectrometer as compared to the fluorescence spectrometer used in this study. This study has established the potential of using fluorescence spectrometer as an inexpensive, rapid, and portable tool for characterizing some chemical and physical properties of northern red oak with high prediction accuracies.

**Table 4. Statistics of measured chemical and physical properties of northern red oak (60 samples)**

|  | Mean* | SD[a] | Min[b] | Max[c] |
|---|---|---|---|---|
| Aerated Bulk Density (g/cm$^3$) | 0.25 | 0.015 | 0.23 | 0.28 |
| Tapped Bulk Density (g/cm$^3$) | 0.35 | 0.023 | 0.31 | 0.38 |
| Ash (%) | 0.18 | 0.013 | 0.15 | 0.20 |
| Extractives (%) | 6.25 | 0.997 | 4.80 | 7.74 |
| Acid Insoluble Lignin (%) | 20.54 | 1.162 | 17.90 | 21.95 |
| Soluble Lignin (%) | 0.034 | 0.015 | 0.003 | 0.063 |
| Total Lignin (%) | 20.60 | 1.028 | 18.00 | 21.97 |
| Holocellulose (%) | 74.8 | 1.45 | 71.48 | 74.93 |

Mean, [a]Standard deviation, [b]Minimum measured value, [c]Maximum maximum measured value

**Table 5. Results of PLS1 calibration and prediction models developed for the chemical and physical properties of northern red oak using the raw fluorescence spectra at two wavelength regions**

| | Wavelength (nm) | [a]PC | [b]$R^2$cal | [c]RMSEC | [d]SEC | [e]$R^2$val | [f]RMSEP | [g]SEP | BIAS |
|---|---|---|---|---|---|---|---|---|---|
| Bulk Density Aerated (g/cm$^3$) | | | | | | | | | |
| | 295-872(A) | 7 | 0.86 | 0.0058 | 0.0058 | 0.85 | 0.0055 | 0.0057 | -5.9E-05 |
| | 400-810 (B) | 7 | 0.87 | 0.0054 | 0.0054 | 0.83 | 0.0058 | 0.0059 | 0.0008 |
| Bulk Density Tapped (g/cm$^3$) | | | | | | | | | |
| | 295-872(A) | 7 | 0.8 | 0.0095 | 0.0097 | 0.87 | 0.0074 | 0.0075 | -0.0012 |
| | 400-810 (B) | 7 | 0.88 | 0.0073 | 0.0073 | 0.92 | 0.0057 | 0.0057 | 0.0016 |
| Ash (%) | | | | | | | | | |
| | 295-872(A) | 1 | 0.45 | 0.0073 | 0.0074 | 0.23 | 0.0104 | 0.0106 | 0.0015 |
| | 400-810 (B) | 1 | 0.42 | 0.0074 | 0.0076 | 0.22 | 0.0105 | 0.0107 | 0.0011 |
| Extractives (%) | | | | | | | | | |
| | 295-872(A) | 6 | 0.85 | 0.3658 | 0.3709 | 0.92 | 0.2693 | 0.2767 | -0.0138 |
| | 400-810 (B) | 6 | 0.78 | 0.4467 | 0.4530 | 0.91 | 0.2894 | 0.2935 | -0.0487 |
| Acid Insoluble  Lignin (%) | | | | | | | | | |
| | 295-872(A) | 5 | 0.61 | 0.0703 | 0.0713 | 0.71 | 0.0676 | 0.0691 | 0.0088 |
| | 400-810 (B) | 5 | 0.62 | 0.0744 | 0.0754 | 0.71 | 0.0693 | 0.0713 | 0.0029 |
| Total Lignin (%) | | | | | | | | | |
| | 295-872(A) | 7 | 0.86 | 0.0409 | 0.0415 | 0.88 | 0.0436 | 0.0446 | 0.0051 |
| | 400-810 (B) | 7 | 0.87 | 0.0427 | 0.0432 | 0.93 | 0.0355 | 0.0346 | 0.0110 |
| Holocellulose (%) | | | | | | | | | |
| | 295-872(A) | 5 | 0.48 | 0.0176 | 0.0178 | 0.48 | 0.0175 | 0.0179 | 0.0016 |
| | 400-810 (B) | 3 | 0.37 | 0.0195 | 0.0197 | 0.23 | 0.0209 | 0.0214 | -0.0014 |

[a]No of principal components, [b]Calibration $R^2$, [c]Root mean square of calibration, [d]Standard error of calibration, [e]Validation $R^2$, [f]Root mean square of validation, [g]Standard error of prediction

**Table 6. Results of PLS1 calibration and prediction models developed for the chemical and physical properties of northern red oak using the first derivative preprocessed full NIR spectra region.**

| | Wavelength (nm) | [a]PC | [b]$R^2$cal | [c]RMSEC | [d]SEC | [e]$R^2$val | [f]RMSEP | [g]SEP | BIAS |
|---|---|---|---|---|---|---|---|---|---|
| Bulk Density Aerated (g/cm$^3$) | | | | | | | | | |
| | 800-2500 (A) | 6 | 0.96 | 0.0030 | 0.0031 | 0.84 | 0.0066 | 0.0068 | 0.0005 |
| BulkDensity Tapped (g/cm$^3$) | | | | | | | | | |
| | 800-2500 (A) | 5 | 0.94 | 0.0055 | 0.0055 | 0.89 | 0.0078 | 0.0072 | -0.0035 |
| Ash (%) | | | | | | | | | |
| | 800-2500 (A) | 4 | 0.62 | 0.0079 | 0.0080 | 0.36 | 0.0119 | 0.0098 | 0.0072 |
| Extractives (%) | | | | | | | | | |
| | 800-2500 (A) | 6 | 0.98 | 0.1498 | 0.1517 | 0.93 | 0.2550 | 0.2519 | 0.0729 |
| Acid Insoluble Lignin (%) | | | | | | | | | |
| | 800-2500 (A) | 4 | 0.82 | 0.0569 | 0.0576 | 0.73 | 0.0631 | 0.0647 | 0.0057 |
| Total Lignin (%) | | | | | | | | | |
| | 800-2500 (A) | 5 | 0.92 | 0.0323 | 0.0327 | 0.70 | 0.0531 | 0.0477 | -0.0262 |
| Holocellulose (%) | | | | | | | | | |
| | 800-2500 (A) | 7 | 0.99 | 0.0036 | 0.0037 | 0.89 | 0.0124 | 0.0124 | -0.0028 |

[a]No of principal components, [b]Calibration $R^2$, [c]Root mean square of calibration, [d]Standard error of calibration, [e]Validation $R^2$, [f]Root mean square of validation, [g]Standard error of prediction
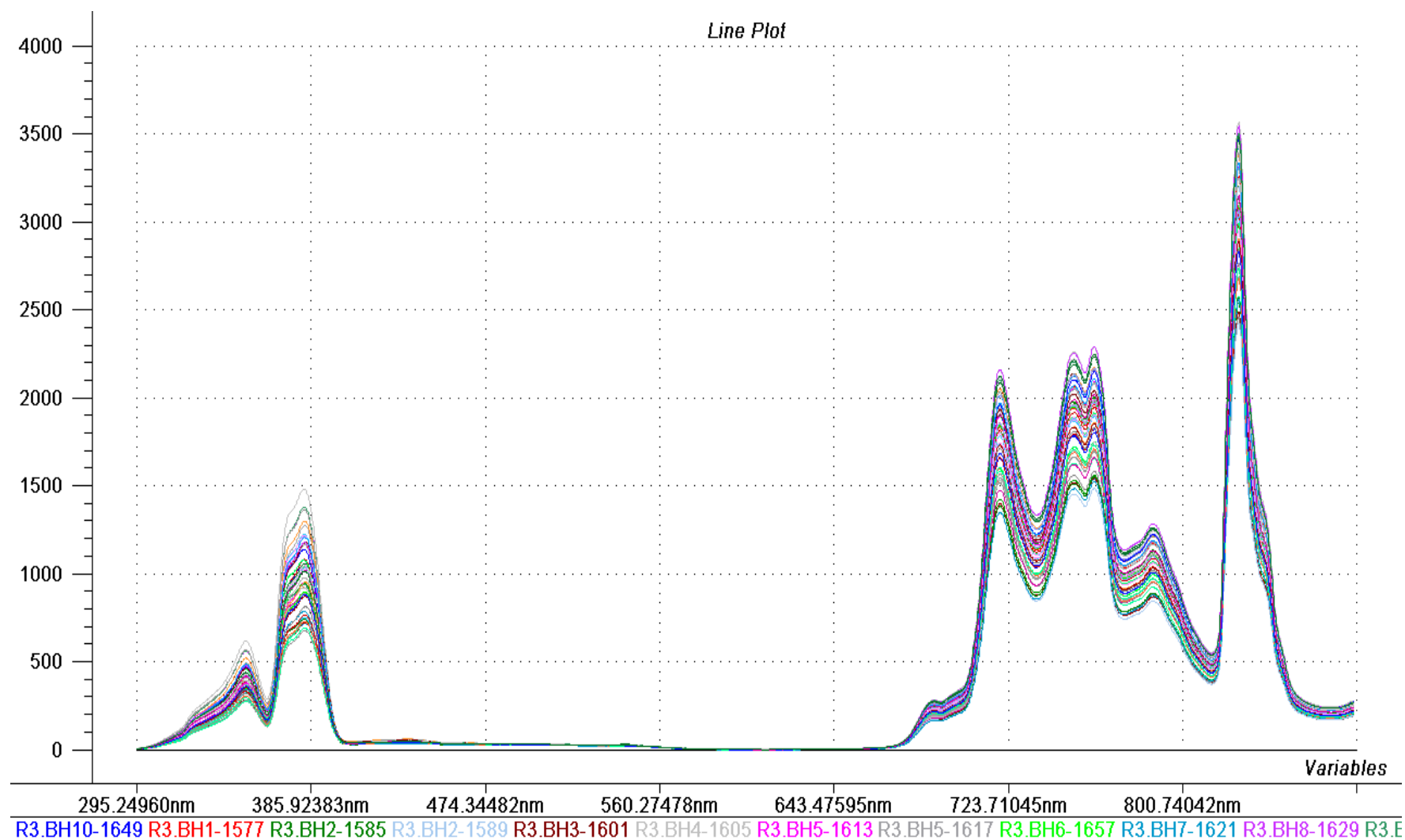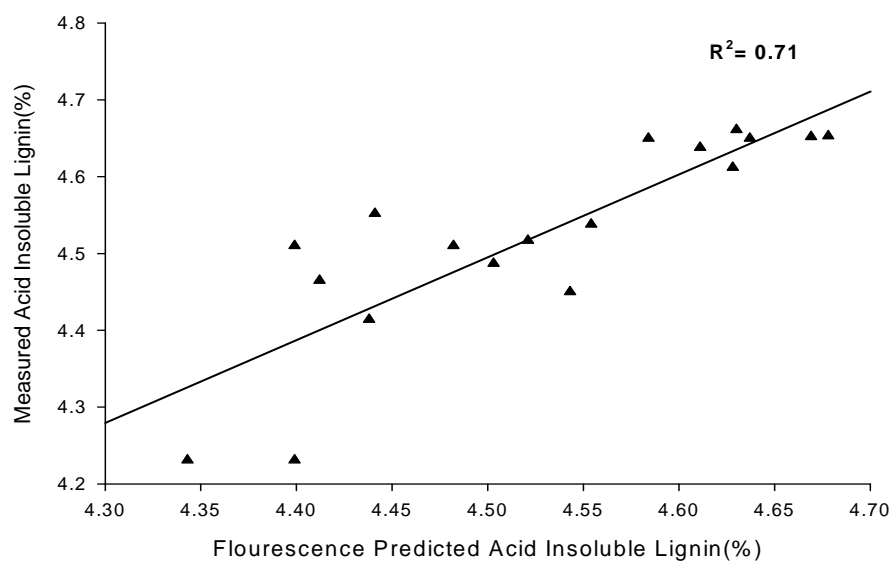
**Figure 5. Representative fluorescence spectra of northern red oak heartwood**
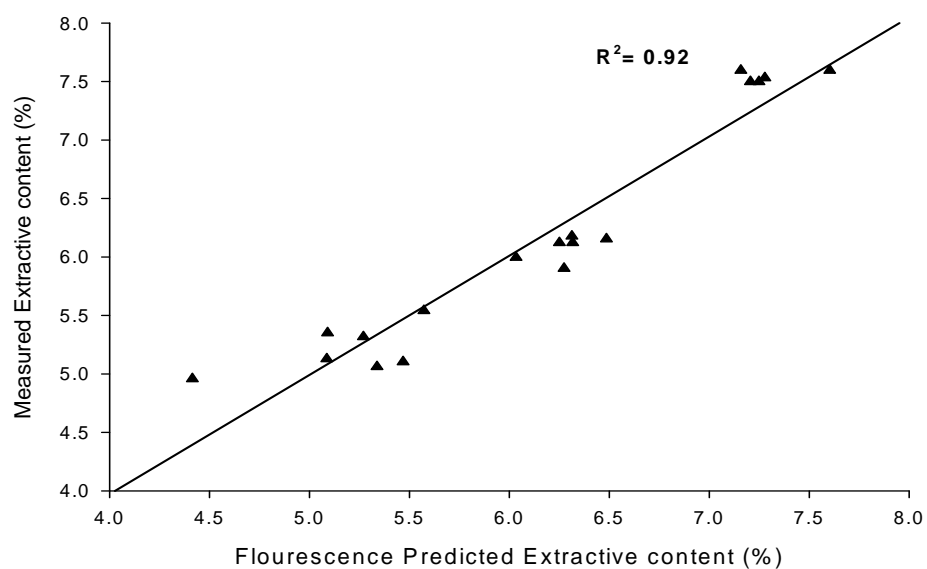
**Figure 6. PLS1 plot of predicted against measured extractive content using the full fluorescence spectra region (295-872nm)**



**Figure 7. PLS1 plot of predicted against measured acid insoluble lignin content of northern red oak using the emission fluorescence spectra region (400-810nm)**

# References

1. Perlack R.D., Wright, L.L., Turnhollow, A.F., Graham, R.L., B.J. Stokes, and D.C. Erbach. Biomass as feed stock for a bioenergy and bioproducts industry: The technical feasibility of a billion-ton annual supply US Department of Energy. 2007, DOE/GO-102995-2135, (http://www.ostl.gov/bridge)

2. Labbe, N.; Lee, Seung-Hawn.; Cho, Hyun-woo.; Jeong, M.K.; Andre, N. Enhanced discrimination and calibration of biomass NIR spectra data using non-linear kernel methods. Bioresource Technology. 2008, 99, 8445-8452

3. C. So, B. Via, L. Groom, Schimleck, L.R.; Shupe, T.; Kelley, S and Rials, T. Near Infrared Spectroscopy in the Forest Products Industry. Forest Product Journal. 2004, 54, 3.

4. Sluiter, A.; Snell, R.; Kelley, S.; Rials, T.; and Groom, L. Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood. Wood Sci Technol, 2004, 38, 257-276.

5. Tsuchikawa, S. A review of recent near infrared research for wood and paper. Applied spectroscopic reviews. 2007, 42:1, 43-71.

6. Adedipe, O.;  Dawson-Andoh, B. Prediction of yellow-poplar (Liriodendron tulipifera) veneer stiffness and bulk density using near infrared spectroscopy and multivariate calibration. J. Near Infrared Spectrosc. 2008, 16, 487-496.

7. Poke, F.; Wright, J.; and Raymond, C. Predicting extractive and lignin contents in Eucalyptus globulus using near infrared reflectance analysis. J. Wood Chem. Technol. 2004, 24, 55-67.

8. Castellan, A.; Ruggiero, R.; Frollini, E.; Ramos, L.A and Chirat, C. Studies on fluorescence of cellulosics. Holzforschung. 2007, 61, 504-508.

9. Olmstead, J.A.; Gray, D.G. Fluorescence spectroscopy of cellulose, lignin and mechanical pulp: A review. J. of pulp and paper sci. 1997,23:12, 571-581.

10. Sum, S.T.; Singleton, D.L.; Paraskevopoulos, G.; Irwin, R.S.; Barbour, R.J.; Sutcliffe, R. Wood sci and technol. 1991. 25, 405-413.

11. Pandey, K.K.; Upreti, N.K.; Srinivasan, V.V. A fluorescence spectroscopic study on wood. Wood Sci and technol. 1998, 32, 309-315

12. Billa, E.; Pastou, A.; Monties, B.; Romero, J and Koukios, E.G. Multivariate chemometric analysis of the fluorescence spectra of eucalyptus wood. Industrial crops and products. 2000, 187-196.

13. Dyer T.S., Wood Fluorescence of Indigenous South African Trees, IAWA Bulletin n.s., 9(1), 1988: 75-87.

14. Krishna S., Chowdhury K.A., Fourescenhce of wood under Ultra-violet Light, Indian Forester, LXI, 1935.

15. Nkansah, K; Dawson-Andoh, B. Rapid Characterization of Biomass: Part1. Fluorescence Spectroscopy characterization of the physical and chemical properties of yellow-poplar (*Liriodendron tulipifera L*). 2009: Unpublished.

16. Lam P.S., S. Sokhansanj, X. Bi, S. Mani, C.J. Lim, A.R. Womac, M. Hoque, J. Peng, T. JayaShankar, L.J. Naimi and S. Nayaran. 2007. Physical charatcterization of wet and dry wheat straw and switchgrass – bulk and specific. density 2007 ASABE Annual International Meeting. Sponsored by ASABE, Minneapolis Convention Center, MN, 17-20 June 2007)

17. Yu A.B.,and J.S. Hall. 1994. Packing of fine powders subjected to tapping. Powder Technology 78: 247–256.

18. Adullah E.C. and D. Geldart. 1999. The use of bulk density measurements as flowability indicators". Powdder technology 102:151-165.

19. Esbensen K.H. 2002.Multivariate data analysis in practice: An introduction to multivariate data analysis and experimental design, 5th edition. CAMO Process AS, Oslo, Norway:598pp

20. Eriksson L., E. Johansson, N. Kettaneh-Wold and S. Wold. 1999. Introduction to multi- and megavariate data analysis using projection methods(PCA  PLS). Umetrics AB, Mea, Sweden: 490pp

21. 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.

22. Pettersen, C. The chemical composition of wood. In: Rowell, R.M., ed. The chemistry of solid wood. Advances in chemistry series. 207 Washington D.C: American Chemical Society; 1984: chapter 2.

23. Nkansah, K.; Dawson-Andoh, B. Rapid Characterization of Biomass: Part2. Near Infrared Spectroscopy characterization of the physical and chemical properties of northern red oak (*Quercus rubra*). 2009: Unpublished.

24. Bro, R. Multiway calibration. Multi-Linear PLS. J. Chemometrics. 1997, 10:47-62.

25. Bro, R. PARAFAC: Tutorial and applications. Chemom. Intell. Lab. System. 1997, 38: 149-171.

26. Bro, R. Multi-way analysis in the food industry: models, algorithms, and applications. Ph.D. Thesis. University of Amsterdam, Netherlands. 1998.

27. Wold, S.K., K. Esbensen and P. Geraldi. 1987. Principal component analysis. Chemom. Intell. Lab. Systems. 2:37-52

28. Martens, H. and T. Naes. Multivariate calibration. Second Edition. John Wiley & Sons, N. York, USA. 1993.

# CHAPTER SIX: Conclusion and recommendation

## Conclusion

The results from this study have established that NIRS and FS can be used online/inline to accurately predict some physical and chemical properties of northern red oak and yellow-poplar. First derivative preprocess of the NIR spectra greatly improved prediction performance of calibration models as compared to similar prediction models developed with the raw NIR spectra of both species. This allowed the much reduced spectra region (1300-1800nm) to be used to predict all chemical and physical properties with comparably similar accuracy as the full NIR spectra. This finding establishes the possibility of building a cheap and portable NIR spectrometer with a reduced wavelength region. Ash content of northern red oak was poorly predicted and the reason was attributed to subtle variance within the laboratory measured values. An increase in sample population may improve the prediction performance of the calibration models of ash content.

The findings in this study also established the potential of using fluorescence spectrometer, which is a more portable, cheaper and sensitive spectroscopic tool when coupled with the appropriate MVDA method could be used as a rapid PAT method to predict important physical and chemical properties of biomass. With the exception of holocellulose content, predicted physical and chemical properties of yellow-poplar exhibited moderate to high model accuracy. Similar results were exhibited in the physical and chemical properties prediction models of northern red oak with the exception of ash content, which were poorly predicted.

Generally the full NIR spectra-based prediction models of all variables considered in this study for both northern red oak and yellow-poplar performed better than the full fluorescence spectra-based models of similar variables. Generally performance of fluorescence spectra-based prediction models showed the potential of FS being used as an accurate prediction PAT tool which is comparable to the performance NIRS based models. An enhance MVDA method of analysis such as three way PARAFAC method may improve the performance of the fluorescence spectra-based models as compared to the two way MVDA analysis used in this study.

## Recommendation

In order to effectively utilize the findings of this study outside a controlled environment (laboratory), the robustness of prediction model needs to be tested and improved. The findings of this thesis may be improved by:

- Increasing the number of samples used in this study. This will help better understand the reasons why NIR-based prediction models of ash content in northern red oak were poor. Similarly holocellulose content prediction models developed with fluorescence spectra may also be evaluated.

- Utilizing an enhanced MVDA tool such as three way PARAFAC method for analyzing fluorescence spectra. Three way analysis may effectively decompose the excitation and emission wavelength region. This will invariably improve the models as compared to two way analysis models.

- Actual field testing (industrial) of calibration models may be necessary in other to utilize it commercially.