2004

# Modeling longleaf pine (Pinus palustris Mill) wood properties using near infrared spectroscopy

Brian Kipling Via
*Louisiana State University and Agricultural and Mechanical College*

# MODELING LONGLEAF PINE (*PINUS PALUSTRIS* MILL) WOOD PROPERTIES USING NEAR INFRARED SPECTROSCOPY

**A Dissertation**
**Submitted to the Graduate Faculty of the**
**Louisiana State University and**
**Agriculture and Mechanical College**
**In Partial Fulfillment of the**
**Requirements for the degree of**
**Doctor of Philosophy**

**in**

**The School of Renewable Natural Resources**

**by**
**Brian K. Via**
**B.S. Virginia Polytechnic Institute, 1996**
**M.S. Virginia Polytechnic Institute, 1998**
**December 2004**

# ACKNOWLEDGEMENTS

My first acknowledgement goes out to my fiancée, Paula Sallas for sticking with me throughout this ordeal, providing more financial support to the family than myself and placing her goals to go to school aside until after I finished school.

Numerous other individuals provided invaluable help during my dissertation. I would like to thank my advisor Dr. Todd Shupe for his assistance and guidance both professionally and as a colleague. From him, I learned to set deadlines and balance research and time constraints without taking shortcuts. He also guided me through various extraneous journal articles not submitted for my dissertation. I would like to thank my co advisor, Dr. Michael Stine for his advice and diversity, which made this dissertation unique from many traditional wood science disciplines. Dr. Les Groom and Dr. Tom Dean deserve attention for their explicit and implicit stressing of quality over quantity, respectively. They not only influenced the quality of my lab work but also unsuspectingly inclined me to send my work to journals abroad, which gave me an international view of scientific expectations. Dr. Chi So also deserves thanks; he managed the lab in Pineville, La. and provided a significant portion of the data for my project. I would also like to thank Mr. Jim Roberds, Research Genticist at USDA Forest Service, Saucier, MS., for his extensive advice on how to do data analysis, by phone and in person. I also need to thank Mr. Roberds crew, led by Larry Lott, for their collection of the increment cores.

A thanks goes out to Dr. Ian Hartley who was my advisor the first year of my Ph.D. His recommendation on me loading up with graduate class work at Mississippi State was helpful to me when I transferred to LSU and enabled me to graduate in just over 3 years.

But also, he influenced me to send my journals abroad and to pursue fundamental relationships through theoretical means.

Dr. Robert Purnell deserves thanks. He is my colleague and was like an advisor for several years at International Paper. He influenced me to always pay attention to quality despite company pressures to produce results. He stressed the importance of large sample size in experimentation which carried on to my work here. Also, he was constructively critical of using NIR to measure microfibril angle and tracheid length and pushed me to find a fundamental explanation for the measurement.

SangYeob Lee, Xiaobo Li, Cheng Piao, Hui Pan, and Yaojian Liu deserve thanks for listening and advising me on tough decisions. From the USDA Forest Service in Pineville, La, Donna Edwards, Gary Smith, and Karen Reed are thanked for their preparation of increment cores, measurement of density, modulus of elasticity, modulus of rupture, and collection of the spectra. They were of particular importance to the success of this project. Also of particular importance to the success of this project was Larry Lott, Lynn Lott, Gay Flurry, E. J. Anderson, and Jack Schonewitz.

I thank Ms. Pat Lefeaux for her assistance and kindness around the office and wish her luck with her love for floral design. I would like to thank Mike Kaller for his statistical insight on experimental design, regression, and SAS. I would like to thank Drs. Wu, Vlosky and deHoop for their influence.

# TABLE OF CONTENTS

ACKNOWLEDGEMENTS…… ii

LIST OF TABLES…….. vii

LIST OF FIGURES…… viii

ABSTRACT…… xi

CHAPTER

1 INTRODUCTION……. 1

2 MULTIVARIATE MODELING OF DENSITY, STRENGTH, AND STIFFNESS FROM NEAR INFRARED SPECTRA FOR MATURE, JUVENILE, AND PITH WOOD OF LONGLEAF PINE (*PINUS PALUSTRIS* MILL)…… 9

2.1 Introduction……... 9

2.2 Materials and Methods…….. 12

2.2.1 Sample Preparation…….. 12

2.2.2 Mechanical Testing…….. 13

2.2.3 Categorization…….. 14

2.2.4 NIR Spectroscopy…… 14

2.2.5 Multivariate Analysis……... 15

2.3 Results and Discussion…….. 17

2.3.1 Density……. 17

2.3.2 MOE and MOR Response to Density…….. 20

2.3.3 Whole Tree Regression Models……... 21

2.3.4 Model Stability……. 24

2.3.5 Modeling Pith Wood…… 25

2.3.6 Predicting Juvenile Properties from the Mature Model……... 26

2.4 Conclusions……... 28

3 ABILITY OF NEAR INFRARED SPECTROSCOPY TO RAPIDLY MONITOR WOOD DENSITY DISTRIBUTION AND VARIATION……... 30

3.1 Introduction……... 30

3.2 Methods and Materials…….. 33

3.3 Results and Discussion…….. 35

3.3.1 Model Validation……. 35

3.3.2 Distribution Modeling…….. 37

3.3.3 Dispersion Modeling…… 40

3.3.4 Discrimination of Density…… 44

3.4 Conclusions……... 46

4 TRACHEID LENGTH PREDICTION IN *PINUS PALUSTRIS* BY MEANS OF NEAR INFRARED SPECTROSCOPY: THE INFLUENCE OF AGE……. 47
4.1 Introduction……... 47
4.2 Materials and Methods…….. 49
4.2.1 Sample Preparation……... 49
4.2.2 Tracheid Length Measurement……. 50
4.2.3 NIR Spectroscopy and Analysis……... 51
4.3 Results and Discussion…….. 52
4.3.1 Whole Tree Modeling……... 52
4.3.2 Age and Height Modeling…… 56
4.4 Conclusions……... 60

5 RELATIONSHIP BETWEEN LIGNIN CONTENT AND MICROFIBRIL ANGLE: THEORETICAL, EMPIRICAL, AND NEAR INFRARED MODELING…….. 62
5.1 Introduction……... 62
5.2 Methods and Materials…….. 64
5.2.1 Sample Preparation……... 64
5.2.2 Chemical Analysis…… 65
5.2.3 Microfibril Angle Measurements……. 65
5.2.4 NIR Spectroscopy……. 66
5.2.5 Calibration Development…….. 66
5.3 Theoretical Model Development……... 67
5.4 Results……... 71
5.4.1 Trait Variation within Trees……. 71
5.4.2 Lignin Content to Microfibril Angle Relationship……... 74
5.5 Discussion…….. 76
5.5.1 Wood Property Variation……. 76
5.5.2 Comparison of Models……. 87

6 THE SENSATIVITY OF NEAR INFRARED SPECTROSCOPY TO TWO WOOD BLUE STAINING FUNGI…….. 82
6.1 Introduction……... 82
6.2 Materials and Methods…….. 85
6.2.1 Increment Core Sampling……. 85
6.2.2 Inoculation and Colonization of *O. minus* and *L. serpens*……... 86
6.2.3 Processing and NIR Scanning of Increment Cores…….. 88
6.2.4 Statistical Analysis……... 88
6.3 Results……... 89
6.3.1 Clear Wood Control versus Stained Wood…….. 89
6.3.2 *L. serpens* versus *O. minus* Treatment…….. 92
6.4 Discussion…….. 93
6.4.1 The Influence of Stain on Spectra…… 93
6.4.2 Differences in Absorbance between *L. serpens* versus *O. minus*… 98
6.4.3 Within and Between Core Variation…… 98

7 PREDICTION OF WOOD MECHANICAL AND CHEMICAL PROPERTIES IN THE PRESENCE AND ABSENCE OF BLUE STAIN AND USING TWO NIR MACHINES……... 100
7.1 Introduction……... 100
7.2 Methods and Materials…….. 102
7.2.1 Sample Preparation……... 102
7.2.2 Mechanical Testing…….. 104
7.2.3 Chemistry Determination……. 105
7.2.4 NIR Spectroscopy……. 105
7.2.5 Multivariate Analysis……... 106
7.3 Results and Discussion…….. 107
7.3.1 Blue Stain Classification……..... 107
7.3.2 Model Building……. 109
7.3.3 Influence of Stain on Test Population for the Slave Machine Scenario……. 113
7.3.4 Verification of Test Population for the Slave Machine Scenario…. 119
7.4 Conclusions……... 121

8 CONCLUSIONS…… 123

REFERENCES……..... 126

APPENDIX: LETTER OF PERMISSION FOR CHAPTER 2…….. 140

VITA…… 141

# LIST OF TABLES

Table 2.1: Associated wavelengths with cellulose, hemicellulose, lignin, kappa number, and hot water extractives……... 11

Table 2.2: A comparison of MLR and PCR model statistics where $PCR_1$ was performed on all wavelengths and $PCR_2$ was performed on the wavelengths listed in Table 2.1…….. 22

Table 2.3: A comparison of regression coefficients for $PCR_1$ and $PCR_2$ models in predicting MOR with standard errors in parentheses……... 25

Table 2.4: A comparison of $MLR_2$ and $PCR_{1,2}$ models for near pith associated wood (n=60)……. 27

Table 3.1: Percent classification by NIR at different bin widths (different precision allowances)……... 44

Table 4.1: Coefficient and error estimates for the principal component (PC) regression model……. 53

Table 5.1: Coefficients, $R^2$, and root mean square error of prediction for all models……. 72

Table 5.2 Experimental mean, standard deviation, and coefficient of variation for lignin for 10 trees at different height and ring numbers. N.A. represents no experimental data due to lack of material……... 73

Table 5.4: Experimental mean, standard deviation, and coefficient of variation for extractives for 10 trees at different height and ring numbers…….. 75

Table 6.1: Chemical or polymer assignments to near infrared wavelengths…….. 84

Table 6.2: The influence of *O. minus* (O), *L. serpens* (L), and control (C) treatment on absorbance between the wavelengths of 350-2500 nm. Significance differences was determined at the$\alpha = 0.05$ level (p-value < 0.05)……. 97

Table 7.1. PCR and MLR calibration and validation correlations for lignin, extractives, modulus of elasticity, modulus of rupture, and density……... 110

Table 7.2. A comparison of variance and mean difference in 5 traits for stain and control treatment within test population 1 and test population 2 using the same MLR model……. 120

# LIST OF FIGURES

Figure 1.1: The electromagnetic spectrum which encompasses all wavelengths between gamma to radio waves. The near infrared region falls just above the visible light region…… 2

Figure 2.1: The breakdown of the tree into disks and increment cores……... 13

Figure 2.2: The response of NIR spectra to wood density (g/cm$^3$) categories where sample size=10……. 15

Figure 2.3: The response of density to the area under the spectra curve with regression and 95% confidence intervals ($R^2$ = 0.71).……. 19

Figure 2.4: The response of density to the ratio of cellulose (2230 nm) to lignin (1708 nm) wavelengths with regression and 95% confidence intervals ($R^2$ = 0.38)… 20

Figure 2.5: The response of MOE to density for pith, juvenile, and mature wood….. 21

Figure 2.6: A residual plot of (a) juvenile wood density using a mature wood density MLR model, (b) juvenile wood MOE using a mature wood MLR model, and (c) juvenile wood MOR using a mature wood MLR model…….. 29

Figure 3.1: Regression analysis plot of air dry density from the area under the spectra curve versus actual density (n=170) ($R^2$ = 0.71)……..... 36

Figure 3.2: Histograms for actual and predicted density for (a) calibration (n=170) and (b) validation data (n=93)…….. 38

Figure 3.3: Actual standard deviation for samples grouped into threes versus that predicted by NIR models (a) calibration (n=170) and (b) validation data (n=93)….. 41

Figure 3.4: Sample size versus estimated standard deviation with the known standard deviation, of the population, plotted as a dashed line…… 43

Figure 3.5: Histogram of density standard deviations for all samples demonstrating the non-normal distribution…….. 43

Figure 4.1: Predicted versus measured tracheid length for calibration and validation samples ($R^2$=0.72)…… 53

Figure 4.2: Wavelength versus 1$^{st}$ derivative spectra of calibration samples (n = 300)……... 54

Figure 4.3: Classification of juvenile and mature wood using principal component 2 versus 3 in a scatter plot……... 55

Figure 4.4: Regression coefficients for principal component 2 across all heights for different age groups…….. 57

Figure 4.5: Diagram showing the $R^2$ value (in boxes) of predicting tracheid length from NIR spectra at each ring and height combination (n = 20 per box)…….. 58

Figure 5.1: Cross section view of microfibril with cellulose elementary fibers in white, hemicellulosic matrix in black, and lignin acting as a sheath (Simplified model and distances adapted after Heyn (1969); Neiduszy and Preston (1970); Fengel (1970); and Donaldson (2001). Note that the $G_{low}$ and $H_{high}$ represent the growth vector of lignin.……..... 67

Figure 5.2: Theoretical growth showing the M vector, radial or tangential growth, and the G vector, growth of lignocellulosic matrix perpendicular to microfibril angle, for (a) zero microfibril angle and (b) some microfibril angle greater than 0 degrees. The target growth plane is assumed to parallel to the tracheid axis……. 68

Figure 5.3: Theoretical sum thickness of lignin, in the radial or tangential direction, for a single microfibril within the S2 layer. The non-linear equation derived from this figure was used for eq. 5.4 while the dashed line was provided to show the near linearity in the data……... 71

Figure 5.4: A scatter plot of experimental microfibril angle and lignin content…….. 75

Figure 5.5: Log maximum likelihood results from Box-Cox transformation procedure with microfibril as the y variable and lignin % as the x variable…… 76

Figure 5.6: A scatter plot of predicted versus actual microfibril angle for the theoretical and near infrared spectra model……. 80

Figure 6.1: a) Absorbance between wavelengths 350 through 2500 for untreated wood and wood treated with *L. serpens* and *O.minus,* b) 1st derivative, and c) 2nd derivative…….. 91

Figure 6.2: a) Absorbance between wavelengths 350 through 2500 for paired data where the spectra of stained wood was subtracted from clear wood to yield the spectra of the *L. minus* and *O. minus*, b) 1st derivative of spectra, and c) 2nd derivative of spectra……. 94

Figure 7.2: The prediction of density for stained (n=101) and unstained (n=101) wood using (a) PCR and (b) MLR……... 114

Figure 7.3: The prediction of MOE for stained (n=101) and unstained (n=101) wood using (a) PCR and (b) PCR from 1st derivative spectra……. 116

Figure 7.4: The prediction of extractives for stained (n=101) and unstained (n=101) wood using (a) MLR and (b) PCR from 1st derivative spectra…… 117

Figure 7.5. The prediction of lignin for stained (n=101) and unstained (n=101) wood using (a) MLR and (b) PCR……... 118

# ABSTRACT

This research demonstrated model development for important wood properties using near infrared spectroscopy (NIR); it considered the effect of outside sources of error, and the ability of NIR to measure fiber morphology.

Strength, stiffness, and density were successfully modeled from wood samples taken throughout 10 longleaf (*Pinus palustris* Mill) trees. Principal components and multiple linear regression were compared for performance in prediction of density, strength, and stiffness. I found both modeling techniques to yield similar prediction accuracies. However, I found that density could be estimated through Beer-Lambert's law since the absorbance at all wavelengths increased with density. Also, 5 of 6 wavelengths needed to predict strength were also needed to predict stiffness lending support that similar chemical morphology controls the covariance between strength and stiffness.

Klason lignin, extractives, and microfiber angle (MFA) were also measured throughout the tree. I found extractives, lignin, and MFA to decrease from the pith outward regardless of height. A theoretical model was built attempting to explain how lignin content and MFA co-vary. Theoretical and empirical spectroscopic models both predicted MFA with nearly similar root mean square error and supported that lignin was a probable factor responsible for the covariance in spectra with MFA.

Tracheid length was another secondary trait investigated. I demonstrated that tracheid length could be predicted with an $R^2$ of 0.71 due to NIR spectra response with age. Accurate tracheid length prediction was possible due to systematic variation of chemistry with age except for at ring 1 and 4 where some other unknown factor was responsible.

Finally, blue stain and machine variability were investigated as two sources of extraneous error. It was of interest to know if the common extraneous error would bias a prediction equation. With proper modeling, I found I could avoid the blue stain variation present in the spectra when modeling modulus of elasticity (MOE), modulus of rupture (MOR), density, lignin, and extractives. However, when a calibration was built from one machine and then applied to a population of scans made from a separate machine, blue stain became problematic and prediction of MOE, MOR, and extractives were biased.

# CHAPTER 1

# INTRODUCTION

Longleaf pine (*Pinus palustris* Mill) is one of the primary southern yellow pine (SYP) species and has greater specific gravity (SG) characteristics when compared to similar aged competing SYP species. Some other major SYP species include loblolly (*Pinus taeda* L.), slash (*Pinus elliottii* Engelm.), and shortleaf (*Pinus echinata* Mill.). Over the last half of the 20$^{th}$ century, loblolly became the dominant species in the South. Most landowners of loblolly pine are pulp and paper companies who also own sawmills and/or composite product mills. As a result, many companies with significant landholdings have performed private research to improve the genetic stock of their loblolly pine resource. Such research reduces costs in the mill and/or improves final product quality. While comparative studies are not available, this has, to some unknown degree, closed the gap in SG between longleaf and loblolly pine.

Today, modeling from the NIR spectral range has shown success in rapidly screening trees for fiber morphology and chemistry. However, the spectra response will vary between species. To date, studies have been performed on species such as loblolly and Monterey pine. However, no one has investigated the ability of NIR to rapidly assess fiber morphology and chemistry in longleaf pine. Since many genetics studies require 500 to 2000 sample trees, this type of technology development is needed for longleaf pine since wood quality response is lacking in the literature.

Some of the most important traits that might be measured by NIR are MFA, tracheid length, lignin, extractives, density, strength, and stiffness. A change in any of these traits will directly influence final product quality whether it be pulp, paper, lumber, oriented

strandboard, plywood, or other composite products. For example, for pulp and paper, an increase in density will increase pulp yield and tear but decrease burst and compression. For solid wood products, an increase in density will increase modulus of rupture (MOR) and modulus of elasticity (MOE). Additionally, a change in density or microfibril angle will influence shrinkage properties (Shupe et al. 1996; Shupe et al. 1998).

There is substantial interest in the improvement of wood properties through genetic selection or a change in silviculture prescription. However, the ability to measure many traits is traditionally difficult and time consuming. As many as 1000 to 2000 trees may be measured in a single study making NIR an attractive measuring tool for many wood properties, particularly if it is accurate and rapid (So et al. 2002). NIR can be described as the region just above the visable region between 700nm to 2500 nm (Figure 1.1).

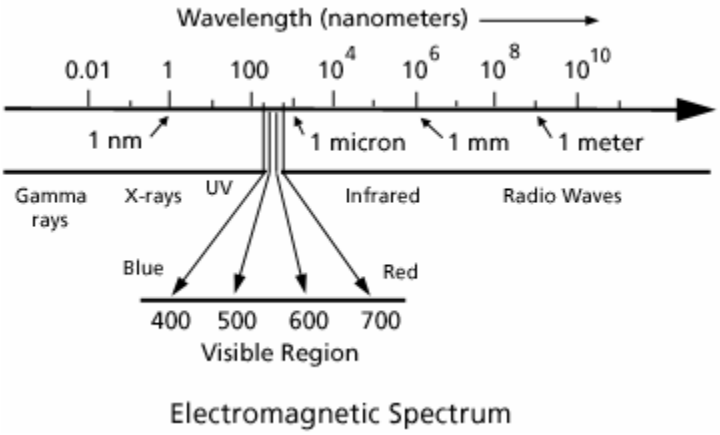



Figure 1.1: The electromagnetic spectrum which encompasses all wavelengths between gamma to radio waves. The near infrared region falls just above the visible light region. The picture was taken from Patterns in Nature (2004).

The precision of NIR is comparable or better than many other non-destructive techniques when estimating wood properties. Ultrasonics is an important alternative for rapid measurement of MOE and MOR. For increment cores, the ultrasonic models for

MOE resulted in an $R^2$ of 0.85 while for MOR the $R^2$ was 0.62 (Yang and Fortin 2001). For tracheid length, more direct methods are available which provides $R^2 > 0.95$ but traditionally take 10 to 20 minutes per sample (Pot et al. 2002). For MFA, $R^2 > 0.98$ are possible with x-ray diffraction but the equipment is costly, extensive expertise is needed, and it still takes 10 to 20 minutes per sample (Cave 1966). For lignin and extractives, the absorbance at infrared wavelengths can be used but the sample must be destroyed (Sun and Tomkinson 2002; Agarwal et al. 2003).

Alternatively, NIR can be used to estimate traits if the precision is acceptable. From my experience, measurements per sample can take 30 to 90 seconds depending on acquisition and operator speed. Strength and stiffness (MOR and MOE) are two important parameters to improve in solid wood products while density, lignin, and extractives variation influence paper yield and strength, all of which can be measured my NIR with $R^{2'}$s generally greater than 0.80 (Schultz and Burns 1990; Sanderson et al. 1996; Schimleck et al. 2001; Schimleck et al. 2002; Bailleres et al. 2002; Gierlinger et al. 2002; Schimleck and Evans 2003). Increased extractives and lignin can lower pulp yield while reduced extractives improves the bonding capacity between fibers. Density improvement in pulpwood is a bit challenging since an increase in density increases pulp yield but decreases paper strength (Via et al. 2004). For solid wood, increased wood density is positively correlated with MOE and MOR making density improvement an important goal. In New Zealand, wood cants have been scanned with NIR and classified for MOE with success (Meder et al. 2003). The ability of NIR to measure MOE and MOR is attributable to the absorbance of light by lignin and cellulose at specific

wavelengths and nondestructive measurements are quite successful on small clear wood (McLellan et al. 1991ab; Gindl et al. 2001; Rials et al. 2002; Kelly et al. 2004ab).

There is also growing interest in using NIR to measure secondary traits such as MFA and tracheid length. Specific wavelengths for tracheid length have not been reported while the following wavelengths has shown to vary with MFA: 1470, 1510, 2458, 2326, and 2082 nm (Schimleck and Evans 2002ab). Also, three of the five wavelengths were sensitive to cellulose suggesting that MFA may partially correlate to cellulose percent (Schimleck and Evans 2002ab). Perhaps during the preparation of the specimen, more cellulose is exposed with decreasing MFA which could in turn influence near infrared spectra (Gindl et. al. 2001).

Spectroscopic measurements were able to account for 70 to 77% of the variation of various tracheid length index measurements on 48-year-old trees where disks of Norway spruce were analyzed (Hauksson et al. 2001). For MFA, spectra accounted for 66 to 83% of the variation (Schimleck et al. 2001ab). Possibly, the ability of NIR to measure MFA was due to systematic alpha-cellulose variation within the tree (Schimleck et al. 2001a). However, intercorrelations between alpha-cellulose, MFA, and tracheid length within a ring at a fixed height (across trees) are needed to remove any confounding data due to tree age and height. For example, Bergander et al. (2002) showed MFA and tracheid length do not correlate when both traits were measured from the same ring of Norway spruce. From pith to bark, Schimleck et al. (2000) demonstrated that different NIR predictive models were needed at different tree heights. This is compounded by the three way interaction of live crown, specific gravity, and MFA (So et al. 2002). Since near infrared can only directly measure the vibration of bonds attributable to major chemical

constituents, it is probable that measurement of MFA and tracheid length are due to a secondary correlation with chemical constituents.

Finally, NIR precision and accuracy is only as good as the samples that represent the calibration set. A large, wide, and uniform distribution is recommended by the author during the calibration stage. Additionally, all sources of variation should be accounted for in the calibration data set. Blue stain is one source of error that may increase the variation in spectra response Lebow et al. 1996; Zulpa et al. 2003). Such influence on the spectra may bias or diminish the ability to model a trait, particularly if the calibration data were collected in a highly controlled laboratory. Also, many laboratories use two or more NIR machines to hasten scan times for a large project such as a tree improvement study. As such, proper transfer of models or independent models is often needed (Fearn 2001). Therefore, it is important to consider the co-effect of machine and stain when measuring samples.

There are many analysis techniques for spectra data but the techniques can be generalized into two major categories: Ordinary least squares regression and biased regression techniques. Least squares regression has been preferred in many situations because the predicted variable can be explained by specific wavelengths. For example, Law and Trachuk (1977) were able to determine the moisture content in wheat with an $r^2$ = 0.98 using simple linear regression of two wavelengths known to represent moisture.

When using NIR, consideration should be given for how to pretreat the spectra and develop calibration models. For biological systems, intercorrelations may exist between wood properties. Furthermore, different wavelengths in most chemical systems intercorrelate. Such correlations between wavelengths can cause problems in modeling.

One way to overcome this is to take the first or second derivative of the raw spectra, which reduces the dependence between wavelengths. Such techniques improve the correlation between independent and dependent variable regardless of whether a regression or other multivariate approach is used afterward. Meder et al. (2002) used the first derivative and improved their $R^2$ from 0.59 to 0.72 in predicting MOE for spectra taken from mini-LVL panels.

One way to eliminate correlations between wavelengths is to use a multivariate technique called principal components analysis (PCA). PCA is a method that reduces a large number of strong multicollinear wavelengths into a smaller number of new orthogonal variables which hopefully retains most of the descriptive variation of the independent variable. These new variables are called eigenvectors. In the analysis, the eigenvector is used to describe the level of variation attributable for each independent variable so that the independent variables may be ranked by contribution to overall variation. The goal is to utilize the fewest principle components (eigenvectors) to account for the majority of variation. The eigenvectors and scores (weighting values for the eigenvectors) matrix are composed and provide a mutually independent axes to describe the data set. Partial least squares (PLS) may then be used to describe correlations.

PLS, in conjunction with PCA, is perhaps the most popular method for developing calibration equations between spectra and dependent variables in paper and forest products (Antti et al. 2000; Antti et al. 1996; Hauksson et al. 2001; Hodges and Krishnagopalan 1999; Meder et al. 2002; Raymond and Schimleck 2002; Wallbäcks et al 1991). The method of PLS is used to create new variables from the original spectra and

then regressing the new variables with the property of interest. It is thus a regression technique that overcomes mutilcollinearity common between the spectra of original data.

One result of PLS is that a small bias is introduced into the derivation of the model equation (Faber 1999). While this may seem less than optimal, these models tend to offer superior error of prediction or calibration versus multiple linear regression (MLR) (the error that occurs when one regress the calibration model with a fresh set of data) under most circumstances (Kowalski 1990; Boister et al. 1996). Furthermore, they offer more realistic equations under extrapolation circumstances while the error of prediction term for MLR models are inflated suggesting more uncertainty (Estienne et al. 2001).

When extrapolation does not occur in the population to be estimated, than MLR can provide an alternative to PLS as long as confidence limits are of no concern. The advantage to MLR is that more interpretation is possible since one can explicitly see the relationship between a wavelength and the dependant variable.

The best way to determine which model best estimates the response is to regress the calibration model using data not in the calibration model (cross validation procedure). Models with significantly lower error of prediction will usually yield more stable predictions (Estienne et al. 2001). When similar prediction errors occur, the more interpretable model should be chosen.

For this dissertation, I looked at the ability of NIR absorbance to predict lignin, extractives, density, modulus of elasticity, modulus of rupture, MFA, and tracheid length. I also tried to interpret the reason for NIR absorbance being sensitive to MFA and trachied length variation. Finally, I investigated the ability of using NIR absorbance to

predict wood chemistry, density, and strength when blue stain and a second machine was introduced causing additional noise in the spectra.

# CHAPTER 2

# MULTIVARIATE MODELING OF DENSITY, STRENGTH, AND STIFFNESS FROM NEAR INFRARED SPECTRA FOR MATURE, JUVENILE, AND PITH WOOD OF LONGLEAF PINE (*PINUS PALUSTRIS* MILL)[1]

## 2.1 Introduction

Softwood lumber is common in building construction where availability and cost are balanced with acceptable strength and stiffness. The juvenile portion of the tree typically exhibits unacceptable strength and stiffness. Juvenile wood comes from xylem wood produced several years after apical meristem production. As the cambium matures the mechanical properties of the wood increase. As a result, development of rapid assessment tools to monitor the mechanical properties of wood may be useful to differentiate between juvenile and mature wood. Proper models would enhance quality control, wood utilization efficiency, research and development, and the theoretical understanding of causal relationships of mechanical properties.

The absorbance at some wavelength(s) in the near-infrared (NIR) region is associated with the fundamental chemical components of wood, which are, in turn, directly or indirectly related to mechanical properties. An example of an indirect relationship is the apparent connection between cellulose associated wavelengths and MFA, a property that impacts modulus of elasticity (MOE) (Schimleck et al. 2001ab). An example of a more direct relationship would be if wood chemistry and MOE were related. For example, for single wood fibers higher in cellulose, a higher response in

[1] Reprinted by permission of the Journal of Near Infrared Spectroscopy.

MOE is present than single wood fibers higher in lignin (Sakurada et al. 1962). Perhaps this plays a role for an aggregate collection of fibers connected by lignin.

The most important macro anatomical characteristic that contributes to strength and stiffness is density. The main determinate of density is well accepted to be the relative amount of lumen to cell wall material present in wood. This explains why the absorbance across all wavelengths shift upward as density increases (Schimleck et al. 2002). Additionally, there may be more subtle variables to influence density that could, in turn, influence spectra response. For example, cellulose has a density of 1.52 to 1.53 g/cm$^3$ for six different softwood species while lignin has a lower density of 1.35 g/cm$^3$ (Kellogg et al. 1975; Jurasek 1995). Hemicellulose has a similar density to cellulose but is perhaps significantly altered during the extraction process, making a quantitative assessment difficult (Kellogg et al. 1975). The relative amount of lignin, cellulose, and hemicellulose largely determine cell wall density and could influence overall density.

The primary chemical components and associated wavelengths of wood are listed in Table 2.1. Because kappa number is a direct function of lignin content (Malkavoara and Alen 1998), it was included in the table and was considered as lignin associated wavelengths. The kappa number is simply a method of determining the amount of delignification that has occurred in chemical pulps. The purpose of Table 2.1 was to define which wavelengths were needed to account for wood chemistry when trying to predict strength and stiffness.

Some work has already shown NIR to successfully estimate wood density, MOE, and MOR (Hoffmeyer and Pedersen 1995; Schimleck et al. 1996; Schimleck et al. 2000; Gindl et al. 2001; Schimleck and Evans 2002; Schimleck et al. 2002; So et al. 2002;

Schimleck et al. 2003ab; Schimleck et al. IN PRESS). However, it is not well understood if different models are needed for different portions of the tree. Schimleck et al. (2000) did use Soft Independent Modeling of Class Analogy (SIMCA) analysis to show that models were significantly different at different tree heights. As a result, one may need different predictive models for pith, juvenile, and mature wood specimens.

Table 2.1: Associated wavelengths with glucose, cellulose, hemicellulose, lignin, kappa number, and hot water extractives.

| Chemical Measurement | Material or solution | Associated Wavelengths (nm) | Reference |
|---|---|---|---|
| Glucose | Human blood | 1100-1300, 1660-1830, 2050-2350 | Malin et al. 1999 |
| Glucose | Cell culture | 2100, 2270, 2320 | Lewis et al. 2000 |
| Kappa number | Softwood and hardwood kraft pulp | 457, 500, 650 | Malkavoara and Alen 1998 |
| Kappa number | *Eucalyptus grandis* kraft pulp | 1722, 1778, 1940, 2100, 2139 | Birkett and Gambino 1989 |
| Kappa number | Mixed eucalyptus kraft + soda | 1759, 1818, 2139, 2190, 2230 | Birkett and Gambino 1989 |
| Kappa number | Mixed pine kraft | 1445, 1680, 1734, 2100, 2270 | Birkett and Gambino 1989 |
| Kappa number | Mixed pine and eucalyptus kraft | 1445, 2100, 2208, 2230, 2270 | Birkett and Gambino 1989 |
| Lignin | Forest foliage 18 hardwood and softwood species | 1438, 1828, 2218, 2386 | McLellan et al. 1991a |
| Lignin | Solid wood alder, bass, cherry, pine, walnut, birch | 1600, 1660, 1505, 1460, 1275, 1230, 870 | Meglen and Kelly 2000 |
| Lignin | Quercus alba foliage | 1438, 1708, 2154, 2320 | McLellan et al. 1991b |
| Cellulose | 14 different pine species | 1722, 1734, 2230, 2310, 2236 | Wright et al. 1990 |
| Cellulose | Forest foliage 18 hardwood and softwood species | 1766, 1960, 1982, 2140 | McLellan et al. 1991a |
| Cellulose | Quercus alba foliage | 1754, 1898, 2076, 1898 | McLellan et al. 1991b |
| Cellulose | Various foods and plants including barley, cereal, forage, flour, wheat | 1207, 1278, 1365, 1431, 1487, 1584, 1707, 1772, 1824, 2088, 2200, 2273, 2336, 2347 | Osborne and Fearn 1986 |
| Cellulose and hemicellulose | Solid wood alder, bass, cherry, pine, walnut, birch | 1425, 1370, 1325, 1160, 1110, 1050, 895, 680 | Meglen and Kelly 2000 |
| Hemicellulose | Various foods and plants including barley, cereal, forage, flour, wheat | 1218, 1278, 1360, 1436, 1492, 1584, 1728, 1778, 1830, 2110, 2186, 2262, 2314, 2380 | Osborne and Fearn 1986 |
| Hot water extractives | Eucalypt wood | 2200, 1700, 1900 | Michell 1995 |

Since the choice of models can play a significant role in prediction efficiency (higher accuracy), it is important to try different modeling techniques. In the wood science discipline, the most popular models are principal components regression (PCR) and partial least squares (PLS), both of which can overcome substantial multicollinearity problems. High multicollinearity between wavelengths is a problem when estimating the variance around the regression coefficient(s) in multiple linear regression (MLR). However, given that MLR is commonly used for calibration of NIR, the variance around the regression coefficients may not matter. Kowalski (1990) found MLR to outperform PLS, PCR, and ridge regression but warned that over fitting could occur if too many wavelengths were chosen. For five separate data sets, MLR performed superiorly until extrapolation occurred (Estienne et al. 2001).

The aim of this paper was to determine if density, strength, and stiffness models could be developed with MLR, principal component regression of all wavelengths ($PCR_1$), and principal component regression of selected wood chemistry wavelengths ($PCR_2$). I also investigated the ability to model density from the area under the spectra and from individual wood chemistry associated wavelengths.

## 2.2 Material and Methods

### 2.2.1 Sample Preparation

Ten longleaf pine trees 41 years in age were selected from a plantation on the Harrison Experimental Forest, owned and maintained by the USDA Forest Service near Saucier, Mississippi (USA). The location was 30.6° north and 89.1° west. The under story of the site was free of competition because of periodic prescribed fires by the Forest

Service staff. Trees were planted approximately 3.66 meters apart from neighboring trees in an equilateral triangle pattern. Each row was spaced equally but offset with the adjacent row forming an equilateral triangle pattern between adjacent trees. Border trees were planted around the overall site. Each tree was cut every 4.57 meters in height, yielding 5 to 7 bolts, each bolt having an accompanying disk cut from the basal end of the bolt. The disk was further processed to yield bending specimens 30.48cm x random width x 1.27 cm, with the ring of interest in the center of the specimen while the disk was used for NIR spectra acquisition. Widths were random because ring thickness varied, and I desired to maintain the same ring count per specimen. I acquired spectra from the radial face from bark to pith to bark strips ripped from the disks. Figure 2.1 illustrates the partitioning of sample from different locations within the tree.

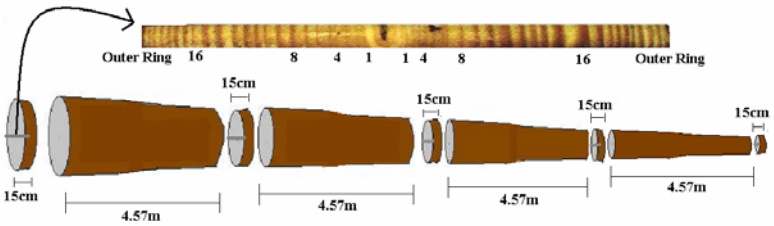



Figure 2.1: The breakdown of the trees into disks and increment cores.

2.2.2 Mechanical Testing

I conditioned the bending specimens in ambient conditions to equilibrium moisture content of 8.0% with a standard deviation of 1.3%. I applied the load in three-point bending on the tangential face at a rate of 0.20 cm $min^{-1}$ on an Instron testing machine. Stress to strain plots was obtained, and the slope of the linear portion of the curve was used to estimate MOE. Density and moisture content were determined from the bending samples. The volume for density was measured with calipers and the weight for density was measured at equilibrium moisture content. Dimensions were measured to the nearest

0.0025 cm. The oven dry and equilibrium moisture content was measured. The weights for both procedures were measured to the nearest 0.001 g.

2.2.3 Categorization

Counting from pith to bark, latewood rings 1, 4, 8, 16, 32, and the last ring on the disk were dissected from 5 to 7 disks, dependent on tree height. Typically, for specimens in the last ring, 5 to 10 rings had to be included 'as the last ring' because of much slower growth rates. I found this necessary because a cross section 1.61 $cm^2$ was needed for MOE and MOR measurement. Spectra for each sample were acquired from each disk, and each bending specimen was measured for density, MOR, and MOE. Rings were further classified into pith, juvenile, and mature wood regions. Pith wood was considered the first year of growth on all disks. For this paper, the juvenile wood zone was classified as ring 4 and 8 for the butt log only and will sometimes be referred to as simply juvenile wood. The mature wood region included the remainder of the tree.

2.2.4 NIR Spectroscopy

I acquired NIR absorbance using a Nexus 670 FTIR spectrometer (Thermo Nicolet Instruments, Madison, WI, USA) with wavelengths between 1000 to 2500 nm. Scans were acquired at 1nm intervals using reflectance spectroscopy. I collected forty scans and averaged them into a single spectrum curve from the north and south position within each tree. The north and south spectra were then averaged together to further increase precision. After averaging, a total of 263 samples/spectra were available for model building and validation. Temperature was controlled at 22 $C^{\circ} \pm 1$ in a laboratory environment with a mean relative humidity of 50%. To demonstrate the effect of density on the spectra, I measured 10 replicates per density group. The density groups were 0.40,

0.50, 0.60, and 0.70 g $cc^{-3}$. I acquired forty scans on the north and south side for a total of 80 scans and then averaged them together for one spectra curve (Figure 2.2). The 5mm spot size was placed as close to the center of the latewood ring with some overlap of the earlywood ring due to NIR spot size.

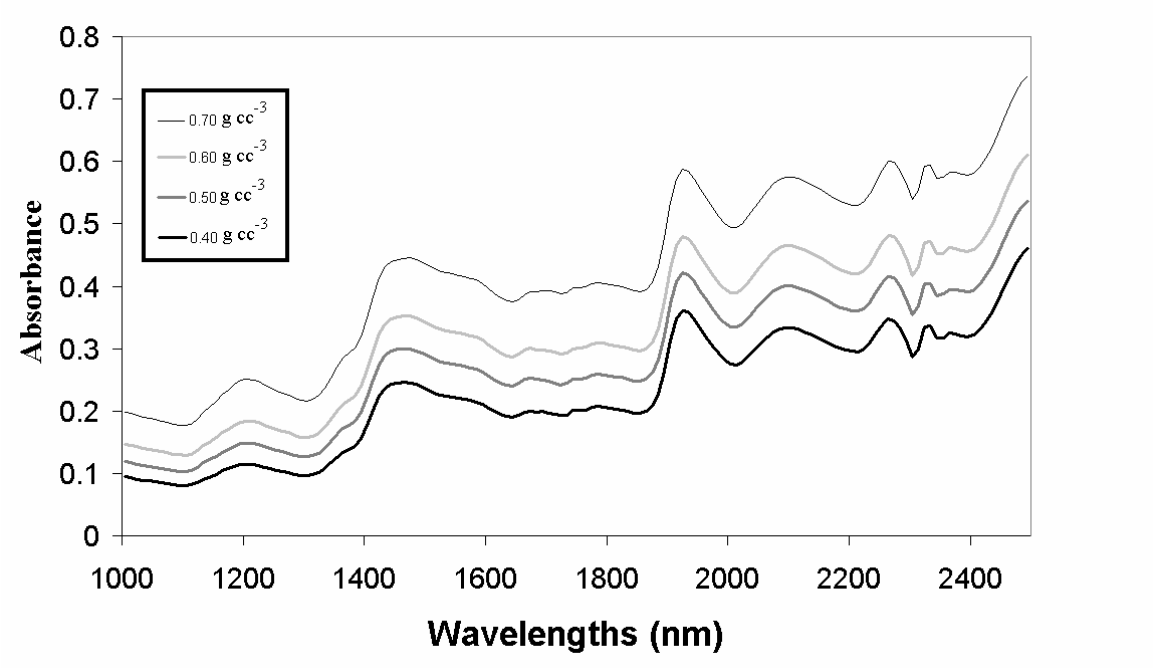



Figure 2.2: The response of NIR spectra to wood density ($g/cm^3$) categories where sample size = 10.

2.2.5 Multivariate Analysis

I performed MLR and PCR using Statistical Analysis Software (SAS 1999) in which PROC REG and PROC PRINCOMP was applied. The algorithms used can be found in manuals provided by the company (SAS 1999). To reduce computation time, I reduced the absorbance values into 10-nm intervals by averaging. Prior analysis showed nearly identical model coefficients when 10-nm intervals were compared to models developed from spectra of 1-nm intervals which agreed with other independent research (Schimleck et al. 2004). I employed regression diagnostics to compare the performance of different

models. One hundred seventy samples were used to develop whole tree models with 93 samples held aside for validation model building. Regression coefficients, root mean square error of validation (RMSEV), PRESS (predicted sum of squares), $R^2$, adjusted $R^2$, and $C_p$ were employed to compare model predictability and can be reviewed in detail by Neter et al. (1996) The equation for PRESS is

$$PRESS = \sum_1^n (Y_i - Y_{predicted}) \qquad (1.1)$$

Where $Y_i$ is the actual value and $Y_{predicted}$ is the value predicted by the model. It should be noted that the $i^{th}$ value was omitted and the regression computed. The smaller PRESS values suggested a model of better predictability. The equation for Mallows $C_p$ was

$$C_p = \frac{SSE_P}{MSE(X_1,...,X_{p-1})} \qquad (1.2)$$

Where $SSE_p$ is the sum of square error using P variables while $MSE(X_1,\ldots,X_{p-1})$ is an unbiased estimator of the variance of the fitted value $Y_{Predicted}$. $C_p$ was a function of the total mean squared error of the regression model; therefore, models with a lower $C_p$ were preferred.

In summary, I found a lower RMSEV, PRESS, and $C_p$ and a higher $R^2$ and adjusted $R^2$ to suggest a better model. It should be noted that the RMSEV, for this paper, was the root mean square error of a separate validation model and is not the root mean square error derived when regressing the calibration model to the validation data. I chose wavelengths for MLR and PCR from Table 2.1. For $PCR_1$, all wavelengths were reduced to principal components and then regressed by linear multiple regression. A second $PCR_2$ model was developed using only selected wavelengths from Table 2.1. The wavelengths were analyzed were 1505, 1665, 1705, 1735, 1825, 1905, 2155, 2205, 2235,

and 2305 nm. These wavelengths for MLR were processed through the backward stepwise selection procedure. All independent variables having a p-value less than 0.10 were maintained in the model as determined by the software default. No spectra pretreatment such as derivatives or multiplicative scatter correction was employed. While a few outliers were detected, none were highly influential enough to change the regression coefficients and thus were maintained in the data set. Given the obvious shift in spectra, with respect to density (Figure 2.2), I estimated the area (nm x absorbance) under each spectra curve using Riemann sums. Linear regression and MLR were performed using area as an independent variable. Because the Type II sum of squares is sequential, I always placed area in the density MLR model first since the baseline shift was the most important parameter in predicting density.

## 2.3 Results and Discussion

### 2.3.1 Density

Figure 2.2 demonstrates the baseline shift in absorbance that occurs with respect to a change in density. Other studies have indicated an increase in absorbance due to increased density but did not quantify the change in absorbance for given wavelength ranges (Gindl et al. 2001; Schimleck et al. 2002) In the lower wavelength region (1000 to 1400nm), increasing the density from 0.4 to 0.7 g/cm$^3$ brought about a 0.1 mean increase in absorbance (Figure 2.2). Thus, as one considers lower to higher wavelengths, the difference increases. At the higher wavelengths around 2200 nm, a 0.3 g/cm$^3$ increase in density increased absorbance by an average of 0.27. While there was increased range at the upper end of the spectrum, there was also increased variance in absorbance per wavelength. When variance in absorbance was plotted at each

wavelength for all wood samples, the variance increased sevenfold (from 0.002 to 0.014) when going from 1000 to 2500 nm. This increase in variance at higher NIR wavelengths would lower the precision of the model.

The baseline shift was primarily accounted for in the first principal component since each wavelength received the same weight within the eigenvector. It is probable that the baseline shift was attributable to some macro (anatomical) characteristic of density. For example, an increase in wheat kernel size caused an upward baseline shift in absorbance values caused by increased concentration of material (Wang et al. 1999). For solid wood density, the macro characteristic that affects density is lumen diameter and cell wall thickness. Having more cell wall material and less lumen diameter implies that more material is available for absorbance across all wavelengths in the NIR region.

Theoretically, I found the baseline shift with density can be explained by the Beer-Lambert law, a theory developed for homogeneous materials (Osborne and Fearn 1986)

$$A = Mcd \qquad (1.3)$$

Where A is absorbance, M is molar absorptivity, c is molar concentration of absorber, and d is sample pathlength. The molor absorptivity indicates how much light, at a particular wavelength, a material will absorb per unit of concentration. When density increases, if molar absorptivity is assumed to stay the same and sample pathlength does not change, then the concentration of the absorber should linearly increase with concentration or density. This relationship explains the increase in absorbance with density in Figure 2.2 and is justified despite the fact that wood is not a homogeneous material. It is noteworthy to point out that the difference in absorbance between 0.7 and

0.6 g/cm$^3$ was greater than at other 0.1 g/cm$^3$ intervals of lower densities (Figure 2.2). This deviation from the Beer-Lambert law at high concentrations is common for various materials (Osborne and Fearn 1986).

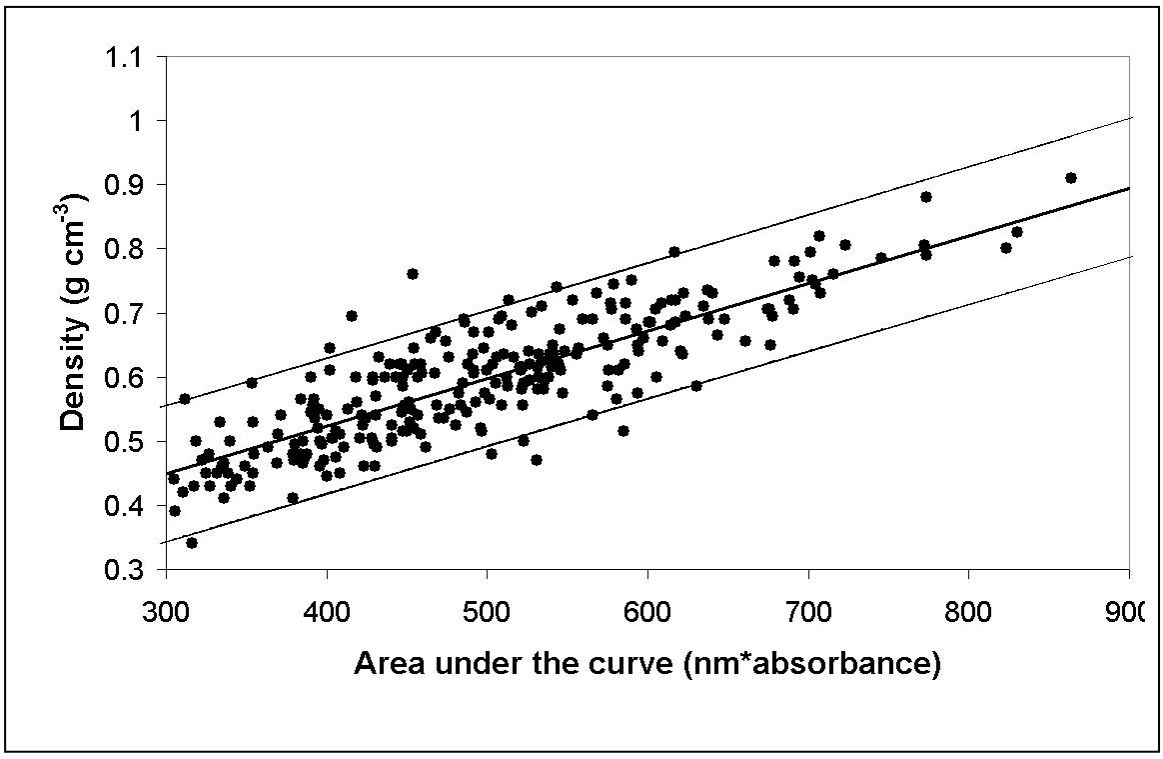



Figure 2.3: The response of density to the area under the spectra curve with regression and 95% confidence intervals ($R^2 = 0.71$).

I computed the area under the curve for the whole NIR region because it would directly relate to a baseline shift in absorbance. The area was regressed against density using least squares linear regression (Figure 2.3). An $R^2$ of 0.71 was found with balanced residuals and proved to be a viable way to model density.

Aside from macro (anatomical) density variation, there may be additional variation in density due to cell wall chemistry. Cellulose is denser than lignin; therefore, 2230 and 1708 nm wavelengths were specifically chosen for cellulose and lignin, respectively, because this combination exhibited the lowest multicolinearity as determined by Pearson

Correlation Coefficient (Neter et al. 1996). Because a baseline shift showed equal effect on the two wavelengths, taking the ratio of absorbance at two wavelengths eliminated the variation attributable to the baseline shift. Figure 2.4 shows a significant relationship ($R^2$ = 0.38) between the absorbance ratio and density, suggesting that lignin and cellulose concentration may have some influence over whole wood density which agrees with other findings (Kellogg et al. 1975).

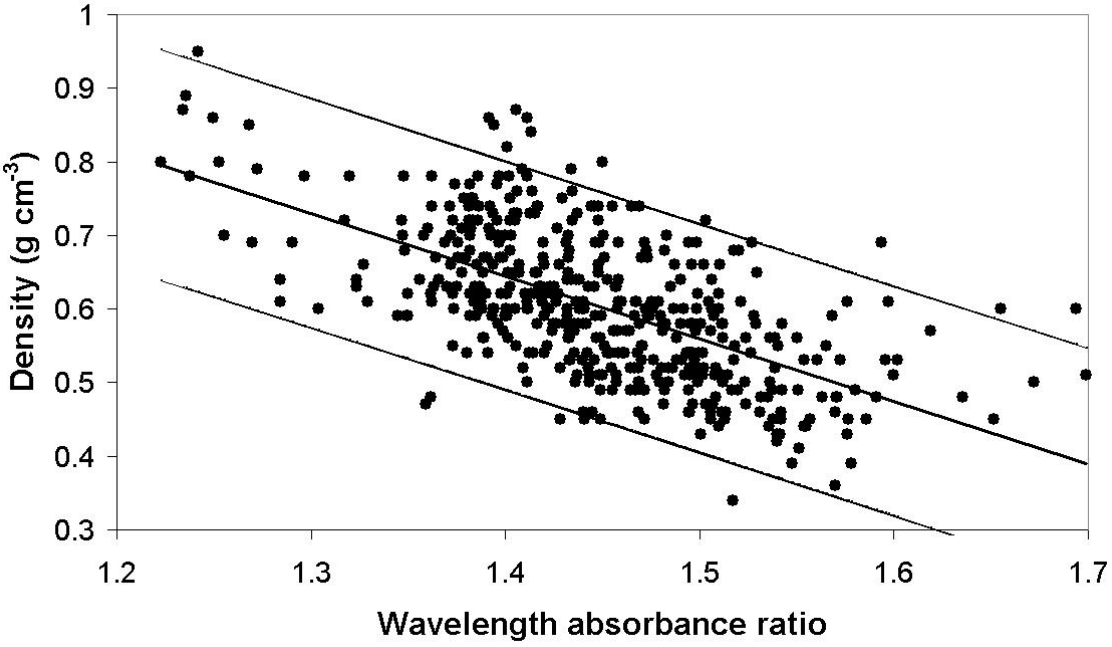



Figure 2.4: The response of density to the ratio of cellulose (2230 nm) to lignin (1708 nm) wavelengths with regression and 95% confidence intervals ($R^2$=0.38).

2.3.2 MOE and MOR Response to Density

Figure 2.5 shows covariance between density and MOE for pith, juvenile, and mature wood. A similar trend was found for MOR for the three zones but the data is not shown. In pith-wood, the density was not a significant predictor of MOE and MOR. Perhaps this lack of significance was attributable to the high concentration of resinous extractives near the pith, which could be visually observed. Excessive extractives would have increased

the apparent density even though they would not contribute to increased strength or stiffness. For juvenile wood, a significant but lower slope occurred when compared to mature wood slopes where the 95% confidence intervals of the slopes were not overlapping. This finding was similar to others who found that for loblolly pine butt logs, density accounts for less variation in MOE in the juvenile wood zone than mature wood for loblolly pine (Pearson and Gilmore 1971; Pearson and Gilmore 1980). Others have reported that MFA was the primary contributor to MOE in the juvenile wood zone (Bendtsen and Senft 1986; Treacy et al. 2000). Given that density has a decreased correlation with MOE and MOR as one goes from mature, to juvenile, to pith, it might be possible that different models are needed for each region.

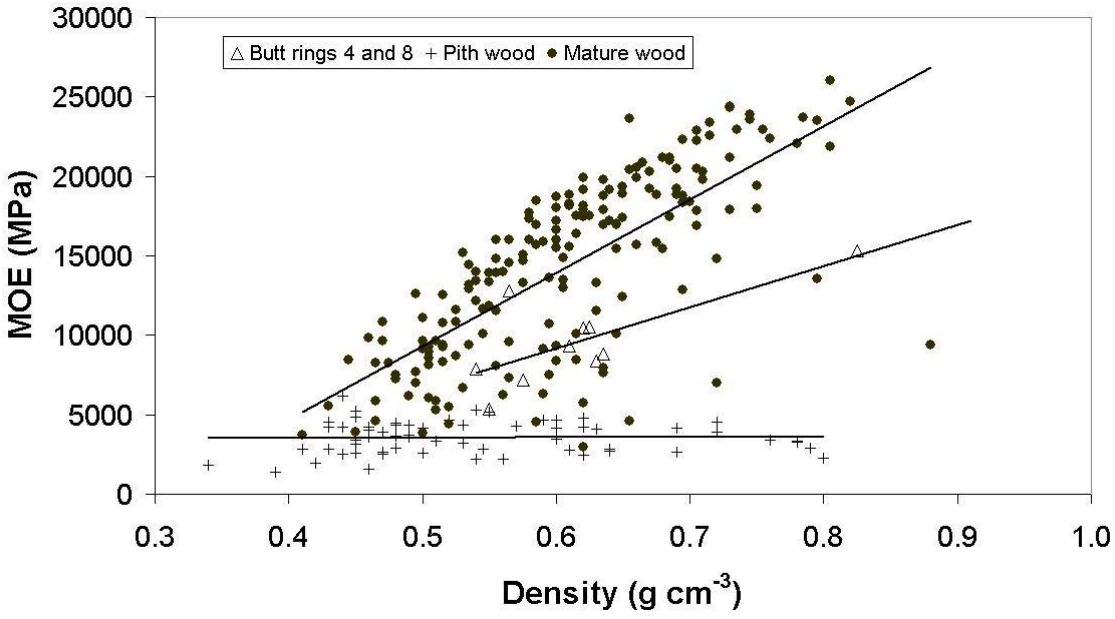



Figure 2.5: The response of MOE to density for pith, juvenile, and mature wood.

2.3.3 Whole Tree Regression Models

For density, I found MLR to perform similarly to $PCR_1$ and $PCR_2$ in $R^2$, adjusted $R^2$, RMSEC, RMSEV, and the PRESS statistic while the $C_p$ (Mallow's statistic) was slightly

higher (Table 2.2). The wavelengths chosen, aside from area and ratio, were 1705, 1825, and 2155 nm. Wavelength 1705 nm has been shown to associate with lignin, cellulose, and hot water extractives, 1825 nm with cellulose and lignin, and 2155 with lignin (Table 2.1).

Table 2.2: A comparison of MLR and PCR model statistics where $PCR_1$ was performed on all wavelengths and $PCR_2$ was performed on the wavelengths listed in Table 2.1.

| Model type | Dependent Variable | Independent Variables | $R^2$ | Adj. $R^2$ | RMSEC N=170 | RMSEV N=93 | Error Ratio | PRESS | $C_p$ |
|---|---|---|---|---|---|---|---|---|---|
| Linear | Density | Area | 0.71 | 0.71 | 0.0530 | 0.0561 | 1.06 | 0.48 | - |
| MLR | Density | Area, 1705, 1825, 2155, Ratio | 0.76 | 0.75 | 0.0485 | 0.0512 | 1.06 | 0.42 | 5.9 |
| | MOE | Area, 1705, 1825, 1905, 2155, 2205, 2385 | 0.86 | 0.86 | 351249 | 398453 | 1.13 | 2.3E13 | 5.2 |
| | MOR | Area, 1705, 1825, 2155, 2205, 2385 | 0.88 | 0.88 | 2523 | 3110 | 1.23 | 1.2E9 | 8.0 |
| $PCR_1$ | Density | 5 PC's | 0.76 | 0.75 | 0.0489 | 0.0507 | 1.04 | 0.42 | 4.0 |
| | MOE | 9 PC's | 0.89 | 0.89 | 315537 | 367562 | 1.16 | 1.9E13 | 10.5 |
| | MOR | 9 PC's | 0.89 | 0.89 | 2401 | 2964 | 1.23 | 1.1E9 | 11.5 |
| $PCR_2$ | Density | 2 PC's | 0.73 | 0.73 | 0.0510 | 0.0515 | 1.01 | 0.45 | 2.6 |
| | MOE | 5 PC's | 0.84 | 0.84 | 373869 | 422768 | 1.13 | 2.6E13 | 5.2 |
| | MOR | 5 PC's | 0.86 | 0.86 | 2680 | 3394 | 1.27 | 1.3E9 | 4.0 |

For the $PCR_1$ density model, 5 principal components were significant in predicting density, while for $PCR_2$ only 2 principal components were significant and only slightly lower in predictive ability (Table 2.2). Of all 3 regression models, MLR $\leq PCR_1 < PCR_2$ for RMSEC, RMSEV, and PRESS statistics, respectively. However, for $C_p$, MLR performed the worst and $PCR_2$ performed the best. In summary, all three regression models appeared to predict density similarly.

For prediction of MOE, the backward stepwise selection procedure did not find area under the absorbance curve to be significant (p value < 0.05). I found this surprising because density is considered to directly influence MOE. This suggests that chemistry influence on stiffness are important and perhaps underestimated. In support of this finding, I found that most of the wood chemistry associated wavelengths needed to predict density were needed to predict MOE. Also, and equally noteworthy, was that cellulose associated wavelengths may be related with MFA which in turn affects MOE, as observed by others (Megraw 1985; Schimleck et al. 2001; Schimleck et al. 2002). It was also possible that NIR was more sensitive to chemistry variation than density which could make wood chemistry appear more important than density when predicting MOE from NIR spectra.

Using $PCR_1$, nine principal components were significant in predicting MOE while $PCR_2$ only found five significant principal components. Even though Table 2.2 suggests that $PCR_1$ is better, perhaps the fewer components needed to predict MOE mean that $PCR_2$ is more stable and will be considered shortly.

Of the three regression procedures, $PCR_1 < MLR_2 < PCR_2$ in RMSEC, RMSEV, and PRESS while MLR and $PCR_2$ had the lowest $C_p$. Just as was the case for all the density models, $C_p$ did not rank the models in the same order as the PRESS statistic. In summary, no best model was conclusive, and all three appear to model MOE adequately.

For MOR, the MLR regression required one less independent variable for prediction than did for the MOE model. However, the remaining significant wavelengths needed to predict MOR were also needed to predict MOE. Just as with MOE, the area under the spectra curve was not useful in predicting MOR. Furthermore, the same wood chemistry

associated wavelengths that were related to density and MOE were also related to MOR suggesting that the same underlying chemical constituent may be responsible for both strength and stiffness.

For $PCR_1$ and $PCR_2$, both MOE and MOR required the same number of principal components. It is interesting to note that the exact same principal components were needed for both MOE and MOR for $PCR_2$, which suggests, just as with the MLR analysis, that the same underlying chemical constituent may influence strength and stiffness. Once again, all three regression models appeared to predict MOE and MOR well with acceptable RMSEV.

2.3.4 Model Stability

So far, I have found the predictive ability of $PCR_1$ and $PCR_2$ to be similar in response. However, another way to determine if a model is appropriate is to compare the regression coefficients of the calibration and validation models. Obviously, similar coefficients are desirable while vastly different coefficients suggest a potential instability and suggests over-fitting of the calibration (Neter et al. 1996). Table 2.3 lists the regression coefficients for the calibration and validation models for the two PCR models predicting MOR. MOR was chosen for demonstration because it had the most independent variables. However, density and MOE gave similar results between $PCR_1$ and $PCR_2$ with $PCR_2$ always having less variables significant.

When I compared calibration and validation regression coefficients for $PCR_1$, different regression coefficients emerged (Table 2.3). The standard errors, for each regression coefficient, were listed to gauge if coefficients significantly differed. Four regression coefficients were significantly different using $PCR_1$. The most striking

difference was that the 8th coefficient changed signs from the calibration to validation model. However, the 8th coefficient only accounted for less than 1% of the total variation. Conversely, when only wood-chemistry-associated wavelengths were used ($PCR_2$), only the 2nd coefficient appeared to significantly differ but not at the magnitude that some of the regression coefficients differed using $PCR_1$. When the potentially unstable principal components were removed from the $PCR_1$ equation, the $R^2$ value dropped from 0.89 to 0.43 for both MOE and MOR. As a result, perhaps $PCR_2$ would be more robust under extrapolation conditions or measurement of specimens from a separate population. As mentioned, density and MOE showed similar instabilities in regression coefficients for $PCR_1$.

Table 2.3: A comparison of regression coefficients for $PCR_1$ and $PCR_2$ models in predicting MOR with standard errors in parentheses.

| $PCR_1$ | | $PCR_2$ | |
|---|---|---|---|
| Calibration set | Validation set | Calibration set | Validation set |
| 14475 (184) | 15638 (307) | 14475 (205) | 15638 (375) |
| 360 (15) | 352 (25) | 1282 (66) | 1244 (120) |
| 1514 (105) | 2920 (209) | 16920 (703) | 21616 (1270) |
| -4233 (195) | -5076 (347) | -5144 (1134) | -3667 (2566) |
| -2181 (526) | -1839 (873) | 9159 (3533) | 4381 (6530) |
| 2822 (1025) | 3766 (1726) | -71482 (10121) | -72453 (24125) |
| 2316 (1336) | 15261 (2616) | | |
| 12705 (1787) | 10342 (4480) | | |
| -7598 (3343) | 3480 (7526) | | |
| -18000 (3749) | -18440 (8025) | | |

### 2.3.5 Modeling Pith Wood

I only developed calibration models for pith wood with the objective of diagnosing if correlations exist. Table 2.4 shows that density was well predicted by NIR spectra for pith-associated wood. However, MOE and MOR were not well predicted for pith wood

using NIR spectra, with the highest $R^2$ equal to 0.18. Perhaps one reason for the poor prediction is the low range of variation in MOE and MOR in pith wood. Most MOE values fell between 1,000 to 4,500 Mpa, while most MOR values fell between 20 to 70 MPa. However, for mature wood, most MOE values fell between 5,000 to 25,000 Mpa, while most MOR values fell between 30 to 230 MPa. Another factor was the high amount of resinous extractives present near the pith. Excessive extractives will heavily influence the spectra curve due to its wide range of chemical compounds. Additionally, increased resin gives erroneous density measurements if one only wants the density of the xylem wood material. Perhaps extracting the pith wood would improve the $R^2$ values. Finally, Figure 2.5 supports the conclusion that something other than density must influence MOE and MOR as one proceeds from mature wood to the first year of growth. Other studies have shown MFA to have increased importance within the juvenile wood zone when predicting MOE and MOR (Megraw 1985; Bendtsen and Senft 1986; Treacy et al. 2000).

Not being able to model the MOR and MOE from pith wood has some practical significance. First, if one wants to estimate MOE and MOR from an increment core for different families, it may not be possible with NIR models in the pith wood region. Second, should NIR ever be used to stress-rate lumber, much care needs to be taken when scanning pith-wood.

### 2.3.6 Predicting Juvenile Properties from the Mature Model

Juvenile wood is synonymous with inferior mechanical properties because of its lower specific gravity and higher mean MFA, particularly at the basal end of the tree. As a result, I feel it is important to know if models developed from whole tree values will

predict juvenile wood properties. Figure 2.6a illustrates that juvenile wood density is under-predicted using MLR, while MOR and MOE were over-predicted in Figures 2.6b and 2.6c. The residuals were computed from the mature wood model as actual juvenile wood response minus predicted response. Had the models been appropriate, approximately 50% of the data would have fallen below, and 50% above, the zero residual axis. When PCR regression was used for the three properties, similar over- and under-prediction errors occurred. Consistent over-prediction errors of MOE and MOR would be misleading if one were using spectra to predict the MOE or MOR of lumber in a production environment. Because many designers require a specific strength for safety reasons, consistent over-prediction of strength would increase the chance of lumber failure in service.

Table 2.4: A comparison of $MLR_{1,2}$ and $PCR_{1,2}$ models for near pith associated wood (n=60).

| Model type | Dependent Variable | Independent Variables | $R^2$ | Adj. $R^2$ | RMSEC |
|---|---|---|---|---|---|
| $MLR_1$ | Density | Area | 0.69 | 0.69 | 0.06 |
| $MLR_2$ | Density | Area, 2155, 2205 | 0.83 | 0.82 | 0.05 |
| | MOE | 1705, 1825 | 0.07 | 0.03 | 145168 |
| | MOR | 1705, 1825 | 0.13 | 0.10 | 1559 |
| $PCR_1$ | Density | 6 PC's | 0.87 | 0.85 | 0.0432 |
| | MOE | 3 PC's | 0.19 | 0.14 | 136818 |
| | MOR | 1 PC | 0.09 | 0.07 | 1585 |
| $PCR_2$ | Density | 5 PC's | 0.83 | 0.81 | 0.05 |
| | MOE | 2 PC's | 0.13 | 0.10 | 139944 |
| | MOR | 1 PC | 0.09 | 0.08 | 1580 |

When I only used the area under a single spectra curve to predict density, there was no significant difference of intercepts and slopes for juvenile and mature wood models.

As a result, it appears more useful to use the area under the spectra curve to estimate wood density from any portion of the tree than to use other multivariate techniques.

## 2.4 Conclusions

MLR accounted for similar proportions of variance as PCR in predicting strength, stiffness, and density from absorbance response. This is important since MLR is easier to implement and interpret for near infrared spectra. PCR, produced a better model when only wavelengths known to associate with lignin and cellulose were included. This is significant since data acquisition and analysis of all available NIR wavelengths are common in PCR.

I found the baseline shift to best for the change in density. I achieved this by regressing the area under the curve to density with least squares regression. Furthermore, the ratio of cellulose to lignin absorbance, from associated wavelengths, was significantly correlated with density suggesting that chemistry has additional impact on wood density independent of baseline shift.

The mature wood samples had a higher slope between density and mechanical properties than juvenile wood, while juvenile wood had a higher slope than pith wood which had a flat slope.

Correspondingly, MOE and MOR were poorly correlated with NIR spectra in the pith wood region while density was strongly correlated. This is important from a practical, perspective suggesting that stress grading pith lumber with NIR may not be suitable. Finally, 5 of 6 of the same wavelengths were significant for prediction of MOE and MOR.

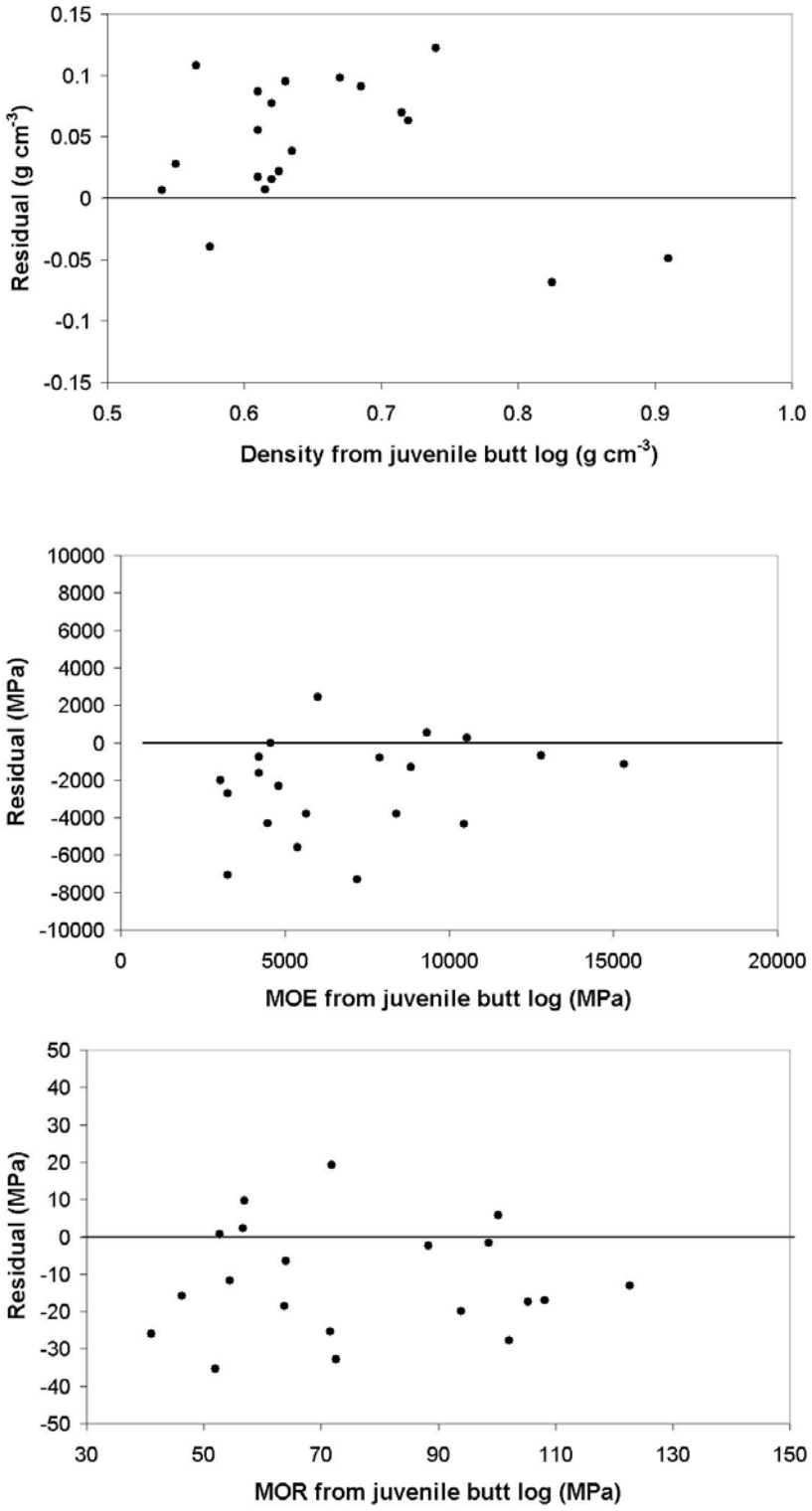



Figure 2.6: A residual plot of (a) juvenile wood density using a mature wood density MLR model, (b) juvenile wood MOE using a mature wood MLR model, and (c) juvenile wood MOR using a mature wood MLR model.

# CHAPTER 3

# ABILITY OF NEAR INFRARED SPECTROSCOPY TO RAPIDLY MONITOR WOOD DENSITY DISTRIBUTION AND VARIATION

## 3.1 Introduction

Wood density is the single best characteristic influencing pulp and wood quality (Kleppe 1970; Kibblewhite 1984; Duffy and Kibblewhite 1989). Being able to monitor the density of the raw material going into the mill is perhaps the next crucial step in efficiency improvement particularly for pulp and paper manufacturers. The pulp and paper industry places a high importance on lowering the variation of pulp to decrease costs and improve quality control. While many technological improvements have been made over the decades, there is still one inherent variation to overcome, the material itself. This paper evaluates the ability of near infrared spectroscopy (NIR) to monitor density distribution.

For paper products, an increase in density will increase pulp yield and tear index, but it will decrease tensile, burst, apparent density, and stretch (Duffy and Kibblewhite 1989; Kibblewhite et al. 1997). Since increased density will bring about mixed improvement in paper strength, Kleppe (1970) suggested that density should be improved while maintaining an acceptable level of strength. For example, linerboard can utilize high density southern pine chips while maintaining an acceptable burst index (Kleppe 1970).

The increase in pulp yield with increased density is due to the higher concentration of cellulose in latewood and the additional porosity of the earlywood zone (Gladstone et al. 1970; Labosky and Ifju 1972). Differential porosity between latewood and earlywood

results in uneven rates of liquor penetration and thus pulp yield, especially for shorter cooking periods (Labosky and Ifju 1972). Latewood can exhibit 2 to 7 percentage points higher yield than earlywood, setting the boundaries for pulp yield variation (Gladstone et al. 1970). As a result, density variation becomes the most important variable to influence most pulp and paper properties.

Improvement of pulp yield or paper strength is obviously desired, but does not add much value if the variation in pulp yield and paper strength is great. One mill study found density variation to double from one month to the next, due to changing chip supply (Farrington 1980). Daily variations are probably larger. To solve this problem, efficient separation of topwood, slabwood, and corewood chips are suggested (Veal et al. 1987). Others support the segregation of species by age and growth rate since the predictive power of these variables on fiber morphology and strength indices are high (Kärenlampi and Suur-Hamari 1997). Mills already blend sawmill pine chips or pine chips from thinnings with hardwood chips to achieve a target chip density. As a result of different material sources, bimodal distributions are likely to occur making the mean density less useful. Some method to monitor this complex variation is needed so that adjustments can be made at the blending station. With such a method, a mill could monitor incoming chips to make sure that fiber supply is providing the density that was paid for. But perhaps most importantly, the variation in density could be lowered so that there are less over and undercooked chips. Such process control ability would result in less rejects from customers while optimizing the use of manufacturing resources.

NIR is now used to estimate solid wood density with $R^2$ values between 0.70 to 0.95 (Hoffmeyer and Pedersen 1995; Schimleck et al. 2001a; Schimleck et al. 2001b;

Schimleck et al. 2002; Schimleck and Evans 2003; Schimleck et al. 2003b; Via et al. 2003). Pulp yield can also be modeled with success when partial least squares regression is the model choice (Michell and Schimleck 1998). Some density calibration equations are sensitive to juvenile and mature wood, suggesting the presence of an over-fit of the equation; however, the area under the spectra curve between 1000 and 2500 nm is insensitive to juvenile wood in response to modeling density (Via et al. 2003). The loss of model sensitivity to juvenile wood suggests that absorbance measured under the curve is the best independent variable for modeling density (Via et al. 2003). Gindl et al. (2001) also found that mean absorbance was highly correlated with wood density. Manufacturing research shows NIR to predict lumber stiffness with appropriate adequacy, although later research cautions against using NIR to stress grade wood originating from the pith (Meder et al. 2003; Thumm and Meder 2001; Via et al. 2003).

Manufacturers need to be able to measure the wood density across a range of species, moisture content, and temperature. For 54 species, NIR can be used to estimate density, with high precision. (Antti et al. 1996; Schimleck et al. 2001; Schimleck et al. 2003). It is also possible to classify wood to a species before predicting density (Schimleck et al. 1996; Tsuchikawa et al. 2003ab). Such a procedure may help to decide which density calibration equation to use for different species.

For moisture, Hoffmeyer and Pedersen (1995) found wood density prediction to be reasonably independent of moisture content below fiber saturation point (FSP). Above FSP, development of calibration models for many chemical and physical properties of chips are not a problem (Axrup et al. 2000). During calibration, removal of many hydroxyl associated wavelengths can result in more robust calibration equations and

would probably work for density (Swierenga et al. 2000). If temperature varies, as in a manufacturing environment, than one needs to either account for the temperature by either removing temperature sensitive wavelengths before calibration or include the full range of temperature into the design of model development (Thygesen and Lundqvist 2000; Wülfert et al. 2000).

NIR may thus be applicable to measuring the density distribution of wood. Similar practice are currently being pursued for other wood properties. For example, histograms of estimated lignin can be successfully plotted to monitor online variation and distribution (Jääskeläinen et al. 2003). While not the objective of this experiment, success at measuring density variation would enable better control of the variation in chip density.

For this paper, my objective was to determine if NIR could predict the variation in air-dry density of solid wood. Thus, I mapped out frequency distribution histograms and validated the NIR generated histograms. Also, density standard deviations were compared with NIR estimated standard deviations of density. Finally, the ability of NIR to classify solid wood into density catagories was investigated.

## 3.2 Methods and Materials

Ten longleaf pine (*Pinus palustris*) trees, 41 years old, were selected from a plantation on the Harrison Experimental Forest which is owned and maintained by the USDA Forest Service near Saucier, Mississippi (USA). Specifically, the location was 30.6º north and 89.1º west. Periodic prescribed fires was applied to the understory, throughout the life of the stand, so as to keep down unwanted vegetation. Trees were planted 3.66 meters apart from neighboring trees in an alternating or equilateral triangle

pattern. Border trees were planted surrounding the overall site. Each of the 10 trees were harvested and cut into bolts every 4.57 meters in height, yielding 5 to 7 bolts. Each bolt had an accompanying disk cut from the basal end of the bolt. The specimens for density measurement were taken from the bolt while the spectra were acquired from the radial face of a strip ripped from the adjoining disk. The spectra acquisition was taken within 30 mm of the actual density measurement and at the same age. Specimens for density measurement were taken at rings 1, 4, 8, 16, 32, and the last ring of each disk. The volume was measured with calipers and the weight was measured at equilibrium moisture content (EMC). The EMC was controlled to a mean of 8.1% and a standard deviation of 1.6%. Dimensions were measured to the nearest 0.0025 cm while weights were measured to the nearest 0.001g.

I acquired NIR absorbance using a Nexus 670 FTIR spectrometer (Thermo Nicolet Instruments, Madison, WI, USA). Scans were acquired at 1 nm intervals between the wavelengths of 1000 and 2500 nm. Forty scans were collected and averaged into one spectrum. During NIR scanning, the temperature was controlled at 22C$^{o}$ ± 1 with a mean relative humidity of 50%.

To reduce data set and computation time, I reduced the spectra to 10-nm intervals by averaging (Schimleck et al. 2004). One hundred seventy samples were used to develop whole tree models while I held aside 93 samples for validation model building. While multiple linear regression models and principal components regression showed slightly higher $R^2$ values, I found the area under the spectra curve to better represent density since the total absorbance in spectra shows increases with an increase in density (Via et al. 2003).

For data analysis, I developed a linear model to predict density from the area under the spectra curve. Frequency distribution histograms were developed for actual and predicted data. For standard deviation estimation, the samples were randomly clustered into groups of 3 and standard deviations in density were computed from the actual and predicted data to determine the ability to predict density variation.

## 3.3 Results and Discussion

### 3.3.1 Model Validation

Multiple linear regression of selected wavelengths and principal components regression yielded the best $R^2$ values when predicting density. However, since the spectra shifted upward as density increased, an interpretable linear model, based on the area under the spectra curve, was chosen to avoid over fitting.

$$D = ß_1 * x + ß_2 + e \qquad (3.1)$$

Where D is the density, $ß_1$ is a constant, x is the area under the curve, $ß_2$ is the intercept, and e is the error term. An $R^2$ value of 0.71 was calculated when the actual density and predicted density were regressed (Figure 3.1). The $ß_2$ slope was also significant in predicting density. While not shown for the multivariate techniques, Eq. 3.1 exhibited better fits throughout the density range even though the $R^2$ value was slightly lower by 0.03. The fit was determined to be superior if the residuals around the regression line fell at equal variance throughout the density range and a mean of zero. Overfit of the multivariate equations was found to be problematic if care was not taken to identify the proper independent wavelengths or principal components. The root mean square error of

calibration was also plotted against the number of variables and the leveling off point help to determine the number of variables needed without overfit.

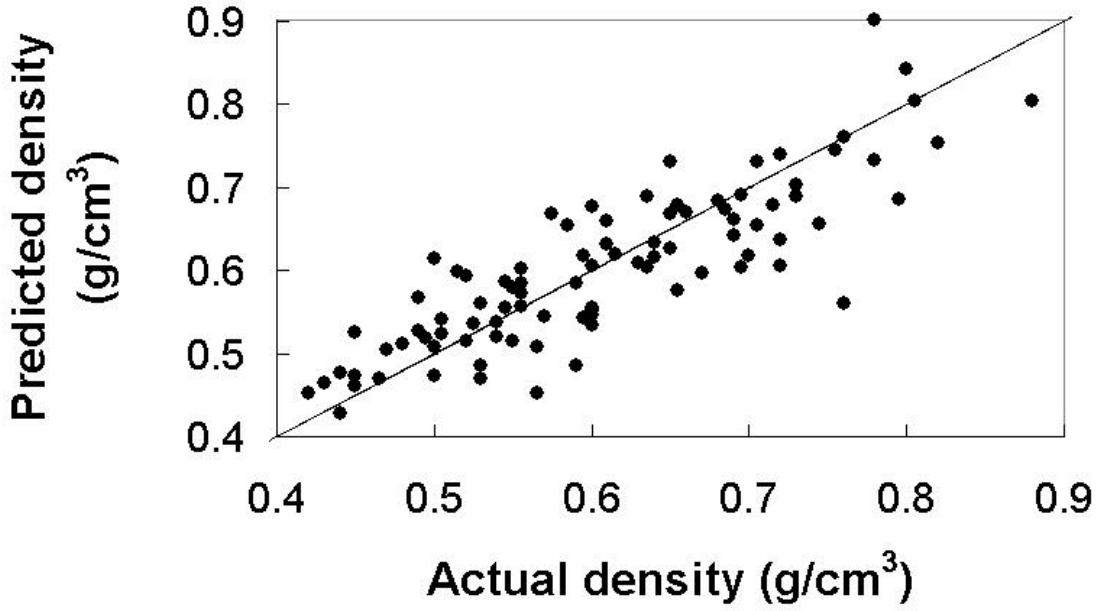



Figure 3.1: Regression analysis plot of air dry density from the area under the spectra curve versus actual density (n=170). The line is a 1:1 ratio such that any offset can be easily viewed ($R^2 = 0.71$).

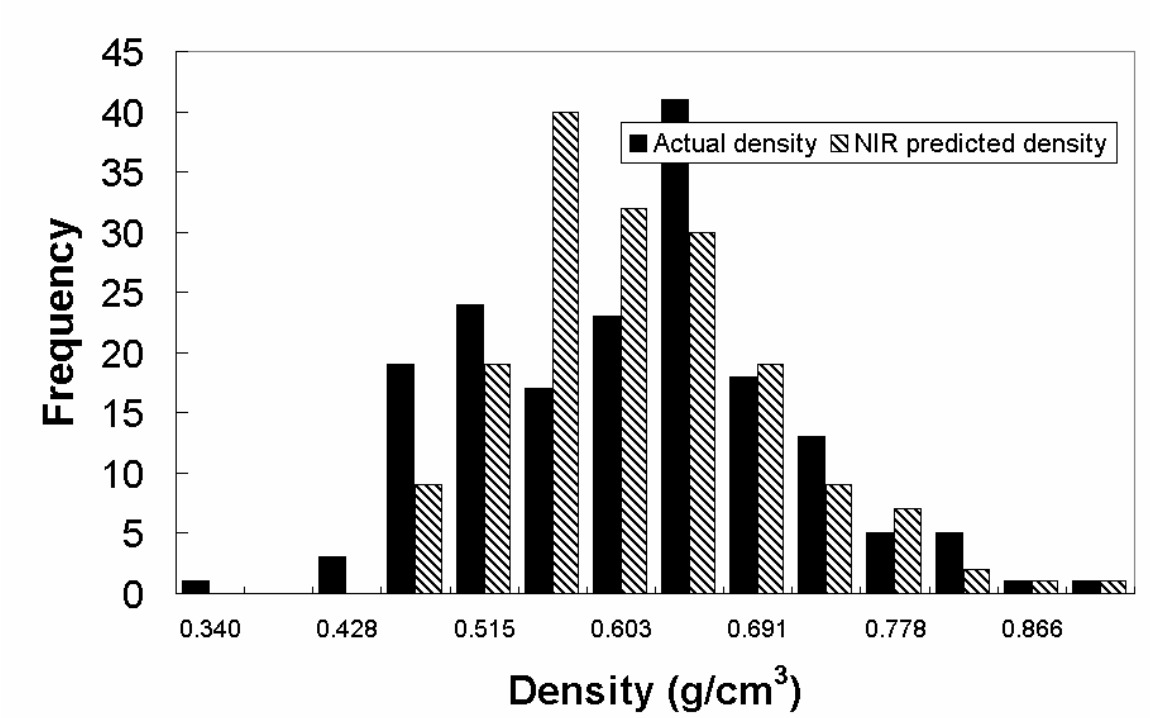
Eq. 3.1 may be important because it may encourage other researchers to follow a similar modeling strategy. The coefficient and intercept of many commercially important species would useful information and may provide the basis for a universal standard. Such information may improve the universal capability of NIR between machines and would provide a foundation of understanding between species. Of course, care would need to be taken to ensure that between machine variance would not significantly bias the results.

### 3.3.2 Distribution Modeling

Mapping out the distribution patterns of density would be very beneficial for many forest products manufacturers. A shift in raw material density, partially due to large differences in log age, could be detected. While not investigated in this study, it may be possible to differentiate between wide ranges of density between species. Schimleck et al. (2001) demonstrated that NIR could easily discriminate wood density independent of species.

I first considered distribution properties for the calibration data. Histograms were developed from the actual density and compared to the density prediction derived from Eq. 3.1. As seen in Figure 3.2a, the histograms for both actual density and NIR predicted density were normally distributed. The actual data appeared more dispersed, as I expected, while the model distribution tended to be fairly smooth, particularly between 0.50 to 0.60 g/cm$^3$. The slightly lower variation in the NIR predicted data was probably attributable to the central limit theorem, which states that the variation in means are lower than the variation in the actual data that comprised the means (Freund and Walpole 1980). Since, a linear model is simply a prediction of mean density for a given absorbance, the predicted density variation was expected to be lower than the actual density variation and the resulting shape and variance was quite similar. But to fully test Eq. 3.1's legitimacy, validation data from 93 hold out samples were tested (Figure 3.2b).

The validation data in Figure 3.2b was well traced by the NIR predicted values. The validation histogram trace was better than the trace shown in the calibration histogram; an event that occurs with good models.



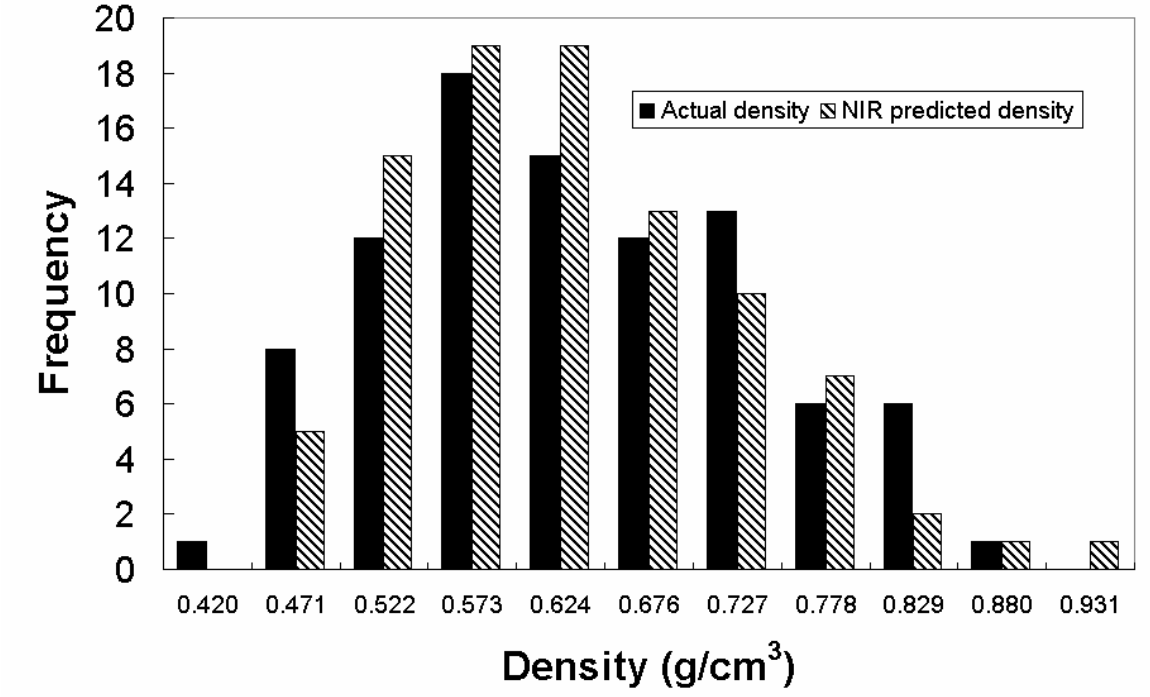




Figure 3.2: Histograms for actual and predicted density for (a) calibration (n=170) b) validation data (n=93)

For the validation data set, a small shift in wood density was apparent between the actual and predicted data values (Fig. 3.2b). In other words, the NIR model histogram appeared to slightly overestimate density as apparent from the shift toward higher density values. This did not occur in the calibration data suggesting a slight bias in the model due to random noise in selection of the validation samples. Another useful outcome was the change in distribution from normal to slightly skewed when going from Figure 3.2a to 3.2b. The model output was mildly skewed to the right with the same variance as the actual data (Fig. 2.2b).

Being able to measure skewness or even bimodal distributions would prove useful in manufacturing. For example, if the distribution is skewed right, a pulp mill may adjust for the median instead of the mean depending on the severity of the skew. For other applications, such as lumber drying, the kiln drying schedule may instead be adjusted for the mean to ensure that no pieces of lumber are above the required 19% or less moisture content. With such tight restrictions, additional time, energy consumption, and warp is minimized.

By plotting out density distributions, manufacturers would also be able to notice subtle or large changes in material variability. Currently, many pulp manufacturers are reluctant to consider categorizing their logs by age, height, or both since such practices would require a heavy process change and financial capital. However, in New Zealand and Australia, such innovative practices are helping to make pulp and forest product manufactures more efficient by lowering the variation in final product properties. Such a visual plot may encourage more inefficient mills to classify their raw material in the log yard in an attempt to lower or control density variation.

It is important to note that I performed this experiment in the confines of a laboratory with a low wood moisture content and temperature range. In a manufacturing environment, the moisture will vary wildly and may be above FSP. Any model developed by a mill should incorporate the complete range of temperature and wood moisture content in their calibration equation. Such practice (Thygesen and Lundqvist 2000; Wülfert et al. 2000) would enable one to account for the variation induced by temperature and moisture when predicting the variable of interest, in this case, density.

### 3.3.3 Dispersion Modeling

Instead of always having to re-map out the complete distribution for every scan, I feel it may be more useful to consolidate the data into a variance or standard deviation estimate. As long as the original density data distribution has not deviated from the Gaussian, such a method can be easier to compute and just as accurate in making decisions. Additionally, if a batch of similar aged logs is processed at the same time, one could adjust the parameters of the process to account for the lower variation in raw material. In a pulp mill, it is not uncommon for a truck load of similar aged logs to be stacked and processed together. It is possible that material of similar age, species, and origin would exhibit lower variation in density. Pine thinnings from a plantation stand would be a good example of processing a similar material with lower variation. As such, Figure 3.3 was built to visually demonstrate the ability of NIR to model the distribution of density.

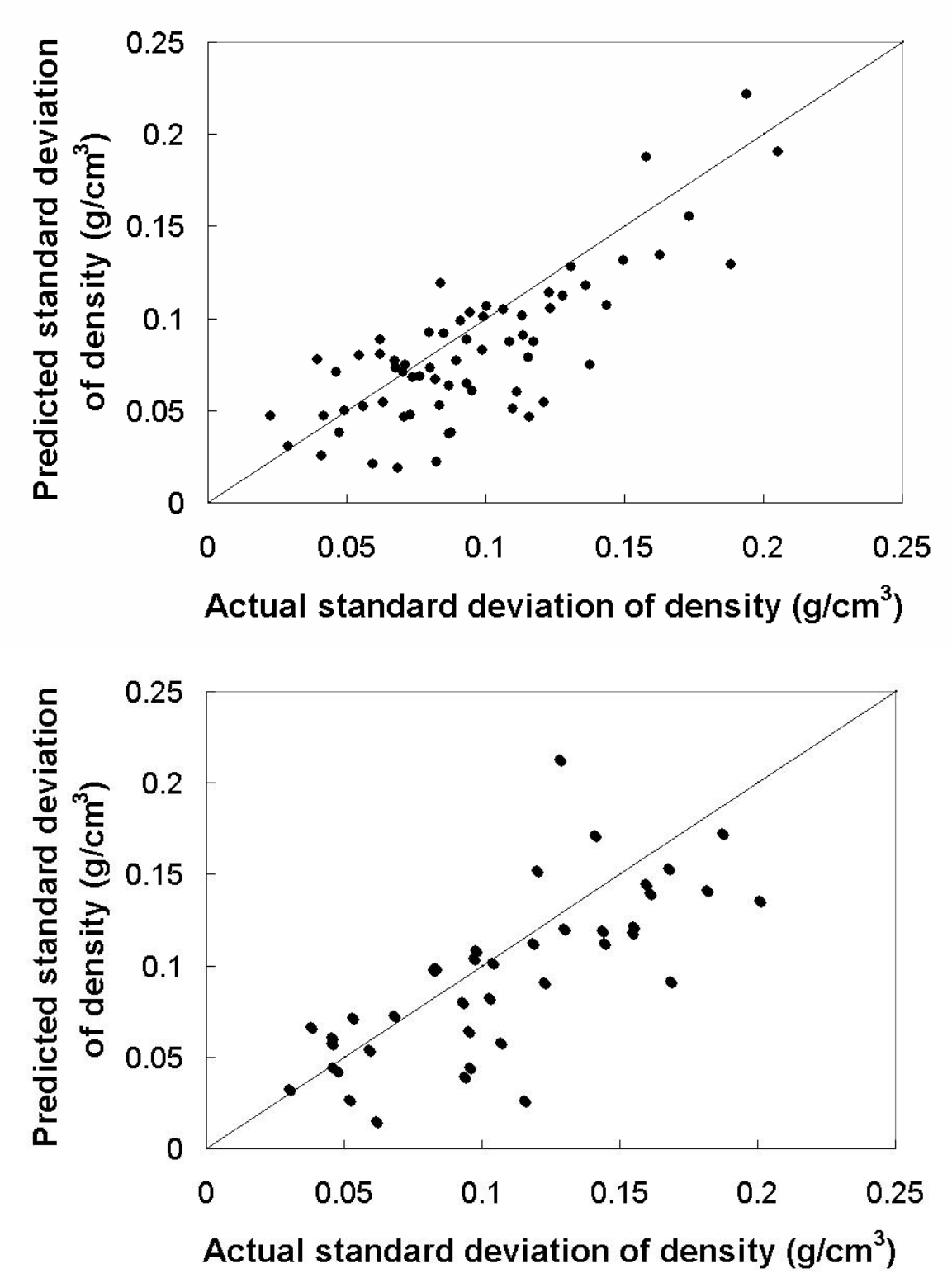



Figure 3.3: Actual standard deviation for samples grouped into threes versus that predicted by NIR models (a) calibration (n=170/3=56) and (b) validation data (n=93/3=31).

Figure 3.3a demonstrates the ability to use absorbance across all NIR wavelengths to model the standard deviation of density for the calibration data. The $R^2 = 0.61$ when the NIR predicted standard deviations were regressed against the actual standard deviations of density. This was quite good considering only three samples were used to calculate each data value for the actual and predicted density. In a manufacturing situation, more than 3 sample measurements are possible.

Figure 3.3b demonstrates the ability of absorbance across all NIR wavelengths to model the standard deviation of density for the validation data. A slightly lower $R^2 = 0.55$ was observed when the NIR predicted standard deviations were regressed against the actual standard deviations of density.

I found it interesting to note that both Figure 3.3a and 3.3b exhibited more residuals on the bottom side of the 1:1 line than on the upper end which was an indication of bias. Figure 3.4 demonstrates the sample size needed to reach the overall population standard deviation. Around 30 data points were needed to estimate the standard deviation to a ± 0.01 accuracy. As can be seen in Figure 3.4, many more data points were needed to improve the estimation of the population standard deviation below ± 0.01 accuracy (± 0.03 g/cm$^3$ precision at 95% confidence interval). Thus, one might develop a control chart using 30 samples to estimate a single density variance data point in time.

Figure 3.5 shows the distribution of the standard deviation to be heavily skewed right. Since the standard deviation can not drop below zero and the probability of lower values being greater was intuitive, a right sided skew was expected. This non-normal property would need consideration when developing a control chart for density. Otherwise, false

alarms could occur when the estimated standard deviation falls outside the 95% confidence limits as assumed under the normal distribution. If a sample size of 30 is used to estimate a single data point, instead of 3 as used in Figure 3.5, the skewness becomes negligible to non-existent and Gaussian control charting could be applied with reasonable estimates.

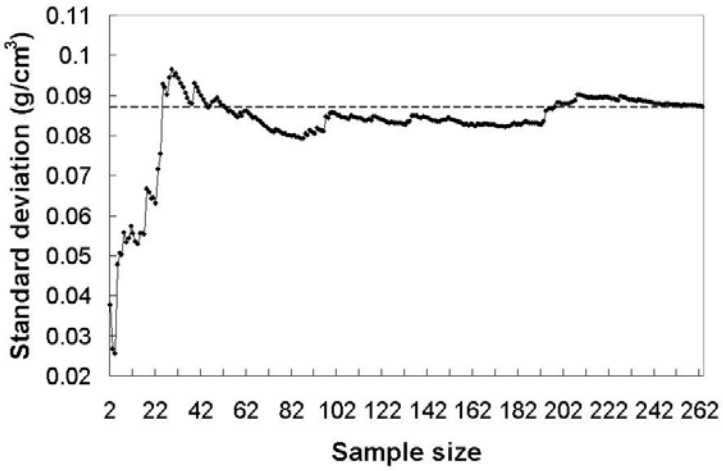



Figure 3.4: Sample size versus estimated standard deviation with the known standard deviation of the population, plotted as a dashed line.

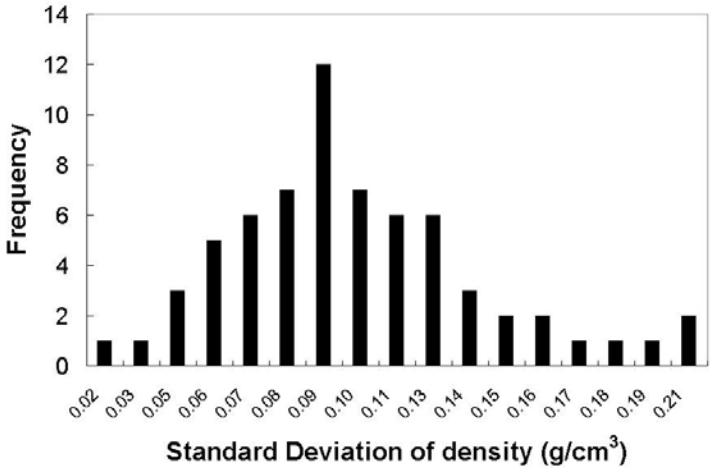



Figure 3.5: Histogram of density standard deviations for all samples demonstrating the non-normal distribution.

### 3.3.4 Discrimination of Density

For some applications, I think it may be desirable to classify the on-line chip density into a given set of bins instead of predicting a specific value. For example, a pulp mill may be satisfied with a low, medium, and high density classification given the wide range in density that may occur from tree to tree. This wide variation might occur if tree age is heterogeneous within the log yard or by the time a group of trees are processed.

Table 3.1 demonstrates the ability to classify according to the number of catagories as a function of the full range of density. Table 3.1 presents the full range of density which fell between 0.35 to 0.9. A partition of 2 resulted in a separation of low, medium, and high density. When this was done, 99 percent of the specimens were properly classified.

Table 3.1: Percent classification by NIR at different bin widths (different precision allowances).

| Density Increments | Number of Partitions | Percent Correct Classification by NIR |
|---|---|---|
| 0.250 | 2 | 99 |
| 0.167 | 3 | 91 |
| 0.125 | 4 | 89 |
| 0.100 | 5 | 74 |
| 0.083 | 6 | 89 |
| 0.071 | 7 | 72 |
| 0.063 | 8 | 76 |
| 0.055 | 9 | 76 |
| 0.050 | 10 | 67 |
| 0.045 | 11 | 78 |

When going from 2 to 11 partitions, the percent correct classification column was expected to drop because the range of density was lower than the expected error. This

did occur, but not as distinctly as expected (Table 3.1). There was also not a smooth transition in percent correction classification from 2 to 11 partitions. Perhaps the reason for this variability was attributable to the change in boundaries as shown in the increment column of Table 3.1. Each time a new partition was added, the increment became lower and all the bins changed to a narrower range resulting in different bin boundaries. Thus, what might be correctly classified under one partition scheme might fall in the error region for another partitioning scheme.

An interpretation of Table 3.1 was that approximately 70% correct classification was to be expected for a 0.05 increment change in density. In other words, to obtain the correct discrimination 70% of the time, a precision of about ± 0.025 g/cm$^3$ was expected. This agreed well with quantitative precision of the residuals, of actual versus predicted density, which fell at 0.06 or ± 0.03 g/cm$^3$, as determined by the 95% confidence interval.

One limitation to the precision estimate was the diameter of the NIR spot size. A 5 mm diameter light beam, in the near infrared range, was used to scan and acquire absorbance data. However, some of the latewood and earlywood rings were as small as 0.3 mm in thickness, especially in the 30 to 40 year old ring region, as measured from the pith outward. A typical difference between early and latewood density was 0.45g/cm$^3$. As a result, an error between what was measured gravimetrically and what was actually scanned occurred even if only a slight difference in proportion of latewood was introduced. The result was actually encouraging because further research may yield improved $R^2$ values for calibration equations simply by using a smaller spot size.

## 3.4 Conclusions

The mapping of density histograms, as predicted from NIR spectra, was nearly equivalent to the actual density histograms for the calibration and validation data. When I investigated the ability to monitor the variation in a sequential manner, I found NIR to be successful at estimating the variation with just 3 measurements. A sample size of 30 is recommended to estimate the standard deviation of density with a precision between ±0.025 to ±0.030 g/cm$^3$. Any measurements greater than 30, yielded only smaller incremental improvement in precision for a given increase in sample size.

These results were based on a tightly controlled moisture and temperature range so that the results could give a baseline for NIR performance. More work, beyond the scope of this paper, would be needed to determine the precision of measuring density above fiber saturation point. However, what was important about this work was that the equation was based on a simple linear regression relationship between absorbance and density. Such a simple equation, with proper interpretation, I feel is beneficial.

# CHAPTER 4

# TRACHEID LENGTH PREDICTION IN *PINUS PALUSTRIS* BY MEANS OF NEAR INFRARED SPECTROSCOPY: THE INFLUENCE OF AGE

## 4.1 Introduction

Rapid measurement of tracheid length is desirable by tree improvement programs who wish to assess tracheid length variation from increment cores. It is well established that tracheid length, in most cases, is more heritable than density, growth rate, or disease resistance (Jackson and Greene 1958; Zobel et al. 1960; Dadswell et al. 1961; Nicholls et al. 1964; Smith 1967; Matziris and Zobel 1973; Loo et al. 1984; Cown et al. 1992; Whiteman et al. 1996; Hannrup and Ekberg 1998; Miranda and Pereira 2002; Pot et al. 2002).

Near infrared spectroscopy (NIR) to measure fiber morphology is emerging as an alternative to sometimes more precise, but time-consuming, measurement techniques for tree improvement groups. For example, NIR and FT-Raman Spectroscopy appears to detect tracheid length by accounting for more than 70% of the variation when tested across rings for Norway spruce (Hauksson et al. 2001; Ona et al. 2003). However, before scanning, the wood was milled to a particle size for both studies. Since tracheid length could be estimated from small particles, it becomes apparent that NIR is dependent on some secondary relationship since tracheids would have been cut to smaller sizes during grinding. To date, no attempt has been made to understand the basis for how a spectroscopic technique is related to tracheid length.

It is clear that tracheid length detection with NIR must have some level of indirect relationship to the NIR signal, particularly since the absorbance capacity at a given

wavelength depends on the stretching or bending nature of bonds between molecules and not a macro feature like tracheid length. One possible connection between tracheid length and NIR spectra is the change in tracheid length and chemistry with age. Often, as the cambium matures, the concentration of cellulose and fiber length increases while the amount of lignin decreases (Bendtsen and Senft 1986; Shupe et al. 1997). Such covariance between traits with age may result in a secondary correlation between the NIR signal and tracheid length.

When the nature of a relationship is not understood, there is increased chance for misleading calibrations (Osborne and Fearn 1986). As a result, one needs to understand the basis for a correlation and when it is appropriate for use. Koubaa et al. (1998) demonstrated that age from the pith accounted for more than 80% of the variation in fiber length for hybrid poplar (*Populus x euramericana*). Many other studies of various softwood species demonstrate similar trends showing a distinct increase in tracheid length at a decreasing rate from pith to bark and from breast height to crown (Megraw 1985). It is possible tracheid length increases with age and various chemistry components change with age resulting in an indirect relationship.

The chemical morphology in the primary wall is also quite different than the secondary wall. Given that primary wall development affects cell length, there may be some relationship between the NIR signal (absorbance at defined wavelengths) and the primary cell wall. For instance, as the primary cell wall develops in length, it needs plasticity which requires an amorphous polymer network (Roelofsen 1969; Cosgrove 1999). Arabinose and galactose are two sugar units common in the primary cell wall of southern pines while mostly absent in the secondary cell wall (Koch 1972).

My primary objective for this work was to explore the potential of using NIR absorbance data to predict tracheid length from a) all samples collected from experimental trees and b) data categorized by age and height. Two thirds of the data, (n = 300) were used to develop calibration equations while the remaining data (n = 153) were used to validate the model. Alternatively, when age and height were held constant, 20 data points were available to develop principal component regressions (PCR) for each age-height group for a total of 25 replicate equations. Trends in principal component coefficients, and $R^2$ values, for age and height were explored. This objective had utility since it determined the role that age and height played in the secondary correlation between tracheid length and NIR absorbance data. I feel such information is needed to help conclude the basis for the correlation between NIR absorbance and tracheid length.

## 4.2 Methods and Materials

### 4.2.1 Sample Preparation

Ten longleaf pine (*Pinus palustris*) trees were harvested from a plantation on the Harrison Experimental Forest, which is maintained and owned by the USDA Forest Service near Saucier, Mississippi (USA), at approximately 30.6 north and 89.1 west coordinates. These 41-year-old trees were bucked into 5 to 7 bolts, depending on tree height, to yield a disk every 4.57 meters in height. The trees were selected to yield the range of diameters and height values for the stand. A double bladed circular saw was used to rip bark to pith to bark strips approximately 20 mm by 20 mm in tangential and longitudinal dimensions. Absorbance in the NIR range was measured at the appropriates section within the strip.

For tracheid separation, a small portion of the latewood, from the remaining part of the disk, was chiseled into random matchstick-sized dimensions and macerated in 1:1 mixture of glacial acetic acid and hydrogen peroxide and heated for 24 hours at 60C$^{o}$ or until bleached white. This was performed on samples at rings 1, 4, 8, 16, and 30 from the pith for both the north and south face of the tree. These ring numbers were chosen since other data from that stand has been collected on those rings. The fibers were then rinsed with 2 to 3 cycles of water; i.e., the glacial acetic acid and hydrogen peroxide were completely replaced by water. The samples were stored in a small vial and then refrigerated at 4.4º C until further processing.

4.2.2 Tracheid Length Measurement

Macerated samples were processed through a Fiber Quality Analyzer, in which each tracheid was suctioned through a flow cell and automated image analysis was used to determine tracheid length. Control samples were periodically run to ensure proper calibration. A minimum of 1,000 tracheids per sample were measured except for a handful of vials that were broken during transportation but still yielded 300 plus tracheids. The mean number of tracheids measured was 4,700 tracheids per sample. A total of 475 samples were run. I calculated length weighted averages using Eq. 4.1 since this is the standard.

$$L_{avg} = \frac{\sum (l_i^{\,2} n_i)}{\sum (l_i n_i)} \qquad (4.1)$$

where $l_i$ represents the tracheid length of the $i^{th}$ bin and $n_i$ represents the total bin tracheid count. The length weighted measurements were used to adjust for the bias caused by the

small distribution of cut fibers inherent when using a 12 mm increment borer during sampling.

4.2.3 NIR Spectroscopy and Analysis

Before scanning, samples were air dried to common moisture content. I acquired NIR absorbance using a Nexus 670 FTIR spectrometer (Thermo Nicolet Instruments, Madison, WI, U.S.A.) at wavelengths between 1000 to 2500 nm at 1 nm intervals using reflectance spectroscopy. I collected forty scans and averaged it into a single spectrum curve. The diameter of the spot size of the scan was 5 mm. The scan was executed on the latewood portion next to the maceration site. I averaged the data set into 10 nm intervals to reduce the memory required for the spectra data (Osborne and Fearn 1986; SAS 1999; Schimleck et al. 2004).

As a first step, I used least square regression to determine how much variation in tracheid length was accounted for by age. I reviewed scatter plots to ensure that a linear trend was appropriate. Then, I used multiple linear regression to determine if peaks and valleys, from the first derivative, were important for determining tracheid length. For all NIR analysis, after preliminary tests, the first derivative yielded superior equations over the original and second derivative data. Therefore, absorbance across all wavelengths was transformed into the first derivative. PROC PRINCOMP was used to run principal components regression, an algorithm in SAS software (1999). A p-value less than 0.05 was used as the determining criteria for a principal component being a significant predictor. After determining which principal components (PCs) were useful in predicting tracheid length, the analysis was ran for a) the whole tree including calibration (n = 300) and validation (n =153) models, and b) each age and height group. For calibration

modeling, two thirds of the data was randomly chosen leaving one third of the data for validation. For each age and height group, the same PCs were used for prediction as those chosen for the overall calibration and validation model.

## 4.3 Results and Discussion

### 4.3.1 Whole Tree Modeling

Figure 4.1 shows the plot of predicted versus measured tracheid length for the calibration and validation samples. A 1:1 diagonal line was drawn to indicate any obvious offset or bias, as standard required by most chemometrics journals. There exists a slightly higher frequency of data points above the diagonal line of approximately a 0.2 mm offset. The calibration and validation data appeared to follow the same slope, which was not deemed to be significantly different at the $\alpha = 0.05$ level; ie, the 95% confidence intervals around the slope estimates overlapped. All tracheid lengths greater than 4.2 mm were below the diagonal line for the validation data indicating an underestimation by the model in this region (Figure 4.1). The underestimation was either due to an over fit of the calibration data or perhaps the data was slightly non linear in behavior above 4.2mm. Using 6 PCs, the calibration model gave a good fit with an $R^2$ value of 0.72 while the validation test set gave a slightly lower $R^2$ value of 0.65.

Table 4.1 lists the coefficients for PC variables used for whole tree regression analysis. These coefficient estimates represent the calibration model used to predict tracheid length in Figure 4.1. When models were built for the validation data, similar coefficients emerged supporting a sound calibration model (Neter et al. 1996). Due to a low correlation with tracheid length (p-value > 0.05), PC 1 was not included in the model. While PC 1 represented the largest axis of variation, for a linear combination of

all wavelengths, it did not prove significant in predictive equations. Inclusion of PC 1 into the backward stepwise selection procedure did not significantly change the coefficients of the remaining PCs and further justified leaving it out of the model.

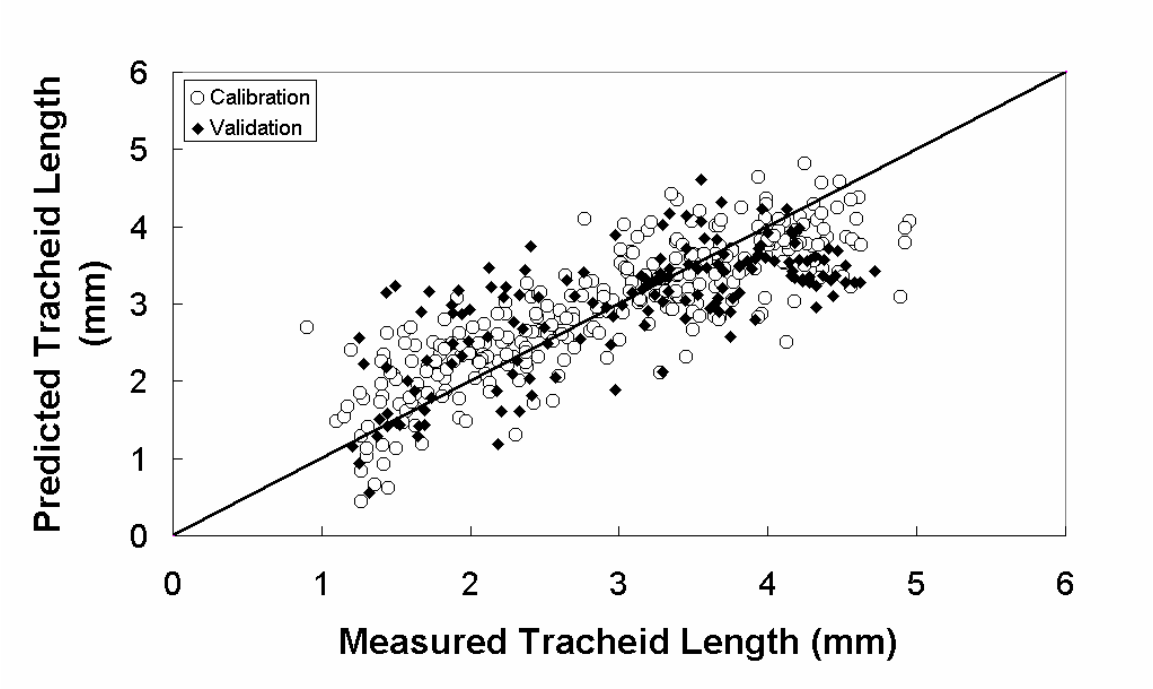



Figure 4.1: Predicted versus measured tracheid length for calibration and validation samples ($R^2 = 0.72$).

Table 4.1: Coefficient and error estimates for the principal component (PC) regression model.

| Variable | Coefficient and Intercept | Standard Error | T-Value | Pr>\|t\| |
|---|---|---|---|---|
| Intercept | 2.969 | 0.031 | 94.31 | <0.0001 |
| PC 2 | 0.080 | 0.005 | 17.55 | <0.0001 |
| PC 3 | 0.138 | 0.009 | 16.08 | <0.0001 |
| PC 5 | 0.144 | 0.015 | 9.48 | <0.0001 |
| PC 6 | 0.123 | 0.022 | 5.64 | <0.0001 |
| PC 7 | 0.134 | 0.026 | 5.19 | <0.0001 |
| PC 9 | -0.150 | 0.032 | -4.64 | <0.0001 |

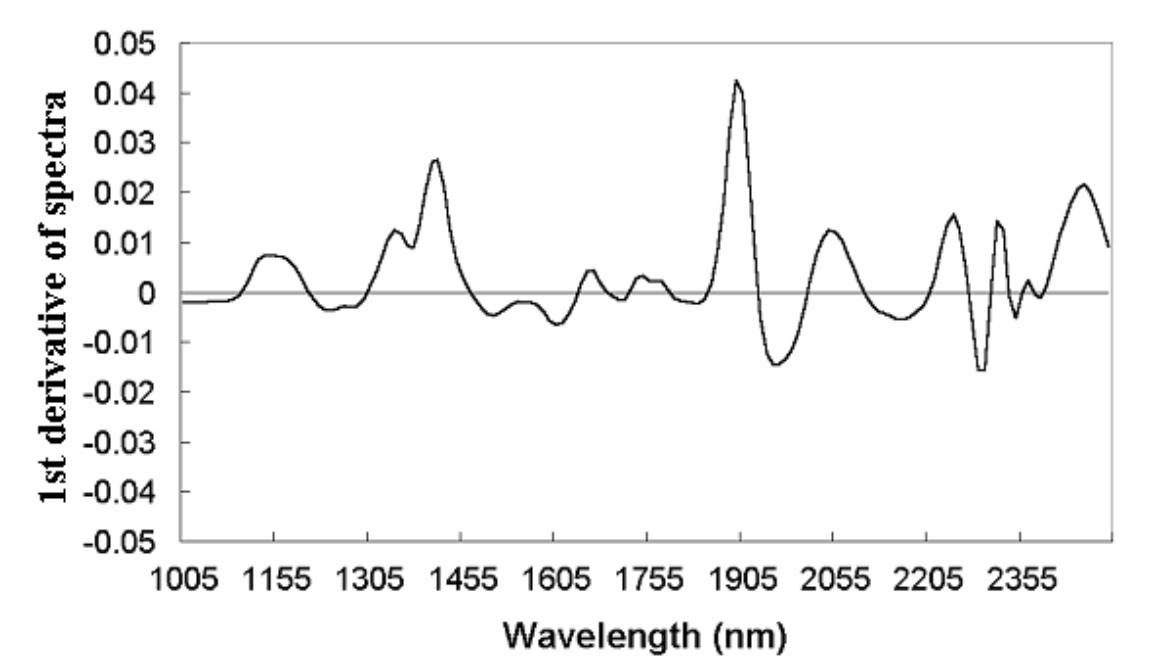



Figure 4.2: Wavelength versus 1st derivative of spectra of calibration samples (n = 300).

By taking the 1st derivative and then calculating PCs, I reduced the statistical noise attributable to high multicollinearity between wavelengths. I identified the extreme peaks and valleys in Figure 4.2 that were most associated with tracheid length. Multiple linear regression, with data not transformed into principal components, revealed all the major peaks and valleys to be significant in predicting tracheid length, with the exception of wavelength 2045 nm. These wavelengths 1895, 1415, 1965, 2245, 2295, 2315, and 2455 nm gave a combined $R^2 = 0.65$ for predicting tracheid length. While regression with PC variables exhibited a more accurate model, multiple linear regression was useful for understanding which wavelengths were fundamentally important in predicting tracheid length. Alternatively, if only one or 2 PC's been needed to predict tracheid length, one could have used the coefficients of the PC linear combination.

The first two PC's, in table 4.1, were plotted against one another to distinguish if clusters occurred as a function of age (Figure 4.3). PC1 was not used for prediction since it did not show a significant p-value. The presence of clusters for juvenile and mature wood indicated that physiological age, represented by ring number or age, may have some influence on the ability to predict tracheid length from NIR absorbance. Rings 16 and 30, throughout the tree, were categorized as mature wood, while juvenile wood was categorized as rings 1, 4, and 8. This grouping scheme was chosen since 2 populations seemed to best provide superior discrimination. As Figure 4.3 shows, PC 2 did most of the separation of mature and juvenile wood. Figure 4.3 was important because it shows that juvenile and mature wood may be separated when the data lay outside the overlap zone. As such, tracheid length increased sharply from juvenile to mature wood which would enable the categorization of tracheid length as short in the juvenile wood zone and long in the mature wood zone.

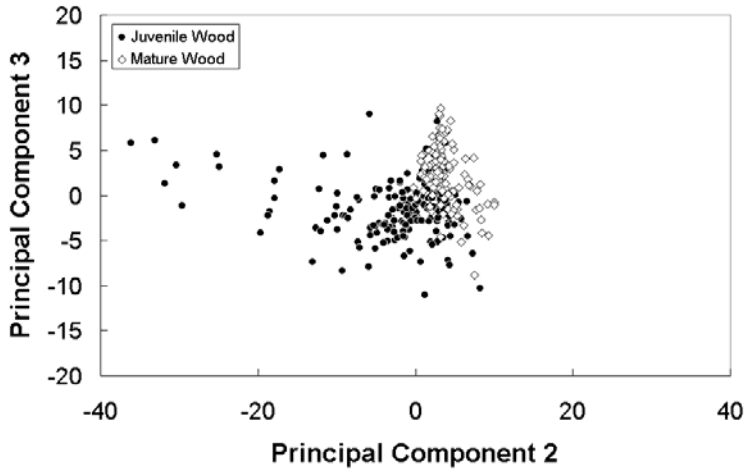



Figure 4.3: Classification of juvenile and mature wood using principal component 2 versus principal component 3 in a scatter plot.

The variability of the juvenile wood data in Figure 4.3 was greater than the mature wood population which was tightly clustered. Sixty-nine percent of the data collected form juvenile wood fell outside the mature wood cluster zone while 31% overlapped. The additional variability of the juvenile wood population implied that many juvenile samples could be separated with certainty when the when outside the overlapping zone (Figure 4.3). Given that tracheid length increased with age and that PC 2 depended on age, I found it evident that the model in Figure 4.1 and Table 4.1 was confounded by age. Interestingly, when age was accounted for by using PC 2, five more PC's were significant in predicting tracheid length (Table 4.1).

4.3.2 Age and Height Modeling

Figure 4.3 shows that radial position within the tree influenced the absorbance in the NIR region which resulted in apparent tracheid length prediction (Table 4.1). I considered it possible that height may play a critical a role. As a result, the apparent correlation in Figure 4.1 may be an artifact of changing chemistry and fiber morphology with age. It is known that lignin decreases, while cellulose and tracheid length increases from pith to bark (Koch 1972; Megraw 1985). So first, I studied the effect of height on regression coefficients. If height were to play a part in predicting tracheid length, then the regression coefficients may change significantly with height. Figure 4.4 demonstrates the coefficient response with height, for different age groups. As revealed, the regression coefficients remained fairly stable regardless of height (bolt number). While the coefficients of rings 1, 4, and 8 appeared to increase between bolt 4 though 6, I found it to be not statistically significant as determined by overlapping 95% confidence intervals

around the slope estimates. I found it interesting interesting to note that juvenile rings 1, 4, and 8 had higher regression coefficients than mature wood rings 16 and 30. For most bolt and age comparisons, these regression coefficients were significantly different. As a result, different models might enhance the ability to predict tracheid length in juvenile and mature wood populations.

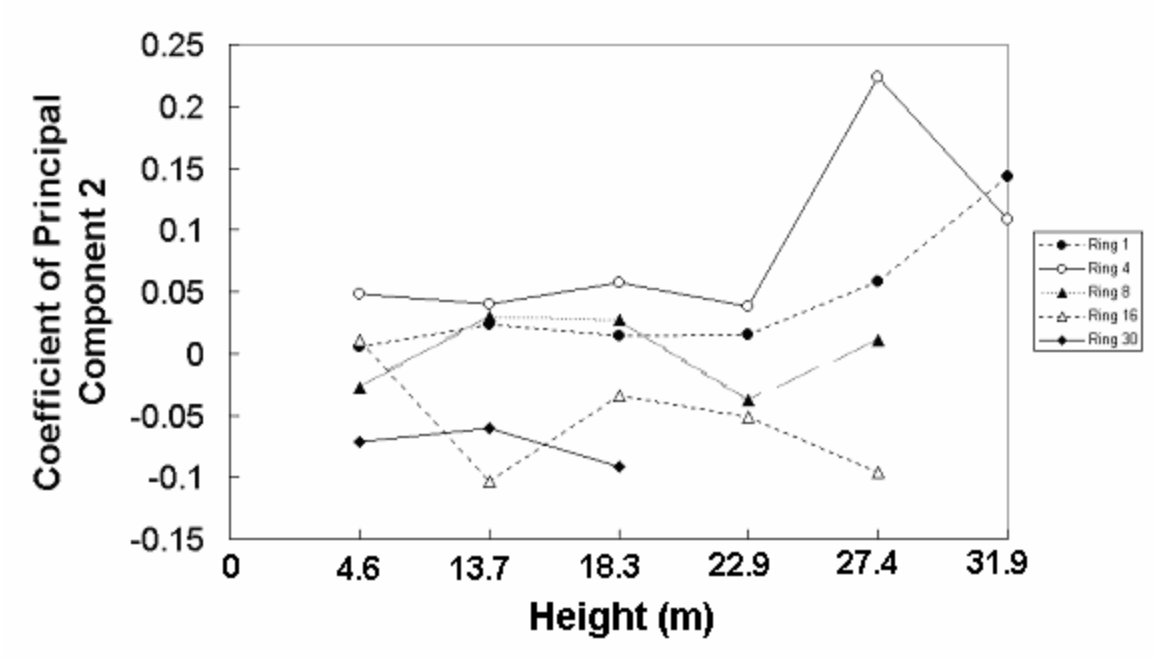



Figure 4.4: Regression coefficients for principal component 2 across all heights for different age groups.

To determine if physiological age had an effect, I ran a least squares regression for the calibration data. Ring number accounted for 66% of the variation in tracheid length using the least squares model. When height was used to predict tracheid length, an $R^2$ of 0.015 resulted supporting that age from pith was much more influential on tracheid length than height.

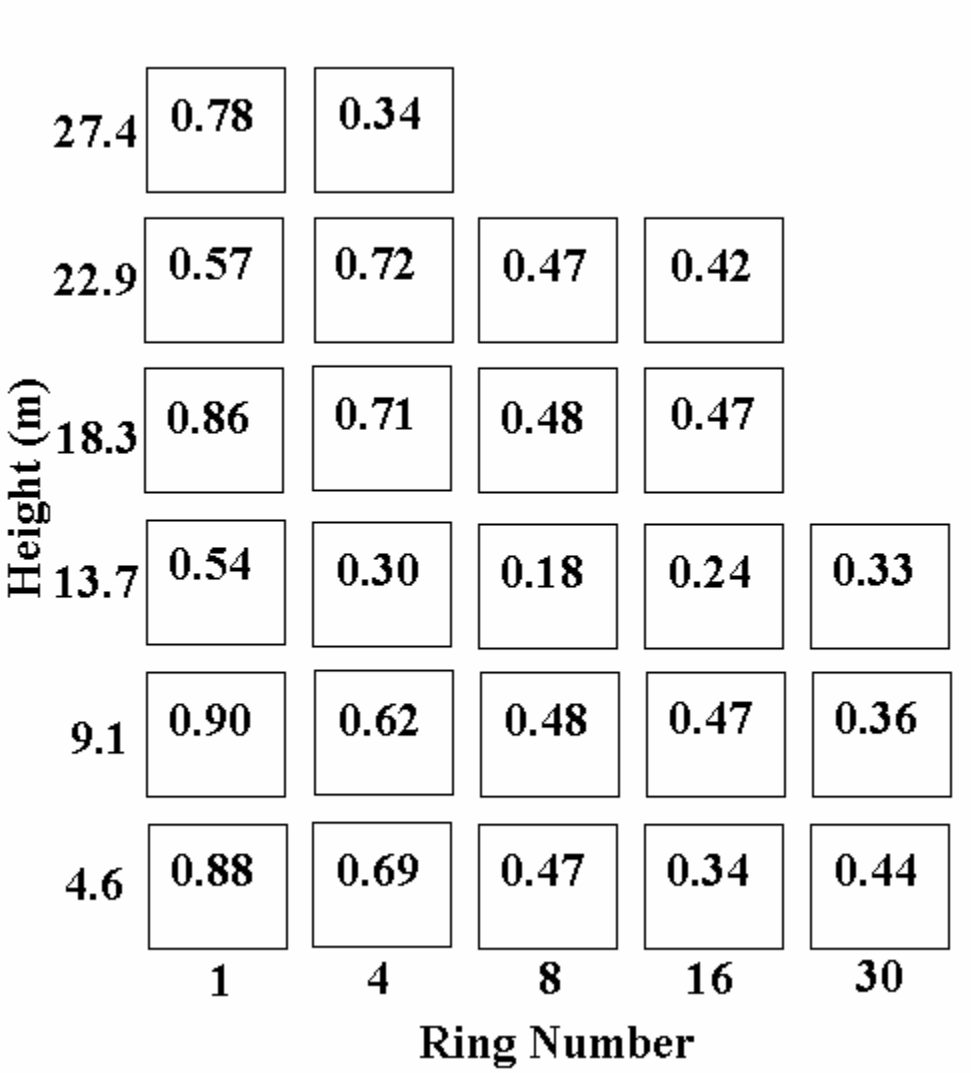



Figure 4.5: Diagram showing the $R^2$ value (in boxes) of predicting tracheid length from NIR spectra at each ring and height combination (n = 20 per box).

To demonstrate the effect of ring number and height on models for tracheid length, I constructed Figure 4.5 which shows the $R^2$ response for a given radial and vertical position. No trend in $R^2$ was apparent with tree height while $R^2$ values generally decreased from pith to bark. Close to the pith, the NIR spectra showed good precision for tracheid length while by ring 8, models were at best moderate and quite often poor. Nevertheless, even for older tree rings, the PCs consistently had p-values less than 0.05, meaning that some fundamental chemical constituent, independent of age, must be

related to tracheid length. The higher $R^2$ values at rings 1 and 4 were unexplainable but might be attributable to the physiological response of the tree. At that time, the cambial initials are growing at a faster rate, resulting in different fiber and chemical morphology (Megraw 1985). The higher $R^2$ value for rings 1 and 4 were opposite of what was expected. Since resin levels were very high during the early years of growth, as discerned visually, increased noise and decreased $R^2$ values were anticipated. Despite the already moderately high correlations at rings 1 and 4, perhaps removal of extractives for all ring numbers would yield better models and merits further research.

Taking the first derivative appeared to partially remedy some of the strong multicollinearity between wavelengths associated with both extractives and tracheid length. I judged this by the decrease in variance inflation factor (VIF), which fell between 10 and 30 for many of the chosen wavelengths. Ten is a common threshold used in multiple regression modeling to evaluate if excessive multicollinearity between variables exist (Neter et al. 1996). While high multicollinearity still existed between wavelengths, the VIF was considerably lower with $1^{st}$ derivative processing. High multicollinearity was undesirable since it escalated the variance of the regression coefficient estimates making interpretation of the model difficult.

The lower $R^2$ values, at ages greater than 8 rings from the pith, suggest that NIR is imprecise for modeling tracheid length, in older portions of the tree. However, improvement of tracheid length in older portions of the tree(s) has limited utility since older trees are harvested for lumber while thinnings go to pulp mills. In genetics studies, replicates are usually no greater than 10 to 20 trees per family, making the precision of the model highly essential. The high correlations, exhibited in the first years of growth of

the cambium, may be high enough to detect differences in means between families with sufficient replication. Given that trees are thinned as early as 8 years into a rotation, measurement and improvement of tracheid length with NIR spectroscopy may be useful.

The significant relationship shown in Figure 4.1 was mostly attributable to an underlying interrelationship between age, wood chemistry, and tracheid length. Specifically, three of the six wavelengths (1415, 1965, and 2315 nm) significant in predicting tracheid length were associated with lignin and cellulose.

## 4.4 Conclusions

The ability to estimate tracheid length from the absorbance in the NIR region was primarily due to the variation in NIR signal with age. When relationships were investigated holding ring position and height constant, the relationship was mostly weak for ages 8 to 30 and moderate to strong for ages 1 to 4. When all PC's were used to predict tracheid length, I showed that absorbance in the NIR region could successfully be used when modeling tracheid length using 6 principal components and yielding an $R^2$ of 0.72 (Figure 4.1).

The modeling of tracheid length, for tree improvement applications needs more justification, particularly for mature wood zones. Without the variation in xylem age, NIR does not generally have an acceptable precision to estimate tracheid length. The exception to the rule was the rings adjacent to the pith which had a mean $R^2 = 0.75$. As a result, NIR spectroscopy may be suitable for tree improvement in this juvenile zone. The predictability of tracheid length ($R^2$) decreased rapidly from ring 1 to 8 and then continued to decrease at a gradual rate. Thus, by ring 30, PC models accounted for approximately 40% of the variation in tracheid length.

Finally, when PC 2 and 3 were plotted I found that juvenile and mature wood could be separated with some overlap. This demonstrated that linkages exist between NIR absorbance and tracheid length when samples from all parts of the trees were analyzed.

# CHAPTER 5

# RELATIONSHIP BETWEEN LIGNIN CONTENT AND MICROFIBRIL ANGLE: THEORETICAL, EMPIRICAL, AND NEAR INFRARED MODELING

## 5.1 Introduction

The composition and orientation of polymers in the cell wall change with age and growth rate. Likewise, tracheid length changes with age and may be a function of polymer orientation in the primary wall. Polymer orientation may be an important factor determining cell length since the cellulosic matrix is formed perpendicular to the cellulose chain (Gunning and Hardham 1982). Similarly, development of the secondary wall (S2 layer) during cell wall thickening occurs in the plane perpendicular to polymer orientation. In the S2 layer, polymer alignment is less variable when compared to the primary wall with microfibrils showing a high degree of order from lamella to lamella.

When cell growth occurs in proximity to the apical meristem, auxin concentrations are high and will influence fiber dimensions (Brown 1970). For example, the addition of the hormone gibberellic acid (GA) results in increased tree growth (Gunning and Hardham 1982) and may influence microfibril angle (MFA) and lignin content. A high MFA and lignin content is needed to resist gravity and wind loads, particularly for young trees (Lichtenegger et al. 1999; Reiterer et al. 1999). Also, a high MFA allows the extensibility needed during wind loading while increased lignin adds bending stiffness resistance to compression forces (Hepworth and Vincent 1998; Köhler 2000, 2002).

With increased radial cell formation, a slower growth rate occurs over time. In turn, the S2 microfibrils are oriented at a lower angle and lignin content commonly decreases (Larson 1966; Shupe et al. 1997). The correlation between lignin content and cellulose is

quite high when isolated from just the S2 layer and avoiding middle-lamella formed lignin (Gindl and Teischinger 2002). Likewise, the relationship between lignin and MFA is strong for 20 coniferous wood species (Saka and Tsuji 1987). Several investigators have reported an increase in lignin and MFA with increased fertilization, irrigation, temperature, and thinning (Wardrop and Dadswell 1955; Erickson and Arima 1974; Shupe et al. 1997; Lindström et al. 1998; Gindl et al. 2000; Wimmer et al. 2002; Sarén et al. 2004; Lundgren 2004). Indeed, the apparent relationship between growth, MFA, and lignin is not passive. McCann et al. (1993) showed that cellulose deposition, and hence MFA, dictates other polymer concentrations, spacing, and orientations. Taylor et al. (1992) found lignin deposition to vary locally with cellulose deposition within the secondary cell wall of treachery elements. As well, the lignin distribution closely follows the microfibrils in the S2 layer for *Picea abies* (Singh and Daniel 2001).

The ability to use the absorbance in the near infrared range to estimate MFA is receiving increasing attention in the scientific community. Such correlation between NIR signal and MFA have been detected both across various ages and within specific density classes for Monterey (*Pinus radiata*), loblolly (*Pinus taeda*), and eucalyptus (*Eucalyptus grandis*) (Schimleck et al. 2001, 2002, 2003, IN PRESS). However, the basis for how NIR can measure MFA is still an uncertainty (Gindl et al. 2001). Wavelengths associated with cellulose have shown considerable variation in response to MFA (Kelly et al. 2004). Such a relation to cellulose associated wavelengths seems logical but presently only empirical equations have been developed. Since the summation of cellulose, hemicellulose, and lignin theoretically sum to > 95% of the total polysaccharide content in cell walls. Assuming extractives content is negligible, only one variable may be

needed to predict MFA. Lignin content is the best candidate since it has been shown to relate to MFA and can easily be estimated from absorbance response in the NIR region (Gindl and Teischinger 2002; Baillères et al. 2002; Poke et al. 2004; Yeh et al. 2004). Development of a model to predict MFA from lignin content could be a first step in explaining how NIR absorbance can be used to estimate MFA. But perhaps more important is an understanding of how MFA and lignin co-vary. Such a relationship would be useful in plant science where both traits are known to influence plant and wood mechanical properties (Hoffmann et al. 2003).

The objectives of this study were four fold: (1) to develop near infrared models for Klason lignin, extractives, and MFA; (2) to measure the klason lignin, extractives content, and MFA from pith to bark and stump to tip; (3) to analyze the relationship between MFA and extractive-free klason lignin through empirical and theoretical means; (4) to compare theoretical and NIR spectroscopy models of MFA with interpretation.

## 5.2 Methods and Materials

### 5.2.1 Sample Preparation

Ten longleaf pine (*Pinus palustris*) trees were harvested from the location 30.6º north and 89.1º west. Periodic fires were prescribed throughout the 41 years of the stand resulting in little competition from understory plants. Seedlings were spaced 3.66 M apart in a equilateral triangle pattern. Each tree was cross cut every 4.57 meters in height, yielding 5 to 7 bolts and accompanying disk. The disk was always taken from the basal direction of the bolt. A 2 mm thick strip was cut from the disk such that the radial face was exposed for later NIR scanning. Likewise, a 10 mm wide strip was taken for microfibril measurement. The 10 mm wide strip was dissected by ring with the

tangential face showing and a mean 1 mm thickness in the radial direction. All specimens were allowed to equilibrate to room humidity and temperature. MFA samples were first air dried to room moisture content and then further dried in an oven at 105 Cº. The remainder of the disk was used for chemical analyses.

5.2.2 Chemical Analyses

The remainder of the disk was separated into latewood fractions for rings 4, 8, 16, and 32 from the pith for all heights of each tree. Therefore, the circumference of the disk was dissected into small slivers (≈ 1 to 3 mm thick by 10 mm long) of wood for each ring and height category. All slivers were milled using a Wiley mill and was allowed to pass between a 40 to 60 mesh screen. I analyzed all samples for alcohol-toluene extractives and Klason lignin by standard test methods (Tappi 1997; Tappi 1998). A replication of three was used for chemistry determination of all samples. Preliminary testing showed lignin content standard error to occur at 0.4% within an average test run (n=5 per run) and 0.69% when all test runs were pooled (n=20). For extractives content, preliminary testing showed extractives content standard error to be 0.48% within an average test run (n=12 per run) and 1.0% when all test runs were pooled (n=60).

5.2.3 Microfibril Angle Measurements

Small molecule X-ray crystallography was performed on a Bruker diffractometer. The T-value, a value closely related to Evans' variance intensity of the (002) arc, was acquired in conjunction with the MFA, a value also provided by Cave as estimated by the equation T = MFA+2*SDA where SDA represents the standard deviation of the flank angle (Cave 1966; Cave personal communication). The beam was aimed such that it was parallel to the radial direction and passed through a mean of 1 mm thickness latewood.

The sample width was parallel to the tangential surface and measured a mean of 5 mm. The height was parallel to the axis of the tracheids and measured a mean of 10 mm.

### 5.2.4 NIR Spectroscopy

I obtained the NIR absorbance measurements along the radial face strip using a Nexus 670 FT-IR spectrometer (Thermo Nicolet Instruments, Madison, WI, U.S.A.). Absorbance readings were acquired between 1000 to 2500 nm wavelengths. Forty scans were acquired at 1 nm intervals using reflectance spectroscopy and then averaged. The 1 nm intervals were next averaged into 10 nm intervals with no loss in spectral resolution (Schimleck et al. 2004). The spot size of the scanning area was 5 mm. The samples were illuminated with a lamp source at approximately 30 degrees to the sample. The path of the light was parallel to the long axis of each sample. A curtain was placed around the scanning area so as to minimize light differences between tests. A reference sample was scanned every 10 minutes to avoid any drift that might occur over time.

### 5.2.5 Calibration Development

I used the loadings from principal components regression (SAS 1999) to determine which wavelengths were most important in predicting lignin, extractives, and MFA. The wavelengths were then analyzed using stepwise regression to determine which wavelengths were significant in predicting the y variable ($\alpha = 0.10$) (SAS 1999). Coefficients and intercepts were then chosen and the model was developed for each trait. Cross validation was used to ensure stable calibration equations. No spectra pretreatment, such as derivatives or multiplicative scatter correction, was employed.

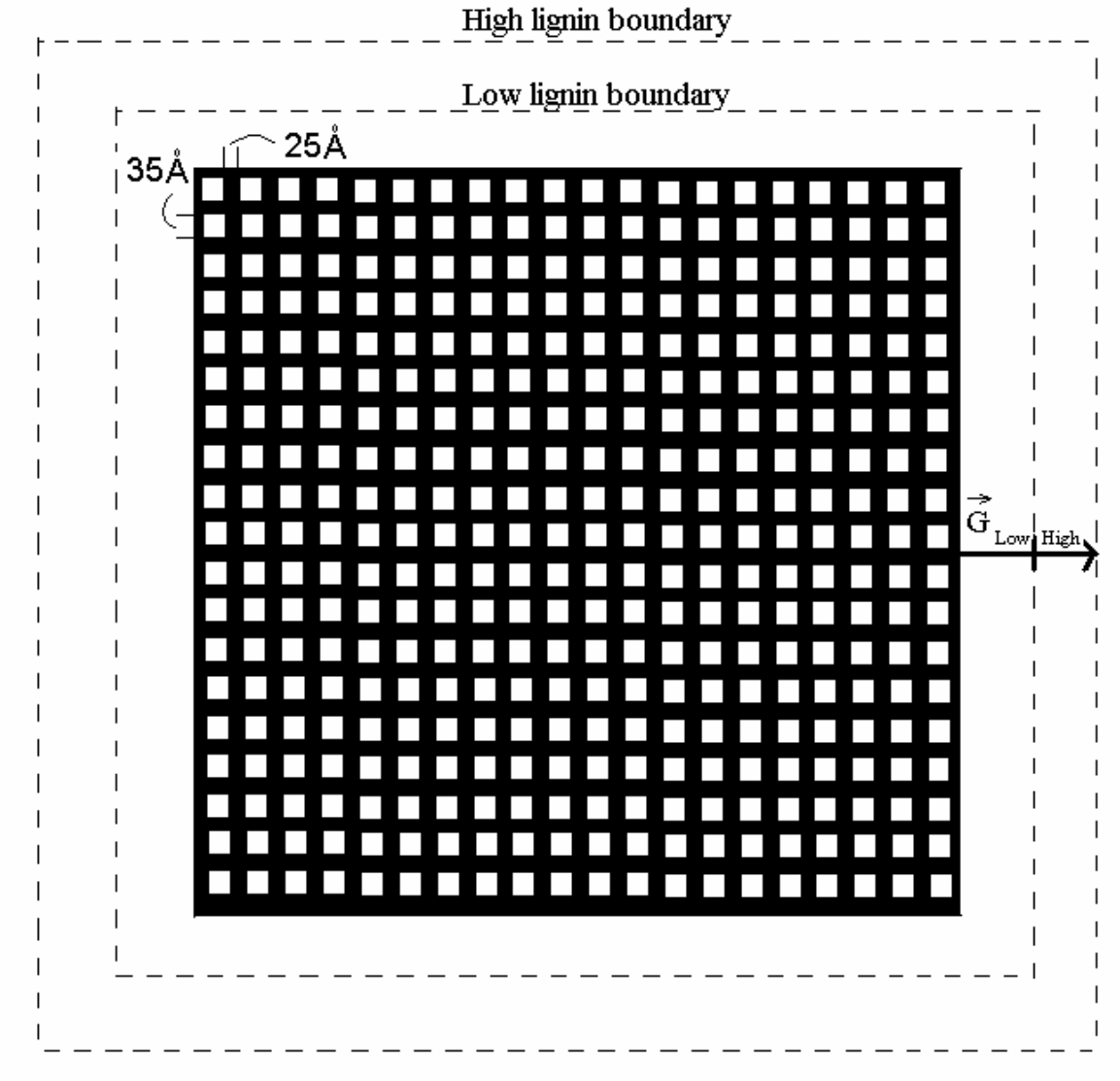



Figure 5.1: Cross section view of microfibril with cellulose elementary fibers in white, hemicellulosic matrix in black, and lignin acting as a sheath (Simplified model and distances adapted after Heyn (1969); Neiduszy and Preston (1970); Fengel (1970); and Donaldson (2001). Note that the $G_{low}$ and $H_{high}$ represent the growth vector of lignin.

## 5.3 Theoretical Model Development

My presumption of the model developed herein is that a response to growth occurs in the lignin matrix perpendicular to the axis of the micelles. Fig. 5.1 and 5.2 show the cross section and longitudinal view of the microfibril model, respectively. When MFA equals 0 degrees (Fig. 5.2a), the growth required to reach the target plane is minimal.

However, as the microfibril axis rotates, the growth needed to reach the target plane increases in a nonlinear manner (Fig. 5.2b). The growth is theorized to occur in the lignin sheath, which surrounds the microfibril (Fig 5.1). The low lignin scenario corresponds with a low growth (G vector low) while high lignin content corresponds with increased growth. To simplify the model, variation in hemicellulose distances, between cellulose crystallites, was kept constant at 25 Å and the cellulose crystallites were assumed to stay constant at 35 Å . The dimensions of the cellulose crystallites were kept constant while the lignin sheath thickness was allowed to vary as a function of MFA. The S1 and S3 layer was assumed to be negligible.

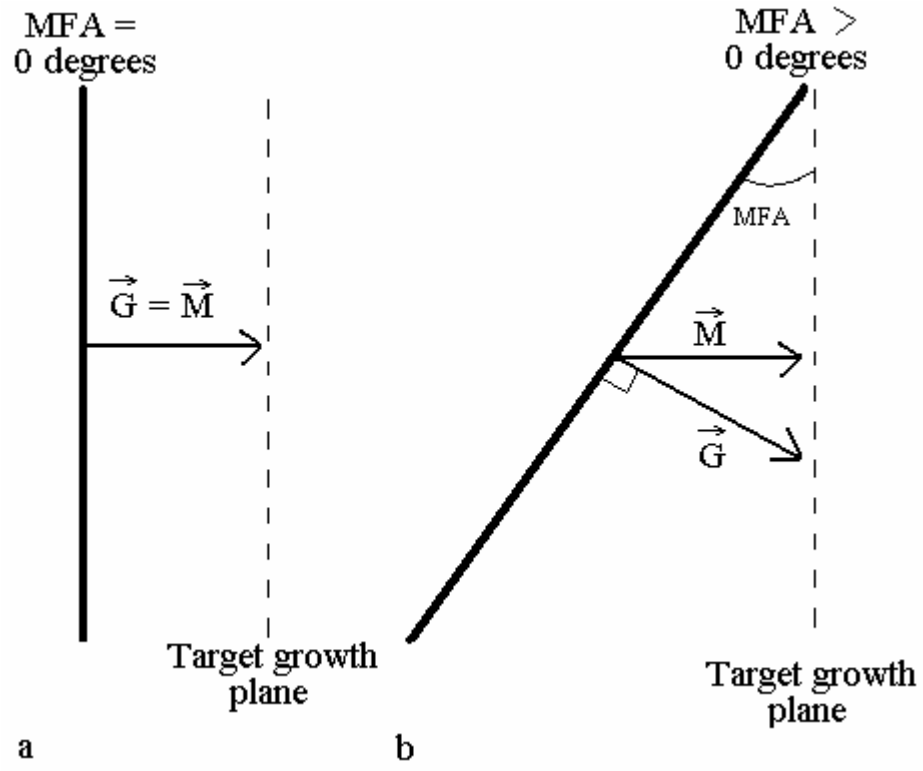



Figure 5.2: Theoretical growth showing the M vector, radial or tangential growth, and the G vector, growth of lignocellulosic matrix perpendicular to MFA, for (a) zero MFA and (b) some MFA greater than 0 degrees. The target growth plane is assumed to parallel to the tracheid axis

The exponential coefficient needed to transform the lignin data such that it linearly corresponded with MFA was found by computing an array of G→ vectors for MFAs ranging from 0 to 50º. Fig. 5.2 demonstrates conceptually how the G→ vector was computed for a given MFA and constant target plane. It should be noted that realistically, the polymer matrix doesn't grow to reach a certain plane. Instead, I assumed a fixed target plane to simulate how far the lignin matrix would have to grow for a higher MFA to reach the same plane, an assumption needed since growth occurs perpendicular to the cellulose matrix (Gunning and Hardham 1982). Without such a constraint, and with the assumption of constant growth regardless of MFA, one would expect lower growth rates in the radial direction for higher MFAs which is not the case (Shupe et al. 1996). A power of approximately 1.9 was found and therefore was rounded off to 2 which represents the two dimensional plane. The sum distance of lignin in the radial direction was computed (Figure 5.3). A constant middle lamella, lumen diameter, and cell wall thickness was assumed. The range of lignin content for longleaf pine was assumed to fall between 25 to 32 percent extractive free lignin content (including the middle lamella) (Koch 1972). A spreadsheet was used to facilitate determination of the dimensions of the lignin sheath at different lignin contents which resulted in eq. 5.1 and Figure 5.3. So for example, for 25% overall lignin content, and assuming 75% of the lignin in wood is concentrated in the S2 layer, the sum distance or thickness of the lignin sheath for one microfibril, and in the radial direction of the microfibril, was computed to be 439 Å. Likewise, 598 Å was found to be the sum thickness of lignin in the radial direction at 32% lignin content for one microfibril. Repetitive calculations were performed to derive the interpolated values shown in Figure 5.3. Next the coefficient, for

the given squared function, needed for the G vector to reach the target plane at intervals between 0 to 50° was computed. The following constraints were thus set: a piece of solid wood with 25% lignin, was assumed to have a 0º MFA, with a sum total lignin radial thickness in one microfibril of 439 Å. Likewise, for wood of 32% lignin, the MFA was assumed to equal 50º with the corresponding computed sum lignin thickness was found to equal 598 Å. As a result, the following constant and intercept was found such that the resulting range in G→ varied from 439 Å to 598 Å when the 2nd power was used (Figure 5.2):

$$G\rightarrow = 439\ \text{Å} + a\alpha^2 \qquad (5.1)$$

Where a is some constant and α is MFA. To determine the constant a, the following equation was solved:

$$598\ \text{Å} = 439\ \text{Å} + a(50^o)^2 \qquad (5.2)$$

The constant labeled "a" was found to be 0.0636. The following equation thus represented the sum thickness of the lignin sheath in the radial direction for Figure 5.1:

$$G\rightarrow = 439\ \text{Å} + 0.0636*\alpha^2 \qquad (5.3)$$

Next, a model to predict the sum thickness of the lignin sheath as a function of overall lignin percent was computed. The coefficients and intercept for the model were determined such that the range of G→ values fell between 439 to 598 Å and is presented in Figure 5.3.

$$G\rightarrow = 4.0704*(L)^2 - 212*(L) + 3199.4 \qquad (5.4)$$

Where L is lignin content percent. Eq. 5.3 was then set equal to Eq. 5.4 to yield the final equation:

$$\alpha = \sqrt{64*(L)^2 - 3333.3*(L) + 43402.52} \qquad (5.5)$$

It should be noted that lignin content was actually a function of MFA according to the model in Figure 5.2. However, to meet objective (4) and to stay consistent throughout the paper, MFA is listed as the dependent variable. Likewise, all graphs were plotted with MFA on the y-axis.

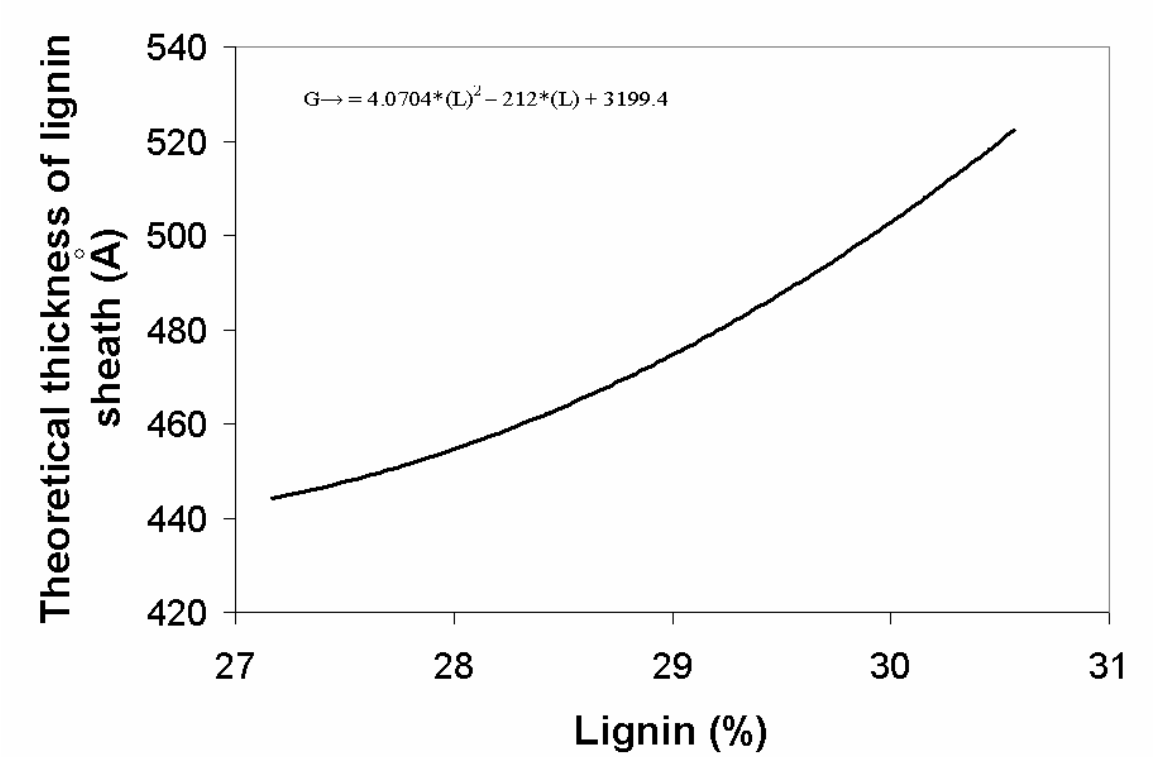



Figure 5.3: Theoretical sum thickness of lignin, in the radial or tangential direction, for a single microfibril within the S2 layer. The non-linear equation derived from this figure was used for eq. 5.4.

## 5.4 Results

### 5.4.1 Trait Variation within Trees

Table 5.1 provides NIR models developed for extractives and lignin, which were in turn used to determine within tree variation of wood chemistry. X-ray diffraction was used to determine within tree variation of MFA. The absorbance at two wavelengths were found to best predict lignin while four wavelengths were needed for extractives.

Table 5.1 Coefficients, $R^2$, and root mean square error of prediction for all models.

| Model | Constants a | b | c | d | e | f | $R^2$ | RMSEP |
|---|---|---|---|---|---|---|---|---|
| $\alpha = \sqrt{64*(L)^2 - 3333*(L) + 43403}$ | - | - | - | - | - | - | 0.40 | 9.80 |
| $A = a + b*(L)$ | 5.7 | -143 | - | - | - | - | 0.40 | 9.35 |
| NIR $\alpha^a$ = 24.57 + a*(w1015[b]) +b*(1135)+ c*(w1375) + d*(w1875) + e*(w1965) + f*(w2165) | 1014 | -314 | -1552 | 626 | -162 | 251 | 0.50 | 9.00 |
| NIR $E^c$ = 2.27 + *a**(w1145) + *b**(1345) + *c**(w1685) + *d**(w2185) | 408 | -797 | 498 | -134 | - | - | 0.71 | 3.27 |
| NIR L = *a* + *b**(w1195) + *c**(w1015) | 32 | 143 | -201 | - | - | - | 0.55 | 2.13 |

[a]NIR model for MFA was only used to compare against theoretical model and not for within tree variation determination reported in Table 5.2 and 5.3.
[b]For example, the absorbance at wavelength 1015 nm
[c]where E represents extractives content

Table 5.2 shows the within tree variation of lignin content. The mean lignin content was highest within the ring adjacent to the pith while just 4 years later, the lignin content commonly dropped by 1 to 2 percentage points. Lignin then continued to decline at a much slower rate from ring 4 to 16 resulting in only a 1 percentage point change. For most tree heights, there was no statistical difference between ring 16 and ring 30 (n=10). The exception was at a height between 22.9 to 27.4 meters where inconsistent lignin patterns seemed to emerge (Table 5.2). Just about every ring and height category had similar coefficient of variation (COV) values with no specific patterns. In short, the age from the pith controlled the variation in lignin while the effect of tree height was negligible.

Table 5.3 shows MFA variation throughout the tree. A mean drop of 10º occurred between years 4 and 8 for most heights. At 4.6 m, a 15º drop occurred between years 8 and 16 while the difference was considerably lower at heights greater than 4.6 m. The MFA continued to drop at a slower rate from 16 to 30 years of age for the lowest half of

the tree while there was not enough material available to make a conclusion at greater tree heights. There was a distinct decrease in MFA with height up to 18.3 m and then the fibril angles either stabilized or increased. The variation in MFA was large, particularly after ring 8. No difference in COV was detected between heights.

Table 5.2 Experimental mean, standard deviation, and coefficient of variation for lignin for 10 trees at different height and ring numbers. N.A. represents no experimental data.

| Extractive-Free Klason Lignin % | | | | | |
|---|---|---|---|---|---|
| Tree | | Ring Number | | | |
| Height (m) | 1 | 4 | 8 | 16 | 30 |
| 4.6 | 30.42<br>0.69 (2.28%) | 28.67<br>1.37 (4.78%) | 28.37<br>1.07 (3.79%) | 27.71<br>1.26 (4.53%) | 27.31<br>1.50 (5.04%) |
| 9.1 | 30.79<br>1.45 (4.69%) | 28.81<br>1.07 (3.73%) | 27.89<br>1.19 (4.31%) | 27.52<br>1.31 (4.78%) | 26.53<br>1.17 (3.94%) |
| 13.7 | 30.35<br>1.97 (6.48%) | 27.78<br>1.19 (4.27%) | 27.80<br>1.39 (5.02%) | 27.17<br>0.65 (2.38%) | 26.6<br>1.02 (3.51%) |
| 18.3 | 30.41<br>1.50 (4.92%) | 28.46<br>0.64 (2.26%) | 28.27<br>0.58 (2.08%) | 27.92<br>1.12 (4.03%) | N.A. |
| 22.9 | 28.57<br>1.87 (6.53%) | 28.75<br>0.56 (1.94%) | 28.40<br>0.87 (3.07%) | 29.07<br>0.75 (2.57%) | N.A. |
| 27.4 | 29.61<br>0.83 (2.81%) | 30.12<br>1.68 (5.58%) | N.A. | N.A. | N.A. |

Table 5.3: Experimental mean, standard deviation, and coefficient of variation for MFA for 10 trees at different height and ring numbers.

| Microfibril angle (º) | | | | | |
|---|---|---|---|---|---|
| Tree | | Ring Number | | | |
| Height (m) | 1 | 4 | 8 | 16 | 30 |
| 4.6 | N.A. | 40.34<br>6.3 (15.7%) | 30.42<br>10.1 (33.2%) | 15.17<br>8.5 (55.8%) | 11.38<br>6.2 (54.3%) |
| 9.1 | N.A. | 30.46<br>14.2 (46.7%) | 17.68<br>8.3 (47.1%) | 13.7<br>7.5 (54.8%) | 10.35<br>7.9 (76.4%) |
| 13.7 | N.A. | 24.78<br>5.3 (21.4%) | 12.13<br>6.1 (50.1%) | 11.1<br>5.15 (46.7%) | 7.6<br>2.3 (30.5%) |
| 18.3 | N.A. | 27.38<br>5.6 (20.1%) | 16.3<br>5.9 (36.6%) | 9.22<br>4.1 (44.7%) | N.A. |
| 22.9 | N.A. | 31.71<br>5.9 (18.5%) | 23.1<br>4.2 (18.2%) | 15.07<br>7.4 (49.1%) | N.A. |
| 27.4 | N.A. | N.A. | N.A. | N.A. | N.A. |

Extractive content decreased with age and height (Table 5.4). The highest extractive content occurred at ring 1 in the lower half of the tree while the lowest extractive content

occurred at rings 16 and 30, regardless of height. The COV for extractive content was high and appeared to be attributable to within ring variation.

Table 5.4: Experimental mean, standard deviation, and coefficient of variation for extractives for 10 trees at different height and ring numbers

| Alcohol Toluene Extractives (%) | | | | | |
|---|---|---|---|---|---|
| Tree Height (m) | | Ring Number | | | |
| | 1 | 4 | 8 | 16 | 30 |
| 4.6 | 30.48<br>15.2 (49.9%) | 17.69<br>13.9 (78.6%) | 11.19<br>8.4 (74.9%) | 3.68<br>1.8 (48.4%) | 2.60<br>1.0 (54.3%) |
| 9.1 | 27.34<br>14.9 (54.5%) | 21.65<br>11.6 (54.2%) | 8.25<br>7.2 (87.5%) | 2.46<br>1.3 (54.3%) | 1.97<br>0.85 (43.1%) |
| 13.7 | 26.7<br>16.5 (61.8%) | 13.47<br>5.17 (38.4%) | 3.52<br>2.49 (70.7%) | 2.53<br>0.89 (35.6%) | 1.67<br>1.2 (73.34%) |
| 18.3 | 18.77<br>7.17 (38.2%) | 8.87<br>4.65 (52.5%) | 3.26<br>1.17 (35.9%) | 2.63<br>0.84 (31.7%) | 2.38<br>1.24 (52.0%) |
| 22.9 | 11.80<br>6.8 (57.4%) | 4.63<br>1.34 (29.0%) | 3.12<br>0.75 (24.1%) | 2.37<br>0.54 (23.0%) | N.A. |
| 27.4 | 5.96<br>0.96 (16.1%) | 4.45<br>0.90 (20.4%) | N.A. | N.A. | N.A. |

### 5.4.2 Lignin Content to Microfibril Angle Relationship

Figure 5.4 experimentally shows MFA to increase with an increase in lignin content but with considerable variation. When the residuals were investigated (not plotted), the variation in the residuals increased from 25 to 29% lignin content and then stayed constant from 29 to 32% lignin content. The $R^2$ (0.40) I observed was not as high as that observed in a separate study where only the lignin in the S2 layer was measured (Table 5.1) (Gindl and Teischinger 2002). The variation in middle lamella lignin composition probably added to the variation resulting in a decreased $R^2$ for this study. The RMSEP for the theoretical model was 9.8 while the empirical model was 9.35. I found extractives content to not influence the MFA to lignin content relationship.

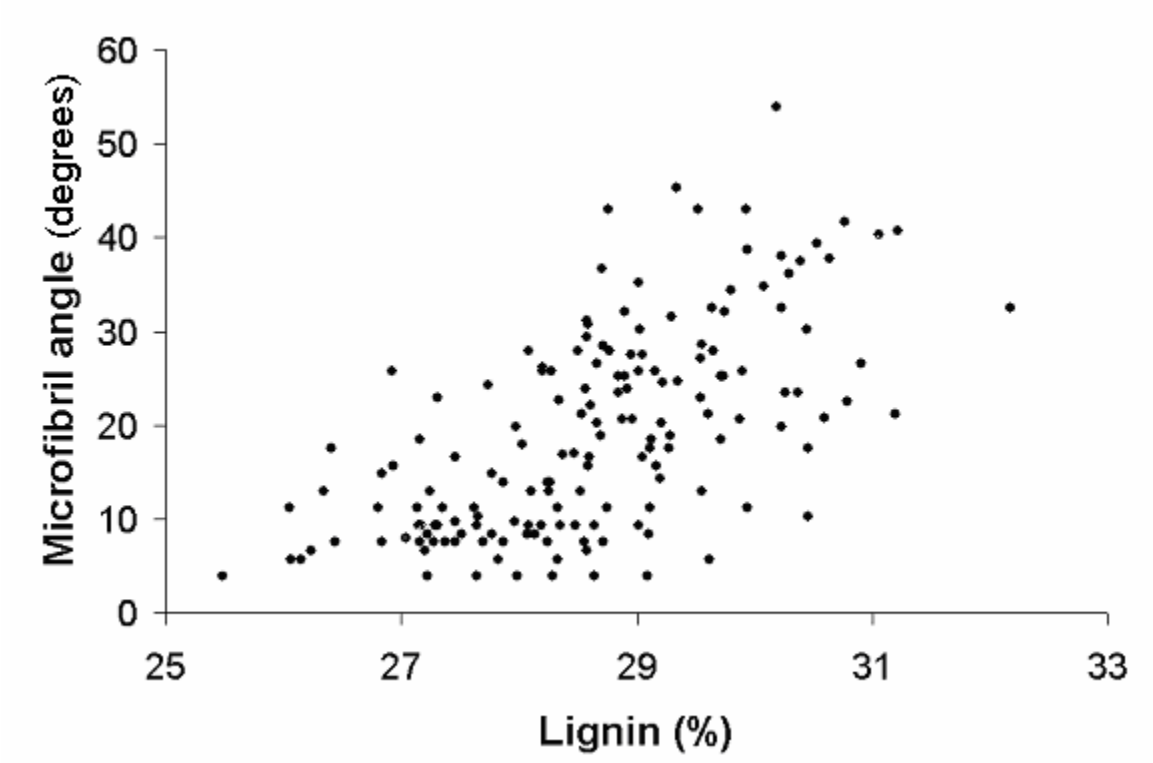



Figure 5.4: A scatter plot of experimental MFA and lignin content.

Figure 5.5 shows the log maximum likelihood results from 25 possible families of transformations using the Box-Cox method (Draper and Hunter 1969). The peak on Figure 5.5, represents the recommended statistical transformation by SAS (1999) that could be applied to lignin content when predicting MFA. The peak occurred at 0.5 which is the square root function, the same transformation I choose by theoretical means (Eq. 5.1 – 5.5).

I constructed an NIR absorbance model to predict MFA from lignin content (Table 5.1). Six wavelengths were needed to predict MFA while the theoretical model needed only lignin content. Figure 5.6 shows a scatter plot comparing the capacity of NIR and theoretical model to predict lignin. There was no significant difference in slopes or intercepts when compared in only these two dimensions (Figure 5.6). The two trend lines

were statistically parallel as determined by the 95% confidence intervals around the coefficient estimates.

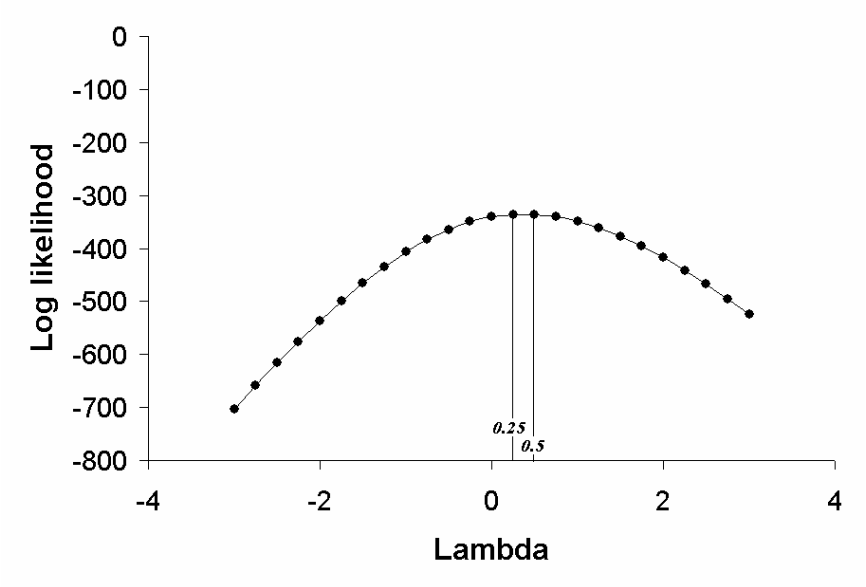



Figure 5.5: Log maximum likelihood results from Box-Cox transformation procedure with microfibril as the y variable and lignin % as the x variable. The 0.5 represents the square root function transformation.

A better $R^2$ (0.70) was achieved for MFA when principal components regression was used with 8 factors to explain the variation. However, for interpretation purposes, and to remain consistent with other NIR models, I used multiple linear regression.

## 5.5 Discussion

### 5.5.1 Wood Property Variation

I found the mean lignin content was consistently higher than that reported for the innerwood and outerwood of loblolly pine (Shupe et al. 1996; Shupe et al.1997). However, the results for this study were almost identical to those reported for longleaf pine where Koch (1972) found lignin to vary between 25 to 30%. There did not appear to be any distinct effect of height on lignin. Such a result was notable since MFA and extractives content showed distinct height differences but higher values of COV.

The reason for such high variation in extractives content was due to within ring variation. Quite often, on solid wood pieces, a 10-mm interval within a ring would show distinct visual differences in color. When the spectra absorbance curve was analyzed at congruent regions, the spectra would drastically rise when the discolored brown areas were scanned. Likewise, the same visual change in color was observed in ground wood, particularly in rings 1 and 4. A discolored brown tint occurred in rings 1 and 4 while samples further from the pith exhibited a light yellow shade. When I observed the darker color, the extractives content almost always measured in excess of 7% (Tappi 1997, 1998).

The reason for the high variation in MFA was less clear. Shupe et al. (1996) showed COV around 10% for both inner and outer wood of loblolly pine while for this study, the COV ranged between 10 to 40%. However, the standard deviation in MFA for this study averaged 6.7 while that for loblolly pine averaged 3 to 5 suggesting that the COV should be carefully interpreted (Shupe et al. 1996). It is noteworthy to mention that Shupe et al. (1996) used a light microscope, which could avoid microfibril deviation near pit regions. However, it should also be noted that these samples were difficult to prepare to a uniform dimension due to the small radial widths of many latewood rings. This made acquisition of the T-value distribution, and hence MFA, more difficult.

### 5.5.2 Comparison of Models

The ability to theoretically model MFA from lignin content (Eq. 5.5) was nearly as good as least squares regression for the experimental data (Table 5.1). Eq. 5.5 slightly better accounted for the change in slopes with increasing lignin content, when compared to least squares regression, by applying the square root to the lignin side of the equation.

However, given the experimental variation in microfibril angle for a given lignin content (Figure 5.4), both a linear and the square root model fell within the confidence intervals derived for the data (Eq. 5.5). Nevertheless, in support for Eq. 5.5 was the results from the Box Cox method which recommended the square root transformation over the linear transformation on the experimental data (Figure 5.4). Such a finding lends support for Eq. 5.5 since the theoretical model and hypothesis tests were developed a priori.

After the experiment was performed, new experimental evidence specific to the S2 layer was published (Gindl et al. 2004). In that study, lignin content was found to increase with increased microfibril angle; however, the variation in lignin was too great between 0 to 35 degrees to discern whether the relationship is linear or similar to Eq. 5.5. Nevertheless, when the compression wood microfibril angle and lignin data were omitted and the estimates (lignin estimated to ±1%) compared to the results of this study, Eq. 5.5 was statistically similar in all coefficients to Gindl et al. (2004) except for the intercept which was 4 degrees offset. When their compression wood data was included, Eq. 5.5 did not acceptably predict the shape of the curve supporting that Eq. 5.5 may not work well under extrapolation conditions for compression wood samples. For our study, compression wood was avoided and was not accounted for in the theoretical model.

The theoretical model developed in Eq. 5.5 did not account for tree height. It should be noted that, in Tables 5.2 and 5.3, MFA decreases with height up to 18.3 m while lignin content remained fairly stable suggesting that height plays an additional role in the relation between lignin and MFA. Not enough data was present to confirm such a hypothesis in this paper but would be worthy of study.

An NIR absorbance model was also built in hopes of shedding light as to how NIR absorbance might correlate with MFA. The NIR model had an RMSEP of only 0.80º lower than the theoretical model; yet, the NIR absorbance model relied on 6 independent variables compared to the one lignin variable in Eq. 5.5. However, since lignin was measured from the absorbance at 2 wavelengths (Table 5.1) then perhaps 2 independent variables were actually involved. Nevertheless, the lower number of variables was desirable and may result in a more robust model across populations, although that hypothesis would have to be tested. The equal slope and intercept, as well as the fairly balanced residuals, support that NIR spectroscopy was apparently dependent on lignin content when predicting MFA (Figure 5.6). Schimleck et al. (2002) found five wavelengths were important in predicting MFA from absorbance, 3 of which were associated with cellulose. However, Schimleck et al. (2002) reported the 1510 and 2458 nm wavelengths to be important in predicting MFA from absorbance response. While neither of these two wavelengths was used in our NIR absorbance models (Table 5.1), principal component regression loading on our data set showed these two wavelengths to be in the vicinity of lignin associated peaks. Additionally, lignin content and cellulose should be somewhat multicollinear since they are measures of percent composition.

The theoretical equation suggested that the physical chemistry dimensions do indeed change with a change in MFA (Figure 5.1). These end results agree with the Emons string model which also predicts an increase in MFA with increased width of the microfibril although the mechanism is different and is analogous to wrapping a string around a tube (Emons et al. 2002).

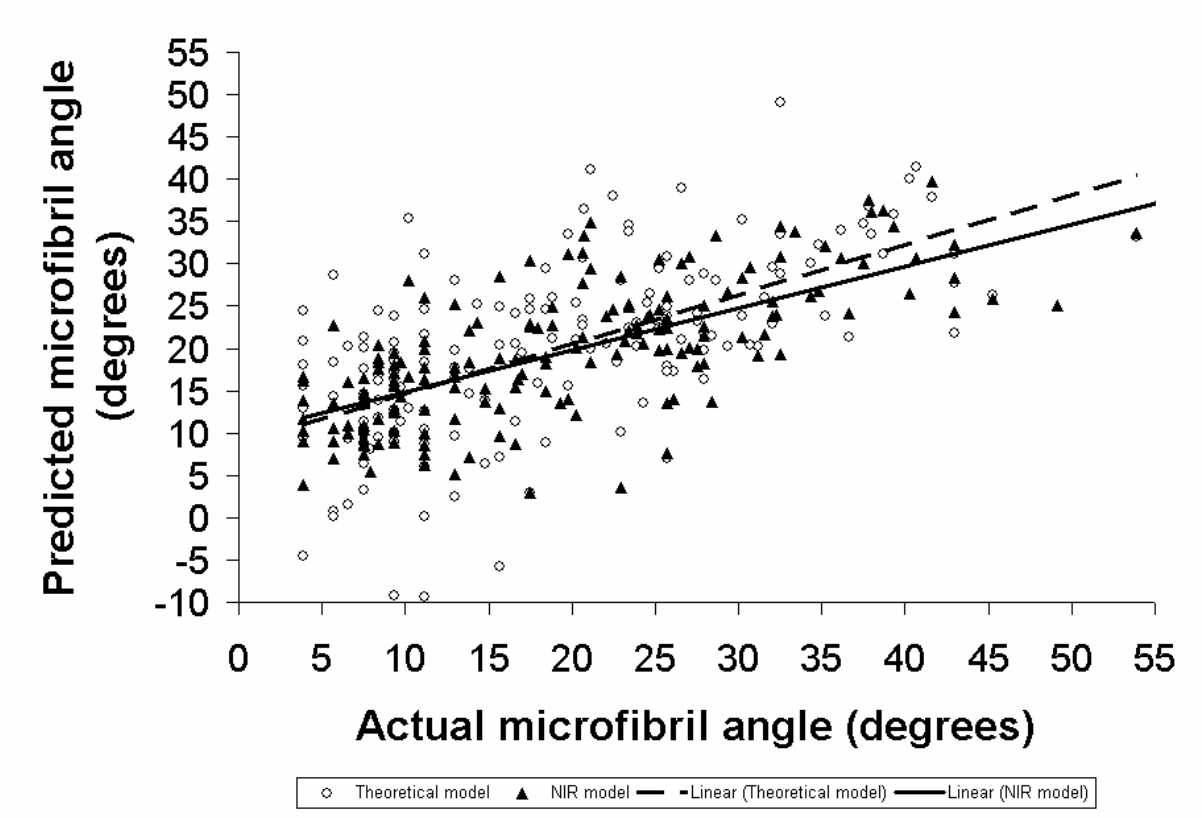



Figure 5.6: A scatter plot of predicted versus actual MFA for the theoretical and near infrared spectra model

Such a relationship between MFA and lignin may be useful in genetics studies if lignin genetically covaries with MFA. However, if a large environmental covariance between lignin and MFA exist, than improper genetic analysis may occur. Pot et al. (2002) give an example of a genetic covariance matrix between wood chemistry and fiber morphology with the best scenario, for indirect selection of a trait, to be zero phenotypic correlation and strong genetic correlation. Similar exploration of the genetic covariance between wood lignin and MFA is recommended before using NIR to estimate MFA. Some tree improvement programs have already used NIR spectroscopy as a tool to help them decrease lignin content through selection (Yeh et al. 2004). In doing so genetics programs may have also lowered MFA if the two are genetically correlated.

I recommend more testing of Eq. 5.5 for more trees at a single age to understand the level of predictability that occurs. Holding age constant can be important since different genes may be important at different ages in their effect on MFA and lignin content. Finally, since the theoretical model was developed specifically for longleaf pine, it is unknown how well it will work for species of different anatomical, physiochemical, and Å level characteristics.

# CHAPTER 6

# THE SENSITIVITY OF NEAR INFRARED SPECTROSCOPY TO TWO WOOD BLUE STAINING FUNGI

## 6.1 Introduction

Near infrared spectroscopy (NIR) is becoming an important tool to quantitatively assess the chemical, mechanical, and physical properties of wood (Hoffmeyer and Pedersen 1995; Antti et al. 1996; Gierlinger et al. 2002; Meder et al. 2002; Schimleck and Evans 2002; Schimleck et al. 2003; Tsuchikawa and Yamato 2003). An understanding of how blue stain influences the absorbance at different wavelengths would be useful knowledge when either modeling the concentration of blue stain or accounting for blue stain co-variation with another predictive variable. Zulpa et al. (2003) found blue stain to influence the absorbance between 190 and 700 nm but only mentioned where maximum absorbance occurred. At 700 nm, nitrogen influences absorbance and could therefore affect the spectra when melanin is present (Table 6.1).

Melanin is the compound left by invading fungi causing wood discoloration. In the tree, melanin can impede the rate of water conductance causing tree mortality or suppression. In wood products, melanin causes an undesirable discoloration that can lower product value. Fungal hyphae are located in parenchyma cells and resin canals but can also grow in tracheids (Liese 1970; Bell and Wheeler 1986; Zink and Fengel 1988; Zink and Fengel 1989; Behrendt and Blanchette 2001). The two most prominent melanin producing genera are *Leptographium* spp. and *Ophiostoma* spp. The former is more virulent to the host, grows twice as fast in media and sapwood, and can withstand an oxygen deficient environment (Uzunovic and Webber 1998; Neal and Ross 1999;

Solheim *et al.* 2001). Alternatively, *Ophiostoma* spp. grows at a moderate rate and occurs more in sapwood than does *Leptographium* spp. which occurs more frequently in the roots (Otrosina *et al.* 1997). These fungi are found in the bole or roots of the tree based upon their association with their insect vectors. *Ophiostoma* spp. most likely does not need to mechanically travel as far as *Leptographium* spp. since *Ophiostoma* spp. attacks the phloem region where water transport may aid in longitudinal movement of the fungi. Solla and Gil (2002) support this theory and found *O. novo-ulmi* infection to increase as vessel size increased. An increase in vessel size can increase sap flow by as much as the $4^{th}$ power (Giordano 1978). However, when the fungi begin to colonize a specific region, such as in diseased roots, the hyphae are believed to obstruct bordered pit pairs and lumens causing cessation of water transport (Nelson 1934; Wilcox 1970). As a result, more research is needed to understand the interaction between fungi and tree moisture transport since stain commonly suppresses tree growth which could result in forest fires or tree mortality. Rapid assessment techniques like NIR may offer a low measurement error but first, the wavelengths most influenced by stain needs to be identified.

Fungi to wood interaction are important in understanding which wavelengths may be influenced in the presence of stain. Wang *et al.* (1995) reported that for *O. piceae* and *O. ainoae*, the resin acid content decreased by up to 67% due to fungal growth. Likewise, in lodgepole pine (*Pinus contorta* Loud.), the presence of *O. piceae* for two weeks reduced the resin acid content by 66% (Gao *et al.* 1994). For quaking aspen (*Populus tremuloides* Michx.) logs, the total extractives content decreased by 30% after inoculation with *O. piceae* and *O. pluriannulatum* (White-McDougall *et al.* 1998).

Table 6.1: Chemical or polymer assignments to near infrared wavelengths.

| Chemical bond or polymer association | Wavelength (nm) | Reference |
|---|---|---|
| Glucose (O-H vibration) | 1480, 1580 | Osborne and Fearn 1986 |
| Cellulose | 1678, 2132, 2332, 2460 | Martin and Aber 1994 |
| Cellulose | 1766, 2140 | McLellan et al. 1991a |
| Cellulose | 1754, 1898, 2076, 2252 | McLellan et al. 1991b |
| Hot water extractives | 2300, 2100, 1900 | Michell 1995; Gierlinger et al. 2002 |
| Lignin | 870, 1230, 1275, 1438, 1460, 1505, 1600, 1660, 1708, 1828, 1445, 2154, 2218, 2386 | McLellan et al. 1991 and 1991b, Meglan and Kelly 2000 |
| Components associated with melanin | | Zink and Fengel 1989; Zink and Fengel 1990 |
| Nitrogen | 700-1100 | Zulpa et al. 2003 |
| Nitrogen | 1650-1770 | Riley and Còcaves 2002 |
| Nitrogen | 1686, 1978, 2170 | Martin and Aber 1994 |
| Nitrogen | 2090, 2174, 2378 | McLellan et al. 1991 |
| $CH_3$ | 740, 900, 1015, 1152, 1360, 1695, 1705, 2280 | Osborne and Fearn 1986 |
| C-H | 1225, 1440 | Osborne and Fearn 1986 |
| C=O | 1920, 1950, 2030, 2294 | Osborne and Fearn 1986 |
| ROH | 738, 970, 1410, 1520, 2080, 2380 | Osborne and Fearn 1986 |

The carbohydrates within wood are another site of reaction for blue stain fungi which might in turn influence NIR spectra. When both *Leptographium* spp. and *Ophiostoma* spp. interact with wood, the xylose and galactose content doubles while the mannose and glucose content decreases (Fleet et al. 2001). Given that glucose unit in the cellulose polymer are in a crystallized state, it is probable that most of the glucose reduction occurs in the hemicellulose matrix. In Norway spruce [*Picea abies* (L.) Karst.], sucrose concentrations decrease immediately after inoculation of *Ceratocystis polonica* (Viiri et al. 2001ab). Terpene concentration also changes sharply with respect to inoculation of *C.*

*polonica* (Viiri et al. 2001b). Finally, for the melanin residue left by the fungi, mid infrared spectroscopy shows significant peaks for OH, C-H, C=O, C-$CH_3$, and C-H bonds (Zink and Fengel 1990). But perhaps most critical to this study, the percentage of nitrogen in the melanin compounds was between 1.5 to 2.5% content (Zink and Fengel 1988). Since the stem, branch, and roots of most woody species commonly fall between 0.02 to 0.12% nitrogen content (Likens and Bormann 1970), the addition of melanin to wood from blue staining fungi may greatly increase the total nitrogen concentration in a localized portion of the wood substrate.

The objective of this research was to determine which regions of the NIR spectra were significantly influenced by the inoculation of *O. minus* and *L. serpens* into longleaf pine (*Pinus palustris*). In addition to the raw spectra, the 1st and 2nd derivative of the data was investigated since it is a common pretreatment in calibration modeling. Finally, there was a desire to determine if the spectra differed in response to *O. minus* versus *L. serpens* stain treatments. Such objectives should be very useful for using NIR to identify when stain is present or to avoid the influence of stain when predicting other wood quality traits.

## 6.2 Materials and Methods

### 6.2.1 Increment Core Sampling

I randomly selected fifty increment cores from longleaf pine trees in a 41 year old plantation on the Harrison Experimental Forest which is owned and maintained by the USDA Forest Service near Saucier, Mississippi (U.S.A.). The location was 30.6º north and 89.1º west and the under-story of the site was free of competition due to periodic fires prescribed over the lifetime of the stand. The increment core specimen went from

bark to pith to bark for all samples. Tree identification numbers were recorded to ensure minimal genetic relation between trees. Five increment cores were randomly chosen to determine if blue staining fungi were present in existing trees. No presence of stain fungi was microscopically found. Furthermore, no blue stain discoloration had occurred on any of the 50 increment cores taken from the site; neither when green nor after air drying. Finally, local forestry experts reported no major bark beetle attacks for 2 years prior to sampling.

I used forty of the fifty increment cores for the final study. Each increment core was broken in half and sawn perpendicular to the core length such that the radial face was exposed. One half of the core was inoculated with the fungus treatment while the other half served as a paired control. Two stain fungi were used for this study, *L. serpens* and *O. minus*.

6.2.2 Inoculation and Colonization of *O. minus* and *L. serpens*

The fungal species I used in this experiment were isolated from longleaf pine roots and southern pine beetles from the Palustris Experimental Forest, USDA Forest Service, Louisiana, U.S.A. (*Leptographium serpens*) and the Bankhead National Forests, Alabama, U.S.A. (*Ophiostoma minus*), respectively. The isolates were cultured on malt extract agar (MEA) (2% malt extract agar), in the dark for comparison to species described in Jacobs and Wingfield (2001), and representative stock cultures were maintained on MEA slants at 4 ºC.

Each isolate was inoculated onto thirty MEA plates and allowed to grow for seven days at 22ºC ± 3. A dipping slurry was created for each isolate to inoculate the cores by blending thirty plates of inoculum with 500 ml $H_2O$ for 1 min. I then dipped the cores in

the inoculum and placed in separately labeled sealed bags. None of the sealed bags were sealed and needle pin holes were punched into the bags to promote ventilation. The labeled bags were placed in a moisture chamber where the humidity was controlled at 95% ± 2(%) and the temperature was 22ºC ± 3(ºC) (Wang 1994; Viitanen 1997; Jacobs et al. 2000).The moisture chamber was a dessicator containing 1000 mL of water below a ceramic plate agitated by a magnetic stir bar. Two dessicator setups were used, one for each fungal treatment. Condensation commonly occurred inside the plastic bags holding the increment core. After a week of growth, each increment core was scrubbed with sterile tissue-wipes to remove excess inoculum and then placed back into the moisture chamber. Sterile rubber gloves were used to handle the stained specimens and core cleaning was completed in a laminar flow hood under sterile conditions to avoid contamination by airborne spores. Growth was allowed to occur for a 6 week duration and then the specimens were taken out of the plastic bags and allowed to air dry (Viitanen and Bjurman 1995). The surface of the specimens dried within a day ensuring that the surface was below fiber saturation point as quickly as possible so as to stop surface growth of any contaminating spores. The entire increment core took less than a week to dry based on weight scale measurements. Nevertheless, two months was allowed before testing to ensure through drying.

I dried the control half of the increment cores for two months at ambient room temperature and humidity. After two months, the increment cores reached equilibrium with the environment as indicated by gravimetric weight.

6.2.3 Processing and NIR Scanning of Increment Cores

I sawed the increment cores, control and inoculated halves, with a fine toothed-band saw after drying. A clean radial face was exposed with minimal deviation from the radial plane. Any sample that showed significant deviation from the radial plane was identified and removed from data analysis. Rings from the bark to pith were counted with the first ring being year 41. Years 40, 30, 22, 17, 12, 9, 7, 5, and 3 were marked on each control and inoculated core. The markings were made away from the scanning area so as to not contaminate the spectra. A total of 164 and 140 rings/data were available for *L. serpens* and *O. minus*, respectively while 304 paired control samples were available.

NIR absorbance was obtained using a FieldSpec FR (Analytical Spectral Devices Inc. Boulder, CO. (USA) at wavelengths between 350 and 2500nm. I collected thirty scans and averaged into a single spectrum. Spectra were obtained from the latewood region of each marked year on the control and inoculated half of the increment core. Since the inoculated (or control) broke down into two clean faces, the cleanest cut face was used for scanning. Any surface burns attributable to the band saw were not considered for NIR scanning.

6.2.4 Statistical Analysis

Significant differences between mean absorbance at a given wavelength were determined using the 95% confidence interval (Ott 1993)

$$\hat{y} \pm t_{\alpha/2} * (\frac{s}{\sqrt{n}}) \qquad (6.1)$$

with $\hat{y}$ being the mean absorbance, *s* being the standard deviation of absorbance for a given wavelength(s), n being sample size, and t being the t-value at a given confidence ($\alpha$

= 0.05). Additionally, the 95% confidence interval was an indication of the precision of the mean absorbance for a given wavelength(s) for the sample populations examined.

I pretreated the spectra by 6 different methods: a) original spectra of wood and stain, b) 1st derivative of wood and stain, c) 2nd derivative of wood and stain, d) original spectra of paired data, e) 1st derivative of paired data, and f) 2nd derivative of paired spectra. The computer software used to create the 1st and 2nd derivative was which used the Savitzky-Golay method of transformation (Unscrambler 1999). For the matched samples, the spectra from the control (clear wood variation in absorbance) were subtracted from the stained wood spectra for the original, 1st, and 2nd derivative. Paired data was possible since half of the increment core was used as a control while the other matching half was used for stain treatment.

I used the F-test to determine if variances differed between the absorbance of solid wood and stained wood (Ott 1993).

$$F = \frac{s^2_{stain}}{s^2_{control}} \qquad (6.2)$$

When the F estimate exceeds a tabulated value for a given (p-value < 0.05), t-value, and degree of freedom, then the variation between the two populations were deemed significantly different. Such information was used to determine if the addition of blue stain added to the variation in absorbance already present in the clear wood.

## 6.3 Results

### 6.3.1 Clear wood (control) versus Stained Wood

Both *L. serpens* and *O. minus* treatments resulted in a similar absorbance response per wavelength for most of the wavelength range (Figure 6.1a). However, the control wood differed from the stained wood for both staining fungi. A substantial difference occurred

between 424–1104 nm (Figure 6.1a). Other differences in mean absorbance between both stain and control treatments were detected between 354-364 (uv) and 1114-1354 nm. The differences were partially detectable due to the large sample size as listed in Table 6.2. The upper region of the spectra (1894-2500) appeared to differ in absorbance, but was not statistically different (p-value > 0.05) due to the increased variance. The variation in absorbance generally increased with increasing wavelength which is typical for spectra collected from wood and made detection of differences in the upper part of the spectra region more difficult for a given sample size.

The variation in absorbance between stained wood and the controls were determined to significantly differ at wavelengths between 424-1324 nm (p-value<0.05). *O. minus* and *L. serpens* treatments exhibited lower variance in absorbance than the control treatment. For wavelengths greater than 1324 nm, there was no statistical difference in variance between the control and stain treatments.

Figure 6.1b shows the mean 1$^{st}$ derivative pretreatment of the spectra. From the graph, all three treatments fell on top of one another throughout the entire wavelength range. I detected significant differences in part due to the large sample size (Table 6.2). However, when the variation in absorbance of wood was compared to the variation in absorbance of wood and stain for the first derivative, there was no significant difference in absorbance for wavelengths greater than 1000 nm. A lack of difference in variance suggests that wood absorbance dominates the variation relative to variation attributable to stain.

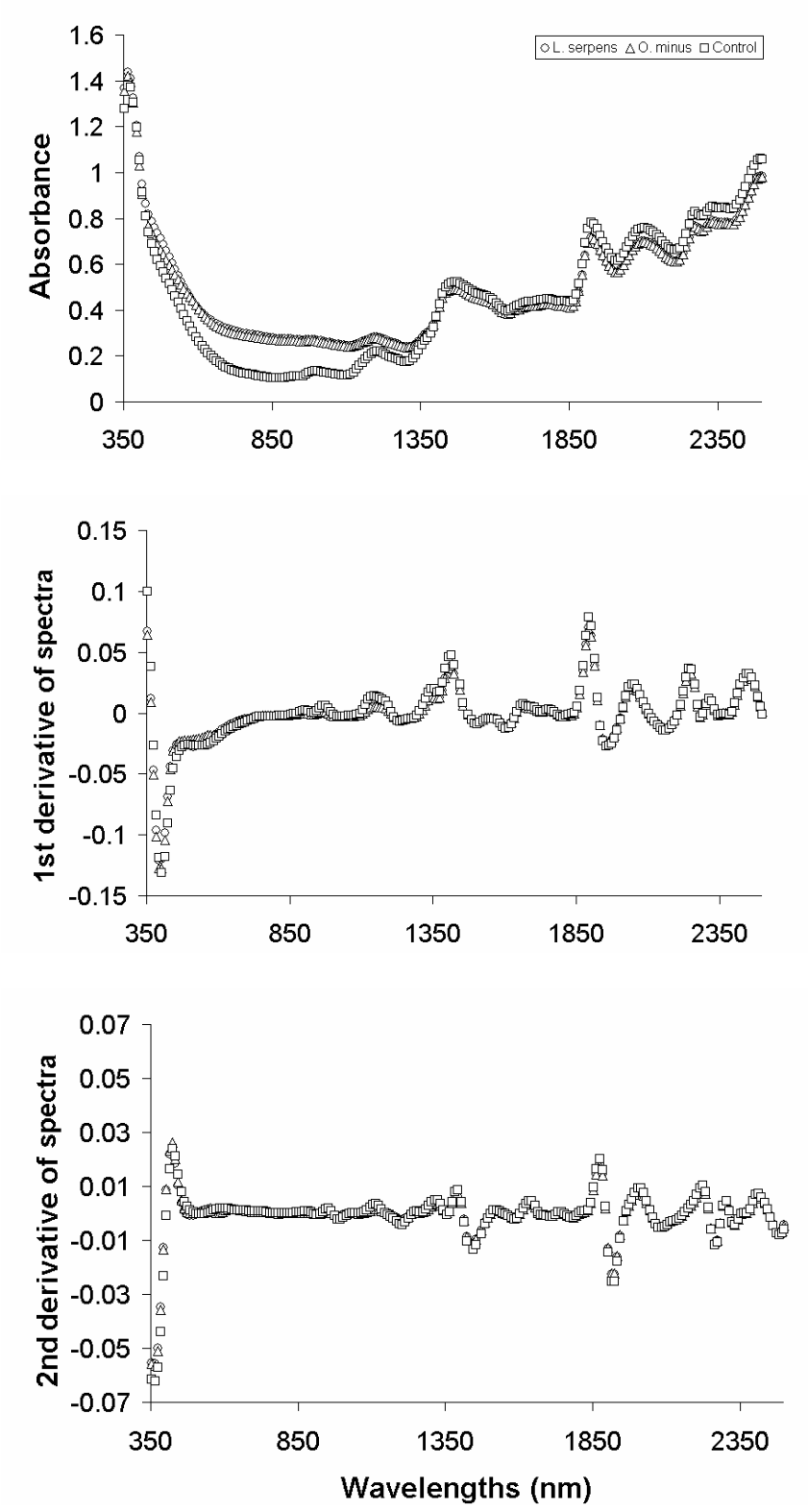



Figure 6.1: a) Absorbance between wavelengths 350 through 2500nm for untreated wood and wood treated with *L. serpens* and *O.minus,* b) 1st derivative of absorbance, and c) 2nd derivative of absorbance.

Figure 6.1c shows the mean absorbance after 2nd derivative pretreatment of the spectra. By my visual assessment, no wavelengths stood out which might classify either stain treatment from the control. However, as Table 6.2 confirms, there were significant differences at various wavelengths throughout the range. The highest difference in absorbance occurred between 374 to 754 nm. The variation in absorbance for each treatment was not plotted on this graph since at a given wavelength, the confidence interval of absorption often varied wildly between treatments when the 2nd derivative was applied. These oscillations were not present in the raw data and 1st derivative pretreatments. The severe fluctuations in confidence intervals were an indication of the second derivative to detect increased noise over signal.

6.3.2 *L. serpens* versus *O. minus* Treatment

To determine the *in-situ* absorbance patterns of *L. serpens* and *O. minus*, I subtracted the spectra of the clear wood from the wood stained spectra. The absorbance at 424-504 nm was significantly different for the two fungi species. The slope of *L. serpens* was steeper than *O. minus* between 414-474 nm as determined statistical tests (p-value < 0.05).

The first derivative of the paired analysis was plotted on Figure 6.2b. At wavelengths between 484-554 nm, respectively. *O. minus* had a significantly greater absorbance than *L. serpens*. *Leptographium serpens* and O. *minus* exhibited a peak and valley of 504 and 564 nm respectively (Figure 6.2b). The valley exhibited the largest statistical difference in absorbance between the two genera. The remainder of the wavelengths, 564-2500 nm showed no significant difference between the *L. serpens* and *O. minus* treatment. Similarly, only between 444-504 nm did *L. serpens* and *O. minus* significantly differ in

absorbance when the 1st derivative pretreatment was applied (Figure 6.2c). When the second derivative pretreatment was applied, *O. minus* and *L. serpens* differed at 474-554 nm as determined by p-values less than 0.05.

When the wood variation in absorbance was not removed from the overall spectra (Figure 6.1a), there was no single region where *L. serpens* and *O. minus* differed. However, when the 1st derivative was computed (Figure 6.1b), *L. serpens* differed from *O. minus* between 474-554 nm which agreed with the paired analysis of Figure 6.2a-c.

## 6.4 Discussion

### 6.4.1 The Influence of Stain on Spectra

The response of spectra to blue stain invasion is significant if it increases error or adds bias to NIR signals. One method to build NIR predictive models is to avoid using wavelengths associated with undesirable contaminates or uncontrollable environmental variables (Thygesen and Lundqvist 2000; Swierenga et al. 2000; Wülfert et al. 2000). In wood, the prediction of mechanical and physical properties from NIR spectra is viable. Since blue stain is a common contaminant of lumber, one needs to be careful when building NIR predictive models. However, using wavelengths most associated with blue stain may be useful in classification of clear wood, stained wood, and perhaps even fungus species of stained wood. Such is particularly true in plant pathology where one studies beetles that transfer the fungi within the tree. Such growth of the fungi causes blue stain and can disrupt the hydraulic functions of the tree stem, branches, and roots. As such, trees often obtain disease and die.

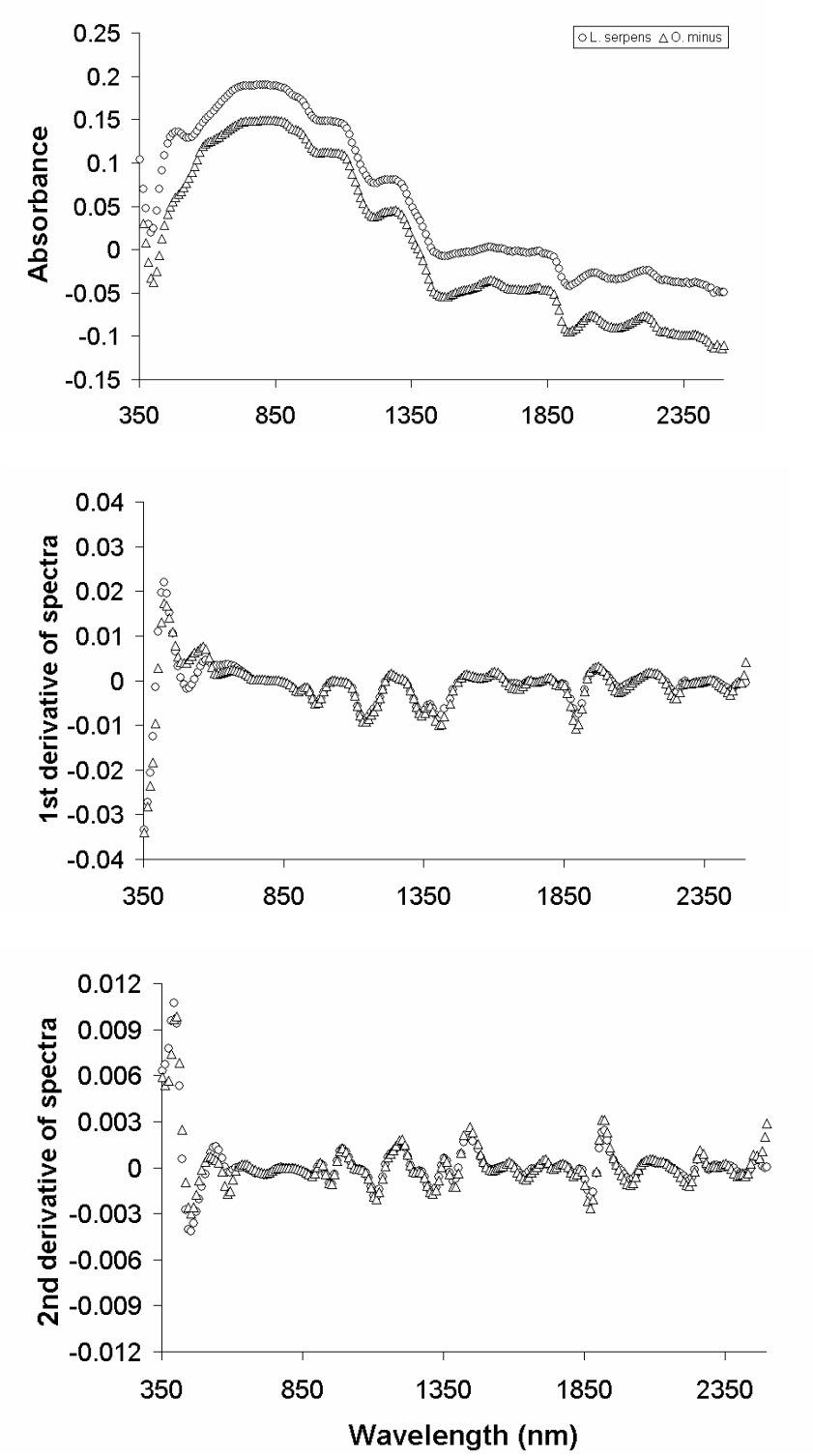



Figure 6.2: a) Absorbance between wavelengths 350 through 2500 for paired data where the spectra of stained wood was subtracted from clear wood to yield the spectra of the *L. minus* and *O. minus*, b) 1st derivative of absorbance, and c) 2nd derivative of absorbance.

As shown in Figure 6.1a, wavelengths 654-1114 nm exhibited the largest difference in mean absorbance for clear wood when compared to wood stained by *L. serpens* and *O. minus*. It has been reported that blue stain is composed of melanin, which possesses much higher nitrogen concentrations than unstained wood (Zink and Fengel 1990). As the fungi compete for nutrients, the nitrogen in the host is concentrated by 20 fold when extracted from the phloem and excreted in feeding chambers (Hodges and Lorio 1969; Eckhardt et al. 2004). The close parallel between Qifa and Jihua (2003) and this study suggest that the increased nitrogen concentration is a potential cause for the increase in absorbance between 654-1114 nm.

Wavelengths greater than 1650 nm can be influenced by nitrogen and potentially blue stain (Table 6.1). However, significant differences in absorbance between clear wood and blue stain infected wood did not occur at wavelengths greater than 1650 nm for the original spectra (Table 6.2). Table 6.1 shows that wavelengths greater than 1650 nm are commonly influenced by lignin, cellulose, extractives, and hemicelluloses. Given that the concentration of wood polymers is of many magnitudes greater than nitrogen concentration (Likens and Bormann 1970), it was of no surprise that blue stain differences could not be detected at wavelengths greater than 1650 nm. However, when the 1$^{st}$ and 2$^{nd}$ derivative was applied, differences between the control and blue stain could be detected at some wavelengths above 1650 nm (Table 6.2). The reason differences was detected was probably attributable to the removal of any baselines shift that occurred in the original spectra.

For wood, an increase in density will shift the spectra upward (Via et al. 2003). Such large variations in the original spectra are likely to dominate absorbance response making

smaller variations, such as nitrogen concentration in melanin, harder if not impossible to detect. Taking the 1st or 2nd derivative is a common spectroscopic technique to reduce baseline variation. Still, while small differences between the control and stain were detectable for the 1st and 2nd derivative for most wavelengths; realistically, this difference at wavelengths greater than 1650 nm may have been somewhat attributable to the high sample size (Table 6.2 and Figure 6.2b-c). Also, the success of detecting such small differences was aided by the controlled growth of *L. serpens* and *O. minus* species in a laboratory environment. The procedure used in the lab lasted 6 weeks under optimal humidity and temperature conditions. Such duration and environment allowed for complete inoculation of two fungi species. The uniform fungi growth resulted in maximal color change with less contrast in stain from ring to ring. When these test cores were visually compared to naturally inoculated cores, I found it obvious that these cores had increased stain concentration which should heighten differences in spectra.

When the variation in clear wood absorbance was compared to the variation in stain absorbance, the variation was found equal at wavelengths greater than 1354 nm. Equal variation is often an important assumption during model building. Additionally, equal variation in absorbance between stain and control wood suggests that absorbance variation was dominated by the wood substrate and not the blue stain. Likewise, when the wood variation was subtracted from the spectra leaving only the absorbance attributable to the in-situ stain, the region above 1354 nm was relatively insensitive to blue stain (Figure 6.2a-c).

Table 6.2: The influence of *O. minus* (O), *L. serpens* (L), and control (C) treatment on absorbance between the wavelengths of 350-2500 nm. Significance differences was determined at the$\alpha = 0.05$ level (p-value < 0.05).

| Treatment to Spectra | Comparison | Significant wavelengths | Sample size |
|---|---|---|---|
| Original | L vs. C | 354-364, 424-1354, 464-1354 | L = 164, C = 304 |
| | O vs. C | 464-1334 | O = 140, C = 304 |
| | L vs. O | 424-504 | L = 164, O = 140 |
| 1st derivative | L and O vs. C | 354-384, 414-474, 554-734, 854-1014, 1064-1204, 1234-1244, 1294-1624, 1694-1714, 1854-2004, 2044-2074, 2124-2194, 2224-2264, 2294-2304 | L and O = 304, C = 304 |
| | L vs. O | 484-554 | L = 164, O = 140 |
| 2nd derivative | L and O vs. C | 374-754, 684-764, 834-1364, 1414-1514, 1924-1964, 2294-2304, 2454-2464 | L and O = 304, C = 304 |
| | L vs. O | 444-504 | L = 164, O = 140 |

The confidence intervals around the mean absorbance became unstable with the application of the second derivative for the stain and control treatments. Differences by an order of magnitude often occurred when the confidence intervals at the same wavelength were compared. Hiukka (1998) noted that the second derivative was very sensitive to noise, just as in this study. It is thus recommended that the noise be addressed when both stained and unstained woods are to be monitored by NIR signal using the 2nd derivative as a pretreatment. The increase in noise to signal was not a

problem when the 1st derivative was used. Furthermore, the noise to signal for stained wood was similar to unstained wood when the 1st derivative was used.

6.4.2 Differences in Absorbance Between *L. serpens* versus *O. minus*

The detection of *L. serpens* versus *O. minus* might be possible if absorbance at a given wavelengths reacts differently to the two stain treatments. To achieve differentiation, the power of the paired experimental design was utilized. While a shift in the 1st derivative occurred, when compared to the original and 2nd derivative spectra, the complete range of wavelengths that differentiated *L. serpens* and *O. minus* was between 424-554 nm. While this fell within the range of spectra influenced by blue stain in Figure 6.1a (424-1354nm), the region of 424-554 nm was a direct influence of the stain whilst the remainder of the range (564-1354 nm) could have been confounded by some stain to wood interaction.

With the exception of 424-554 nm, the mean absorbance across all wavelengths had an equal slope for *O. minus* and *L. serpens* spectra (Figure 6.2a). The parallelism between absorbance suggests that some degree of similarity in chemical structure exist between the two species. While not plotted, the variation in absorbance at each wavelength was considerable with *O. minus* and *L. serpens* overlapping in standard deviation below 424 nm and above 554 nm.

6.4.3 Within and Between Core Variation

When I visually observed the heartwood of most cores, the heartwood appeared resistant to stain. Observable stain usually began to occur at rings 9 through 40. However, when the data was partitioned by sapwood and heartwood, blue stain was detectable in the heartwood region despite the lack of stain as observed by eye. As a

result, NIR may be a practical method to detect stain otherwise unobservable. Such rapid detection is needed by tree physiologists to determine the ratio of trees infected by fungi.

The range of stain variation between increment cores was large with some increment cores being only slightly stained while most were heavily stained. The occurrence of some lightly stained cores suggests that some trees had increased resistance to stain. The lightly blue stained cores usually possessed lower absorbance in the 564-1354 nm range than the heavily stained cores which was probably attributable to decreased Nitrogen content. Qualitative assessment suggested this to be a possibility and merits attention in future work.

NIR appeared to be a promising tool to identify when stain is present in wood. Future research is needed to determine if NIR can distinguish between blue stained wood from control wood for different populations and species. For those wishing to monitor other traits like stiffness, density, lignin, or extractives, then blue stain associated wavelengths should be avoided during the calibration process in hopes of minimizing error attributable to stain.

# CHAPTER 7

# PREDCITION OF WOOD MECHANICAL AND CHEMICAL PROPERTIES IN THE PRESENCE AND ABSENCE OF BLUE STAIN AND USING TWO NIR MACHINES

## 7.1 Introduction

There is substantial interest in the improvement of wood properties through genetic selection or a change in silviculture prescription. However, the ability to measure many traits is traditionally difficult and time consuming. As much as 1000 to 2000 trees may be measured in a single study making NIR an attractive measuring tool for many wood properties, particularly if it is accurate and rapid (So et al. 2002).

Strength and stiffness (MOR and MOE) are two important parameters to improve in solid wood products and composites while density, lignin, and extractives variation influence paper yield and strength (Schultz and Burns 1990; Sanderson et al. 1996; Shupe et al. 1999; Schimleck et al. 2001; Schimleck et al. 2002; Bailleres et al. 2002; Gierlinger et al. 2002; Schimleck and Evans 2003). Increased extractives and lignin can lower pulp yield while reduced extractives improves the bonding capacity between fibers (Schimleck and French 2002) Density improvement in pulpwood is a bit challenging since an increase in density increases pulp yield but decreases paper strength (Via et al. 2004). For solid wood, increased wood density is positively correlated with MOE and MOR making density improvement an important goal. In New Zealand, wood cants have been scanned and classified for MOE with fair success (Meder et al. 2003). The ability of NIR to measure MOE and MOR is attributable to the absorbance of light by lignin and cellulose at specific wavelengths and nondestructive measurements are quite successful

on small clear wood (McLellan et al. 1991; Gindl et al. 2001; Rials et al. 2002; Kelly et al. 2004ab).

The ability to measure lignin, cellulose, and extractives can be improved by grinding the wood sample to some mesh size with lower analytical error than wet chemistry (Shultz and Burns 1990; McLellan et al. 1991). However, grinding is timely and destructive; as a result, many breeders are interested in using NIR to measure solid wood samples.

The measurement of these traits with NIR can be biased if some external variation influences the spectra. Blue stain is one potential source of variation in wood and commonly occurs in lumber, chips, and trees. Blue stain is thus a potential source of bias when measuring mechanical and chemical properties with NIR. The two most prominent staining species are *Leptographium* and *Ophiostoma* (Ross et al. 1992; Kreber and Morrell 1993; Uzunovic and Webber 1998; Eckhardt et al. 2004) Stain deposits can be detected at wavelengths between 190 to 800 nm (Lebow et al. 1996; Zulpa et al. 2003) However, it is unknown how the influence of stain will affect the predictability of NIR at wavelengths greater than 800 nm. If the stain associated wavelengths are known, one may remove those wavelengths from calibration modeling (Swierenga et al. 2000; Thygesen and Lundqvist 2000; Wulfert et al. 2000). If the color change causes a shift in spectra (Gillon et al. 1999), then pretreatments such as derivatives or multiplicative scatter correction may be needed. Multiplicative scatter correction is a transformation method used to compensate for additive or interaction effects in the spectra.

The stain can also increase the nitrogen content in wood since the residual pigment has high concentrations of nitrogen (Zink and Fengel 1988). Therefore, the absorbance of

light by additional nitrogen may have consequences in the prediction of mechanical and chemical properties. Modeling may particularly become difficult if many factors are needed or interactions between wavelength absorbance occur.

My objective in this research was to investigate the influence of blue stain on the predictability of MOE, MOR, density, lignin, and extractives for a) master NIR machine and b) slave NIR machine with master calibrations. I chose principal component (PCR) and multiple linear regression (MLR) procedures for interpretation and usually have similar predictive ability to partial least squares regression (Estienne et al. 2001; Smith et al. 1998; Smith et al. 2001; Wentzell and Montoto 2003). Also, multiple linear regression was attractive because it may avoid blue stain associated variation that would be captured in principal components regression when all wavelengths were included. Multiplicative scatter correction was attempted but did not adequately remove blue stain variation. The 1$^{st}$ derivative pretreatment was applied when models between blue stain and control treatments exhibited different means and variance.

## 7.2 Methods and Materials

### 7.2.1 Sample Preparation

For this study, I did not consider the pith wood in model building or prediction since NIR has shown poor performance in predicting MOE and MOR for pith wood (Via et al. 2003). Ten longleaf pine (*Pinus palustris*) trees were harvested at a stand age of 41 years. The site was managed at the Harrison Experimental Forest by the United States Forest Service in Saucier, Mississippi (USA) at a location 30.6º north and 89.1º west. Periodic fires were prescribed to limit the density of understory between neighboring trees. Trees were positioned 3.66 m apart in an equilateral triangle position. Each tree

was bucked every 4.57 m in height yielding five to seven bolts and cookie disks. The bolts were used to make bending specimens 30.48 cm x random width x 1.27 cm. Spectra were acquired from the radial face from pith to bark from strips ripped from the disk. The strips were approximately 2 mm thick x random length.

I used two populations of increment cores for testing the effect of stain on spectra response. For test population 1, twenty increment cores were selected from a library of 1800 increment cores. These increment cores were distinctive because one side had blue stain while the other side was clear of pigmentation. The matched samples samples were cut into half and represented the control and blue stain treatment. This resulted in a sample size of 101 for the control and 101 for the blue stained side.

For population 2, a validation population, 39 increment cores were artificially inoculated in the laboratory with *Leptographium serpens* and *Ophiostoma minus*. A total of 300 sample points were available for the control and stained treatment. The *Leptographium serpens* was obtained from longleaf pine roots and southern pine beetles from the Palustris Experimental Forest which is owned and maintained by the USDA Forest Service, Louisiana, U.S.A. Likewise, the *Ophiostoma minus* was sampled from Bankhead National Forests, Alabama, U.S.A. The isolates were cultured on malt extract agar (MEA) (2% malt extract agar), and representative stock cultures were maintained on MEA slants at 4ºC.

Each isolate was inoculated onto thirty MEA plates and allowed to grow for seven days at 22ºC ± 3 (ambient room temperature). An inoculation slurry was created for each isolate. The thirty plates of inoculum were shredded in a blender with 500 mL $dH_2O$ for 60 seconds. The cores were then briefly submerged in the inoculum and placed in

separately labeled but unsealed ziplock bags. At this stage, I controlled the humidity and temperature at 95% ± 2 and 22ºC ± 3 as established by (Wang 1994; Viitanen 1997; Jacobs et al. 2000) respectively. The moisture chamber was a dessicator containing 1000 mL of tap water below a ceramic plate stirred by a magnetic stir bar. Two dessicator setups were used, one for each fungal treatment. After a week of growth, I scrubbed each increment core with sterile tissue-wipes to remove excess inoculum and then placed back into the moisture chamber. Sterile rubber gloves were used to handle the stained specimens and core cleaning was completed in a laminar flow hood under sterile conditions. Fungi growth was allowed to occur for a six week duration and then the specimens were taken out of the plastic bags and allowed to air dry. The surface of the specimens dried within a day, as monitored by touch, ensuring that the surface was below fiber saturation point as quickly as possible so as to stop surface growth of any contaminating spores. The entire increment core took less than a week to dry based on weight scale measurements. The control half of the sample was air dried for two months at ambient room temperature and humidity. The sample reached equilibrium with the environment when gravimetric weight became constant.

7.2.2 Mechanical Testing

All specimens were conditioned to an equilibrium moisture content of 8% with a standard deviation of 1.3%. Loading was applied on the tangential face in a three-point bending setup at a rate of 0.20 cm $min^{-1}$. The linear slope of stress and strain plots was used to determine MOE. MOR, moisture content, and air dry density were determined by standard test methods (ASTM 1993). Density was taken as the mass per volume of wood.

Volumes were measured with calipers to the nearest 0.0025 cm. Weights were measured to the nearest 0.001 g on a digital balance.

7.2.3 Chemistry Determination

I ground the disks in a Wiley mill at latewood rings 4, 8, 16, 32 from the pith and screened between 40 to 60 mesh particle size. I analyzed all samples for alcohol-toluene extractives and Klason lignin by standard test methods (Tappi 1997) The last ring was also processed and represented the 41$^{st}$ year from the time of planting. Samples were sealed in a plastic bag to minimize the variation in moisture content between tests. A replicate of 3 was used and the mean reported for each extractive and lignin measurement. After analysis, samples that exhibited strong leverage in regression equations were not determined to be outliers unless supportive evidence from notes supported elimination.

7.2.4 NIR Spectroscopy

I used a Nexus 670 FT-IR spectrometer (Thermo Nicolet Instruments, Madison, WI, USA) as a master machine to acquire absorbance between 1000 and 2500 nm. Likewise, I used a slave machine to acquire data between 1000 and 2500 nm for the experimental samples (ASD Field Spec – Analytical Spectral Devices Spectrometer – Boulder, CO, USA). Scans were acquired at 1 nm intervals using reflectance spectroscopy. A single spectrum was acquired as the average of 40 scans. The north and south spectra were averaged together to obtain one spectrum per sample. A total of 250 samples were available for model building and cross validation by a statistical software package (SAS 1999). The absorbance spectra were transferred from the spectrometer to an intermediate software package and then to SAS. The intermediate software package (Unscrambler

1999) was used to compute first derivatives before exporting the data into SAS and was used to average the spectrum into 10 nm intervals. Therefore, any wavelength reported from this study was accurate to ±5 nm. The reduction to 10 nm intervals has no detrimental influence on model precision (Schimleck et al. 2004).

7.2.5 Multivariate Analysis

I used MLR and PCR procedures in SAS using the PROC REG and PROC PRINCOMP for regression and principal components determination, respectively. Statistical diagnostic tools were applied to determine the predictive power and number of variables needed for each model. A standard cross validation (leave-one-out) was performed in UnScrambler to determine the root mean square error of calibration and prediction (RMSEC and RMSEP). The RMSEC from the UnScrambler software was compared to SAS RMSEC to ensure that the same model was computed by both software packages. The $R^2$ and adjusted $R^2$ were used to determine both the amount of variance accounted for by the model and if an overfit might have occurred by inclusion of too many variables. Additionally to avoid overfit, the Mallows $C_p$ was computed as an index to determine the maximum predictive power with minimal factors (Neter et al. 1996).

$$C_p = \frac{SSE_p}{MSE(X_{1.....}, X_{p-1})} - (n - 2p) \qquad (7.1)$$

where SSE is the sum of square error for the fitted subset with p parameters, MSE is the mean square error of the model, and X is the independent variable (wavelengths or principal components). The $C_p$ procedure is a method to chose a submodel with the lowest possible number of factors (p) and where $C_p$ is less than p. Wavelengths for MLR were predetermined by the influential regression coefficients as deemed important by principal components regression analysis. For PCR, all wavelengths were reduced to

principal components and then regressed by multiple linear regression in SAS. MLR was chosen to see if individual wavelengths might be used to avoid the influence of stain. All independent variables were left in the model at p-values less than 0.15. The first derivative was applied in situations were PCR and MLR did not show identical distributions and means for a trait. The area under the curve was not used to determine density since blue stain variation may confound the raw spectra due to a baseline shift that might occur with both density and blue stain variation (Gillon et al. 1999; Via et al. 2003).

## 7.3 Results and Discussion

### 7.3.1 Blue Stain Classification

Figure 7.1 demonstrates the potential to classify stain and unstained wood. The segregation of blue stain and control samples occurred along the axis of the 2$^{nd}$ principal component (PC), which accounted for 9% of the total variation. There was no clustering along the 1$^{st}$ PC axis and the variation in both groups was equal while the range was slightly larger for the non blue stained group. Eighty-eight percent of the blue stained specimens fell above zero for the 2$^{nd}$ PC while 92% of the control specimens fell below zero. A third PC (3$^{rd}$ dimension) was plotted using Unscrambler software and no significant segregation occurred in that dimension. The 3$^{rd}$ PC accounted for 1% of the overall variation and thus 99% of the total variation in spectra was accounted for by the first three principal components.

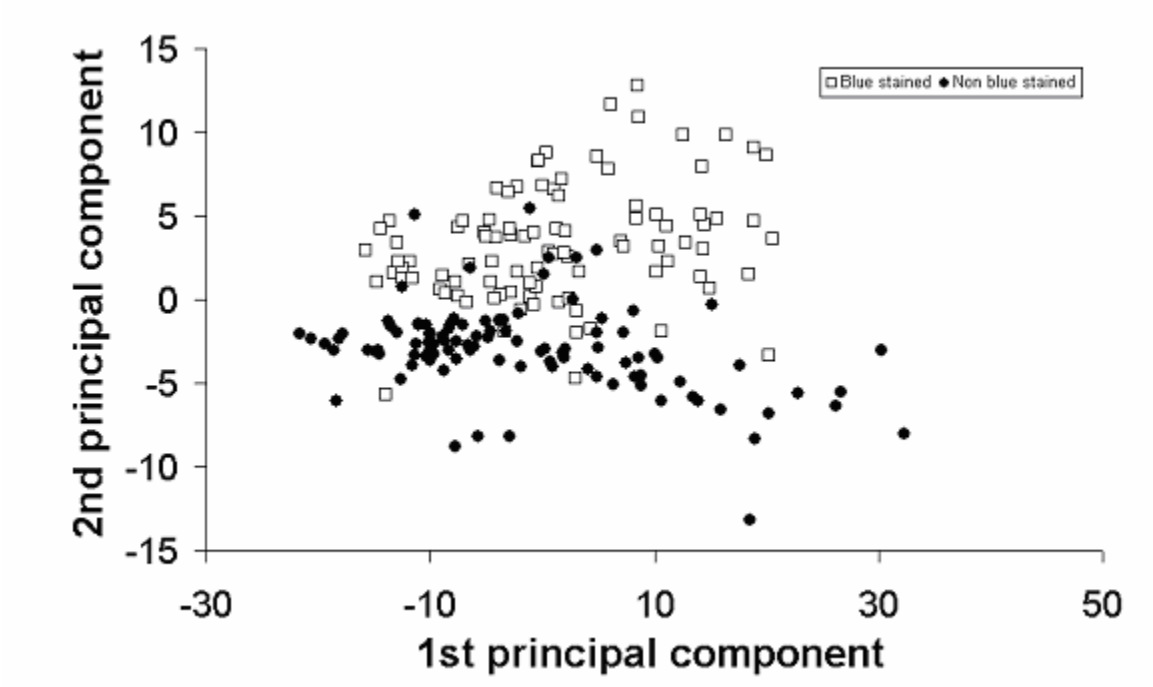



Figure 7.1: The segregation of blue stain and control samples when the 1st and 2nd principal components were plotted. Principal components were developed from the raw spectra between 1000 to 2500 nm. The 1st and 2nd principal component accounted for 89% and 9% of the overall variance respectively.

The 2nd PC accounted for the blue stain variation in both the raw spectra and after the 1st derivative was applied. Applying the 2nd derivative to the raw spectra removed any segregation along the 2nd PC. Unless the 2nd derivative is applied, removal of the 2nd PC from modeling was thus a viable consideration; however, the 2nd PC was important in the modeling of all 5 traits. As a result, removal of this PC was thus not an option for this data set making blue stain an embedded bias when modeling with PCR.

The cores were taken during heightened bark beetle attacks which transports the staining fungus. Stain colonization within the tree can suppress moisture flow and cause disease or stress during droughts. Any method to rapidly identify the occurrence of stain in trees might be useful for forest managers, plant pathologists, or tree geneticists.

### 7.3.2 Model Building

Table 7.1 (master NIR machine category) demonstrates the predictive ability for five traits when the master machine was used to both build the calibration equation and to scan the experimental samples. I found that blue stain did not have a significant influence on the prediction of the five traits when the clear wood and stained wood were compared as shown by the non significant sign in (Table 7.1). As such, I concluded that one can build models to avoid the influence of stain, by picking wavelengths not sensitive to stain. Such a possibility was useful since stain may add additional noise to the spectra which might bias predictions if not accounted for during calibration development.

The remainder of the paper will focus on the response of a slave machine when using calibrations built from the master machine. It should be noted that one should never assume that transfer of a calibration equation is appropriate. In fact, the results of this paper show how critical such a practice may be. As such, this research demonstrates the complexities that occur when this assumption is made. As expected, more problems occurred for this scenario.

Table 7.1 demonstrates the predictive power of the slave machine for when calibration models were built from the masters machine. Using Mallow's $C_p$ diagnostic, I determined that lignin prediction typically required less factors or wavelengths than MOE, MOR and extractives. The number of variables needed to model density was less certain although two wavelengths proved almost as successful as 5 PC's but with higher RMSEC and RMSEP (Table 7.1). The more factors needed to predict a trait suggests increased complexity within the independent matrix or between wavelengths in MLR. A higher number of factors (or independent variables) was expected for MOE and MOR

since density, MFA, cellulose, and lignin can all influence these two traits (Gindl et al. 2001; Meder et al. 2003; Kelly et al. 2004). The higher number of factors needed to predict extractives supports a wider variation in chemical compounds in extractives relative to lignin.

Table 7.1. PCR and MLR calibration and validation correlations for lignin, extractives, modulus of elasticity, modulus of rupture, and density.

| Traits | Model | Ind. variables | N= | $R^2$ | Adj. $R^2$ | RMSEC | RMSEV | $C_p$ |
|---|---|---|---|---|---|---|---|---|
| Density | PCR | 5 PC's | 250 | 0.71 | 0.69 | 0.0442 | 0.0499 | 3.4 |
| | MLR | 1495, 1885 nm | 250 | 0.72 | 0.69 | 0.054 | 0.054 | 5.1 |
| MOE | PCR | 6 PC's | 250 | 0.86 | 0.85 | 2048 | 2392 | 12.1 |
| | PCR[a] | 4 PC's | 250 | 0.87 | 0.86 | 3043 | 3274 | 5.0 |
| | MLR | 1395, 1695, 1925, 1995 nm | 250 | 0.80 | 0.78 | 2797 | 2861 | 3.6 |
| MOR | PCR | 6 PC's | 250 | 0.87 | 0.86 | 16.4 | 19.0 | 7.6 |
| | PCR[a] | 5 PC's | 250 | 0.88 | 0.87 | 19.2 | 20.2 | 5.4 |
| | MLR | 1395, 1695, 1925, 1995 nm | 250 | 0.83 | 0.80 | 21.2 | 21.7 | 3.7 |
| Extractives | PCR | 4 PC's | 107 | 0.72 | 0.70 | 3.06 | 3.48 | 9.7 |
| | PCR[a] | 5 PC's | 107 | 0.64 | 0.61 | 3.30 | 3.04 | 8.0 |
| | MLR | 1405, 1685, 2185 nm | 107 | 0.69 | 0.68 | 3.11 | 3.27 | 7.8 |
| Lignin | PCR | 4 PC's | 80 | 0.55 | 0.50 | 1.83 | 1.99 | 3.8 |
| | MLR | 1935, 2265 nm | 80 | 0.44 | 0.40 | 2.01 | 2.07 | 9.5 |

[a]A first derivative pretreatment was applied to the spectra

To determine the optimal number of factors, the $C_p$ should be equal to or less than the number of factors (p) or ($C_p - p \leq 0$) (Neter et al. 1996). Table 7.1 demonstrates this did not always occur. If $C_p$ was greater than the number of factors (p) for the best submodels, then the probability of overfit was increased and is an indication that wavelengths other

than what was needed to predict y was most likely entrenched in the independent matrix. During the model building stage only the important wavelengths should be included to improve the reliability of $C_p$. Of course, this was not the case for PCR since all wavelengths were used to build PC scores. As a result, having a $C_p$ less than p may sometimes be a severe requirement for PCR regression if all wavelengths are used. For this study, the $C_p$ of submodels were fairly stable when applied to the test population, even when $C_p$ was slightly greater than p. Specifically, gross differences approximately greater than or equal to 5 ($C_p - p \leq 5$) were usually the limit. When differences exceeded 5, the model predictability diminished in almost all cases when applied to the test population. The exception to this rule was lignin where the $C_p = 9.5$ while $p = 2$. Better models for lignin with $C_p \approx 2$ and with p=2 were possible with the inclusion of wavelength 1015 nm, but this wavelength was heavily influenced by blue stain and thus was removed. Likewise, removal of the 1015 nm wavelength resulted in lower $R^2$ values for MLR (Table 7.1).

When optimal wavelengths for MLR were predetermined by PCR regression coefficients, the $C_p$ dropped considerably below or equal to p suggesting that MLR may have provided better models than PCR. Conversely, PCR is more stable when slight extrapolations occur in the test population (Estienne et al. 2001). For our calibration samples, 10 trees were used to build these models since the harvest of more trees from the same stand were not possible. As expected, I found that within tree variation was more than between tree variation for all mechanical and chemical properties. Such a criteria was important since calibration ranges need to exceed those experienced in the prediction population (Schultz and Burns 1990). Still, it was quite possible that some

combination of factors or wavelengths for a single tree were outliers when compared to the overall population of 1800 trees. As will be seen later, PCR sometimes worked quite better than MLR perhaps demonstrating the ability of PCR to model under slight extrapolation conditions. As a result, sampling of more trees is encouraged when feasible but was not detrimental to this study.

All traits in Table 7.1 exhibited $R^2 > 0.7$ except for lignin where the $R^2$ ranged between 0.44 to 0.55 for MLR and PCR regression respectively. When partial least squares (PLS) regression was considered, similar results were obtained to Kelly et al. (2004) for both lignin and extractives. Better correlations between solid wood lignin and NIR spectra have been achieved with transmission NIR (Yeh et al. 2004). In some situations, PLS regression yields higher $R^2$ values and lower RMSEC-RMSEP than MLR and PCR (Wentzell and Montoto 2003). However, MLR and PCR were preferred for this study since interpretation of underlying data was easier.

It should be remembered that I built these models using solid wood samples where chemical and anatomical properties can change systematically within a ring. Much higher $R^2$ values are sometimes possible with lower RMSEC-RMSEP when samples are ground before measuring the absorbance with NIR (McLellan et al. 1991; Sanderson et al. 1996; Bailleres et al. 2002; Poke et al. 2004).Grinding the samples probably averages out any systematic variation present in solid wood. For extractives, one study found the $R^2$ to improve from 0.75 to 0.95 after grinding solid wood samples (Gierlinger et al. 2002). Additionally, removing extractives before scanning can improve the predictive ability of lignin since lignin and extractives can share analogous absorption bands and thus may help to explain our low $R^2$ values (Bailleres et al. 2002). Nevertheless, solid

unextracted wood was investigated here since that is how most laboratories would prefer to scan the wood material.

7.3.3 Influence of stain on test population for the slave machine scenario

I found the prediction of density in stained specimens was better using PCR than MLR (Figure 7.2(a) and 7.2(b)). The variation of the blue stained wood appeared somewhat larger than density (Figure 7.2(a)). However, the F-test statistic showed both variances to be equal. When MLR was used (Figure 7.2(b), the median was about the same for both stain and control wood. On the other hand, the mean was higher for the control due to the difference in variation of distributions.

It should be understood that the NIR spot size was usually greater than the width of the ring of interest. Between rings 30 through 40, sometimes as much as 5 rings fell within the 5 mm spot size for NIR scanning. To minimize this error, the probe was placed in the center of each ring with earlywood overlap on either side. As such, this explains the wide variation in density. Had the spot size been smaller than the latewood ring width for all rings then a very narrow distribution of densities would have been expected since pine earlywood and latewood density differ greatly while the variation in density within earlywood or latewood is very narrow (Bendtsen 1978).

MOE exhibited a difference in means regardless of the model used (Figure 7.3). Although MOR was not shown, it exhibited a similar response to MOE for all models. Even the 1$^{st}$ and 2$^{nd}$ derivative did not completely eliminate the influence of blue stain on the MOE and MOR models. For MOE, the blue stained wood was overpredicted by 13,000 MPa (Figure 7.3). When the first derivative was applied and PCR was determined, the mean MOE was under predicted by 10,000 MPa for the blue stained

wood. Likewise, when predicting MOR, differences in means between 6,000 to 17,000 MPa occurred depending on which model was chosen. However, when a PCR model was used, the 1st derivative pretreatment did yield very similar distributions with identical variations for both MOE and MOR (Figure 7.3(b)). The equivalent shape in distribution was an improvement over no pretreatment where both the mean and variance differed between the stain and control wood.

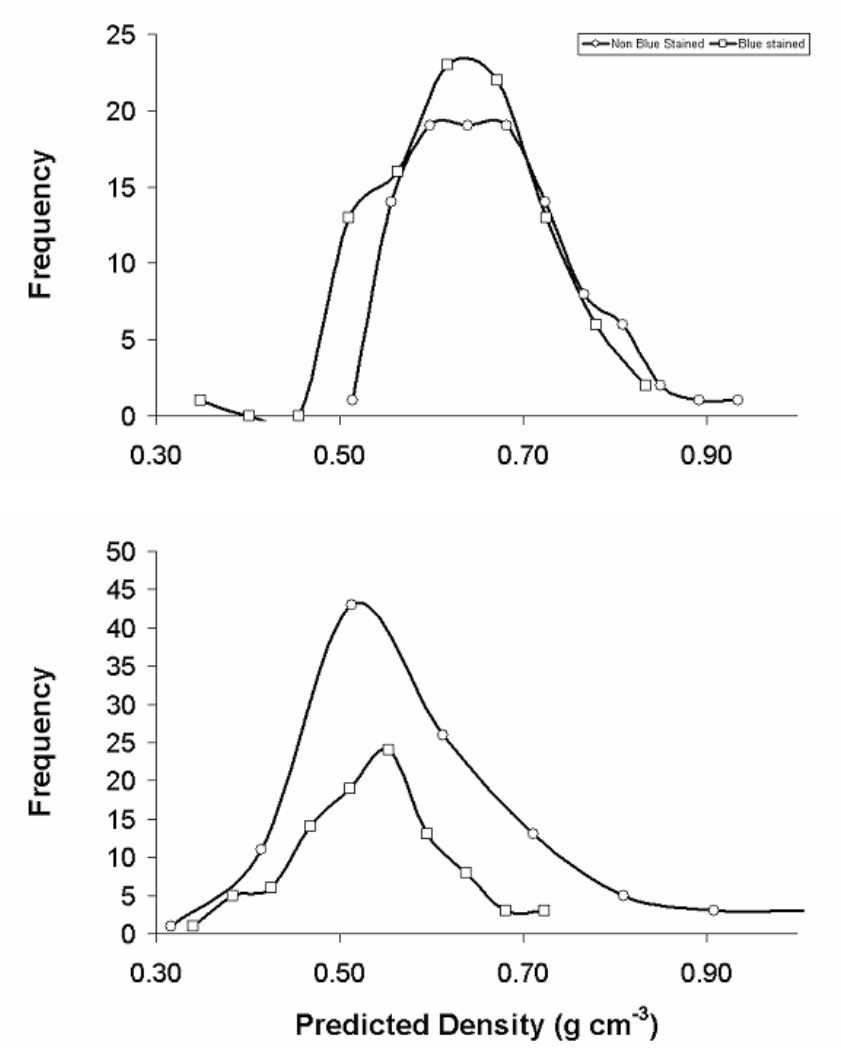



Figure 7.2: The prediction of density for stained (n=101) and unstained (n=101) wood using (a) PCR and (b) MLR.

As Figure 7.3 demonstrates, there was no way to avoid the error introduced by blue stain when predicting MOE (and also MOR). In fact, MLR yielded worst results for most submodels when predicting these mechanical properties. These poor results were unexpected by MLR because I hypothesized that a selection of specific wavelengths would hopefully avoid those wavelengths also influenced by stain. For all PCR models, a minimum of 5 factors was recommended by the $C_p$ diagnostic. The increase in the number of factors with respect to lignin and density (Table 7.1) was one complexity that made it difficult to eliminate blue stain influence. I found that the chances of blue stain influence on the model increased as the number of factors increased. In other words, as the number of factors increased, the chance of finding a model not influenced by blue stain became less likely.

One solution to minimize the MOE and MOR error attributable to stain would be to include the natural variation of stain within the calibration model. A second option would be to first classify the wood for stain (Figure 7.1), and then adjust the mean MOE or MOR if blue stain occurred. For example, in Figure 7.3b, the distribution shape between the stain and control wood was identical but with a differing mean. If the sample fell in the blue stain zone (Figure 7.1) then one could adjust the predicted mechanical property for that sample. Such an adjustment may be needed in a manufacturing environment when an extraneous source of variation is present. However, classification of the stain prior to modeling would be needed which may prove unfeasible in an industrial setting.

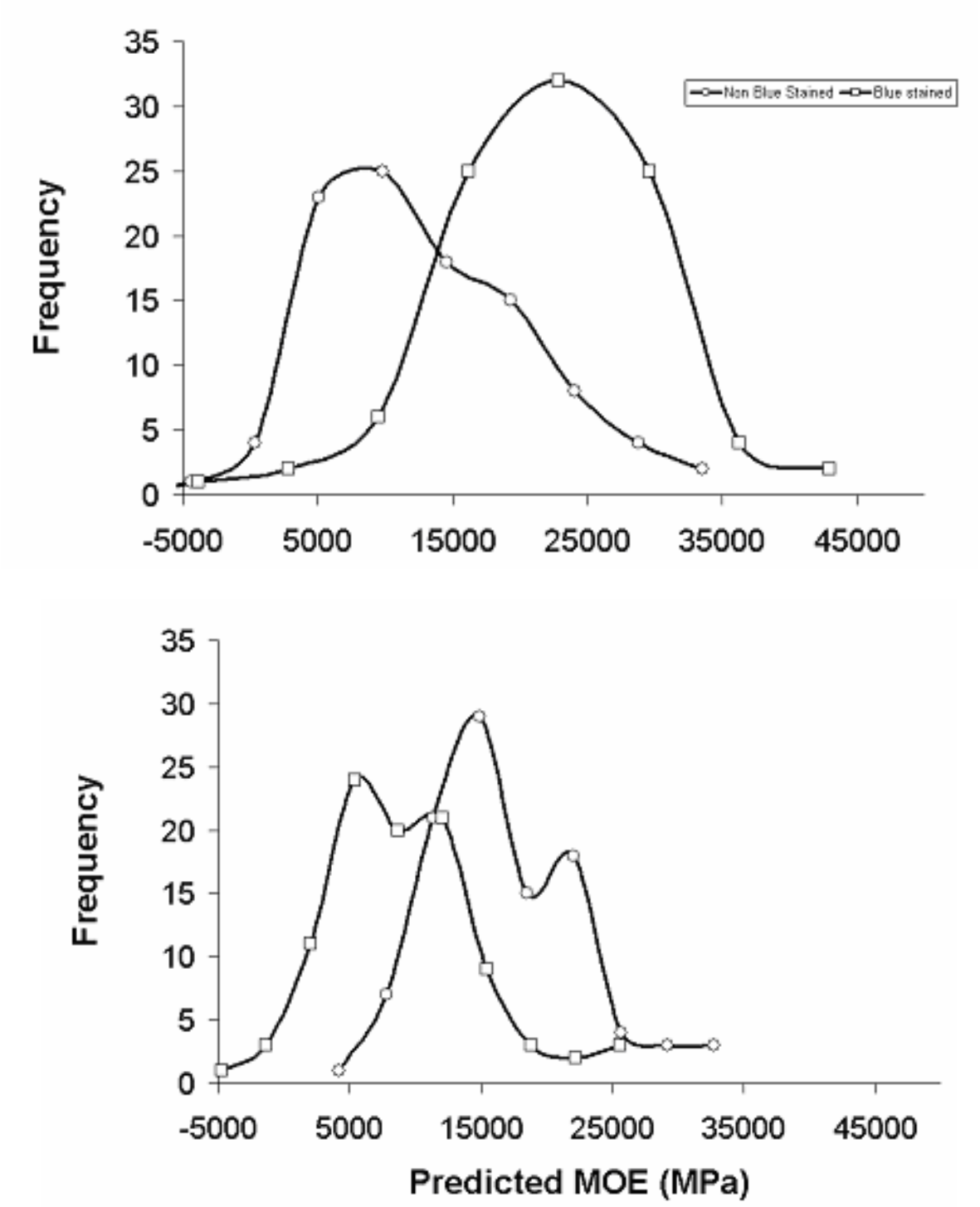



Figure 7.3: The prediction of MOE for stained (n=101) and unstained (n=101) wood using (a) PCR and (b) PCR from 1st derivative spectra.

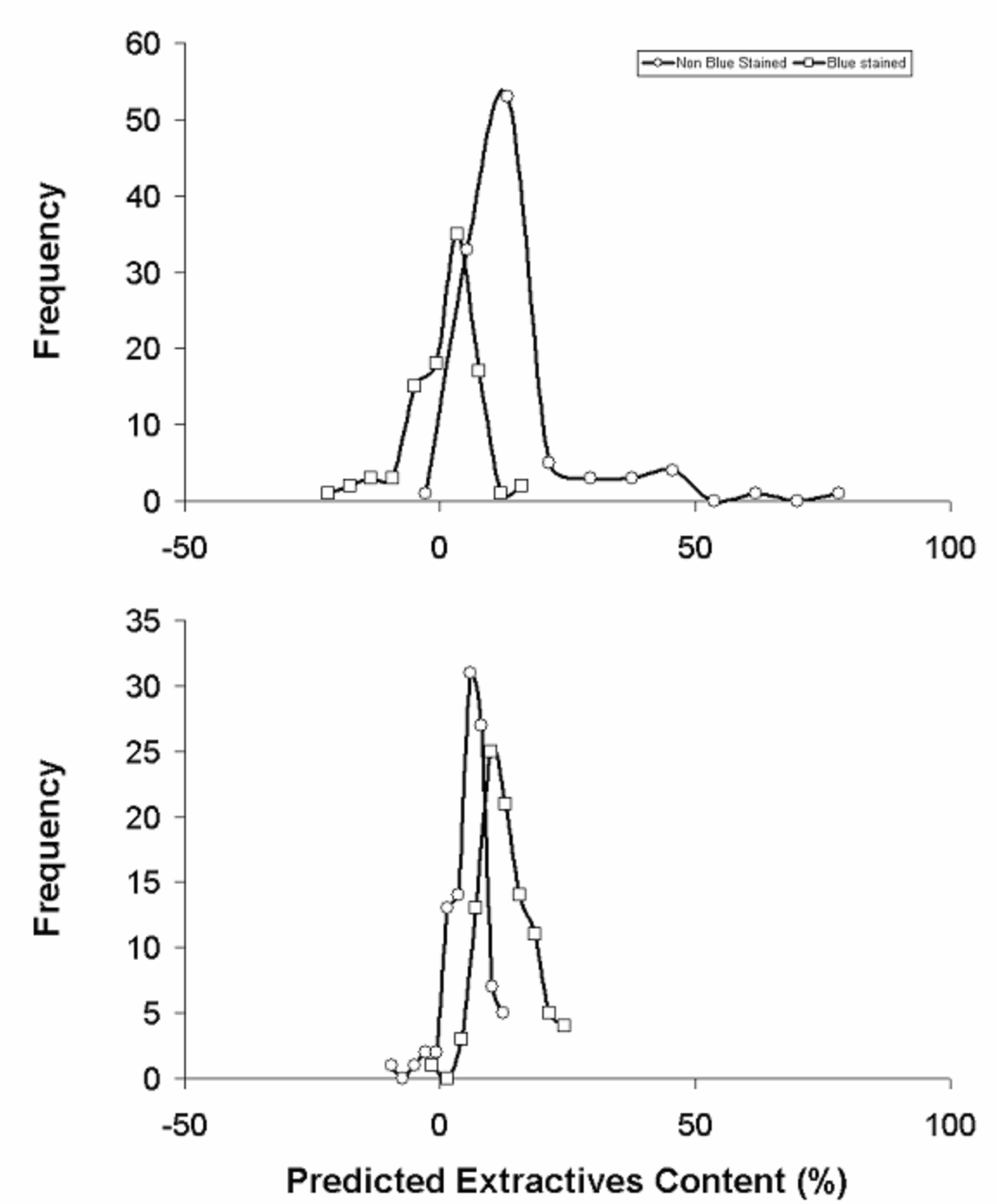



Figure 7.4: The prediction of extractives for stained (n=101) and unstained (n=101) wood using (a) MLR and (b) PCR from 1st derivative spectra.

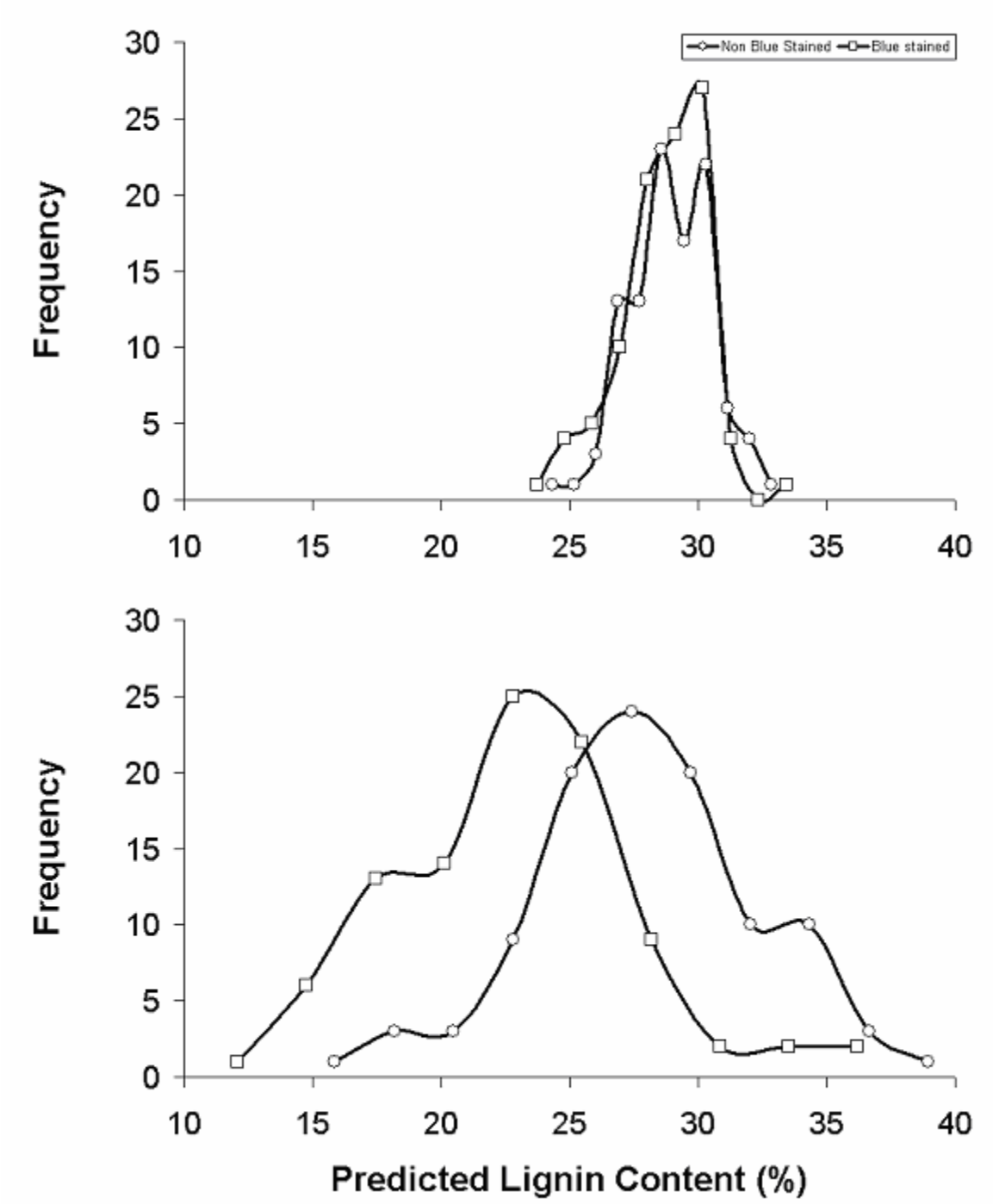



Figure 7.5. The prediction of lignin for stained (n=101) and unstained (n=101) wood using (a) MLR and (b) PCR.

For extractives content, the MLR predicted values fell within the range of the calibration data for the control treatment. Regrettably, MLR did not accurately predict extractives for the blue stained treatment with over 30% of the values being less than zero

% (Figure 7.4(a)). For the original spectra, PCR faired no better and a first derivative was thus applied to the original spectra (Figure 7.4(b)). Using PCR on the first derivative still had bias, but like the mechanical properties, exhibited a similar histogram shape between the two treatments. Also, the extractives content was similar to the mechanical properties in that 5+ factors were commonly recommended for prediction, as determined by the $C_p$ diagnostic. As discussed earlier, such an increase in factors made it more difficult to avoid the influence of stain.

For lignin content, I found MLR worked nicely and avoided the influence of stain with a similar mean and distribution (Figure 7.5(a)). In this case, 2 to 4 independent variables were recommended by $C_p$ making it more likely to find a group of factors uninfluenced by the stain. Unlike other traits, PCR performed poorly with an underprediction in the blue stain treatment by 5 percentage points (Figure 7.5(b)). Also, the variation of the predicted data fell outside the range of the calibration data, a sure sign of overfit. In Table 7.1. as Figure 7.5b shows, the MLR model behaved better under blue stain conditions and is a practical example of how building a representative model can be more important than the resulting $R^2$. For lignin it was quite apparent that the precision of the NIR model was more accurate than the wet lab chemistry used to develop the model as seen in other work (McLellan et al. 1991ab).

### 7.3.4 Verification of Test Population for the Slave Machine Scenario

The blue stained samples I discussed so far were not control inoculated in a lab meaning that the origin and type of stain was unknown. As such, variation in stain intensity was great suggesting that absorbance response could vary considerably within a blue stain treatment. Another population, where the fungi growth was controlled, was

tested to understand the mean change in absorbance that would occur for a minimal level of blue stain variation.

In Table 7.2, the results were strikingly similar between this experiment and the 1st experiment with all five traits yielding similar means, variance patterns, and histogram shapes. As expected, population 2 (laboratory inoculated) sometimes exhibited slightly lower variation in predictive response due to the fairly uniform stain. Nevertheless, the variation in concentration of stain in the field may have differed from ring to ring, tree to tree, and the species of stain was unknown. The naturally inoculated increment cores were thus of important interest. However, one could also use the laboratory treated cores since both stain and control populations gave equal results. The mean difference between the stain and control for density, MOE, MOR, extractives, and lignin were statistically equal. One could thus develop models and adjust the predicted mean for those instances where the stained histogram shifts distinctively but maintains variance and histogram shape.

Table 7.2. A comparison of variance and mean difference in 5 traits for stain and control treatment within test population 1 and test population 2 using the same MLR model. The slave machine group was when a calibration from the master was used on spectra acquired from the slave machine while the master machine group was when only the master machine was used for calibration and scanning. Significant difference ($\alpha = 0.05$) denoted by * while non significant difference denoted by n.s. The symbol >, <, = denotes significantly greater, less, or no significant difference.

| Trait | (Stain – Control) difference between means | | | Variance | |
|---|---|---|---|---|---|
| | Pop. 1 | | Pop. 2 | Pop. 1 | Pop. 2 |
| | Slave | Master | Slave | Slave | Slave |
| Density ($g/cm^3$) | -0.03 n.s. | + 0.01 n.s. | -0.05 n.s. | Stain = Control | Stain = Control |
| MOE (MPa) | + 7166 * | +173 n.s. | + 6501 * | Stain = Control | Stain = Control |
| MOR (MPa) | +5909 * | - 0.1 n.s. | +5500 * | Stain > Control | Stain > Control |
| Extractives (%) | - 11.3 * | +0.78 n.s. | -9.2 * | Stain < Control | Stain < Control |
| Lignin | + 0.1 n.s. | + 0.19 n.s. | + 0.7 n.s | Stain = Control | Stain = Control |

## 7.4 Conclusions

My most important finding was that the introduction of blue stain did not influence prediction models for all five traits when both scanning and calibration was performed on the same NIR machine. Such a finding suggests that one can pick wavelengths not associated with the stain during model building.

The introduction of blue stain had mixed results on NIR models for five separate traits when a slave machine was used for scanning while calibration equations were built from the master machine. With the exception of lignin, PCR gave better models when applied to the test population indicating that PCR does a better predictive job under slight extrapolation conditions, especially as the number of factors increased. For the slave machine scenario, there were three sources of known extrapolation (a) 10 trees were used to represent a population of 1800 trees, (b) the inclusion of blue stain, and (c) the introduction of the slave machine. As the number of factors recommended by $C_p$ increased past 5, the stability of the models diminished and rubbish predictive values were common. For this study, the harvest of 10 trees to mill 250 samples for mechanical testing was the upper limit of feasibility. For large scale operations such as plantation forestry or forest products manufacturing, probably more than 10 trees would be needed to yield acceptable equations.

The performance of the Mallow's $C_p$ statistic to predict which and how many variables were needed, for both MLR and PCR, was quite useful when double-checked against a test population. It should be noted that since paired samples were collected, then equivalent material properties were assumed and thus the difference in means between traits should be an accurate method to demonstrate a stain effect. With

justification through a test population, the traditional rule of ($C_p - p \leq 0$) appeared quite conservative when picking factors. For this study, a ($C_p - p \leq 5$) seemed to qualitatively be the transition point between stable and unstable models when clear wood calibrations were applied to a blue stained treatment for the slave machine scenario. Of course, this was not an absolute rule because on occasion very low $C_p$ values gave uncharacteristically poor models while very high $C_p$ values gave surprisingly good results, which was the case for the lignin prediction.

Finally, for the slave machine scenario, the influence of stain was most damaging to the mechanical property and extractive calibrations but was not a factor on selected models for lignin and density. Since lignin and density required fewer factors than MOE, MOR, and extractives, it was apparent that as the number of factors increased, it became increasingly difficult to avoid extraneous variation not present in the calibration set (in this case blue stain). For MOE, MOR, and extractives, the models followed the same distribution when the first derivative was first applied and PCR used. Given the segregation of stained and control samples along the PC2 axis (Figure 7.1) and given that similar histogram shapes occurred using PCR on 1$^{st}$ derivative pretreatment (Figure 7.3 and 7.4 (b)), one could feasibly adjust by some constant value when blue stain is present.

# CHAPTER 8

# CONCLUSIONS

Near infrared spectroscopy is a powerful technique to measure multiple wood properties simultaneously but is at the infancy of its development. My research presented demonstrates the ability to model modulus of elasticity, modulus of rupture, lignin, extractives, MFA, and tracheid length from the spectra acquired when longleaf pine increment cores were scanned. As such, the utility of this technique is targeted for use by genetics or silviculturists who might need a rapid technique to estimate multiple traits without destroying the sample.

Models I developed for MOE and MOR exhibited $R^2$ values near 0.90 while extractives explained the variation with an $R^2$ of 0.83. All three traits showed nearly equal performance when retested on the validation data set. Models developed for density and tracheid length exhibited an $R^2$ around 0.70 and dropped to 0.68 when the models were tested on the validation set. The estimate of density was primarily a function of the area under the curve based on Beer Lambert's Law while tracheid length was modeled as an indirect correlation with ring number from the pith. The estimate of modulus of elasticity and modulus of rupture was associated with lignin and cellulose related wavelengths, but more importantly, I determined that the same 4 of 5 wavelengths needed to predict MOE and MOR and thus helps to explain how the two were linearly related. Lignin provided the largest variation in prediction between models and yielded respective $R^2$ values of 0.40 to 0.55 when multiple linear regression and principal components regression was used. However, when additional models, such as partial least squares (PLS) regression were explored, I found an $R^2$ around 0.70 was possible for

lignin. For the remainder of the traits, PLS was not superior in $R^2$ when compared to MLR and PCR and so PLS was not reported to maintain consistency.

I found Microfibril angle (MFA) to correlate to Klason lignin content and as such, lignin variation was determined to be the link between NIR and MFA. A theoretical model was developed based on the covariance of lignin and MFA that might occur with increased tree growth. The theoretical model yielded promising results although more work is needed to justify its validity.

During experimentation, I found that blue stain and machine variation can add experimental error. As such, two sources of variation in the spectra matrix was investigated to determine the resistance of the calibration equations to bias. Blue stain and machine variability were introduced into the data set. Blue stain was found to not significantly influence the prediction of all 5 traits (density, alcohol-toluene extractives, Klason lignin, modulus of elasticity, and modulus of rupture) when specific models were chosen and one machine used. However, when a second machine was introduced to scan the prediction cores but calibrations from the master machine were maintained, I found that blue stain biased the ability to predict modulus of rupture, modulus of elasticity, and alcohol-toluene extractives.

Finally, an overall conclusion of this dissertation was that successful models can be built to predict traits from increment cores. However, it is probable that sample preparation and selection are the two most important variables for preparing samples for a large population of trees. If the samples prepared from a new population are different from those of calibration, then erroneous results may occur (Schultz and Burns 1990). But also, subtle differences between populations of trees, particularly at different

geographies, suggest that considerable time needs to go into calibration to ensure that the largest and representative range of variation is present. Future research is needed in the area of modeling multiple wood quality traits under multiple sources of variation for long periods of time since sources of variation may drift with time. Wood orientation, temperature, humidity, stain, or other biological invasions may change the spectra and introduce enough uncontrolled experimental error to make the models invalid.

# REFERNCES


Agarwal, U.P., I.A. Weinstock, and R.H. Atalla. 2003. FT-Raman spectroscopy for direct measurement of lignin concentrations in kraft pulps. Tappi J. 2(1): 22-26.

Antti, H., M. Sjöström, and L. Wallbäcks. 1996. Multivariate calibration models using nir spectroscopy on pulp and paper industrial applications. J. Chemometr. 10(5-6): 591-603.

Antti, H, D. Alexandersson, M. Sjöström, and L. Wallbäcks. 2000. Detection of kappa number distributions in kraft pulps using NIR spectroscopy and multivariate calibration. Tappi J. 83(3): 102-108.

Axrup, L., K. Markides, and T. Nilsson. 2000. Using miniature diode array NIR spectrometers for analyzing wood chips and bark samples in motion. J. Chemometr. 14(5-6): 561-572.

American Society for Testing Materials (ASTM). 1993. Standard methods of testing small clear specimens of timber. ASTM D 143-83, Philadelphia, PA, USA.

American Society for Testing Materials (ASTM). 1993. Standard test methods for specific gravity of wood and wood-base materials. ASTM D 2395-83, Philadelphia, PA, USA.

Baillères, H., F. Davrieux, and F.H. Pichavant. 2002. Near infrared analysis as a tool for rapid screening of some major wood characteristics in a eucalyptus breeding program. Ann. For. Sci. 59(5-6): 479-490.

Behrendt, C.J. and R.A. Blanchette. 2001. Biological control of blue stain in pulpwood: mechanisms of control used by *Phlebiopsis gigantean*. Holzforschung 55(3): 238-245.

Bell, A.A. and M.H. Wheeler. 1986. Biosynthesis and functions of fungal melanins. Annu. Rev. Phytopathol. 24: 411-451.

Bendtsen, B.A. 1978. Properties of wood from improved and intensively managed trees. For. Prod. J. 28 (10): 61-72.

Bendtsen, B.A., and J. Senft. 1986. Mechanical and anatomical properties in individual growth rings of plantation grown eastern cottonwood and loblolly pine. Wood Fib. Sci. 18(1): 23-38.

Bergander, A., J. Brändström, D. Daniel, and L. Salmén. 2002. Fibril angle variability in early wood of norway spruce using soft rot cavities and polarization confocal microscopy. J. Wood Sci. 48(4): 255-263.

Birkett, M.D. and M.J.T. Gambino. 1989. Estimation of pulp kappa number with near-infrared spectroscopy. Tappi J. 72(9): 193-197.

Boister, K.L., M.E. Martin, and J.D. Aber. 1996. Determination of carbon fraction and nitrogen concentration in tree foliage by near infrared reflectance: a comparison of statististical methods. Can. J. For. Res. 26(4): 590-600.

Brown, C.L. 1970. Physiology of wood formation in conifers. Wood Sci. 3: 8-22.

Cave, I.D. 1966. Theory of X-ray diffraction method for the measurement of microfibril angle in wood. For. Prod. J. 16: 37-42.

Cosgrove, D.J. 1999. Enzymes and other agents that enhance cell wall extensibility. Ann. Rev. Plant Phys. 50: 391-417.

Cown, D.J., G.D.Young, and R.D. Burdon. 1992. Variation in wood characteristics of 20-year-old half-sib families of *Pinus radiata*. New Zeal. J. For. Sci. 22(1): 63-76.

Dadswell, H.E., J.M. Fielding, W.P. Nicholls, and A.G. Brown. 1961. Tree-to-tree variations and the gross heritability of wood characteristics of *Pinus radiata.* Tappi J. 44(3): 174-179.

Donaldson, L.A. 2001. A three-dimensional computer model of the tracheid cell wall as a tool for interpretation of wood cell wall ultrastructure. IAWA J. 22(3): 213-233.

Draper, N.R. and W.G. Hunter. 1969. Transformations: some examples revisited. Technometrics 11: 23-40.

Duffy, G.G., and R. Kibblewhite. 1989. A new method of relating wood density, pulp quality, and paper properties. Appita J. 42(3): 209-214.

Eckhardt, L.G., R.A. Goyer, K.D. Klepzig, and J.P. Jones. 2004. Interactions of *Hylastes* species (Coleoptera: Scolytidae) with *Leptographium* species associated with loblolly pine decline. J. Econ. Entomol. 97(2): 468-474.

Emons, A.M.C., J.H.N. Schel, and B.M. Mulder. 2002. The geometrical model for microfibril deposition and the influence of the cell wall matrix. Plant Biol. 4(1): 22-26.

Erickson, H.D. and T. Arima. 1974. Douglas-fir wood quality studies part II: effects of age and stimulated growth on fibril angle and chemical constituents. Wood Sci. Tech. 8(4): 255-265.

Estienne, F., L. Pasti, V. Centne, B. Walczak, F. Despagne, D. J. Rimbaud, O.E. de Noord, and D.L. Massart. 2001. A comparison of multivariate techniques applied to experimental NIR data sets. Part II: Predictive ability under extrapolation conditions. Chemometr. Int. Lab. Sys. 58(2): 195-211.

Faber, N.M. 1999. A closer look at the bias-variance trade-off in multivariate calibration. J. Chemometr. 13(2): 185-192.

Farrington, A. 1980. Wood and digester factors affecting kraft pulp quality and uniformity. Appita J. 34(1): 40-46.

Fearn, T. 2001. Standardisation and calibration transfer for near infrared instruments: a review. J. Near Infrared Spectrosc. 9(4): 229-244.

Fengel, D. 1970. Ultrastructural behavior of cell wall polysaccharides. Tappi J. 53(3): 497-503.

Fleet, C., C. Breuil, and A. Uzunovic. 2001. Nutrient consumption and pigmentation of deep and surface colonizing sapstaining fungi in *Pinus contorta*. Holzforschung 55(4): 340-346.

Freund, J.E. and R.E. Walpole. 1980. Mathematical statistics: third edition. Prentice-Hall, Inc., Englewood Cliffs, NJ. 548 pp.

Gao, Y., C. Breuil, and T. Chen. 1994. Utilization of triglycerides, fatty acids and resin acids in lodgepole pine wood by a sapstaining fungus *Ophiostoma piceae*. Mater. Organismen 28 (2): 105-118.

Gierlinger, N., M. Schwanninger, B. Hinterstoisser, and R. Wimmer. 2002. Rapid determination of heartwood extractives in *Larix sp*. by means of Fourier transform near infrared spectroscopy. J. Near Infrared Specrosc. 10(3): 203-214.

Gillon, D., C. Houssard, and R. Joffre. 1999. Using near-infrared reflectance spectroscopy to predict carbon, nitrogen and phosphorus content in heterogeneous plant material. Oecologia 118(2): 173-182.

Gindl, W., M. Grabner, and R. Wimmer. 2000. The influence of temperature on latewood lignin content in treeline Norway spruce compared with maximum density and ring width. Trees 14(7): 409-414.

Gindl, W., A. Teischinger, M. Schwanninger, and B. Hinterstoisser. 2001. The relationship between near infrared spectra of radial wood surfaces and wood mechanical properties. J. Near Infrared Spectrosc. 9(4): 255-261.

Gindl, W. and A. Teischinger. 2002. Axial compression strength of Norway spruce related to structural variability and lignin content. Compos. A 33(12): 1623-1628.

Gindl, W., H.S. Gupta, T. Schöberl, H.C. Lichtenegger, and P. Fratzl. 2004. Mechanical properties of spruce wood cell walls by nanoindentation. Appl. Phy. A: Mat. Sci. Proc. 79: 2069-2073.

Giordano, R., A. Salleo, S. Salleo S., and F. Wanderlingh. 1978. Flow in xylem vessels and poiseuilles law. Can. J. Bot. 56(3): 333-338.

Gladstone, W.T., A.C. Barefoot, and B.J. Zobel. 1970. Kraft pulping of earlywood and latewood from loblolly pine. For. Prod. J. 20(2): 17-24.

Gunning, B. and A.R. Hardham. 1982. Microtubules. Ann. Rev. Plant Phys. 33: 651-698.

Hannrup, B. and I. Ekberg. 1998. Age-age correlations for tracheid length and wood density in *Pinus sylvestris*. Can. J. For. Res. 28(9): 1373-1379.

Hauksson, J.B., G. Bergqvist, U. Bergsten, M. Sjöström, and U. Edlund. 2001. Prediction of basic wood properties for norway spruce. Interpretation of near infrared spectroscopy data using partial least squares. Wood Sci. Tech. 35(6): 475-485.

Hepworth, D.G. and J.F.C. Vincent. 1998. Modeling the mechanical properties of xylem tissue from tobacco plants (Nicotiana tabacum 'Samsun') by considering the importance of molecular and micromechanisms. Ann. Bot. 81(6): 761-770.

Heyn, A.N.J. 1969. The elementary fibril and supermolecular structure of cellulose in soft wood fiber. J. Ultra Mol. Struct. 26: 52-68.

Hiukka, R. 1998. A multivariate approach to the analysis of pine needle samples using NIR. Chemometr. Int. Lab. Sys. 44(1-2): 395-401.

Hodges, J.D. and P.L. Lorio. 1969. Carbohydrate and nitrogen fractions of the inner bark of loblolly pines under moisture stress. Can. J. Bot. 47: 1651-1657.

Hodges, R.E., and G. A. Krishnagopalan. 1999. Near-infrared spectroscopy for on-line analysis of white and green liquors. Tappi J. 82(9): 101-106.

Hoffmeyer, P. and J.G. Pedersen. 1995. Evaluation of density and strength of Norway spruce by near infrared reflectance spectroscopy. Holz Roh Werkst. 53(3): 165-170.

Hoffmann, B., B. Chabbert, B. Monties, and T. Speck. 2003. Mechancal, chemical and x-ray analysis of wood in the two tropical lianas Bauhinia guianensis and Condylocarpon guianense: variations during ontogeny. Planta 217(1): 32-40.

Hoffmeyer, P. and J.G. Pedersen. 1995. Evaluation of density and strength of Norway spruce by near infrared reflectance spectroscopy. Holz Roh Werkst. 53(3): 165-170.

Jääskeläinen, A. S., M. Nuopponen, P. Axelsson, M. Tenhunen, M. Löija, and T. Vuorinen. 2003. Determination of lignin distribution in pulps by FTIR ATR spectroscopy. J. Pulp Pap. Sci. 29(10): 328-331.

Jackson, L.W.R. and J.T. Greene. 1958. Tracheid length variation and inheritance in slash and loblolly pine. For. Sci. 4(4): 316-318.

Jacobs, K. and M.J. Wingfield. 2001. *Leptographium* species: tree pathogens, insect associates and agents of blue-stain. APS Press, St. Paul, MN., pp224.

Jacobs, K., M.J. Wingfield, N.V. Pashenova, and V.P. Vetrova. 2000. A new *Leptographium* species from Russia. Mycol. Res. 104: 1524-1529.

Jurasek, L. 1995. Toward a 3-dimensional model of lignin structure. J. Pulp Pap. Sci. 21(8): 274-279.

Kärenlampi, P., and H.S. Hamari. 1997. Classified wood raw materials for diversified softwood kraft pulps. Pap. Puu-Pap. Tim. 79(6): 404-410.

Kellogg, R.M., C.B.R. Sastry, and R.W. Wellwood. 1975. Relationships between cell-wall composition and cell wall density. Wood Fib. Sci. 7(3): 170-177.

Kelly, S.S., T.G. Rials, B. Snell, L.H. Groom, and A. Sluter. 2004a. Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood. Wood Sci. Tech. 38(4): 257-276.

Kelly, S.S., T.G. Rials, L.R. Groom, and C.L. So. 2004b. Use of near infrared spectroscopy to predict the mechanical properties of six softwoods. Holzforschung 58 (3): 252-260.

Kibblewhite, R.P. 1984. Radiata pine wood and kraft pulp quality relationships. Appita J. 37(9): 741-747.

Kibblewhite, R.P., R. Evans, and M.J.C. Riddell. 1997. Handsheet property prediction from kraft-fibre and wood-tracheid properties in eleven radiate pine clones. Appita J. 50(2): 131-138.

Kleppe, P.J. 1970. The process of, and products from, kraft pulping of southern pine. For. Prod. J. 20(5): 50-59.

Koch P. 1972. Utilization of the southern pines. Vol. 1: The raw material. U.S. Department of Agriculture, Washington D.C. pp.1663.

Köhler, L., T. Speck, and H. Spatz. 2000. Micromechanisms and anatomical changes during early ontogeny of two lianescent Aristolochia species. Planta 210(5): 691-700.

Köhler, L. and H.C. Spatz. 2002. Micromechanics of plant tissues beyond the linear-elastic range. Planta 215(1): 33-40.

Koubaa A., R.E. Hernández, M. Beaudoin, and J. Poliquin. 1998. Interclonal, intraclonal, and within-tree variation in fiber length of poplar hybrid clones. Wood Fib. Sci. 30(1): 40-47.

Kowalski, K.G. 1990. On the predictive performance of biased regression methods and multiple linear regression. Chemometr. Int. Lab. Sys. 9(2): 177-184.

Kreber, B. and J.J. Morrell. 1993. Ability of selected bacterial and fungal bioprotectants to limit fungal stain in ponderosa pine sapwood. Wood Fib. Sci. 25(1): 23-34.

Labosky, P. and G. Ifju. 1972. A study of loblolly pine growth increments. Part II. Pulp yield and related properties. Tappi J. 55(4): 530-534.

Loo, J.A., C.G. Tauer, and J.P. Van Buijtenen. 1984. Juvenile-mature relationships and heritability estimates of several traits in loblolly pine (*Pinus taeda*). Can. J. For. Res. 14(6): 822-825.

Larson, P.R. 1966. Changes in chemical composition of wood cell walls associated with age in Pinus Resinosa. For. Prod. J. 16: 37-45.

Law, D.P. and Tkachuk R. 1977. Determination of moisture content in wheat by near-infrared diffuse reflectance spectrophotometry. Cereal Chem. 54(4): 874-881.

Lebow, P.K., C.C. Brunner, A.G. Maristany, and D.A. Butler. 1996. Classification of wood surface features by spectral reflectance. Wood Fib. Sci. 28(1): 74-90.

Lewis, C.B., R.J. Mcnichols, A. Gowda, and G.L. Cote. 2000. Investigation of near-infrared spectroscopy for periodic determination of glucose in cell culture media *in situ*. App. Spectro. 54(10): 1453-1457.

Lichtenegger, H., A. Reiterer, S.E. Tschegg, and P. Fratzl. 1999. Variation of cellulose microfibril angles in softwoods and hardwoods- a possible strategy of mechanical optimization. J. Struct. Biol. 128(3): 257-269.

Liese, W. 1970. Ultrastructural aspects of woody tissue disintegration. Annu Rev Phytopathol. 8: 231–257.

Likens, G.E. and F.H. Bormann. 1970. Chemical analysis of plant tissues from the Hubbard Brook ecosystem in New Hampshire. Yale University: Bulletin No. 79 pp. 24.

Lindström, H., J.W. Evans, and S.P. Verrill. 1998. Influence of cambial age and growth conditions on microfibril angle in young Norway spruce (Picea abies [L.] Karst.). Holzforschung 52(6): 573-581.

Lundgren, C. 2004. Microfibril angle and density patterns of fertilized and irrigated n norway spruce. Silva Fen. 38(1): 107-117.

Malkavaara, P. and R. Alén. 1998. A spectroscopic method for determining lignin content of softwood and hardwood kraft pulps. Chemometr. Int. Lab. Sys. 44(1-2): 287-292.

Malin, S.F., T.L. Ruchti, T.B. Blank, S.N. Thennadil, and S.L. Monfre. 1999. Noninvasive prediction of glucose by near-infrared diffuse reflectance spectroscopy. Clin. Chem. 45(9): 1651-1658.

Martin, M.E. and J.D. Aber. 1994. Analysis of forest foliage III: Determining nitrogen, lignin and cellulose in fresh leaves using near infrared reflectance data. J. Near Infrared Spectrosc. 2: 25-32.

Matziris, D.I., B.J. Zobel. 1973. Inheritance and correlations of juvenile characteristics in loblolly pine (*Pinus taeda L.*). Silvae Genet. 22:38-45

McCann, M.C., N.J. Stacey, R. Wilson, and K. Roberts. 1993. Orientation of macromolecules in the walls of elongating carrot cells. J. Cell Sci. 106: 1347-1356.

McLellan, T.M., J.D. Aber, M.E. Martin, J.M. Melillo, and K.J. Nadelhoffer. 1991. Determination of nitrogen, lignin, and cellulose content of decomposing leaf material by near infrared reflectance spectroscopy. Can. J. For. Res. 21(11): 1684-1688.

McLellan, T.M., M.E. Martin, J.D. Aber, J.M. Melillo, K.J. Nadelhoffer, and B. Dewey. 1991b. Comparison of wet chemistry and near infrared reflectance measurements of carbon-fraction chemistry and nitrogen concentration of forest foliage. Can. J. For. Res. 21: 1689-1693.

Meder, R., A. Thumm, and H. Bier. 2002. Veneer stiffness predicted by NIR spectroscopy calibrated using mini-LVL test panels. Holz Roh Werkst. 60(3): 159-164.

Meder, R., A. Thumm, and D. Marston. 2003. Sawmill trial of at-line prediction of recovered lumber stiffness by NIR spectroscopy of *Pinus radiate* cants. J. Near Infrared Specrosc. 11(2): 137-143.

Meglen, R. and S. Kelley. 2000 National Renewable Energy Lab, USA Document.

Megraw, R.A. 1985. Wood Quality Factors in Loblolly Pine. Tappi Press, Atlanta. pp.88.

Michell, A.J. 1995. Pulpwood quality estimation by near-infrared spectroscopic measurements on eucalypt woods. Appita J. 48(6): 425-428.

Michell, A.J., and L.R. Schimleck. 1998. Developing a method for the rapid assessment of pulp yield of plantation eucalypt trees beyond the year 2000. Appita J. 51(6): 428-432.

Miranda, I. and H. Pereira. 2002. Variation of pulpwood quality with provenances and site in *Eucalyptus globulus*. Ann. For. Sci. 59(3): 283-291.

Neal, T.A. and D.W. Ross. 1999. Pathogenicity to western larch (*Larix occidentalis*) of two fungi, *Ophiostoma pseudotsugae* and *Leptographium abietinum*, associated with the Douglas fir beetle (Coleoptera: Scolytidae) Ag. Forest Entomol. 1:203-207.

Nelson, R.M. 1934. Effects of bluestain fungi on southern pines attacked by bark beetles. Phytopathol. Z. 4:327-353.

Neter, J., M.H. Kutner, C.J. Nachtsheim, and W. Wasserman. 1996. Applied Linear Statistical Models. 4th Ed. Richard D. Irwin, Inc., Burr Ridge, Illinois, USA, pp1408.

Nicholls, J.W.P., H.E. Dadswell, and J.M. Fielding.1964. The heritability of wood characteristics of *Pinus radiata*. Silvae Genet. 13: 68-71.

Nieduszy, I. and R.D. Preston. 1970. Crystallite size in natural cellulose. Nature 225(5229): 273-274.

Ona, T., T. Sonoda, J. Ohshima, S. Yokota, and N. Yoshizawa. 2003. A rapid quantitative method to assess *Eucalyptus* wood properties for kraft pulp production by FT-Raman spectroscopy. J. Pulp Pap. Sci. 29(1): 6-10.

Osborne, B.G. and T. Fearn. 1986. Near infrared spectroscopy in food analysis. Longman Scientific and Technical, Essex, England, pp200.

Otrosina, W.J., N.J. Hess, S.J. Zarnoch, T.J. Perry, and J.P. Jones 1997. Blue-stain fungi associated with roots of southern pine trees attacked by the southern pine beetle, *Dendroctonus frontalis*. Plant Dis. 81(8): 942-945.

Ott, R.L. 1993. An introduction to statistical methods and data analysis. Wadsworth, Inc., Belmont, CA., pp1152.

Patterns in Nature. 2004. http://acept.la.asu/PiN/rdg/color/color.shtml

Pearson, R.G. and R.C. Gilmore. 1971. Characterization of the strength of juvenile wood of loblolly pine. For. Prod. J. 21(1): 23-31.

Pearson, R.G. and R.C. Gilmore. 1980. Effect of fast growth rate on the mechanical properties of loblolly pine. For. Prod. J. 30(5): 47-54.

Poke, F.S., J.K. Wright, and C.A. Raymond. 2004. Predicting extractives and lignin contents in Eucalyptus globulus using near infrared reflectance analysis. J. Wood Chem. Tech. 24(1): 55-67.

Pot D., G. Chantre, P. Rozenberg, J.C. Rodrigues, G.L. Jones, H. Pereira, B. Hannrup, C. Cahalan, and C. Plomion. 2002. Genetic control of pulp and timber properties in maritime pine (*Pinus pinaster* Ait.). Ann. For. Sci. 59(5-6): 563-575.

Raymond, C.A. and L.R. Schimleck. 2002. Development of near infrared reflectance analysis calibrations for estimating genetic parameters for cellulose content in *Eucalyptus globules*. Can. J. For. Res. 32(1): 170-176.

Reiterer, A., H. Lichtenegger, S. Tschegg, and P. Fratzl. 1999. Experimental evidence for a mechanical function of the cellulose microfibril angle in wood cell walls. Philos. Mag. A 79(9): 2173-2184.

Rials, T.G., S.S. Kelley, and C.L. So. 2002. Use of advanced spectroscopic techniques for predicting the mechanical properties of wood composites. Wood Fib. Sci. 34(3): 398-407.

Riley, M.R. and L.C. Cánaves. 2002. FT-NIR spectroscopic analysis of nitrogen in cotton leaves. Appl. Spectro. 56(11): 1484-1489.

Roelofsen RA. 1969.The Plant Cell Wall. Gerbrüder Borntraeger, Berlin, pp335.

Ross, D.W., P. Fenn, and F.M. Stephen. 1992. Growth of southern pine-beetle associated fungi in relation to the induced wound response in loblolly pine. Can. J. For. Res. 22(12): 1851-1859.

Saka, S. and M. Tsuji. 1987. The relationship between microfibril orientation in the tracheid $S_2$ layer and the lignin content of coniferous woods. Cell. Chem. Technol. 21(3): 225-231.

Sakurada, I., Y. Nukushima, and T. Ito. 1962. Experimental determination of the elastic modulus of crystalline regions in oriented polymers. J. Polymer Sci. 57: 651-660.

Sanderson, M.A., F. Agblevor, M. Collins, and D.K. Johnson. 1996. Compositional analysis of biomass feedstocks by near infrared reflectance spectroscopy. Biomass Bioenerg. 11(5): 365-370.

SAS, Statistical Analysis Software. 1999. version 7. Raleigh, N.C.

Sarén, M.P., R. Serimaa, S. Andersson, P. Saranpää, J. Keckes, and P. Fratzl. 2004. Effect of growth rate on mean microfibril angle and cross-sectional shape of tracheids of Norway spruce. Trees 18(3): 354-362.

Schimleck, L.R., A.J. Michell, and P. Vinden. 1996. NIR spectroscopy and principal components analysis. Appita J. 49(5): 319-324.

Schimleck, L.R., A.J. Michell, and C.A. Raymond. 2000. Effect of site on the within-tree variation of wood properties of eucalypts as determined by NIR spectroscopy and multivariate analysis. Appita J. 53(4): 318-322.

Schimleck, L.R., R. Evans, and J. Ilic. 2001a. Application of near infrared spectroscopy to a diverse range of species demonstrating wide density and stiffness variation. IAWA J. 22(4): 415-429.

Schimleck, L.R., R. Evans, and J. Ilic . 2001b. Estimation of *Eucalyptus delegatensis* wood properties by near infrared spectroscopy. Can. J. For. Res. 31(10): 1671-1675.

Schimleck, L.R. and R. Evans. 2002a. Estimation of microfibril angle of increment cores by near infrared spectroscopy. IAWA J. 23(3): 225-234.

Schimleck, L.R. and R. Evans. 2002b. Estimation of wood stiffness of increment cores by near infrared spectroscopy: the development and application of calibrations based on selected cores. IAWA J. 23(3): 217-224.

Schimleck, L.R., R. Evans, and A.C. Matheson. 2002. Estimation of *Pinus radiata* D. Don clear wood properties by near-infrared spectroscopy. J. Wood Sci. 48(2): 132-137.

Schimleck, L.R. and R. Evans. 2003. Estimation of air-dry density of increment cores by near infrared spectroscopy. Appita J. 56(4): 312-317.

Schimleck, L.R., R. Evans, and J. Ilic. 2003b. Application of near infrared spectroscopy to the extracted wood of a diverse range of species. IAWA J. 24(4): 429-438.

Schimleck L.R., Mora C., and R.F. Daniels. 2003. Estimation of the physical wood properties of green Pinus taeda radial samples by near infrared spectroscopy. Can. J. For. Res. 33(12): 2297-2305.

Schimleck, L., R. Stürzenbecher, P.D. Jones, and R. Evans. 2004. Development of wood property calibrations using near infrared spectra having different spectral resolutions. J. Near Infrared Spectrosc. 12(1): 55-61.

Schimleck, L., R. Evans, D.P. Jones, R.F. Daniels, G.F. Peter, and A. Clark. IN PRESS. Estimation of microfibril angle and stiffness by near infrared spectroscopy using sample sets having limited wood density variation. IAWA J. IN PRESS.

Schultz, T.P. and D.A. Burns. 1990. Rapid secondary analysis of lignocellulose – Comparison of near infrared (NIR) and Fourier Transform Infrared (FTIR). Tappi J. 73 (5): 209-212.

Shupe, T.F., E.T. Choong, and C.H. Yang. 1996. The effects of silvicultural treatments on the chemical composition of plantation-grown loblolly pine wood. Wood Fib. Sci. 28(3):295-300.

Shupe T.F., C.Y. Hse, E.T. Choong, and L.H. Groom. 1997. Differences in some chemical properties of innerwood and outerwood from five silviculturally different loblolly pine stands. Wood Fib. Sci. 29(1): 91-97.

Shupe, T.F., E.T. Choong, M.D. Gibson, G.A. Grozdits, and O.V. Harding. 1998. The effects of previous drying on shrinkage and moisture content of some southern bottomland hardwoods. Wood Fib. Sci. 30(3): 273-280.

Shupe, T.F., C.Y. Hse, E.T. Choong, and L.H. Groom. 1999. Effect of silvicultural practice and wood type on loblolly pine particleboard and medium density fiberboard properties. Holzforschung 53(2): 215-222.

Singh, A.P. and G. Daniel. 2001. The S2 layer in the tracheid walls of Picea abies wood: inhomogeneity in lignin distribution and cell wall microstructure. Holzforschung 55(4): 373-378.

Smith, G.D., A. Jervis, M. Lennartsson, and W.F. Bourne. 2001. Laboratory methods of estimating potentially mineralizable nitrogen in organic potting mixes. II. Development of near infrared reflectance spectroscopy method. Commun. Soil Sci. Plant Anal. 32 (17-18): 2769-2781.

Smith, K.F., R.J. Simpson, and R.D. Armstrong. 1998. Seasonal variation in the herbage yield and nutritive value of perennial ryegrass (Lolium perenne L.) cultivars with high or normal herbage water-soluble carbohydrate concentrations grown in three contrasting Australian dairy environments. Aust. J. Exp. Agr. 38(8): 821-830.

Smith W.J. 1967. The heritability of fibre characteristics and its application to wood quality improvement in forest trees. Silvae Genet. 16: 41-50.

So, C.L., L.H. Groom, T.G. Rials, R. Snell, S.S. Kelly, and R. Meglen. 2002. Rapid assessment of the fundamental property variation of wood. Outcalt, Kenneth W., (eds) In: Proceedings of the eleventh biennial southern silvicultural research conference. Gen. Tech. Rep. SRS-48. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station, 622pp.

Solheim, H., Krokene, P., and Långström, B. (2001) Effects of growth and virulence of associated blue-stain fungi on host colonization behaviour of the pine shoot beetles *Tomicus minor* and *T. piniperda*. Plant Pathol. 50(1): 111-116.

Solla, A. and L. Gil. 2002. Xylem vessel diameter as a factor in resistance of *Ulmus minor* to *Ophiostoma novo-ulmi*. For. Pathol. 32(2): 123-134.

Statistical Analysis Software (SAS) 2001. version 8.2. Cary, North Carolina, USA.

Sun, R.C. and J. Tomkinson. 2002. Extraction and characterization of lipophilic extractives from wheat straw. II. Spectroscopic and thermal characterizations. Cell. Chem. Technol. 36(1-2): 93-103.

Swierenga, H., F. Wülfert, O.E. de Noord, A.P. de Weijer, A.K. Smilde, and L.M.C. Buydens. 2000. Development of robust calibration models in near infra-red spectrometric applications. Anal. Chem. Acta 411(1-2): 121-135.

Tappi Standard. 1997. Solvent extractives of wood and pulp. T 204 cm-97, Atlanta, GA, U.S.A.

Tappi Standard. 1998. Acid-insoluable lignin in wood and pulp. T 222 om-98, Atlanta, Ga, U.S.A.

Taylor, J.G., Owen, T.P., L.T. Koonce, and C.H. Haigler. 1992. Dispersed lignin in tracheary elements treated with cellulose synthesis inhibitors provides evidence that molecules of the secondary cell wall mediate wall patterning. Plant J. 2(6): 959-970.

Thumm, A., and R. Meder. 2001. Stiffness prediction of radiate pine clearwood test pieces using near infrared spectroscopy. J. Near Infrared Specrosc. 9(2): 117-122.

Thygesen, L.G., and S.O. Lundqvist. 2000. NIR measurement of moisture content in wood under unstable temperature conditions. Part 1. Thermal effects in near infrared spectra of wood. J. Near Infrared Specrosc. 8(3): 183-189.

Treacy, M., J. Evertsen, and A. Dhubháin. 2000. National Council for Forest Research Development (COFORD) Document. Project Number 96125.

Tsuchikawa, S., K. Inoue, J. Noma, and K. Hayashi. 2003. Application of near-infrared spectroscopy to wood discrimination. J. Wood Sci. 49(1): 29-35.

Tsuchikawa, S. and K. Yamato. 2003. Discriminant analysis of wood-based materials with weathering damage by near infrared spectroscopy. J. Near Infrared Specrosc. 11(5): 391-399.

Unscrambler software. 1999. version 7.5 Vika, Norway.

Uzunovic, A. and J.F. Webber. 1998. Comparison of bluestain fungi grown *in vitro* and in freshly cut pine billets. Eur. J. For. Path. 28(5): 323-334.

Veal, M.A., G.R. Marrs, and M. Jackson. 1987. Control over the quality of loblolly pine chips. Tappi J. 70(1): 51-54.

Via, B.K., T.F. Shupe, L.H. Groom, M. Stine, and C.L. So. 2003. Multivariate modeling of density, strength and stiffness from near infrared spectra for mature, juvenile and pith wood of longleaf pine (*Pinus palustris*). J. Near Infrared Specrosc. 11(5): 365-378 (CHAPTER 2).

Via, B.K., M. Stine, T. Shupe, C.L. So, and L. Groom. 2004. Feasibility of Genetic Improvement of Fiber Length and Coarseness Based on Paper Product Performance and Material Variability: A Review. IAWA J.: IN PRESS.

Viiri, H., P. Niemelä, V. Kitunen, and E. Annila. 2001. Soluble carbohydrates, radial growth and vigour of fertilized Norway spruce after inoculation with blue-stain fungus, *Ceratocystis polonica.* Trees 15(6): 327-334.

Viiri, H., E. Annila, V. Kitunen, and P. Niemelä. 2001b. Induced responses in stilbenes and terpenes in fertilized Norway spruce after inoculation with blue-stain fungus, *Ceratocystis polonica*. Trees 15(2): 112-122.

Viitanen, H.A. and J. Bjurman.1995. Mold growth on wood under fluctuating humidity conditions. Mater. Organismen 29(1): 27-46.

Viitanen, H.A. 1997. Modelling the time factor in the development of mould fungi – the effect of critical humidity and temperature conditions on pine and spruce sapwood. Holzforschung 51(1): 6-14.

Wallbäcks, L., U. Edlund, B. Nordén, and Berglund. 1991. Multivariate characterization of pulp using solid-state $^{13}C$ NMR, FTIR, and NIR. Tappi J. 74(10): 201-206.

Wardrop, A.B. and H.E. Dadswell. 1955. The development of the conifer tracheid. Holzforschung 7: 33-39.

Wang, D., F.E. Dowell, and R.E. Lacey. 1999. Single wheat kernel size effects on near-infrared reflectance spectra and color classification. Cereal Chem. 76(1): 34-37.

Wang, Q. 1994. Growth on mould and stain fungi on wood-based boards in relation to temperature and relative humidity. Mater. Organismen 28(8): 81-103.

Wang, Z., T. Chen, Y. Gao, C. Breuil, and Y. Hiratsuka. 1995. Biological degradation of resin acids in wood chips by wood-inhabiting fungi. App. Environ. Microb. 61(1): 222-225.

Wentzell, P.D. and L.V. Montoto. 2003. Comparison of principal components regression and partial least squares regression through generic simulations of complex mixtures. Chemometr. Int. Lab. Sys. 65(2): 257-279.

White-McDougall, W.J., R.A. Blanchette, and R.L. Farrell. 1998. Biological control of blue stain fungi on *Populus tremuloides* using selected *Ophiostoma* isolates. Holzforschung 52(3): 234-240.

Wilcox, W.W. (1970) Anatomical changes in wood cell walls attacked by fungi and bacteria. Bot. Rev. 36(1):1-28.

Wimmer, R., G.M. Downes, and R. Evans. 2002. Temporal variation of microfibril angle in Eucalyptus nitens grown in different irrigation regimes. Tree Physiol. 22(7): 449-457.

Whiteman P.H., J.N. Cameron, and A. Farrington. 1996. Breeding trees for improved pulp and paper production – a review. Appita J. 49(1): 50-53.

Wright J.A., M.D. Birkett, and M.J.T. Gambino. 1990. Prediction of pulp yield and cellulose content from wood samples using near infrared spectroscopy. Tappi J. 73(8): 164-166.

Wülfert, F. W.T. Kok, O.E. de Noord, and A.K. Smilde. 2000. Linear techniques to correct for temperature-induced spectral variation in multivariate calibration. Chemometr. Int. Lab. Sys. 51(2): 189-200.

Yang, J.L. and Y. Fortin. 2001. Evaluating strength properties of Pinus radiata from ultrasonic measurements on increment cores. Holzforschung 55(6): 606-610.

Yeh, T.F., H.M. Chang, and J.F. Kadla. 2004. Rapid prediction of solid wood lignin content using transmittance near-infrared spectroscopy. J. Ag. Food Chem. 52(6): 1435-1439.

Zink, P. and D. Fengel. 1988. Studies on the colouring matter of blue-stain fungi. Part 1. General characterization and the associated compounds. Holzforschung 42(4): 217-220.

Zink, P. and D. Fengel. 1989. Studies on the colouring matter of blue-stain fungi. Part 2. Electron microscopic observations of the hyphae walls. Holzforschung 43(6): 371-374.

Zink, P. and D. Fengel. 1990. Studies on the colouring matter of blue-stain fungi. Part 3. Spectroscopic studies on fungal and synthetic melanins. Holzforschung 44(3): 163-168.

Zobel B., E. Thorbjornsen, and F. Henson. 1960. Geographic, site, and individual tree variation in wood properties of loblolly pine. Silvae Genet. 9:149-158.

Zulpa, G., M.C. Zaccaro, F. Boccazzi, J.L. Parada, and M. Storni. 2003. Bioactivity of intra and extracellular substances from cianobacteria and lactic acid bacteria on “Wood blue stain” fungi. Biol. Control 27(3): 345-348.

# APPENDIX: LETTER OF PERMISSION




# Journal of Near Infrared Spectroscopy


Mr Brian K Via
School of Renewable Natural Resources
Louisiana State University Agricultural Center
Baton Rouge, Louisiana, 70803
United States of America

Dear Brian,

I am pleased that you are about to submit your thesis for examination for studies involving near infrared spectroscopy.

It is a common practice in some countries to bind into the thesis a hard copy of each paper which was published on relevant material by the candidate prior to submission of the thesis.

I hereby advise that the Journal of Near Infrared Spectroscopy agrees that you may add a copy of the following paper to your thesis –

*Via, BK, Shupe,TF, Groom, L, Stine, M and So, CL (2003) Multivariate modeling of density, strength, and stiffness from near infrared spectra for mature, juvenile, and pith wood of longleaf pine (Pinus palustris). J Near Infrared Spectrosc. 11(5), 365-378.*

All the best for the defense of your thesis.

Graeme D Batten
Editor-in-Chief

17th September 2004


Published by NIR Publications
Partners: A.M.C. Davies and I. Michael

www.nirpublications.com
VAT Registration No.: GB582 5597 00

# VITA

The author was born in Roanoke, Virginia (USA), in 1972. He received his degree of bachelor of science in 1996 at Virginia Polytechnic Institute (Virginia Tech) in forest products marketing and management. In 1998, he received his degree of master of science at Virginia Tech in wood mechanics. He then became a wood quality project leader at International Paper where he led a research program between 1999 to 2002. After that, he began his doctorial studies at Mississippi State University in the summer of 2002 in wood physics. However, after 1 year, his advisor left and he transferred to Louisiana State University to continue his studies in wood quality/physics. At the fall of 2004 commencement, the author will obtain his Doctor of Philosophy degree.