

ASSESSMENT OF PINE BIOMASS DENSITY THROUGH MID-INFRARED SPECTROSCOPY AND MULTIVARIATE MODELING

Brian K. Via,^{a,b*} Oladiran Fasina,^b and Hui Pan^c

The assessment of wood biomass density through multivariate modeling of mid-infrared spectra can be useful for interpreting the relationship between feedstock density and functional groups. This study looked at predicting feedstock density from mid-infrared spectra and interpreting the multivariate models. The wood samples possessed a random cell wall orientation, which would be typical of wood chips in a feedstock process. Principal component regression and multiple linear regression models were compared both before and after conversion of the raw spectra into the 1st derivative. A principal component regression model from 1st derivative spectra exhibited the best calibration statistics, while a multiple linear regression model from the 1st derivative spectra yielded nearly similar performance. Earlywood and latewood based spectra exhibited significant differences in carbohydrate-associated bands (1000 and 1060 cm^{-1}). Only statistically significant principal component terms (alpha less than 0.05) were chosen for regression; likewise, band assignments only originated from statistically significant principal components. Cellulose, lignin, and hemicellulose associated bands were found to be important in the prediction of wood density.

Keywords: FTIR; Biomass; Spectroscopy; Process monitoring; Nondestructive; Cellulose; Lignin; Hemicellulose

Contact information: a: School of Forestry and Wildlife Sciences, Auburn University, 3301 Forestry and Wildlife Sciences Bldg. Auburn, AL 36804 USA; b: Department of Biosystems Engineering, Auburn University, 209 Tom E. Corley Building, Auburn, AL 27695-8005 USA; c: Calhoun Research Station, LSU AgCenter, 321 Highway 80 E., Calhoun, LA 7122 ; *Corresponding author: bkv0003@auburn.edu

INTRODUCTION

Vibrational spectroscopy has recently received increased attention because of its versatility in the determination of the chemical composition in biological materials and in process quality control. Vibrational spectroscopy includes near infrared reflectance (NIR) and Fourier Transform Infrared Reflectance (FTIR) spectroscopy. NIR has been utilized more than FTIR because it requires little to no sample preparation, predictions can be made in seconds, and multiple properties can be predicted from a single spectrum. NIR has therefore been used in several applications such as quantifying the properties of non-woody including corn stover, miscanthus, switchgrass, and corn barley (Hodgson et al. 2010; Liu et al. 2010, Sohn et al. 2007) and woody biomass (Nkansah et al. 2010; Yao et al. 2010) and for assessing the density of wood (Schimleck et al. 2001; Via et al. 2003, 2005a).

Fourier Transform Infrared Spectroscopy (FTIR) is now receiving increased attention because of the availability of the ATR-diamond reflectance method that has

made it possible for the spectra of solid samples to be acquired in seconds. In addition, the peaks of FTIR spectra are more interpretable, resulting in easier qualitative analysis. In the past, assessment of specific peaks was performed to determine the effect of a perturbation on functional groups. Now FTIR data can be coupled with multivariate analysis to build prediction equations of the trait of interest. Even though interpretation of functional groups in the near infrared region is also possible through multivariate modeling (Via et al. 2009), FTIR is still superior in peak resolution for the raw spectra, and this may translate into better resolution of global and local peaks during multivariate modeling. These advantages make FTIR a viable option for laboratory analysis of biological materials.

One hurdle for using FTIR spectroscopy on solid wood samples could be the cellular orientation in wood and its effect on reflectance, absorbance, and/or transmission of light (Tsuchikawa and Tsutsumi 1999). For instance, when using NIR spectroscopy to scan solid wood samples, there can be significant differences in absorbance and shapes of the spectra between the tangential, radial, and transverse surfaces (Defo et al. 2007; Schimleck et al. 2005). The few studies that have involved the use of FTIR for density assessment of wood were carried out on ground samples or samples where the surface presentation was controlled and band assignments were made (Freer et al. 2003; Meder et al. 1999; Ruiz et al. 2005).

Band assignments can be of particular importance in quality control, because control charting methods can be utilized to determine when the process is out of control based on shifts in functional groups or principal components that represent key functional groups (Geladi et al. 2004; Kauper and Ferri 2004). Any shift in the functional groups could be an indication of a change in feedstock quality. Impacts of functional group shift on density can be understood through model investigation and precision adjustments to the process can be made accordingly.

The objective of this study was to utilize FTIR spectroscopy and multivariate modeling to predict solid wood density. Other goals were to a) identify those functional groups important in the prediction of density, b) compare principal components regression (PCR) and multiple linear regression (MLR) performance, and c) determine if transforming the spectra with a first derivative pretreatment would improve both the prediction and interpretation of models.

EXPERIMENTAL

Materials Selection and Density Measurement

Southern pine (*Pinus* spp.) wood samples were generated from a rotating knife in the long direction of the original wood axis at a local wood manufacturing plant. Collection was done over the course of a day to ensure a wide range of samples and to increase the likelihood that each flake was independent from one another. A flake was considered independent if it was not adjacent to another one within the tree. Four hundred fifty samples were selected at random with the constraint that a significant area of the sample be mostly free of fracture and defects such as knots. Samples were further milled with a band saw (Delta model 40-570) to a dimension of approximately 5 to 10

mm parallel to grain, 5 to 10 mm perpendicular to grain, and 0.8 to 1 mm in thickness. The variability in sample dimensions was necessary to ensure that a complete earlywood and latewood zone was present within each sample (which were then scanned later with FTIR). The earlywood is that wood which is produced during the spring time and has a very low density, while latewood is that which is produced during the summer and has a very high density. The samples were then placed into a dessicator where they were completely dried to bone dry density with Drierite (CaSO_4) at 20°C and then each one was placed into a separate ziplock type bag (150 x 70 mm). The weights were measured to 0.0001 g, dimensions were measured to 0.01 mm, and these measurements were utilized in the calculation of density. The dimension in any plane was based on the average of 3 measurements. After calculating the density of these 450 samples, the lowest 33 and highest 32 in density were selected. The medium density group were also selected ($n=33$) around the mean density.

FTIR Analysis

Mid-IR spectra were collected between 4000 and 650 cm^{-1} using a PerkinElmer Spectrum model 400 (Perkin Elmer Co., Waltham, MA) outfitted with a single reflectance ATR diamond. The earlywood and latewood zones of each sample were each scanned four times and at a resolution of 4 cm^{-1} . All scans were carried out at a temperature of $22^\circ\text{C} \pm 1$. Because each sample was stored in a separate ziplock type bag, they did not pick up any excess moisture during temporary opening of the dessicator. Scanning of the sample occurred immediately after withdrawal from the bag.

Multivariate Modeling and Spectra Preprocessing

Prior to multivariate modeling, each spectrum was adjusted to a mean = 0 and a standard deviation = 1. In addition to the raw spectra, the 1st derivative of the raw spectrum was computed to see if baseline variation could be removed, resulting in improved regression diagnostics. The 1st derivative was computed by computing the slope between every two consecutive points/wavenumber interval in a spreadsheet.

Prior to regression, the absorbance data obtained from the FTIR were reduced to 10 cm^{-1} intervals by averaging. Preliminary analysis to the data sets found that averaging to 10 cm^{-1} intervals yielded similar model coefficients and was necessary to reduce the data set to a manageable size by the SAS (2010) software. After reduction of absorbance data by averaging, the new spectral matrix consisted of 98 rows (number of samples) and 336 lines (mean absorbance for every 10 cm^{-1} interval between 4000 to 650 cm^{-1})

For MLR, the procedure PROC REG was used to regress the wavenumbers against density. The following model form was chosen for regression:

$$\text{Oven dry density} = \beta_0 + \beta_1 W_1 + \dots + \beta_i W_i + \varepsilon \quad (1)$$

where W_i represents the absorbance at the i th wavenumber with a maximum $i=11$, B_0 represents the intercept, B_i represents the coefficient, and ε represents the error.

The wavenumbers used for MLR model building were based on published band assignments in the mid-infrared region for chemical bonds that can exist in wood (Esmeraldo et al. 2010; Jonoobi et al. 2009; Jungnikl et al. 2008; Muller et al. 2009;

Pandey 1999; Qu et al. 2010; Rana et al. 2008; Singha and Rana 2010). These band assignments include: 1738 cm^{-1} C=O stretch in hemicelluloses, 1650 cm^{-1} C-O stretch in lignin, 1504 and 1600 cm^{-1} C=C stretching vibration in lignin, 1456 cm^{-1} asymmetric bending in CH_3 of lignin, 1425 cm^{-1} C-H deformation in lignin and carbohydrates, 1375 cm^{-1} , C-H deformation in cellulose and hemicelluloses, 1152 cm^{-1} C-O-C vibration in cellulose and hemicelluloses, 1104 cm^{-1} O-H association with cellulose and hemicelluloses, 1048 cm^{-1} C-O stretching vibration in cellulose and hemicelluloses, and 898 cm^{-1} C-H deformation in cellulose.

The predictive model that fit closest to the experimental data was chosen using the following statistical diagnostics: root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP), R^2 , adjusted R^2 , and variance inflation factor (VIF) (Neter et al. 1990). The lower the values of RMSEC, RMSEP, and VIF, and the higher the values of R^2 , the better the fit of the model to experimental data. For model validation, the predicted sum of squares (PRESS) was computed and converted into RMSEP (Casal et al. 1996). The PRESS procedure used a leave one out strategy for each i^{th} data point and then estimated the sum of squares error across $n-1$ iterations. The stepwise selection procedure was utilized to determine which independent wavenumbers were important in predicting density. The stepwise selection procedure was compared to other selection procedures such as: Akaike's information criteria, Bayesian information criteria, and Mallows' C_p statistic to protect against model overfit (Akaike 1974; Schwarz 1978; Neter et al. 1990). The default selection criteria for multiple variable models was typically $\alpha = 0.15$ (software default), although for MLR we had to adjust the selection criteria, within the stepwise procedure, to $\alpha = 0.01$ level to eliminate significant VIF problems. MLR on specific wavenumbers was deemed to yield interpretable models if the VIF factors were less than 10. A low VIF was an indication that multicollinearity between the independent wavenumbers were not influencing the regression coefficients and thus interpretation of the slopes could be possible. It should be noted that changing the α to 0.01 penalized the modeler during model building because it increased the chance of under-fitting the data.

All wavenumbers (4000 to 650 cm^{-1}) were used in PCR model building. Each PC was the sum of the linear combination (wavenumber*corresponding eigenvector), resulting in 336 coefficients/weights that were then utilized to compute the PC_i . The principal components from the spectra were determined through PROC PRINCOMP, which is a standard procedure that by default does not rotate the factors. Then PROC REG was utilized on the principal components to develop the PCR model. The same procedures used for MLR were then used for PCR. The model for PCR took on the following form:

$$\text{Oven dry density} = \beta_0 + \beta_1 PC_1 + \dots + \beta_i PC_i + \varepsilon \quad (2)$$

where PC_i represents the i^{th} principal component.

The coefficients (eigenvectors) of the PCR model obtained from the stepwise selection procedure were significant with 0.15 confidence level, which was the default level in SAS. However, a stricter criterion was utilized for selection of the top three principal components in which loading interpretation was performed to determine which

bands where important ($p\text{-value} \leq 0.0001$). The stepwise selection procedure was compared to other selection procedures such as: Akaike's information criteria, Bayesian information criteria, and Mallow's Cp statistic to protect against model overfit.

RESULTS AND DISCUSSION

Comparison of MLR to PCR

Figure 1 shows the relationship between RMSEP and the number of principal components utilized in the model. The PCR + 1st derivative pretreatment performed the best but with the tradeoff of requiring many factors (10). The final values of RMSEP and other statistical parameters for each model and pretreatment combination can be seen in Table 1. The MLR (with no pretreatment) performed the worst (in RMSEP) with a RMSEP value that was 35% higher than the best performing PCR + 1st derivative model. However, the MLR (with no pretreatment) did utilize a lower number of factors than the PCR + 1st derivative pretreatment, which makes it difficult to make a direct comparison. Had the same number of factors been utilized, similar RMSEP were possible, but then there would be an increased risk of overfit.

Similar rankings for RMSEC and RMSEP were obtained after final model selection (Table 1). Based on the values of RMSEC and RMSEP, PCR + 1st derivative was selected as the best model for in-depth interpretation and development. However, MLR was further assessed due to the lack of availability of PCR regression to some users.

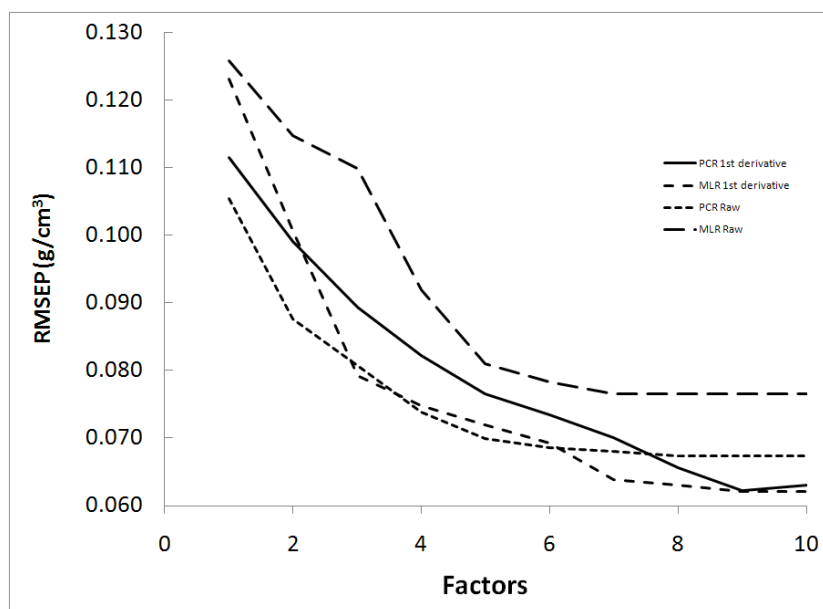


Fig. 1. Model selection based on root mean square error of calibration (RMSEP) versus the number of factors (principal components). Multiple linear regression (MLR) and principal components regression (PCR) was compared with no processing and after taking the 1st derivative.

Figure 2a demonstrates the predictive capability of MLR and PCR when no pretreatment was applied. The MLR model exhibited a little more error than PCR for the raw data, as seen by the increase in scatter around the 1:1 line and the increased RMSEP in Table 1. But after applying the 1st derivative, both MLR and PCR exhibit similar scatter (Fig. 2b) as indicated by the RMSEP in Table 1. The competitiveness of MLR to PCR in prediction, after applying the 1st derivative, was encouraging for situations where PCR is not available. But if interpretation of the trends between dependent and independent variables are important, then PCR is still superior due to the high VIF factors that can occur with MLR. A VIF factor greater than 10 is commonly used as the

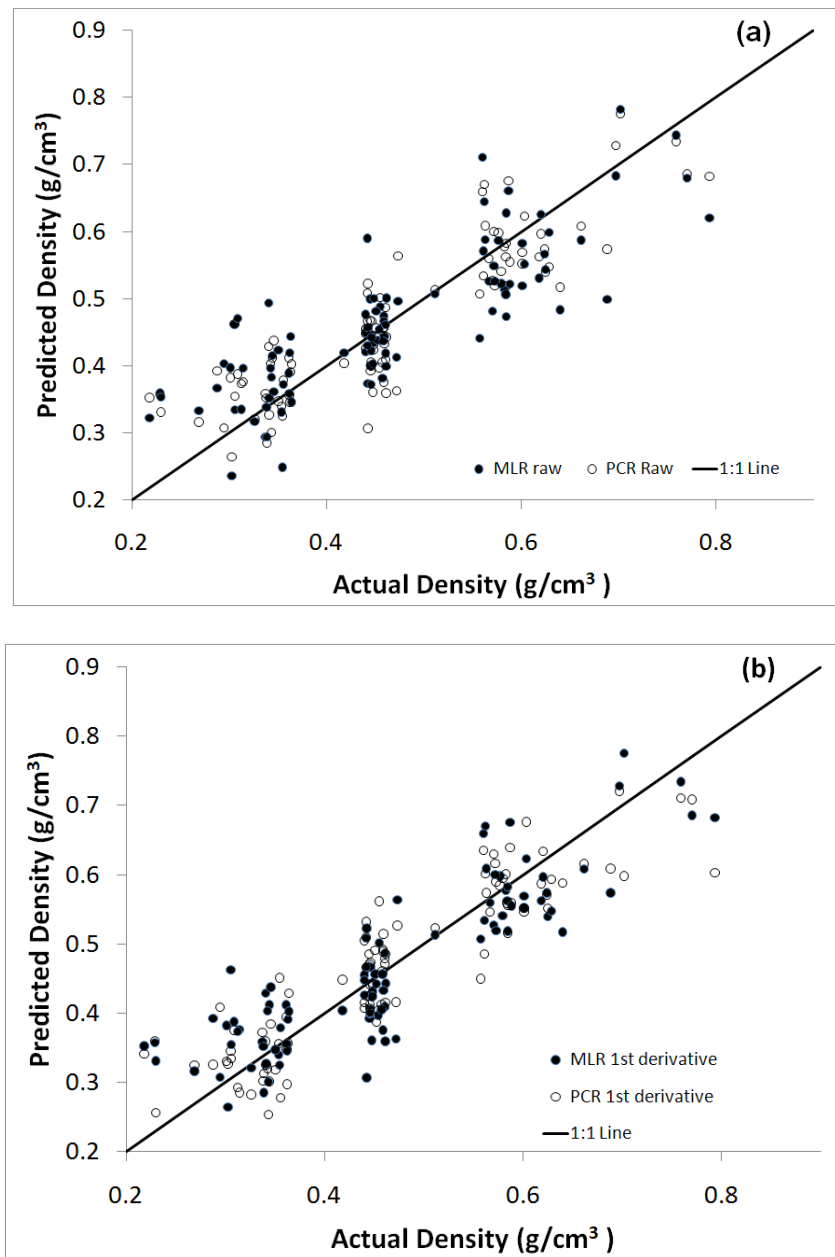


Fig. 2. Density estimated by MLR and PCR versus actual density for (a) raw/no pretreatment spectra ($n=98$) and (b) 1st derivative spectra ($n=98$)

threshold to determine if the covariance between independent variables is too high (Neter et al. 1990). A high VIF means that the independent variables interfere with each other to such degree that interpretation of model coefficients is not possible. The VIF factor for MLR on 1st derivative spectra was 8.88 and was deemed acceptable for interpretation of the coefficients. However, it should be noted that the alpha had to be restricted to 0.01 before acceptable VIF's were found.

Table 1. Model Validation in which the Best Model was Selected Based on the Lowest RMSEP. The R^2 , Adj- R^2 , and VIF were based on the calibration model.

Model + Pretreatment	Factors	RMSEC (g/cm ³)	RMSEP (g/cm ³)	R^2	Adj- R^2	Maximum VIF
PCR + No Pretreatment	8	0.0646	0.0708	0.77	0.75	0
MLR + No Pretreatment	6	0.0742	0.0812	0.69	0.67	1028
PCR + 1 st Derivative	10	0.0581	0.0654	0.82	0.80	0
MLR + 1 st Derivative	7	0.0608	0.0666	0.79	0.78	8.88

Several studies have been carried out that compare NIR multivariate modeling techniques, while a limited amount of studies have been reported for FTIR. For example, MLR has proven to be competitive with PCR for NIR spectroscopy in predicting strength, stiffness, and density of wood (Via et al. 2003), but in a later study it was found that PCR was more robust under extrapolation conditions (Via et al. 2005b). In both studies, the RMSEC values for density were found to be between 0.0485 to 0.0510 g/cm³ for MLR and PCR regression. Higher values of RMSEC (0.0581 to 0.0742 g/cm³) were obtained for the FTIR based models in this study. Ruiz et al. (2005) explained the lower performance of mid-IR spectroscopy in predicting density. They compared the ability of NIR and FTIR to predict solid wood density of *Eucalyptus globulus*. An R^2 value of 0.94 was obtained when predicting density with NIR, but this value dropped to 0.84 when predicting density with FTIR. This drop was attributed to a better signal to noise ratio for the NIR equipment. It should be mentioned that a comparable R^2 value of 0.82 was obtained in this study even though the samples analyzed by Ruiz et al. (2005) were ground and sieved for tighter laboratory control, whereas the samples used in this study were solid and lacked control of surface orientation, which resulted in an increase in the number of factors necessary for prediction. The result obtained from this study is important because this study provides a method to predict and interpret density from PCR and MLR models for samples typical of a manufacturing process. Therefore, despite the random orientation of the tracheids within the samples, manufacturers may be able to monitor key functional groups within the process through control charting techniques and then refer to the model to understand which functional group is responsible for shifts in feedstock density.

The density in a sample is highly dependent on the percentage of earlywood (or latewood) in a sample. The earlywood density in this study typically fell near 0.3 g/cm³, while samples with all latewood had a density around 0.75 g/cm³. This study found the 1060 cm⁻¹ wavenumber to be important in distinguishing latewood from earlywood due to the C-O deformation in carbohydrates (Kotilainen et al. 2000). Other researchers who

have used FTIR to predict density found the nearby 1065 cm^{-1} wavenumber to be critical in predicting density due to lignin-associated structures (Nuopponen et al. 2004, 2006). It should be mentioned that the other researchers that have used FTIR to predict wood density carried out their studies on ground wood samples (e.g. Nuopponen et al. 2004; Ruiz et al. 2005; Meder et al. 1999). However grinding of the sample did not necessarily improve the predictability of the models. For example Meder et al. (1999) sieved the ground wood particles in order to obtain a tightly distributed particle size. A calibration R^2 of 0.87 with 4 factors was obtained. However, during the validation stage, the R^2 dropped to 0.60. In this study, the adjusted R^2 after validation only dropped by an average of 0.02 for all four models.

Another important finding of this study was that MLR could be a viable alternative to PCR when modeling biomass density from FTIR spectra. If one wants to directly use predetermined wavenumbers for modeling, it is likely that high covariances between wavenumbers in percent transmittance will occur. After taking the 1st derivative, we did find significant reductions in VIF but still, many models exhibited unacceptable VIF numbers (>10) when the typical stepwise selection method was used. Also, we had to reduce the alpha to 0.01 to obtain models with acceptable VIF, which in turn limits the number of wavenumbers available for modeling. As wavenumbers become further apart (often $> 500\text{ cm}^{-1}$) range, they were less likely to interfere with one another during modeling. This resulted in forbidding the use of several wavenumbers that were closer than 500 cm^{-1} apart. Thus, it may be more difficult to utilize exact wavenumbers of interest during calibration and instead one is at the leniency of which wavenumbers are less correlated. Nevertheless, these results suggest that MLR can be utilized for modeling density for interpretation purposes, but care needs to be taken during model development. If one is willing to sacrifice some interpretation, such as that which might be necessary for biomass manufacturing, MLR becomes competitive to PCR in performance. MLR may thus be useful for biomass applications where calibration equations can easily be programmed into common spreadsheets. On the other hand, if manufacturers can afford software and control charting tools that can utilize PCR, then interpretation and control charting of key functional groups may be an advantage.

Further PCR Development and Interpretation

After generation and comparison of many PCR models, the best PCR model (Table 1) was chosen for both prediction and interpretation purposes. This PCR model required the 1st derivative pretreatment and 10 factors plus an intercept term. Table 2 gives the details of this model, including coefficients and level of significance associated with each factor. PC 2, 3, and 5 were determined to be important in predicting density based on the t-statistic and total variance accounted for by each PC (not shown). As such, the loadings across all wavenumbers was plotted for PC2 (Fig. 3a), PC3 (Fig. 3b), and PC5 (Fig. 3c). Significant global and local peaks from these three graphs are also listed in Table 3 with their band assignments. Consequently, most of the wavenumbers that were found to be important through PCR (Table 3 and Fig. 3a, b, and c) were also used for MLR, where the PC's were selected *a priori* to the performed analysis (see methods section). All of the important wavenumbers found important through PCR could be traced back to cellulose, hemicellulose, and lignin polymers.

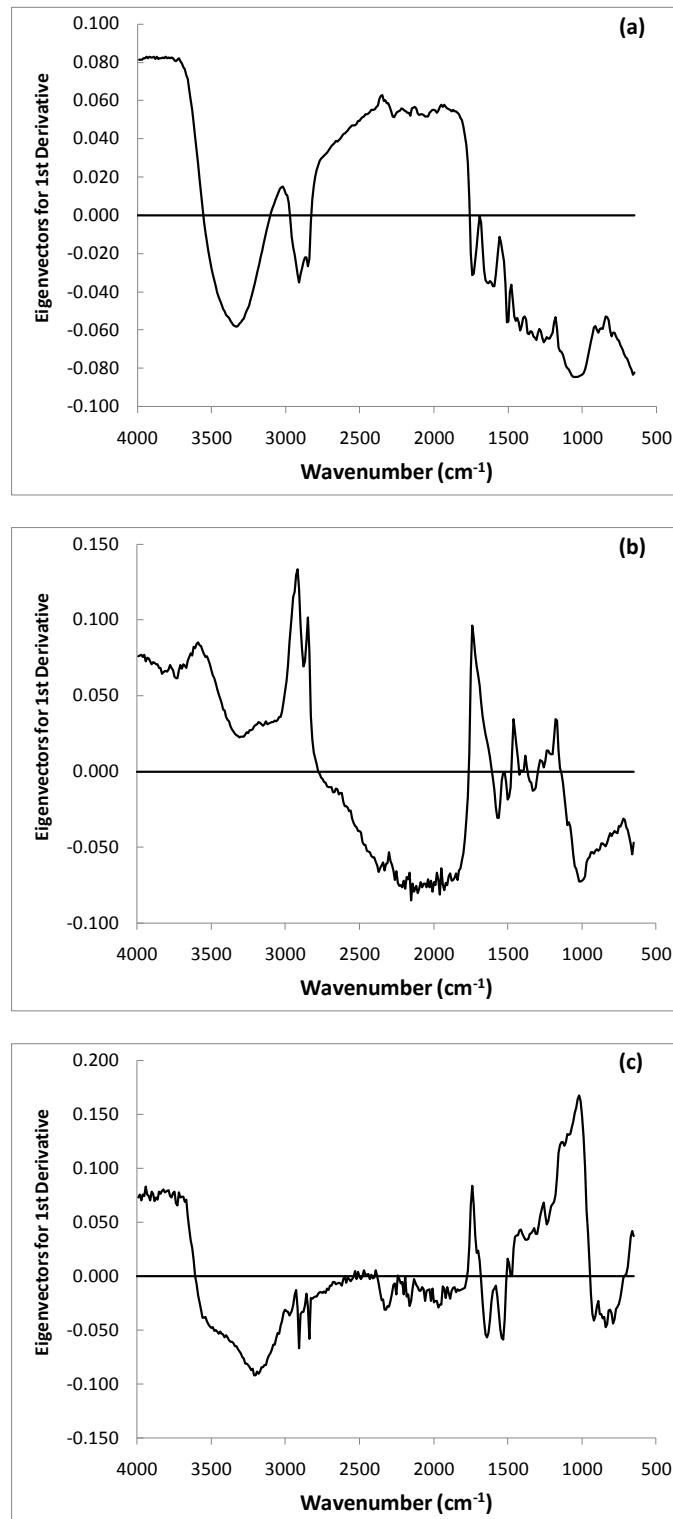


Fig. 3. Eigenvector loading on 1st derivative spectra for (a) principal component 2, (b) principal component 3, and (c) principal component 5.

Of particular interest were those wavenumbers that showed up as being important in the loadings for 2 independent factors (Table 3), implying that these band assignments may be of higher importance in predicting density. For example, the 1048 cm^{-1} was attributable to the C-O stretching vibration in hemicellulose and cellulose. The 1736 cm^{-1} was attributable to the C=O stretching vibration in hemicellulose. The sensitivity of FTIR to hemicellulose signals is beneficial, particularly from a modeling standpoint. FTIR appears to better detect hemicellulose signals than NIR. For example, one study found NIR to be more sensitive to lignin and cellulose than hemicellulose (Via et al. 2009).

Table 2. Best PCR Calibration Model to Predict Density

Parameter	Estimate	Standard Error	t-statistic	p-value
β_0	0.465	0.00587	79.27	<0.0001
β_1	0.0039	0.00053	7.26	<0.0001
β_2	-0.0052	0.00091	-5.73	<0.0001
β_3	-0.0167	0.00144	-11.6	<0.0001
β_4	-0.0069	0.00177	-3.91	= 0.0002
β_5	-0.0075	0.00190	-3.96	= 0.0001
β_6	0.00726	0.00196	3.71	0.0004
β_7	0.01822	0.00205	8.90	<0.0001
β_8	0.00888	0.00223	3.98	0.0001
β_9	0.00453	0.00237	1.91	0.05
β_{10}	0.01301	0.00246	5.29	<0.0001

Table 3 demonstrates the important wavenumbers in predicting density. When compared to a similar study on ground wood (Nuopponen et al. 2006), over 50% of the wavenumbers that were significant in that study matched the wavenumbers deemed important through PCR analysis in this study. In addition, it was also found that four wavenumbers were important in two factors. This indicates that these are the four most important wavenumbers for predicting density of woody biomass (Table 3). Two of these four wavenumbers were also highlighted by Nuopponen (2006) as being important in predicting density from ground wood. These wavenumbers found by Nuopponen (2006) were the C-O stretching vibration in hemicellulose and cellulose (1048-1050 cm^{-1}) and the C=O stretching vibration (1736 cm^{-1}) in hemicelluloses.

In another study predicting density of *E. globulus* for ground wood, 2 more significant wavenumbers arose as being important, which also agreed with this study (Table 3) (Freer et al. 2003). The 1504 cm^{-1} wavenumber due to the C=C lignin bond was critical in both studies. Likewise, the 1736 cm^{-1} wavenumber was once again important in demonstrating the sensitivity of FTIR to hemicellulose signals. The sensitivity of 1736 cm^{-1} for this study and other gymnosperm and angiosperm woody plants demonstrates the utility of utilizing similar functional groups across plant species (Freer et al. 2003; Nuopponen et al 2006). This is especially useful, since the hemicellulose between softwoods and hardwoods differ in concentration and distribution of branched heteropolysaccharides.

Interpretation of important wavenumbers on density can be important because it enables the quality control manager to better understand the quality of the wood and the interrelationships between functional groups and density. This can be important to product quality, and likely different genus and/or species will have slightly different coefficients and overall models. But future studies would be necessary to confirm this possibility. Likewise, in pine, juvenile wood concentration will be an important factor for biomass processing due to the recent shift from mature to juvenile wood in production. Increased juvenile wood in pine will have less cellulose and more lignin, which will have a negative impact on density and perhaps product quality. This consequence may be detectable through the monitoring of key principal components through T^2 Hotelling. T^2 Hotelling is a multivariate quality control technique that can handle simultaneous shifts in multiple PC's attributable to changes in the concentration of underlying functional groups (Marengo et al. 2003).

Table 3. Band Assignment for Important Wavenumbers extracted from Statistically Significant Principal Components (p-value \leq 0.0001) through Regression Analysis on 1st Derivative Spectra

<u>Wavenumber</u>			
2 nd PC	3 rd PC	5 th PC	Band Origin
1048	1048		C-O stretching vibration in hemicellulose and cellulose
1504			C=C stretching vibration in lignin
2935	2935		C-H stretching vibration
	730		No sources available
	1160		C-O-C bridge stretching in cellulose
	1320		C-O vibration in syringyl ring
	1456		Asymmetric CH ₃ bending in lignin
	1736	1736	C=O stretching vibration in hemicellulose
	2842	2842	C-H stretching vibration
		896	C-H deformation vibration cellulose
		1030	Aromatic C-H in plane deformation
		1648	C-O stretching vibration in lignin
		3230	No sources available

Effects of Earlywood and Latewood on Spectra

Figure 4a demonstrates the effect of earlywood and latewood on the spectra. There were clear shifts between 4000 to 3660 cm^{-1} , 3160 to 3030 cm^{-1} , 2810 to 1820 cm^{-1} , and 1030 cm^{-1} . Given the lack of control over surface presentation, it was apparently difficult to partition out peaks due to specific functional groups (Fig. 4a).

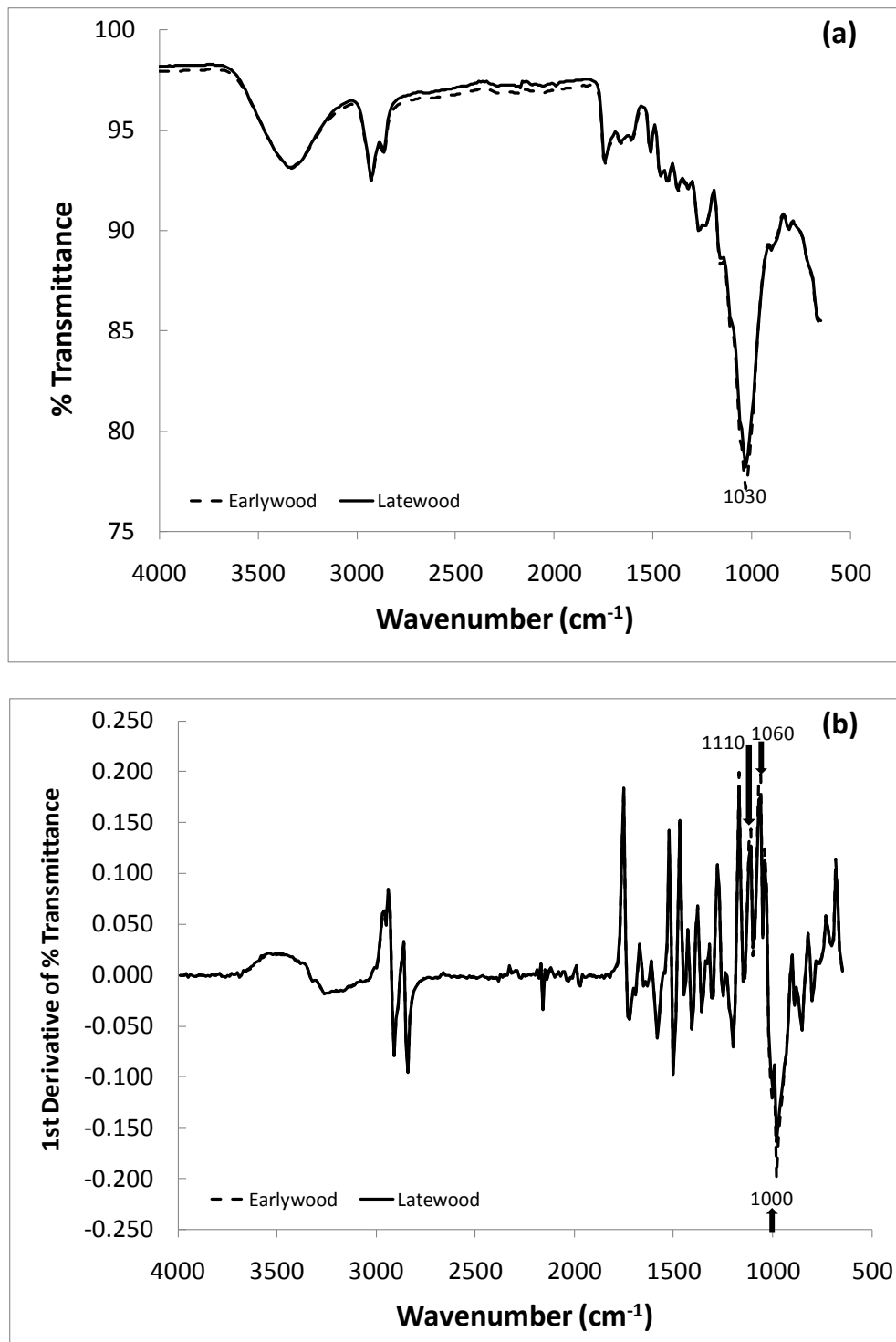


Fig. 4. FTIR spectra for earlywood (n=98) and latewood (n=98) for (a) unprocessed raw spectra and (b) after the 1st derivative pretreatment

Further data analysis that involved taking the 1st derivative resulted in three peaks that were distinctively different for earlywood and latewood (Fig. 4b). The peak at 1000 and 1060 cm^{-1} differed between earlywood and latewood due to the C-O bond present in carbohydrates (Goncalves et al. 1998; Kotilainen et al. 2000). The peak at 1068 cm^{-1} was found to be due to arabinose concentration (Hori and Sugiyama 2003), while the peak at 1110 cm^{-1} was due to the O-H association of cellulose.

CONCLUSIONS

1. Capable models were developed from FTIR spectra to predict the density of solid wood biomass. The biomass possessed a random tracheid orientation with respect to the FTIR beam. This random surface axis is considered more representative of an industrial feedstock than similar studies in the literature where the wood was ground or the tracheid axis was controlled during milling of the samples.
2. Multiple linear regression could be used to build models, using an approach that is easier to utilize during programming or in common spreadsheet software; however, if interpretation of the coefficients are of interest, then the alpha had to be set to 0.01 and the first derivative was necessary to remove the covariance between adjacent wavenumbers.
3. Principal component 3 (after 1st derivative processing) was the most statistically significant in prediction of biomass density based on t-value and p-value results. Of the loadings in principal component 3, the top 4 bands were identified at 1048, 1736, 2842, and 2935 cm^{-1} . Respectively, these were attributable to the C-O stretching vibration in cellulose and hemicellulose, the C=O stretching vibration in hemicellulose, and the C-H stretching vibration (2842 and 2935 cm^{-1}) of cellulose. The asymmetric CH_3 bending in lignin at 1456 cm^{-1} was also influential.

ACKNOWLEDGMENTS

This research was in part alignment with 5 year goals set forth in Hatch – Project Number ALA031-1-09020 and Mcintire-Stennis - Project Number ALAZ00051. A special thanks goes to Marjorie Gentry and Ginger Phillabaum for their assistance in ensuring these two proposals met agency requirements.

REFERENCES CITED

- Akaike H. (1974). "A new look at the statistical model identification," *IEEE Trans Automatic Control* 19(6), 716-723.
- Casal, V., MartinAlvarez, P. J., and Herraiz, T. (1996). "Comparative prediction of the retention behaviour of small peptides in several reversed-phase high-performance liquid chromatography columns by using partial least squares and multiple linear regression," *Analytica Chimica Acta* 326(1-3), 77-84.

- Defo, M., Taylor, A. M., and Bond, B. (2007). "Determination of moisture content and density of fresh-sawn red oak lumber by near infrared spectroscopy," *Forest Prod. J.* 57(5), 68-72.
- Esmeraldo, M. A., Barreto, A. C. H., Freitas, J. E. B., Fachine, P. B. A., Sombra, A. S. B., Corradini, E., Mele, G., Maffezzoli, A., and Mazzetto, S. E. (2010). "Dwarf-green coconut fibers: A versatile natural renewable raw bioresource. Treatment, morphology, and physicochemical properties," *BioResources* (<http://www.bioresources.com>), 5(4), 2478-2501.
- Freer, J., Ruiz, J., Peredo, M. A., Rodriguez, J., and Baeza, J. (2003). "Estimating the density and pulping yield of e-globulus wood by drift-mir spectroscopy and principal components regression (pcr)," *J. Chil. Chem. Soc.* 48(3), 19-22.
- Geladi, P., Sethson, B., Nystrom, J., Lillhonga, T., Lestander, T., and Burger, J. (2004). "Chemometrics in spectroscopy - Part 2. Examples," *Spectrochim Acta B* 59(9), 1347-1357.
- Goncalves, A. R., Esposito, E., and Benar, P. (1998). "Evaluation of *Panus tigrinus* in the delignification of sugarcane bagasse by ftir-pca and pulp properties," *J. Biotechnol.* 66(2-3), 177-185.
- Hodgson, E. M., Lister, S. J., Bridgwater, A. V., Clifton-Brown, J., and Donnison, I. S. (2010). "Genotypic and environmentally derived variation in the cell wall composition of miscanthus in relation to its use as a biomass feedstock," *Biomass Bioenerg* 34(5), 652-660.
- Hori, R., and Sugiyama, J. (2003). "A combined ft-ir microscopy and principal component analysis on softwood cell walls," *Carbohydr Polym* 52(4), 449-453.
- Jonoobi, M., Harun, J., Shakeri, A., Misra, M., and Oksman, K. (2009). "Chemical composition, crystallinity, and thermal degradation of bleached and unbleached kenaf bast (*hibiscus cannabinus*) pulp and nanofibers," *BioResources* (<http://www.bioresources.com>), 4(2), 626-639.
- Jungnikl, K., Paris, O., Fratzl, P., and Burgert, I. (2008). "The implication of chemical extraction treatments on the cell wall nanostructure of softwood," *Cellulose* 15(3), 407-418.
- Kauper, P., and Ferri, D. (2004). "From production to product: Part 2, FT-IR spectroscopy as a tool for quality control in the preparation of alkaline solutions of bagasse lignin," *Ind. Crop. Prod.* 20(2), 159-167.
- Kotilainen, R. A., Toivanen, T. J., and Alen, R. J. (2000). "FTIR monitoring of chemical changes in softwood during heating," *J. Wood Chem. Technol.* 20(3), 307-320.
- Liu, L., Ye, X. P., Womac, A. R., and Sokhansanj, S. (2010). "Variability of biomass chemical composition and rapid analysis using FT-NIR techniques," *Carbohydr. Polym.* 81(4), 820-829.
- Marengo, E., Robotti, E., Liparota, M. C., and Gennaro, M. C. (2003). "A method for monitoring the surface conservation of wooden objects by Raman spectroscopy and multivariate control charts," *Anal. Chem.* 75(20), 5567-5574.
- Meder, R., Gallagher, S., Mackie, K. L., Bohler, H., and Meglen, R. R. (1999). "Rapid determination of the chemical composition and density of *Pinus radiata* by PLS modelling of transmission and diffuse reflectance FTIR spectra," *Holzforschung* 53(3), 261-266.

- Muller, G., Schopper, C., Vos, H., Kharazipour, A., and Polle, A. (2009). "FTIR-ATR spectroscopic analyses of changes in wood properties during particle- and fibreboard production of hard- and softwood trees," *BioResources* (<http://www.bioresources.com>), 4(1), 49-71.
- Neter, J., Wasserman, W., and Kutner, M. H. (1990). *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs*, Irwin, Homewood, IL.
- Nkansah, K., Dawson-Andoh, B., and Slahor, J. (2010). "Rapid characterization of biomass using near infrared spectroscopy coupled with multivariate data analysis: Part 1. Yellow-poplar (*Liriodendron tulipifera* L.)," *Bioresource Technol.* 101(12), 4570-4576.
- Nuopponen, M., Wikberg, H., Vuorinen, T., Maunu, S. L., Jamsa, S., and Viitaniemi, P. (2004). "Heat-treated softwood exposed to weathering," *J. Appl. Polym. Sci.* 91(4), 2128-2134.
- Nuopponen, M. H., Birch, G. M., Sykes, R. J., Lee, S. J., and Stewart, D. (2006). "Estimation of wood density and chemical composition by means of diffuse reflectance mid-infrared fourier transform (drift-mir) spectroscopy," *J. Agr. Food Chem.* 54(1), 34-40.
- Pandey, K. K. (1999). "A study of chemical structure of soft and hardwood and wood polymers by ftir spectroscopy," *Journal of Applied Polymer Science* 71(12), 1969-1975.
- Qu, P., Tang, H. W., Gao, Y. A., Zhang, L. P., and Wang, S. Q. (2010). "Polyethersulfone composite membrane blended with cellulose fibrils," *BioResources* (<http://www.bioresources.com>), 5(4), 2323-2336.
- Rana, R., Mueller, G., Naumann, A., and Polle, A. (2008). "FTIR spectroscopy in combination with principal component analysis or cluster analysis as a tool to distinguish beech (*fagus sylvatica* l.) trees grown at different sites," *Holzforschung* 62(5), 530-538.
- Ruiz, J., Rodriguez, J., Baeza, J., and Freer, J. (2005). "Estimating density and pulping yield of e-globulus wood: Comparison of near-infrared (NIR) and mid-infrared (MIR)," *J. Chil. Chem. Soc.* 50(3), 565-568.
- Schimleck, L. R., Evans, R., and Ilic, J. (2001). "Estimation of *Eucalyptus delegatensis* wood properties by near infrared spectroscopy," *Can. J. Forest Res.* 31(10), 1671-1675.
- Schimleck, L. R., Sturzenbecher, R., Mora, C., Jones, P. D., and Daniels, R. F. (2005). "Comparison of *Pinus taeda* L. Wood property calibrations based on NIR spectra from the radial-longitudinal and radial-transverse faces of wooden strips," *Holzforschung* 59(2), 214-218.
- Schwarz, G. (1978). "Estimating the dimension of a model," *Ann. Stat.* 6, 461-464.
- Singha, A. S., and Rana, R. K. (2010). "Effect of pressure induced graft copolymerization on the physico-chemical properties of bio-fibers," *BioResources* (<http://www.bioresources.com>), 5(2), 1055-1073.
- Sohn, M., Himmelsbach, D. S., Barton, F. E., Griffey, C. A., Brooks, W., and Hicks, K. B. (2007). "Near-infrared analysis of ground barley for use as a feedstock for fuel ethanol production," *Appl. Spectrosc.* 61(11), 1178-1183.

- Tsuchikawa, S., and Tsutsumi, S. (1999). "Directional characteristics model and light-path model for biological material having cellular structure," *Appl. Spectrosc.* 53(2), 233-240.
- Via, B. K., Shupe, T. F., Groom, L. H., Stine, M., and So, C. L. (2003). "Multivariate modelling of density, strength and stiffness from near infrared spectra for mature, juvenile and pith wood of longleaf pine (*pinus palustris*)," *J. Near Infrared Spec.* 11(5), 365-378.
- Via, B. K., So, C. L., Shupe, T. F., Stine, M., and Groom, L. H. (2005a). "Ability of near infrared spectroscopy to monitor air-dry density distribution and variation of wood," *Wood Fiber Sci.* 37(3), 394-402.
- Via, B. K., So, C. L., Shupe, T. F., Eckhardt, L. G., Stine, M., and Groom, L. H. (2005b). "Prediction of wood mechanical and chemical properties in the presence and absence of blue stain using two near infrared instruments," *J. Near Infrared Spec.* 13(4), 201-212.
- Via, B. K., So, C. L., Shupe, T. F., Groom, L. H., and Wikaira, J. (2009). "Mechanical response of longleaf pine to variation in microfibril angle, chemistry associated wavelengths, density, and radial position," *Compos. Part A-Appl. S.* 40(1), 60-66.
- Yao, S., Wu, G. F., Xing, M., Zhou, S. K., and Pu, J. W. (2010). "Determination of lignin content in *Acacia* spp. Using near-infrared reflectance spectroscopy," *BioResources* (<http://www.bioresources.com>), 5(2), 556-562.

Article submitted: November 26, 2010; Peer review completed: December 23, 2010;
Revised article accepted: January 17, 2011; Published: January 22, 2011.