Multivariate modelling of density, strength and stiffness from near infrared spectra for mature, juvenile and pith wood of longleaf pine (*Pinus palustris*)

Brian K. Via,[a] Todd F. Shupe,[a] Leslie H. Groom,[b] Michael Stine[a] and Chi-Leung So[b]

[a]*School of Renewable Natural Resources, Louisiana State University Agricultural Center, Baton Rouge, Louisiana, USA*

[b]*USDA Forest Service, Southern Research Station, Pineville, Louisiana, USA*

In manufacturing, monitoring the mechanical properties of wood with near infrared spectroscopy (NIR) is an attractive alternative to more conventional methods. However, no attention has been given to see if models differ between juvenile and mature wood. Additionally, it would be convenient if multiple linear regression (MLR) could perform well in the place of more complicated multivariate models. Therefore, the purpose of this paper was to model the strength, stiffness and density of mature and juvenile longleaf pine to NIR spectra with MLR and principal component regression (PCR). MLR performed almost as well as PCR when predicting density, modulus of rupture (MOR) and modulus of elasticity (MOE). Choosing wavelengths associated with wood chemistry and developing principal components gave better predictive models (PCR_2) than when all NIR wavelengths were used (PCR_1). Models developed from mature wood did not predict wood properties from juvenile wood adequately, suggesting that separate models are needed. However, for density prediction, the area under the spectral curve appeared to be insensitive to mature and juvenile wood differences. Five of the six wavelengths associated with MOE were also associated with MOR, perhaps accounting for how MOE and MOR might be related. For pith wood, MOE and MOR were poorly related to NIR spectra, while density was strongly correlated. This inability to predict mechanical properties in the pith-wood zone warrants attention for those manufacturers interested in using near infrared to stress rate lumber within a mill.

Keywords: multivariate, principal component, multiple linear, regression, density, modulus, NIR, spectroscopy, wood, lumber

Introduction

Softwood lumber is common in building construction where availability and cost are balanced with acceptable strength and stiffness. The portion of the tree that, typically, yields unacceptable strength and stiffness is juvenile wood. Juvenile wood comes from xylem wood produced several years after apical meristem production. As the wood matures, the mechanical properties increase. As a result, development of rapid assessment tools to monitor the mechanical properties of wood is

needed to differentiate between juvenile and mature wood. Proper models would enhance quality control, wood utilisation 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 microfibril angle, a property that impacts modulus of elasticity (MOE).[1] An example of a more direct relationship would be if wood chemistry and MOE were related. For example, for single wood fibres higher in cellulose, a higher response in MOE is present than single wood fibres higher in lignin.[2] Perhaps this plays a role for an aggregate collection of fibres connected by lignin (solid wood).

The most important macro characteristic to contribute 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 shifts upward as density increases.[3,4] Additionally, there may be more subtle variables to influence density that could, in turn, influence spectral 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}.[5,6] Hemicellulose has a similar density to cellulose but is, perhaps, significantly altered during the extraction process, making a quantitative assessment difficult.[5] 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 1. By associated wavelengths, we mean the wavelengths where absorbance response has been shown to relate to some chemical constituent in wood. Because the kappa number is a direct function of lignin content, it was included in the table and was considered as lignin-associated wavelengths.

Some work has already shown NIR to estimate wood density, MOE and MOR successfully.[3,4,15-18] However, it is not well understood if different models are needed for different portions of the tree. Schimleck *et al.*[15] did use soft independent modelling 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.

Since the choice of models can play a significant role in prediction efficiency, it is important to try different modelling 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 multi-collinearity 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[19] 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.[20]

In this study, ten longleaf pine trees of varying diameter and height were randomly chosen and felled. Density, MOE and modulus of rupture (MOR) were measured by conventional methods from five to seven heights at 4.57 M intervals. 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). The ability to model density from the area under the spectra and from individual wood chemistry associated wavelengths was also investigated.

Material and methods

Sample preparation

Ten longleaf pine trees, 41 years in age, were selected from a plantation on the Harrison Experimental Forest which was owned and maintained by the USDA Forest Service near Saucier, Mississippi (USA). The location was 30.6° north and 89.1° west.

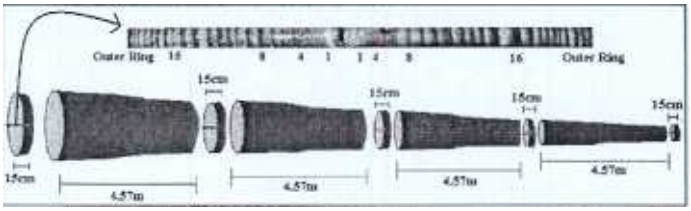

Figure 1. The breakdown of the tree into disks and increment cores.

The understory of the site was free of competition because of periodic prescribed fires by the Forest Service staff. Trees were planted approximately 3.66 M apart from neighbouring trees in an equilateral triangle pattern. In other words, 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 M in height, yielding five to seven bolts, each bolt having an accompanying disk cut from the basal end of the bolt (Figure 1). The bolt was further processed to yield bending specimens 30.48 cm × random width × 1.27 cm, with the ring of interest in the centre of the specimen, while the disk was used for NIR spectra acquisition. Widths were random because ring thickness varied and it was desired to maintain the same ring count per specimen. Spectra was acquired from the radial face from bark to pith to bark strips ripped from the disks.

Mechanical testing

The bending specimens were conditioned in ambient conditions to an equilibrium moisture content of 8.0% with a standard deviation of 1.3%. The load was applied 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 were 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 callipers 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.

Categorisation

Counting from pith to bark, latewood rings 1, 4, 8, 16, 32 and the last ring on the disk were quantified from five to seven disks, dependent on tree height (disks were cut every 4.57 M in height). Typically, for specimens in the last ring, five to ten rings had to be included 'as the last ring' because of much slower growth rates. This was necessary because a cross-section of 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 Rings 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.

NIR spectroscopy

NIR absorbance was obtained using a Nexus 670 FT-IR spectrometer (Thermo Nicolet Instruments, Madison, WI, USA) with wavelengths between 1000 to 2500 nm. Scans were acquired at 1 nm intervals using reflectance spectroscopy. Forty scans were collected and averaged into a single spectrum curve. 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 ± 1 in a laboratory environment with a mean relative humidity of 50%.

To demonstrate the effect of density on the spectra, ten replicates per density group were measured. The density groups were 0.40, 0.50, 0.60 and 0.70 g cc^{-3}. Forty scans were acquired on the north and south side for a total of 80 scans and then averaged together for one spectral curve.

Multivariate analysis

MLR and PCR were performed using Statistical Analysis Software (SAS) in which PROC REG and PROC PRINCOMP were applied.[21] To reduce data set and computation time in SAS, the absorbance values were reduced to 10 nm intervals by averaging. Prior analysis showed nearly identical model coefficients when 10 nm intervals were compared with models developed from spectra of 1 nm intervals. Regression diagnostics were employed to compare the predictive power of different models. One hundred and 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*), predicted sum of squares (*PRESS*), R^2, adjusted R^2 and C_p were employed to compare model predictability and can be reviewed in detail by Neter *et al.*[22] The equation for *PRESS* is

$$PRESS = \sum_i^n (Y_i - Y_{predicted})$$

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

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

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 reression model; therefore, models with a lower C_p were preferred.

In summary, a lower *RMSEV*, *PRESS* and C_p and a higher R^2 and adjusted R^2 suggests 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. Wavelengths for MLR and PCR were chosen from Table 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 1. The most influential wavelengths 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. No spectral pre-treatment, 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), the area (nm*absorbance) under each spectral curve was estimated by Reimann sums. Linear regression and MLR were performed using area as an independent variable. Because the Type II sum of squares is sequential, area was always placed in the density MLR model first, since the baseline shift was the most important parameter in predicting density.

Results and discussion

Density

Figure 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.[4,18] In the lower wavelength region (1000 to 1400 nm), increasing the density from 0.4 to 0.7 g cm^{-3} brought about a 0.1 mean increase in absorbance and can be viewed in Figure 2 ($\alpha = 0.05$). 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 brought about a 0.27 mean increase in absorbance. While there was increased range at the upper end of the spectrum, there was also increased

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

Chemical measurement	Material or solution	Associated wavelengths (nm)	Reference
	Softwood and hardwood kraft pulp	457, 500, 650	3
Kappa number	Eucalyptus grandis kraft pulp	1722, 1778, 1940, 2100, 2139	4
Kappa number	Mixed pine kraft	1445, 1680, 1734, 2100, 2270	4
Kappa number	Mixed pine and eucalyptus kraft	1445, 2100, 2208, 2230, 2270	4
Lignin	Forest foliage 18 hard-wood and softwood species	1438, 1828, 2218, 2386	5,6
Lignin	Solid wood alder, bass, cherry, pine, walnut, birch	1600, 1660, 1505, 1460, 1275, 1230, 870	7
Lignin	Quercus alba foliage	1438, 1708, 2154, 2320	5,6
Cellulose	14 different pine species	1722, 1734, 2230, 2310, 2236	8
Cellulose	Forest foliage 18 hard-wood and softwood species	1766, 1960, 1982, 2140	5,6
Cellulose	Quercus alba foliage	1754, 1898, 2076, 1898	5,6
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	9
Cellulose and hemicellulose	Solid wood alder, bass, cherry, pine, walnut, birch	1425, 1370, 1325, 1160, 1110, 1050, 895, 680	7
	Various foods and plants including barley, cereal, forage, flour, wheat	1218, 1278, 1360, 1436, 1492, 1584, 1728, 1778, 1830, 2110, 2186, 2262, 2314, 2380	9
Hot water extractives	Eucalypt wood	2200, 1700, 1900	10

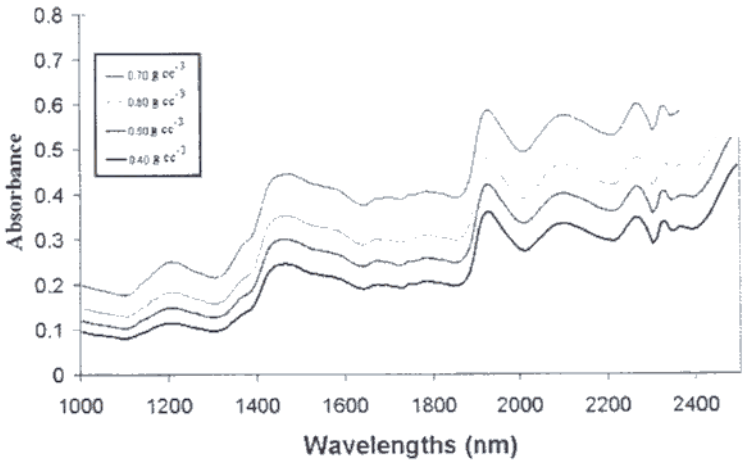

Figure 2. The response of NIR spectra to wood density (g cm^{-3}) categories where $n = 10$.

variance in absorbance per wavelength. When variance in absorbance was plotted at each wavelength for all wood samples, the variation 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 characteristic of density. For example, an increase in wheat kernel size caused an upward baseline shift in absorbance values caused by increased material.[23] 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, the baseline shift with density can be explained by the Beer–Lambert law.[13]

$$A = \underline{M}cd$$

where A is absorbance, M is molar absorptivity, c is molar concentration of absorber and d is sample path-length. The molar absorbtivity 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 path-length does not change, then the concentration of the absorber should increase linearly with concentration or density. This relationship explains the increase in absorbance with density in Figure 2. 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). This deviation from the Beer–Lambert Law at high concentrations is common for various materials.[13]

The area under the curve for the whole NIR region was computed because it would directly relate to a baseline shift in absorbance. The area was regressed against density using least squares linear regression (Figure 3). An R^2 of 0.71 was achieved with balanced residuals and proved to be a viable way to model density.

Aside from macro density variation, there may be additional variation in density due to cell wall chemistry. It was pointed out in the Introduction that cellulose is more dense than lignin. Therefore, from Table 1, the 2230 and 1708 nm wavelengths were chosen specifically for cellulose and lignin, respectively, because this combination exhibited the lowest multi-colinearity as determined by Pearson Correlation Coefficient.[22] Because a baseline shift showed equal effect on the two wavelengths, taking

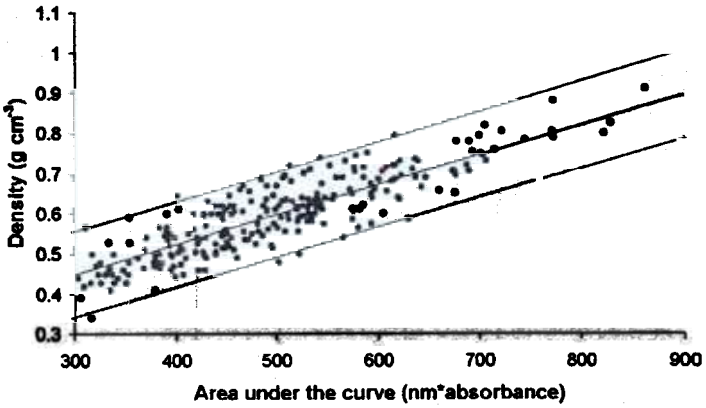

Figure 3. The response of density to the area under the spectral curve with regression and 95% confidence intervals.

the ratio of absorbance at two wavelengths eliminated the variation attributable to the baseline shift. Figure 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.

MOE and MOR response to density

Figure 5 shows the influence of density on MOE for pith, juvenile and mature wood. While not plotted, a similar trend was found for MOR for the three zones. 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 with mature wood (p-value < 0.01). 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.[24,25] Others have reported that microfibril angle was the primary contributor to

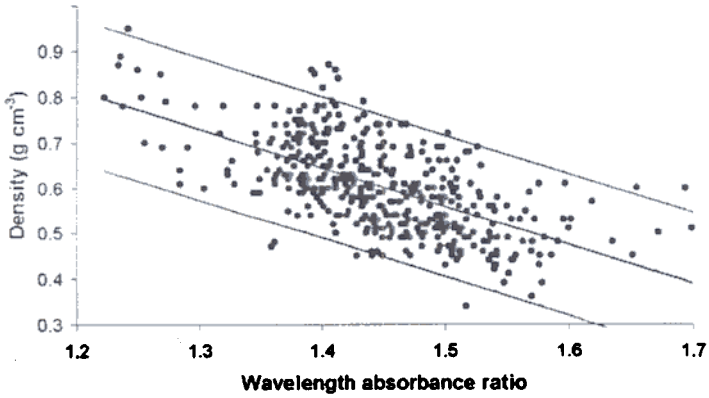

Figure 4. The response of density to the ratio of cellulose (2230 nm) to lignin (1708 nm) wavelengths with regression and 95% confidence intervals.

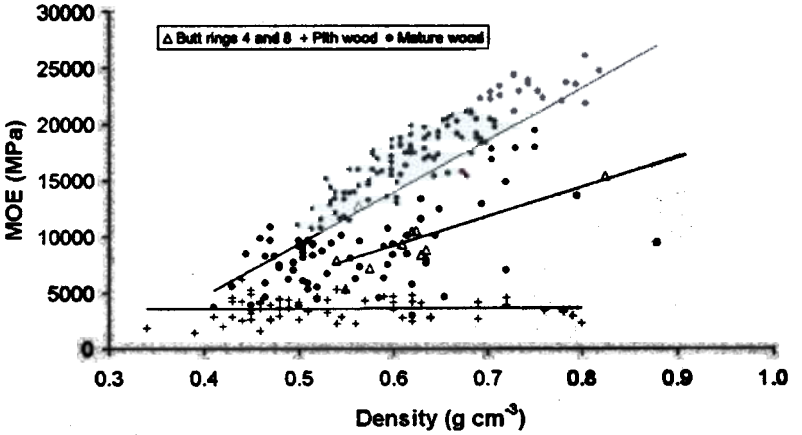

Figure 5. The response of MOE to density for pith, juvenile and mature wood.

MOE in the juvenile wood zone.[26,27] Given that density has a decreased influence on MOE and MOR as one goes from mature, to juvenile, to pith, it might be possible that different models are needed.

Whole tree regression models

For density, MLR performed 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). 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 1).

For the PCR_1 density model, five principal components were significant in predicting density, while for PCR_2, only two principal components were significant and only slightly lower in predictive ability (Table 2). Of all three regression models, $MLR \leq PCR_1 < PCR_2$ for *RMSEC*, *RMSEV* and *PRESS* statistics. 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). This was surprising because density is considered to directly influence MOE. This suggests that chemistry influence on stiffness is important and perhaps underestimated. In support of this finding, it was found that most of the wood chemistry associated wavelengths needed to predict density were needed to predict MOE. Alternatively, and equally noteworthy, was that cellulose wavelengths may be associated with microfibril angle which, in turn, affects MOE, as observed by others.[1,4] 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 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

Table 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 1.

Model type	Dependent variable	Independent variables	R^2	Adj. R^2	*RMSEC* $N = 170$	*RMSEV* $N = 93$	Error ratio	*PRESS*	C_p
LR	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

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 spectral 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*.

Model stability

So far, the predictive ability of PCR_1 and PCR_2 have appeared similar. 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.[22] Table 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 significant variables.

When calibration and validation regression coefficients were compared for PCR_1, different regression coefficients emerged (Table 3). The standard

Table 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)		

errors, for each regression coefficient, was listed to gauge if coefficients differed significantly. 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 earlier, density and MOE showed similar instabilities in regression coefficients for PCR_1.

Modelling pith wood

Only calibration models were developed for pith wood with the objective of diagnosing correlation. Table 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, as visually observed. Excessive extractives will heavily influence the spectral 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 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 microfibril angle to have increased importance within the juvenile wood zone when predicting MOE and MOR.[26,27]

Table 4. A comparison of MLR_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	0.6942	0.6889	0.0629
MLR_2	Density	Area, 2155, 2205	0.8307	0.8216	0.0476
	MOE	1705, 1825	0.0684	0.0357	145168
	MOR	1705, 1825	0.1326	0.1021	1559
PCR_1	Density	6 PC's	0.8681	0.8532	0.0432
	MOE	3 PC's	0.1870	0.1435	136818
	MOR	1 PC	0.0875	0.0717	1585
PCR_2	Density	5 PC's	0.8289	0.8130	0.0488
	MOE	2 PC's	0.1343	0.1039	139944
	MOR	1 PC	0.0940	0.0784	1580

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. Second, should NIR ever be used to stress-rate lumber, much care needs to be taken when scanning pith-wood.

Predicting juvenile properties from mature model

Juvenile wood is synonymous with inferior mechanical properties because of its lower specific gravity and higher mean microfibril angle, particularly at the basal end of the tree. As a result, it is important to know if global models developed from whole tree values will predict juvenile wood properties. Figure 6(a) illustrates that juvenile wood density is under-predicted using MLR, while MOR and MOE were over-predicted in Figures 6(b) and 6(c). 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.

When only the area under the spectra was used 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 spectral curve to estimate wood density from any portion of the tree than to use other multivariate techniques.

Conclusions

MLR performed as well as PCR in predicting strength, stiffness and density. This is important since MLR is easier to implement and interpret. For PCR, it was found that a better model could be built if only wavelengths known to associate with lignin and cellulose were utilised. This is significant, since data acquisition and analysis of all available NIR wavelengths are common in PCR.

The wood density appeared to be best modelled by accounting for the baseline shift that occurs with

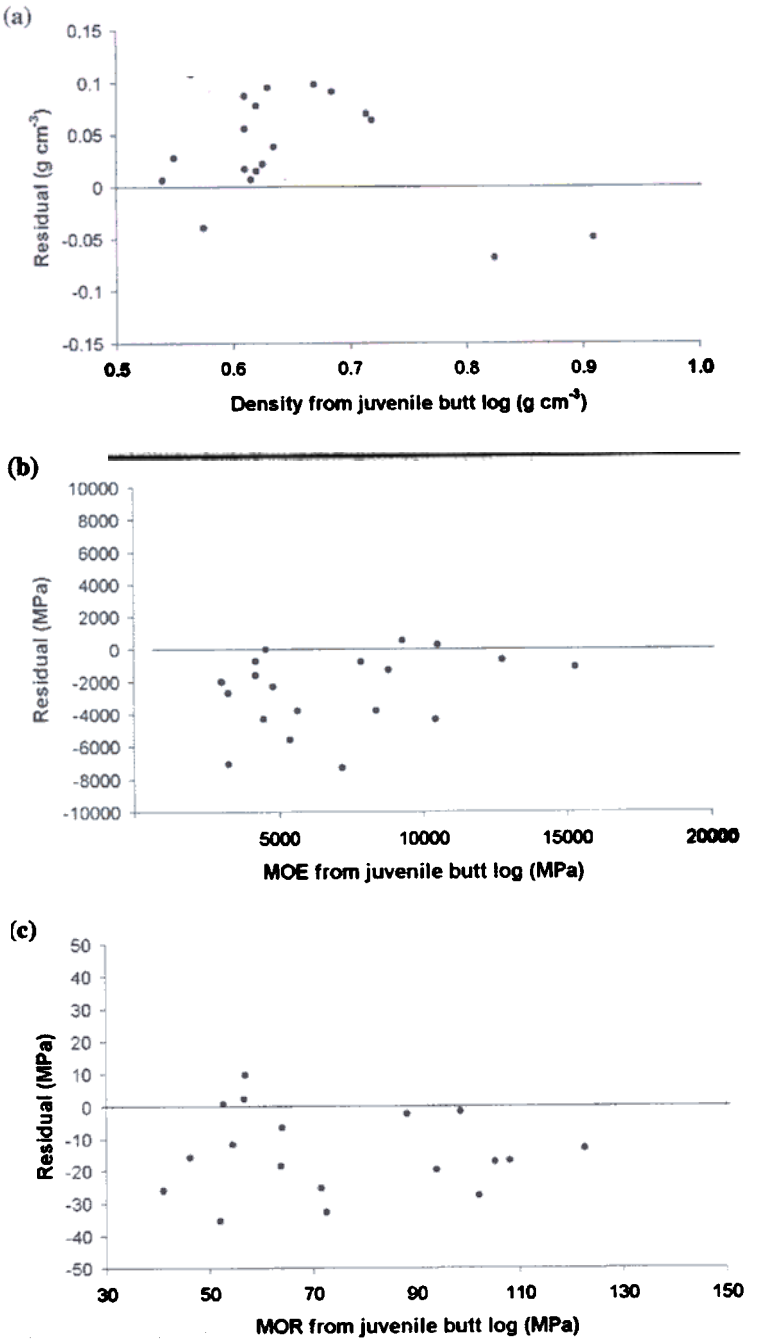

Figure 6. A residual plot of (a) juvenile wood density using a mature wood density MLR model, (b) juvenile wood MOE using a mature wood MLR model and (c) juvenile wood MOR using a mature wood MLR model.

a change in density. This was achieved by regressing the area under the curve to density with least squares regression. Furthermore, the ratio of cellulose to lignin 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 high slope between density and mechanical properties, while juvenile wood had a moderate slope and pith wood 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 is not suitable. While not explored in this study, based on the Introduction, perhaps microfibril angle may account for a higher proportion of the variation in the pith region. Finally, when modelling MOE and MOR, five of six of the same wavelengths were significant for prediction and perhaps sheds some light as to why MOE and MOR are strongly correlated.

Acknowledgements

This study was funded by the USDA National Research Initiative Competitive Grants Program Agreement No. 2001-35103-10908. This paper (No. 03-40-1192) is published with the approval of the Director of the Louisiana Agricultural Experiment Station. The authors gratefully acknowledge the contribution of Mr Jim Roberds and Mr Larry Lott, USDA Forest Service, Southern Research Station, Southern Institute of Forest Genetics. Dr Laurie Schimleck, University of Georgia, is also thanked for his critical review of this paper.

References

1. L.R. Schimleck, R. Evans and J. Ilic, *IAWA J.* **22**, 415 (2001).
2. I. Sakurada, Y. Nukushima and T. Ito, *J. Polymer Sci.* **57**, 651 (1962).
3. L.R. Schimleck and R. Evans, *IAWA J.* **23**, 217 (2002).
4. L.R. Schimleck, R. Evans and A.C. Matheson, *J. Wood Sci.* **48**, 132 (2002).
5. R.M. Kellogg, C.B.R. Sastry and R.W. Wellwood, *Wood Fiber Sci.* **7**, 170 (1975).
6. L. Jurasek, *J. Pulp Pap. Sci.* **21**, J274 (1995).
7. P. Malkavaara and R. Alén, *Chemometr. Intell. Lab. Syst.* **44**, 287 (1998).
8. M.D. Birkett and M.J.T. Gambino, *Tappi J.* **72**, 193 (1989).
9. T.M. McLellan, J.D. Aber, M.E. Martin, J.M. Melillo and K.J. Nadelhoffer, *Can. J. For. Res.* **21**, 1684 (1991).
10. T.M. McLellan, M.E. Martin, J.D. Aber, J.M. Melillo, K.J. Nadelhoffer and B. Dewey, *Can. J. For. Res.* **21**, 1689 (1991).
11. R. Meglen and S. Kelley, National Renewable Energy Lab, USA Document (2000).
12. J.A. Wright, M.D. Birkett and M.J.T. Gambino, *Tappi J.* **73**, 164 (1990).
13. B.G. Osborne and T. Fearn. *Near Infrared Spectroscopy in Food Analysis.* Longman Scientific and Technical, Harlow, UK (1986).
14. A.J. Michell, *Appita J.* **48**, 425 (1995).
15. L.R. Schimleck, A.J. Michell and C.A. Raymond, *Appita J.* **53**, 318 (2000).
16. C. So, L.H. Groom, T.G. Rials, R. Snell, S.S. Kelly and R. Meglen, *Proceedings of the eleventh biennial southern silvicultural research conference.* Gen. Tech. Rep. SRS-48. (2002).
17. P. Hoffmeyer and J.G. Pedersen, *Holz Roh Werkst.* **53**, 165 (1995).
18. W. Gindl, W.A. Teischinger, M. Schwanninger and B. Hinterstoisser, *J. Near Infrared Spectrosc.* **9**, 255 (2001).
19. K.G. Kowalski, *Chemometr. Intell. Lab. Syst.* **9**, 177 (1990).
20. F. Estienne, L. Pasti, V. Centne, B. Walczak, F. Despagne, D.J. Rimbaud, O.E. de Noord and D.L. Massart, *Chemometr. Intell. Lab. Syst* **58**, 195 (2001).
21. Statistical Analysis Software (SAS) version 8.2. Cary, North Carolina, USA (2001).
22. J. Neter, M.H. Kutner, C.J. Nachtsheim and W. Wasserman, *Applied Linear Statistical Models, 4th Edn.* Richard D. Irwin, Inc., Burr Ridge, Illinois, USA (1996).
23. D. Wang, F.E. Dowell and R.E. Lacey. *Cer. Chem.* **76**, 34 (1999).

24. R.G. Pearson and R.C. Gilmore, *For. Prod. J.* **21,** 23 (1971).
25. R.G. Pearson and R.C. Gilmore, *For. Prod. J.* **30,** 47 (1980).
26. B.A. Bendtsen and J. Senft, *Wood Fiber Sci.* **18,** 23 (1986).
27. M. Treacy, J. Evertsen and A. Dhubháin, National Council for Forest Research and Development (COFORD) Document Project Number 96125 (2000).

Appendix

The list summarises acronyms used in the paper.

B_i	*i*th regression coefficient
C_p	Mallows C_p statistic
MLR	multiple linear regression
MOR	modulus of rupture
MOE	modulus of elasticity
NIR	near infrared spectroscopy
PCR	principal components regression
PCR_1	principal components regression using absorbance from all possible wavelengths
PCR_2	principal components regression only using wavelengths associated to wood chemistry
PRESS	predicted residual sum of squares
RMSEC	root mean square error of calibration
RMSEV	root mean square error of validation
SIMCA	soft independent modelling class analogy

Received: 19 May 2003
Accepted: 2 October 2003
Web Publication: