Received: 15 April 2010 ■ *Revised: 11 September 2010* ■ *Accepted: 15 September 2010* ■ *Publication: 30 November 2010*

JNIRS JOURNAL OF NEAR INFRARED SPECTROSCOPY

Special Issue on Wood and Wood Products

Global near infrared models to predict lignin and cellulose content of pine wood

Gary R. Hodge and William C. Woodbridge

Camcore, Department of Forestry and Environmental Resources, NC State University, 3221 Jordan Hall II, Raleigh, NC 27695, USA.
E-mail: grh@gw.ncsu.edu or wcwoodbr@gw.ncsu.edu

Global near infrared models to predict lignin and cellulose content of pine wood were developed using 517 samples for lignin and 457 samples for cellulose. Samples came from seven different pine species, including tropical species (*Pinus caribaea*, *P. oocarpa*, *P. maximinoi*, *P. patula* and *P. tecunumanii*) and temperate species (*P. radiata* and *P. taeda*) from five different countries (Brazil, Colombia, Chile, South Africa and the USA). The global models were tested on an independent validation data set and had excellent fits for lignin [correlation coefficient (r^2) = 0.97 and standard error of prediciton (*SEP*) = 0.44] and good fits for cellulose (r^2 = 0.82 and *SEP* = 1.08). Subsets of the data were used to develop smaller multi-species, multi-site calibrations that could be tested on independent datasets containing different species not included in the calibration model. For calibrations based on four or more species, predictions from those models on independent datasets were generally good, with only slight degradation in r^2 and *SEP* relative to the calibration R^2 and *SECV*. The results suggest that global calibrations could be valuable in tree breeding programmes to rank species and genotypes for lignin and cellulose content. Species-specific models were developed for two species (*P. tecunumanii* and *P. taeda*) which had sufficient numbers of observations; the global calibrations gave predictions as good as the species-specific calibrations.

Keywords: NIR, global calibrations, lignin, cellulose, breeding, indirect selection, wood quality

Introduction

Near infrared (NIR) spectroscopy can be used to provide rapid indirect assessments of chemical properties of various materials. NIR is being utilised or investigated for a broad range of uses in the forestry and forest products industry.[1] An early application of NIR in the forestry field was to conduct foliar analyses of nitrogen in a variety of forest tree species.[2,3] Today it is more commonly used to assess wood properties, including the prediction of pulp yield and lignin and cellulose content of wood.[4–8] More recently, NIR has been shown to have utility to measure solid wood properties such as density, microfibril angle and modulus of elasticity.[9–15] There are a number of advantages with the use of NIR compared to traditional laboratory methods of measuring wood properties. These include rapid speed, reliable and easy measurement of NIR spectra, multiple analyses resulting from one scanning operation and non-destructive sampling.[16] These features are particularly appealing to tree breeders interested in genetic improvement of wood properties, as breeding programmes can often have hundreds or thousands of candidates of interest, and often work with more than one species, multiplying the potential workload.

Tree breeders have known for many years that most wood quality traits are under a high degree of genetic control. Despite this, there has been limited use of selection pressure on these traits in most operational breeding programmes, even though these traits can have significant impact on mill profitability.[7,17,18] Measuring wood properties, particularly chemical traits such as pulp yield or lignin content, can be both slow and expensive. In a tree breeding programme, any technique used to assess wood quality must be fast, inexpensive and reliable. NIR spectroscopy appears to offer that combination of attributes.

ISSN: 0967-0335
doi: 10.1255/jnirs.902

Camcore is a 30-year-old international university–industry research partnership working in the area of gene conservation and tree improvement. There are 28 active and four associate Camcore members in 17 countries. Most of these organisations have a very diverse landbase, necessitating the use of several pine species (as well as eucalypt and other hardwood species) in their plantation and tree improvement programmes. The pine species include a number of tropical and sub-tropical pines from Central America and Mexico, as well as temperate species such as *P. radiata* (native to California, USA) and *P. taeda* (native to the south-eastern USA). Global NIR models (i.e. robust, multi-site, multi-species calibration models) able to predict wood chemical properties would be very valuable for joint Camcore tree breeding programmes, as well as internal company selection and breeding efforts. Previous research has demonstrated that robust NIR models could be developed to predict lignin content of tropical pine wood samples.[19] Models, including samples from multiple regions (i.e. Brazil and Colombia) and multiple tropical species (*P. caribaea*, *P. oocarpa*, *P. tecunumanii*, *P. patula* and *P. maximinoi*), had very good fits, with r^2 around 0.95 and standard errors of prediction (*SEP*) around 0.40%. In this current study, the work is extended to include both lignin and cellulose content, as well as additional pine species and regions, specifically *P. radiata* samples from Chile, *P. patula* samples from South Africa and *P. taeda* samples from three geographical regions in south-eastern USA. The objectives of this study were:

1. to develop global NIR spectroscopy models to predict lignin and cellulose content of wood samples from different tropical, subtropical and temperate pine species grown in different locations around the world
2. to examine how well a multi-site, multi-species calibration could be extended to or extrapolated to an independent data set, possibly with samples from different species and/or regions
3. to compare predictions from a species-specific calibration to the predictions for that species derived from a global calibration model.

Materials and methods

Wood samples

Wood samples were collected from seven species of pines (*Pinus caribaea*, *P. maximinoi*, *P. oocarpa*, *P. tecunumanii*, *P. patula*, *P. radiata* and *P. taeda*) in five different countries: Brazil, Chile, Colombia, South Africa and the USA (Table 1).

In Brazil, samples of *P. caribaea*, *P. oocarpa* and *P. tecunumanii* were collected from progeny tests established by Aracruz and samples of *P. caribaea*, *P. maximinoi* and *P. tecunumanii* from progeny tests established by Cenibra. In Colombia, samples of *P. maximinoi*, *P. patula* and *P. tecunumanii* were collected from progeny tests established by Smurfit Kappa Cartón de Colombia in Colombia. The tests from Brazil and Colombia ranged in age from 13 years to 21 years, and trees representing four to six provenances and 15 to 32 different families were sampled from each species at each site. Trees were felled and discs approximately 2 cm thick were taken at breast height (1.3 m). The discs were air dried, and then shipped to North Carolina State University, Raleigh, NC, USA. Small wedges (approximately 15° to 20°) from the breast height discs were taken and split into "juvenile" and "mature" sections, producing two samples per tree. For the Brazil samples, growth rings were visible and wood laid down in the 10th rings and beyond were considered "mature". For the Colombia samples, which were grown very near the equator, growth rings were not well-defined. Since it was not possible to use growth rings to section the wedges into juvenile and mature wood, an estimated separation point was used. Trees sampled at the Buenos Aires site (elevation 2040 m) were 13 years old and wood in the outer ¼ of the wedge was considered mature. Trees sampled at La Paz and La Arcadia (elevation 1750 m) were 17–21 years old and wood in the outer ⅓ of the wedge was considered mature. Juvenile and mature sections from each tree were chopped into small sticks, ground in a Wiley mill, and passed through a 1 mm screen.

All wood sampling in Chile, South Africa and the USA was done with 12 mm bark-to-bark increment cores taken at breast height. The *P. radiata* samples from Chile came from plantations at 10 different sites ranging in age from 11 years to 23 years. The *P. patula* samples from South Africa came from nine families in a single progeny test aged 14 years. The *P. taeda* samples from the USA came from 163 families in three different regions (Northern, Coastal and Piedmont), and from multiple progeny test sites in each region, ranging in age from eight years to 16 years. For all trees sampled with increment cores, a single pith-to-bark core was used for this study and all cores were processed in the same manner. First, the 1 cm section of the core nearest to the pith was excised and discarded. This was done as the wood near the pith was thought to be more variable and less typical of the remaining core or the whole tree than material laid down when the cambium was older. Next, the remaining core was sliced into thin discs which were then ground in a Wiley mill and passed through a 1 mm screen.

Near infrared scanning

Wood meal samples were oven dried at 50°C for 12 h and then 4 g of wood meal of each sample were scanned using a Foss NIRSystems Model 6500 Spectrometer equipped with a spinning sample module (www.foss-nirsystems.com, Laurel, Maryland, USA). Reflectance was measured over the range of 400–2500 nm at 2-nm intervals. A total of 32 scans were completed and averaged to produce a single reflectance spectrum for each sample.

Wet-laboratory chemistry

Klason lignin content (i.e. acid-insoluble lignin content) was determined on two 1-g samples of woodmeal following TAPPI (Technical Association of the Pulp and Paper Industry) procedure T222 (http://www.tappi.org/Hide/branded-links/Educational-Resource-4021-425225/Standards/TMs/Fibrous-

Table 1. Samples used for the development of global NIR calibration models for lignin and cellulose content of pine wood.

Country	Species	Site/region	Latitude	Elevation (m)	Precipitation (mm)	Age (years)	Cellulose	Lignin	Data subsets			
									Temperate	Tropical	Mix 1	Mix 2
Brazil	*P. caribaea*	Aracruz	19° 48′ S	25	1500	15	20	40	—	X	X	—
	P. oocarpa	Aracruz	19° 48′ S	25	1500	17	22	62	—	X	—	X
	P. tecunumanii	Aracruz	19° 48′ S	25	1500	17	38	64	—	X	—	X
	P. caribaea	Fazenda Primavera	18° 38′ S	850	1400	16	19	—	—	X	—	X
	P. maximinoi	Fazenda Primavera	18° 38′ S	850	1400	16	25	—	—	X	X	—
	P. tecunumanii	Fazenda Primavera	18° 38′ S	850	1400	16	41	—	—	X	—	X
Chile	*P. radiata*	Primavera	37° 38′ S	160	1100	11–23	31	50	X	—	—	X
Colombia	*P. maximinoi*	Buenos Aires	2° 30′ N	2040	2100	13	12	30	—	X	X	—
	P. patula	Buenos Aires	2° 30′ N	2040	2100	13	14	30	—	X	X	—
	P. tecunumanii	Buenos Aires	2° 30′ N	2040	2100	13	19	38	—	X	—	X
	P. maximinoi	La Paz	2° 31′ N	1750	2100	17	5	30	—	X	X	—
	P. tecunumanii	La Paz	2° 31′ N	1750	2100	21	35	54	—	X	—	X
South Africa	*P. patula*	Maxwell	30° 03′ S	1350	817	14	27	27	—	X	X	—
United States	*P. taeda*	Coastal	28–35° N	5–100	1200–1500	9–18	49	40	X	—	—	X
	P. taeda	Northern	32–36° N	20–200	1100–1400	8–15	55	31	X	—	X	—
	P. taeda	Piedmont	30–35° N	100–400	1100–1400	8–16	45	21	X	—	X	—
Five countries	Seven species	—	—	—	—	—	457	517	—	—	—	—

Materials/Acid-Insoluble-Lignin-in-Wood-and-Pulp-Test-Method-T-222-om-06.aspx). This technique involves two phases of hydrolysis with H_2SO_4, first with 72% H_2SO_4 for 2 h at 20°C and then with 3% H_2SO_4 performed on a hot plate, with a 4 h boiling period. The insoluble residue remaining is Klason lignin content.

The α-cellulose content was determined following TAPPI procedure T429 (http://www.tappi.org). This method consists of treating the sample with 17.5% sodium hydroxide for 10 min. Distilled water is then added to reduce the sodium hydroxide to 7.3% and allowed to stand for 1 h. The α-cellulose is collected by filtering and determined volumetrically after oxidation with potassium dichromate. In this study, the procedure began with a 5 g sample of woodmeal to assess holocellulose content and then two 1.5 g samples of holocellulose used to assess α-cellulose content. For the remainder of the manuscript, "cellulose" will refer to α-cellulose content. A total of 457 samples were assessed for cellulose content and 517 samples for lignin content. There were 340 samples with wet-laboratory data for both cellulose and lignin content.

Statistical analysis

Calibration equations were determined correlating NIR spectra to lignin and cellulose content measured using wet chemistry. All models were developed using Foss NIRSystems Vision® software (www.foss-nirsystems.com/software.html, Laurel, Maryland, USA) using the following steps. A mathematical pretreatment was applied to all spectra in order to minimise the effects of particle size variation among samples.[20] Both multiplicative scatter correction (MSC) and standard normal variate (SNV) transformation were examined for both cellulose and lignin models. Preliminary investigations indicated that MSC gave slightly better results for lignin models, and that SNV gave slightly better results for cellulose models. Consequently, all lignin models were developed using MSC and all cellulose models were developed using SNV. Following those pretreatments, spectra were then converted to the 2nd derivative of Savitzky–Golay smoothed spectra (using seven points in the convolution window and a quadratic convolution polynomial). Partial least squares (PLS) equations were then developed using the wavelength range of 1100–2500 nm. Coefficients of determination and standard error were calculated for both the calibration (R^2) and validation (r^2) data sets, with notation as follows: *SEC* = standard error of calibration (calibration data set), *SEP* = standard error of prediction (validation data set). The calibration data set was also partitioned into four sets for cross-validation, with the equation fit on ¾ of the calibration set and cross-validated on the remaining ¼ of the data. This was repeated four times, so that all observations were part of the independent ¼ of the data and the average standard error is the standard error of cross-validation (*SECV*). The ratio of performance to deviation (*RPD*, the ratio of the standard deviation of the laboratory measurements to the standard error of the NIR assessments)[21] was also calculated. RPD_c, RPD_{cv} and RPD_p will be used to indicate the *RPD* of calibration, cross-validation and prediction, respectively. The number of factors in the final model was determined by minimising the predicted residual sum of squares (*PRESS*).[22] The *PRESS* statistic is defined as the sum of the squared prediction errors for all observations, where the residual for each observation is computed as the difference between the laboratory value and the prediction from a calibration model that was fit with a data set excluding that observation. Increasing the number of factors in a model will initially decrease *PRESS*, but eventually the calibration model will be over-fitted, observations used for cross-validation will be poorly predicted, and *PRESS* will increase.

Models were developed for a number of different data sets.

1. Global models for cellulose and lignin were developed using 67% of the available observations as a calibration data set, and 33% of the observations as an independent validation data set.
2. Overall global models were then developed using all 457 observations for cellulose and all 517 observations for lignin.
3. After development of the global models, models were developed for certain subsets of the data, and then used to make predictions for completely different independent data sets, for example, different species and/or different regions.
   i. Temperate species models were developed using the data for *P. radiata* and *P. taeda,* and Tropical species models were developed using the data for *P. caribaea, P. maximinoi, P. oocarpa, P. patula* and *P. tecunumanii*. The Temperate models were then used to predict cellulose and lignin for the Tropical species, and vice versa.
   ii. Models for two independent species mixes were developed. Both mixes contained samples of temperate and tropical–sub-tropical species. Mix 1 contained samples of *P. caribaea*, *P. maximinoi*, *P. patula* and *P. taeda* from the Northern and Piedmont regions and Mix 2 contained samples of *P. oocarpa*, *P. radiata*, *P. tecunumanii* and *P. taeda* from the Coastal region. The Mix 1 models were then used to predict cellulose and lignin for the Mix 2 dataset, and vice versa.
4. Species-specific models for *P. taeda* and *P. tecunumanii* were developed in order to compare their fit with the more broadly-based global models. The species-specific models were fitted using the same observations for fitting the calibration (67% of the data set) and for validation (33% of the data), as were used for the global models. Models were compared by examining r^2 and *SEP* of the two types of models.

Results

Chemical analyses

Wet-chemistry determinations of lignin content for a particular sample had a standard error of ±0.25%. Wet-chemistry determinations of cellulose content for a particular sample had a standard error of ±0.98%. Across the whole dataset, the range in lignin values was from 22% to almost 33%, while for cellulose the range was from 34% to almost 48% (Figure 1). For the 340 samples that were assessed for both traits,

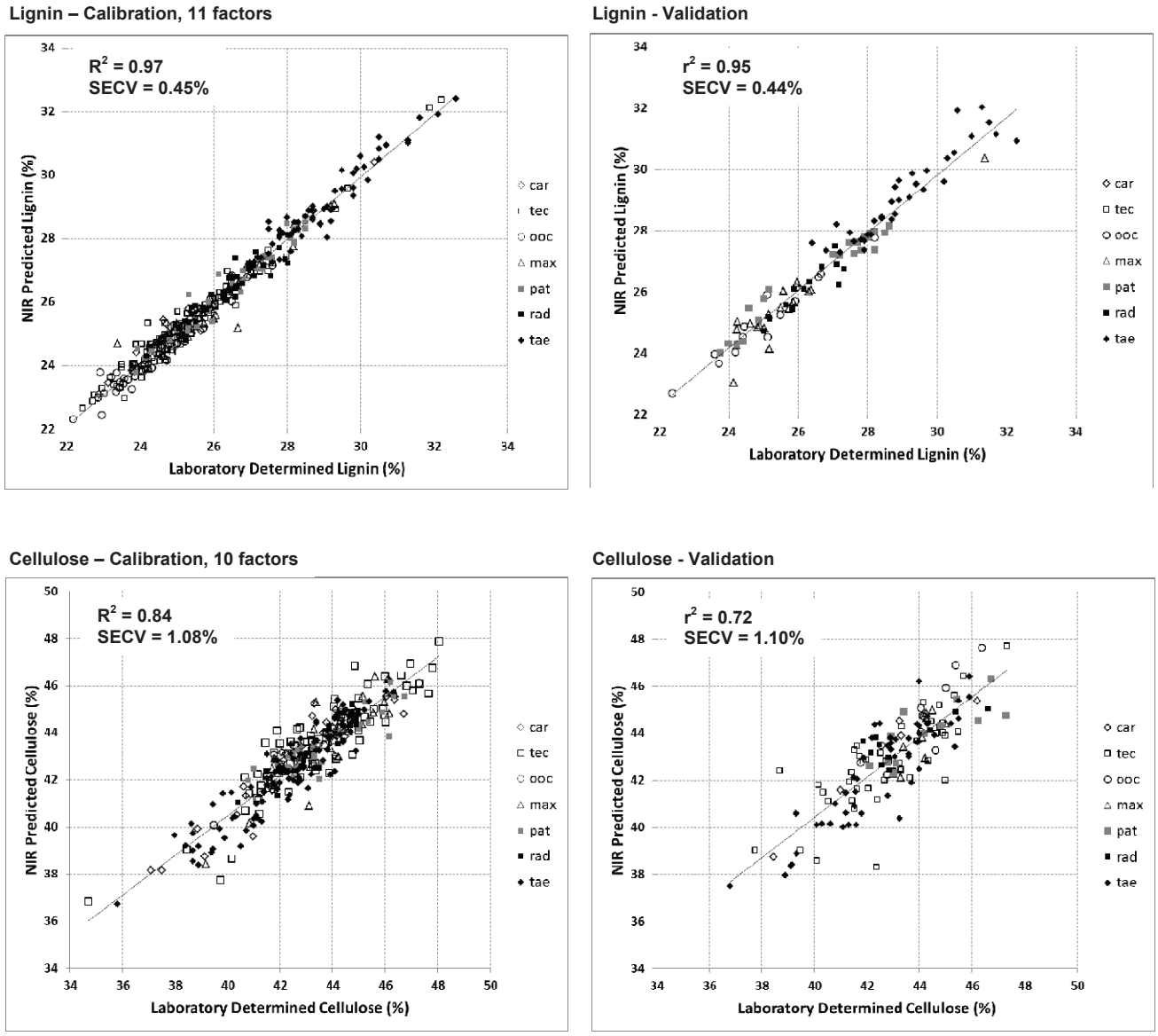

Figure 1. Scatterplots for calibration models and validation datasets for lignin and cellulose. Laboratory-determined chemical content is on the *x*-axis, and NIR predicted value is on the *y*-axis.

there was a negative correlation between lignin and cellulose content ($r = -0.53$, $P < 0.0001$).

Global calibration: validation models

Lignin

The global lignin calibration model gave an excellent overall fit, with 11 factors producing an $R^2 = 0.97$, $SEC = 0.34$, $SECV = 0.45$ and $RPD_c = 5.91$ (Table 2). For each of the seven species, the sub-set of laboratory values and NIR predictions were compared, and correlations, prediction errors and *RPD* were determined. The model fitted well across all seven species–sub-sets, with R^2 ranging from 0.88 for *P. radiata* to 0.97 for *P. caribaea* (Table 2). Mean lignin values for species ranged from 24.5% for *P. oocarpa* to 29.0% for *P. taeda*. The calibration model predicted species means extremely accurately, with species mean for NIR-predicted lignin within ±0.1% of the mean laboratory value in all cases.

The calibration equation also fit the validation dataset extremely well, with an overall $R^2 = 0.95$, $SEP = 0.44$ and $RPD_v = 4.63$ (Table 2). The *SEP* for the validation dataset matched the *SECV* for the calibration dataset almost exactly. The *SEP* and *SECV* compared reasonably well with the laboratory standard error of 0.25%, which should be the theoretical minimum for calibration models. As in the calibration

Table 2. Fit statistics for global calibration–validation models for lignin and cellulose using 67% of all data for calibration and 33% of all data for validation.

Data set	Lignin—calibration							Lignin—validation				
	Factor	R^2	*SEC*	*SECV*	RPD_c	Mean	NIR mean	r^2	*SEP*	RPD_p	Mean	NIR Mean
GLOBAL	11	0.97	0.34	0.45	5.91	26.1	26.1	0.95	0.44	4.63	26.4	26.4
caribaea		0.97	0.29		5.84	25.2	25.3	0.83	0.36	2.44	25.0	25.0
maximinoi		0.91	0.41		3.27	25.8	25.8	0.89	0.53	2.97	25.5	25.4
oocarpa		0.94	0.30		3.99	24.5	24.4	0.94	0.37	4.04	25.1	25.2
patula		0.96	0.30		4.94	26.6	26.6	0.95	0.46	3.78	26.6	26.6
radiata		0.88	0.34		2.92	26.4	26.4	0.82	0.35	2.32	26.2	26.1
taeda		0.92	0.38		3.53	29.0	29.1	0.87	0.53	2.74	29.0	29.1
tecunumanii		0.96	0.31		5.22	25.3	25.4	0.94	0.39	4.24	25.6	25.7
Data set	Cellulose—calibration							Cellulose—validation				
	Factor	R^2	*SEC*	*SECV*	RPD_c	Mean	NIR mean	r^2	*SEP*	RPD_p	Mean	NIR mean
GLOBAL	10	0.84	0.83	1.08	2.53	43.1	43.1	0.72	1.10	1.80	43.0	43.0
caribaea		0.88	0.84		2.92	42.9	43.1	0.93	0.76	3.80	42.4	42.8
maximinoi		0.81	0.78		2.20	44.0	43.9	0.62	0.70	0.93	44.0	43.7
oocarpa		0.92	0.65		3.03	43.7	43.8	0.76	0.84	1.57	44.1	44.6
patula		0.77	0.82		1.96	43.7	43.5	0.61	1.09	1.59	44.3	44.0
radiata		0.87	0.50		2.79	43.7	43.6	0.62	1.04	1.52	43.3	43.6
taeda		0.85	0.78		2.54	42.2	42.2	0.78	1.03	1.99	42.4	42.3
tecunumanii		0.81	0.94		2.29	43.3	43.4	0.60	1.32	1.52	42.8	42.7

dataset, all species sub-sets were fit very well, with r^2 ranging from 0.83 for *P. caribaea* to 0.94 for *P. tecunumanii* (Table 2 and Figure 1). Also of interest was the fact that the species means for wet-laboratory lignin and NIR-predicted lignin were almost identical, again never differing by more than ±0.1%.

Cellulose

The overall cellulose calibration model was also quite good. The optimum model used 10 factors to produce an $R^2 = 0.84$, *SEC* = 0.83, *SECV* = 1.08 and $RPD_c = 2.53$. For the seven species, R^2 ranged from 0.77 for *P. patula* to 0.92 for *P. oocarpa* (Table 2). There was a smaller range among species for laboratory cellulose content than for lignin content, with mean cellulose values ranging from 42.2% for *P. taeda* to 44.0% for *P. maximinoi*. Despite this, the species mean NIR-predicted cellulose for the calibration data sets matched the laboratory values almost exactly, with differences from 0 to ±0.2%.

The fit statistics for the validation data set showed some degradation from the calibration fit statistics. Overall r^2 was 0.72, compared to 0.84 for the calibration data set, but the *SEP* of the validation data set was again nearly identical to the *SECV* from the calibration data set (*SEP* = 1.10, *SECV* = 1.08) and compared favourably with the laboratory standard error of 1.01%. The species means for laboratory cellulose and NIR-predicted cellulose matched up nicely, with a correlation of 0.92 and a maximum difference of 0.5% (Table 2).

Global calibrations: all data

The calibration models using all 517 observations for lignin and all 457 observations for cellulose gave very similar fit statistics to the calibration models using 67% of the data (Table 3). R^2 were essentially identical and *SECV* decreased slightly for both lignin (from 0.45% to 0.44%) and cellulose (from 1.10% to 1.08%).

The correlation between NIR-predicted lignin and cellulose content was –0.58, similar to the observed correlation between laboratory lignin and cellulose content (–0.53). Since wood is made up of lignin, cellulose, hemicellulose and extractives, increases in biomass directed to one of these components will necessarily mean less biomass directed toward the other components. In view of the negative correlation between lignin and cellulose, one might expect that some wavelengths would be useful in predicting both traits. Table 4 presents R^2 and *SEC* for single wavelengths for both lignin and cellulose. The region 1668 nm to 1672 nm has been found to be associated with lignin content in both pines and eucalypts,[23,24] and the wavelength of 1670 nm had a strong correlation with lignin content ($r^2 = 0.68$) in this study. Other wavelengths with high correlations with lignin content were 1482 nm and 1772 nm, with r^2 of 0.73 and 0.62, respectively. All of these wavelengths were also informative for cellulose content, although 1484 nm gave a slightly better correlation with cellulose content than did 1482 nm, and 1668 nm was slightly better than 1670 nm. In this study, some wavelengths gave good correlations with

Table 3. Fit statistics for global calibration models for lignin and cellulose using all data from five countries and seven species.

Data set	Lignin							Cellulose						
	Factor	R^2	*SEC*	*SECV*	RPD_c	Mean	NIR mean	Factor	R^2	*SEC*	*SECV*	RPD_c	Mean	NIR mean
GLOBAL	12	0.97	0.34	0.44	6.04	26.2	26.2	10	0.82	0.89	1.08	2.33	43.0	43.0
caribaea		0.97	0.26		5.34	25.1	25.1		0.88	0.88		2.83	42.9	43.0
maximinoi		0.92	0.40		3.55	25.7	25.7		0.76	0.79		1.92	44.0	44.0
oocarpa		0.94	0.32		3.98	24.6	24.6		0.84	0.65		2.53	43.9	44.0
patula		0.96	0.33		4.85	26.6	26.5		0.76	0.87		1.94	43.9	43.7
radiata		0.90	0.31		3.09	26.3	26.4		0.79	0.70		2.13	43.6	43.4
taeda		0.93	0.36		3.90	29.0	29.0		0.83	0.82		2.44	42.3	42.3
tecunumanii		0.96	0.33		4.92	25.4	25.5		0.76	1.03		2.04	43.1	43.2

cellulose content, but not lignin, for example, 2384 nm with an r^2 of 0.29 and 0.01, for cellulose and lignin, respectively.

In a prior study with unextracted woodmeal samples from Brazil and Colombia, the wavelengths 1144 nm, 2154 nm and 2388 nm had strong correlations with lignin content.[19] In the current study, with the addition of new species and more sampling environments, 1144 nm was predictive of lignin content, but not cellulose content. Wavelength 2154 nm was not strongly correlated with either trait, but 2158 nm was informative for both lignin (r^2 = 0.58) and cellulose (r^2 = 0.24). In this study, wavelength 2388 nm was not correlated with either trait; however, wavelength 2384 nm was relatively strongly correlated with cellulose.

Extrapolation of models to independent datasets

Validation of NIR models is typically done with an independent set of samples, similar to those used to create the calibration.[20] It is clear from the above results that the global models in this study could be used to predict values for samples from the same or similar populations. To examine the question of how well these models could be extended or extrapolated to independent datasets with samples from different geographical regions and/or different species, independent models for the two temperate species (*P. radiata* and *P. taeda*), and the five tropical species (*P. caribaea*, *P. maximinoi*, *P. oocarpa*, *P. patula* and *P. tecunumanii*) were developed and used to predict lignin and cellulose for the other data set. In addition, two other independent mixes of species were created (Table 1). Lignin and cellulose models were calculated for each mix, and then the calibration was used to predict the other data set. Results of these extrapolations can be seen in Table 5 and Figure 2.

Table 4. Fit statistics for the prediction of lignin and cellulose contents for selected single wavelengths.

Wavelength (nm)	Lignin		Cellulose	
	r^2	*SEC* (%)	r^2	*SEC* (%)
1144	0.47	1.44	0.03	2.01
1482	0.73	1.03	0.42	1.56
1484	0.61	1.23	0.43	1.55
1487	0.47	1.43	0.33	1.68
1668	0.62	1.21	0.27	1.76
1670	0.68	1.11	0.25	1.77
1772	0.62	1.21	0.26	1.76
2154	0.04	1.93	0.13	1.91
2158	0.58	1.28	0.24	1.79
2172	0.53	1.36	0.09	1.96
2384	0.01	1.96	0.29	1.73
2388	0.04	1.93	0.13	1.92

The Tropical and Temperate calibration models and the Mix 1 and Mix 2 models behaved similarly to the full dataset calibration models for those traits: for lignin, R^2 was around 0.94–0.97, with *SECV* around 0.40–0.50% and for cellulose R^2 was around 0.70–0.91 with *SECV* around 0.90–1.20% (Table 5). When extrapolated to an independent data set, in three of the four cases there was a good fit for lignin and moderately good fit for cellulose. In one case, the extrapolation of the Temperate calibrations to the Tropical dataset, there was a significant reduction in fit.

The Temperate calibration models were based on only two species and, when extrapolated to the Tropical dataset, gave r^2 = 0.41, *SEP* = 1.59% and RPD_p = 0.92 for lignin and r^2 = 0.43, *SEP* = 1.53% and RPD_p = 1.28 for cellulose. While the r^2 values are highly statistically significant ($p < 0.0001$) and correspond to correlation coefficients of r = 0.64 and r = 0.66 for lignin and cellulose, respectively, the *RPD* ratios are well below 2.0, and spectroscopists would deem these models unsatisfactory.[21] However, for the extrapolation of the Tropical calibrations to the Temperate dataset, the fit statistics were much better: r^2 = 0.88, *SEP* = 0.71% and RPD_p = 2.69 for lignin; and r^2 = 0.70, *SEP* = 1.09% and RPD_p = 1.80 for cellulose. Considering that these values result from an extrapolation of tropical species to temperate species, these values compare well to the validation

Table 5. Fit statistics for independent datasets containing samples of different species and/or geographic regions. Calibrations were fit on each dataset, and then used to predict lignin and cellulose values for the other corresponding dataset.

Calibration model	Lignin						Cellulose					
	Factor	R^2	*SEC*	RPD_c	*SECV*	RPD_{cv}	Factor	R^2	*SEC*	RPD_c	*SECV*	RPD_{cv}
Tropical	7	0.94	0.36	4.10	0.45	3.25	10	0.83	0.81	2.43	1.19	1.65
Temperate	6	0.95	0.43	4.45	0.53	3.60	9	0.88	0.66	2.96	0.93	2.11
Mix 1	8	0.97	0.38	5.78	0.51	4.30	11	0.91	0.65	3.05	1.02	1.95
Mix 2	10	0.97	0.29	6.06	0.41	4.33	5	0.70	1.08	1.83	1.16	1.71

Prediction equation	Validation dataset	Lignin					Cellulose				
		r^2	*SEP*	RPD_p	Mean	NIR mean	r^2	*SEP*	RPD_p	Mean	NIR mean
Temperate	Tropical	0.41	1.59	0.92	25.4	28.6	0.43	1.53	1.28	43.4	44.3
Tropical	Temperate	0.88	0.71	2.69	28.2	28.0	0.70	1.09	1.80	42.	41.7
Mix 2	Mix 1	0.92	0.66	3.32	26.8	26.7	0.60	1.37	1.45	43.0	42.0
Mix 1	Mix 2	0.92	0.51	3.48	25.7	25.8	0.55	1.29	1.54	43.2	43.6

statistics for the full global model: $r^2 = 0.95$, *SEP* = 0.44% and $RPD_p = 4.63$ for lignin; and $r^2 = 0.72$, *SEP* = 1.10% and $RPD_p = 1.80$ for cellulose (Table 2).

For the two species mixes, the results of extrapolating one calibration model to the other dataset gave similar results. Lignin model extrapolations gave better fit statistics than cellulose models. For example, using Mix 2 calibrations to predict Mix 1 data, for lignin gave $r^2 = 0.92$, *SEP* = 0.66% and $RPD_p = 3.32$. For cellulose, $r^2 = 0.60$, *SEP* = 1.37% and $RPD_p = 1.45$ (Table 5).

The results suggest that a global calibration, even one based on only four species, as in the two mixes examined in this study, can be used to predict values for an independent dataset (very precisely for lignin, less so for cellulose), even if it contains species not included in the calibration dataset (Figure 2). The extrapolated models generally ranked the species correctly relative to one another for mean lignin and cellulose content. For the Tropical calibration extrapolated to the Temperate dataset, mean lignin content for *P. radiata* and *P. taeda* were 26.3% and 29.0%, respectively. The Tropical calibration gave a mean NIR-predicted lignin content for *P. radiata* and *P. taeda* of 26.0% and 29.1%, respectively (Figure 3). Similarly, the Tropical to Temperate extrapolation for cellulose ranked the two temperate species correctly with *P. radiata* having 1.3% more cellulose than *P. taeda* (mean NIR-predicted cellulose of 42.7% and 41.4%, compared to mean cellulose content of 43.6% and 42.3%, respectively).

The Mix 1 to Mix 2 extrapolations and the Mix 2 to Mix 1 extrapolations gave nearly perfect estimates of species means for lignin and good species rankings for cellulose (Figure 3). There was only one case where a species was clearly incorrectly ranked: the mean cellulose content of *P. radiata* predicted by the Mix 1 calibration extrapolated to the Mix 2 dataset. The mean cellulose content for *P. radiata* was 0.4% higher than *P. tecunumanii*, but the Mix 1 calibration gave a mean NIR prediction of cellulose content 0.6% lower than *P. tecunumanii* (Figure 3).

There was no bias apparent in the lignin extrapolations; however, there was a tendency for cellulose extrapolations to result in slightly biased predictions. The Tropical-to-Temperate and the Mix 2-to-Mix 1 NIR cellulose predictions were biased low relative to the laboratory values, while the Mix 1-to-Mix 2 NIR cellulose predictions were biased high relative to laboratory values (Figures 2 and 3). This means that extrapolation predictions might not give "true" cellulose content, but should, nevertheless, give precise relative rankings of species and genotypes.

Species-specific calibration models relative to global models

Species-specific models were developed for two species with a substantial number of observations: *P. taeda* and *P. tecunumanii*. As observed with other datasets in this study, the lignin models were excellent ($R^2 = 0.96$ and 0.98, *SEC* = 0.23% and 0.38% and $RPD_c = 5.83$ and 4.26, respectively), and cellulose models were good ($R^2 = 0.81$ and 0.70, *SEC* = 0.86% and 1.10% and $RPD_c = 2.30$ and 1.96, respectively) (Table 6). The species-specific calibrations did use fewer factors than the global models (four to eight factors versus 10 to 11 factors, respectively). However, the global model dataset (seven species and 457 observations for cellulose and 517 observations for lignin, Table 1) was much more variable and complex than the species-specific models (for *P. taeda*, one species, 149 observations for cellulose and 92 for lignin; for *P. tecunumanii*, one species, 133 observations for cellulose and 156 for lignin, Table 1). It seems reasonable to expect that the global model would require some additional factors in the model. Somewhat unexpectedly, the global models gave predictions essentially as good as the species-specific models, with r^2 and *SEP* values for the validation data sets nearly identical. For *P.*

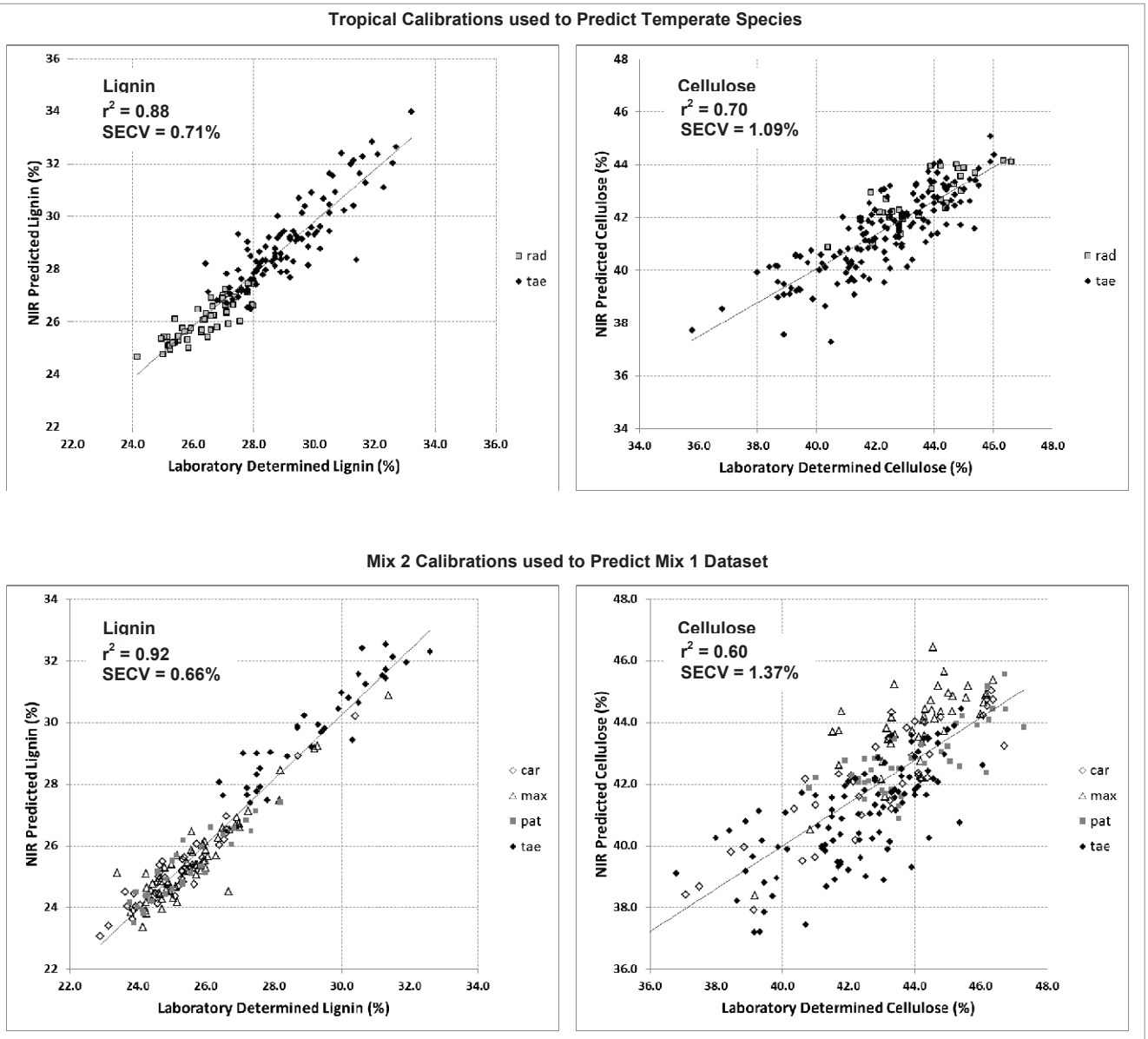

Figure 2. Extrapolations of NIR calibrations to predict lignin and cellulose for independent datasets. Upper panels: Tropical species calibrations (based on *P. caribaea*, *P. maximinoi*, *P. oocarpa*, *P. patula* and *P. tecunumanii*) used to predict values for Temperate species (*P. taeda* and *P. radiata*). Lower panels: Mix 2 calibrations (based on *P. caribaea*, *P. maximinoi*, *P. patula* and Coastal *P. taeda*) used to predict Mix 1 values (*P. tecunumanii*, *P. oocarpa*, *P. radiata* and Northern and Piedmont *P. taeda*).

taeda lignin, the global model *SEP* was 0.53%, slightly higher than the *SEP* for the *P. taeda* model of 0.45%. For cellulose, the two *SEP* values were 1.02% for the *P. taeda* model and 1.03% for the global model (Table 6). For *P. tecunumanii*, the results were similar: for lignin, the global model *SEP* for the *P. tecunumanii* validation dataset was 0.39%, slightly lower than the *SEP* for the *P. tecunumanii* model of 0.47%. For cellulose, the *SEP* values were 1.30% and 1.32% for the *P. tecunumanii* and the global models, respectively (Table 6). In summary, there is no evidence in this study to support the idea that predictions from species-specific calibration models will always be better than those from a robust global calibration model.

Discussion

Lignin, cellulose and pulp yield

In this study, the objective was to develop NIR models for lignin and cellulose content. The true target or breeding objective for pulp and paper producers would actually be pulp yield. Some work has been done with pulp yield for pines,[4] but more effort has gone into developing NIR models to predict pulp yield for eucalyptus species.[6,7,25–28]

Measuring pulp yield is typically slow and expensive, which is the main reason a technique such as NIR spectroscopy is attractive. However, it is necessary to measure pulp yield on a large number of samples in order to develop good calibration

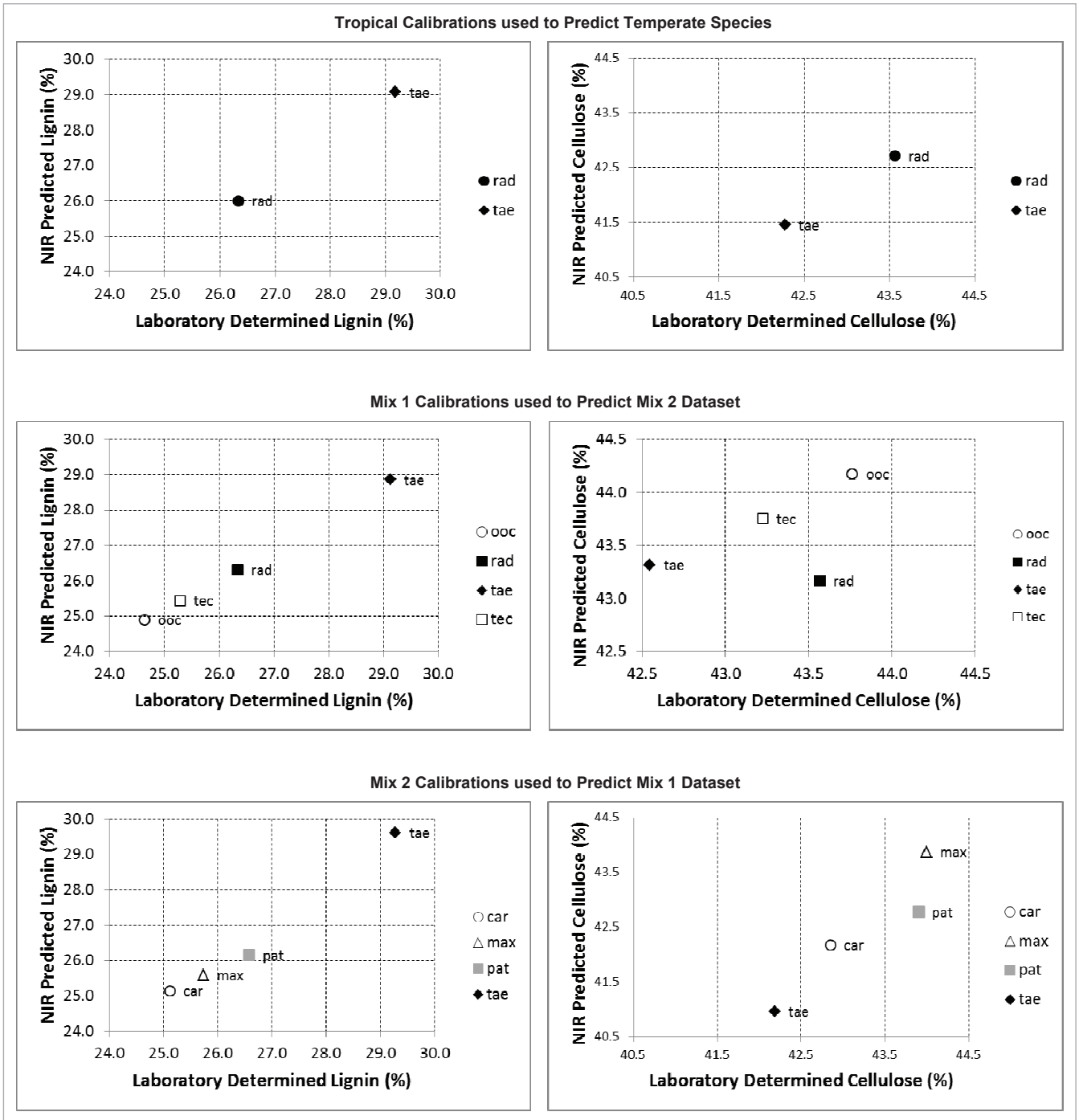

Figure 3. Species means for laboratory lignin and cellulose versus NIR predictions of lignin and cellulose from extrapolations of calibrations to independent datasets. Upper panels: Tropical species calibrations (based on *P. caribaea*, *P. maximinoi*, *P. oocarpa*, *P. patula* and *P. tecunumanii*) used to predict values for Temperate species (*P. taeda* and *P. radiata*). Middle panels: Mix 1 calibrations (based on *P. tecunumanii*, *P. oocarpa*, *P. radiata* and Northern and Piedmont *P. taeda*) used to predict Mix 2 values (*P. caribaea*, *P. maximinoi*, *P. patula* and Coastal *P. taeda*). Lower panels: Mix 2 calibrations (based on *P. caribaea*, *P. maximinoi*, *P. patua* and Coastal *P. taeda*) used to predict Mix 1 values (*P. tacunumanii*, *P. oocarpa*, *P. radiata* and Northern and Piedmont *P. taeda*).

models, and the time and cost of the wet-laboratory work can be substantial. Specific pulping conditions will vary between mills, and the final kappa number (residual lignin) will also vary between mills depending on the product objective. Measuring secondary traits such as lignin and cellulose content is faster and less expensive than measuring pulp yield[8] and these traits are highly correlated to pulp yield.[29–32] Selection for increased cellulose to lignin ratio offers the potential for large economic

Table 6. Comparison of global calibration–validation models (fitting data on seven species) versus single-species models for lignin and cellulose. The global and the single-species models use the same 67% of the data for calibration and the same 33% of the data for validation.

Equation	Data set	Lignin—calibration						Lignin—validation				
		Factor	R^2	*SEC*	RPD_c	Mean	NIR mean	r^2	*SEP*	RPD_p	Mean	NIR mean
taeda	*taeda*	8	0.97	0.23	5.83	29.0	29.0	0.91	0.45	3.23	29.0	29.1
Global	*taeda*	11	0.92	0.38	3.53	29.0	29.1	0.87	0.53	2.74	29.0	29.1
tecunumanii	*tecunumanii*	4	0.95	0.38	4.26	25.3	25.3	0.92	0.47	3.52	25.6	25.7
Global	*tecunumanii*	11	0.96	0.31	5.22	25.3	25.4	0.94	0.39	4.24	25.6	25.7
Equation	Data set	Cellulose—calibration						Cellulose—validation				
		Factors	R^2	*SEC*	RPD_c	Mean	NIR Mean	r^2	*SEP*	RPD_p	Mean	NIR Mean
taeda	*taeda*	5	0.82	0.86	2.30	42.2	42.2	0.75	1.02	2.01	42.4	42.2
Global	*taeda*	10	0.85	0.78	2.54	42.2	42.2	0.78	1.03	1.99	42.4	42.3
tecunumanii	*tecunumanii*	6	0.85	1.10	1.96	43.3	43.3	0.59	1.30	1.54	42.8	42.8
Global	*tecunumanii*	10	0.81	0.94	2.29	43.3	43.4	0.60	1.32	1.52	42.8	42.7

gains,[33] and good global calibrations of the important constituents should provide breeders with the capacity to improve the wood quality of their commercial species. Breeders might wish to select on cellulose/lignin ratio, or perhaps could consider a selection index giving slightly more weight to the lignin prediction due to its higher precision.

Calibrations

In this study, lignin calibrations were better than the corresponding cellulose calibrations in every case. The global model for lignin had an $R^2 = 0.97$ and *SECV* = 0.44 across a range of values from 22% to almost 33% lignin. For cellulose, the values were $R^2 = 0.82$ and *SECV* = 1.08 across a range from 34% to almost 48% cellulose. R^2 values are affected both by the range of values and standard errors of prediction. The standard errors of prediction and cross-validation in this study are similar to those observed in other studies. Garbut *et al.*[5] obtained quite low *SEC* for a lignin calibration set of 45 eucalypt samples (*SEC* = 0.35%) and *SEP* for a validation data set of 41 samples (*SEP* = 0.22%, after adjustment for bias and skew). For a study of lignin content in bamboo, Chinese fir, pawlonia and poplar, *SEP* on validation data sets ranged from 0.28% to 0.81%.[34] For pine species, Wright *et al.*[4] reported on a cellulose model developed with a small dataset of 29 trees representing 14 pine hybrids and *P. taeda*, *P. elliottii* and *P. patula* in South Africa. Fit statistics for the calibration were $R^2 = 0.73$ and *SEC* = 2.4%. Raymond and Schimleck[8] reported on a cellulose model for *E. globulus* developed with 120 trees (90 calibration, 30 validation) across three sites. Fit statistics were good, with $R^2 = 0.77$, *SEC* = 0.95% and *SEP* = 0.88%. Although not directly comparable, standard errors on pulp yield prediction models for eucalypts look similar to the errors on cellulose predictions reported in this study. For example, Schimleck *et al.*[35] reported *SECV* on the order of 0.73% to 1.23% for *E. nitens*. Downes *et al.*[28] report similar values for a large global model of 720 samples from more than 40 different eucalypt species; the calibration model for pulp yield had a *SEC* = 1.25% and a *SECV* of 1.36%.

Almost certainly, much of the difference in the models for lignin and cellulose in this study was due to the difference in laboratory error for the two traits. Lignin values were quite precisely estimated in the laboratory, while cellulose standard errors were approximately four times as large (SE_{lab} = 0.25% and 0.98% for lignin and cellulose, respectively). Better laboratory techniques and/or larger numbers of laboratory replications on a given wood sample to reduce the laboratory standard error for cellulose would very likely have resulted in better NIR calibration models. It is encouraging to note reports in the literature that calibrations produced with "noisy" reference values have been used to make very good predictions for validation datasets with more precisely estimated reference values.[36] But even without considering that possibility, the global cellulose model in this study (with a correlation between laboratory and predicted value of over 0.90) is certainly good enough to be of substantial use to breeders.

The global calibrations in this study have high R^2 (0.97 and 0.82), which should give confidence to breeders considering their use to assess lignin and cellulose for the purposes of selection. Some breeders may be concerned about the standard errors of prediction, particularly for cellulose. However, it is important to remember that these standard errors apply to a measurement on a single 4 g sample of woodmeal. For genetic analyses, multiple offspring from a given parent or full-sib family are typically sampled to provide estimates of parental or family genetic value, while for clones, multiple ramets can be sampled. Without genetic parameter estimates for a tree species of interest, one cannot directly calculate sample size per family or clone needed for optimum selection precision. But

as one example, in a study with *Eucalyptus nitens*, Schimleck *et al*.[37] found high heritability (h^2 from 0.60 to 0.70) and low levels of genotype × environment interaction for both laboratory and NIR-determined cellulose content. With these genetic parameters, 10 trees per family assessed on each of two sites would result in family heritability above 0.80, which would produce very precise rankings of parental genotypes. And, of course, for individual tree or clonal selection, it would also be possible to evaluate more than one 4 g sample for a single tree. As an illustration, consider the scenario where a clone is assessed for cellulose content by sampling five ramets and assume that the *SEP* for cellulose is 1.20%. The standard error of the clonal mean for predicted cellulose would then be quite low, approximately $1.20\% \div \sqrt{5} = 0.37\%$.

Utility of global near infrared calibrations

Global NIR calibrations (i.e. multiple-site, multiple-species models) for wood chemical properties would be of great benefit to tree breeders. If new calibrations need to be developed for every species–site combination, the utility of NIR would be much reduced. As mentioned above, a very good global model to predict pulp yield in eucalypts has been developed,[28] based on 720 samples from 40 eucalypt species and covering a range of pulp yields from less than 40% to more than 60%, with an $R^2 = 0.91$ and *SECV* of 1.36%. The results of this study also suggest that global calibrations for lignin and cellulose content of pine wood are possible and, furthermore, that a robust global model can be extrapolated to independent datasets to precisely rank genotypes, even for sites and/or species that were not included in the original calibration. In this study, the extrapolation that performed the worst was the Temperate calibration extrapolated to the Tropical dataset. The Temperate model could not fairly be called robust, as it was based on only two species, and with the bulk of the observations from *P. taeda* from the south-eastern USA. There are some clear differences in how wood is laid down in temperate pine species compared to tropical pine species: the tropical species *P. tecunumanii* and *P. maximinoi* have been shown to have less latewood, a higher density earlywood and a more uniform pith-to-bark density profile than temperate species, even when grown in the same environment.[38,39] Nevertheless, the Tropical-to-Temperate extrapolation performed quite well in this study, as did the Mix 1-to-Mix 2 and Mix 2-to-Mix 1 extrapolations. In all of those cases, the calibration models were based on four or more species sampled in three or more distinct geographical regions in the world, and might be expected to be relatively robust. For those extrapolations, relative to the calibration dataset, the independent validation dataset showed a slight decrease in r^2 and an increase in *SEP* compared to the R^2 and *SECV* of the calibration dataset. The results imply that use of the full global models described in this study to predict lignin and cellulose for new independent datasets would likely produce $r^2 \approx 0.90$ and $SEP \approx 0.50\%$ for lignin, and $r^2 \approx 0.70$ and $SEP \approx 1.20\%$ for cellulose.

Contrary to expectation, species-specific models for *P. taeda* and *P. tecunumanii* did not produce predictions with better fit statistics for those species than did the predictions for those species from the global model. For a tree breeder wishing to use NIR models to predict lignin and cellulose, a well-developed species-specific model, based on many wet laboratory observations, would certainly be preferable to a global model. However, if a good species-specific model is not available, a breeder could either choose to develop such a model (which is expensive and time consuming), or use a well-developed global model. In this study, the global model gave predictions with accuracy and precision essentially equal to those from species-specific models.

Once a global model has been developed, it could probably be progressively and incrementally improved by the addition of data from new species and new geographical regions. It has been demonstrated that the addition of just a few samples from a new dataset into the calibration model can greatly improve the predictions for the remainder of that dataset.[40] Thus, one approach for any new prediction project would be to use an "enhanced calibration", for example, add from 10 to 50 samples from the new dataset and then use the modified global calibration to predict all other samples in the dataset.[16] However, even without such enhancement, the global models developed in this study could be used directly by breeders to rank species, provenances, families or individual genotypes with a high degree of precision.

Acknowledgements

The authors would like to acknowledge the contributions of Smurfit Kappa Cartón de Colombia, Aracruz Celulose, Celulose Nipo-Brasileira S.A. (Cenibra), Sappi Forests and CMPC Forestal Mininco, and the NC State University–Industry Tree Improvement Program for providing wood samples for this study and related research. Thanks are due to all Camcore members for continuing support of the work in conservation, breeding and research, and to the departments of Forestry and Wood & Paper Science at NC State University. Thanks also go to José Luis Romero, Mike Tighe, Christine Wiecek, Alexander Innes, David Lopez and Stephanie Hafer for all of their hard work in the field and the laboratory.

References

1. C.L. So, B.K. Via, L.H. Groom, L.R. Schimleck, T.F. Shupe, S.S. Kelley and T.G. Rials, "Near infrared (NIR) spectroscopy in the forest products industry", *Forest Products J.* **54(3),** 8 (2004).

2. C.A. Wessman, J.D. Aber, D.L. Peterson and J.M. Melillo, "Foliar analysis using near infrared reflectance spectroscopy", *Can. J. Forest Res.* **18(1),** 6 (1988). doi: 10.1139/x88-002

3. T. McClellan, M. Martin, J. Aber, J. Melillo, K. Nadelhoffer and B. Dewey, "Comparison of wet chemistry and near infrared reflectance measurements of

carbon-fraction chemistry and nitrogen concentration of forest foliage", *Can. J. Forest Res.* **21,** 1689 (1991). doi: 10.1139/x91-233

4. J.A. Wright, M.D. Birkett and M.J.T. Gambino, "Prediction of pulp yield and cellulose content from wood samples using near infrared reflectance spectroscopy", *Tappi J.* **73(7),** 164 (1990).

5. D.C.F. Garbutt, M.J. Donkin and J.H. Meyer, "Near infrared reflectance analysis of cellulose and lignin in wood", *Paper Southern Africa* **April,** 45 (1992).

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

7. A.J. Michell and L.R. Schimleck, "Developing a method for the rapid assessement of pulp yield of plantation eucalypt trees beyond the year 2000", *Appita J.* **51(6),** 428 (1998).

8. C.A. Raymond and L.R. Schimleck, "Development of near infrared reflectance analysis calibrations for estimating genetic parameters for cellulose content in *Eucalyptus globules*", *Can. J. Forest Res.* **32(1),** 170 (2002). doi: 10.1139/x01-174

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

10. L.R. Schimleck and R. Evans, "Estimation of microfibril angle of increment cores by near infrared spectroscopy", *IAWA J.* **23(3),** 225 (2002).

11. L.R. Schimleck and R. Evans, "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 (2002).

12. S.S. Kelley, T.G. Rials, R. Snell, L.H. Groom and A. Sluiter, "Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood", *Wood Sci. Technol.* **38,** 257 (2004). doi: 10.1007/s00226-003-0213-5

13. S.S. Kelley, T.G. Rials, L.R. Groom and C.L. So, "Use of near infrared spectroscopy to predict the mechanical properties of six softwoods", *Holzforschung* **58,** 252 (2004). doi: 10.1515/HF.2004.039

14. P.D. Jones, L.R. Schimleck, G.F. Peter, R.F. Daniels and A. Clark, "Nondestructive estimation of *Pinus taeda* L. wood properties for samples from a wide range of sites in Georgia", *Can. J. Forest Res.* **35,** 85 (2005). doi: 10.1139/x04-160

15. P.D. Jones, L.R. Schimleck, R.F. Daniels, A.C. Clark III and R.C. Purnell, "Comparison of *Pinus taeda* L. whole-tree wod property calibrations using diffuse reflectance near infrared spectra obtained using a variety of sampling options", *Wood Sci. Technol.* **42,** 385 (2008). doi: 10.1007/s00226-008-0178-5

16. L.R. Schimleck, "Near infrared spectroscopy: A rapid, non-destructive method for measuring wood properties and its application to tree breeding", *New Zealand J. Forest Sci.* **38,** 14 (2008).

17. N.M.G. Borrahlo, P.P. Cotterill and P.J. Kanowski, "Breeding objectives for pulp production of eucalyptus globulus under different industrial cost structures", *Can. J. Forest Res.* **23,** 648 (1993). doi: 10.1139/x93-085

18. J. Lopez, J.L. Gomide and R. Phillips, "Influence of eucalyptus wood properties on the financial performance of a modeled Brazilian pulp mill", *O Papel* **70(7),** 53 (2009).

19. G.R. Hodge and W.C. Woodbridge, "Use of near infrared spectroscopy to predict lignin content in tropical and sub-tropical pines", *J. Near Infrared Spectrosc.* **12,** 381 (2004). doi: 10.1255/jnirs.447

20. H. Mark, "Quantitative spectroscopic calibration", in *Encyclopedia of Analytical Chemistry*, Ed by R.A. Myers. John Wiley & Sons, Chichester, UK, p. 13587 (2000).

21. P.C. Williams and D.C. Sobering, "Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds", *J. Near Infrared Spectrosc.* **1,** 25 (1993). doi: 10.1255/jnirs.3

22. S. Weisberg, *Applied Linear Regression*, 2nd Edn. John Wiley & Sons Inc., New York, USA, p. 324 (1985).

23. B.K. Via, C.L. So, L.G. Eckhardt, T.F. Shupe, L.H. Groom and M. Stine. "Response of near infrared diffuse reflectance spectra to blue stain and wood age", *J. Near Infrared Spectrosc.* **16,** 71 (2008). doi: 10.1255/jnirs.756

24. A.J. Michell and L.R. Schimleck, "NIR Spectroscopy of woods from *Eucalyptus globules*", *Appita J.* **49(5),** 23 (1996).

25. B.L. Greaves, N.M.G. Borrahlo and C.A. Raymond, "Breeding objective for plantation eucalypts grown for production of Kraft pulp", *Forest Sci.* **43,** 465 (1997).

26. L.R. Schimleck and A.J. Michell, "Determination of within-tree variation of kraft pulp yield using near-infrared spectroscopy", *Tappi J.* **81,** 229 (1998).

27. C.A. Raymond, L.R. Schimleck, A. Muneri and A.J. Michell, "Non-destructive sampling of *Eucalyptus globulus* and *E. nitens* for wood properties. III. Predicted pulp yield using near infrared reflectance analysis", *Wood Sci. Technol.* **35,** 203 (2001). doi: 10.1007/s002260100092

28. G.M. Downes, F. Catela and R. Meder, "Developing and evaluating a multisite and multispecies NIR calibration for the prediction of Kraft pulp yield in eucalypts", *Southern Forests* **71,** 155 (2009).

29. A.F.A Wallis, R.H. Wearne and P.J. Wright, "Chemical analysis of polysaccharides in plantation eucalypt woods and pulps", *Appita J.* **49,** 427 (1996).

30. A.F.A Wallis, R.H. Wearne and P.J. Wright, "Analytical characteristics of plantation eucalypt woods relating to kraft pulp yields", *Appita J.* **49,** 258 (1996).

31. P.D. Kube and C.A. Raymond, "Prediction of whole-tree basic density and pulp yield using wood core samples in *Eucalyptus nitens*", *Appita J.* **55(1),** 43 (2002).

32. N.D. Kien, T.H. Quang, G. Jansson, C. Harwood, D. Clapham and S. von Arnold, "Cellulose content as a selection trait in breeding for kraft pulp yield in

Eucalyptus urophylla", *Ann. Forest Sci.* **66,** 711 (2009). doi: 10.1051/forest/2009064

33. G.F. Peter, D.E. White, R. de la Torre, R. Singh, D. Newman, O. Alfranca and A. Casimiro Herruzo, "The value of forest biotechnology: a cost modelling study with loblolly pine and kraft linerboard in the south-eastern USA", *Int. J. Biotechnol.* **9(5),** 415 (2007).

34. A. Huang, G. Li, F. Fu and B. Fei, "Use of visible and near infrared spectroscopy to predict Klason lignin content of bamboo, Chinese fir, Paulownia, and poplar", *J. Wood Chem. Technol.* **28,** 194 (2008). doi: 10.1080/02773810802347008

35. L.R. Schimleck, P.D. Kube, C.A. Raymond, A.J. Michell and J. French, "Estimation of whole-tree kraft pulp yield of *Eucalyptus nitens* using near infrared spectra collected from increment cores", *Can. J. Forest Res.* **35,** 2797 (2006). doi: 10.1139/x05-193

36. J. Rodriguez, A. Alves, H. Pereira, D. da Silva Perez, G. Chantre and M. Schwanninger, "NIR PLSR results obtained by calibration with noisy, low-precision reference values: Are the results acceptable?", *Holzforschung* **60,** 402 (2006). doi: 10.1515/HF.2006.063

37. L.R. Schimleck, P.D. Kube and C.A. Raymond, "Genetic improvement of kraft pulp yield in *Eucalyptus nitens* using cellulose content determined by near infrared spectroscopy", *Can. J. Forest Res.* **34,** 2363 (2004). doi: 10.1139/x04-119

38. F.S. Malan, "The quality and wood properties of four provenances of South-African grown *Pinus tecunumanii*", *Ann. Forest Sci.* **51,** 203 (1994). doi: 10.1051/forest:19940301

39. U. Klock, "Qualidade da madeira juvenil de *Pinus maximinoi*" H.E. Moore", PhD thesis, Universidade Federal do Parana, Curitiba, Brazil, p. 291 (2000).

40. L.R. Schimleck, P.D. Kube, C.A. Raymond, A.J. Michell and J. French, "Extending near infrared reflectance (NIR) pulp yield calibrations to new sites and species," *J. Wood Chem. Technol*. **26,** 299 (2006). doi: 10.1080/02773810601076683