Technical report 171

Developing and evaluating a global near-infrared calibration for the prediction of kraft pulp yield in eucalypts

by Geoff Downes, Frank Catela and Roger Meder

July 2007

COOPERATIVE RESEARCH CENTRE FOR FORESTRY
College Road, Sandy Bay, Tasmania
Private Bag 12, Hobart, Tasmania 7001 Australia

www.crcforestry.com.au

Executive summary

Over recent years the application of near-infrared spectroscopy (NIR) to the prediction of wood properties has been demonstrated in many proof-of-concept studies. Previous work has demonstrated that NIR measurements can be used to predict basic density from wood meal, chainsaw dust and solid wood, as well as micro fibril angle and modulus of elasticity in solid samples. For over a decade, the prediction of kraft pulp yield (KPY) has been a constant research focus, and numerous small studies have demonstrated this potential. However, because of the cost of obtaining calibration samples with known KPY, sample numbers are typically less than 100.

While the potential for NIR prediction of KPY is well recognised, the shift to routine commercial use has not occurred. There still remains considerable scepticism in the research and industry communities about the use of NIR. Concern is typically expressed in two areas:

- the consistency, accuracy and precision of predictions
- the need to prepare a separate calibration for each site and /or species group.

To elevate NIR from proof-of-concept to pilot scale, a large multi-site, multi-species calibration was developed over a range of iterative cycles to determine whether KPY in eucalypts can be predicted from a single calibration independent of site and species and also to identify the potential limits of accuracy and precision. While the project itself is ongoing, this paper reports the results of the first seven testing cycles. The NIR calibration was expanded from an initial sample set of 104 mixed eucalypt samples to over 700 samples covering more than 40 species from sites across Australia (predominantly temperate sites).

This paper reports the prediction statistics of seven datasets used to build a global calibration, and an analysis of the performance of the final calibration using two independent data sets that were available for testing but not inclusion. The paper also discusses the expected potential accuracy / precision that can be expected from NIR within the context of the accuracy / precision of the actual KPY data used in developing the calibration. The potential to use NIR in the KPY prediction of standing trees is also considered.

Introduction

Together with tree volume and wood density, KPY is an important indicator of wood quality in plantation eucalypts (Greaves *et al.* 1997, Downes *et al.* 1997). Small changes, over a large resource base, translate into large impacts on enterprise profitability.

Wood properties are the end-product of a range of interacting factors including, genetics, environment and silviculture. To improve the wood properties of plantation-grown trees an understanding of how these factors interact to influence the wood produced is required. The achievement of this goal can only be accomplished by testing large numbers of samples. Unfortunately standard methods of measuring KPY are prohibitively expensive.

Methods for the rapid, cost-effective measurement of KPY have focussed on NIR spectroscopy (Michell 1995, Michell and Schimleck 1998*)*. NIR spectra consist of overtone and combination bands of the fundamental stretching vibrations of organic functional groups and contain chemical and physical information about a sample (Kaye 1954, Shenk *et al.* 1992). NIR spectroscopic analysis involves measuring the NIR spectra for a large number of samples, independently measuring the property of interest using classical methods (e.g. lab pulping), then developing a regression calibration that links the spectra to the parameter of interest. This calibration model is then used to predict the property using spectra of a new set of samples to validate the calibration (Martens and Næs 1984, Thomas 1994).

While the potential of NIR to predict KPY has been repeatedly demonstrated (Michell 1995, Michell and Schimleck 1998, Schimleck and Michell 1998, Schimleck *et al.* 2000, Downes *et al.* 2006), the transition into routine commercial use has not occurred to any significant extent. A predominant reason is that the majority of the work to date has been proof-of-concept studies which have involved small sample sets (generally <100) from a restricted range of sites and species. This has generated the perception that NIR can only be used in this manner and requires calibration development each time it is used, which offsets the cost-effectiveness it promises. There are also concerns about the consistency, accuracy and precision of the predictions.

This paper describes the progressive development and testing of a multi-species, multi-site NIR calibration model for the prediction of eucalypt KPY at kappa 18. The approach taken was to develop an initial model from two sets of available samples, each set representing a different set of samples analysed for KPY by a different pulp laboratory. Further prediction–validation cycles were implemented as new data sets became available.

The objective was to determine whether a single generic calibration model could be developed that would allow the prediction of KPY in any eucalypt species regardless of species or site. It is

believed that, if this is possible, it will ultimately require the building of a calibration with several thousand observations, to adequately represent within the model the variety of chemistries of individual trees.

Methods

Seven distinct sample sets with known KPY were used in this study (Table 1). These sets represented samples from a wide range of sites and species. Some sample sets were obtained from research studies, and where appropriate the reference to the study is indicated in Table 1. However a large proportion of the samples came from routine testing of samples from a wide range of sources. For each sample pulped, a representative sub-sample was made available for NIR analysis.

Table 1. Description of sample sets available for building a large multi-site, multi-species calibration

Set	Pulp laboratory	Number of samples	Species represented*
1	A	106	Mixed species, native forest and plantation
2	B	171	Mixed species, native forest and plantation
3	B	70	Mixed species, native forest and plantation
4	A	110	Plantation grown *E. globulus*.
5	B	60	Plantation grown *E. globulus*
6	B	19	Plantation grown *Corymbia citriodora* ssp *variegata*
7	B	87	Mixed species, native forest and plantation
8	C	50	Plantation grown *E. grandis* hybrids
9	A	20	Plantation grown *E. nitens*

*many samples were only described as mixed eucalypts where species were not known. Hence this list should be considered as a minimal list of represented species.

Pulping conditions

Pulp values were obtained from three different pulp laboratories here after referred to as laboratory A, B and C.

Pulp laboratory A

Chips were thoroughly riffled to ensure adequate mixing prior to sub-sampling for testing. All samples were stored in sealed plastic bags. Laboratory pulping properties were measured using a Haato 12 autoclave air pulping digester and associated equipment. Pulping conditions are given in Table 2. Each pulping value used in this study was based on three to four separate cooks to allow the calculation of a pulp yield corrected to kappa 18, together with chemical demand. Kappa number was measured as per conditions specified in Australian, New Zealand standard AS/NZ 1301.201. Basic density of wood chips was determined in duplicate as per Australian, New Zealand standard AS/NZ 1301.001 using the seven-day soaking option.

Table 2. Pulping conditions used by laboratory A

Nominal time to/at temperature	90/90 min
Cook temperature	170°C
Liquor ratio	3.5:1
Sulphidity	25 per cent
Per cent sodium hydroxide (NaOH) charge	variable to give kappa 18
'H' factor	2,865
Activation energy	147.5 kilojoule per mole
Wood charge	300 grammes oven dried equivalent

* The 'H' factor programme controls actual times.

Pulp laboratory B

The sample was screened to remove over and undersized chips and the remainder equilibrated overnight as a minimum prior to pulping. Oven-dry measurement for moisture content was carried out in triplicate by drying at 105^{o}C for two days (Tappi method T258). Chip basic density is also carried out in triplicate.

Pulping was done in a hot air oven containing six digester vessels (three litres), which are rotated and heated by passing air over a bank of electrical heaters. The chips were kraft cooked, using sodium hydroxide and sodium sulphide as cooking chemicals.

The cook was performed at 165^{o}C taking 1.75 hours to reach temperature. Approximately 2 hours were spent at temperature (1300 H-factor-area under time-temp curve above 100^{o}C). Cooked chips were washed, disintegrated, and spun dry, crumbed and weighed. Oven-dry weights of pulps were determined in triplicate as per Tappi method T412.

Pulp laboratory C

Only one of the validation sets (set 8–50 samples) was pulped by this laboratory. Detailed pulping conditions were not made available to the authors. Unlike the other laboratories the pulp values supplied were not corrected to kappa 18 but represented a range of kappa numbers, varying from 16.0 to 17.2 with an average of 16.5.

NIR spectra acquisition and calibration development

Wood chip samples were ground into wood meal (Downes *et al.* 1997) and spectra were acquired between 4,000 and 15,000 wave numbers (687 – 2,500 nm) on a Bruker MPA FT-NIR instrument. Analysis of the spectra within the Bruker QUANT routine within the OPUS 5.5 software package utilised projection to latent structures (PLS) regression (Martens and Næs 1989). After a systematic study of different spectral processing options, two types of calibration models were developed and tested for each validation cycle. In the first no spectral pre-processing was performed, while in the second, the first derivatives of the spectra were used. For each calibration the optimal wave number

range and number of latent variables was determined using the optimisation process within the Quant software. The model with the lowest root mean square error of cross validation (RMSECV) and number of factors was selected. Obvious outliers were removed based on the difference between observed and predicted values.

Calibration development started with sets one and two which represented sample sets where the KPY reference data was determined by two different pulp laboratories (A and B; Table 1). Calibrations from sets one and two were used to cross predict each other and then combined into a single calibration to determine the potential to utilise the data from both laboratories.

The samples represented over 40 different species of eucalypt (Table 3) from both plantation and native forest sources.

Predicted versus actual plots were prepared in Microsoft Excel and the correlation coefficient determined. The root mean square error of estimation (RMSEE), root mean square error of prediction (RMSEP), standard error of predicted residuals (SEP) and bias were calculated as per Næs *et al.* (2004).

Table 3. Eucalypt species represented in the calibration (numbers in brackets indicate the samples represented

E. arenacea (1)	*E. cypellocarpa (5)*	*E. gomphocephala (1)*	*E. occidentalis (1)*	*E. viminalis (4)*
E. argophloria (1)	*E. dalrympleana (7)*	*E. goniocalyx (1)*	*E. ovata (1)*	*E. punctata (11)*
E. baxteri (1)	*E. delegatensis (8)*	*E. grandis (4)*	*E. petiolaris (1)*	*E. pilularis (2)*
E. blakeyi (1)	*E. diversicolor (20)*	*E. macrorhyncha (1)*	*E. porosa (1)*	*E. polybractea (1)*
E. bridgesiana (1)	*E. dunii (1)*	*E. meulleriana (4)*	*E. radiata (5)*	*Corymbia sp (28)*
E. chlorolada (1)	*E. dwyeri (1)*	*E. miniata (1)*	*E. regnans (11)*	
E. cladocalyx (1)	*E. fasciculosa (1)*	*E. nesophylla (1)*	*E. rubida (1)*	
E. conica (1)	*E. fastigata (5)*	*E. nitens (3)*	*E. sieberi (5)*	
E. cosmophyla (1)	*E. globulus (244)*	*E. obliqua (9)*	*E. tetradonta (1)*	

Results

A preliminary study showed that most sample sets generated calibrations that explained over 90 per cent of the variance in the data (Table 4). The OPUS 5.5 QUANT software (Bruker 2005) used an optimisation procedure in which a range of spectral pre-processing methods and wave number ranges were evaluated to determine the optimal calibration based on RMSECV and the number of latent variables (principal components) required. It was evident that, in general, most of the pre-processing options gave reasonable calibrations, and that either no spectral processing (NSP) or the first derivative of the spectra performed the best. Consequently further work on the calibration focussed on these two approaches only, and calibrations using both approaches were compared.

Several of the data sets had small numbers of samples (19 and 20 in sets six and nine respectively) and the cross-validation models in these sets generally performed poorly. In general, the first derivative models required fewer latent variables to explain similar amounts of variance.

Table 4. Performance calibration statistics for each available sample set

Model calibration set	Pulp laboratory	Calibration (r^2)	RMSEE	Cross validation (r^2)	RMSECV	No. factors	Wave number range	Outliers excluded
No spectral processing								
Set 1	A	0.98	0.59	0.92	1.14	12	12,501-7,498; 6,102-4,598	7
Set 2	B	0.94	1.15	0.93	1.23	13	7,502-5,446; 4,602-4,247	14
Set 3	B	0.94	1.05	0.89	1.33	9	7,502-6,098; 4,601-4,247	2
Set 4	A	0.85	1.08	0.80	1.19	7	7,502-6,098	6
Set 5	B	0.95	0.82	0.87	1.21	11	6,102-4,598	1
Set 6	B	0.98	0.32	0.00	1.42	10	6,102-5,446; 4,602-4,247	1
Set 7	B	0.98	0.58	0.94	1.21	17	7,502-4,247	0
Set 8	C	0.9	0.48	0.70	0.71	12	7,502-6,098; 5,450-4,598	0
Set 9	A	0.92	1.12	0.56	1.96	8	5,454-4,247	0
First derivative								
Set 1	A	0.91	1.23	0.92	1.35	6	7,502-6,098; 5,450-4,247	4
Set 2	B	0.95	1.05	0.92	1.25	12	7,502-4,247	12
Set 3	B	0.94	1.03	0.92	1.14	4	6,102-5,774	3
Set 4	A	0.82	1.12	0.77	1.21	4	7,502-5,446	5
Set 5	B	0.98	0.57	0.90	1.10	11	7,502-6,098; 5,450-5,022	0
Set 6	B	0.99	0.16	0.56	0.93	7	6,800-6,098	1
Set 7	B	0.97	0.82	0.93	1.21	7	7,502-5,446	2
Set 8	C	0.85	0.51	0.74	0.63	5	7,502-5,446; 4,602-4,247	1
Set 9	A	0.92	1.02	0.17	2.68	6	5,454-4,247	0

Consistency within a laboratory

A fundamental issue in the application of NIR as a predictive tool is whether calibrations are unique to a given laboratory. To what extent can a calibration based on data from laboratory A be used to predict values produced by laboratory B? The first two sample sets were pulped by different laboratories. Set one comprised 110 samples from a range of eucalypt species that had been evaluated for KPY by laboratory A. These samples had been pulped to a range of different kappa numbers (Table 1) and the KPY values corrected to kappa 18. A second set of 274 samples, from laboratory B, had also been analysed to give KPY at kappa 18. Like those described above, they represented a broad range of eucalypt species, predominantly from regrowth forests across the south-east region of Australia. The calibrations developed within each set were reasonable (Figure 1 a, b) indicating that the calibration data quality was acceptable for this study.

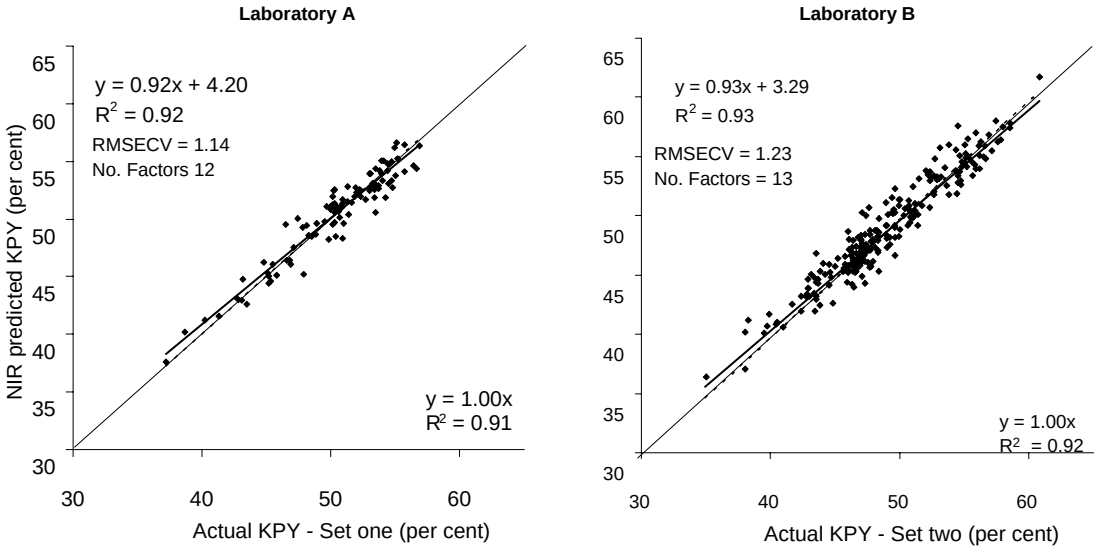

Figure 1. NIR-predicted vs laboratory-measured KPY calibration plots for (b) full cross-validation models for laboratory A and (b) laboratory B samples. Descriptive statistics are shown in Table 4

Cross prediction between two different pulp laboratories

To determine whether the data from the two laboratories could be combined, the KPY for samples within each set was predicted using the NIR calibration developed from the other laboratory (Figure 2). These results indicated that the calibrations from the two laboratories gave reasonable KPY predictions of each other's pulp data, and that the spectra could be combined into a single calibration (Figure 3). Both NSP and first derivative pre-processing options gave calibrations with similar statistics, raising the question as to whether one is more robust in prediction than the other.

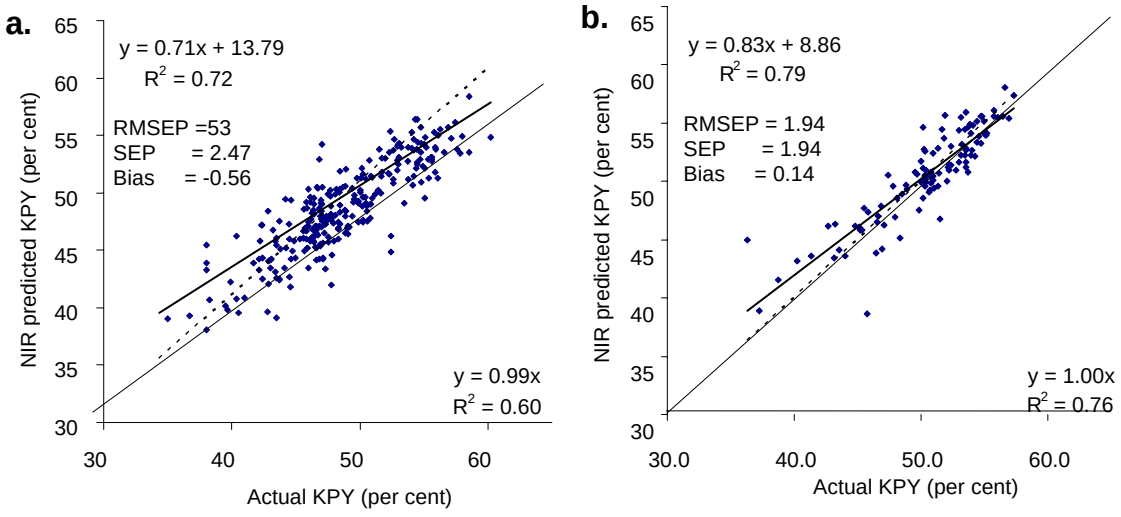

Figure 2. NIR-predicted vs laboratory-measured KPY for (a) set two predicted using set one calibration model and (b) set one predicted using set two calibration model

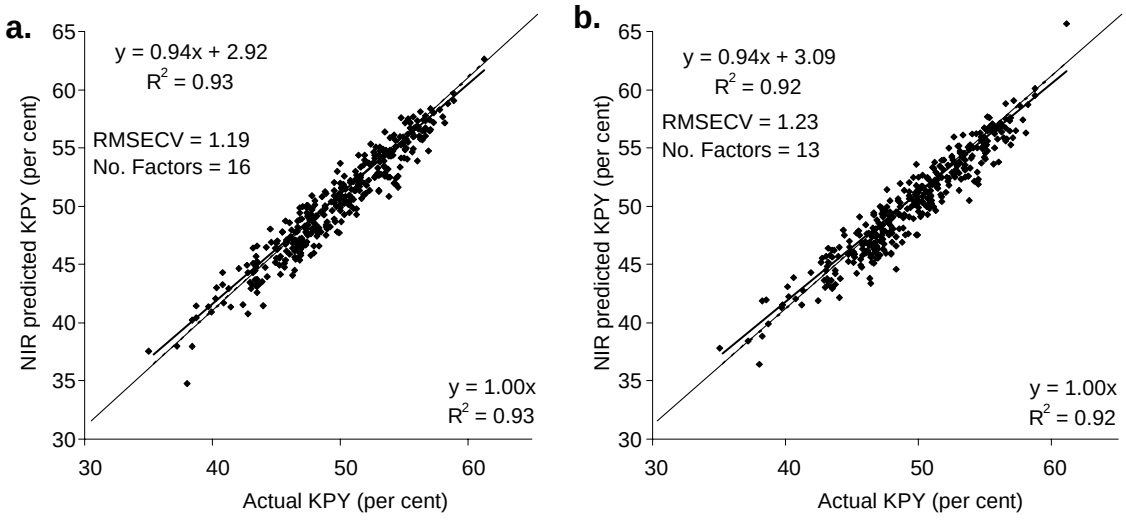

Figure 3. NIR-predicted vs laboratory-measured KPY plots of full cross validation models from combining sets one and two (a) no spectral pre-processing (b) first derivative pre-processing

Iterative calibration development and validation

Five additional sample sets (sets three to seven) were used to test and develop the calibration further. Of these, one used data arising from laboratory A and four used data from laboratory B. These sample sets were used in the order they became available to test the robustness of prediction, followed by expansion of the calibration. The descriptions of each calibration model are described in Table 5, while the predictive performance of each testing cycle is shown in Table 6.

The variance explained by the calibration and cross-validation models was consistent and stable as the sample numbers increased. The optimal spectral ranges varied were generally within the wave-number range 7,502–4,247 (1,330–2,350 nm). Calibrations based on raw spectra with no pre-processing consistently required about three to four more factors than calibration models using the first derivative of the spectra, and excluded fewer samples as outliers (Table 5).

Table 5. Performance statistics of the calibrations through each development iteration

Model calibration set	No. samples	Calibration set range	No spectral processing Calibration (r^2)	RMSEE	Cross validation (r^2)	RMSECV	No. factors	Wave number range	Outliers excluded
Set 1-2	384	25.9	0.94	1.11	0.93	1.19	16	7,502-4,247	18
Set 1-3	454	25.9	0.93	1.19	0.92	1.28	16	6,102-4,247	17
Set 1-4	564	25.9	0.91	1.27	0.9	1.33	16	7,502-4,247	17
Set 1-5	624	26.2	0.92	1.29	0.91	1.37	17	7,502-4,247	17
Set 1-6	644	26.3	0.92	1.29	0.91	1.35	16	7,502-4,247	17
Set 1-7	728	26.3	0.93	1.24	0.91	1.35	20	7,502-4,247	21

Model calibration set	First derivative Calibration (r^2)	RMSEE	Cross validation (r^2)	RMSECV	No. factors	Wave number range	Outliers excluded
Set 1-2	0.95	1.06	0.92	1.23	13	7,502-4,247	19
Set 1-3	0.93	1.16	0.91	1.29	12	7,502-4,598	22
Set 1-4	0.93	1.16	0.91	1.29	13	7,502-4,247	22
Set 1-5	0.92	1.21	0.91	1.29	15	7,502-4,247	22
Set 1-6	0.92	1.26	0.91	1.31	14	7,502-4,247	21
Set 1-7	0.92	1.25	0.91	1.36	14	7,502-4,247	20

The first derivative calibration consistently explained more variance in the prediction sets than the NSP calibration (Table 6). The difference was typically less than five per cent, but significant ($p=0.03$). A concern for the application of NIR as a predictor for KPY was the variability in bias among the prediction sets. This value is a measure of difference between the mean of the actual KPY predicted values and the mean of the predicted values. This is also evident in comparing the RMSEP and SEP values. The former estimates error which included the bias, while the latter removes it first. Thus the first is a combined measure of accuracy and precision, while the latter is a measure of precision only. While in general the NSP calibrations had smaller bias, the change during development appeared relatively random. It remains to be seen whether expanding the calibration with additional sample sets will reduce or stabilise it.

Table 6. Predictive statistics of the developing calibration for each sample set during development

Calibration set	Validation set	No. validation samples	Sample range	Prediction range	Prediction (r^2)	RMSEP	SEP	bias
			No spectral processing					
1-2	3	70	14.6	20.1	0.75	2.48	2.17	-1.21
1-3	4	110	13.5	13.8	0.65	5.72	1.69	-5.47
1-4	5	60	15.2	10.3	0.83	1.6	1.61	-0.03
1-5	6	19	5.9	6.6	0.55	2.11	0.97	-1.89
1-6	7	87	21.9	19.1	0.90	3.75	1.54	-3.42
			First derivative					
1-2	3			16.0	0.68	2.64	2.48	-0.95
1-3	4			14.0	0.68	6.72	1.64	-6.52
1-4	5			10.2	0.87	1.93	1.41	-1.34
1-5	6			6.9	0.60	1.95	0.92	-1.73
1-6	7			18.7	0.91	4.39	1.41	-4.12

Independent test data

Two additional data sets were used for a final independent evaluation of prediction performance. Neither of these sample sets was available for inclusion in the calibration model. The first was supplied by Votorantim (Brazil) consisting of 50 samples of *E. grandis* hybrids that had been pulped to varying kappa numbers between 16.5 and 18 (Gabriela pers. comm.[1]). The pulp data was generated by a third laboratory (laboratory C) using trees grown in completely different environments to those used to generate the calibrations. The second sample set consisted of 20 *E. nitens* samples grown in Tasmania, and pulped by laboratory A as part of a CRC for Forestry project (D. Stackpole pers. comm.). Cross-validation models for each of these sample sets are shown in Table 4.

In the Brazilian grown *E. grandis* hybrids, up to 76 per cent of the variance in KPY was explained. Interestingly the bias seemed generally to diminish as the sample numbers within the calibration used for prediction was increased. As an additional test, calibrations were developed in which only data from a single laboratory was included to determine whether these calibrations explained greater variance. This was not evident; however they did differ in terms of bias.

The predictions made with the different calibrations varied in terms of bias as well as variance explained. In general the variance explained tended to increase until the sample set seven was added to the model. On average the two calibration types explained similar amounts of variance, had similar SEP but the calibration based on raw NSP spectra had less bias than the first derivative

[1] Ana Gabriela Monnerat Carvalho Bassa, Votorantim Celulose e Papel, CDTC, Jacareí – SP, Brazil.

calibration. Also of interest is that the laboratory specific calibrations explained similar amounts of variance, but the laboratory B calibration had much lower bias.

Table 7. Predictive statistics for sample set eight (Votorantin)

Calibration	Prediction range	Prediction (r^2)	RMSEP	SEP	bias
	No spectral preprocessing				
Set 1	4.4	0.48	2.27	0.95	2.1
Set 2	6.1	0.64	3.81	0.81	3.7
Set 1-2	5.4	0.72	3.93	0.70	-3.9
Set 1-3	5.2	0.62	3.22	0.80	-3.1
Set 1-4	5.0	0.75	0.67	0.67	-0.1
Set 1-5	5	0.76	0.72	0.66	-0.3
Set 1-6	5	0.74	0.76	0.68	-0.4
Set 1-7	5.6	0.53	1.06	0.90	0.6
Laboratory A only	4	0.62	2.37	0.81	2.2
Laboratory B only	5.4	0.70	0.75	0.71	0.3
	First derivative preprocessing				
Set 1	4.7	0.52	5.67	0.91	5.6
Set 2	6.2	0.64	3.19	0.81	-3.1
Set 1-2	5.3	0.71	0.80	0.70	0.4
Set 1-3	5.6	0.71	5.88	0.70	-5.8
Set 1-4	4.6	0.7	1.73	0.72	1.6
Set 1-5	4.7	0.76	1.04	0.66	0.8
Set 1-6	5	0.71	0.76	0.71	-0.3
Set 1-7	5.5	0.49	1.58	0.97	-1.3
Laboratory A only	5	0.58	3.34	0.85	3.2
Laboratory B only	4.1	0.58	0.97	0.85	-0.5

The second data set consisted of 20 samples and had been pulped as a check on the predictions made from 2,100 increment core samples from an *E. globulus* breeding trial in Tasmania, Australia. As such the predictions (Table 8) are made from cores sampled at breast height, relative to pulp values obtained from whole-tree chip samples. On average both calibration types explained 82-83 per cent of the variance and had similar SEP. The bias was markedly higher than that seen in previous tests and is probably due in large part to the predicted KPY of the core versus the actual KPY of the whole tree. Interestingly the calibration based on samples from laboratory A tended to explain slightly more variance and have lower bias than the calibration containing only samples from laboratory B.

Table 8. Predictive statistics for sample set nine (CRC for Forestry)

Calibration	Prediction range	Prediction (r^2)	RMSEP	SEP	bias
		No spectral preprocessing			
Set 1	11.1	0.85	5.84	1.24	-5.7
Set 2	11.4	0.84	2.79	1.67	2.3
Set 1-2	11.3	0.81	7.19	1.60	-7
Set 1-3	9.6	0.77	11.06	1.56	-11
Set 1-4	10.6	0.81	4.46	1.47	-4.2
Set 1-5	9.9	0.81	5.21	1.44	-5.0
Set 1-6	9.4	0.85	3.84	1.38	-3.6
Set 1-7	10	0.83	4.04	1.47	-3.8
Laboratory A only	7.8	0.87	2.62	1.11	-2.4
Laboratory B only	10.7	0.85	4.85	1.41	-4.7
		First derivative preprocessing			
Set 1	7.5	0.86	1.20	1.17	0.4
Set 2	10.1	0.81	7.91	1.69	-7.7
Set 1-2	10.2	0.81	5.69	1.52	-5.5
Set 1-3	10.2	0.84	10.30	1.54	-10.2
Set 1-4	9.3	0.81	4.02	1.42	-3.8
Set 1-5	9.7	0.83	4.90	1.45	-4.7
Set 1-6	9.5	0.84	4.55	1.56	-4.4
Set 1-7	10.4	0.87	4.43	1.43	-4.2
Laboratory A only	7.9	0.86	2.86	1.19	-2.6
Laboratory B only	9.1	0.81	3.62	1.90	-3.4

Current calibration description

As of April 2007, the eucalypt KPY calibration is based on 720 samples from more than 40 different species from across Australia (Figure 4) albeit dominated by *E. globulus* (Table 3), a cool temperate species. Future work will increase the representation of sub-tropical species and involve more comprehensive evaluations using samples for independent validation as they become available.

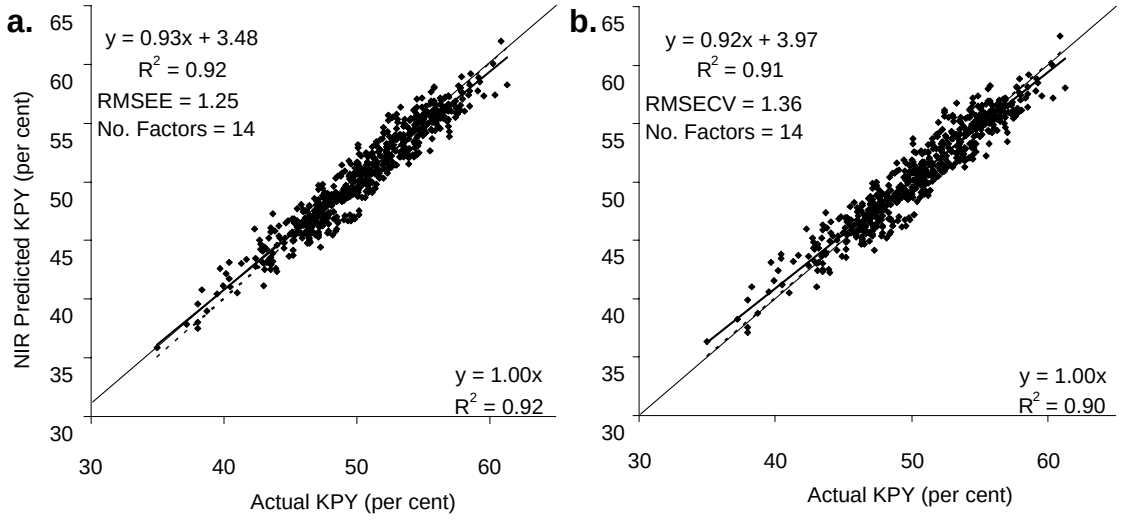

Figure 4. The current (a) calibration and (b) cross-validation models represent over 720 samples from more than 40 different species

Discussion

The development of a single generic calibration for the prediction of KPY in eucalypts seems possible. Good quality pulping data from two independent laboratories were combined into a single generic calibration which allowed reasonably precise predictions of KPY values provided by a third laboratory. Predicted values that explain more than 70 per cent of the variation in laboratory values was achieved. Explaining more than 80 per cent should not be uncommon. However, while predictions may have acceptable precision, their accuracy may be a concern. The fluctuations in bias seemed to be large and variable from one development cycle to the next.

Many studies have demonstrated the potential to use NIR for the prediction of KPY in eucalypts and some of these have explored multi-site calibrations (Michell and Schimleck 1998, Schimleck *et al.* 2000) for *E. globulus* and *E nitens*. Similarly multi-site and species calibrations have been shown from spectra collected from green chips (Wright *et al.* 2003). Schimleck *et al.* (2006) compared seven different species and five hybrids from three different locations in Brazil, producing a calibration with poor calibrations statistics. The suggested reasons for this were that the actual KPY data was not pulped to a constant kappa number, but varied between 16.7 and 18.8.

In general, attempts to develop broader, more generic calibrations have been few, limited in both the species and site range, and represented by relatively small numbers of samples. This is probably due in large part to the expense of obtaining good calibration data, as well as the commercial sensitivity that often surrounds it. As far as can be determined, the calibration development

described here represents the most extensive yet reported for KPY. It is also unique in that it combines samples where the pulp data has been produced by two different laboratories.

Prediction precision

Over the testing cycle, the precision of NIR predictions tended to improve, even though the variance explained by the cross-validation models for each calibration remained consistently high (r^2 ~0.92). This was apparent in the predictions of the Brazilian test set (set eight), apart from the final calibration where the variance explained dropped from 74 per cent to 53 per cent. Interestingly the final calibration performed well in predicting the CRC for Forestry (set nine) samples. A subsequent analysis of 30 samples from Chile (not shown) also indicated that this final calibration gave precise predictions.

The reasonably high levels of variance explained indicate that the NIR predicted data is suitable for ranking or discriminating differences between individuals or populations, as required in selection for breeding. Precision (variance explained) should be improved by expanding the calibration set by include more samples from different species and regions. However given the expense of obtaining actual KPY data, the generation of these calibrations is likely to be slow.

Prediction accuracy

Prediction accuracy refers to the closeness of the prediction to the true value. bias is a measure of this (Downes *et al.* 1997). Ideally predictions should be both accurate and precise resulting in small RMSEP and SEP values. As the number of samples in the calibration set increased (Table 6), bias did not show any consistent tendency to decrease or stabilise. However, as a general trend, the bias in the predictions from set eight tended to diminish as the later, more extensive calibrations were used (Table 7). Over the last four iteration cycles, bias was consistently less than one per cent. This is despite the fact that the KPY data originated from a third laboratory (Laboratory C) with varying kappa numbers (16.5–18). In contrast the bias evident in predicting values from set nine (Table 8) was consistently large, often exceeding -5 per cent. This was presumably due at least in part to the predictions based on cores taken at 1.1 m, compared to the actual KPY data prepared from whole-tree samples. Cores over-sample the younger, inner wood at the expense of outer wood, which generally has higher pulp yield (Schimleck and Michell 1998, Downes *et al.* 2000).

For all practical purposes, the true KPY value in these studies is considered to be the value supplied by the wet chemistry assessment.

There are several reasons why an NIR calibration might lack accuracy;

- The laboratory-determined pulp yield data is not accurate
- The sample from which the NIR spectra was obtained poorly represents the sample used to generate the pulp yield data
- The range of KPY in the calibration data set is limited and/or
- The range of variation in the prediction set is limited or
- NIR doesn't work well for KPY which is not a chemical property as such but the result of a complex industrial process.

One of the constraints in assessing the robustness of NIR predictions is uncertainty about the variance in the actual KPY data generated by traditional pulping methods. For commercial reasons these figures could not be made available to the authors. However it is evident that analyses of this kind can be subject to variation and this will impact on the observed accuracy of NIR-based predictions. One commercial laboratory finds that the actual pulp yield data they supply is accurate to ± 0.5 per cent, and in another study KPY values had a standard deviation of 0.5 per cent (Clarke, 1995).

KPY is known to vary within trees both radially and longitudinally (Downes *et al.* 1997, Schimleck and Michell 1998). Thus when selecting a sample for NIR analysis, care must be taken to ensure that it is representative of the sample used in developing the calibration. If a whole-tree chip sample is used, the sub-sample taken for NIR must be from a well-mixed sample and large enough to be representative prior to grinding. Similarly one would expect that, dependent on the pattern of variation within a tree, predictions made from a sub-sample taken from a specific location within a stem (e.g. increment core), when compared to actual KPY values from the whole-tree, would display a given bias or offset (e.g. set nine, Table 8). The prediction of KPY in cores exhibited relatively large bias compared to actual KPY data made from whole-tree chip samples. While the predictions were consistently lower than the actual, the rankings of trees based on NIR-predicted KPY should still be sufficient for identifying quartiles.

Effect of range

While it was not consistently true, there was some indication that the range in the prediction set influenced the variance explained, with narrower ranges having less variance explained (Table 6). This might be expected and the effect of narrowing the range on the variance explained in the final calibration (sets one to seven) was examined (Figure 5). The actual variance explained is tied

closely to the range of the actual or predicted KPY values, the scatter about the mean and the range over which the data extends.

If the calibration range of the predicted set is narrow, one can plot the variance explained in terms of what might be expected from the calibration data (Figure 5). This series of plots was produced using the calibration shown in Figure 4, and limiting the predicted (left-hand plots) and actual (right-hand plots) data ranges. Thus if the prediction range is less than 10 per cent, a prediction (r^2) of 0.8 or better might be considered good.

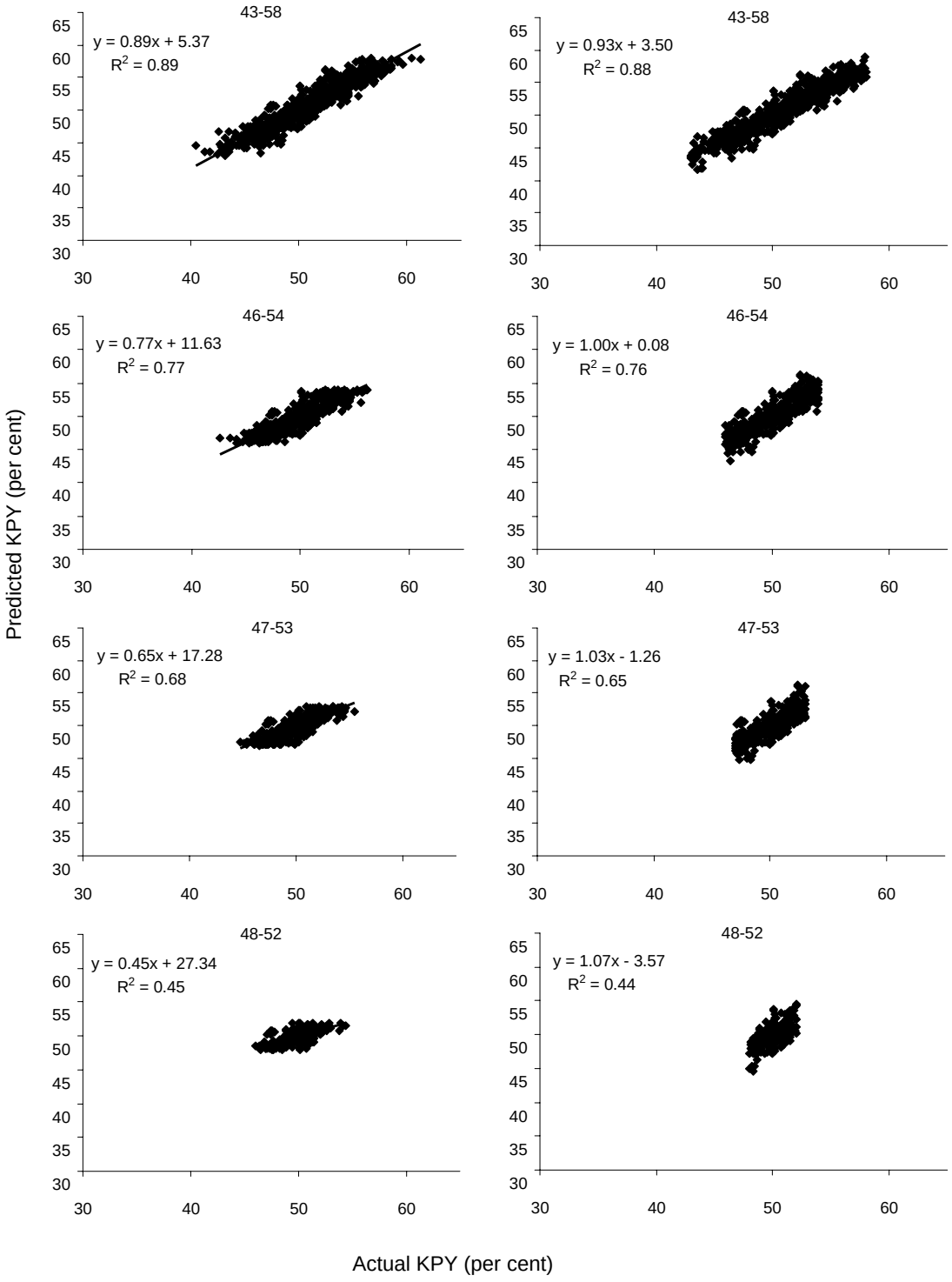

Figure 5. Effect of reducing the predicted range (left side) and actual range (right side) on the variance explained.

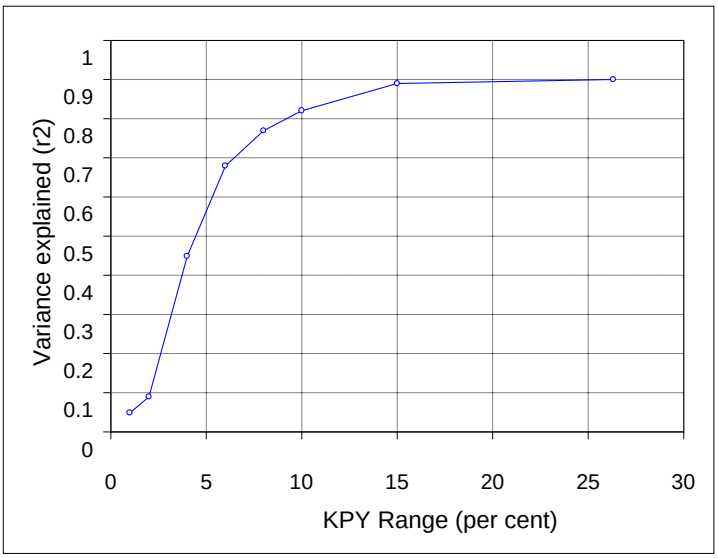

Figure 6. Effect of reducing calibration range on the variance explained

Confidence intervals

How accurate is an individual KPY value, either actual or predicted? The number of samples (data points) required to estimate the true mean of a population is a function of the variance among individual samples from that population (Downes *et al.* 1997). If laboratory KPY determination has a standard deviation of 0.5 per cent, then one can have 95 per cent confidence that the true value is between ±1.96 standard deviations (i.e. within a range of ~ ±1 per cent) of a single individual value obtained from the laboratory. To reduce the confidence range to ~ ±0.5 per cent requires the laboratory KPY analysis procedure to have a standard deviation of ~0.25 per cent.

The practical application of the various NIR assessment methods depends to a large extent on the required accuracy and precision of the predicted values. Some studies may need only know the relative ranking among individual or family means (Raymond *et al.* 2001 a, b), in which case accuracy is less important than precision. However if the actual KPY of a given tree or stand over time is needed, than accuracy is also important (that is to say, bias should be minimal and not change with successive measures over time).

When comparing NIR predictions with laboratory KPY data, the variance in the laboratory KPY values needs to be considered. Table 9 indicates the 95 per cent confidence intervals for each calibration developed in this study. For example if the actual KPY data has a standard deviation of 0.5, one can be confident that any given value is within ± 1 per cent of the true value, 95 per cent of the time. By comparison, individual predictions obtained from calibration set one to seven will be within ± 2.65 per cent of the true value, 95 per cent of the time. This range is large, even though the

r^2 is relatively high at 0.91. If the user specifies that the confidence interval required is ±1 per cent, one can be confident that predictions will be in this range 54 per cent of the time.

Table 9. Probabilities of accurate individual predicted KPY values

Calibration set	Cross validation (r^2)	RMSECV	95 per cent confidence interval	Probability (±1.0 per cent) (per cent)	Probability (±0.5 per cent) (per cent)
1-2	0.93	1.19	2.33	60	33
1-3	0.92	1.28	2.51	57	31
1-4	0.9	1.33	2.61	55	30
1-5	0.91	1.37	2.69	54	29
1-6	0.91	1.35	2.65	54	29
1-7	0.91	1.35	2.65	54	29

If traditional chemical pulping analyses can determine KPY to within ± 1 per cent, 95 per cent of the time and to ± 0.5 per cent 66 per cent of the time, with what level of confidence can one predict KPY within this range? Predicted KPY from spectra obtained from sets one to seven, can be expected to be within ± 1 per cent of the true value, only 54 per cent of the time. This is a function of the variance explained in the cross-validation model and does not take into account the bias that may be evident when applied to an independent prediction sample set. Caution should be exercised in comparing these intervals with those used for industry standard wet chemical methods as no reliable estimate of the actual standard deviation of this method was available. This would require the replicate pulping of the same sample of chips to obtain a true standard deviation value over time.

Conclusion

NIR predictions of KPY will never replace traditional methods as NIR calibrations require actual data for both calibration development and prediction validation. In addition, there will always be a demand for determining the actual value, rather than a prediction of the actual value. However for many applications, NIR complements and enhances the cost-effectiveness of traditional pulping, especially where non-destructive evaluation (NDE) is required. Certainly for purposes of ranking families/clones, NIR offers the ability to screen large numbers of individuals in a non-destructive and cost-effective manner.

The remainder is more discussion than conclusion. We envisage a need to increase sample numbers in the calibration to at least several thousand, in order to maximise the robustness of KPY prediction. In order to be truly global for all eucalypts we need to extend the species range to include more sub-tropical and tropical species across a wide range of sites and climatic conditions.

Alternatively if the demand is for *E. globulus* and *nitens* only, a greater representation of these species should improve prediction accuracy.

The potential to improve the model using PLS 2 (CAMO 1998), utilising variance in chip basic density and, if possible, cellulose content, needs to be explored. Including these properties should allow more spectral variance within the sample set to be explained, thereby producing more robust predictions. Chemical determination of cellulose might provide a cheaper way of checking NIR-predicted KPY if this is accepted as a suitable surrogate.

A further development will be to shift the predictions from a wood meal sample to a woodchip sample, thereby reducing the cost of sample preparation by avoiding the need to grind samples, which currently costs at least AUD $10 per sample. Additional work is underway to explore the ability to predict KPY directly from solid wood samples (i.e. increment cores) both green and air-dry using portable, in-field instruments (Downes *et al.* 2006).

Acknowledgements

We acknowledge the use of pulping data provided by Nigel England, of Oji Paper Co. Ltd plantations in data set four. Data set one and five were obtained from a project co-founded by Ensis, the Forest and Wood Products Research and Development Corporation and industry partners. Data set eight was made available by Ana Gabriela Monnerat Carvalho Bassa, Votorantim. Data set nine was made available by the Des Stackpole and Brad Potts of the CRC for Forestry.

References

Bruker (2005) OPUS 5.5 © (Bruker Optik: Ettlingen, Germany)

Camo (1998) The unscambler - user manual. (CAMO ASA: Trondheim, Norway)

Clarke C (1995) Variation in growth, wood, pulp and paper properties of nine eucalypt species with commercial potential in South Africa. PhD, University College of North Wales, Bangor, UK.

Downes GM, Hudson IL, Raymond CA, Dean GH, Michell AJ, Schimleck LS, Evans R, Muneri A (1997) 'Sampling plantation eucalypts for wood and fibre properties.' (CSIRO Publishing: Melbourne, Australia)

Downes GM, Evans R, Schimleck LR, Fritts HC (2000) The commercial cambium: understanding the origin of wood property variation. In 'Cell and Molecular Biology of Wood Formation.' (Eds RA Savidge, JR Barnett, R Napier) pp. 325-336. (BIOS Sci.: Oxford) ISBN 1 85996 123 1

Downes GM, Meder R, Luff R, Catela F (2006) 'Quality assessment of Australian woodchips Project 4: Rapid field-based assessment of pulpwood quality in standing trees.' CSIRO FFP Client Report 1658, FWPRDC Project PN04.2008

Downes GM, Worledge D, Schimleck LR, Harwood CE, French JRJ, Beadle CL (2006) The effect of growth rate and irrigation on the basic density and Kraft pulp yield of *Eucalyptus globulus* and *E. nitens*. *New Zealand Journal of Forestry* **51**, 13-22.

Greaves BL, Borrahlo NMG, Raymond CA (1997) Breeding objective for plantation eucalypts grown for production of Kraft pulp. *Forest Science* **43**, 465-472.

Kaye W (1954) Near-infrared spectroscopy. 1-Spectral identification and analytical applications, *Spectrochimica Acta* **6,** 257.

Martens H, Næs T (1989) Multivariate Calibration. (John Wiley and Sons, Chichester: UK)

Martens H, Næs T (1984) Multivariate calibration. 1. Concepts and distinctions, *Trends Anal. Chem.* **3**, 204-210.

Michell AJ (1995) Pulpwood quality estimation by near-infrared spectroscopic measurements on eucalypt woods, *Appita Journal* **48**, 425.

Michell AJ, Schimleck LR (1998) Developing a method for the rapid assessment of pulp yield of plantation eucalypt trees beyond the year 2000, *Appita Journal* **51**, 428

Næs T, Isaksson T, Fearn T, Davies T (2004) A user-friendly guide to multivariate calibration and classification. (NIR publications, Chichester, UK)

Raymond CA, Schimleck LR, Muneri A, Michell AJ (2001a) Non destructive sampling of *Eucalyptus globulus* and *E. nitens* for wood properties III. Predicted pulp yield using near infrared reflectance analysis. *Wood Science and Technology* **35,** 203-215.

Raymond CA, Schimleck LR, Muneri A, Michell AJ (2001b) Genetic parameters and genotype by environment interactions for pulp yield predicted using near infrared reflectance analysis and pulp productivity in *Eucalyptus globulus*. *Forest Genetics* **8**, 213-224.

Schimleck LR, Workman JJ Jr, (2004) Analysis of timber and paper. In 'Near-infrared spectroscopy in agriculture, A Tri-Societies Monograph'. (Eds C Roberts, JJ Workman, J Reeves) pp. 635-646. American Society of Agronomy, Crop Science Society of America and Soil Science Society of America. (Madison: Wisconsin, USA)

Schimleck LR (2007) Near infrared spectroscopy: a rapid, non-destructive method for measuring wood properties and its application to tree breeding. *New Zealand Journal Forestry Science* (submitted)

Schimleck LR, Michell AJ (1998) Determination of within-tree variation of kraft pulp yield using near-infrared spectroscopy. *Tappi Journal* **81**(5), 229-236.

Schimleck LR, Raymond CA, Beadle CL, Downes GM, Kube PD, French J (2000) Applications of NIR spectroscopy to forest research, *Appita Journal* **53**, 458.

Shenk JS, Workman JJ Jr, Westerhaus O (1992) Application of NIR spectroscopy to agricultural products, In 'Handbook of Near-Infrared Analysis' (Eds. Burns DA, Ciurczak EW) pp. 385-386. (Marcel Dekker Inc.: New York)

So C-L, Via BK, Groom LH, Schimleck LR, Shupe TF, Kelley SS, Rials TG (2004) Near-infrared (NIR) spectroscopy in the forest products industry. *Forest Products Journal* **54**(3), 8-18.

Thomas EV (1994) A primer on multivariate calibration. *Analytical Chemistry* **66**, 795A.

Tsuchikawa S (2007) A review of recent near-infrared research for wood and paper. *Applied Spectroscopy Reviews* **42**, 43-71.

Wright J, Wearne R, Clarke N, Brimmer P, French J (2003) 'Quality assessment of Australian woodchips Project 3: rapid assessment of the pulpwood quality of export chips.' CSIRO Forestry and Forest Products Client Report 1296.