High-value wood resources

Technical Report 212

Developing and assessing a woodmeal calibration for the Polychromix Phazir™

G Downes

CRC for Forestry
Researching sustainable forest landscapes

Technical Report 212

Developing and assessing a woodmeal calibration for the Polychromix Phazir™

G Downes[1, 2]

Public report

[1] CRC for Forestry
[2] CSIRO Ecosystem Sciences

June 2011

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

www.crcforestry.com.au

Contents

Summary

The use of hand-held portable near-infrared (NIR) spectrometers such as the Polychromix Phazir™ may provide an affordable method for forest managers to obtain predictive measures of cellulose content and Kraft pulp yield in standing trees from various types of wood samples. To be useful for businesses that do not have NIR-related capability in house, the method must give precise predictions using a simple sampling protocol requiring minimal knowledge of the technology. A first step is to determine how well the hand-held spectrometer can predict wood properties from the same types of samples (i.e. woodmeal) that are known to yield successful calibrations for predicting wood properties using more expensive laboratory NIR instruments.

This report describes the development and testing of a model for predicting pulp yield and cellulose content in air-dry woodmeal from *Eucalyptus nitens* from spectra collected using the portable NIR. The method allowed the samples to be measured without removing them from the plastic bags in which they were stored. Although developed using *E. nitens* samples only, the model was able to predict 80–90% of the variance in samples from two other eucalypt species (*E. pellita* and *E. globulus*), as well as from an independent set of *E. nitens* samples.

The model is a starting point for the development of a more broadly-based model (representing more sites and species) that allows industry partners of the CRC for Forestry to gain experience and confidence in the technology. The model accommodates the chemistry of the plastic bags and allows the pulp yield and cellulose content of woodmeal to be predicted by the Phazir™ whether the woodmeal is in the plastic bag or not.

Introduction

The Polychromix Phazir™ is a multi-purpose, hand-held NIR spectrometer suitable for many chemometric measurements. It is finding increasing use in a range of commercial applications from raw materials inspection to chemical identification. It is of interest to forest managers and tree breeders because of its portability (and therefore potential for field use) and relatively low cost, making its purchase more feasible for individual companies.

The potential of the Phazir™ for measuring Kraft pulp yield (KPY), basic density and dry matter content in standing trees was demonstrated across 40 sites in Tasmania, Victoria, South Australia and Western Australia (Downes et al. 2009b; Meder et al. 2009). Measurements, based on spectra collected from green chainsaw dust, and green and dry increment cores explained commercially useful levels of variance when comparing among trees and sites.

Meder et al. (2010) used the Phazir™ to predict KPY in woodmeal presented to the instrument through the base of glass vials, analogous to that used for laboratory NIR instruments (Downes et al. 2009a; Downes et al. 2010). The calibrations were relatively poor, but improved markedly when spectra were collected directly from the woodmeal. The illuminating light source of the Phazir™ is considerably weaker than that of the laboratory instrument, and the vial absorbed too much of the light.

Ziploc™ bags are often used in the field to transport wood samples to the laboratory, as well as for long-term storage of dry samples; consequently, for field use, obtaining spectra from samples already in these bags is preferable. Initial tests indicated that collecting spectra from green wood samples through the Ziploc™ bag was possible, but often gave mixed results. It has been difficult to determine the reasons for poor performance of some trial calibration models of green drill frass (GDF), but might include:

- varying sample moisture content
- excessive coarseness of the particle size, resulting in unrepresentative spectra
- spectral properties of the Ziploc™ bag affecting the calibration
- general variability experienced in developing a multisite and species calibration.

Because we know that NIR prediction of CC and KPY from woodmeal samples works well using laboratory NIR instruments (Downes et al. 2009a; Downes et al. 2010), the effect on calibration performance of spectra obtained with and without the Ziploc™ bag can be assessed using woodmeal samples, without the other potentially confounding effects of GDF.

The two major objectives of this study were to determine:

- whether the Polychromix Phazir™ can be used to predict CC and KPY in *Eucalyptus nitens* woodmeal with acceptable accuracy and precision
- the extent to which obtaining spectra through Ziploc™ bags affects the predictive ability of calibrations.

To address these objectives, the study sought to answer the following questions:

- Does a calibration model built using samples in Ziploc™ bags explain less variance than a calibration model built using spectra obtained directly from woodmeal?
- Can a calibration model built from spectra obtained directly from woodmeal be used to predict CC and KPY from spectra collected through a Ziploc™ bag?
- Can a calibration be constructed using both spectra obtained with and without the Ziploc™ bag, such that the effect of the Ziploc™ bag is compensated for in the calibration model?
- For a well-mixed woodmeal sample, how many spectra are required to adequately represent the sample?

Methods

Two sets of woodmeal samples from an existing CRC for Forestry project[1] were used. Set 1 included samples from 47 trees that had known, laboratory-determined KPY and CC. Set 2 included a further 32 samples where only laboratory CC was known. NIR-predicted CC and KPY for all samples were determined for each sample from spectra obtained using the Bruker MPA NIR instrument at the University of Tasmania and applying the existing global calibrations (Downes et al. 2009a; Downes et al. 2010). All samples were provided by Gunns Ltd as woodmeal[2].

Phazir™ NIR spectra (935 – 1796 nm) were obtained for each sample both through the Ziploc™ bag and directly from the woodmeal surface (Figure 1). At least 8, and typically more than 12, spectra were obtained from each of the 79 samples, for each presentation type (Ziploc™ and direct) yielding a total of 1768 spectra.

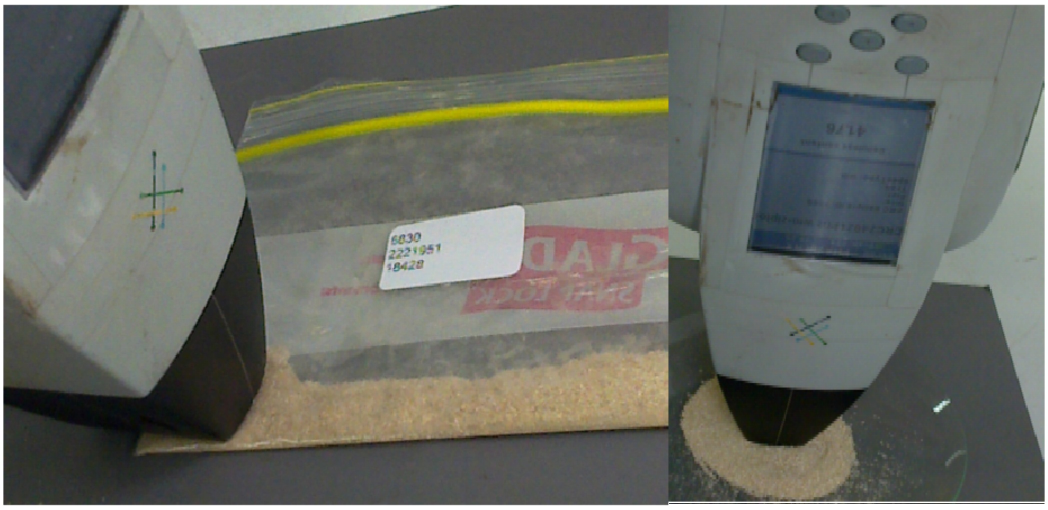


Figure 1. Spectra were collected using the Phazir™ either through the Ziploc™ bag (left) or directly from the woodmeal surface (right)

[1] This project covered the prediction of CC and KPY in 750 samples of *E. nitens* used in the FWPA 07.3025 study investigating whether acoustic wave velocity in standing trees and logs predicts KPY. These trees were also used in a further study with Dr Simon Southerton, CSIRO, developing marker-aided selection for KPY. The laboratory KPY was determined by Gunns Fibre Technologies utilising a single cook per tree (typically at least 3 cooks are used) and estimating the KPY at kappa 18.

[2] Woodmeal samples were delivered in 'GLAD™ Snaplock, or Ziploc™, bags. It is unknown whether the use of other brands of plastic bags affects the prediction capability.

The study sequence was as follows:

1. Obtain Phazir™ spectra from set 1 samples through the Ziploc™ bag. Remove spectra that look atypical[3].
2. Use these spectra to build a Ziploc™ calibration model against the laboratory CC values using the Polychromix method generator software (Version 3.101 R1) and load it into the Phazir™.
3. Obtain spectra directly from samples in set 1, predicting CC using the Ziploc™ calibration.
4. Compare these predicted values with laboratory CC values.
5. Use the latter to build a calibration model combining Ziploc™ and directly acquired spectra.
6. Predict the values of sample set 2 through the Ziploc™ bag and compare with laboratory CC values.
7. Predict the values of sample set 2 directly and compare with laboratory CC values.
8. Build and describe a suitable model combining sets 1 and 2 for further assessment.

The model in step 5 was built to determine whether it could account for the standard effect of the Ziploc™ bag, allowing the model to be used on spectra regardless of whether they were obtained through the bag or directly from the woodmeal. This calibration identified nine factors[4] as optimal. The effect of constraining the calibration model to five factors was assessed and the resultant four calibrations combined into a single script and uploaded into the Phazir™. This allowed predictions for each of the four calibrations to be made at the time the spectra were collected from the unknown samples.

Phazir™-predicted values were compared with laboratory CC values, but comparisons were also made with laboratory KPY, and NIR-predicted CC and NIR-predicted KPY using the spectra already obtained on the laboratory Fourier transform–NIR system (MPA, Bruker, Ettlingen, Germany) and the global calibrations developed for it (Downes et al. 2009a; Downes et al. 2010).

[3] On occasion the Phazir™ may be removed from the sample slightly early or the trigger starting spectra collection pressed too early resulting in spectra that are qualitatively different from correctly acquired spectra.

[4] 'Factors' are used in multivariate statistical methods such as principal components or partial least squares analysis to reduce the number of variables (e.g. all the wavelengths in the spectrum) to a more manageable number by identifying factors or components that explain the most variance between the spectra and the calibration data.

Results and discussion

The laboratory values for the two sample sets (Table 1) indicate moderate ranges of CC and KPY in the two sets.

Table 1. Laboratory data descriptions for the two woodmeal sample sets

	Set 1	**Set 2**
Number of samples	47	32
Laboratory KPY (%)	53.9	–
Max.	57.7	–
Min.	50.3	–
Range	7.4	–
Laboratory cellulose content (%)	41.7	41.0
Max.	44.5	44.2
Min.	38.7	37.6
Range	5.9	6.6
Laboratory NIR-predicted KPY (%)	52.8	51.6
Max.	56.1	54.5
Min.	49.9	48.5
Range	6.2	6.0
Laboratory NIR-predicted cellulose content (%)	42.0	41.3
Max.	44.4	44.6
Min.	39.6	37.9
Range	4.8	6.7

KPY = kraft pulp yield; NIR = near infrared

Ziploc™ calibration

The spectra obtained from set 1 through the Ziploc™ bags were used to build a calibration model (Figure 2) that had a cross-validation r^2 of 0.83 and a root mean standard error of cross-validation (RMSECV) of 0.56%. The software identified six factors as optimal. Some individual spectra (44 of 630) were rejected as outliers based on the Hotelling T2 statistic and visual assessment. Typically, these exhibited an unusual form or variability. Each of the 49 samples was still represented by a minimum of 10 spectra.

Spectra were pre-processed using the Savitsky-Golay first derivative with three smoothing points. Although a variety of other pre-processing options were tried, they were not found to improve the calibration. Optimising the wavelength range identified the range 1102–1749 nm as optimal.

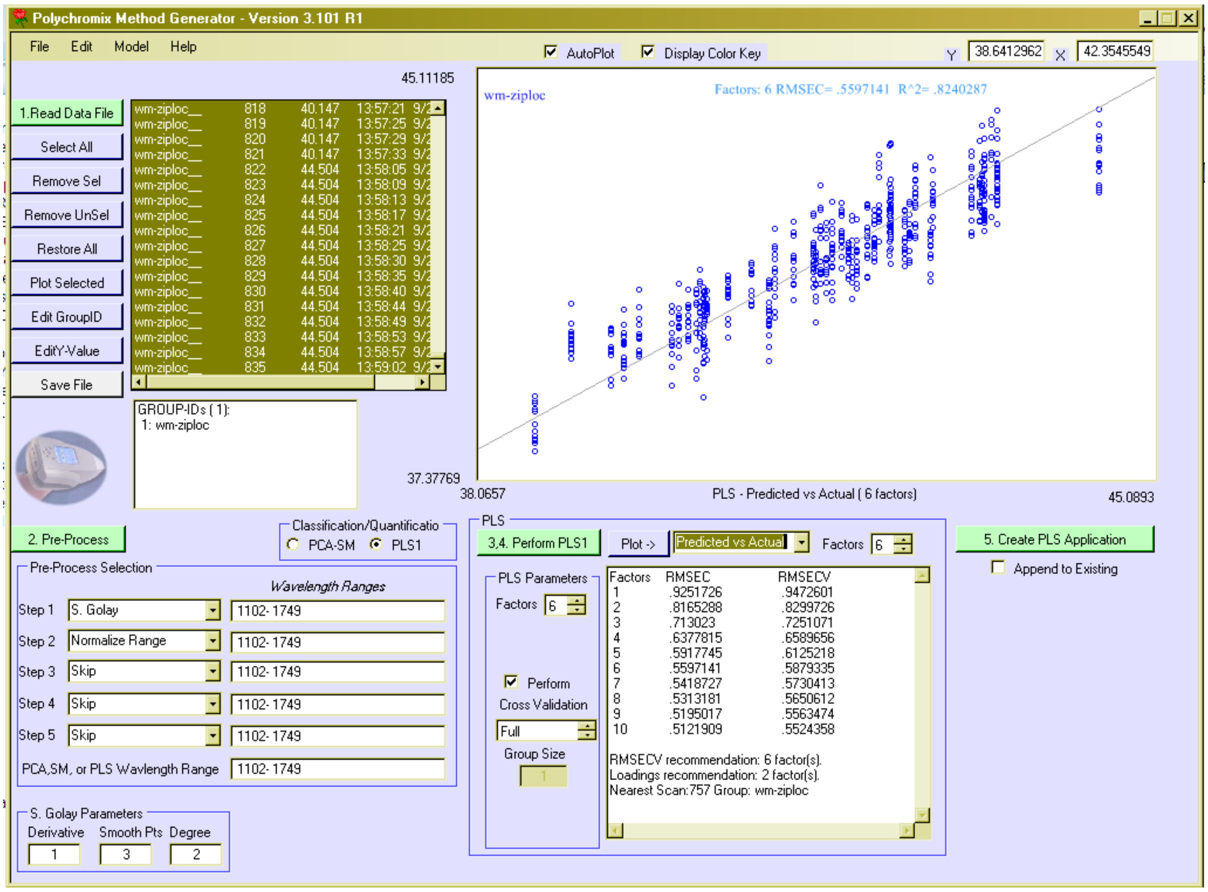


Figure 2. Screen image of the calibration model; calibration r^2 = 0.824; RMSEC = 0.560, cross-validation r^2 = 0.806; RMSECV = 0.588

Spectra from set 1 obtained directly from the woodmeal surface were used to predict CC using the calibration model shown in Figure 3. Cellulose content predictions were reasonably precise (r^2 = 0.78) but with a consistent over-prediction bias of approximately 2% CC. The prediction of laboratory KPY was poorer (r^2 = 0.62)[5].

[5] The laboratory KPY values for this sample set were determined from a single cook per sample so individual sample values would be expected to have lower precision of estimation compared to the conventional method involving three or more cooks per sample.

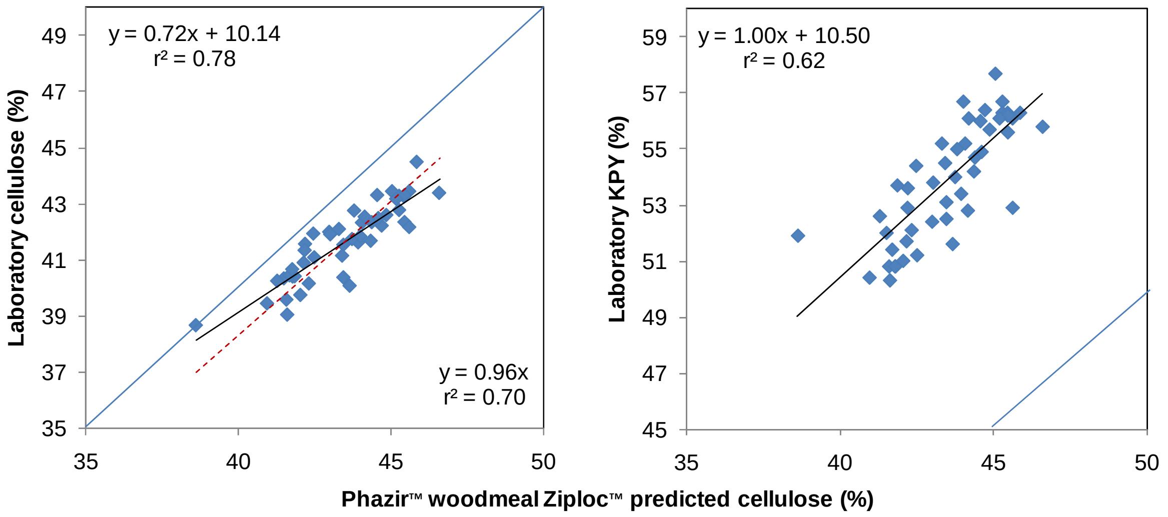


Figure 3. The Phazir™ calibration based on Ziploc™ spectra accounted for 78% of the variance in the laboratory cellulose content (left) and 62% of the laboratory KPY (right) using spectra obtained directly from the woodmeal

Direct and combined calibration

Pre-processed spectra were subjected to PCA (Figure 4), which indicated that the spectra grouped into two discrete clusters defined by type (direct and Ziploc™). This helped to identify spectral outliers (95 out of 1768), which were sequentially removed from the dataset.

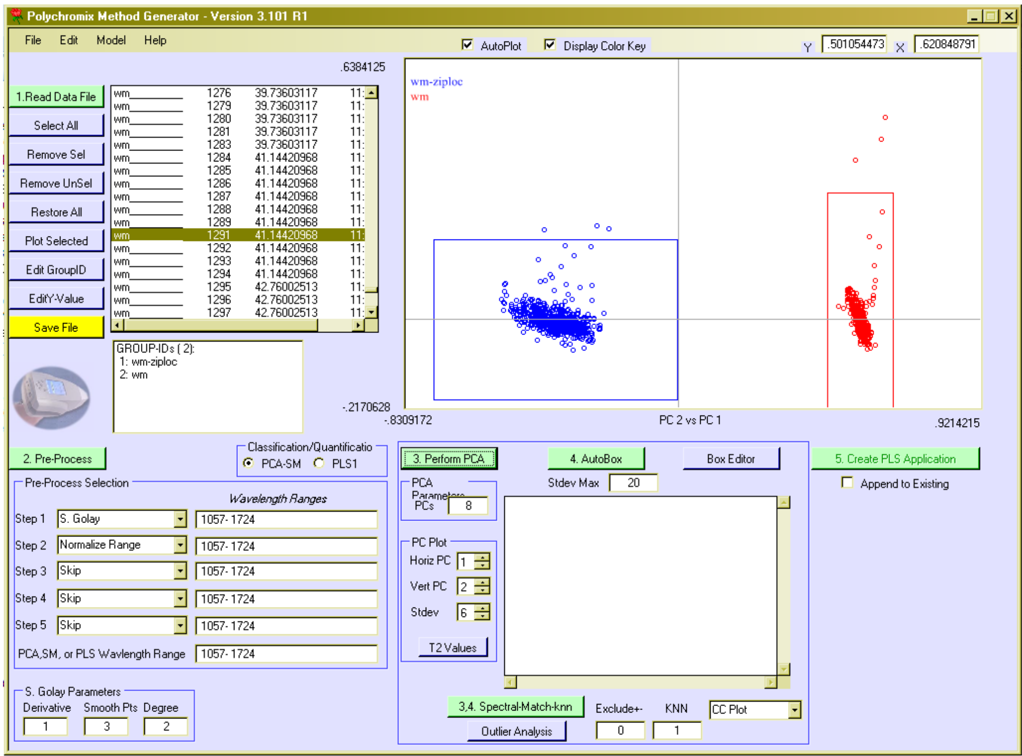


Figure 4. User interface of the PMG software (supplied by Polychromix) showing PCA analysis of the spectra discriminated between spectra obtained directly from woodmeal and those obtained through the Ziploc™ bag

The ‘direct’ spectra were then added to the ‘Ziploc™’ spectra and a combined calibration prepared. From the combined spectra set, a calibration model using first derivative spectra between the wavelength range 1057–1724 nm had a cross-validation r^2 of 0.81 with a RMSECV of 0.59. The software identified nine factors as optimal based on the change in RMSECV. To assess the effect of different factors, a secondary model was produced with the same features but with the number of factors constrained to five.

Another calibration model was developed using only the spectra obtained directly from woodmeal, which used the wavelength range 1148–1789 nm, had seven factors, a cross-validation r^2 of 0.83 and an RMSECV of 0.57.

In all, four separate calibrations, based on spectra from the 47 samples in set 1, were produced for the prediction of CC:

- Ziploc™ spectra only
- direct spectra only
- combined Ziploc™ and direct spectra—nine factors
- combined Ziploc™ and direct spectra—five factors.

These were combined into a single script that predicted CC four times from each spectrum.

Assessing calibration performance

Spectra from samples in set 2 were collected through the Ziploc™ bag and the CC predicted using the above four calibrations. The predictions made by all four calibrations explained more than 80% of the variance in the laboratory cellulose data. The Ziploc™ calibration explained the most variance (90%) but over-predicted by, on average, 2%. The direct calibration explained 82% but on average under-predicted by 4%. The combined calibrations explained 84% and 88% of the variance for the five-factor and nine-factor calibrations, respectively. These exhibited relatively smaller bias, over-predicting by 1.1% and 1.3%, respectively.

Unexpectedly, predictions made from spectra collected directly from the woodmeal (Figure 5a) tended to explain slightly less variance in the laboratory data than did predictions based on the calibration prepared from spectra obtained through the Ziploc™ bags (Figure 5a). However, the predictions made by the direct calibration had an average bias of only 0.5%. Overall, the combined spectra model, combining spectra taken directly and through Ziploc™ bags, had a relatively small bias (Figure 5c and 5d), with the nine-factor model explaining 11% more variance.

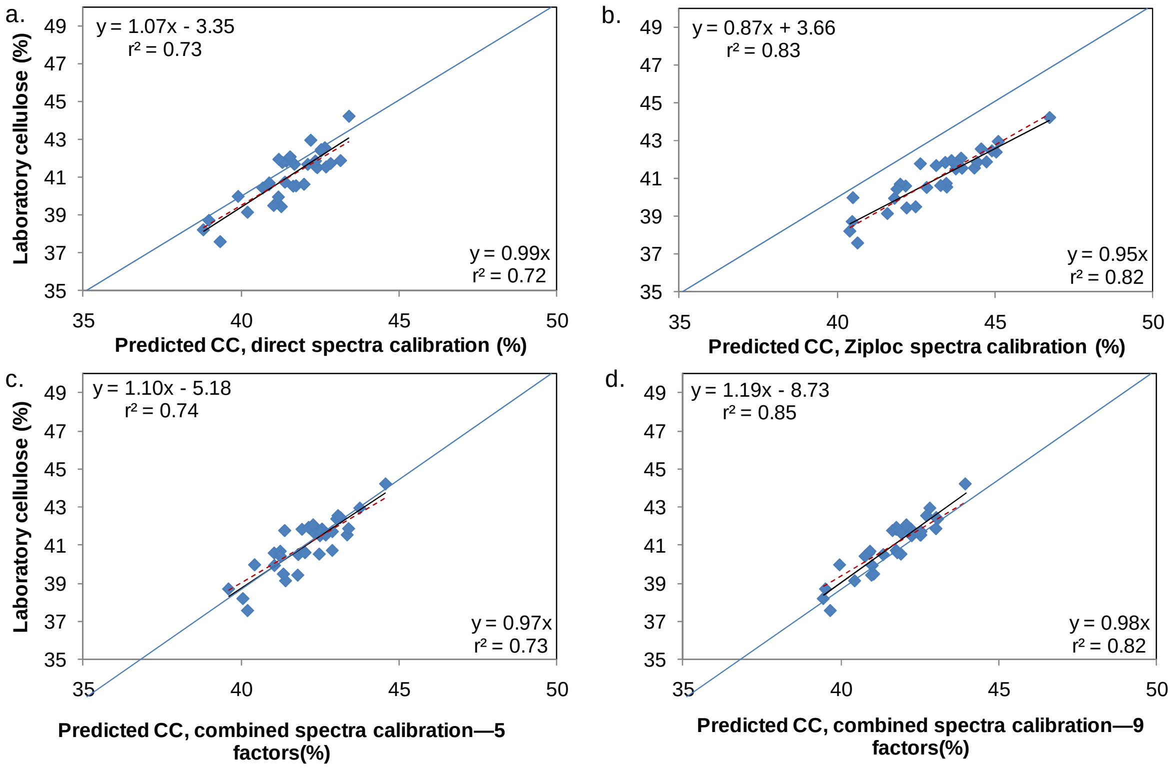


Figure 5. Cellulose content (CC) predicted from spectra obtained directly from woodmeal (a) explained less variance in the laboratory values compared to those obtained through Ziploc™ bags (b) but with less bias; combining both spectra types was able to explain useful variance in the spectra using models with five factors (c) and nine factors (d)

As an additional comparison, the Phazir™-predicted values were regressed against laboratory KPY (Figure 6a) and CC values (Figure 6b) obtained from laboratory analysis and predicted using the laboratory NIR multisite and species calibration (Downes et al. 2010). The comparison with the predictions from the Bruker KPY NIR values were both accurate (bias = 0.4%) and precise, explaining 6% less variance (r^2 = 0.79) than the comparison with the laboratory CC values (r^2 = 0.85, Figure 6b).

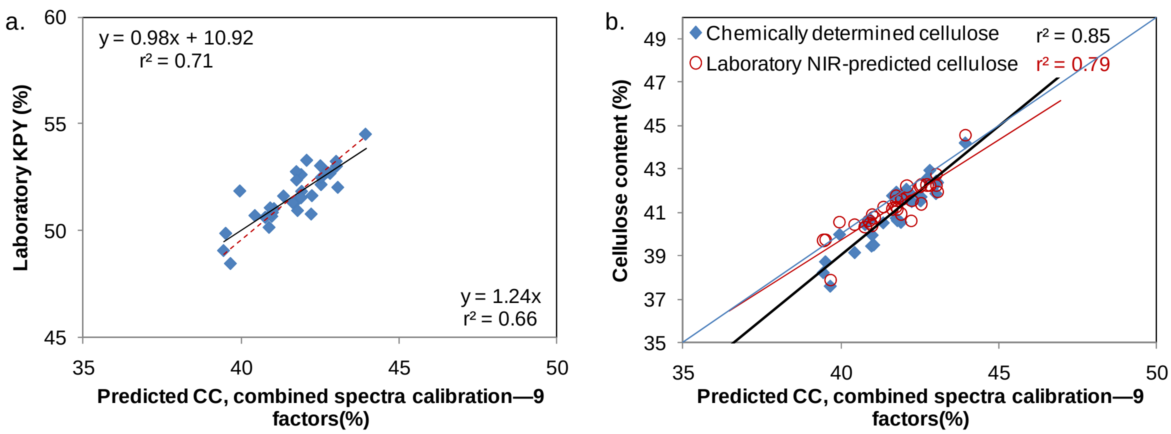


Figure 6. Cellulose content (CC) predictions from the Phazir™ set 1 calibration model explained acceptable levels of variance in spectra from the set 2 woodmeal samples for laboratory KPY (a) and both laboratory-analysed and Bruker NIR-predicted CC (b)

Number of spectra required per sample

In these studies, between 7 and 19 spectra were collected from each woodmeal sample. For each individual spectrum a predicted CC value was generated.

Based on the approach described in Downes et al. (1997), the number of spectra required to adequately predict the woodmeal value to a given level of accuracy was determined as follows. The CC value for each individual spectrum for each sample in set 2 was predicted using the Phazir™ set 1 calibration. The variance explained (r^2) in the laboratory values by a mean CC value obtained from averaging the predicted values for varying numbers of spectra was calculated and plotted in Figure 6. It was estimated that, in general, a single spectrum would adequately represent most samples. Occasional samples had greater variability and required up to four or five spectra. Comparisons between spectra used to estimate the sample mean and the laboratory value indicate that one to three spectra per woodmeal sample is sufficient (Figure 7). Given the relative speed of acquisition, we suggest three spectra per sample be taken routinely from woodmeal samples.

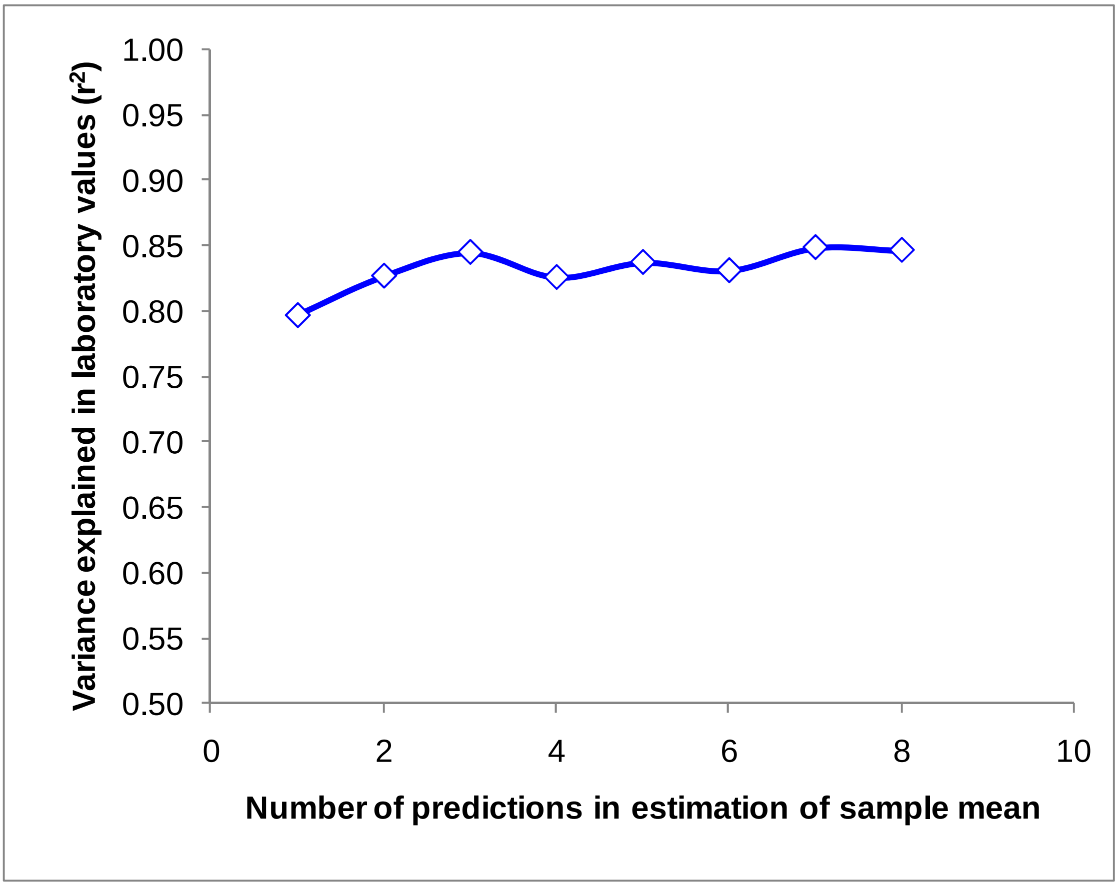


Figure 7. Comparisons between the laboratory and NIR-predicted cellulose content values, where the latter were calculated based on varying number of spectra per sample, indicated that 1–3 spectra would optimally represent the sample

Combined calibration

The spectra from the two sample sets and both spectral types were combined into a single calibration, in which the two spectral types were clearly identified by PCA (Figure 4). The PMG software recommended that a calibration using four factors, across a wavelength range of 1039–1683 nm was optimal. This had a cross-validation r^2 of 0.71 and an RMSECV of 0.74.

The performance of this calibration was applied to an independent set of samples, including other eucalypt species (Figure 8). The four-factor Phazir™ calibration described above explained approximately 60% of the variation in CC in the combined sample set, but was markedly better when applied to individual species. This suggests that a combined calibration built using samples from a range of species may improve predictions overall by providing a wider representation of wood chemistry in the calibration model.

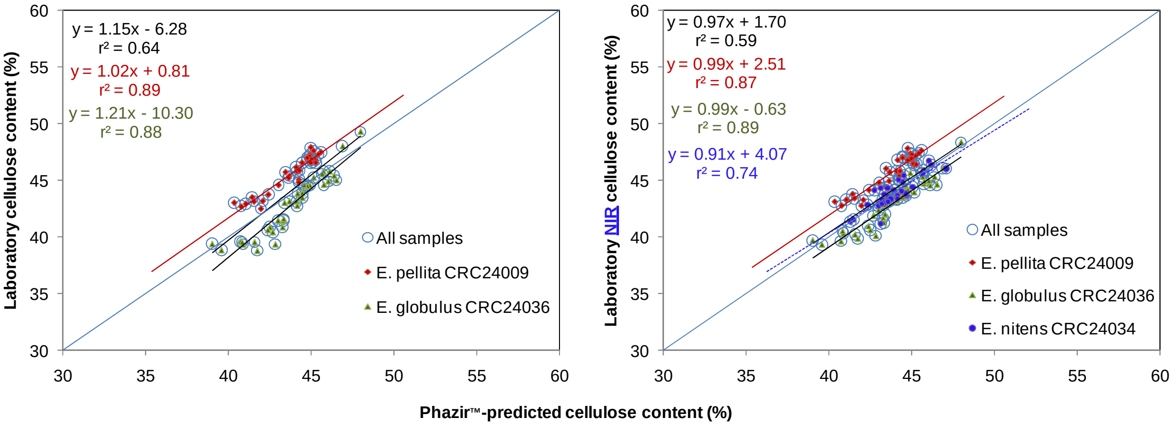
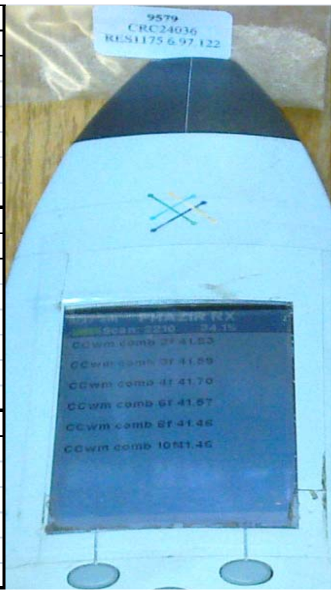


Figure 8. Comparisons of the laboratory (left) and NIR-predicted (right) cellulose content values and those predicted from the four-factor Phazir™ calibration based on 79 *E. nitens* samples (sets 1 and 2) combining spectra obtained directly and through a Ziploc™ bag

As an additional comparison, the effect of calibrations with varying numbers of factors were assessed by building calibrations restricted to 2, 3, 4, 6, 8 and 10 factors into a single application, and uploaded into the Phazir™ (Table 2). Increasing the number of factors to more than four gave little or no increase in the explained variance.

Table 2. Increasing the number of factors beyond four had little effect on the variance explained. Combining calibrations into a single application (script) allows a single spectrum to be used to predict a property multiple times, or have multiple properties predicted (see image on right)

	Variance explained against laboratory cellulose content			
No. factors	All samples (N=68)	*E. pellita* (N=32)	*E. nitens* (N=33)	*E. globulus* (N=36)
2	0.69	0.84		0.92
3	0.66	0.89		0.91
4	0.64	0.89		0.88
6	0.61	0.91		0.89
8	0.63	0.91		0.88
10	0.63	0.91		0.88
	Variance explained against laboratory NIR-predicted cellulose content			
	All samples (N=101)			
2	0.58	0.82	0.73	0.90
3	0.58	0.88	0.75	0.91
4	0.59	0.87	0.74	0.89
6	0.57	0.89	0.77	0.89
8	0.59	0.89	0.78	0.89
10	0.59	0.88	0.78	0.89
	Variance explained against laboratory NIR-predicted KPY			
2	0.62	0.70	0.73	0.82
3	0.66	0.79	0.76	0.85
4	0.62	0.83	0.77	0.86
6	0.64	0.86	0.80	0.86
8	0.63	0.85	0.81	0.86
10	0.63	0.85	0.81	0.86



KPY = Kraft pulp yield; NIR = near infrared

Conclusions

The Phazir™ was able to predict CC and KPY in woodmeal with commercially useful levels of precision and accuracy. More than 60–70% of variance in CC and KPY among samples was accounted for in blind predictions of an independent sample set (set 2), with ranges of 5–7 % in CC and KPY). Obtaining spectra through Ziploc™ bags had no obvious deleterious effect on calibration performance, as shown by the answers to the following specific questions that were investigated:

- Does a calibration model built using samples in Ziploc™ bags explain less variance than a calibration model built using spectra obtained directly from woodmeal?

Calibrations built from Ziploc™ spectra appear to predict equally as well as calibrations built from spectra collected directly, with no reduction in variance explained.

- Can a calibration model built from spectra obtained directly from woodmeal be used to predict CC and KPY from spectra collected through a Ziploc™ bag?

Yes, but with a tendency to under-predict. Conversely, a calibration built from Ziploc™ spectra equally tended to over-predict the values from both direct and Ziploc™ spectra.

- Can a calibration be constructed using both spectra obtained with and without the Ziploc™ bag, such that the effect of the Ziploc™ bag is compensated for in the calibration model?

Yes, calibrations that combined the spectra types lost a small amount of precision (percentage of variance explained) but performed equally well with both spectra types.

- For a well-mixed woodmeal sample, how many spectra are required to adequately represent the sample?

Comparisons between spectra used to estimate the sample mean and the laboratory value indicate that three spectra per woodmeal sample are optimal.

References

Downes, G.M., N. Ebdon, C. Hicks and R. Meder 2009a. Developing and evaluating a multi-site and species NIR calibration for the prediction of Kraft pulp yield in eucalypts. *Southern Forests*. 72:151.

Downes, G.M., I.L. Hudson, C.A. Raymond, G.H. Dean, A.J. Michell, L.S. Schimleck, R. Evans and A. Muneri 1997. Sampling plantation eucalypts for wood and fibre properties. CSIRO Publishing, Melbourne, Australia

Downes, G.M., R. Meder, N. Ebdon, D. Menz and C. Hicks 2009b. FWPA PN07.3027: Quality Assessment of Australian woodchips Project 5: Field-based application of acoustic velocity and NIR as predictors of pulp quality across site and species, p. 33.

Downes, G.M., R.A. Meder and C.E. Harwood 2010. A multi-site, multi-species calibration for the prediction of cellulose content in eucalypt woodmeal. *J. Near Infrared Spectrosc.* . 18:381–387.

Meder, R., D. G.M., J. Brawner and N. England 2010. Understanding within tree variation of KPY to establish protocols for in-field sampling for NIR prediction. In Appita 2010, Melbourne.

Meder, R., D. G.M. and N. Ebdon 2009. Experiences in field assessment of Kraft pulp yield in plantation Eucalypts. In Tappi, Memphis, Tennessee.