Progress in the use of near-infrared absorption spectroscopy as a tool for the rapid determination of pulp yield in plantation eucalypts

By Nelson L. Sefara, Denise Conradie, and Philip Turner

Forestry & Forest Products Research Centre
CSIR
P O Box 17001
4013 Congella
South Africa

Presented at IUFRO Conference “Forest Genetics for the next Millenium”,
IUFRO Working Party 2.08.01
Durban, South Africa
8 – 13 October 2000

Presented at TAPPSA Conference “African Paper week ‘2000 and beyond’”
Durban, South Africa
17 – 20 October 2000

Published in TAPPSA Journal, November 2000, pp 15-17.
Abstract

The feasibility of utilizing near-infrared spectroscopy as a tool for predicting pulp yield in *E. grandis* Tag5 clonal material is demonstrated. Three sampling methodologies, namely wedges, chips and sawdust, were evaluated in order to determine the method that most accurately reflects pulp properties of the whole tree. The results from this study showed that good correlations between NIR spectra and screened pulp yields were achieved for all three sampling methods. NIR spectra of wedges had the highest $R^2$ value of 0.898. Based on these calibration models, higher prediction ability was achieved with normal spectra. The regression line of the prediction captured 83.6% of the variation, which corresponds to a mean residual of 0.666. Furthermore, the results showed that the average NIR predicted pulp yield for the samples in the prediction set was very similar to the actual pulp yield determined by the digester.

Keywords: Near-infrared spectroscopy, pulp yield, *E. grandis* Tag 5 clonal material
Introduction

Near-infrared spectroscopy has been recognized as a powerful analytical technique for rapid determination of various constituents in many agricultural and raw materials. The approach involves the acquisition of a reflectance spectrum after near-infrared radiation is allowed to penetrate a ground sample of the analyte. The resulting NIR spectral information is then calibrated against measurements obtained using conventional analytical techniques using linear regression.

In recent years, there has been a growing interest in the development of NIR applications to the pulp and paper industry (Birkett et al., 1988a, Easty et al., 1990, Michell, 1995). Work on a variety of eucalypts has been focused on developing NIR models to quantify and characterise chemical components of wood such as lignin and cellulose (Garbutt et al., 1992) and also for pulp properties (pulp yield, Kappa number) (Easty et al., 1990, Birkett et al., 1988b).

The potential application of NIR spectroscopy to predict pulp yield from ground wood samples of eucalypts has been reported in the literature (Michell et al., 1995). Birkett and Gambino demonstrated that NIR spectra can be used to estimate pulp yield of E. grandis timber (Birkett et al., 1988a). Using a series of filters at different wavelengths, correlations of calibration as high as 0.91 were achieved. Wright and coworkers obtained correlation coefficients of 0.88 using a combination of four wavelengths (Wright et al., 1990). However, the correlation coefficient for the prediction was fairly poor. Work on E. globulus samples by Michell (Michell, 1995) showed that improved correlation coefficients (r = 0.99) could be achieved when full wavelength spectrometers are used. This work was extended to cover other species such as E. nitens and the quality of the models were comparable with those obtained from E. globulus (Schimleck et al., 1996, Schimleck et al., 1997).

The aim of this paper was to assess the utility of NIR spectroscopy to predict pulp yield of E. grandis Tag5 clonal material grown in South Africa. The primary reason for selecting Tag5
material lies in the lower variability of the material compared to seedling material. Three sampling methods, namely, wedges, chips, and sawdust were evaluated in order determine the method that most accurately reflected properties of the whole tree. The sampling methodology with the highest $R^2$ (and lowest standard error of calibration) were evaluated for the ability to predict pulp yield using independent samples that were left out of the calibration.

**Material and Methods**

The samples used for this study were *Eucalyptus grandis* Tag5 clonal material. Site-index classes were identified using Mondi Forests mensuration methods. Samples with a site index of below 15.5 were classed as poor, those between 16 and 18.5 as medium, those between 19 and 21 as medium-good and above 21 as good. This process resulted in four site index classes. From each site index, five trees were sampled from five different ages (5, 6, 7, 8 and 9 years) resulting in a total of 100 trees (4 sites-indices x 5 age groups x 5 trees/age group). Two sets of sampling methods were carried out on the same tree; one set for pulp analysis and the other set for the NIR spectrometer.

For pulp samples, 20 mm thick discs were cut at 1-metre intervals up the height of the tree. The discs were chipped together and pulped in an electrically heated rotating digester using the Kraft process. Chips of individual trees were pulped at an active alkaline charge of 18% and sulphidity of 25%. Pulping conditions were selected to achieve a kappa number of between 20 and 22. Prior to pulping, a 100 g portion of the chipped material was sampled and Wiley-milled into fine particles and screened through a 40 – 60 mesh resulting in sample with a range between 250 and 500 µm in size. A mixed 10 g sample was scanned on the NIR spectrophotometer.

Twenty-millimetre thick wood disks were cut at 5%; breast-height (BH); 15%; 35% and 65% height up the tree. At each height, a wedge (1/4 of the area) was cut and all wedges from the same tree were combined, chipped and Wiley-milled into fine sawdust particles and screened
through a 40 – 60 mesh. A 10 g sample was used for the NIR analysis. During the sawing of
the wedges, sawdust particle from the cuttings were collected. The sawdust for the different
heights up the tree was combined and properly mixed. A 10 g sample of this material was
scanned into the NIRA.

Diffuse reflectance spectra for each sample were acquired using the NIRSystems Model 6500
scanning spectrophotometer. All spectra were collected at 2-nm intervals over the 400 to
2500 nm spectral range. Each spectrum is based on 32 co-added scans. Reflectance (log
(1/R)) spectra belonging to the same sampling method (i.e. wedges, sawdust and chips) were
put together into one file along with the corresponding pulp yield. The data was divided into
the calibration set which was used to develop the model and the performance of the model was
then tested using the validation set. A multivariate partial least squares (PLS) regression
analysis relating the pulp yield and spectra were carried out using the ISI software supplied by
NIRSystems. A cross-validation method was conducted in order to determine the optimum
number of factors (latent variables) needed to describe all the variation in the data. Different
combinations of derivative techniques and spectral range were also evaluated to obtained an
optimised model.

Results and Discussion

Figure 1 shows an average NIR spectrum of the 100 spectra obtained from the wedges method.
Average spectra obtained from sawdust and chips methods displayed spectral features (peaks)
that are similar to those shown in Figure 1. Most of the major peaks are concentrated in the
region between 1100 and 2500 nm and consist of broad bandwidths. The appearance of these
broad bands suggests the presence of underlying overlapping minor peaks. The presence of
underlying small bands can be resolved by applying a second derivative algorithm to the
spectra.

Laboratory measured pulp yield from the digester showed a significant difference in the yield
between sites. The impact of age on pulp yield was very small. Pulp yield for *E. grandis* Tag5 clonal material ranged from 45.40% on the poorest site to 52.40% on the best site, with an average pulp yield of 48.86%. A Partial Least Squares (PLS) linear regression was used to relate the pulp yield of each tree with the corresponding NIR spectra. Models were developed using normal NIR spectra and second derivative to assess the different in the spectral treatment.

Table 1 lists a summary of the calibration results from chips, wedges, and sawdust data samples for normal and second derivative NIR spectra. The $R^2$, standard error of calibration (SEC) and the optimum number of PLS factors used to develop the calibration are listed. For all three sampling methods, optimal number of PLS factors needed to build the model was approximately equal. Models developed using second derivative spectra showed better correlations with pulp yield across the three sampling methods. This occurrence is attributed to the fact that overlapping peaks are better resolved in second derivative, therefore giving a better correlation with pulp yield. Overall, calibration results from the wedges showed higher $R^2$ (and lower SEC) of 0.898 than chips ($R^2 = 0.889$) and sawdust ($R^2 = 0.844$).

Since the wedges sampling methodology yields the calibration model with the highest $R^2$, the calibration was tested for the ability to predict pulp yield using spectra left out in the calibration set. In order to do this, the original calibration data consisting of 100 spectra was randomly separated into individual calibration and validation sets. The calibration set consisted of (80%) of the data set and the remaining 20% were used to test the predictability of the model (built using 80% of the data). The accuracy of the models was evaluated by comparing the predicted pulp yield to the known laboratory values. Calibration models were also developed using both normal and second derivative spectra in order to determine the effect of the spectral treatment on the calibration and prediction.

Table 2 summarises the details of the calibration models for both normal and second derivative spectra. $R^2$ values listed in this table are slightly higher than those from calibrations developed
using the all spectra in the original data set (Table 1). While the number of PLS factors have not changed, improved R² values are obtained. For normal spectra the correlation coefficient increased from 0.887 to 0.907 with a subsequent decline in the SEC from 0.605 to 0.528. A similar trend is also observed for models based on second derivative spectra.

Figure 2 displays two correlation plots corresponding to models developed using normal (top) and second derivative spectra (bottom). Both correlation plots showed a fairly even distribution of points (samples) along the diagonal, indicating that the model is not biased. The actual calibration model for the normal spectra (top plot) is linear and results in R² of 0.907 with a slope of 0.918 and a y-intercept of 4.01. The corresponding model based on second derivative spectra resulted in an improved correlation model with R² of 0.912 and a slope and y-intercept of 0.918 and 4.02, respectively. The results imply that pulp yield is well modelled by the NIR spectra.

In Figure 2, over 85% of the samples in the prediction set show residual (i.e. laboratory measured pulp yield minus NIRA estimated pulp yield) of less than one. The average residual for normal spectra and second derivative spectra was 0.666 and 0.725, respectively. Although, the calibration model developed using normal spectra is slightly poorer than that from second derivative spectra, the former method gave good predictions as depicted by the lower standard error of prediction (SEP). The second derivative calibration gives considerable scatter about the unity line. The correlation line for the prediction data for the normal spectra has of an R² of 0.836 and an SEP of 0.829.

An overall analysis of the samples in the validation set shows that the pulp yields predicted from the NIR are very close to those determined from the digester. The mean pulp yield determined by NIR (49.069) and the mean from the digester (49.053) are virtually the same. This implies that we can predict pulp yield of E. grandis Tag5 clonal material from a wide range of ages and sites to within 0.666 magnitude of the pulp yield determined by the digester. Compared to the error associated with pulp yield determination from the digesters at the CSIR
(0.25 percentage points), this method shows great potential for becoming a commercial screening tool for predicting pulp yield. The results from this study demonstrate clearly how accurate the NIRA can be at predicting the pulp yield of a stand of trees.

**Conclusions**

The feasibility of developing calibration models that relate pulpwood properties and NIR spectra of *E. grandis* was demonstrated. The results from these studies have shown that the pulp yield can be predicted with relatively high precision from NIR spectra of wedges and chips. The average percent error in the predictability of this method is very similar to that obtained from digesters. For *E. grandis* Tag 5 clonal material, we can conclude that NIR spectroscopy is accurate enough to replace digesters for routine pulp yield predictions.

**Acknowledgments**

Mondi Forests is thanked for providing the materials for this study. This work would not have been possible if it wasn’t for the help of Jeff Male, Annalie Havenga, Edward Khumalo, Mike Letsoalo and Daniel Gama. This work forms part of the output generated from the research cooperative on Eucalypt pulp quality, sponsored by Mondi, Sappi and CSIR.
References


Table 1: Results from the calibration models obtained from wedges, chips and sawdust samples.

<table>
<thead>
<tr>
<th>Sampling Method</th>
<th>Spectra Type</th>
<th>$R^2$</th>
<th>Standard Error of Calibration (SEC)</th>
<th>PLS factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wedges</td>
<td>Normal Spectra</td>
<td>0.887</td>
<td>0.605</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2nd derivative spectra</td>
<td>0.898</td>
<td>0.575</td>
<td>5</td>
</tr>
<tr>
<td>Chips</td>
<td>Normal Spectra</td>
<td>0.883</td>
<td>0.618</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2nd derivative spectra</td>
<td>0.889</td>
<td>0.601</td>
<td>4</td>
</tr>
<tr>
<td>Sawdust</td>
<td>Normal Spectra</td>
<td>0.847</td>
<td>0.712</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2nd derivative spectra</td>
<td>0.845</td>
<td>0.719</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 2: R-squared values from the calibration models.

<table>
<thead>
<tr>
<th>Spectral Type</th>
<th>Calibration Set (~80%)</th>
<th>Validation Set (~20%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$</td>
<td>Standard Error of Calibration (SEC)</td>
</tr>
<tr>
<td>Normal spectra</td>
<td>0.907</td>
<td>0.528</td>
</tr>
<tr>
<td>2nd derivative spectra</td>
<td>0.912</td>
<td>0.516</td>
</tr>
</tbody>
</table>

**average of the absolute values of the residual.
Figure 1: Average NIR reflectance spectrum for the wedges of *E. grandis* Tag5 clonal material.
Figure 2: Pulp yield correlation plots corresponding to models developed using normal (top) and second derivative spectra (bottom). In both plots, the open circles represent samples in the calibration set which were used to generate the regression line for the models and the closed circles represent data used to test the prediction ability of the models. The diagonal line is shown for comparison and it represents line of equality between predicted and actual pulp yield.