External Validation for NIR Spectroscopy of *Eucalyptus* spp. Wood

Humberto Fantuzzi Neto : Full Professor, UFES, Brazil, hfantuzzi@yahoo.com.br
Flaviana Reis Milagres: Ph.D Candidate, UFV, Brazil, flaviana.milagres@ufv.br
José Lívio Gomide : Full Professor, UFV, Brazil, jlgomide@ufv.br
Andréia S. Magaton, Full Professor, UFRB, Brazil, anmagaton@yahoo.com.br
Bruno F. Homem de Faria: Graduate Student, UFV, Brazil, brunohomemfaria@hotmail.com

Abstract

Near infrared spectroscopy (NIRS) is a low cost, short time consuming technique and requires no chemicals reagents. Several publications about NIRS have shown the importance of cross validation and external validation for calibration of NIRS models. External validation is more appropriate to validate NIRS models because uses samples that were not used to develop the models. The objective of this study was to analyze external validation of NIRS models developed for wood basic density, lignin and extractives contents of 7 years old *Eucalyptus* wood. The NIRS models were developed using partial least squares (PLS) regression and mathematical transformations. The selected NIRS models presented 0.90, 0.95 and 0.90 as coefficients of determination (R^2) for basic density, total lignin and extractives content, respectively. Twenty five samples were used for external validation of the selected models. The models showed good predictive capabilities, resulting in 83% of the predictions with residual below or equal to 17 kg/m³ for wood basic density, 82% of the predictions with residual below or equal to 1% for lignin and 60% of the predictions with residual below or equal to 0.5% for extractives.

Keywords: *Eucalyptus*; wood quality; wood chemistry; NIR Spectroscopy; external validation.

Introduction

The Near Infrared Spectroscopy (NIRS) has been extensively used as an analytical technique because of its low cost, short time and no requirement of chemical reagents. Those characteristics are very important to accelerate forest breeding programs, reducing the waiting time for clone selection and extensively reducing costs for analysis of thousands individuals [1].

The infrared region of electromagnetic spectrum ranges from 750 to 2500 nm and this analytical technique monitors the molecular vibrations that are closely associated with linkages in different molecular structures [2]. The molecular vibrations occur in molecules that have O-H, N-H and C-H linkages and provide physical and chemical information about chemical substances [3].

NIRS calibration requires data from conventional laboratory analysis to correlate with near-infrared spectrum. The success of the technique essentially requires analytical data obtained from well established and precise laboratory methods. Model validation is essential after calibration for the characteristic being determined by spectroscopy [4]. Samples for validation must have values that contain the same data range used for calibration. The NIRS model precision can be determined by cross-validation and external validation methods [5].

In cross-validation one or more samples are removed from the group samples used to establish the calibration model and the remaining samples are used for prediction. The samples removed are returned to the original group a new set of samples is removed for new prediction. This process is repeated until all samples in the calibration set have passed by the set of prediction. It is suggested that cross-validation should be used only when the number of samples are limited due, for example, to high cost of laboratory analysis [2].

For external validation, to check the effectiveness of prediction of the characteristic of interest, are extracted randomly from the set of calibration a number of samples, called external validation group. NIRS calibration is performed with remaining samples to obtain the regression equations used to predict the validation samples. Calibration model is then used to estimate the characteristics values of the validation group from its near-infrared spectra. That way it is possible to compare the estimated values with values determined in laboratory [6].

Use of a separate set of validation (external validation) is therefore a more direct and precise approach. It is advisable to employ a set of data as similar as possible to the calibration set. It is recommend using the external validation whenever possible because the models obtained usually provide better results and precision [7].

The objective of this study was development and utilization of NIRS models, after external validation, to predict basic density, total lignin and extractive contents of *Eucalyptus* clones woods.

Experimental

Physical-chemicals characteristics

Seventy five clones of *Eucalyptus* wood, 7 years old, from Mato Grosso do Sul and São Paulo States, Brazil, were used for this study. Three trees for each clone representing the commercial average DBH and height were harvested for each clone. Five 50 cm long logs were cut from each tree at different heights (base, 25, 50, 75 and 100%). The 5 logs representing each clone were cut into chips which were screened (16x16mm and 5x5mm screens) and chips with cutting defects, knots and pieces of bark were manually removed. The chips were air-dried, homogenized in a rotary mixer and finally stored in polythene bags to obtain and maintain uniform moisture content.

Samples of chips representing each clone were used for wood basic density determination and chemical analysis. Sawdust for chemical analysis was produced in Wiley mill, screened to obtain 40/60 mesh particle size, acclimatized in a temperature and moisture controlled room (25 °C and 50%, respectively) and finally stored in tight closed jars.

Methods used to determine the basic density, total lignin and extractive contents are described in Table 1.

Table 1. Methodology for wood physical-chemicals characteristics

Characteristics	Methodology
Basic Density (kg/m³)	ABNT (NBR 11941)
Total Lignin (%)	Gomide and Demuner [8]; Goldschmid [9]
Extractives (%)	Tappi T264 om-82 [10]

NIR spectra measurement

The Wiley mill sawdust was further processed in a cyclone mill to obtain more uniform and finer particles, acclimatized (25 ºC and 50%) and then used to obtain NIRS spectra. The equipment used to collect the spectra was a NirSystem-5000 spectrophotometer from FOSS. The NIR spectra were obtained in wavelengths range 1100-2500nm at 2nm intervals, totaling 700 wavelengths per sample. Thirty-two scans were accumulated for each sample and the results were averaged. After the spectrum had been obtained, the “spin sample” was emptied, repacked and a duplicate spectrum obtained. The duplicate spectra were averaged.

Calibrations for wood quality characteristics

After collection of NIR spectra, correlations were determined between spectra and laboratory determinations for wood basic density, lignin and extractives contents. NIRS models were established for each wood quality characteristics. From the total 75 samples, 50 were randomly used for calibration and the remaining 25 samples were used for external validation. Samples with high leverage and high residual variance, noticeably different from the rest of the samples, were detected as outliers in graphical analysis and were excluded from the models.

The Unscrambler 9.6® program was used for model calibration and analysis were carried out by partial least squares (PLS) regression with a maximum of 15 factors or latent variables (LV) and full cross validation. Mathematical transformations as first and second derivatives of the spectral data were also tested as proposed by Savitzky and Golay [11].

Selection of calibrations models and their performance in external validation

Selection of calibration models were done by coefficient of determination (R^2) and standard error of calibration (SEC, determined from the residuals of the final calibration). Models with higher values of R^2, lower calibration errors and a lower number of latent variables were selected. The selected models were tested by external validation [6] using a group of 25 samples that did not participate in models calibration. The predicted values obtained from the models were evaluated by determination coefficient of prediction (R^2_p), standard error of prediction (SEP) and the residuals (laboratory values subtracted from the estimated NIRS values).

Results and Discussion

In Table 2 are reported the statistical parameters of wood characteristics for samples used for models calibration and external validation, including mean, maximum, minimum, standard deviation and coefficient of variation.

The calibration samples presented wood basic density, lignin and extractives contents (average, minimum, maximum) that represent the variability usually reported for *Eucalyptus* species in Brazil. Lignin content of the samples used in this study ranged from 24.3 to 30.9%. It is known that lignin content of *Eucalyptus* grown in Brazil usually is in the range of 24-32%, much higher than in Northern Hemisphere hardwood specie. The high variability of extractives content (1.2-7.3%) resulted in a high coefficient of variation (18%). The coefficients of variation for basic density and lignin content were low, demonstrating the homogeneity of these characteristics.

Samples used for external validation presented characteristics similar to the ones used for models development (Table 2), a very important parameter required to obtain good predictions of the desired characteristic.

Table 2. Characteristics of samples used for model calibration and external validation

	Characteristics	Maximum	Minimum	Mean	SD	CV (%)
Calibration	Basic Density (kg/m³)	555	429	493	27.5	6
	Total Lignin (%)	30.9	24.3	27.4	1.7	6
	Extractives (%)	7.3	1.2	3.1	1.3	18
Validation	Basic Density (kg/m³)	547	451	493	23,0	5
	Total Lignin (%)	30.1	24.4	27.4	1.8	7
	Extractives (%)	5.6	1.4	3.0	1.3	23

NIRS calibration and validation

In Figure 1A is graphically presented wood basic density determined in laboratory and estimated by NIRS model. The coefficient of determination (R^2) showed a high predictive ability of the model (0.90) with a standard error of calibration of only 9 kg/m³. Schimleck et al. [12] estimated the wood basic density of *Eucalyptus globulus* using NIRS technique and reported R^2 values ranging from 0.62 to 0.80 and SEC values ranging from 25 a 33 kg/m³.

To confirm the predictive performance of the NIRS model, samples that had not been used for development of the models were tested and the residuals of this external validation are shown in Figure 1B. External validation of the predicted wood basic densities presented a coefficient of determination (R^2_p) of 0.72 and a standard error of prediction (SEP) of 13 kg/m³. As can be seen in Figure 1B, 83% of the samples showed residues in the range of ± 17 kg/m³, close to those found in the laboratory determination, and 57% of the residual values were

between ± 10 kg/m³.

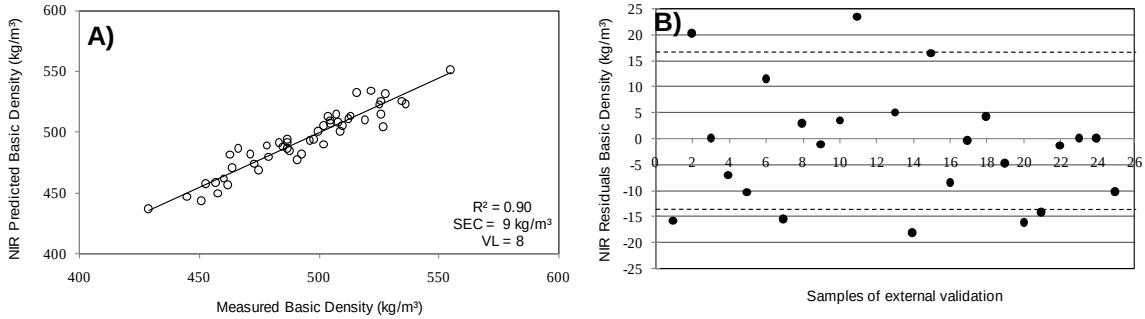

Figure 1. A) Relationship between laboratory and NIR-estimated wood basic density; B) Residuals for wood basic density in external validation.

The NIRS lignin content model showed a R² value of 0.95 and 0.4% SEC (Figure 2A), better values than found by Tyson et al. [13] who studied *Eucalyptus* wood and found 0.76 R² and 0.7% SEC and they considered them as good results for prediction. NIRS study of *Eucalyptus* lignin content published by Schimleck et al. [14] reported 0.79 R² and 0.8% SEC. Analyses of residuals in Figure 2B demonstrated that 82% of the residues were within ± 1% of lignin content and 45% were within ± 0.5%. It should be considered that, for laboratory analysis, lignin content variation smaller than or equal to 1% can be considered acceptable between two replicates.

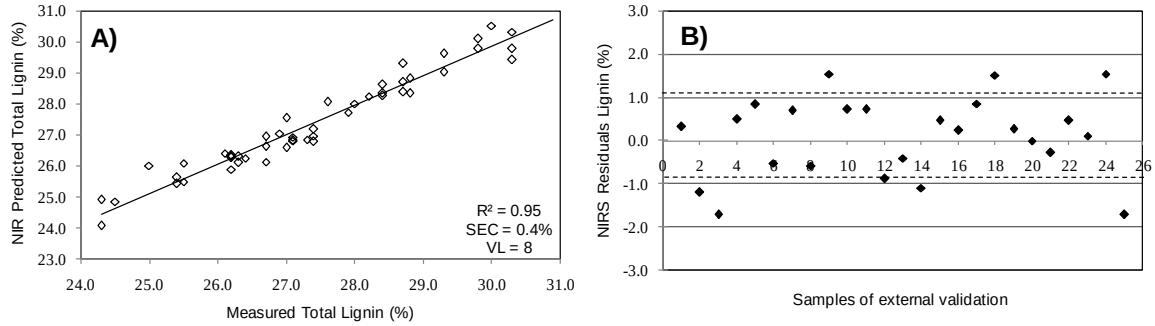

Figure 2. A) Relationship between laboratory and NIRS-estimated lignin content; B) Residuals for lignin content in external validation.

The NIRS model developed for extractives content showed 0.90 R² and 0.4% SEC (Figure 3A), similar to values described by Baillères et al. [15] who reported 0.87 and 0.2%, respectively, for R² and SEC. Figure 3B shows that 76% of the samples presented residuals within ± 0.6%, value close the 0.5% variation considered acceptable in laboratory analysis.

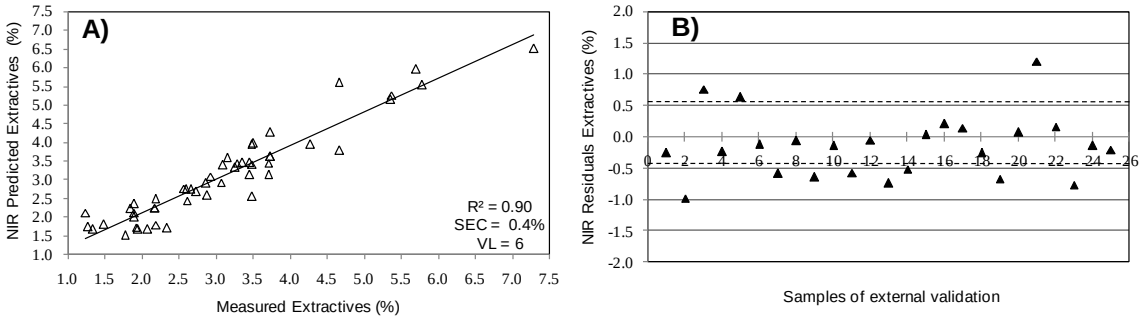

Figure 3. A) Relationship between laboratory and NIRS-estimated extractives content. B) Residuals for extractives content in external validation.

Conclusions

All models developed in this study (basic density, lignin and extractives contents) presented good predictive performance and their errors were within values considered acceptable in laboratory analysis. Therefore, NIRS technology can be considered as an analytical tool for fast and reliable estimates of wood basic density and chemical composition, which are highly used in breeding programs of *Eucalyptus* spp for kraft pulp production.

References

1. Schimleck, L. R., Stürzenbecher, R., Jones, P. D., Evans, R., Development of wood property calibrations using near infrared spectra having different spectral resolutions. Journal of Near Infrared Spectroscopy 12: 55–61 (2004).
2. Pasquini, C. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. Journal of the Brazilian Chemical Society, 14 (2):198 (2003).
3. Schimleck, L. R., Evans, R., Ilic J., Estimation of *Eucalyptus delegatensis* wood properties by near infrared spectroscopy. Canadian Journal of Forest Research, Canada, 31: 1671-1675 (2001).
4. Gemperline, P. Practical guide to chemometrics. 2.ed. Boca Raton: Taylor and Francis, pp. 520 (2006).
5. WORKMAN, J.; WEYER, L. Practical guide to interpretive near-infrared spectroscopy. New York: CRC Press, pp.307 (2007).
6. Burns, D.A., Clurczak, E.W. Handbook of near-infrared analysis. 3rded. Boca Raton: CRC, pp.808 (2008).
7. Coscione, A. R. O uso de calibração multivariada para a determinação espectrofotométrica simultânea de alumínio e ferro: aplicação na análise de plantas e solos. Doctoral Thesis, State University of Campinas, Campinas, São Paulo, Brazil, (2001).
8. Gomide, J. L. and Demuner, B. J., O Papel 47 (8): 36 (1986).
9. Goldschimid, O., Ultraviolet spectra. In: Lignins, Occurrence, formation, structure and reactions (Ed: K. V. Sarkanen, and C. H. Ludwig), John Wiley and Sons, New York, USA, pp. 241-266, (1971).
10. Tappi Standard Procedures. Tappi Press. Atlanta, USA, (1998).
11. Savitzky, A., Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry, 36: 1627-1639 (1964).
12. Schimleck, L. R., Michell, A. J., Raymond, C. A., Muneri, A. Estimation of basic density of Eucalyptus globulus using near-infrared spectroscopy. Canadian Journal of Forest Research, 29: 194-201 (1999).
13. Tyson, J. A., Schimleck, L. R., Aguiar, A. M., Abad, J. I. M., Rezende, G. D. S. P. Adjusting

near infrared wood property calibrations for central Brazil to predict the wood properties of samples from southern Brazil. Journal Appita, 62 (1): 46-51 (2009).
14. Schimleck, L.R., Rezende, G.D.S.P., Demuner, B.J., Downes, G.M. Estimation of whole-tree wood quality traits using near infrared spectra collected from increment cores. Appita Journal, 59: 231-236 (2006).
15. Baillères, H., Davrieux, F., Ham-Pichavant, F. Near infrared analysis as a tool for rapid screening of some major wood characteristics in a eucalyptus breeding program. Annals of Forest Science, 59 (5-6): 479-490 (2002).