

Desenvolvimento e validação de modelos multivariados empregando espectroscopia no infravermelho próximo para estimativa de características da madeira de eucalipto set

Development and validation of multivariate models employing near-infrared spectroscopy for eucalyptus wood characteristics estimations

Autores/Authors*: Celio Pasquini¹
Cláudio José Gonçalves Carneiro²
Permínio Muniz Bomfim³

Palavras-chave: Calibração multivariada, densidade básica, espectroscopia no infravermelho próximo (NIR), extrativos, lignina total, madeira de eucalipto, qualidade de madeira.

RESUMO

A Espectroscopia no Infravermelho Próximo (NIR) tem sido empregada no desenvolvimento de modelos multivariados para estimar o teor de lignina total, densidade básica e teor de extrativos em etanol/tolueno em serragem de madeira de eucalipto. A avaliação multivariada, feita por meio de Análise de Componentes Principais (PCA) dos espectros NIR obtidos para 926 amostras de madeira, mostra que a espectroscopia NIR reconhece os diferentes materiais genéticos. Essa diferenciação entre os espectros de materiais genéticos deve ser considerada na elaboração de modelos globais quantitativos multivariados. Os modelos de Regressão de Quadrados Mínimos Parciais (PLS) para previsão de lignina total, densidade básica e teor de extrativos em madeira de eucalipto foram construídos e validados segundo as Normas ASTM 1655-05. Os resul-

Keywords: Basic density, eucalyptus wood, extractives, multivariate calibration, near-infrared (NIR) spectroscopy, total lignin, wood quality.

ABSTRACT

Near-Infrared (NIR) Spectroscopy has been employed in the development of multivariate models in order to estimate the total lignin content, basic density and extractives in ethanol/toluene in eucalyptus wood sawdust. The multivariate evaluation made by means of Principal Components Analysis (PCA), of NIR spectra obtained for 926 wood samples, shows that the NIR spectroscopy recognizes the different genetic materials. This distinction among the spectra of several genetic materials must be considered when quantitative global multivariate models are developed. The Partial Least Square Regression (PLS) models to predict total lignin, basic density and extractive content in eucalyptus wood were developed and validated according to ASTM 1655-05. The

*Referências dos Autores: / Authors' references:

- 1 - Instituto de Química – Universidade Estadual de Campinas - Caixa Postal 6154 – CEP: 13084-971 - Campinas – SP – pasquini@iqm.unicamp.br - Brasil/Brazil
- 2 - Suzano Papel e Celulose – Mucuri – Bahia – email: carneiro@suzano.com.br - Brasil/Brazil
- 3 - Suzano Papel e Celulose – Mucuri – Bahia – email: permindio@suzano.com.br – Brasil/Brazil

tados de validação externa, expressos pelo Erro Padrão de Previsão (SEP) e pelo coeficiente de correlação entre os resultados estimados pela espectroscopia NIR e aqueles obtidos por meio dos métodos padrão (r), realizada com 160 amostras não empregadas na elaboração do modelo, foram: Lignina Total, SEP = 1,2%, $r = 0,800$; Densidade Básica, SEP = 26 kg m⁻³, $r = 0,930$; Extrativos, SEP = 0,38%, $r = 0,864$. Os modelos foram considerados satisfatórios segundo as Normas ASTM 1655-05.

INTRODUÇÃO

A demanda por resultados analíticos que são empregados na definição da qualidade da madeira que se destina à produção de celulose e papel tem aumentado continuamente. Os principais fatores que induzem essa demanda são: a preocupação constante com a avaliação e padronização da qualidade da matéria prima, que chega ao pátio e é enviada à produção, e os programas de melhoria genética.

Os métodos padrão empregados na obtenção dos parâmetros associados à qualidade da madeira de eucalipto são validados e reconhecidamente produzem resultados confiáveis. Porém, todos eles são trabalhosos e requerem um longo tempo de execução, o que inviabiliza seu uso quando a demanda analítica é alta.

A busca por técnicas alternativas que possam fazer frente a essa alta demanda promovida pelos programas de melhoria genética e de qualidade de matéria prima tem, em consequência, sido intensificada. Dentre as técnicas elegíveis destaca-se a Espectroscopia no Infravermelho Próximo (NIR)[1,2], que apresenta todas as características necessárias a um método rápido, não destrutivo e que demanda pouca quantidade de amostra, a qual, por sua vez, precisa ser submetida a procedimentos mínimos e rápidos de pré-tratamento[3-12].

A espectroscopia NIR tem sido empregada junto a produtos agro-industriais, como cereais e madeira, por meio do uso da técnica de refletância difusa[2]. Nessa técnica, amostras sólidas são apresentadas ao instrumento (espectrofotômetro) que incide sobre elas a radiação eletromagnética na região espectral entre 800 a 2500nm. A interação da radiação nessa faixa de comprimento de onda com as espécies químicas constituintes da madeira, por exemplo, é medida por meio da relação: $R(\lambda) = \text{LOG}(R_a/R_r)$, na qual $R(\lambda)$ é a refletância difusa da amostra no comprimento de onda λ , R_a é a intensidade da radiação medida na presença da amostra e R_r é a intensidade medida para um material de referência (spectralon ou cerâmica), ambas medidas nesse mesmo comprimento de onda. Uma vez que essa relação não expressa a refletância difusa real da amostra, mas sim uma aproximação desta, a geometria do sistema de medida e da amostra deve ser

results of the external validation, expressed by the Standard Error of Prediction (SEP) and by the correlation coefficient (r) between the results estimated by NIR spectroscopy and those obtained by the standard methods, carried out with additional 160 samples not employed in the modeling stage, were: Total Lignin, SEP = 1.2%, $r = 0.800$; Basic Density, SEP = 26 kg m⁻³, $r = 0.930$; Extractives, SEP = 0.38%, $r = 0.864$. The models were considered to be satisfactory according to ASTM 1655-05.

INTRODUCTION

The demand for analytic results used to define the wood quality destined for pulp and paper production has continuously increased. The main factors inducing this demand are: the constant concern about evaluation and standardization of the quality of the raw material arriving at the yard and being sent to production, and the genetic improvement programs.

The standard methods employed to obtain the parameters associated with the eucalyptus wood quality are validated and recognizably produce reliable results. However, all of them involve hard work and require a long time of execution, which makes their use impracticable when the analytical demand is high.

Therefore, the search for alternative techniques that may be up to this high demand promoted by the genetic improvement and raw material quality standardization programs has been enhanced. The Near-Infrared Spectroscopy (NIR)[1,2] distinguishes itself among the eligible techniques, because it presents all characteristics required for a quick and nondestructive method, demanding just a small amount of samples, which must only undergo minimum and quick pretreatment procedures [3-12].

The NIR spectroscopy has been employed with agribusiness products, such as cereals and wood, by means of the diffuse reflectance technique[2]. In this technique, solid samples are presented to the instrument (spectrophotometer), from which they receive the electromagnetic radiation in the spectral region, between 800 and 2500nm. The interaction of radiation in this wave length range with the chemical wood constituting species, for instance, is measured by means of the ratio: $R(\lambda) = \text{LOG}(R_a/R_r)$, where $R(\lambda)$ is the diffuse reflectance of the sample in the wave length λ , R_a is the radiation intensity measured in the presence of the sample and R_r is the intensity measured for a reference material (spectralon or ceramics), both measured on this same wave length. Since this ratio does not express the actual diffuse reflectance of the sample, but just an approximation

o mais reprodutível possível. Na prática, essa necessidade pode ser suprida pulverizando a amostra de madeira, reduzindo-a a uma serragem fina (granulometria $< 0,50\text{mm}$) e acondicionando-a em um recipiente (célula de medida) adequado para apresentação ao equipamento.

O espectro de refletância gerado resulta principalmente da interação da radiação eletromagnética resultante da indução de transições energéticas vibracionais referentes aos sobretons de modos vibracionais e suas combinações provenientes principalmente de ligações químicas constituídas pelos grupos de átomos C-H, N-H e O-H. A presença destes grupos de átomos na totalidade das espécies químicas constituintes da madeira assegura que seu espectro de refletância NIR contém informações analíticas que podem ser empregadas na elaboração de modelos quali/quantitativos, visando seu uso na determinação de parâmetros de interesse da indústria de celulose e papel.

No entanto, o mesmo fato que assegura que um espectro de refletância de madeira contém informações sobre todos os seus constituintes, também impede o seu aproveitamento direto, da forma convencional univariada, ou seja, por meio de utilização de modelos simples que se fundamentam na variação da refletância medida em um único comprimento de onda e sua correlação com a propriedade ou conteúdo da espécie de interesse. Essa limitação provém da alta sobreposição das refletâncias promovidas pelos mesmos grupos de átomos (O-H, por exemplo), presentes em diferentes moléculas em diferentes conformações.

De fato, a única forma possível de se aproveitar analiticamente as informações que os espectros de refletância NIR contém se dá por meio do uso de técnicas que empregam um grande número (talvez todas) das variáveis disponíveis. Essas técnicas são denominadas técnicas multivariadas.

A área da ciência que se dedica ao desenvolvimento e aplicação das técnicas multivariadas é chamada Quimiometria[13]. A Quimiometria congrega uma série de ferramentas estatísticas multivariadas que permitem a extração da informação presente nos espectros NIR e sua correlação com as propriedades de interesse associadas à qualidade da madeira. Programas quimiométricos têm sido disponibilizados para permitir o uso amigável dessas técnicas e a interpretação segura dos resultados.

Dentre as técnicas multivariadas, destacam-se a Análise de Componentes Principais (PCA) e a Regressão de Quadrados Mínimos Parciais (PLS)[14]. A primeira (PCA) é empregada em uma análise qualitativa dos espectros de refletância e permite acessar as diferenças e semelhanças entre eles. Um bom programa quimiométrico permite empregar o conhecimento prévio existente sobre as amos-

thereto, both measuring system and sample geometry should be the most reproducible as possible. In practice this need may be helped by grinding the wood sample, so as to reduce it to fine sawdust (granulometry $< 0.50\text{mm}$) and placing it in a suitable container (measuring cell) for presentation to the equipment.

The reflectance spectrum generated results especially from the interaction of the electromagnetic radiation resulting from the induction of vibrational energetic transitions concerning the overtones of vibrational modes and their combinations, mainly originating from chemical bonds formed by the atom groups C-H, N-H and O-H. The presence of these atom groups in the totality of the chemical wood constituting species ensures that their NIR reflectance spectrum contains analytical information that may be used to develop quali/quantitative models, aiming to utilize them for determining parameters of interest to the pulp and paper industry.

However, the same fact ensuring that a wood reflectance spectrum contains information on all its constituents also prevents it from being taken direct advantage of in the conventional univariate way, i.e., by using simple models based on the variation of the reflectance measured on a single wave length and its correlation with the property or contents of the species of interest. This limitation originates from the high overlapping of the reflectances promoted by the same atom groups (O-H, for instance), present in different molecules in different conformations.

In fact, the only possible way of analytically taking advantage of the information contained in the NIR reflectance spectra is by making use of techniques employing a great number (maybe all of them) of the variables available. These techniques are called multivariate techniques.

The science area dealing with the development and application of the multivariate techniques is called Chemometrics [13]. The Chemometrics gathers a series of multivariate statistical tools allowing to extract the information present in the NIR spectra and its correlation with the properties of interest, associated with the wood quality. Chemometric programs have been made available to allow using amicably these techniques and the reliable interpretation of the results.

The Principal Components Analysis (PCA) and the Partial Least Square Regression (PLS)[14] distinguish themselves among the multivariate techniques. The first one (PCA) is used for a qualitative analysis of the reflectance spectra and allows accessing the differences and similarities between them. A good

tras - como sua origem geográfica, idade, características genéticas -, e associá-lo aos resultados da PCA. Desta forma, podem-se estabelecer conclusões relevantes sobre as causas das possíveis semelhanças, da definição de agrupamentos e tendências entre os espectros de refletância das amostras. Essas conclusões balizam também as decisões sobre a necessidade e possibilidade ou não de elaboração de modelos quantitativos locais ou globais.

A segunda técnica (PLS) permite a elaboração de modelos quantitativos multivariados e tem sido a técnica de escolha em praticamente todas as aplicações da espectroscopia NIR baseadas em medidas de refletância difusa. O principal requisito para utilização adequada da técnica PLS junto a amostras de madeira - e amostras naturais de forma geral -, é a obediência ao critério de representatividade das amostras de calibração e validação em relação à população das amostras que se pretende analisar com os modelos construídos. Para amostras naturais, a obediência a esse critério pressupõe que toda a variabilidade significativa associada a fatores regionais, sazonais, à hibridização e material genético das árvores das quais a madeira foi extraída seja contemplada nos conjuntos de amostras empregados na construção e validação do modelo quantitativo elaborado por PLS. A PCA pode auxiliar bastante no aspecto de seleção das amostras representativas.

A complexidade dos procedimentos de calibração multivariados e suas exigências levaram, no ano de 1995, a ASTM (American Standards for Testing and Materials) a lançar uma norma chamada de Práticas Padrão para Análises Quantitativas Multivariadas por Infravermelho (incluindo o NIR) (ASTM 1655-05)[15], a qual foi recentemente revista e que guia o analista no processo de calibração multivariada. A norma é bastante completa e dá ênfase especial à seleção das amostras representativas e ao processo de validação e manutenção de modelos multivariados, como aqueles de interesse da indústria de celulose e papel. De forma implícita à existência dessa norma, está o reconhecimento da utilidade dos procedimentos analíticos baseados em espectroscopia NIR e calibração multivariada.

Este trabalho visou o desenvolvimento de modelos quantitativos multivariados para previsão de três parâmetros de interesse na determinação da qualidade da madeira de eucalipto destinada à fabricação de celulose: Lignina Total (LT), Extrativos em etanol/tolueno (ET) e Densidade Básica (DB). Inicialmente, realizou-se uma Análise de Componentes Principais para se poder observar a variabilidade dos espectros de refletância difusa das amostras analisadas e a associação dessa variabilidade a informações prévias sobre as amostras. Com base nesses resultados, foram construídos e validados modelos PLS

chemometric program allows applying the previous knowledge existing about the samples - such as their geographic origin, age, genetic characteristics -, and associating it with the PCA results. In this way, relevant conclusions can be drawn about the causes of the possible similarities, the definition of groupings and tendencies among the reflectance spectra of the samples. These conclusions also delimit the decisions about the need and possibility or not of developing local or global quantitative models.

The second technique (PLS) allows developing multivariate quantitative models and has been the technique chosen in practically all NIR spectroscopy applications based on diffuse reflectance measurements. The main requirement for the suitable utilization of PLS technique with wood samples - and natural samples in general -, is the adherence to the criterion of representativeness of the calibration and validation samples with regard to the sample population intended to be analyzed with the developed models. For natural samples, the adherence to this criterion presupposes that all significant variability associated with regional and seasonal factors, with hybridization and genetic material of the trees the wood was extracted from, is contemplated in the sets of samples used in the construction and validation of the quantitative model developed by PLS. PCA can help a good deal with regard to selection of representative samples.

The complexity of the multivariate calibration procedures and their requirements led ASTM (American Standards for Testing and Materials) to introduce in 1995 a standard called Standard Practices for Multivariate Quantitative Infrared Analyses (including NIR) (ASTM 1655-05)[15], which was recently revised and guides the analyst as to the multivariate calibration process. The standard is rather complete and puts special emphasis on selecting the representative samples and on the process of validation and maintenance of multivariate models, like those of interest to the pulp and paper industry. Implicitly to the existence of this standard is the acknowledgement of the usefulness of the analytical procedures based on NIR spectroscopy and multivariate calibration.

This work aimed at the development of multivariate quantitative models for predicting three parameters of interest for determining the eucalyptus wood quality destined for pulp manufacturing: Total Lignin (LT), Extractives in ethanol/toluene (ET) and Basic Density (DB). Initially, a Principal Components Analysis was conducted in order to be able to observe the variability of the diffuse reflectance spectra of the an-

para previsão do teor de lignina total e de extrativos, assim como da densidade básica das madeiras de eucalipto, empregando-se a Norma ASTM 1655-05. Os resultados obtidos são comparados àqueles recentemente relatados na literatura técnica.

EXPERIMENTAL

Os espectros de refletância das amostras de madeira foram obtidos na região de 1100 a 2500 nm em um espectrofotômetro FOSS/NIRsystem modelo 5000. Uma célula circular padrão fornecida pelo fabricante desse equipamento foi utilizada na obtenção dos espectros de refletância difusa. Os espectros foram obtidos como média de 50 varreduras.

As amostras empregadas tanto para calibração como validação contemplaram a variedade genética e de origem geográfica esperada em futuras previsões a serem feitas pelos modelos multivariados baseados na espectroscopia NIR, além da variedade necessária dos valores dos parâmetros de interesse. Como é usual nesses casos de amostras naturais, uma grande quantidade de amostras que apresentava valores das propriedades em um intervalo estreito foi identificada para todos os três parâmetros. O número de amostras empregado nesse estudo foi de 926 para a construção dos modelos e 160 para sua validação.

Todas as amostras foram compostas por cavacos oriundos de toretes de árvores de eucalipto coletadas entre os anos de 2002 a 2004. As amostras foram pulverizadas inicialmente em moinho de facas e finalmente em moinho tipo UDY para granulometria menor que 0,50mm. Antes da obtenção de seus espectros, as amostras permaneceram por três horas em contato com a atmosfera do laboratório, cuja temperatura e umidade foram controladas em (24 ± 1) °C e (54 ± 2) %, respectivamente.

Os valores de referência para os três parâmetros de interesse foram obtidos por meio de análises realizadas com o uso de métodos de referência padrão para a indústria de papel e celulose.

Procedeu-se a uma avaliação interlaboratorial da reprodutibilidade dos métodos de referência com a participação de três laboratórios. Três amostras de madeira na forma de cavacos foram enviadas aos laboratórios que efetuaram as determinações, empregando métodos padrão.

Os espectros obtidos no espectrofotômetro NIR system 5000 foram exportados e armazenados no padrão de formatação de arquivos digitais NAS e acessados pelo programa quimiométrico utilizado nesse trabalho. O pré-tratamento dos espectros de refletância e elaboração dos modelos PCA e PLS foram realizados com o uso do programa quimiométrico UNSCRAMBLER versão 9.5

alyzed samples and the association of this variability with previous information about the samples. Based on these results, PLS models were constructed and validated, in order to predict the total lignin content, as well as that of extractives, in addition to the basic density of the eucalyptus woods, by following ASTM 1655-05 Standard. The results obtained are compared to those recently reported in the technical literature.

MATERIALS AND METHODS

The wood sample reflectance spectra were obtained in the range of 1100 to 2500 nm, by means of a 5000 model FOSS/NIR system spectrophotometer. A circular standard cell supplied by the manufacturer of this equipment was used to obtain the diffuse reflectance spectra. The spectra were obtained as an average of 50 scanings.

The samples used both for calibration and validation considered the genetic variety and that of geographic origin expected in future predictions to be made by the multivariate models based on NIR spectroscopy, in addition to the necessary variety of the values of the parameters of interest. As it is usual in these cases of natural samples, a great amount of samples presenting property values in a narrow interval was identified for all three parameters. The samples utilized in this study amounted to 926 for the model construction and to 160 for its validation.

All samples consisted of chips coming from small eucalyptus tree logs, collected between 2002 and 2004. The samples were initially ground in a knife chipper and finally in a UDY type mill to a granulometry smaller than 0.50 mm. Prior to obtaining their spectra, the samples remained for three hours in contact with the lab atmosphere, the temperature and humidity of which were controlled at (24 ± 1) °C and (54 ± 2) %, respectively.

The reference values for the three parameters of interest were obtained by means of analyses carried out by using standard reference methods for the pulp and paper industry.

An inter-laboratory evaluation of the reproducibility of the reference methods was performed with the participation of three laboratories. Three wood samples in the form of chips were sent to the laboratories, where determinations were carried out by use of standard methods.

The spectra obtained by the 5000 model NIR system spectrophotometer were exported and filed in the formatting standard of digital NAS files and accessed by the chemometric program used in this work. The reflectance spectra pretreatment and the develop-

(CAMO, Noruega).

Várias formas de pré-tratamento foram preliminarmente investigadas, incluindo ferramentas como MSC (Multiplicative Signal Correction), 1ª e 2ª derivadas e filtros polinomiais. Finalmente, após avaliações dos efeitos dessas formas de pré-tratamento, optou-se pelo seguinte protocolo aplicado a todos os espectros: 1) Obtenção da primeira derivada empregando-se o método polinomial de Savitzk-Golay com polinômio de 2º Grau

e janelas de 5 elementos à esquerda e à direita do ponto central e, 2) Os espectros derivados foram centrados na média.

RESULTADOS E DISCUSSÃO

A Figura 1 mostra os espectros de refletância para o conjunto de calibração constituído por 926 amostras. Os espectros apresentam-se como aqueles típicos de produtos naturais, mostrando regiões nas quais a água contribui significativamente para a absorção da radiação. Mesmo nessas amostras previamente secas e para as quais se espera um teor médio de umidade inferior a 5%. A alta intensidade da absorção das ligações O – H da água se faz notar nas regiões entre 1380 – 1600 nm e, principalmente, entre 1850 – 2100 nm. As demais ocorrências espectrais podem ser atribuídas aos constituintes principais da madeira, tais como celulose, lignina e materiais extraíveis em solventes orgânicos. Nesse caso, as ligações C – H e O – H de caráter menos polar do que as ligações da água contribuem significativamente.

A flutuação de linha base observada entre os espectros originais da Figura 1, provém do espalhamento da radiação NIR pelas amostras que não apresentam, por exemplo, uma granulometria uniforme e que podem apresentar variações estruturais que contribuem para esse efeito. Essa flutuação certamente não carrega informações relacionadas à composição química da madeira e sua minimização pode contribuir na etapa de elaboração dos modelos, pois esses irão considerar as variações espectrais que são realmente significativas para a previsão dos parâmetros de interesse.

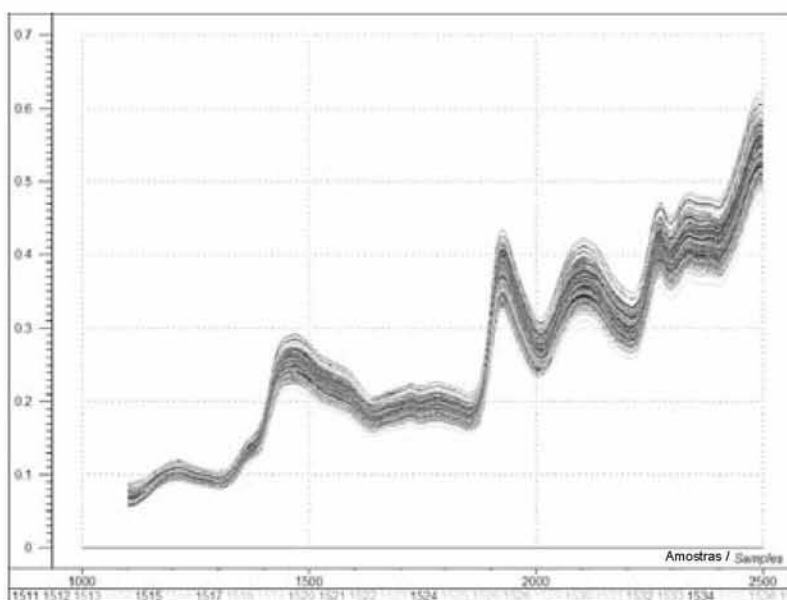


Figura 1. Espectros de refletância obtidos para as 926 amostras de madeira que constituem o conjunto de calibração

Figure 1. Reflectance spectra obtained for the 926 wood samples the calibration set consists of

ment of PCA and PLS models were performed using the chemometric program UNSCRAMBLER, version 9.5 (CAMO, Norway).

Several pretreatment modes were preliminarily investigated, including such tools as MSC (Multiplicative Signal Correction), 1st and 2nd derivatives and polynomial filters. Finally, after evaluating the effects of these pretreatment modes, the following protocol, applied to all

spectra, was selected: 1) Obtention of the first derivative by using the polynomial method of Savitzk-Golay with 2nd degree polynomial and 5-element windows to the left and to the right of the central point; and 2) The derived spectra were centered on the average.

RESULTS AND DISCUSSION

Figure 1 shows the reflectance spectra for the calibration set consisting of 926 samples. The spectra present themselves as those typical of natural products, showing regions in which water significantly contributes to the radiation absorption. Even at these previously dried samples, for which an average moisture content lower than 5% is expected, the high intensity of absorption of O – H bonds of the water can be observed in the ranges of 1380 to 1600 nm and especially between 1850 and 2100 nm. The remaining spectral occurrences may be attributed to the main wood constituents, such as cellulose, lignin and extractable materials in organic solvents. In this case, the C – H and O – H bonds of less polar character than the water bonds contribute in a significant way.

The baseline fluctuation observed between the original spectra of Figure 1 originates from NIR radiation scattering by the samples that do not present, e.g., a uniform granulometry and may present structural variations contributing to this effect. This fluctuation certainly does not carry any information related to the chemical composition of the wood and its minimization may contribute at the developing stage of the models,

A primeira derivada dos espectros parece constituir uma forma efetiva de pré-tratamento, como pode ser observado na Figura 2, por meio de uma definição clara de linha base dos espectros modificados.

O conjunto original e o das primeiras derivadas dos espectros das 926 amostras de calibração foram analisados empregando-se a técnica PCA, com o objetivo de se identificar a existência de agrupamentos entre as amostras. A PCA mostra que 6 componentes principais são necessários para considerar 96% da variabilidade inicial existente nos dados espectrais após serem submetidos ao pré-tratamento por meio da primeira derivada.

A Figura 3 mostra os resultados da distribuição dos scores nas três primeiras componentes principais resultantes da PCA sobre todo o conjunto de dados espectrais originais. Os scores representam a projeção dos espectros das amostras no espaço de dimensão reduzida, produzido pelas novas variáveis (componentes principais). Pode ser observada uma distribuição dos espectros das amostras que define ao menos três grandes grupos.

Esses grupos refletem, em grande parte, a variedade genética das amostras de eucalipto analisadas. No caso, as três primeiras componentes explicam cerca de 85% da variância total dos dados espectrais. Três grupos de amostras podem ser identificados claramente e atribuídos à

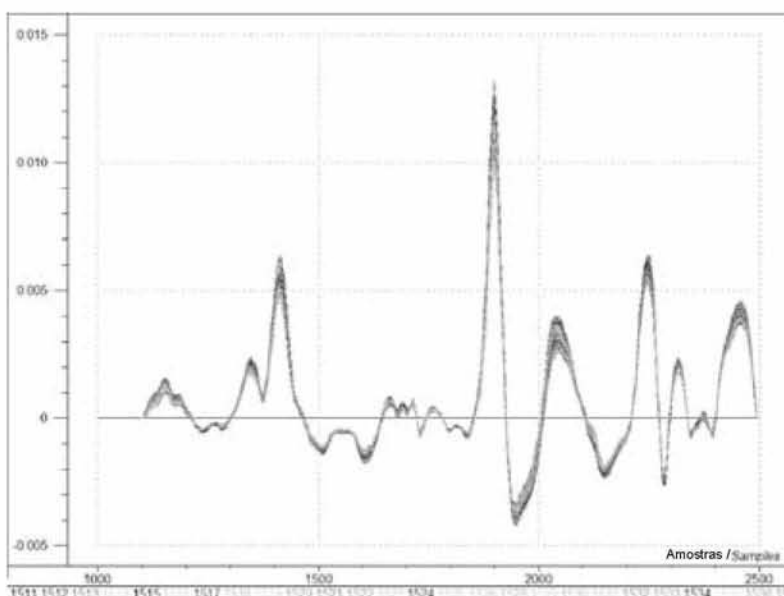


Figura 2. Primeira derivada dos espectros de refletância das 926 amostras de madeira empregadas na elaboração dos modelos

Figure 2. First derivative of the reflectance spectra of the 926 wood samples used for developing the models

as these ones will consider the spectral variations that are actually significant to predict the parameters of interest.

The first spectra derivative seems to be an effective pre-treatment mode, as it can be observed in Figure 2, by means of a clear baseline definition of the modified spectra.

The original set and that of the first derivatives from the spectra of the 926 calibration samples were analyzed by using the PCA technique,

in order to identify the existence of groupings among the samples. The PCA technique shows that 6 main components are required to consider 96% of the initial variability existing in the spectral data after undergoing the pretreatment by means of the first derivative.

Figure 3 shows the results of the score distribution in the first three main components resulting from the PCA technique over the whole set of original spectral data. The scores represent the projection of the sample spectra in the reduced dimension space, produced by the new variables (main components). A sample spectra distribution can be observed that defines at

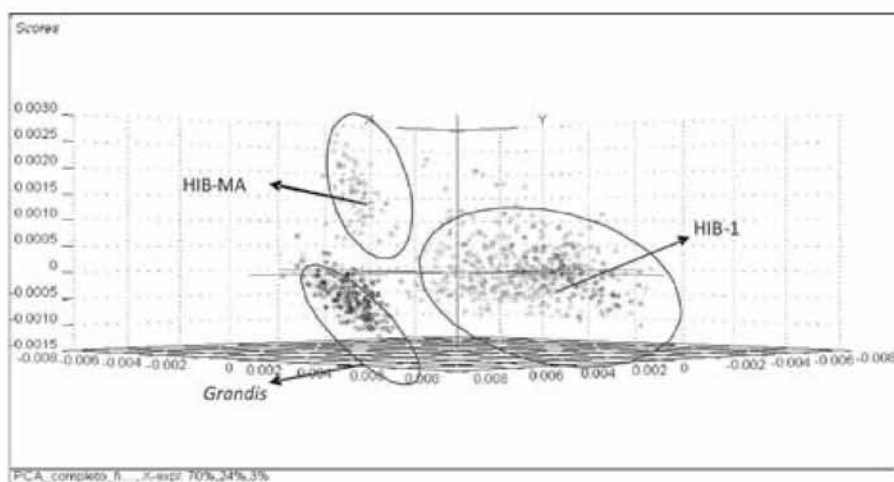


Figura 3. Resultado da análise de componentes principais do conjunto de espectros originais de calibração

Figure 3. Result of the analysis of main components of the set of original calibration spectra

least three big groups. These groups largely reflect the genetic variety of the eucalyptus samples analyzed. In the present case, the first three components explain approximately 85% of the total variance of the spectral data. Three groups of samples can be

variedade genética dos materiais avaliados.

Essa análise preliminar dos dados permite concluir que o material genético das árvores abatidas para constituir o presente conjunto de calibração afeta os espectros NIR de refletância difusa. No entanto, o número de amostras pertencentes a cada classe é relativamente alto e inclui a variabilidade da composição química dentro de cada classe. Desta forma, um modelo global robusto, capaz de prever as propriedades de interesse em amostras das três classes principais identificadas no conjunto de calibração, pode ser construído e avaliado. É importante ressaltar que se uma ou outra classe estivesse representada com deficiência, dificilmente poder-se-ia construir um modelo global.

Antes de se proceder com a elaboração dos modelos de regressão, os resultados da avaliação da reprodutibilidade interlaboratorial dos métodos empregados na obtenção dos valores de referência para o teor de lignina, extrativos em álcool/tolueno e densidade básica foram analisados. Essa análise é muito importante para se poder estabelecer o limite do desempenho para os modelos baseados em espectroscopia NIR.

A Tabela 1 mostra os resultados obtidos para três amostras analisadas em três laboratórios diferentes. Pode-se observar que a repetibilidade intra-laboratórios é muito boa, mantendo os desvios padrão para cada série de determinações dos três parâmetros dentro do esperado, considerando-se a reprodutibilidade relatada pelos métodos padrão. Porém, diferenças inter-laboratórios, que freqüentemente excedem aquela esperada

clearly identified and attributed to the genetic variety of the evaluated materials.

This preliminary data analysis allows concluding that the genetic material of the trees felled to form the present calibration set affects the NIR diffuse reflectance spectra. Nevertheless, the number of samples belonging to each class is relatively high and includes the variability of the chemical composition within each class. Thus, a robust global model capable of predicting the properties of interest in samples from the three main classes identified in the calibration set can be constructed and evaluated. It is important to stress that if one or other class was deficiently represented, a global model could be hardly constructed.

Before proceeding to develop the regression models, the results of the evaluation of the inter-laboratory reproducibility of the methods employed for obtaining the reference values for the lignin content, extractives in alcohol/toluene and basic density were analyzed. This analysis is very important in order to be able to establish the performance limit for the NIR spectroscopy based models.

Table 1 shows the results obtained for three samples analyzed at three different labs. It can be observed that the intra-laboratory repeatability is very good, keeping the standard deviations for each series of determinations of the three parameters within the expectations, considering the reproducibility reported by the standard methods. However, inter-laboratory differences,

Parâmetros Parameters	Amostra Sample	Lab. I	dp n=6	Lab. II	dp n=6	Lab. III	dp	Reprodutividade (referência) Reproducibility (reference)
Lignina total (%)	1	31,9	0,21	30,4	0,27	35,2	—	0,9
Total lignin (%)	2	30,1	0,94	26,9	0,35	27,9	—	
	3	31,9	0,21	29,3	0,27	33,2	—	
Extrativos álcool/tolueno (%)	1	2,32	0,24	2,37	0,12	1,80	—	0,3 (estimada) (estimated)
Extractives alcohol/toluene (%)	2	1,51	0,65	1,25	0,09	0,90	—	
	3	0,54	0,16	1,00	0,06	1,00	—	
Densidade básica (kg.m ⁻³)	1	657	17	638	8	648	—	6 (repetibilidade) (repeatability)
Basic density (kg.m ⁻³)	2	428	13	406	13	454	—	
	3	517	8	492	7	528	—	

Tabela 1. Resultados médios e desvios padrão (dp) para determinações interlaboratoriais de três amostras de madeira de eucalipto

Table 1. Average results and standard deviations (dp) for inter-laboratory determinations of three eucalyptus wood samples

pelas normas, podem ser observadas para todos os parâmetros analisados.

Esses resultados devem ser avaliados com cautela em vista do pequeno número de amostras analisadas. Porém, eles servem de referência para se antecipar qual seria a performance ideal dos modelos quantitativos multivariados que foram desenvolvidos para a previsão dos mesmos parâmetros. Por exemplo, tendo sido observada uma diferença média absoluta entre os resultados dos laboratórios para a determinação de lignina total igual a 2,3%, não se deve esperar que os modelos multivariados ofereçam uma exatidão muito superior a esse valor, uma vez que esses modelos empregam esses valores de referência na sua elaboração. Por outro lado, como as amostras empregadas na elaboração dos modelos foram processadas em um único laboratório, espera-se que resultados significativamente melhores que os previstos pela concordância interlaboratorial possam ser obtidos.

Empregou-se, em todos os casos, a técnica multivariada de regressão por quadrados mínimos parciais (PLS1) para se produzir os modelos quantitativos.

Os procedimentos necessários à obtenção dos modelos descritos a seguir foram aqueles recomendados pela Norma ASTM E 1655-05. A detecção e eliminação de amostras anômalas (outliers) foram realizadas também conforme as Normas da ASTM E 1655-00 e se fundamentaram na análise dos scores nos componentes principais mais relevantes e no gráfico de influência (resíduo x leverage), gerado após análise por PCA do conjunto de 926 espectros empregados no processo de calibração. Durante a elaboração dos modelos PLS, além de se avaliar o comportamento dos scores recorreu-se aos gráficos das relações X-Y anômalas nos primeiros 6 componentes principais. Poucas amostras foram retiradas em cada caso (máximo de 23).

Na validação interna dos modelos construídos empregou-se a técnica de validação cruzada completa (full cross-validation). A validação externa foi feita com um conjunto de 160 amostras que não foram empregadas na confecção dos modelos. A validação seguiu as recomendações das Normas ASTM 1655-05.

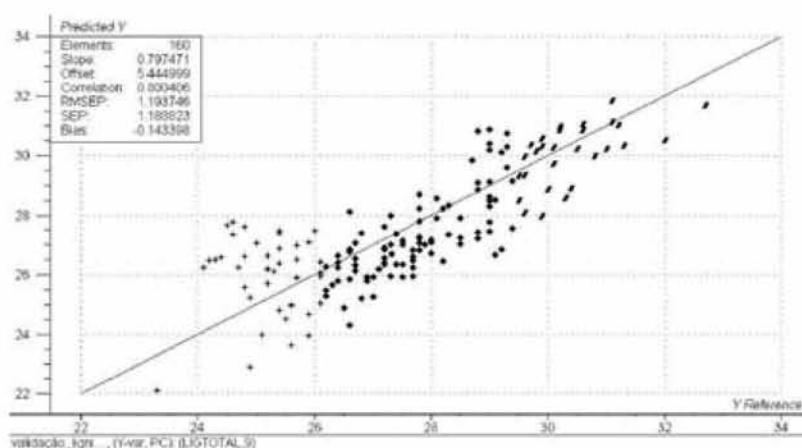


Figura 4. Resultados da validação externa para o modelo PLS construído para lignina empregando toda a faixa espectral (1100 – 2500 nm)

Figure 4. Results of the external validation for the PLS model constructed for lignin, by using the whole spectral range (1100 – 2500 nm)

which frequently exceed that expected by the standards, can be observed for all parameters analyzed.

These results should be cautiously evaluated, considering the small number of samples analyzed. Nevertheless, they serve as reference for anticipating which would be the optimum performance of the multivariate quantitative models developed for predicting the same parameters. For instance, an absolute average difference between the lab results for total lignin determination equal to 2.3% having been observed, it should not be expected that the multivariate models will offer an accuracy much higher than this value, since these models employ these reference values in their development. On the other hand, as the samples used to develop the models have been processed at a single laboratory, it is expected that significantly better results than those predicted by the inter-laboratory agreement can be achieved.

The multivariate regression technique by partial least squares (PLS1) was used in all cases to produce the quantitative models.

The procedures required to obtain the models described in the following were those recommended by ASTM E 1655-05. The detection and elimination of abnormal samples, the so-called outliers, have been also carried out in accordance with ASTM E 1655-00 and were based on the score analysis in the most relevant main components and on the influence graph (residue x leverage), generated after a PCA analysis of the set of 926 spectra used in the calibration process. During PLS model development, besides evaluating the score behavior, use was also made of the abnormal X-Y ratio graphs at the first 6 main components. A small amount of samples was taken in

each case (23 at the most).

The full cross-validation technique was used for the internal validation of the constructed models. The external validation was carried out based on a set of 160 samples that have not been used for preparing the models. The validation followed the recommendations of ASTM 1655-05.

A análise de componentes principais (PCA), realizada no conjunto de calibração contendo 926 amostras, revela que a água exerce uma influência significativa na variabilidade dos espectros de refletância. Os loadings associados à segunda componente principal (a qual explica cerca de 25% da variância dos dados originais) de uma PCA feita com os espectros originais, sem serem submetidos à primeira derivada, apresentam os maiores valores nas regiões entre 1380-1600 e 1850-2100 nm. Essas regiões referem-se claramente à influência da presença de umidade residual nas amostras de serragem de madeira. Desta forma, avaliou-se os modelos para determinação de lignina, extrativos e densidade em duas situações. Na primeira, considerou-se todo o espectro no processo de modelagem e na segunda, excluiram-se as regiões espectrais mencionadas acima, as quais podem ser atribuídas à presença da água.

Esta comparação mostrou que somente para o caso dos modelos destinados à determinação da densidade básica houve uma variação significativa dos resultados obtidos. Para esse caso, o número de variáveis latentes necessárias diminuiu significativamente (de 13 para 5), mantendo-se praticamente constantes os valores de SEP e RMSEP obtidos na validação externa e interna, respectivamente.

Os desempenhos dos modelos finais obtidos para os três parâmetros são ilustrados nas Figuras 4, 5 e 6.

A Tabela 2 mostra uma avaliação final do desempenho dos modelos PLS construídos com base nos espectros de refletância de serragem de madeira de eucalipto. Todos os modelos foram validados empregando-se a Norma ASTM

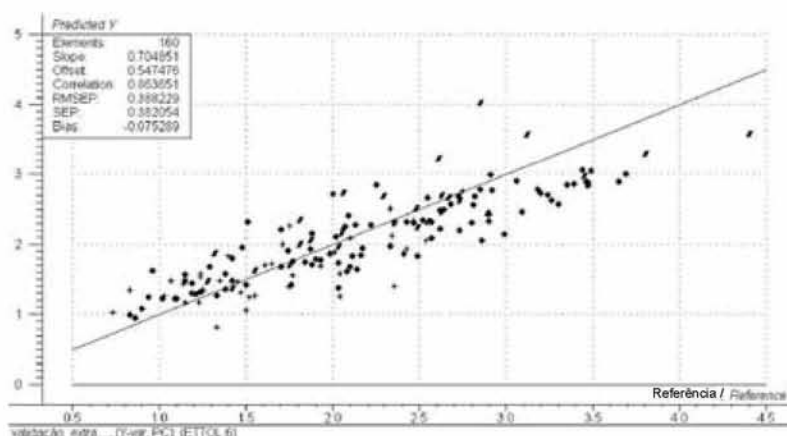


Figura 5. Resultados da validação externa para o modelo PLS para extrativos empregando toda a faixa espectral (1100 – 2500 nm)

Figure 5. Results of the external validation for the PLS model for extractives, by using the whole spectral range (1100 – 2500 nm)

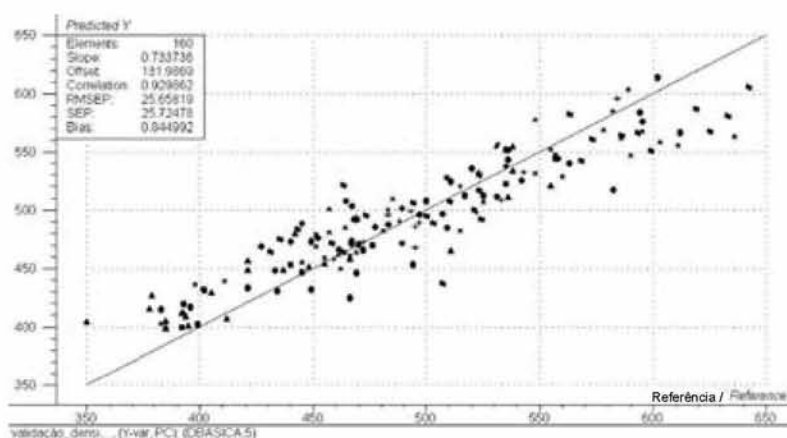


Figura 6. Resultados da validação externa para o modelo PLS para densidade básica empregando as variáveis espectrais após remoção das faixas espectrais relacionadas à contribuição da água

Figure 6. Results of the external validation for the PLS model for basic density, by using the spectral variables after removing the spectral ranges related to water contribution

The Principal Components Analysis (PCA), conducted on the calibration set consisting of 926 samples, indicates that water exerts a significant influence on the reflectance spectra variability. The loadings associated with the second principal component (which explains about 25% of the variance of the original data) of a PCA carried out with the original spectra, without being submitted to the first derivative, present the highest values in the ranges of 1380 to 1600 and 1850 to 2100. These ranges clearly refer to the influence of the presence of residual moisture in the wood sawdust samples. Thus, the models for determining lignin, extractives and density have been evaluated in two situations. In the first one, the whole spectrum was considered in the modeling process, while in the second one the above-mentioned spectral regions, which may be attributed to the presence of water, have been excluded.

This comparison showed that only in the case of the models destined for basic density determination there was a significant variation in the results obtained. For this case, the number of latent variables required decreased significantly (from 13 to 5), while the SEP and RMSEP values, obtained at the external and the internal validation, respectively, remained practically constant.

The performances of the final models obtained for the three parameters are illustrated in Figures 4, 5 and 6.

Table 2 shows a final evaluation of the performance of PLS models constructed based on the reflec-

Tabela 2. Resultados da validação interna e externa dos modelos elaborados

Table 2. Results of the internal and external validation of the developed models

Parâmetro / <i>Parameter</i>	R.E.	V.L.	Validação interna <i>Internal validation</i>		Validação externa <i>External validation</i>		
			RMSEP	r	Bias	SEP	r
Lignina total, % <i>Total lignin, %</i>	Total	9	1,15	0,762	-0,14	1,18	0,800
Extrativos em etanol/tolueno, % <i>Extractives in ethanol/toluene, %</i>	Total	6	0,38	0,904	-0,08	0,38	0,864
Densidade básica, kg m ⁻³ <i>Basic density, kg.m⁻³</i>	Parcial <i>Partial</i>	5	30	0,890	0,84	26	0,930

R.E. = região espectral: Total (1100 – 2500 nm); parcial, excluindo as absorções da água: (1380 – 1600 e 1850 – 2100 nm); V.L. = número de variáveis latentes; RMSEP = erro padrão médio obtido através de validação cruzada; SEP = erro padrão médio obtido no processo de validação externa

R.E. = spectral region; Total (1100 – 2500 nm); partial (excluding water absorptions: (1380 – 1600 and 1850 – 2100 nm); V.L. = number of latent variables; RMSEP = average standard error obtained through cross-validation; SEP = average standard error obtained by the external validation process

Tabela 3. Comparação entre os resultados obtidos por esse trabalho e outros trabalhos encontrados na literatura técnica recente / Table 3. Comparison between the results obtained by this work and other works mentioned in the recent technical literature

Parâmetro <i>Parameter</i>	Ref. (nm)	R.E.	SEP	r	V.L.	Pré-tratamento <i>Pretreatment</i>	Matriz <i>Matrix</i>
Lignina total (%) <i>Total lignin (%)</i>	Este <i>This work</i>	1100-2500	1,18	0,800	9	1ª. derivada <i>1st derivative</i>	Eucalipto (diversos) <i>Eucalyptus (miscellaneous)</i>
	6	1100-2500	0,46	0,96	-	2ª. derivada <i>2nd derivative</i>	<i>Pinus</i>
	7	800-1800	0,87	0,716	5	1ª. derivada <i>1st derivative</i>	<i>Pinus loblolly</i> , 33 anos <i>Pinus loblolly</i> , 33 years
	8	-	0,60	0,89	-	-	-
	10	400-2400	0,80	0,97	3	1ª. derivada <i>1st derivative</i>	Eucalipto (30-60 mesh) <i>Eucalyptus (30-60 mesh)</i>
	11	400-2500	0,44	0,95	-	-	<i>Pinus</i> e Eucalipto <i>Pinus and Eucalyptus</i>
Extrativos em etanol/tolueno (%) <i>Extractives in ethanol/toluene (%)</i>	Este <i>This work</i>	1100-2500	0,38	0,864	6	1ª. derivada <i>1st derivative</i>	Eucalipto (diversos) <i>Eucalyptus (miscellaneous)</i>
	8	-	0,60	0,99	-	-	-
	10	400-2400	0,30	0,87	8	1ª. derivada <i>1st derivative</i>	Eucalipto (30-60 mesh) <i>Eucalyptus (30-60 mesh)</i>
	11	400-2500	0,67	0,42	-	-	<i>Pinus</i> e Eucalipto <i>Pinus and Eucalyptus</i>
Densidade básica (kg. m ⁻³) <i>Basic density (kg m⁻³)</i>	Este <i>This work</i>	1100-2500	26	0,930	5	1ª. derivada <i>1st derivative</i>	Eucalipto (diversos) <i>Eucalyptus (miscellaneous)</i>
	3	1100-2500	27-33	0,80	6	2ª. derivada <i>2nd derivative</i>	<i>Eucalyptus globulus</i> <i>Eucalyptus globulus</i>
	4	-	19	0,94	8	Dados originais <i>Original data</i>	<i>Spruce - Norway</i> <i>Spruce - Norway</i>

R.E. = região espectral; SEP = Erro padrão médio obtido no processo de validação externa; r = coeficiente de regressão; V.L. = número de variáveis latentes

R.E. = spectral region; SEP = average standard error obtained by the external validation process; r = regression coefficient; V.L. = number of latent variables

1655-05 e mostram vieses (bias) não significativos e desempenhos adequados para a avaliação da qualidade da madeira destinada a produção de celulose e papel. A análise comparativa dos resultados contidos na Tabela 1 e os erros padrões obtidos na validação externa, constantes da Tabela 2, mostra a importância dos resultados dos métodos de referência na qualidade dos modelos multivariados obtidos pela técnica PLS aplicada aos espectros NIR.

A literatura internacional registra uma série de contribuições relacionadas ao uso da espectroscopia NIR na previsão de parâmetros de interesse da indústria de papel e celulose. A Tabela 3 resume os resultados encontrados na literatura e permite a sua comparação com os resultados obtidos neste trabalho.

A comparação dos resultados deve ser feita de forma cautelosa, pois muitos trabalhos relacionados na Tabela 3 realizaram a previsão de parâmetros com modelos construídos a partir de um conjunto restrito de amostras pertencentes a um mesmo tipo de madeira, enquanto este trabalho incluiu na elaboração dos modelos uma grande variedade de amostras (inclusive genética), tornando-os mais gerais e robustos. Além disso, vários trabalhos limitam-se a divulgar seus resultados sem revelar informações relevantes à construção dos modelos, como: o número de variáveis latentes, o tipo de pré-tratamento espectral e a faixa espectral empregada.

De uma forma geral, no entanto, há concordância dos resultados obtidos por este trabalho e aqueles encontrados na literatura considerando-se os parâmetros investigados, ou seja, teores de lignina total e de extrativos em etanol/tolueno e densidade básica da madeira, o que atesta o desempenho da técnica NIR.

CONCLUSÕES

Os resultados permitiram concluir que os espectros de refletância difusa de amostras de madeiras de eucalipto, obtidos na região do infravermelho próximo, contêm informações sobre a variedade genética das árvores das quais foram extraídas. Esse fato indica a necessidade de contemplar essa variedade nos conjuntos de amostras empregados em métodos de calibração multivariada.

Uma vez atendendo essa necessidade, como demonstrado nesse trabalho, os modelos de regressão multivariados resultantes mostram desempenho em termos de exatidão restrito apenas pela qualidade dos resultados dos métodos de referência empregados na etapa de calibração. Portanto, esses modelos podem ser empregados na substituição dos métodos convencionais via úmida com ganhos significativos em relação ao tempo e custo necessários a uma determinação de parâmetros de interesse do controle de qualidade da madeira destinada à produção de celulose e papel. ▲

tance spectra of eucalyptus wood sawdust. All models have been validated based on ASTM 1655-05 and show biases that are not significant and performances suitable for evaluating the quality of the wood destined for pulp and paper production. The comparative analysis of the results contained in Table 1 and the standard errors obtained at the external validation, indicated in Table 2, show the importance of the results of the reference methods in the quality of the multivariate models obtained by the PLS technique applied to NIR spectra.

The international literature records a series of contributions related to the use of NIR spectroscopy for predicting parameters of interest to the pulp and paper industry. Table 3 summarizes the results found in the literature and allows comparing them with the results obtained in the present work.

The comparison of the results should be cautiously made, as many works listed in Table 3 have predicted parameters based on models constructed starting from a restricted set of samples belonging to the same kind of wood, whereas the present work included a great variety (also genetic) of samples for developing the models, making them more general and robust. Moreover, several works confine themselves to divulge their results without disclosing relevant information for the model construction, such as: the number of latent variables, the type of spectral pretreatment and the spectral range used.

In general, however, there is agreement between the results obtained by the present work and those cited in the literature, considering the investigated parameters, i.e., total lignin content, as well as that of extractives in ethanol/toluene, and basic density of the wood, which attests to the NIR technique performance.

CONCLUSIONS

The results allowed concluding that the diffuse reflectance spectra of eucalyptus wood samples, obtained in the near infrared region, contain information on the genetic variety of the trees they have been extracted from. This fact indicates the need of considering this variety in the sets of samples used in multivariate calibration methods.

Once this requirement is met, as demonstrated in the present work, the resulting multivariate regression models show a performance in terms of accuracy, which is only limited by the quality of the results of the reference methods employed at the calibration stage. Therefore, these models can be used to replace the conventional wet-assay methods with significant gains with regard to time and cost required for determining parameters of interest to the quality control of the wood destined for pulp and paper production. ▲

REFERÊNCIAS / REFERENCES

1. A. ALVES, M. SCHWANNINGER, H. PEREIRA, D. S. PERES, G. CHANTRE e/and J. RODRIGUES - Determination of the lignin content of Maritime pine wood with NIR-PLSR-models based on noisy reference: Can predicted values be more precise than the reference values?, Abstracts Book, CAC-2004, Lisbon, PT.
2. ASTM 1655-05, Standard Practices for Infrared Multivariate Quantitative Analysis (2005).
3. C.A. RAYMOND e/and L.R. SCHIMLECK - Development of near infrared reflectance analysis calibration for estimating genetic parameters for cellulose content in Eucalyptus globules, Can. J. For. Res., 32:170-176(2002).
4. F.S. POKE, J.K. WRIGHT, e/and C.A. RAYMOND - Predicting extractives and lignin contents in Eucalyptus globulus using near infrared reflectance analysis, J. Wood Chem Tech., 24:55-67(2004).
5. H. BAILLERES, F. DAVIRIEUS, e/and F. HAN-PICHAVANT - Near infrared analysis as a tool for rapid screening of some major wood characteristics in a eucalyptus breeding program, Am. For. Sci., 59:479-490(2002).
6. L.R. SHIMLECK, A.J. MICHEL, C.A. RAYMOND, e/and A. MURERI - Estimation of basic density of Eucalyptus globulus using near infrared spectrometry, Can. J. For. Res., 29:194-201(1999).
7. L.R. SCHIMLECK, C.A. RAYMOND, C.L. BEADLE, G.M. DOWNES, P.D. KUBE, e/and J. FRENCH - Applications of NIR spectroscopy to forest research, Appita Journal, 53:458-464(2000).
8. NAES, T., ISAKSSON, T., FEARN, T., AND DAVIES, T. - Multivariate Calibration and Classification, NIR Publications: Chichester (2002).
9. N.L. SEFARA, D. CONRADIE, e/and P. TURNER - Progress in the use of near infrared absorption spectroscopy as a tool for the rapid determination of pulp yield plantation of eucalyptus, TAPPSA Journal, NOV:15-17(2000).
10. OTTO, M., CHEMOMETRICS - Statistics and Computer Application in Analytical Chemistry, Wiley-VCH: New York (1998).
11. PASQUINI, C. - Near Infrared spectroscopy: Fundamentals, Practical Aspects and Analytical Applications, J. Braz. Chem. Soc., 14: 198-219 (2003).
12. P. HOFFMEYER, e/and J.G. PEDERSEN - Evaluation of density and strength of Norway spruce by near infrared reflectance spectroscopy, Holz. Roh. Werkst., 53:165-170(1995).
13. T.F. YEH, H.M. CHANG, e/and J.F. KADLA - Rapid prediction of solid wood lignin content using transmittance near infrared spectroscopy, J. Agric. Food Chem., 52:1435-1439(2004).
14. WILLIAMS, P. e/and NORRIS, K. eds. - Near-Infrared Technology in Agricultural and Food Industries, American Association of Cereal Chemistry, USA, 2nd. Ed. (2001).
15. W.L.E. MAGALHÃES, J.C.D.; PEREIRA, G.I.B. MUNIZ, U. KLOCK e/and J.R.M. DA SILVA - Determinação de propriedades de madeira através do infravermelho próximo.

“O PROGRAMA DE PESQUISA E DESENVOLVIMENTO DA SUZANO CONTA COM O APOIO DA ‘FINANCIADORA DE ESTUDOS E PROJETOS – FINEP’” (www.finep.gov.br)

“THE RESEARCH AND DEVELOPMENT PROGRAM OF SUZANO HAS THE SUPPORT OF ‘FINANCIADORA DE ESTUDOS E PROJETOS – FINEP’” (www.finep.gov.br)