

DESENVOLVIMENTO DE MODELOS DE CALIBRAÇÃO NIRS PARA MINIMIZAÇÃO DAS ANÁLISES DE MADEIRAS DE *Eucalyptus* spp.

DEVELOPMENT OF NIRS CALIBRATION MODELS FOR MINIMIZATION OF *Eucalyptus* spp WOOD ANALYSIS

Leonardo Chagas de Sousa¹ José Lívio Gomide² Flaviana Reis Milagres³ Diego Pierre de Almeida⁴

RESUMO

Foi avaliada a técnica de seleção de amostras de madeira de *Eucalyptus* spp. pelo algoritmo de Kennard-Stone para desenvolvimento de modelos de calibração NIRS (Espectroscopia de Infravermelho próximo), objetivando minimizar o número de amostras, mas mantendo a precisão dos modelos. Foram utilizadas 3.369 amostras de madeiras de *Eucalyptus* spp. para desenvolvimento de modelos NIRS para densidade básica, teor de lignina e teor de extrativos em álcool-tolueno. Os modelos de calibração desenvolvidos com a totalidade das amostras para predição dos parâmetros de qualidade da madeira foram comparados com modelos desenvolvidos utilizando apenas 1.000, 500, 200 e 100 amostras selecionadas pelo algoritmo de Kennard-Stone. As análises dos parâmetros estatísticos comprovaram a similaridade dos modelos, com exceção dos modelos desenvolvidos com apenas 100 amostras, demonstrando a eficiência desta técnica no desenvolvimento de calibrações NIRS, possibilitando considerável economia de tempo e de custo das análises.

Palavras-chave: algoritmo; predição; amostragem; qualidade da madeira.

ABSTRACT

The Kennard-Stone algorithm was used to select *Eucalyptus* spp. wood samples for development of NIRS (Near-Infrared Spectroscopy) calibration models aiming to minimize number of samples but maintaining the model precisions. A large number of *Eucalyptus* spp. wood samples (3369 samples) were used to develop NIRS calibration models for the wood basic density, the lignin content and the ethanol-toluene extractives. The models developed with the total number of samples were compared with models developed using only 1000, 500, 200 and 100 samples, which were selected using the Kennard-Stone algorithm. Analysis of the models statistics parameters confirmed the similarity of all models, with exception of the 100 sample models, demonstrating the possibility of substantial savings in time and costs for wood laboratory analysis.

Keywords: algorithm; prediction; sampling; wood quality.

INTRODUÇÃO

Um grande problema na escolha aleatória de um conjunto de amostras para desenvolvimento de modelos de calibração NIRS é que muitas vezes as amostras selecionadas não representam com precisão a variabilidade dos indivíduos a serem avaliados. Normalmente, os parâmetros analisados

nas amostras disponíveis tendem a seguir um padrão de distribuição normal, o que faz com que os modelos desenvolvidos tenham baixa capacidade de predição nos extremos das curvas. Em muitas situações, como nos programas de melhoramento das empresas florestais, onde indivíduos com valores extremos são desejados, os modelos de calibração desenvolvidos podem não ser satisfatórios.

1. Engenheiro Florestal, Doutor em Ciência Florestal pela Universidade Federal de Viçosa, Av. P.H. Rolfs, s/n, Campus, CEP 36570-000, Viçosa (MG). leonardochagas@vicosa.ufv.br
2. Engenheiro Florestal, PhD., Professor Titular do Departamento de Engenharia Florestal, Universidade Federal de Viçosa, Av. P.H. Rolfs, s/n, Campus, CEP 36570-000, Viçosa (MG). jlgomide@ufv.br
3. Engenheira Florestal, MSc., Doutoranda em Ciência Florestal, Universidade Federal de Viçosa, Av. P.H. Rolfs, s/n, Campus, CEP 36570-000, Viçosa (MG). flavianamilagres@yahoo.com.br
4. Engenheiro Florestal, Mestrando em Tecnologia de Celulose e Papel, Universidade Federal de Viçosa, Av. P.H. Rolfs, s/n, Campus, CEP 36570-000, Viçosa (MG). diego.almeida@ufv.br

Recebido para publicação em 17/11/2009 e aceito em 30/07/2010

A acurácia de um modelo estatístico para prever as características tecnológicas da madeira por espectroscopia no infravermelho próximo (NIRS) depende, fundamentalmente, das amostras usadas para geração do modelo e da qualidade das análises laboratoriais utilizadas para a caracterização da mesma.

A técnica utilizada neste estudo para seleção de amostras foi o algoritmo de Kennard-Stone que se baseia na distância euclidiana para selecionar amostras que representem toda a variabilidade do conjunto de dados estudados. Sousa et al. (2007) mostraram a melhoria de um modelo de calibração desenvolvido para predição do teor de lignina Klason quando as amostras foram submetidas a esta técnica de seleção para redução do conjunto inicial de dados.

Técnicas precisas de seleção de amostras são, também, muito importantes para evitar análises químicas desnecessárias que, normalmente, são de alto custo e morosas e, às vezes, constituem de amostras que não contribuem para o aumento da capacidade de predição dos modelos desenvolvidos.

Uma possibilidade de seleção de amostras para geração de modelos de calibração NIRS, principalmente quando há um elevado número das mesmas disponíveis, consiste na utilização de métodos que se baseiam na escolha de espectros. As amostras correspondentes aos espectros selecionados é que serão, então, submetidas a análises laboratoriais. Uma das técnicas de escolha é o algoritmo Kennard-Stone, desenvolvido pelos pesquisadores R.W. Kennard e L.A. Stone, em 1969. Esse algoritmo tem como objetivo a seleção de um subconjunto de amostras X a partir de um conjunto mais amplo de amostras Y , sendo que as amostras selecionadas para formar o subconjunto X possam representar o máximo da variabilidade do conjunto Y (KENNARD e STONE, 1969).

Segundo Dantas Filho (2007), para assegurar uma distribuição uniforme do subconjunto de amostras representadas pelo espaço de dados baseado na resposta instrumental, o algoritmo de Kennard-Stone segue um procedimento orientado, no qual novas seleções são realizadas em regiões do espaço distante das amostras já selecionadas. Para esse propósito, o algoritmo emprega a distância euclidiana entre os vetores x para cada par (p, q) de amostras, calculada como abaixo:

$$Dx(p, q) = \sqrt{\sum_{j=1}^J [x_p(j) - x_q(j)]^2} \quad p, q \in [1, N]$$

Para dados espectrais, $x_p(j)$ e $x_q(j)$ são as respostas instrumentais nos j -ésimos comprimentos de onda para as amostras p e q , respectivamente, e J representa o número de comprimentos de ondas no espectro. Segundo, ainda, Dantas Filho (2007), a seleção inicia-se escolhendo o par (p_1, p_2) de amostras para os quais a distância $Dx(p_1, p_2)$ seja maior. Em cada interação subsequente, o algoritmo seleciona a amostra que apresentar a maior distância em relação a alguma amostra já selecionada. Este procedimento é repetido até o número de amostras que se deseje selecionar seja atingido.

A Figura 1 representa uma ilustração do critério de escolha de amostras adotado pelo algoritmo de Kennard-Stone. Nesta Figura as amostras foram selecionadas num conjunto total de seis amostras. As amostras 1, 2 e 4 foram selecionadas nessa ordem por serem consideradas as mais representativas em relação ao espaço amostral estudado, considerando as distâncias existentes entre cada uma delas.

A grande vantagem desta técnica, se comprovada suficientemente precisa, seria a obtenção dos espectros de todas as amostras, de maneira relativamente rápida e simples, e diminuição no número das análises laboratoriais, que demandam longo tempo e são dispendiosas.

Este estudo consistiu em testar uma técnica de seleção de amostras, baseada na distância euclidiana entre os espectros, através do algoritmo de Kennard-Stone, e utilizar apenas as amostras selecionadas para desenvolver modelos de predição e analisar se os novos modelos apresentavam a mesma precisão que o gerado com todas as amostras.

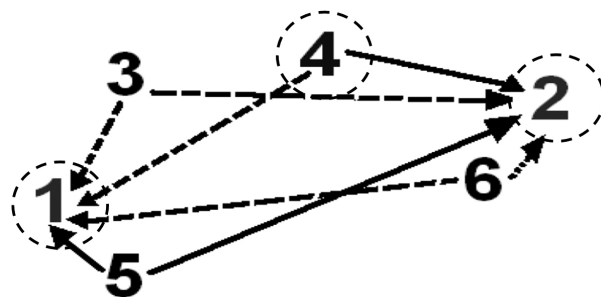


FIGURA 1: Exemplo de seleção de amostras pelo algoritmo de Kennard-Stone.

FIGURE 1: Example of sample selection based on Kennard-Stone algorithm.

MATERIAL E MÉTODO

Obtenção e preparo das amostras

Para o desenvolvimento desse trabalho foi utilizado um banco de dados com 3.369 amostras de madeiras de eucalipto, aos sete anos de idade, provenientes de empresas nacionais de celulose localizadas em diferentes regiões geográficas do Brasil, incluindo os estados de São Paulo, Bahia, Maranhão, Minas Gerais e Mato Grosso do Sul. Estas amostras fazem parte do banco de dados e arquivos do Laboratório de Celulose e Papel da Universidade Federal de Viçosa (LCP/UFV).

As amostras apresentam grande variabilidade de materiais genéticos e locais de coleta, contemplando várias características edafoclimáticas, incluindo também a alta qualidade tecnológica dos clones de *Eucalyptus* spp., atualmente utilizados no Brasil pelas principais empresas produtoras de celulose *kraft* estudadas por Gomide et al. (2005). Essa heterogeneidade é de grande importância para a construção de modelos de calibração NIRS que representem a grande diversidade dos plantios florestais nacionais de *Eucalyptus* spp.

Os parâmetros de qualidade da madeira avaliados nesse estudo foram a densidade básica, o teor de lignina Klason, abrangendo as frações insolúveis e solúveis em ácido, e os extrativos solúveis em álcool-tolueno.

As amostras de madeiras de eucalipto utilizadas neste estudo foram fornecidas pelas empresas em forma de cavacos ou de toretes, sendo os toretes transformados em cavacos em picador laboratorial do LCP/UFV. Os cavacos foram classificados em peneira com aberturas de 40x40 mm e 16x16 mm e, em seguida, foi realizada uma classificação manual para eliminar pedaços de cascas, cavacos com defeitos de corte e madeira com nós. Depois de selecionados, os cavacos foram secos ao ar e armazenados em sacos de polietileno para uniformização e conservação do teor de umidade.

Amostras representativas dos cavacos foram transformadas em serragens, utilizando-se moinho Wiley, e depois classificadas em peneiras de 40 e 60 mesh. Parte da serragem 40/60 mesh foi utilizada para as análises químicas laboratoriais e outra parte foi moída em moinho tipo ciclone que possibilita a obtenção de serragem com baixa granulometria (atravessa peneira de 100 mesh), resultando em material mais homogêneo. Esta serragem foi

acondicionada em sala com temperatura e umidade relativas constantes, aproximadamente 25°C e 50% e então usadas para obtenção dos espectros NIR.

Determinação da densidade básica e análises químicas

A densidade básica das madeiras foi determinada conforme especificado pela Associação Brasileira de Normas Técnicas (NBR 11941), o teor de lignina segundo metodologia de Gomide e Demuner (1986) e o teor de extrativos conforme norma TAPPI (T 204 om-97).

Obtenção de espectros NIR

Os espectros NIR foram coletados em espectrofotômetro NirSystem-5000 FOSS, utilizando as serragens obtidas no moinho ciclone.

Para cada amostra foi utilizada uma repetição, desta maneira a obtenção de um espectro NIR foi representado pela média de duas leituras por amostra, sendo calculada pelo programa The Unscrambler 9.6®. Os espectros NIR foram obtidos em comprimentos de onda variando de 1100 a 2500 nm, com resolução de 2 nm, totalizando 700 variáveis por amostra. O espectrofotômetro NirSystem-5000 realiza a varredura da amostra 32 vezes e o espectro gerado representa a média destas 32 varreduras. Como foram utilizadas duas leituras para cada amostra, o espectro médio obtido representava 64 varreduras.

Após a realização das análises laboratoriais de caracterização tecnológica das madeiras e da aquisição dos espectros NIR, foram desenvolvidos modelos de calibração. Os espectros NIR foram utilizados sem nenhuma transformação ou pré-tratamento espectral.

Utilizando a totalidade dos espectros (3.369 espectros), foram desenvolvidos modelos de calibração, com o auxílio do *software* The Unscrambler 9.6®, para densidade básica, teor de extrativos em álcool/tolueno e lignina total. Numa segunda etapa, foram gerados novos modelos para as mesmas características de qualidade da madeira, mas com menores números de amostras selecionadas pela técnica descrita neste estudo, utilizando o *software* Matlab. A seleção foi realizada em três etapas. Na primeira etapa foram selecionadas 1.000 amostras, na segunda etapa foram selecionadas 500 e, finalmente, na última etapa, foram selecionadas 200 amostras, onde foram desenvolvidos modelos de calibração para as três características de qualidade da madeira.

Seleção das calibrações

Desenvolvidos os modelos de predição das características das madeiras, pelo uso do *software* The Unscrambler 9.6®, a precisão dos modelos foi analisada pela validação cruzada. A validação cruzada é baseada na retirada de uma amostra do conjunto de dados e o modelo é estimado com o restante dos dados, fazendo-se a predição da amostra que não participou do modelo. Esta técnica é repetida para todas as amostras, uma a uma, proporcionando a validação cruzada completa. Posteriormente, foi realizada a validação externa, baseada em amostras que não participaram da calibração. Neste trabalho foram utilizadas para validação externa 100 amostras. Os modelos calibrados para predição adotaram os seguintes critérios de seleção:

- Coeficiente de correlação do modelo na validação cruzada (R) dado pela fórmula:

$$R = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{(n\sum X^2 - (\sum X)^2)(n\sum Y^2 - (\sum Y)^2)}}$$

- Erro padrão médio da predição (RMSEP) dado pela fórmula:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

Em que y_i é o valor conhecido da amostra i , \hat{y}_i é o valor estimado e n é o número de amostras.

O RMSEP mede a eficiência do modelo de calibração na predição da propriedade de interesse em um conjunto de amostras desconhecidas, diferente das amostras que compõem o lote de calibração.

Outro critério de seleção seria o número de fatores ou variáveis latentes (VLs), sendo que nem todas as VLs possuem informações úteis, normalmente, as últimas VLs modelam ruído inerente aos dados e, assim, a eliminação dessas

VLs frequentemente diminui a relação sinal/ruído (FILHO, 2003).

O RPD (Ratio of Performance to Deviation) representa a relação entre o desvio padrão dos valores da propriedade analisada pelo método convencional (SD) e o erro padrão da validação cruzada (SEP). Esta estatística fornece uma base de padronização do erro padrão da predição (WILLIAMS e SOBERING, 1993) e torna possível a comparação entre calibrações de diferentes propriedades.

De acordo com Schimleck e Evans (2004), para seleção de árvores em programas de melhoramento, considera-se suficiente um RPD de 2,5 para utilizações no campo das ciências florestais. Schimleck et al. (2003) afirmam que um RPD maior que 1,5 é considerado satisfatório para leituras e predições preliminares (screenings).

RESULTADOS E DISCUSSÃO

Na Tabela 1 é mostrada a amplitude de variação das características de qualidade das madeiras utilizadas para calibração NIRS e a Tabela 2 mostra a amplitude de variação das 100 amostras utilizadas para validação externa, sendo que, para todos os modelos gerados o grupo de validação externa foi o mesmo.

Como observado na Tabela 1, os parâmetros avaliados apresentaram grande diferença entre os valores máximos e mínimos, o que possibilita o desenvolvimento de modelos de calibração com alta capacidade de predição, abrangendo toda a faixa de variabilidade encontrada nos parâmetros estudados. Na Tabela 2 podem ser observados os dados usados na validação externa, obtidos manualmente de forma a abranger toda a variação de valores máximos e mínimos, fazendo com que os modelos desenvolvidos pudessem ser validados em toda sua faixa de cobertura. Nas Figuras 2, 3 e 4 são apresentados os histogramas de distribuição dos parâmetros avaliados com todos os dados utilizados.

TABELA 1: Características das madeiras das 3.369 árvores utilizadas.

TABLE 1: Wood characteristics of the 3,369 trees.

Parâmetros	Lignina Total (%)	Densidade básica (Kg m ⁻³)	Extrativos Álcool/tolueno (%)
Média	27,4	462	2,0
Máximo	33,8	696	7,8
Mínimo	21,4	330	0,6
CV (%)	5,7	9	48,0

Em que: CV = Coeficiente de variação.

TABELA 2: Características das 100 amostras utilizadas para validação externa.

TABLE 2: Wood characteristics of 100 samples used for external validation.

Parâmetros	Lignina Total (%)	Densidade básica (Kg m ⁻³)	Extrativos Álcool/tolueno (%)
Média	27,3	469	1,9
Máximo	33,4	685	7,6
Mínimo	21,7	338	0,7
CV (%)	7,2	10	46,5

Em que: CV = Coeficiente de variação.

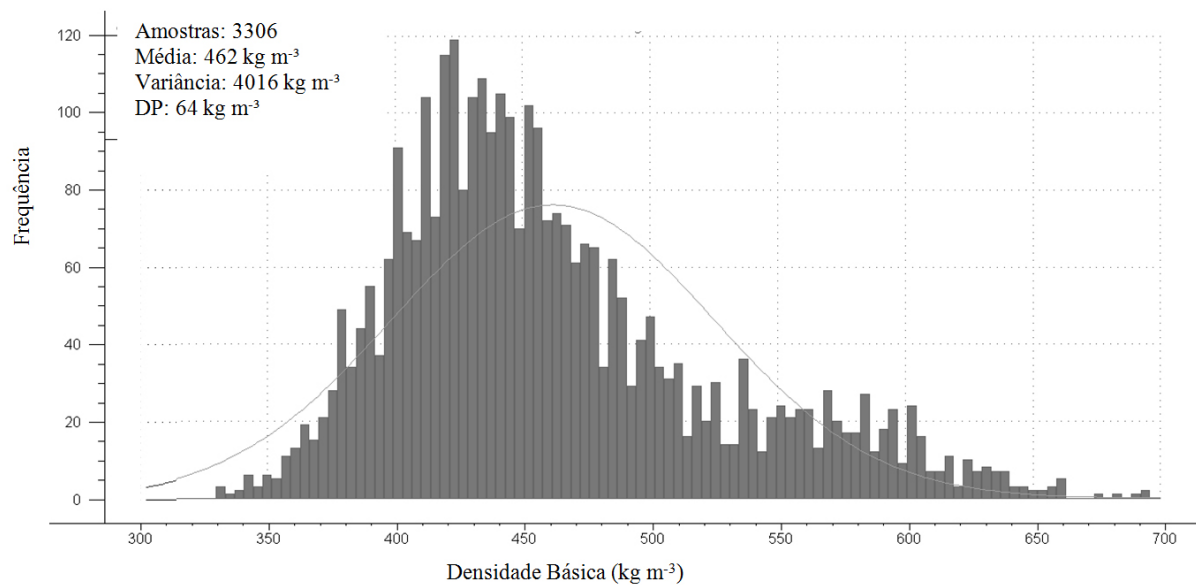


FIGURA 2: Histograma da distribuição das densidades básicas das madeiras.

FIGURE 2: Distribution histogram of wood basic density.

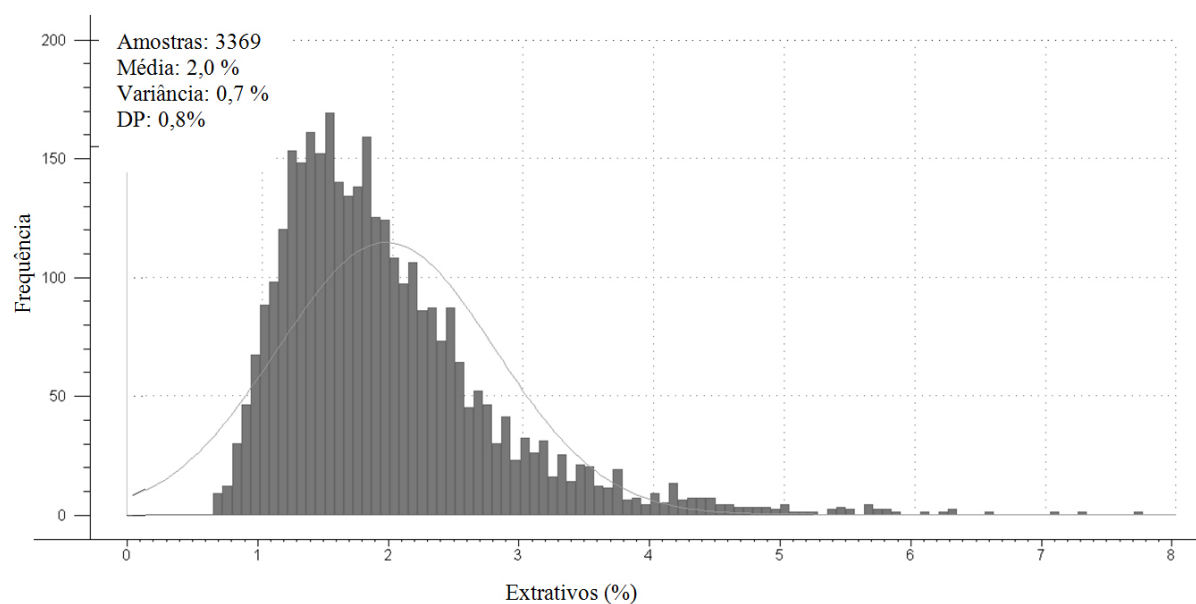


FIGURA 3: Histograma da distribuição dos teores de extrativos solúveis em álcool-tolueno.

FIGURE 3: Distribution histogram of alcohol-toluene extractive content.

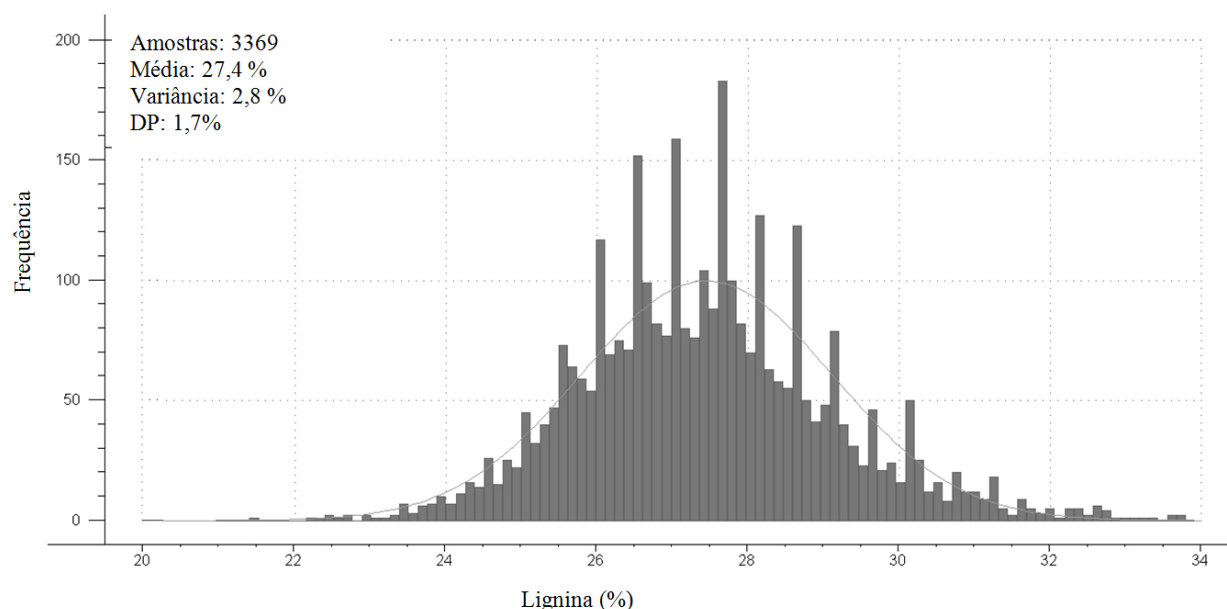


FIGURA 4: Histograma da distribuição dos teores de lignina Klason.

FIGURE 4: Distribution histogram of Klason lignin content.

Os histogramas acima mostram claramente que para todos os parâmetros estudados a concentração de amostras em torno da média foi muito grande, ou seja, ocorrência de muitos valores redundantes em torno da média. Isso indica que, para o desenvolvimento de modelos de predição, muitas amostras poderiam ser descartadas e o modelo desenvolvido deveria apresentar precisão semelhante. O menor número de amostras importantes para desenvolvimento dos modelos possibilitaria economizar grande quantidade de análises químicas, que são caras e demandam longos tempos.

Na Tabela 3 são mostrados os parâmetros estatísticos dos modelos gerados para as três características de qualidade das madeiras analisadas, utilizando todas as amostras.

Como podem ser observados na Tabela 3, os modelos de calibração desenvolvidos com todas as amostras apresentaram boa correlação, com baixo RMSEP e bom RPD, indicando que estes modelos NIRS podem ser

utilizados para predição de amostras de madeiras de eucalipto provenientes de várias regiões do Brasil.

O modelo desenvolvido para o teor de lignina Klason foi o que apresentou o menor coeficiente de correlação, quando comparado com os outros modelos. Tem sido frequentemente relatado na literatura que, dentre os parâmetros de qualidade da madeira, o teor de lignina klason é um dos que proporciona modelos mais precisos para predição desta propriedade da madeira (POKE e RAYMOND, 2006; YEH et al., 2004). Uma possível explicação é que para o desenvolvimento do modelo de predição deste estudo foram utilizados teores de lignina que poderiam ser considerados outliers, uma vez que eram valores extremos, e também deve ser levado em consideração, segundo Barthus (1999), que outliers podem ser oriundos da coleta de espectros ou dos dados laboratoriais, mas que neste trabalho não foram retirados para possibilitar comparação com o modelo desenvolvido apenas com as amostras selecionadas.

TABELA 3: Parâmetros estatísticos dos modelos de calibração desenvolvidos com todas as amostras.

TABLE 3: Statistical parameters for the calibration models developed with all samples.

Parâmetros	N. Amostras	Correlação Cruzada	RMSEP	VLs	RPD
Extrativos	3369	0,89	0,42 (%)	9	3,4
Lignina	3369	0,75	1,14 (%)	9	3,7
Densidade	3306	0,89	34 kg m ⁻³	9	3,5

Em que: RMSEP (Root Mean Square Error of Prediction) = Raiz quadrada do erro médio de predição; VLs = Variáveis Latentes; RPD (Ratio of Performance of Deviation) = Taxa de performance do desvio.

Na Tabela 4 estão os resultados da primeira etapa, onde foram selecionadas 1.000 amostras e desenvolvidos modelos de calibração para as três características de qualidade da madeira.

Como podem ser observadas na Tabela 4, as precisões dos modelos foram melhores que as obtidas com a totalidade das amostras disponíveis. É importante destacar que os números de fatores ou variáveis latentes, dos modelos de 1.000 amostras, foram inferiores aos dos modelos completos. Um fator de grande importância nas técnicas de seleção de amostras é que devem ser retirados primeiramente os *outliers* detectados na análise de componentes principais. Essa remoção é importante, pois o algoritmo é baseado em distâncias, ou seja, o algoritmo seleciona as primeiras amostras mais extremas do conjunto de calibração e se as amostras que apresentam grandes distâncias forem anômalas

e não forem removidas elas serão as primeiras a serem selecionadas, prejudicando a eficiência da técnica de seleção (HONORATO et al., 2007).

Na Tabela 5 são mostrados os valores estatísticos dos modelos de predição desenvolvidos com 500 amostras selecionadas.

A Tabela 5 mostra, também, a eficiência da seleção de amostras mesmo quando o número de amostras selecionadas foi de apenas cerca de 15% do total de amostras iniciais. Os erros de predição dos modelos desenvolvidos, apesar de relativamente altos, podem ser considerados aceitáveis, pois se aproximam dos erros de repetibilidade e reprodutibilidade encontrados em laboratórios.

Na Tabela 6 são mostrados os valores estatísticos dos modelos de predição quando foram selecionadas apenas 200 amostras.

TABELA 4: Parâmetros estatísticos dos modelos de calibração desenvolvidos com 1.000 amostras selecionadas pelo algoritmo de Kennard-Stone.

TABLE 4: Statistical parameters for the calibration models using 1,000 samples selected by the Kennard-Stone algorithm.

Parâmetros	Correlação Cruzada	RMSEP	VLs	RPD
Extrativos	0,92	0,34 (%)	6	3,7
Lignina	0,86	1,01 (%)	5	3,5
Densidade	0,91	32 kg m ⁻³	6	3,6

Em que: RMSEP (Root Mean Square Error of Prediction) = Raiz quadrada do erro médio de predição; VLs = Variáveis Latentes; RPD (Ratio of Performance of Deviation) = Taxa de performance do desvio.

TABELA 5: Parâmetros estatísticos dos modelos de calibração desenvolvidos com 500 amostras selecionadas pelo algoritmo de Kennard-Stone.

TABLE 5: Statistical parameters for the calibration models using 500 samples selected by the Kennard-Stone algorithm.

Parâmetros	Correlação Cruzada	RMSEP	VLs	RPD
Extrativos	0,91	0,36 (%)	6	3,4
Lignina	0,87	1,07 (%)	5	3,2
Densidade	0,90	33 kg m ⁻³	6	3,4

Em que: RMSEP (Root Mean Square Error of Prediction) = Raiz quadrada do erro médio de predição; VLs = Variáveis Latentes; RPD (Ratio of Performance of Deviation) = Taxa de performance do desvio.

TABELA 6: Parâmetros estatísticos dos modelos de calibração desenvolvidos com 200 amostras selecionadas pelo algoritmo de Kennard-Stone.

TABLE 6: Statistical parameters for the calibration models using 200 samples selected by the Kennard-Stone algorithm.

Parâmetros	Correlação	RMSEP	VLs	RPD
Extrativos	0,89	0,40 (%)	5	3,1
Lignina	0,85	1,23(%)	5	3,0
Densidade	0,88	35 kg m ⁻³	5	3,2

Em que: RMSEP (Root Mean Square Error of Prediction) = Raiz quadrada do erro médio de predição; VLs = Variáveis Latentes; RPD (Ratio of Performance of Deviation) = Taxa de performance do desvio.

A Tabela 6 mostra que as correlações e os erros das validações externas dos modelos desenvolvidos com 200 amostras foram ligeiramente inferiores aos obtidos com 500 amostras. Ainda assim estes modelos podem ser considerados satisfatórios para predição das características de qualidade da madeira. A grande vantagem destes modelos é que eles foram gerados com apenas 6% do conjunto total de amostras disponíveis (3.369 amostras), mostrando que a representatividade das amostras é mais importante que um elevado número de amostras.

Os resultados obtidos neste estudo mostram que a técnica de seleção de amostras é de fundamental importância, pois possibilita reduzir drasticamente o número de análises realizadas em laboratório para desenvolvimento de calibrações NIRS, sem perda de precisão e cobrindo toda a variabilidade encontrada no conjunto de dados a ser avaliado.

Nesse trabalho também foram desenvolvidos modelos de calibração para os três parâmetros avaliados selecionando apenas 100 amostras. Entretanto, os resultados encontrados para este baixo número de amostras não foram satisfatórios, mostrando ser necessário utilizar um número mínimo de amostras, acima de 100, no caso deste estudo, para que os modelos de calibração desenvolvidos fossem representativos da variabilidade do conjunto de dados. Provavelmente para saber ao certo este número mínimo de amostras um estudo mais detalhado entre 100 e 200 amostras deverá ser feito.

CONCLUSÕES

O elevado número de dados avaliados apresentava grande variabilidade dos parâmetros de qualidade da madeira, o que possibilitou a aplicação do algoritmo de Kennard-Stone para explorar o potencial de seleção. Os modelos desenvolvidos com todo o conjunto de amostras (3.369 amostras) e os desenvolvidos com 1.000, 500 e 200 amostras não apresentaram diferenças significativas. Todos os modelos apresentaram boas correlações, baixos RMSEP e alto RPD, mostrando a eficiência do algoritmo na seleção de amostras, pois permitiu reduzir o número de dados e proporcionar grande economia de tempo e custos na determinação das características tecnológicas de madeiras de eucalipto.

REFERÊNCIAS BIBLIOGRÁFICAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS

TÉCNICAS-ABNT. Rio de Janeiro. **Norma NBR n. 11941**. Março 2003.

BARTHUS, R. C. **Aplicação de métodos quimiométricos para análises de controle de qualidade de óleos vegetais utilizando espectroscopias no infravermelho e Raman**. 1999. 89 f. Dissertação (Mestrado em Química Analítica)-Universidade Estadual de Campinas, Campinas, 1999.

DANTAS FILHO, H. A. **Desenvolvimento de técnicas quimiométricas de compressão de dados e de redução de ruído instrumental aplicadas a óleo diesel e madeira de eucalipto usando espectroscopia NIR**. 2007. 158 f. Tese (Doutorado em Química) Universidade Estadual de Campinas, Campinas, 2007.

FILHO, P. A. D. A. **Estudo comparativo entre técnicas de inteligência artificial e modelos lineares em determinações quantitativas no infravermelho próximo**. 2003. 111 f. Tese (Doutorado em Química)-Universidade Estadual de Campinas, Campinas, São Paulo, 2003.

GOMIDE, J. L. et al. Caracterização tecnológica para produção de celulose da nova geração de clones de *Eucalyptus* do Brasil. **Revista Árvore**, Viçosa, v. 29, n. 1, p.129-137, 2005.

GOMIDE, J. L.; DEMUNER, B. J. Determinação do teor de lignina em material lenhoso: Método Klason modificado. **O Papel**, São Paulo, v. 47, n. 8, p. 36-38, 1986.

KENNARD, R. W.; STONE, L. A. Computer Aided Design of Experiments, **Technometrics**, v.11, n. 1, p. 137 – 148, Feb. 1969.

HONORATO, F. A. et al. Transferência de calibração em métodos multivariados. **Química Nova**, São Paulo, v. 30, n. 5, p. 1301-1312, set./out. 2007.

POKE, S. F.; RAYMOND, C. A. Predicting Extractives, Lignin, and Cellulose Contents Using Near Infrared Spectroscopy on Solid Wood in *Eucalyptus globulus*. **Journal of Wood Chemistry and Technology**, v. 26, n. 2, p. 187–199, July 2006.

SCHIMLECK, L. R.; EVANS, R. Estimation of *Pinus radiata* D. Don tracheid morphological characteristics by near infrared spectroscopy. **Holzforschung**, v. 58, n. 1, p.66-73, 2004.

SCHIMLECK, L. R.; DORAN, J. C.; RIMBAWANTO, A. Near infrared spectroscopy for cost-effective screening of foliar oil characteristics in a *Melaleuca cajuputi* breeding population. **Journal of Agricultural and Food Chemistry**, v. 51, p. 2433-2437, 2003.

SOUSA, L. C.; GOMIDE, J. L.; BALLEIRINI R. S. Importância da amostragem no uso da espectroscopia NIRS para caracterização de madeiras. In: COLÓQUIO INTERNACIONAL DE CELULOSE, 7., 2007, Belo Horizonte. **Proceedings...** Belo Horizonte, 2007.

TECHNICAL ASSOCIATION OF THE PULP AND PAPER INDUSTRY. Standard methods of technical association of the pulp and paper industry. Atlanta, 2002.

YEH, T.; CHANG, H.; KADLA, J. F. Rapid prediction of solid wood lignin content using transmittance Near-Infrared Spectroscopy. **Journal of Agricultural and Food Chemistry**, v. 52, n. 6, p. 1435-1439, Feb. 2004

WILLIAMS, P. C.; SOBERING, D. C. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. **Journal of Near Infrared Spectroscopy**, v. 1, n. 1, p. 25-33, 1993.