



**EMILLY GRACIELLY DOS SANTOS BRITO**

**VALIDAÇÃO INDEPENDENTE DE MODELOS BASEADOS  
EM ESPECTROS NO NIR PARA ESTIMATIVA DE  
DENSIDADE BÁSICA DA MADEIRA**

**LAVRAS – MG  
2022**

**EMILLY GRACIELLY DOS SANTOS BRITO**

**VALIDAÇÃO INDEPENDENTE DE MODELOS BASEADOS EM ESPECTROS NO  
NIR PARA ESTIMATIVA DE DENSIDADE BÁSICA DA MADEIRA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Ciência e Tecnologia da Madeira, área de concentração Ciência e Tecnologia da Madeira, para obtenção do título de Mestre.

Prof. Dr. Paulo Ricardo Gherardi Hein  
Orientador

Dr. Fernanda Maria Guedes Ramalho  
Coorientadora

**LAVRAS – MG  
2022**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca  
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Brito, Emilly Gracielly dos Santos.

Validação independente de modelos baseados em espectros no  
NIR para estimativa de densidade básica da madeira / Emilly  
Gracielly dos Santos Brito. - 2021.

45 p. : il.

Orientador(a): Paulo Ricardo Gherardi Hein.

Coorientador(a): Fernanda Maria Guedes Ramalho.

Dissertação (mestrado acadêmico) - Universidade Federal de  
Lavras, 2021.

Bibliografia.

1. Fenotipagem. 2. Predição. 3. Massa específica. I. Hein, Paulo  
Ricardo Gherardi. II. Ramalho, Fernanda Maria Guedes. III. Título.

**EMILLY GRACIELLY DOS SANTOS BRITO**

**VALIDAÇÃO INDEPENDENTE DE MODELOS BASEADOS EM ESPECTROS NO  
NIR PARA ESTIMATIVA DE DENSIDADE BÁSICA DA MADEIRA**

**INDEPENDENT VALIDATION OF SPECTRUM-BASED MODELS IN THE NIR  
FOR BASIC WOOD DENSITY ESTIMATION**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Ciência e Tecnologia da Madeira, área de concentração Ciência e Tecnologia da Madeira, para obtenção do título de Mestre.

APROVADA em 10 de dezembro de 2021

Dr. Alla Motta Couto – UEMS

Dr. Bruno Charles Dias Soares – UFLA

Prof. Dr. Paulo Ricardo Gherardi Hein  
Orientador

Dr. Fernanda Maria Guedes Ramalho  
Coorientador

**LAVRAS – MG  
2022**

*Á minha mãe Elisangela pelo imenso amor e dedicação que  
sempre teve comigo.*

*Ao meu pai Eivaldo por ser o meu maior exemplo de  
perseverança.*

*Dedico*

## **AGRADECIMENTOS**

À Deus, pelas infinitas bênçãos e vitórias, e por ter me sustentado e guiado em todos os caminhos da minha vida.

Aos meus queridos pais Elisângela e Erivaldo e meu pequeno irmão Enzho, pelo amor incondicional, por todo apoio e incentivo, e por terem me ensinado e ser corajosa em minhas jornadas, externo meus mais sinceros sentimentos de gratidão e admiração.

Ao Tiago, meu amado companheiro, por toda compreensão e paciência, pelo apoio e pelos sacrifícios que fez para me ajudar nessa conquista.

Ao Dr. Paulo Hein, Dr. Fernanda Maria e toda equipe NIR da UFLA pela oportunidade, por todos os ensinamentos e toda dedicação que tiveram a esse trabalho.

Aos colegas e funcionários da CTM que tornaram essa jornada mais leve e mais feliz, em especial minhas queridas amigas Dáfilla Brito, Gabrielle Andrade e Thayrine Ribeiro por terem me ajudado do início ao fim.

Às empresas Cenibra S/A e Suzano S/A pelo fornecimento do material de estudo desta pesquisa.

À Universidade Federal de Lavras e ao programa de Pós-graduação em Ciência e Tecnologia da madeira pela estrutura e suporte técnico e científico.

À Coordenação de Aperfeiçoamento de Pessoal do Nível Superior (CAPS) pelo suporte financeiro.

## RESUMO

As indústrias de base florestal têm interesse no desenvolvimento de tecnologias capazes de estimar a densidade de suas matérias-primas de forma rápida, confiável e automatizada. O objetivo desse estudo foi analisar a robustez de modelos na estimativa da densidade básica da madeira por meio de validações independentes a partir de espectros no infravermelho próximo (NIR). Para isso, espectros no NIR de amostras de madeira de lotes independentes foram aplicados à modelos preditivos pré-estabelecidos. Para o lote amostral A, madeira de discos retirados da base e a 1,30 m do solo de clones comerciais de *Eucalyptus* spp. de 5 anos de idade cultivados em espaçamento de 3x1, 3x2, 3x3, e 3x3 m, foi utilizada para determinação da densidade básica e aquisição de espectros no NIR. Amostras com dimensões de 20 x 20 x 25 mm foram retiradas dos discos de madeira na medula, região intermediária e alburno. A aquisição espectral foi feita com o auxílio de uma sonda de fibra ótica nas faces transversal e radial da madeira, e a densidade básica determinada em laboratório pela razão entre massa absolutamente seca e o volume saturado. A base de dados do lote amostral B é representada por clones de *Eucalyptus* spp. de 6,5 anos de idade plantados em espaçamento único de 9 m<sup>2</sup>. Foram usados corpos de prova com dimensões de 25 x 25 x 50 mm. Aquisição de espectros no NIR foi realizada com uso de uma sonda de fibra ótica e uma esfera de integração na face transversal da madeira. A determinação da densidade básica também realizada pela razão entre massa seca e volume saturado. Regressão dos mínimos quadrados parciais foi estabelecida considerando os espectros no NIR como variável independente (x) e a densidade básica como variável dependente (y). Os modelos foram validados nos lotes independentes de madeiras, considerando as diferenças de vias de aquisição espectral, clones, idades, espaçamentos e sítios. O modelo desenvolvido com espectros medidos usando esfera de integração nas madeiras do lote B (modelo 3) apresentaram estimativas com melhor correlação com o valor determinado em laboratório ( $R_{cv}^2 = 0.81$ ). Porém, quando aplicado no outro lote, a relação entre valores observados e preditos foi nula ( $R_p^2 = 0$ ). O modelo 3 foi capaz de identificar 7 das 15 amostras de madeiras mais leves e 8 das 15 amostras de madeiras mais densas. Por outro lado, o modelo desenvolvido com espectros medidos por fibra ótica nas madeiras do Lote A (modelo 1) apresentaram estimativas com correlação mediana com o valor determinado em laboratório ( $R_{cv}^2 = 0.53$ ). No entanto, este modelo gerou estimativas com  $R_p^2$  de 0.50 com valores observados. O modelo 1 foi capaz de identificar 7 das 15 amostras de madeiras mais leves e 11 das 15 amostras de madeiras mais densas.

**Palavras-chave:** Fenotipagem, predição, massa específica, assinatura espectral, estatística multivariada.

## ABSTRACT

Forest-based industries are interested in developing technologies capable of estimating the density of their raw materials quickly, reliably and automatically. The objective of this study was to analyze the robustness of models in estimating the basic density of wood by means of independent validations from near-infrared (NIR) spectra. For this, NIR spectra of wood samples from independent lots were applied to pre-established predictive models. For sample lot A, wood from discs taken from the base and 1.30 m from the ground of commercial clones of *Eucalyptus spp.* 5-year-old animals grown in 3x1, 3x2, 3x3, and 3x3 m spacing were used to determine the basic density and acquire NIR spectra. Samples with dimensions of 20 x 20 x 25 mm were taken from the wooden discs in the pith, intermediate region and sapwood. The spectral acquisition was performed with the aid of a fiber optic probe on the transverse and radial faces of the wood, and the basic density was determined in the laboratory by the ratio between absolutely dry mass and saturated volume. The database of sample lot B is represented by clones of *Eucalyptus spp.* 6.5 years old planted in a single spacing of 9 m<sup>2</sup>. Specimens with dimensions of 25 x 25 x 50 mm were used. The acquisition of spectra in the NIR was performed using a fiber optic probe and an integration sphere on the transverse face of the wood. The determination of the basic density was also performed by the ratio between dry mass and saturated volume. Partial least squares regression was established considering the NIR spectra as the independent variable (x) and the basic density as the dependent variable (y). The models were validated in independent wood lots, considering the differences in spectral acquisition pathways, clones, ages, spacings and sites. The model developed with spectra measured using an integration sphere in wood from batch B (model 3) presented estimates with better correlation with the value determined in the laboratory ( $R_{cv}^2 = 0.81$ ). However, when applied to the other lot, the relationship between observed and predicted values was null ( $R_p^2 = 0$ ). Model 3 was able to identify 7 of 15 samples of lighter woods and 8 of 15 samples of denser woods. On the other hand, the model developed with spectra measured by optical fiber in the woods of Lot A (model 1) presented estimates with median correlation with the value determined in the laboratory ( $R_{cv}^2 = 0.53$ ). However, this model generated estimates with  $R_p^2$  of 0.50 with observed values. Model 1 was able to identify 7 out of 15 samples of lighter woods and 11 of 15 samples of denser woods.

**Keywords:** Phenotyping, prediction, specific mass, spectral signature, multivariate statistics.

## LISTA DE FIGURAS

- Figura 1.** Discos de madeira de árvores de *Eucalyptus* aos 5 anos de idade nos espaçamentos de 3x1, 3x2, 3x3 e 3x4 m.....23
- Figura 2.** Ilustração da confecção de amostras do lote A para determinação da densidade básica e aquisição dos espectros do NIR.....24
- Figura 3.** Espectrômetro da marca Bruker, modelo Vector 22/N, Bruker Optik GmbH, Ettlingen, Germany e aquisição de espectros por sonda de fibra ótica e esfera de integração.....25
- Figura 4.** Esquema de validação independente entre os lotes de madeira A e B.....26
- Figura 5.** Valores de coeficiente de determinação da regressão linear entre valor medido e valor estimado ( $R^2_p$ ) das validações independentes entre os lotes A e B..... 31
- Figura 6.** Relação entre os valores medidos (eixo x) e os valores estimados (eixo y) pelos modelos 1 do lote A e modelo 3 do lote B em validação cruzada.....34
- Figura 7.** Relação entre os valores medidos (eixo x) e os valores estimados (eixo y) das validações independentes W e X.....37
- Figura 8.** Relação entre os valores medidos (eixo x) e os valores estimados (eixo y) das validações independentes Y e Z.....38

## LISTA DE TABELAS

<b>Tabela 1.</b> Consumo de madeira de <i>Eucalyptus</i> spp. para uso industrial em 2018.....	14
<b>Tabela 2.</b> Calibração de modelos baseados nos espectros do NIR para estimativa da densidade básica da madeira de <i>Eucalyptus</i> em validações cruzadas com diferentes números de <i>outliers</i> para os lotes A e B.....	28
<b>Tabela 3.</b> Coeficiente de determinação da regressão linear entre valor medido e valor estimado na validação independente entre os lotes A e B com diferentes percentagens <i>outliers</i> .....	30
<b>Tabela 4.</b> Capacidade preditiva do modelo 1 ( $R^2_{cv}$ 0,53) do lote A (Suz fib) para identificar as 15, 30 e 45 amostras mais e menos densas de 216 amostras totais.....	33
<b>Tabela 5.</b> Capacidade preditiva do modelo 3 do lote B (Cnb esf) para identificar as 15, 30 e 45 amostras mais e menos densas de 273 amostras totais.....	33
<b>Tabela 6.</b> Capacidade preditiva do modelo 1 do lote A (Suz fib) aplicado nos espectros do lote B (Cnb esf) e vice-versa para identificar as 15, 30 e 45 amostras mais e menos densas de 277 amostras totais.....	35
<b>Tabela 7.</b> Capacidade preditiva do modelo 3 do lote B (Cnb esf) aplicado nos espectros do lote A (Suz fib) e vice-versa para identificar as 15, 30 e 45 amostras mais e menos densas de 228 amostras totais.....	36

## LISTA DE SIGLAS

CP	Componentes principais
DAP	Diâmetro a altura do peito
NIR	Espectroscopia no infravermelho próximo
PCA	Análise de componentes principais
PLS	Regressão pelos mínimos quadrados parciais
$R^2c$	Coefficiente de determinação de calibração
$R^2cv$	Coefficiente de determinação de validação cruzada
RMSEC	Raiz quadrática média do erro padrão da calibração
RMSEcv	Raiz quadrática média do erro padrão da validação cruzada
$R^2p$	Coefficiente de determinação da regressão linear
VL	Variáveis latentes

## SUMÁRIO

1	INTRODUÇÃO .....	12
2	OBJETIVOS .....	13
2.1	Geral.....	13
2.2	Específicos .....	13
3	REVISÃO DE LITERATURA.....	14
3.1	Gênero <i>Eucalyptus</i> .....	14
3.2	Densidade básica da madeira .....	15
3.3	Espectroscopia no infravermelho próximo – NIR .....	16
3.4	O uso do NIR na caracterização tecnológica da madeira.....	17
3.5	Estatística multivariada para espectros no NIR .....	18
3.5.1	Análise de Componentes Principais .....	18
3.5.2	Regressão pelos Mínimos Quadrados Parciais .....	19
3.5.3	Calibração e validação de modelos .....	19
3.6	Fatores que influenciam na robustez dos modelos .....	20
3.6.1	Seleção do comprimento de onda.....	20
3.6.2	Tratamento matemático .....	21
3.6.3	Número de variáveis latentes.....	21
3.6.4	Identificação e eliminação de <i>outlier</i> .....	22
4	MATERIAL E MÉTODOS .....	22
4.1	Origem do material e preparo das amostras.....	23
4.1.1	Lote A.....	23
4.1.2	Lote B.....	24
4.2	Aquisição espectral .....	24
4.3	Determinação da densidade básica .....	25
4.4	Calibração e validação dos modelos .....	25
4.5	Validação independente entre os lotes A e B .....	26
5	RESULTADOS E DISCUSSÃO .....	27
5.1	Dados de referência .....	27
5.2	Validações cruzada e independente entre os lotes A e B e o efeito da exclusão de outlier na capacidade preditiva dos modelos.....	28
5.3	Acurácia dos modelos ao classificar amostras de maior e menor densidade em validações cruzadas e independentes .....	32
6	CONCLUSÃO .....	38
7	REFERÊNCIAS.....	39

## 1 INTRODUÇÃO

A densidade básica é uma propriedade tecnológica que tem se destacado como um importante indicativo de qualidade desse material. Devido sua facilidade de obtenção e importância para o estudo da matéria lenhosa, a densidade se tornou uma das propriedades mais difundidas, influenciando diretamente na escolha do material genético para utilização industrial (RIBEIRO e ZANI FILHO, 1993). Assim, o entendimento dessa propriedade é de fundamental importância frente os diversos usos da madeira, considerando sua direta interferência na eficiência dos processos de obtenção de subprodutos.

A determinação da densidade básica se tornou uma atividade importante nas indústrias de base florestal. É de grande interesse dessas indústrias que essas demandas sejam supridas de forma rápida, econômica e automatizada, dessa forma, tem-se a procura de novos métodos para a caracterização de propriedades tecnológicas. Essa crescente busca por rápidas ferramentas para o estudo do lenho estimulou o desenvolvimento de novos recursos capazes de determinar a densidade básica (HEIN et al., 2009).

Nesse cenário, destaca-se a espectroscopia no infravermelho próximo (NIR), baseada na espectroscopia vibracional, cujo principal objetivo é obter informações qualitativas e/ou quantitativas, por meio da interação de ondas eletromagnéticas com constituintes químicos de amostras orgânicas (PASQUINI, 2003). Amplamente utilizada em diferentes setores industriais, a técnica tem se mostrado eficiente na estimativa de propriedades da madeira. Especificamente para a densidade básica, estudos que utilizaram essa técnica obtiveram resultados satisfatórios para diferentes espécies, como *Eucalyptus* spp (ARRIEL et al., 2019; BALDIN et al., 2020) e *Pinus* spp (NISGOSKI, 2005; LAZZAROTTO et al., 2016).

Dentre as principais vantagens para a utilização do NIR estão as rápidas análises em tempo real, de forma não destrutiva, não invasiva, e sem necessidade de pré-tratamento nas amostras, possibilitando assim a reutilização das mesmas em demais experimentos. Contudo, o processo de construção de modelos estatísticos para a predição da propriedade de interesse, requer correlação entre dados espectrais e valores de referência, geralmente obtidos por métodos convencionais de laboratório, muitas vezes de forma destrutiva, invasiva e mais prolongada. (MUNIZ et al., 2012).

Apesar da tecnologia NIR apresentar desempenho satisfatório na caracterização da madeira como mostram estudos para discriminação de espécies nativas e plantadas (RAMALHO et al., 2018); variação espacial das propriedades da madeira (HEIN et al., 2016);

caracterização química da madeira (ESTOPA, et al., 2017) e qualidade da madeira e polpa celulósica (SANTOS et al., 2009), assume-se que é necessário avaliar o potencial da técnica em situações reais, aplicando modelos preditivos a base de dados totalmente independente para o qual o modelo não foi previamente treinado. Dessa forma, essas novas abordagens auxiliam na expansão do uso de recursos que a técnica possui.

Portanto, o objetivo desse estudo foi ajustar regressões para estimativa da densidade básica da madeira de *Eucalyptus* spp, com base nos espectros do NIR e validá-las em lotes totalmente independentes. Para isso, modelos foram desenvolvidos a partir de dois lotes distintos de amostras de madeira em termos de variação clonal, idades, tratamentos silviculturais e sítios. Assim, o modelo desenvolvido com base nas informações de um lote foi aplicado no outro lote, e vice-versa, a fim de avaliar o comportamento preditivo dos modelos quando aplicados em madeira desconhecida.

## **2 OBJETIVOS**

### **2.1 Geral**

Avaliar a robustez da regressão por mínimos quadrados parciais para estimativa da densidade básica da madeira a partir de espectros no NIR por meio de validação independente entre lotes de amostras distintos.

### **2.2 Específicos**

- 1) Avaliar o efeito da exclusão de outliers no desempenho preditivo dos modelos em validações cruzadas dentro do mesmo lote de madeira e em validações independentes em lotes totalmente desconhecidos e;
- 2) Avaliar a acurácia dos modelos independentes aplicados na base de dados original para selecionar as amostras com maior e menor densidade básica da madeira com base nos espectros no NIR.

### 3 REVISÃO DE LITERATURA

#### 3.1 Gênero *Eucalyptus*

As primeiras ações para introduzir o gênero *Eucalyptus* no Brasil ocorreram no século 19, contudo, no cenário comercial, as explorações iniciaram no século 20, contribuindo para a instauração do que hoje é conhecido como florestas plantadas (EMBRAPA, 2016). A ampla variedade de espécies, o rápido crescimento, alta produtividade e fácil adaptação a diferentes variações ambientais tornaram o *Eucalyptus* o gênero mais plantado e estudado no Brasil. Em conjunto a isso, Salazar (2012) destaca que a crescente demanda por produtos florestais impulsionou o desenvolvimento das indústrias do setor florestal, cuja expansão do uso de *Eucalyptus* spp se deu dentre outros fatores, pelo desenvolvimento de pesquisas de programas de hibridação e clonagem.

No ano de 2019 o Brasil possuía 9 milhões de hectares de florestas plantadas, sendo 6,97 milhões de hectares de plantios de *Eucalyptus* spp., o que representa 77% da área total, localizados principalmente nos estados de Minas Gerais, São Paulo e Mato Grosso do Sul. No segmento de celulose, o Brasil se manteve na segunda colocação no ranking mundial de produtores, totalizando 19,7 milhões de toneladas, sendo 16,9 milhões de toneladas de celulose produzidas a partir da madeira de *Eucalyptus* spp. (IBÁ, 2020).

De acordo com os relatórios do IBÁ, em 2018, o consumo de madeira de *Eucalyptus* spp. para uso industrial foi de 166 milhões de m<sup>3</sup>, destinadas principalmente para a produção de celulose, lenha e carvão conforme Tabela 1.

Tabela 1 - Consumo de madeira de *Eucalyptus* spp. para uso industrial em 2018

Segmento	Milhões m <sup>3</sup>
Celulose e papel	79,9
Painéis reconstituídos	6,6
Indústria madeireira	6,0
Carvão	23,4
Lenha industrial	47,1
Madeira tratada	1,4
Outros	1,6
<b>Total</b>	<b>166</b>

Fonte: adaptado de IBÁ (2019).

Pertencente ao grupo vegetal das angiospermas (folhosas), *Eucalyptus* spp. apresenta uma estrutura anatômica evoluída e complexa, com presença de vasos para condução, parênquima axial e radial para armazenamento e transporte de substâncias, e fibras como estrutura de sustentação (BURGER e RICHTER, 1991). Tais características, bem como as demais propriedades da madeira dessas espécies são largamente estudadas tanto no âmbito acadêmico quanto no empresarial, com o objetivo de atribuir suas melhores aplicações.

### **3.2 Densidade básica da madeira**

A densidade da madeira é resultante da combinação de diferentes elementos, tais como a dimensão das fibras, espessura da parede das fibras, volume de parênquimas e vasos e da proporção de lenho inicial e tardio (FOELKEL, 1971). Possui estreita relação com as demais propriedades e suas variações ocorrem em função das diferentes dimensões e proporções dos tecidos na madeira (BURGER e RICHTER, 1991). Para Latorraca e Albuquerque (2000) essas variações da densidade podem ocorrer entre diferentes espécies ou ainda por fatores externos, como os tratamentos silviculturais utilizados no povoamento, bem como as condições ambientais em que as árvores se desenvolvem.

Além das variações entre espécies distintas, há variações da densidade na mesma árvore, em diferentes posições do lenho, que ocorrem em função do percentual de lenho inicial e tardio dos anéis de crescimento (TRIANOSKI, 2013), das proporções de madeira juvenil e adulta (VIDAURRE, et. al., 2011) ou ainda da idade das árvores, gerando assim variações tanto no sentido radial quanto no sentido longitudinal (ALZATE, et al., 2005). Essas variações de densidade se tornam de certo modo um desafio para as indústrias que utilizam a madeira como matéria prima básica.

No processo de caracterização da madeira, a determinação da densidade e a variação dentro da árvore é de primordial importância no entendimento da qualidade do material lenhoso (OLIVEIRA, 2005). Nos setores industriais, o amplo uso da madeira requer densidade básica mais uniforme (HEIN et al., 2009), pois, de modo geral, madeiras que possuem menor variabilidade quanto a densidade do tronco, apresentam melhor desempenho quando submetidas a operações de processamentos mecânicos, além de expressarem maior homogeneidade e menor instabilidade em outras propriedades tecnológicas, como as físico-mecânicas (OLIVEIRA et al., 2005).

### 3.3 Espectroscopia no infravermelho próximo – NIR

A espectroscopia constitui-se de um conjunto de técnicas capazes de avaliar a interação entre radiações eletromagnéticas e diferentes matérias. Dentre as técnicas de espectroscopia utilizadas para obtenção de informações físico-químicas de amostras, o infravermelho próximo se destaca por ser um método de análises simultâneas (HEISE, 2009). É um tipo de espectroscopia vibracional, que ocorre na faixa de comprimento de onda de 700 a 2.500 nm ou de número de onda de 13.300 a 4.000  $\text{cm}^{-1}$ , cujo objetivo principal é obter informações qualitativas e/ou quantitativas (PASQUINI, 2003).

A região do infravermelho está localizada abaixo da região do espectro visível e acima da região das micro-ondas, sendo ainda subdividida em três partes: próximo, intermédio e longínquo (COSTA, 2017), das quais, o infravermelho próximo é a região do espectro eletromagnético que mais se aproxima da região visível, por isso a sua nomenclatura “próximo” (LIMA e BAKKER, 2011). Sua descoberta ocorreu em 1880 pelo músico e astrônomo Frederick William Herschel, quando estudava a frangibilidade de raios invisíveis do sol e o potencial de cores prismáticas para aquecer e iluminar objetos, (SCHWANNINGER; RODRIGUES; FACKLER, 2011). Em seus experimentos o astrônomo observou que para além do espectro visível aos olhos humanos, havia radiação capaz de aumentar a temperatura de um termômetro convencional (WHEELER, 1929).

Entretanto, nenhuma aplicabilidade para essa descoberta havia sido feita até o ano de 1881, quando o primeiro espectro na região do infravermelho próximo foi mensurado e interpretado em compostos orgânicos, onde foi observado que as moléculas de hidrogênio são o fundamento de absorção no NIR nesse tipo de amostra (WHEELER, 1929). Após sua descoberta, foram necessários cerca de 150 anos até as primeiras aplicações analíticas do NIR, o que pode ser atribuído a falta de equipamentos comerciais para explorar essa região espectral, a dificuldade de interpretação e atribuição de bandas de absorção, além da ausência de técnicas matemáticas que auxiliassem na obtenção de informações analíticas (PASQUINI, 2018).

Contudo, esse cenário mudou entre os anos 80 e 90, quando essa tecnologia se tornou amplamente disponível no mercado industrial, pois, a união de diversos fatores como (1) o surgimento de métodos quimiométricos eficazes, (2) as rotinas de avaliação e (3) o desenvolvimento de instrumentos para a técnica lançaram a espectroscopia no infravermelho próximo a uma nova era, tornando-a uma importante e usual ferramenta de controle de qualidade e processos industriais (BURNS; CIURCZAK, 2008).

### 3.4 O uso do NIR na caracterização tecnológica da madeira

No Brasil, os primeiros trabalhos realizados com o NIR ocorreram no ano de 1991 por uma instituição de pesquisa brasileira na área de ciências agrárias (PASQUINI, 2003). Atualmente, a técnica é largamente estudada para diferentes finalidades, e vem sendo utilizada nos setores industriais e de pesquisa de produtos florestais. Ramalho et al. (2018) avaliaram a eficácia da técnica NIR para a discriminação de madeiras de espécies nativas e plantadas. A partir da Análise de Componente Principais – PCA foi possível observar que o tratamento matemático da primeira derivada evidenciou a existência de dois grupos de madeira, indicado a presença de materiais vegetais com características químicas, físicas e anatômicas diferentes, possibilitando a distinção de madeira nativa e plantada pela espectroscopia no NIR.

Abjaud et al. (2017) obtiveram resultados satisfatório para a caracterização química da madeira de *Eucalyptus benthamii*. Os autores utilizaram a análise de regressão pelos mínimos quadrados parciais – PLS e observaram boa correlação entre valor medido e predito para teor de lignina, extrativos e carboidratos totais, considerando os valores estatisticamente aceitáveis e com baixo erro. Portanto, o uso do NIR foi considerado uma alternativa viável para a caracterização química, com melhores previsões estatísticas para os extrativos ( $R^2_{cv}$  0,82), possibilitando avaliação precoce de tais características e o uso em programas de melhoramento genético para a indústria de papel e celulose.

Para avaliar a qualidade da madeira e da polpa celulósica, Santos et al. (2009) realizaram aquisição de espectros no NIR em cavacos em diferentes umidades (33% e 10%), em polpa desagregada e em polpa na forma de celulose. A construção dos modelos multivariados foi feita utilizando a PLS. Os autores observaram que ao utilizar o NIR, a amostragem em cavacos gerou estatística preditiva eficaz na estimativa da densidade com 97% de precisão, não havendo influência do teor de umidade na eficiência dos modelos preditivos. Modelos desenvolvidos para o número kappa, rendimento depurado, viscosidade e teor de ácidos hexenurônicos foram considerados satisfatórios e de alta precisão principalmente para a determinação do número kappa, e de modo geral, com melhores previsões para aquisições de espectros em polpa na forma de folha.

Zanuncio et al. (2018) utilizaram o NIR para prever as propriedades da madeira de *Eucalyptus* e *Pinus* termicamente tratados com 125, 150, 175, 200 e 225 °C. Foi observado que na faixa de comprimento de 15,500 a 7.000  $\text{cm}^{-1}$  é possível observar influência do tratamento térmico na natureza dos espectros do NIR, tendo em vista que essa região espectral é mais sensível a coloração da madeira, com melhor absorbância nas madeiras tratadas a 225 °C.

Modelos desenvolvidos por meio da PLS foram satisfatórios para estimar propriedades físicas da madeira, com melhores previsões para o teor de umidade ( $R^2_{cv}$  0,94). Para o módulo de elasticidade e módulo de ruptura a melhor estatística observada foi considerando todas as amostras em grupo único, pois quando a predição foi feita separadamente por gênero, a eficiência do modelo diminuiu.

### **3.5 Estatística multivariada para espectros no NIR**

O desenvolvimento de pesquisas com o uso de muitas variáveis impulsionou o surgimento de técnicas capazes de processar e correlacionar esses dados. Com isso, a quimiometria surgiu em resposta a demanda de se obter métodos estatísticos e matemáticos capazes de compactar grandes matrizes em informações essenciais, sendo, portanto, uma interseção entre química, estatística e matemática (FERREIRA, 2008). Dentre os métodos utilizados para análise de dados de natureza multivariada, tem-se a estatística multivariada.

A estatística multivariada engloba um conjunto de técnicas poderosas (BEE, 1987), possibilitando o estudo de fenômenos complexos, uma vez que realiza tratamentos simultâneos em diferentes variáveis (BAKKE, 2008). Nesse conjunto, dentre as análises que objetivam a extração de informações químicas das variáveis, destacam-se elas a Análise de Componentes Principais (PCA) e a Regressão Pelos Mínimos Quadrados Parciais (PLS). A relação da quimiometria e do NIR é de ganho mútuo, pois, o uso da quimiometria proporciona obtenção de informações mais robustas, além de ampla aplicabilidade da técnica, ao passo que o NIR gera novos desafios para o aprimoramento das técnicas quimiométricas (PASQUINI, 2003).

#### **3.5.1 Análise de Componentes Principais**

A análise de componentes principais – PCA é uma técnica de modelagem de redução de dados, pois, transforma um conjunto inicial de variáveis em um conjunto final menor, contendo grande parte das informações do conjunto inicial e sem significativa perda de informações, a esse novo conjunto dá-se o nome de Componentes Principais – CP (HONGYU, 2015). Para Borin (2003) a base de dados de menor dimensão possibilita a extração de informações relevantes.

Utilizando a PCA em um conjunto de dados por meio de combinações lineares, as variáveis originais geram Componentes Principais (CPs), das quais a primeira componente apresenta variância máxima, ou seja, obtém maior quantidade de informações do que a subsequente, e assim sucessivamente (MOITA NETO e MOITA, 1997). As CPs geradas a

partir da PCA apresentam ortogonalidade, de modo que não se correlacionam entre si, assim, esse novo conjunto reduzido representa as variáveis mais importantes da matriz anterior, logo, na regressão pelos mínimos quadrados parciais (PLS) estão contidas somente as informações de maior relevância e melhor representatividade. (GOUVÊA et al., 2011).

### **3.5.2 Regressão pelos Mínimos Quadrados Parciais**

A análise de regressão pelos Mínimos Quadrados Parciais (PLS, do inglês *Partial Least Squares*) é uma técnica analítica para dados multivariados, na qual os problemas de análises multivariadas são resolvidos por meio de uma sequência dos mínimos quadrados simples, podendo ser utilizada em diversos problemas de modelagem (BURNS e CIURCZAK, 2008). Permite a manipulação de matriz de dados contendo grande número de variáveis e baixa quantidade de amostras, proporcionando uma correlação linear entre as variáveis espectrais e o valor da propriedade a ser estimada (PASQUINI, 2003).

Dessa forma, na PLS são geradas combinações capazes para extrair o maior número de informações possíveis das variáveis explicativas (x) com o intuito de prever as variáveis respostas (y) (SILVEIRA, 2014). Durante a obtenção dos modelos pela regressão dos mínimos quadrados parciais, as variáveis latentes, ou componentes principais, são geradas em forma de compostos lineares, por uma sequência dos mínimos quadrados, alternando algoritmos e corrigindo erros de linearidade na matriz, para assim extrair informações preditivas das amostras (DIJKSTRA, 2010).

### **3.5.3 Calibração e validação de modelos**

Modelos de calibração validados podem ser utilizados para prever os valores de determinada propriedade. É necessário determinar os valores de referência pelo método tradicional para obtenção dos valores reais correspondentes a propriedade de interesse; obter os espectros das amostras utilizando uma ferramenta NIR; correlacionar matematicamente os espectros do NIR com os valores de referência e validar os modelos para verificar relação estatisticamente aceitável (BURNS e CIURCZAK, 2007).

O objetivo da calibração é gerar um modelo que melhor relacione os espectros do NIR com os valores obtidos de forma independente (BEE e KOWALSKI, 1987). A resposta instrumental é arranjada em uma matriz x, enquanto os valores de referência ficam dispostos em um vetor y (BRAGA e POPPI, 2004). Nessa etapa, a obtenção de espectros representativos quanto a variabilidade das amostras, e a determinação correta e precisa da propriedade (valores

de referência) pelo método convencional em laboratório, influenciam significativamente na construção de modelos estatisticamente confiáveis e aceitáveis (PASQUINI, 2003).

A última etapa da construção do modelo é a validação, cujo objetivo principal é testar o desempenho do modelo calibrado. Essa é uma etapa crítica e indispensável para o uso de modelos em rotinas de ensaios em laboratório, pois, assegura a reprodutibilidade da calibração e sua robustez (GEMPERLINE, 2006). A confiabilidade do modelo preditivo pode ser avaliada por meio da validação cruzada, onde são utilizadas as mesmas amostras da calibração, ou por validação externa, quando são utilizadas amostras diferentes da calibração.

Na validação cruzada, a otimização do número adequado de variáveis do modelo é feita pela remoção de uma ou de parte das amostras da calibração, em seguida, é feita a reconstrução do modelo de modo que as amostras excluídas não estejam contidas no conjunto (PASQUINI, 2003). Em contrapartida, na validação externa um novo conjunto de amostras, diferente das que foram utilizadas na calibração é utilizado para testar o modelo (BURNS e CIURCZAK, 2007), o que gera maior demanda de material, necessidade de aquisição de maior quantidade de espectros e maior demanda por análises convencionais (SOTELO, 2006).

### **3.6 Fatores que influenciam na robustez dos modelos**

#### **3.6.1 Seleção do comprimento de onda**

Na análise de uma amostra, a radiação que é refletida da superfície é analisada como comprimento de onda (EMBRAPA, 2006). Grande quantidade de espectros acarreta no acúmulo de regiões espectrais que não fornecem informações úteis, podendo muitas vezes representar ruídos, correspondentes a regiões de baixo desempenho, baixa absorção dos constituintes de interesse, além de regiões de não linearidade (MAGALHÃES, 2014).

Nessa etapa, são selecionadas as regiões espectrais com comprimentos de onda que possuem melhor correlação com os compostos químicos de interesse (COSTA, 2017), assim, a seleção do comprimento de onda objetiva a redução dos ruídos para a otimização dos modelos. Em conjunto com demais ferramenta de otimização, a seleção do comprimento de onda contribui expressivamente para melhoria de modelos preditivos para densidade básica da madeira (HEIN et al., 2010). Contudo, a seleção deve ser feita de forma minuciosa, visto que o maior número de informações espectrais contribui para modelos de maior capacidade preditiva (MAGALHÃES, 2014).

### 3.6.2 Tratamento matemático

Espectros ainda que pertencentes as mesmas amostras podem apresentar algumas interferências, equivalentes ao espalhamento de luz e profundidade de penetração de luz. A eliminação ou redução desse impasse é feita na etapa de construção dos modelos, por meio do uso de tratamentos matemáticos aplicados nos espectros (ZENI, 2005). Para Ribeiro (2012) esse pré-processamento melhora a eficiência dos modelos multivariados, pois, reduz ou elimina a variação de informações não favoráveis. Dentre as operações matemáticas para correção dos impactos nos espectros, destaca-se o uso da primeira derivada.

A aplicação da primeira derivada nos espectros originais possibilita a eliminação dos desvios lineares e de sobreposições de bandas (GEMPERLINE, 2006; SOUZA et al., 2012) além, da exclusão dos efeitos de linha de base, que podem ocorrer em função das diferentes condições do ambiente no momento da aquisição dos espectros, como por exemplo a temperatura (ANDRADE, 2009). Contudo, a primeira derivada apresenta como desvantagem a redução da relação sinal/ruído, pois, pode haver a inclusão de ruídos nos espectros, sendo necessário nesses casos, a utilização de um alisamento eficiente para a remoção dos mesmos (PEDRO, 2002).

### 3.6.3 Número de variáveis latentes

As variáveis latentes são novas combinações lineares, resultado da decomposição das matrizes das variáveis originais, durante a construção do modelo pela PLS (LI, MARTIN e MORRIS, 2001). Nessa técnica, a escolha do número de variáveis latentes (VL) é de grande importância (MORGANO et al., 2007) uma vez que, o número ideal das variáveis em um modelo permite o alcance de melhores previsões, o que torna o modelo mais eficiente do ponto de vista estatístico, além de possibilitar a quantidade ideal de informações na modelagem, evitando, portanto, que o modelo seja subestimado ou superestimado quanto a capacidade preditiva (SEVERINO JUNIOR, 2011).

Na geração dos modelos, a seleção de demasiado número de variáveis latentes pode ocasionar a diminuição de previsões e gerar o acúmulo de informações irrelevantes, como algumas interferências e/ou ruídos. Em contrapartida, a pequena quantidade de variáveis latentes pode reduzir o número de informações importantes das amostras, ocasionando assim um ajustamento dos modelos não satisfatório (SEVERINO JUNIOR, 2011). Dessa forma, o número ideal de variáveis latentes é o que gera menor variância residual, quando utilizado o método da validação cruzada (BARTHUS, 1999).

### 3.6.4 Identificação e eliminação de *outlier*

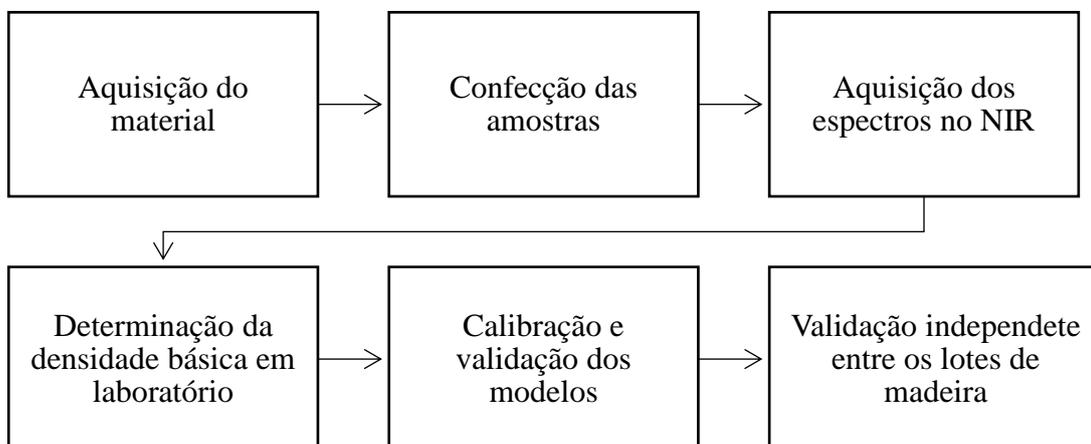
Em uma análise de dados é possível observar amostras que se diferem estatisticamente da maioria do conjunto, a essas amostras dá-se o nome de *outliers*. Tendem a representar alterações das condições experimentais ou erros de transcrição, logo, não se encaixam bem no modelo. Durante o ajuste dos modelos, quando amostras anômalas ocorrem em grupos, a exclusão de apenas um *outlier* por vez poderá não mudar consideravelmente o modelo, sendo necessário utilizar ferramentas que excluam todo o grupo de forma simultânea (GEMPERLINE, 2006).

Entretanto, ainda que diferentes, essas amostras não devem ser apenas descartadas do modelo durante a otimização, é necessário considerar uma análise criteriosa em função do comportamento diferente do *outlier*, pois, esses comportamentos distintos podem indicar alguma característica diferente e importante do material, não contida nas demais amostras. Portanto, em situação semelhante é importante considerar a inclusão de mais amostras dessa classe para que se obtenha uma maior variabilidade e representatividade no conjunto (PASQUINI, 2003).

## 4 MATERIAL E MÉTODOS

O fluxograma abaixo contém o resumo das atividades a serem desenvolvidas para atingir os objetivos desta pesquisa.

Fluxograma 1 – Resumo da metodologia utilizada



## 4.1 Origem do material e preparo das amostras

### 4.1.1 Lote A

Foram utilizados híbridos de *Eucalyptus urophylla* x *Eucalyptus grandis* de um plantio comercial da empresa Suzano, localizado no município de Itamarandiba – MG, Brasil, destinado a produção de celulose. O plantio é constituído de quatro variedades clonais distribuídos em espaçamentos de 3, 6, 9, e 12m<sup>2</sup>. As árvores apresentam 5 anos de idade e em média 8 cm de Diâmetro a Altura do Peito – DAP nos espaçamentos menores (3 e 6 m<sup>2</sup>) e 16 cm em média nos espaçamentos maiores (9 e 12 m<sup>2</sup>) como mostra a Figura 1.

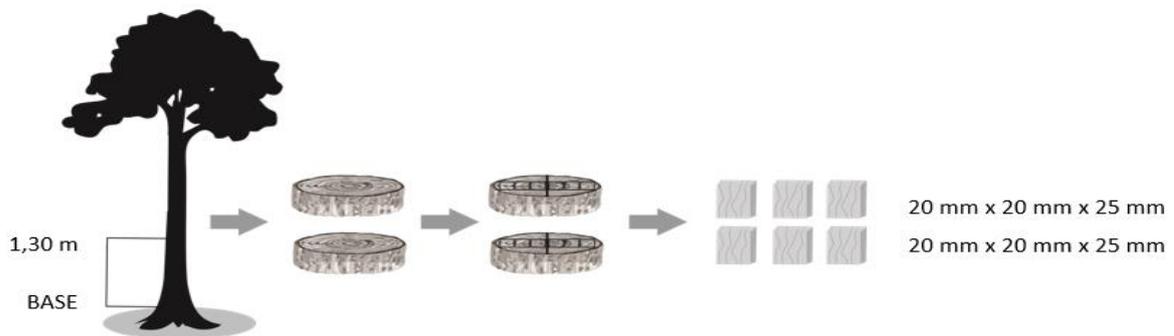
Figura 1 – Discos de madeira de árvores de *Eucalyptus* aos 5 anos de idade nos espaçamentos de 3x1, 3x, 3x3 e 3x4 m.



Fonte: do autor (2021).

Uma árvore de cada clone foi selecionada em cada espaçamento, totalizando 16 indivíduos. Discos de madeira foram obtidos da base e a 1,30 m do solo, considerando a altura comercial em 0, 25, 50, 75 e 100 %. O seccionamento das amostras foi feito considerando as regiões da medula, intermediária e albarno, com uso de uma serra fita no Laboratório de Usinagem da Madeira – UFLA (Figura 2). Ao final do processo as amostras apresentavam dimensões de 20 x 20 x 25 mm nos sentidos radial, tangencial e longitudinal respectivamente, sendo direcionadas para a aquisição espectral e determinação da densidade básica em laboratório.

Figura 2 – Preparo das amostras do lote A para determinação da densidade básica e aquisição dos espectros do NIR.



Fonte: do autor (2021).

#### 4.1.2 Lote B

A base de dados do lote B é composta pelas informações espectrais e de densidade básica previamente publicados por Costa et al. (2018). Essa amostragem de madeira é representada por híbridos de *Eucalyptus urophylla* x *Eucalyptus grandis* com 6,5 anos de idade, provenientes de um plantio comercial da empresa Vallourec Florestal Ltda, situada na cidade de Paraopeba – MG, Brasil, além de clones de híbridos de *Eucalyptus urophylla* x *Eucalyptus grandis* oriundos da empresa Cenibra S/A, no município de Belo Oriente – MG, Brasil. Ambos estabelecidos em talhões adjacentes de um terreno plano, com espaçamento único de 9 m<sup>2</sup>. Toras de madeiras foram obtidas para confecção de amostras com dimensões de 25 x 25 x 50 mm no sentido radial, tangencial e longitudinal respectivamente, usinadas com uma serra fita (COSTA et al., 2018).

#### 4.2 Aquisição espectral

As amostras permaneceram em sala condicionada a 20 °C para a aquisição dos espectros e apresentaram cerca de 10,40% de umidade. Foi utilizado um Espectrômetro (marca Bruker, modelo MPA, Bruker Optik GmbH, Ettlingen, Germany) (Figura 2) conectado a um computador contendo o programa *Opus* versão 7.5 no Laboratório de Biomateriais – UFLA. Na gama de 12.500 cm<sup>-1</sup> a 3.500 cm<sup>-1</sup> com resolução espectral de 8 cm<sup>-1</sup> em modo de reflexão difusa, foi feita aquisição de uma assinatura espectral na face transversal e outra na face radial de todas as amostras. Para o lote A utilizou-se somente a fibra ótica como via de aquisição. Para o lote B utilizou-se tanto a fibra ótica quanto a esfera de integração (COSTA et al. 2018). A resposta espectral da madeira foi convertida em números utilizando o programa *Opus*.

Figura 3 – Espectrômetro no NIR com transformada de Fourier e dois métodos de aquisição espectral: esfera de integração e fibra ótica.



Fonte: do autor (2021)

#### 4.3 Determinação da densidade básica

O mesmo lote amostral utilizado para a aquisição dos espectros no NIR foi utilizado para determinação da densidade básica em laboratório. Foi realizada uma adaptação do método de imersão em água, conforme procedimentos estabelecidos pela NBR 11 941 (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS – ABNT, 2003).

#### 4.4 Calibração e validação dos modelos

O ajuste dos modelos foi feito utilizando o software *The Unscrambler*. Regressão pelos mínimos quadrados parciais – PLS foi estabelecida considerando os espectros no NIR como variável independente (y), e a densidade básica obtida em laboratório como variável dependente (x). Na calibração foram considerados espectros das faces transversais e radiais das amostras de madeira. Para otimização dos modelos, foi aplicado aos espectros originais o tratamento matemático da primeira derivada, no intuito de remover informações não relevantes. Foi feita seleção de 1.300 absorvâncias (entre 9000 a 4000  $\text{cm}^{-1}$ ), para assim excluir regiões espectrais com informações que não contribuem para construção dos modelos (entre 12000 a 9000  $\text{cm}^{-1}$ ).

Após a calibração dos modelos, estes foram validados somente por validação cruzada randômica. O teste de incerteza de Martens foi utilizado para indicar no gráfico de coeficiente de regressão as regiões espectrais de maior relevância na geração dos modelos. No gráfico de resíduos foram identificados e excluídos *outliers* para remover amostras com comportamento distinto, bem como, avaliar o efeito desse tratamento na capacidade preditiva de diferentes

modelos, observando o maior coeficiente de determinação de validação cruzada ( $R^2_{cv}$ ), o número de variáveis latentes da calibração (LV) e o erro da validação cruzada (RMSE) como mostra Equação 2.

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

Onde:

$\hat{y}_i$  = valor estimado pelo modelo da amostra i

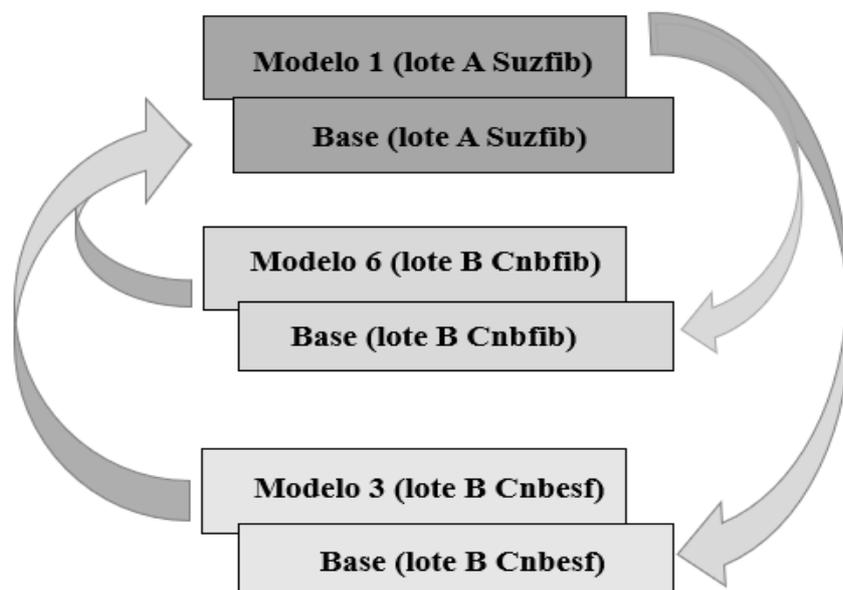
$y_i$  = valor de referência da amostra i

NP = número de amostras do lote de validação cruzada

#### 4.5 Validação independente entre os lotes A e B

A análise da capacidade preditiva dos modelos para densidade básica da madeira de *Eucalyptus* quando aplicado ao material externo, foi feita inicialmente utilizando os modelos 1, 3 e 6 da tabela 2, no programa *The Unscrambler*. Modelos gerados a partir de espectros do lote de A (modelo 1) foram aplicados aos espectros do lote B (base de dados desconhecido do modelo) obtidos tanto por sonda de fibra ótica quanto por esfera de integração. Processo inverso também foi realizado aplicando modelos do lote B (3 e 6) obtidos pelas duas vias de aquisição, nos espectros do lote A obtidos por sonda de fibra ótica, conforme ilustrado na Figura 6.

Figura 4 – esquema de validação independente entre os lotes de madeira A e B



Lote A Suzfib – modelos/base de dados do lote A gerados a partir de espectros coletados pela fibra ótica em amostras da Suzano; Lote B Cnb fibra – modelos/base de dados do lote B gerados a partir de espectros coletados pela fibra ótica em amostras da Cenibra; Lote B Cnb esfera – modelos/base de dados do lote B gerados a partir de espectros coletados pela esfera de integração em amostras da Cenibra. Fonte: do autor (2021).

O resultado dessas validações independentes nos parâmetros pré-estabelecidos acima foi plotado em uma tabela dinâmica no programa *Excel* versão 2019. Utilizando a tabela foi avaliada a acurácia dos modelos para selecionar as amostras de maior e menor densidade tanto nas validações cruzadas, quanto nas validações independentes entre os lotes de madeira. Assim, por meio de análise visual e cálculo de porcentagem foi possível comparar o percentual de acerto da densidade estimada pelos modelos treinados para a base de dados conhecida, com o percentual de acerto dos modelos treinados para a base de dados desconhecida. O efeito da exclusão de *outliers* nas validações independentes foi avaliado por meio da obtenção da correlação entre os valores de referência e os valores estimados pelos modelos com diferentes números de *outliers*.

## 5 RESULTADOS E DISCUSSÃO

### 5.1 Dados de referência

O valor médio da densidade básica da madeira de *Eucalyptus* sp. do lote A foi de 0,470 g/cm<sup>-3</sup>, variando de 0,394 g.cm<sup>-3</sup> a 0,534 g.cm<sup>-3</sup>. De acordo com Costa et al. (2018) o valor médio para densidade da madeira de *Eucalyptus* sp. do lote B foi de 0,462 g.cm<sup>-3</sup> com variação de 0,294 g.cm<sup>-3</sup> e 0,679 g.cm<sup>-3</sup>. Resultado semelhante foi encontrado por Milagres et al. (2013) ao avaliarem o efeito da idade na construção de modelos no NIR para estimativa de propriedades da madeira de *Eucalyptus* spp., observando valor médio para densidade de 0,432 g.cm<sup>-3</sup>, mínimo de 0,367 g.cm<sup>-3</sup> e máximo de 0,483 g.cm<sup>-3</sup>.

Outra pesquisa realizada por Neto et al. (2020) obteve valor médio de 0,512 g.cm<sup>-3</sup> ao analisarem características do carvão e madeira de *Eucalyptus* sp., enquanto Magalhães (2020) ao estudar a densidade básica da madeira de clones de *Eucalyptus* em diferentes espaçamentos encontrou valores de 0.375 g/cm<sup>-3</sup> a 0.445 g/cm<sup>-3</sup>. Observa-se, portanto, que valores semelhantes aos do lote A e B foram obtidos por demais autores que estudaram propriedades da madeira de *Eucalyptus*, assim, a densidade observada em ambos os lotes corrobora com as descritas em literatura.

## 5.2 Validações cruzada e independente entre os lotes A e B e o efeito da exclusão de outlier na capacidade preditiva dos modelos

A tabela 2 mostra a estatística preditiva dos modelos do lote A e B em diferentes percentuais de exclusão de amostras anômalas em validações cruzadas. Percebe-se uma relação diretamente proporcional entre a exclusão de amostras anômalas e a capacidade preditiva dos modelos, pois, em ambos os lotes de madeiras o maior percentual de exclusão gerou maior coeficiente de determinação tanto de calibrações quanto de validações, entretanto, não alterou expressivamente o erro padrão de calibração e validação. Esse resultado é esperado, tendo em vista que a principal função desse tratamento é eliminar amostras que não contribuem beneficemente para a melhor calibração do modelo.

Tabela 2 – Calibração de modelos baseados nos espectros do NIR para estimativa da densidade básica da madeira de *Eucalyptus* em validações cruzadas com diferentes percentagens de *outliers* para os lotes A e B.

Modelo	Lote amostral	Total de mostras	Out %	LV	R <sup>2</sup> c	RMSEC	R <sup>2</sup> cv	RMSECV
1	A Suz fib	216	6%	6	0.58	33	0.53	35
2	A Suz fib	203	12%	6	0.65	29	0.60	32
3	B Cnb esf	273	2%	9	0.84	26	0.81	28
4	B Cnb esf	277	0.07%	9	0.82	28	0.78	31
5	B Cnb fib	265	5%	9	0.72	33	0.67	37
6	B Cnb fib	264	6%	7	0.73	33	0.68	36

A Suz fib – modelo gerado a partir de espectros obtidos por sonda de fibra ótica de amostras do lote A Suzano; B Cnb esf - modelo gerado a partir de espectros obtidos por esfera de integração de amostras do lote B Cenibra; B Cnb fib - modelo gerado a partir de espectros obtidos por sonda de fibra ótica de amostras do lote B Cenibra; Out % - porcentagem de outliers excluídos do modelo; LV – número de variáveis latentes;  $R^2c$  – coeficiente de determinação da calibração; RMSEC - raiz quadrática média do erro padrão da calibração ( $g/cm^3$ );  $R^2cv$  – coeficiente de determinação da validação cruzada; RMSECV – raiz quadrática média do erro padrão da validação cruzada ( $g/cm^3$ ).

Resultado similar foi observado por Hein, Lima e Chaix (2009) ao avaliarem otimizações baseadas no NIR para a madeira de *Eucalyptus*. PLS foi utilizada para relacionar propriedades da madeira e os espectros no NIR. Os autores constataram que o modelo com exclusão de outliers apresentou  $R^2cv$  de 0,77 enquanto o modelo sem a exclusão de outlier apresentou de  $R^2cv$  de 0,71. No mesmo lote de amostra, mas submetido ao tratamento matemático da primeira derivada, o modelo com exclusão de outlier apresentou  $R^2cv$  de 82 enquanto o modelo sem exclusão de outlier apresentou 78 de  $R^2cv$ .

Em estudo para discriminar espécies similares de madeiras Amazônicas (*Carapa guianensis* Aubl, *Cedrela odorata*, *Erismia uncinatum* Warm, *Micropholis melinoniana* Pierre, *Hymenea coubaril* L. e *Swietenia macrofila* King), Soares et al. (2017) utilizaram NIR portátil e PLS-DA. Na otimização dos modelos, os autores observaram que com a exclusão de 1,3% de amostras consideradas anômalas na calibração, os modelos apresentaram taxas de acerto de 90% variando de 90,4% a 99,7%, demonstrando elevada capacidade preditiva.

Na tabela 3 consta estatística preditiva das validações independentes entre os lotes A e B. Observa-se mesma tendência da validação cruzada em cada lote quanto ao efeito da exclusão de outliers. Não foi observado variação alguma no  $R^2c$  e  $R^2cv$  dos modelos 1 e 2 do lote A, quando utilizados em um lote desconhecido, mesmo considerando todas as diferenças citadas, incluindo a das vias de aquisição dos espectros do lote B (sonda de fibra ótica e esfera de integração). Nos modelos do lote B também não houve variação em tais parâmetros quando utilizados nos espectros do lote A. Os modelos com maior percentual de exclusão de outliers permanecem com maior  $R^2cv$ . Portanto, no que se refere a eliminação de outliers, a variabilidade entre os lotes estudados não afetou os parâmetros estatísticos dos modelos, sendo mantidos quando realizada validação independente.

Entretanto, observa-se que o coeficiente de determinação da regressão linear entre os dados da validação Y é 0. Nas validações X e Z os valores de  $R^2p$  foram bem próximos, sendo 0,30 e 0,25 respectivamente, enquanto na validação W foi observado  $R^2p$  0,50.

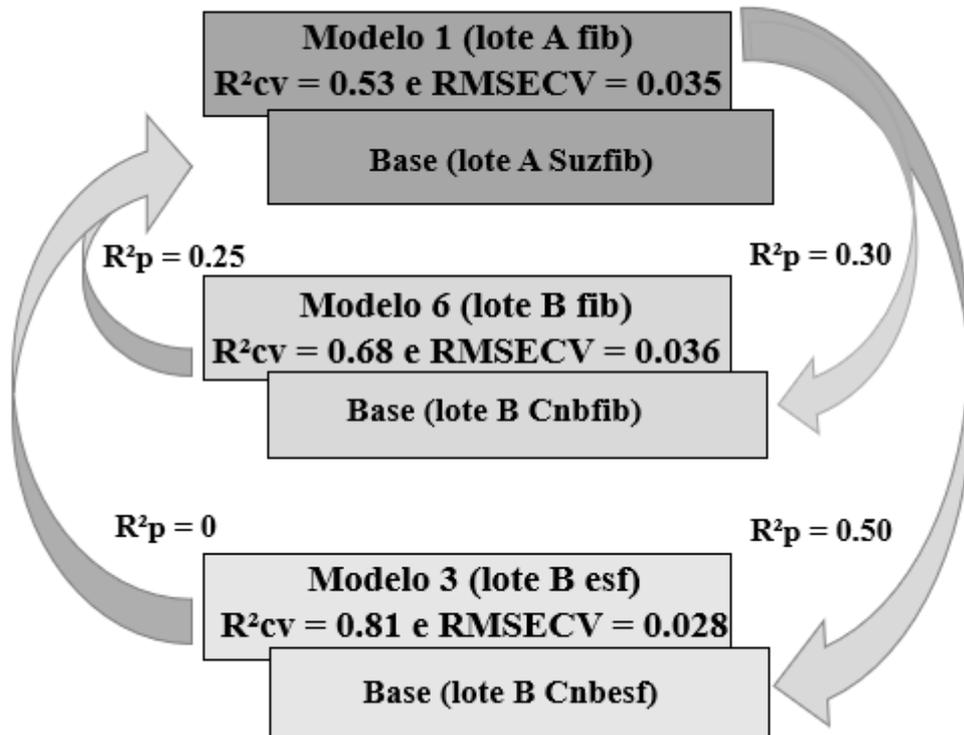
Tabela 3 – Coeficiente de determinação da regressão linear entre valor medido e valor estimado na validação independente entre os lotes A e B com diferentes percentagens de *outliers*.

Validação	Out %	LV	R <sup>2</sup> c	RMSEC	R <sup>2</sup> cv	RMSEcv	R <sup>2</sup> p
Validação W:							
Modelo 1 lote A (Suz fib) em espectros do lote B (Cnb esf)	6	6	0.58	33	0.53	35	0.50
Validação X:							
Modelo 1 lote A (Suz fib) em espectros do lote B (Cnb fib)	6	6	0.58	33	0.53	35	0.30
Validação Y:							
Modelo 3 lote B (Cnb esf) em espectros do lote A (Suz fib)	2	9	0.84	26	0.81	28	0
Validação Z:							
Modelo 6 lote B (Cnb fib) em espectros do lote A (Suz fib)	6	7	0.73	33	0.68	36	0.25

Suz – Suzano; fib – sonda de fibra ótica; Cnb – Cenibra; esf – esfera de integração; Out % – percentagem de outliers excluídos do modelo; LV – número de variáveis latentes; R<sup>2</sup>c – coeficiente de determinação da calibração; R<sup>2</sup>cv – coeficiente de determinação da validação cruzada; RMSEC - raiz quadrática média do erro padrão da calibração (g/cm<sup>3</sup>); RMSEcv – raiz quadrática média do erro padrão da validação cruzada (g/cm<sup>3</sup>); R<sup>2</sup>p - coeficiente de determinação da regressão linear entre valor medido e valor estimado.

Na Figura 5 é ilustrado o esquema das validações W, X, Y e Z e seus respectivos coeficientes de determinação da regressão linear entre valor medido e valor estimado. De maneira geral, a melhor correlação obtida foi na validação W quando aplicado modelo 1 do lote A (Suz fib) nos espectros do lote B (Cnb esf) observando R<sup>2</sup>p de 0,50, entretanto, quando processo inverso é realizado e modelo 3 do lote B (Cnb esfera) é aplicado aos espectros do lote A (Suz fibra) o R<sup>2</sup>p é 0.

Figura 5 – Valores de coeficiente de determinação da regressão linear entre valor medido e valor estimado ( $R^2_p$ ) das validações independentes entre os lotes A e B.



Lote A Suzfib – modelos/base de dados do lote A gerados a partir de espectros coletados pela fibra ótica em amostras da Suzano; Lote B Cnb fibra – modelos/base de dados do lote B gerados a partir de espectros coletados pela fibra ótica em amostras da Cenibra; Lote B Cnb esfera – modelos/base de dados do lote B gerados a partir de espectros coletados pela esfera de integração em amostras da Cenibra.  $R^2_{cv}$  – coeficiente de determinação da validação cruzada;  $RMSECV$  – raiz quadrática média do erro padrão da validação cruzada;  $R^2_p$  - coeficiente de determinação da regressão linear entre valor medido e valor estimado. Fonte: do autor (2021).

Apesar do modelo 3 do lote B apresentar o maior  $R^2_{cv}$ , o valor de  $R^2_p$  dessa validação foi o menor observado, o que pode ser explicado por uma comparação das áreas de aquisição espectral. Conforme descreveu Costa et al. (2018), a área varrida pela sonda de fibra ótica é muito menor (~1 mm de diâmetro) enquanto a esfera integradora mede a reflexão da radiação numa área com diâmetro de 10 mm. Assim, o maior tamanho de ponto da esfera integradora fornece melhor representação da amostra de madeira, particularmente cobrindo potencialmente mais de um único anel de crescimento. Além disso, há significativa variação genética, de idade e sítio entre os clones dos lotes avaliados. Tais fatores conjuntamente podem ter gerado grande incompatibilidade entre o modelo e os espectros, contribuindo assim para o  $R^2_p$  inexistente.

Resultado similar foi encontrado ainda por Costa et al. (2018) ao realizarem teste semelhante em estudo da densidade básica da madeira de *Eucalyptus*. Um modelo gerado por

espectros obtidos a partir da esfera de integração foi aplicado em espectros obtidos por sonda de fibra ótica, entretanto, no mesmo lote de madeiras em superfície transversal usinado por serra fita. Mesmo utilizando amostras que não possuem variabilidade quanto a genética, sítio ou idade, foi observado baixa correlação ( $R^2$  0,42) em relação ao teste inverso, quando aplicado um modelo de sonda de fibra ótica em espectros de esfera de integração ( $R^2$  0,81). Com isso, constatou-se que o modelo a partir de esfera de integração não apresentou bom ajuste para os espectros a partir da sonda de fibra ótica.

Quando utilizado o modelo 1 do lote A (Suz fibra) aplicado aos espectros do lote B (Cnb fibra) e depois modelo 6 do lote B (Cnb fibra) em espectros do lote A (Suz fibra), não houve grande variação nos valores de  $R^2_p$ , sendo 0,30 e 0,25 respectivamente. Nesse caso, em ambos os lotes os espectros no NIR foram obtidos pela mesma via de aquisição, o que pode ter gerado compatibilidade nos valores da regressão linear e ainda baixa variação entre eles. Para essas validações, o maior  $R^2_{cv}$  (modelo 6 com  $R^2_{cv}$  0,68) também não contribuiu para o maior  $R^2_p$ . Portanto, observa-se que não houve relação entre o coeficiente de determinação de validação cruzada e o coeficiente de determinação de regressão linear.

### **5.3 Acurácia dos modelos ao classificar amostras de maior e menor densidade em validações cruzadas e independentes**

Nas tabelas 4 e 5 constam a capacidade preditiva dos modelos 1 (lote A) e modelo 3 (lote B) ao selecionarem amostras de maior e menor densidade em validações cruzadas. No lote A ao ranquear as amostras mais leves, o modelo 1 ( $R^2_{cv}$  0,53) classificou corretamente 7 quando consideradas 15 amostras (46,67 %), 15 consideradas 30 amostras (50 %) e 24 considerando 45 amostras (53,33 %). Ao classificar as amostras mais densas, o modelo selecionou corretamente 11 de 15 amostras (73,33 %), 17 de 30 amostras (56,67 %) e 27 de 45 amostras (60 %). Observa-se que nos três tipos de ranqueamento, o modelo apresentou melhor desempenho ao selecionar as amostras mais densas, com maior percentual de acertos considerando as 15 primeiras amostras do lote de madeiras (73,33 %).

No lote B, ao classificar as amostras mais leves, o modelo 3 ( $R^2_{cv}$  0,81) ranqueou corretamente 7 de 15 amostras (46,67 %), 21 de 30 (70 %) e 29 de 45 (64,44 %). Ao considerar as amostras mais densas, o modelo classificou corretamente 8 de 15 (53,33 %), 17 de 30 (56,67 %) e 30 de 45 (66,67). De modo geral, o modelo do lote B também ranqueou mais corretamente as amostras mais densas, entretanto o maior percentual de acerto do modelo foi de 70 % ao classificar 21 de 30 amostras mais leves.

Tabela 4 – Capacidade preditiva do modelo 1 ( $R^2_{cv}$  0,53) do lote A (Suz fib) para identificar as 15, 30 e 45 amostras mais e menos densas de 216 amostras totais.

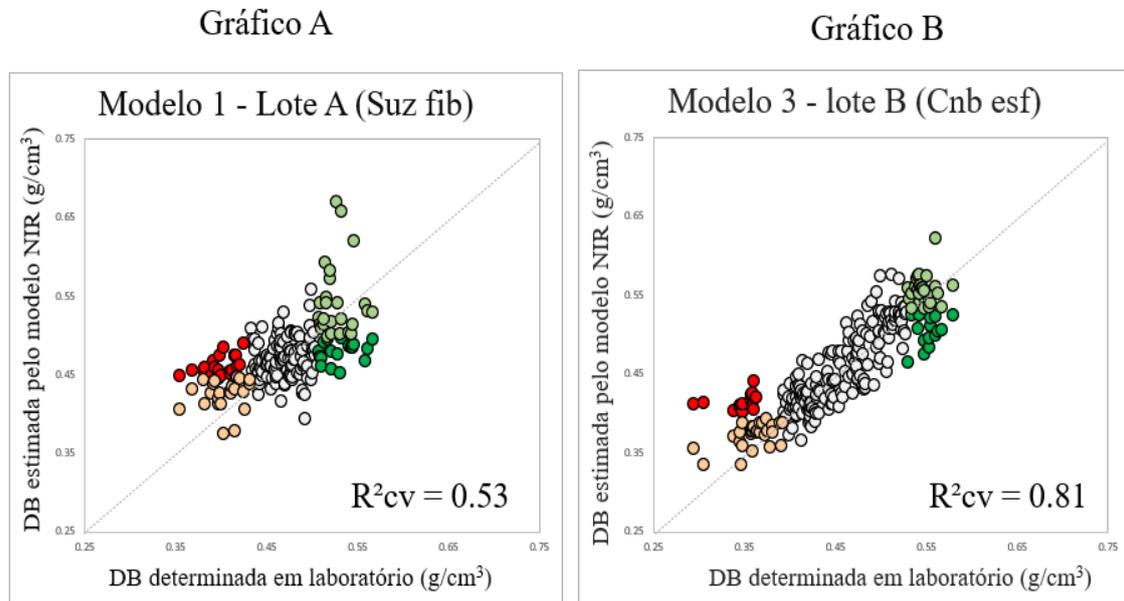
Nº de amostras alvo	Característica	Total de amostras	Amostras selecionadas (%)	Nº de acertos	Acertos (%)
15	menor densidade	216	6.94	7	46.67
30		216	13.89	15	50.00
45		216	20.83	24	53.33
15	maior densidade	216	6.94	11	73.33
30		216	13.89	17	56.67
45		216	20.83	27	60.00

Tabela 5 – Capacidade preditiva do modelo 3 do lote B (Cnb esf) para identificar as 15, 30 e 45 amostras mais e menos densas de 273 amostras totais.

Nº de amostras alvo	Característica	Total de amostras	Amostras selecionadas (%)	Nº de acertos	Acertos (%)
15	menor densidade	273	5.49	7	46.67
30		273	10.99	21	70.00
45		273	16.48	29	64.44
15	maior densidade	273	5.49	8	53.33
30		273	10.99	17	56.67
45		273	16.48	30	66.67

A figura 6 mostra a relação entre os valores medidos e preditos pelos modelos 1 (lote A) e 3 (lote B) da tabela 2. Os valores estimados pelos modelos do NIR em ambos os lotes estão bem próximos dos valores reais obtidos em laboratório, havendo, portanto, pouca dispersão dos dados. De modo geral, o modelo 3 do lote B apresentou menor dispersão e maior  $R^2_{cv}$  (0,81). No gráfico A a maior dispersão ocorreu nas amostras mais densas, enquanto no gráfico B houve maior dispersão das amostras mais leves.

Figura 6 – Relação entre os valores medidos (eixo x) e os valores estimados (eixo y) pelo modelo 1 e 3 da Tabela 2.



● - 45 amostras mais leves determinadas em laboratório; ○ - 45 amostras mais leves estimadas pelo modelo; ● - 45 amostras mais densas determinadas em laboratório; ● - 45 amostras mais densas estimadas pelo modelo; DB – densidade básica.

Fonte: Do autor (2021).

Ao avaliarem diferentes propriedades da madeira com o NIR, diversos autores constataram que na maioria das vezes a via de aquisição esfera de integração apresentou melhores calibrações de modelos em relação a sonda de fibra ótica. Arriel et al (2019) utilizaram espectros do NIR para estimar a densidade básica em diferentes tipos de amostras de madeira de *Eucalyptus*, com modelos ajustados pela PLS. Nesse estudo, o melhor modelo ( $R^2_{cv}$  0,77 e RPD 1,24) foi desenvolvido a partir de espectros de esfera de integração, que também apresentou gráficos com menor dispersão dos dados, indicando maior proximidade entre valor real e valor predito.

Santos (2017) usou o NIR para monitorar a dessecção de água em amostras de madeira de *Eucalyptus*. Observou que o melhor modelo ajustado pela PLS foi oriundo de espectros da esfera de integração ( $R^2_{cv}$  0,96) em face transversal e usinado por serra fita. Inferiu que essa via de aquisição é a mais indicada para realizar leituras no NIR, especificamente em madeiras com umidade entre 0% e 40%, possibilitando melhores condições para criação de melhores modelos para estimativa da umidade da madeira.

Na tabela 6 e 7 estão os acertos dos modelos no ranqueamento das amostras leves e pesadas das validações W, X, Y e Z.

Tabela 6 – Capacidade preditiva do modelo 1 do lote A (Suz fib) aplicado nos espectros do lote B obtidos por esfera de integração (validação W) e sonda de fibra ótica (validação X) para identificar as 15, 30 e 45 amostras mais e menos densas de 277 amostras totais do lote B.

Nº de amostras alvo	Característica	Total de amostras	Amostras selecionadas (%)	Nº de acertos	Acertos (%)
Validação W					
15	menor densidade	277	5.42	1	6.67
30		277	10.83	11	36.67
45		277	16.25	17	37.78
15	maior densidade	277	5.42	5	33.33
30		277	10.83	12	40.00
45		277	16.25	22	48.89
Validação X					
15	menor densidade	277	5.42	2	13.33
30		277	10.83	7	23.33
45		277	16.25	20	44.44
15	maior densidade	277	5.42	3	20.00
30		277	10.83	7	23.33
45		277	16.25	18	40.00

Fonte: Do autor (2021).

Tabela 7 – Capacidade preditiva do modelo 3 do lote B (Cnb esf) aplicado nos espectros do lote A (Suz fib) (validação Y) e modelo 6 do lote B (Cnb fib) nos espectros do lote A (Suz fib) (validação Z).

N° de amostras alvo	Característica	Total de amostras	Amostras selecionadas (%)	N° de acertos	Acertos (%)
Validação Y					
15	menor densidade	228	6.58	1	6.67
30		228	13.16	4	13.33
45		228	19.74	11	24.44
15	maior densidade	228	6.58	1	6.67
30		228	13.16	4	13.33
45		228	19.74	6	13.33
Validação Z					
15	menor densidade	228	6.58	6	40.00
30		228	13.16	12	40.00
45		228	19.74	21	46.67
15	maior densidade	228	6.58	5	33.33
30		228	13.16	12	40.00
45		228	19.74	17	37.78

Fonte: Do autor (2021).

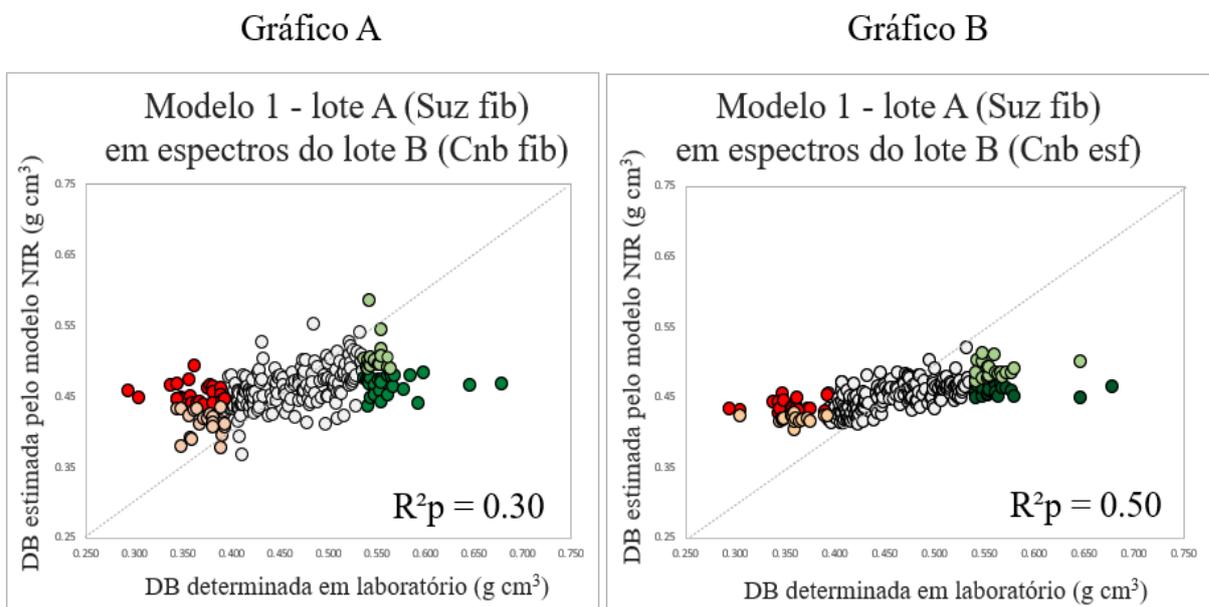
De modo geral, a validação que melhor ranqueou as amostras foi a W, com percentual de acerto de 37,78 % de amostras leves e 48,89 % de amostras pesadas. Essa validação foi a que também apresentou maior  $R^2_p$  (0,50). Em seguida, a validação X e Z, ou seja modelos e espectros a partir da sonda de fibra ótica, apresentaram resultados muito semelhantes, sendo 44,44 % de acerto para amostras leves e 40 % para amostras pesadas e 46,67 % em amostras leves e 37,78% em amostras pesadas, respectivamente. Na validação cruzada, o modelo 1 do

lote A (Suz fib) utilizado nas validações W e X, quando utilizado em sua própria base de dados, aumenta o percentual de acerto em 8,89 % em amostras leves e 20 % em amostras pesadas.

A validação que apresentou menor desempenho no ranqueamento das amostras em função da densidade foi a Y. É notório um baixo índice de acerto do modelo na base externa utilizada, sendo 24,44 % de acerto das amostras leves e apenas 13,33 % das amostras pesadas. Na validação cruzada, em sua base de dados original, o modelo 3, utilizado nessa validação, aumenta em 40 % o acerto em amostras leves e 53,34 % em amostras pesadas em relação aos acertos da validação independente.

Nas Figura 7 e 8 constam os gráficos com a relação dos valores medidos e preditos das validações independentes. A correlação inexistente entre o modelo 3 e a base de dados do lote A (validação Y) é evidenciada no gráfico E, onde é possível observar grande dispersão dos dados. A melhor correlação e menor dispersão é observado no gráfico D, mostrando significativa proximidade entre valores medidos e estimados.

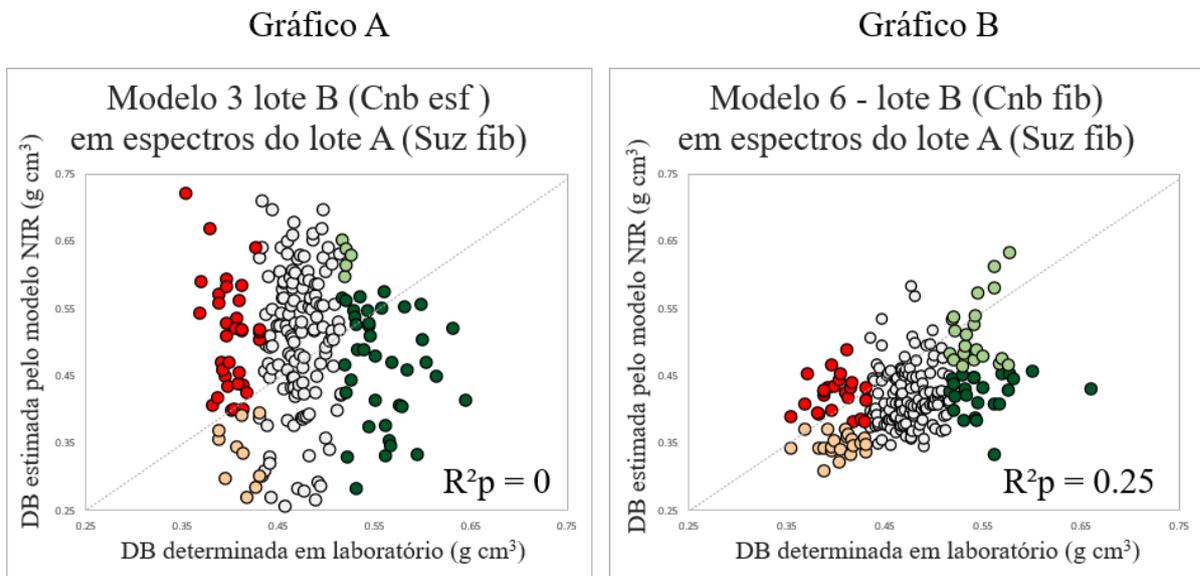
Figura 7 – Relação entre os valores medidos (eixo x) e os valores estimados (eixo y) das validações independentes W e X.



● - 45 amostras mais leves determinadas em laboratório; ○ - 45 amostras mais leves estimadas pelo modelo; ● - 45 amostras mais densas determinadas em laboratório; ○ - 45 amostras mais densas estimadas pelo modelo; DB – densidade básica.

Fonte: Do autor (2021).

Figura 8 – Relação entre os valores medidos (eixo x) e os valores estimados (eixo y) das validações independentes Y e Z.



● - 45 amostras mais leves determinadas em laboratório; ○ - 45 amostras mais leves estimadas pelo modelo; ● - 45 amostras mais densas determinadas em laboratório; ● - 45 amostras mais densas estimadas pelo modelo; DB – densidade básica.

Fonte: Do autor (2021).

## 6 CONCLUSÃO

Os modelos desenvolvidos para estimar a densidade básica com base nos espectros de NIR para as duas base de dados apresentaram estatísticas variadas, com  $R^2cv$  variando de 0.53 a 0.81 e raiz quadrática média do erro variando de 0.028 a 0,037 g/cm<sup>3</sup>.

O modelo desenvolvido com espectros medidos usando esfera de integração nas madeiras da Cenibra (Modelo 3) apresentaram estimativas com melhor correlação com o valor determinado em laboratório ( $R^2cv = 0.81$ ). Porém, quando aplicado no outro lote, a relação entre valores observados e preditos foi nula ( $R^2p = 0$ ).

Por outro lado, o Modelo desenvolvido com espectros medidos por fibra ótica nas madeiras da Suzano (modelo 1) apresentaram estimativas com correlação mediana com o valor determinado em laboratório ( $R^2cv = 0.53$ ). No entanto, este modelo gera estimativas com  $R^2p$  de 0.50 com valores observados.

O Modelo 3 apresenta a melhor performance preditiva observada, quando aplicado em madeiras semelhantes àquelas utilizadas para construção do modelo, mas teve desempenho

ruim quando aplicado em lotes de madeiras desconhecidas. O Modelo 3 foi capaz de identificar 7 das 15 amostras de madeiras mais leves e 8 das 15 amostras de madeiras mais densas.

Por outro lado, o Modelo 1 apresentou performance preditiva mais baixa, porém as estimativas geradas em madeiras desconhecidas foram mais correlacionadas aos valores conhecidos. O Modelo 1 foi capaz de identificar 7 das 15 amostras de madeiras mais leves e 11 das 15 amostras de madeiras mais densas.

## 7 REFERÊNCIAS

ALVES, R. C.; OLIVEIRA, A. L. C.; CARRASCO, E. V. M. Propriedades físicas da madeira de *Eucalyptus cloeziana* f. Muell. **Floresta e Ambiente**, v. 24, 2017.

ALZATE S. B. A.; TOMAZELLO, M.; PIEDADE, S. M. S. Variação longitudinal da densidade básica da madeira de clones de *Eucalyptus grandis* Hill ex Maiden, E. saligna Sm. e E. grandis x urophylla. **Scientia forestalis**, n. 68, p.87-95, 2005.

ANDRADE, G. H. **Estudo da espectroscopia na região do infravermelho médio e próximo para previsão das propriedades do petróleo e emulsão de petróleo do tipo água em óleo**. 2009. 144 f. Dissertação (Mestrado em Ciências) – Escola de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.

ANDRADE, C. R. Calibrações por meio da técnica da espectroscopia no infravermelho próximo para teor de extrativos totais, de lignina total e holocelulose usando espectros obtidos na madeira sólida. **Ambiência Guarapuava**, v.7 n.1 p. 39 – 49, 2011

ARRIEL, T. G. et al. Developing near infrared spectroscopic models for predicting density of *eucalyptus* wood based on indirect measurement. **Cerne**, v. 25, n. 3, p.294-300, 2019.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **Normas técnicas NBR 11941**. Brasília, DF, 2003. 6p.

BAKKE, H. A.; LEITE, A. S. M.; SILVA, L. B. Estatística multivariada: aplicação da análise fatorial na engenharia de produção. **Revista Gestão Industrial**, v. 04, n. 04, p. 01 – 14, 2008.

BALDIN, T. et al. Qualidade da madeira de *Eucalyptus benthamii* para produção de celulose por espectroscopia no infravermelho próximo (NIRS). **Scientia Forestalis**, v. 48, n. 126, 2020.

BARTHUS, R. C. **Aplicação de métodos quimiométricos para análises de controle de qualidade de óleos vegetais utilizando espectroscopias no infravermelho e Raman**. 1999. 112 p. Dissertação. (Mestrado em Química) – Instituto de Química, Universidade Estadual de Campinas, Campinas, 1999.

BEE, K. R.; KOWALSKI, B. R. An introduction to multivariate and calibration analysis. **Analytical Chemistry**, v. 59, n. 17, 1987.

BEEBE, K. R.; KOWALSKI, B. R. An introduction to multivariate calibration and Analysis. **Analytical chemistry**, v. 59, n. 17, 1987.

BORIN, A. **Aplicação de quimiometria e espectroscopia no infravermelho próximo no controle da qualidade de lubrificantes**. 2003. 104 p. Dissertação. (Mestrado em Química) – Instituto de Química, Universidade Estadual de Campinas, Campinas, 2003.

BRAGA, J. W. B.; POPPI, R. J. Validação de modelos de calibração multivariada: uma aplicação na determinação de pureza polimórfica de carbamazepina por espectroscopia no infravermelho próximo. **Química Nova**, v. 27, n. 6, 2004.

BURGER, L. M.; RICHTER, H. G. **Anatomia da madeira**. São Paulo: Nobel. 1991. 150 p.

BURNS, D. A.; CIURCZAK, E. W. **Handbook of Near-Infrared Analysis**. 3. ed. Boca Raton: CRC Press, 2008. 150 p.

CASTRO, A. F. N. M. et al. Análise multivariada para seleção de clones de eucalipto destinados à produção de carvão vegetal. **Pesquisa agropecuária brasileira**, Brasília, v.48, n.6, p.627-635, 2013

COSTA, E. V. S. **Validação de modelos baseados em Espectroscopia no Infravermelho Próximo para estimativa da densidade básica da madeira**. 2016. 78 f. Dissertação. (Mestrado em Ciência e Tecnologia da Madeira) – Faculdade de Ciências Agrárias, Universidade Federal de Lavras, Lavras, 2016.

COSTA, E.V.S. et al. Influence of spectral acquisition technique and wood anisotropy on the statistics of predictive near infrared-based models for wood density. **Journal of Near Infrared Spectroscopy**, v. 26, n. 2, p. 106-116, 2018b

COSTA, L. R.; TRUGILHO, P. F.; HEIN, P.R.G. Evaluation and classification of Eucalypt charcoal quality by near infrared spectroscopy. **Biomass and Bioenergy**, v.112, p. 85-92, 2019

COSTA, L. R. **Aplicação da Espectroscopia no Infravermelho Próximo para avaliação do carvão vegetal**. 2017. 62 f. Dissertação (Mestrado em Ciência e Tecnologia da Madeira) – Faculdade de Ciências Agrárias, Universidade Federal de Lavras, Lavras, 2017.

COUTO, A. M. et al. Amostras longitudinais alternativas para a determinação da densidade básica em clones de *Eucalyptus* sp. **Floresta e Ambiente**, v. 19, n. 12, p. 184 – 193, 2012.

EMBRAPA. **Ministério da Agricultura, Pecuária e Abastecimento. Eucalipto no Semiárido brasileiro**. Petrolina, 2016. 42 p. (EMBRAPA/PE. Documentos online).

ESTOPA, R. A. et al. Caracterização química da madeira de *Eucalyptus benthamii* por meio de espectroscopia NIR. **O papel**, vol. 78, n. 2, p. 75 – 81 fev., 2017.

FERREIRA, M. M. C. **Quimiometria: conceitos, métodos e aplicações**. Campinas: Editora da Unicamp, 2015.

FOELKEL, C. E. B.; BRASIL, M. A. P.; BARRICHELO, L. E. G. Métodos para determinação da densidade básica de cavacos para coníferas e folhosas. **IPEF**, n. 2, v. 3, p. 65-74, 1971.

GEMPERLINE, P. **Practical Guide to Chemometrics**. 2. ed. Boca Raton: Taylor and Francis, 2006. 520 p.

GOUVÊA, A. F. G. et al. Determinação da densidade básica da madeira de *Eucalyptus* por diferentes métodos não destrutivos. **Revista Árvore**, v.35, n.2, p.349-358, 2011.

HEIN, P. R. G.; LIMA, J. T.; CHAIX, G. Otimização de calibrações baseadas em espectroscopia no infravermelho próximo para estimativa de propriedades da madeira de *Eucalyptus*. **Floresta**, Curitiba, v. 40, n. 3, p. 615-624, 2010.

costa

HEIN, P. R. G. LIMA, J. T.; CHAIX, G. Robustness of models based on near infrared spectra to predict the basic density in *Eucalyptus urophylla* wood. **Journal of near infrared spectroscopy**, v. 17, n. 3, p. 141-150, 2009

HEIN, P. R. G. et al. Near infrared spectroscopy for estimating wood basic density in *Eucalyptus urophylla* and *Eucalyptus grandis* **Cerne**, Lavras, v. 15, n. 2, p. 133-141, 2009

HEIN, P. R. G. et al. Spatial variation of wood density, stiffness and microfibril angle along *Eucalyptus* trunks grown under contrasting growth conditions. **Trees**, v. 30 p. 871–882, 2016. DOI 10.1007/s00468-015-1327-8.

HEISE, H. M. Donald A. Burns, Emil W. Ciurczak (Eds.): Handbook of near-infrared analysis, 3rd ed. **Analytical and Bioanalytical Chemistry**, v. 393, p. 1387–1389, 2009.

HONGYU, K.; SANDANIELO, V. L. M.; OLIVEIRA JUNIOR, G. J. Análise de Componentes Principais: resumo teórico, aplicação e interpretação. **Engineering and Science**, v. 5, n. 1, 2015.

INDÚSTRIA BRASILEIRA DE ÁRVORES (IBÁ). Relatório IBÁ 2020. Disponível em > <https://iba.org/datafiles/publicacoes/relatorios/relatorio-iba-2020.pdf>< Acesso em 02 nov. 2020.

LATORRACA, J. V. F.; ALBUQUERQUE, C. E. C. Efeito do rápido crescimento sobre as propriedades da madeira. **Floresta e Ambiente**, v. 7, n. 1, p. 279 - 291, 2000.

LAZZAROTTO, M. et al. Espectroscopia no infravermelho próximo para estimativa da densidade básica de madeiras de *Pinus*. **Ciência da madeira**, v. 7, n. 3, p. 119-126, 2016.

LI, Y. VIA, B. K.; LI, Y. Lifting wavelet transform for Via-NIR spectral data optimization to predict wood density, **Elsevier**, 2020

LI, B.; MARTIN, E.; MORRIS, J. Latent variable selection in partial least squares modelling. **Dynamics and control of process systems**, 2001.

LIMA, A.; BAKKER, J. Espectroscopia no Infravermelho Próximo para a monitorização da perfusão tecidual. **Revista Brasileira de Terapia Intensiva**. n. 23, v.3, p. 341-351, 2011.

LOBÃO, M. S. et al. Agrupamento de espécies florestais por análises univariadas e multivariadas das características anatômica, física e química das suas madeiras. **Scientia Forestalis**. Piracicaba, v. 39, n. 92, p. 469-477, 2011

MAGALHÃES, W. L. E. et al. **Uso da espectroscopia no infravermelho próximo (NIR) para predição não-destrutiva de densidade básica da madeira de *Pinus taeda***. Colombo: Embrapa, 2006. 4 p. (Embrapa. Comunicado Técnico, 159).

MAGALHÃES, D. F. **Aplicações da espectroscopia de infravermelho próximo na monitorização de processos farmacêuticos**. 2014. 127 f. Dissertação. (Mestrado em Química Tecnológica) – Faculdade de Ciências, Universidade de Lisboa, Lisboa, 2014.

MANCINI, T.; RAMALHO, F. M. G.; TRUGILHO, P. F.; HEIN, P. R. G. Estimation of total extractive content of wood from planted and native forests by near infrared spectroscopy. **Forest-Biogeosciences and Forestry**, v. 14, n. 1, 2021

MILAGRES, F. R. et al. Influência da idade na geração de modelos de espectroscopia NIR, para predição de propriedades da madeira de *eucalyptus* spp. **Revista Árvore**, Viçosa, v.37, n.6, p.1165-1173, 2013.

MORGANO, M. A. et al. Determinação de açúcar total em café cru por espectroscopia no infravermelho próximo e regressão por mínimos quadrados parciais. **Química Nova**, v. 30, n. 2, p. 346-350, 2007.

MUÑIZ, G. I. B. et al. Fundamentos e estado da arte da espectroscopia no infravermelho próximo no setor de base florestal. **Ciência Florestal**, Santa Maria, v. 22, n. 4, p. 865-875, 2012.

NETO, J. M. M.; MOITA, G. C. Uma introdução à análise exploratória de dados multivariados. **Química Nova**, v. 21, n. 4, p. 467-469, 1997.

NETO, R. A. et al. Estimating hardness and density of wood and charcoal by near-infrared spectroscopy. **Wood Science and Technology**, n. 55, p. 215–230, 2020.

NISGOSKI, S. **Espectroscopia no infravermelho próximo no estudo de características da madeira e papel de *Pinus taeda* L.** 2005. 174 f. Tese. (Doutorado em Engenharia Florestal) – Faculdade de Ciências Agrárias, Universidade Federal do Paraná, Curitiba, 2005.

OLIVEIRA, J. T. S.; HELLMEISTER, J. C. TOMAZELLO, M. Variação do teor de umidade na madeira de sete espécies de Eucalipto. **Revista Árvore**, Viçosa, v. 29, n. 1, p. 115-127, 2005.

PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. **Analytica Chimica Acta**, v. 1026, p.1-29, 2018.

PASQUINI, C. Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. **Journal of the Brazilian Chemical Society**, v. 14, n. 2, p. 198-219, 2003.

PEDRO, A. M. K. **Determinação simultânea e não-destrutiva de sólidos totais e solúveis, licopeno e beta-caroteno em produtos de tomate por espectroscopia no infravermelho próximo utilizando calibração multivariada**. 2004. 118 f. Dissertação. (Mestrado em Química) – Instituto de Química, Universidade Estadual de Campinas, Campinas, 2004

RAMALHO, F. M. G.; Andrade, J. M.; HEIN, P. R. G. Rapid discrimination of wood species from native forest and plantations using near infrared spectroscopy. **Forest Systems**, v. 27, n. 2, 2018.

RIBEIRO, L. F. **Uso de espectroscopia no infravermelho e análise multivariada para previsão de ácidos graxos em linhaça dourada e marrom**. 2012. 119 f. Dissertação. (Mestrado em Engenharia de Alimentos) – Setor de Tecnologia, Universidade Federal do Paraná, Curitiba, 2012.

RIBEIRO, F. A.; ZANI FILHO, J. Variação da densidade básica da madeira em espécies/procedências de *Eucalyptus* spp. **IPEF**, n.46, p.76-85, jan./dez.1993.

SALAZAR, M. M. **Caracterização do transcriptoma e parede celular de três espécies de *Eucalyptus* com importância industrial**. 2012. 125 f. Tese. (Doutorado em Genética e Biologia Molecular), Instituto de Biologia, Universidade Estadual de Campinas, Campinas, 2012.

SANTOS, R. B.; GOMIDE, J. L.; SOUSA, L. C. Predição de qualidade da madeira e da polpa celulósica por técnica de espectroscopia de infravermelho próximo (NIRS). **Revista Árvore**, Viçosa-MG, v.33, n.4, p.759-767, 2009

SANTOS, L. M. **Monitoramento da dessorção de água na madeira por espectroscopia no infravermelho próximo**. 2017. 57 f. Dissertação. (Mestrado em Ciência e Tecnologia da Madeira) – Departamento de Ciências Florestais, Universidade Federal de Lavras, Lavras, 2017.

SEVERIDO JUNIOR, G. **Aplicação de métodos quimiométricos em análises de formas farmacêuticas sólidas lamivudina (3TC) + zidovudina (AZT) baseados na espectroscopia de absorção na região do ultravioleta e do infravermelho próximo**. 2011. 134 f. Tese. (Doutorado em Ciências Farmacêuticas) – Centro de Ciências da Saúde, Universidade Federal de Pernambuco, Recife, 2011.

SILVEIRA, F. G. **Abordagem geométrica do método dos quadrados mínimos parciais com uma aplicação a dados de seleção genômica**. 2014. 178 f. Tese. (Doutorado em Estatística e Experimentação Agropecuária), Universidade Federal de Lavras, Lavras, 2014.

SILVA, D. A. et al. Avaliação das Propriedades Energéticas de Resíduos de Madeiras Tropicais com Uso da Espectroscopia NIR. **Floresta e Ambiente**, v. 21, n. 4, p. 561-568, 2014.

SCHWANNINGER, M.; RODRIGUES, J. C.; FACKLER, K. A review of band assignments in near infrared spectra of wood and wood components. **Jornal of Near Infrared Spectroscopy**, v.19, p. 287–308, 2011.

SOARES, L. F.; et al. Avaliação de espectrômetro NIR portátil e PLS-DA para a discriminação de seis espécies similares de madeiras amazônicas. **Química Nova**, v. 40, n. 4, p. 418-426, 2017.

SOTELO, F. F. **Aplicação de espectroscopia de infravermelho próximo na caracterização de petróleo. Simulação de uma unidade de destilação atmosférica**. 2006. 154 f. Tese.

(Doutorado em Engenharia Química), Escola politécnica, Universidade de São Paulo, São Paulo, 2006.

SOUZA, D. M.; MADARI, B. E.; GUIMARÃES, F.F. Aplicação de técnicas multivariadas e inteligência artificial na análise de espectros de infravermelho para determinação de matéria orgânica em amostras de solos. **Química Nova**, v. 35, n. 9, p. 1738-1745, 2012.

TRIANOSKI, R. et al. Variação longitudinal da densidade básica da madeira de espécies de *Pinus* tropical. **Floresta**, Curitiba, v. 43, n. 3, p. 503 - 510, jul./set. 2013.

VIDAURRE, G. et al. Lenho juvenil e adulto e as propriedades da madeira. **Floresta e Ambiente**, Rio de Janeiro, v. 18, n. 4, p. 469-480, 2011.

VINZI, V. E. et al. **Handbook of Partial Least Squares**. Ed. Springer, 2010. 791 p.

WHEELER, O. H. Near Infrared Spectra of organic compounds. **Chemical Reviews**, v. 59, n. 4, p. 629-666, 1959.

ZANUNCIO, A. J. V. et al. Determination of heat-treated eucalyptus and *Pinus* wood properties using NIR spectroscopy. **Journal of Tropical Forest Science**, p. 117-125, 2018

ZENI, D. **Determinação de cloridrato de propanol em medicamentos por espectroscopia no infravermelho com calibração multivariada**. 2005. 64 f. Dissertação (Mestrado em Química) – Departamento de Química, Universidade Federal de Santa Maria, Santa Maria, 2005.