



**UnB**

Instituto de Química  
Programa de Pós-Graduação em Química (PPGQ)

DISSERTAÇÃO DE MESTRADO

**CORREÇÃO DO EFEITO DA UMIDADE EM ESPECTROS NIR  
EMPREGANDO EPO PARA O DESENVOLVIMENTO DE MODELOS PLS-DA  
PARA DISCRIMINAÇÃO DE MADEIRAS**

**Aluna:** Daniele Cristina Gomes da Cunha Kunze

**Orientador:** Prof. Dr. Jez Willian Batista Braga

**Coorientadora:** Dr<sup>a</sup>. Tereza Cristina Monteiro Pastore

**Brasília, DF**

**2020**



**UnB**

Instituto de Química  
Programa de Pós-Graduação em Química (PPGQ)

DISSERTAÇÃO DE MESTRADO

**CORREÇÃO DO EFEITO DA UMIDADE EM ESPECTROS NIR  
EMPREGANDO EPO PARA O DESENVOLVIMENTO DE MODELOS PLS-DA  
PARA DISCRIMINAÇÃO DE MADEIRAS**

Dissertação apresentada ao Instituto de Química da Universidade de Brasília como parte dos requisitos exigidos para a obtenção do Título de Mestre em Química.

**Aluna:** Daniele Cristina Gomes da Cunha Kunze

**Orientador:** Prof<sup>o</sup>. Dr. Jez Willian Batista Braga

**Coorientadora:** Dr<sup>a</sup>. Tereza Cristina Monteiro Pastore

**Brasília, DF**

**2020**

*Dedico este trabalho à minha mãe,  
Teresa Cunha.*

## AGRADECIMENTOS

Ao meu orientador, Prof. Dr. Jez Braga, por toda sua atenção, conhecimento, confiança, incentivo e ensinamentos. Tenho comigo uma imensa gratidão e admiração pela pessoa e pelo cientista que o compõem. É um exemplo de orientador, encontrando as melhores maneiras de alcançar o melhor do aluno.

À Dr.<sup>a</sup> Tereza Pastore, minha coorientadora, por todo auxílio, sempre buscando os melhores recursos para alcançarmos nossos objetivos. Muito obrigada por todas as conversas e por me encantar pela química ambiental.

Ao Grupo de Automação, Quimiometria e Química Ambiental (AQQUA) do IQ/UnB, professores, alunos e técnica, pelo ambiente de trabalho amigável e caloroso.

À equipe de trabalho: Hugo Rocha, Priscila Veras e Rafaela Dornelas, pela contribuição na realização e conclusão deste trabalho.

À UnB e ao Programa de Pós-graduação em Química, pelo apoio e oportunidade.

Ao Serviço Florestal Brasileiro (SFB) e Laboratório de Produtos Florestais (LPF), pela infraestrutura e colaboração que permitiram a execução da pesquisa.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pela bolsa de mestrado concedida (processo 88882.383351/2019-01).

Aos órgãos de fomento: FAPDF (processo 193.001.643/2017), CAPES (projeto NEQUIFOR, processo AUXPE no 3509/2014), CITES e União Europeia (processo S1-32QTL-000018), pelo apoio financeiro.

Aos amigos que conheci graças ao mestrado: Hugo Rocha, Guilherme Borges, Hadassa Ramos e Artur Amorim, pelo acolhimento e auxílio durante meu processo de mudança de cidade, pelas conversas e pelas brincadeiras, vocês tornaram a jornada mais leve.

Aos amigos que, mesmo de longe, estavam ao meu lado enfrentando os problemas, celebrando as vitórias e rindo da vida. Em especial a Pedro Torres e Sarah Cristina, vocês fazem muita falta.

À minha família, em especial às minhas tias Cílinha e Socorro, aos meus tios Walter e José Henrique, às minhas primas/irmãs Luciana, Taciana e Marina, e aos meus sobrinhos Fernando e Miguel, o apoio de vocês foi fundamental para mim.

Ao companheiro que a vida me deu, Marlos Santana, por aquele sorvete de banana que mudou a minha vida. Obrigada por estar presente nos bons e maus momentos, pelo incentivo e por nunca duvidar de mim (mesmo quando eu mesma duvidava).

À minha mãe, Teresa Cunha, mulher mais forte que conheço. Minha melhor amiga, maior incentivadora e razão pela qual cheguei até aqui. Palavras não descrevem todo o amor que sinto por ti.

A todos que contribuíram de alguma forma para este trabalho, agradeço.

## SUMÁRIO

<b>1. INTRODUÇÃO E OBJETIVOS</b> .....	1
<b>2. FUNDAMENTAÇÃO TEÓRICA E REVISÃO BIBLIOGRÁFICA</b> .....	3
<b>2.1. Madeira</b> .....	3
<b>2.1.1. Estrutura Anatômica</b> .....	4
<b>2.1.2. Composição Química</b> .....	4
<b>2.1.3. Umidade</b> .....	7
<b>2.1.4. Métodos de Identificação de Madeira</b> .....	8
<b>2.2. Espectroscopia NIR</b> .....	11
<b>2.3. Quimiometria</b> .....	13
<b>2.3.1. Ortogonalização dos Parâmetros Externos – EPO</b> .....	13
<b>2.3.2. Mínimos Quadrados Parciais para Análise Discriminante – PLS-DA</b> .....	14
<b>2.3.3. Figuras de Mérito</b> .....	18
<b>3. MATERIAIS E MÉTODOS</b> .....	20
<b>3.1. Parte I: Avaliação da eficiência do EPO e criação de modelos de predição</b> .....	20
<b>3.1.1. Seleção e preparo de amostras</b> .....	20
<b>3.1.2. Obtenção do teor de umidade</b> .....	21
<b>3.1.3. Medidas NIR</b> .....	22
<b>3.1.4. Condicionamento</b> .....	22
<b>3.1.5. Análise multivariada</b> .....	23
<b>3.1.5.1. Organização das matrizes de dados e seleção de amostras</b> .....	23
<b>3.1.5.2. Correção da umidade utilizando EPO</b> .....	24
<b>3.1.5.3. Criação dos modelos PLS-DA flexíveis</b> .....	25
<b>3.2. Parte II: Avaliação dos modelos de predição criados</b> .....	28

3.2.1.	Seleção de amostras .....	29
3.2.2.	Tratamento umidificador.....	29
3.2.3.	Teor de Umidade, Medidas NIR e Processo de Secagem.....	30
3.2.4.	Análise Multivariada.....	30
4.	<b>RESULTADOS E DISCUSSÃO.....</b>	<b>33</b>
4.1.	<b>Parte I: Avaliação da eficiência do EPO e criação de modelos de predição.....</b>	<b>33</b>
4.1.1.	Avaliação preliminar via PCA da correção do efeito da umidade por EPO em amostras de madeira .....	33
4.1.2.	Determinação dos parâmetros utilizados no EPO e no PLS-DA para o desenvolvimento do método de discriminação .....	37
4.1.3.	Escolha do pré-processamento aplicado aos espectros .....	39
4.1.4.	Resultados para o modelo de discriminação por PLS-DA utilizando conjunto de treinamento com amostras em umidade ambiente e validado com amostras que passaram pela estufa (AE) .....	43
4.2.	<b>Parte II: Avaliação dos modelos de predição criados.....</b>	<b>50</b>
4.2.1.	Resultados do tratamento das amostras nas caixas umidificadoras..	50
4.2.2.	Resultados das predições para a discriminação das amostras com maior teor de umidade .....	52
5.	<b>CONCLUSÕES.....</b>	<b>57</b>
6.	<b>BIBLIOGRAFIA .....</b>	<b>59</b>
7.	<b>APÊNDICES.....</b>	<b>68</b>
7.1.	<b>APÊNDICE A - Gráfico de curvas de nível de valores de RMSEC para modelos corrigidos por EPO da matriz de treinamento na condição (a) ambiente e (b) estufa, predizendo os dados da matriz de treinamento robusta na condição de pré-processamento 1 (EPO: bruto; PLS-DA: <i>Mean Center</i>). .....</b>	<b>68</b>
7.2.	<b>APÊNDICE B - Gráfico de curvas de nível de valores de RMSEC para modelos corrigidos por EPO da matriz de treinamento na condição (a) ambiente</b>	

e (b) estufa, predizendo os dados da matriz de treinamento robusta na condição de pré-processamento 2 (EPO: SNV; PLS-DA: SNV + <i>Mean Center</i> ).....	68
<b>7.3. APÊNDICE C - Gráfico de curvas de nível de valores de RMSEC para modelos corrigidos por EPO da matriz de treinamento na condição (a) ambiente e (b) estufa, predizendo os dados da matriz de treinamento robusta na condição de pré-processamento 4 (EPO: 1ª derivada; PLS-DA: 1ª derivada + <i>Mean Center</i>).....</b>	<b>69</b>

## LISTA DE ABREVIATURAS E ACRÔNIMOS

amb	Amostras que estavam em umidade ambiente
AN	Andiroba
AQUA	Laboratório de Automação, Quimiometria e Química Ambiental
ASTM	Sociedade Americana para Testes e Materiais, do inglês <i>American Society for Testing and Materials</i>
<i>c</i>	Número de componentes do EPO
CE	Cedro
CH	Cedrinho
CITES	Convenção sobre o Comércio Internacional das Espécies da Fauna e da Flora Silvestres Ameaçadas de Extinção, do inglês <i>Convention on International Trade in Endangered Species of Wild Fauna and Flora</i>
CVCE	Erro de classificação empregando validação cruzada, do inglês <i>Cross Validation Classification Error</i>
CX	Curupixá
DART-MS	Análise Direta em Tempo Real por Espectrometria de Massas, do inglês <i>Direct Analysis in Real Time Mass Spectrometry</i>
EPO	Ortogonalização dos parâmetros externos, do inglês <i>External Parameter Orthogonalisation</i>
esp	Espécie estudada
est	Amostras que passaram pelo processo de secagem na estufa
FIR	Infravermelho distante, do inglês <i>far-infrared</i>

FN	Número absoluto de falsos negativos
FP	Número absoluto de falsos positivos
FT-NIR	Transformada de Fourier no infravermelho próximo
GTTN	Rede Global de Rastreamento de Madeira, do inglês <i>Global Timber Tracking Network</i>
IAWA	Associação Internacional de Anatomistas de Madeira, do inglês <i>International Association of Wood Anatomists</i>
incon	Número de amostras com resultado de análise inconclusivo
LPF	Laboratório de Produtos Florestais
M	Mogno
MIR	Infravermelho médio, do inglês <i>mid-infrared</i>
NIR	Infravermelho próximo, do inglês <i>near-infrared</i>
NIRS	Espectroscopia no infravermelho próximo, do inglês <i>Near Infrared Spectroscopy</i>
nm	Nanômetro
mm	Milímetro
OSC	Correção de sinal ortogonal, do inglês <i>Orthogonal Signal Correction</i>
PCA	Análise das componentes principais, do inglês, <i>Principal Component Analysis</i>
PLS-DA	Análise discriminante pelo método de mínimos quadrados parciais, do inglês <i>Partial Least Squares for Discriminant Analysis</i>
hard-PLS-DA	Análise PLS-DA rígida

soft-PLS-DA	Análise PLS-DA flexível
RMSECV	Raiz quadrada do erro médio quadrático de validação cruzada, do inglês <i>Root Mean Square Error of Cross Validation</i>
RMSEC	Raiz Quadrada do Erro Quadrático Médio de Calibração, do inglês <i>Root Mean Square Error of Calibration</i>
rob	Robusto, união das condições amb e est
SFB	Serviço Florestal Brasileiro
SNV	Padronização normal de sinal, do inglês <i>Standard Normal Variate</i>
SVD	Decomposição em valores singulares, do inglês <i>Singular Value Decomposition</i>
TEF	Taxa de eficiência
TFN	Taxa de falsos negativos
TFP	Taxa de falsos positivos
TINDEF	Taxa de amostras com resultado indefinido
tr	Amostras de treinamento
Unb	Universidade de Brasília
UNODC	Escritório das Nações Unidas para Drogas e Crime, do inglês <i>United Nations Office on Drugs and Crime</i>
val	Amostras de validação
VL	Variáveis latentes
VN	Número absoluto de verdadeiros negativos
VP	Número absoluto de verdadeiros positivos

## LISTA DE FIGURAS

<b>Figura 1)</b> Estrutura macroscópica do tronco de uma árvore: casca, floema, câmbio, albarno, cerne e medula <sup>7</sup> (adaptado). .....	4
<b>Figura 2)</b> Estrutura da unidade repetidora da celulose .....	5
<b>Figura 3)</b> Parte do espectro eletromagnético com destaque a região do visível ao infravermelho (adaptado <sup>50</sup> ). (UV): ultravioleta; (NIR): infravermelho próximo; (MIR): infravermelho médio; (FIR): infravermelho distante. ....	11
<b>Figura 4)</b> Aparelho medidor de umidade para madeiras DL 2000, Digisystem™. (Fonte: Digisystem <sup>78</sup> ).....	21
<b>Figura 5)</b> Espectrofotômetro microPHAZIR™ RX Analyzer, da ThermoFisher Scientific™. (Fonte: ThermoFisher Scientific <sup>79</sup> ) .....	22
<b>Figura 6)</b> Amostras de madeira na estufa para processo de secagem.....	22
<b>Figura 7)</b> Processo de separação de matrizes por espécie. ( <i>esp</i> ) cada uma das espécies estudadas; ( <i>amb</i> ) amostras que estavam em umidade ambiente; ( <i>est</i> ) amostras que passaram pelo processo de secagem na estufa; ( <i>tr</i> ) amostras de treinamento; ( <i>val</i> ) amostras de validação.....	23
<b>Figura 8)</b> Esquema para criação do modelo de treinamento, considerando a condição em que apenas o pré-processamento <i>Mean Center</i> foi utilizado.....	25
<b>Figura 9)</b> Esquema para a análise dos conjuntos de validação no modelo de predição otimizado. ....	26
<b>Figura 10)</b> Fluxograma para identificação da classe de cada replicata. ....	27
<b>Figura 11)</b> Fluxograma para identificação da classe por amostra. ....	28
<b>Figura 12) (a) e (b)</b> Amostras dispostas nas caixas umidificadoras com teor de umidade interno de 97,5%.....	29
<b>Figura 13)</b> Processo de separação de matrizes e pré-processamento por espécie. ( <i>esp</i> ) cada uma das espécies estudadas; ( <i>amb</i> ) amostras que estavam em umidade ambiente; ( <i>est</i> ) amostras que passaram pelo processo de secagem na estufa; ( <i>val</i> ) amostras de validação.....	30

<b>Figura 14)</b> Espectros brutos referentes as amostras de curupixá ( <i>Micropholis melinoniana</i> Pierre) antes (em azul) e após secagem (em vermelho). .....	34
<b>Figura 15)</b> Espectros pré-processados referentes as amostras de curupixá ( <i>Micropholis melinoniana</i> Pierre) antes (em azul) e após secagem (em vermelho). .....	35
<b>Figura 16)</b> Gráficos dos (a) escores e (b) pesos com dados centrados na média para os espectros de curupixá pré-processados antes (♦) e após (■) secagem. ....	35
<b>Figura 17)</b> Espectros pré-processados e corrigidos com EPO referentes as amostras de curupixá antes (em azul) e após secagem (em vermelho). .....	36
<b>Figura 18)</b> Gráficos dos (a) escores e (b) pesos com dados centrados na média para os espectros de curupixá pré-processados e corrigidos com EPO antes (♦) e após (■) secagem. ....	36
<b>Figura 19)</b> Gráfico de curvas de nível de valores de RMSEC para modelos corrigidos por EPO da matriz de treinamento na condição (a) ambiente e (b) estufa, predizendo os dados da matriz de treinamento robusta na condição de pré-processamento 3 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + <i>Mean Center</i> ). ....	37
<b>Figura 20)</b> Gráficos T <sup>2</sup> versus Q referentes às amostras de validação sem correção e com correção do efeito da umidade por EPO na condição de pré-processamento 3 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + <i>Mean Center</i> ) no modelo AE. (♦) andiroba; (■) cedro; (▲) cedrinho; (▼) curupixá; (★) mogno. ....	43
<b>Figura 21)</b> Valores de classe estimados para o conjunto de validação para discriminação da espécie andiroba em relação às demais espécies antes e após correção com EPO, respectivamente, na condição de pré-processamento 3 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + <i>Mean Center</i> ) e no modelo AE. (♦) andiroba; (■) cedro; (▲) cedrinho; (▼) curupixá; (★) mogno; (---) Limiar de discriminação. ....	44
<b>Figura 22)</b> Valores de classe estimados para o conjunto de validação para discriminação da espécie cedro em relação às demais espécies antes e após correção com EPO, respectivamente, na condição de pré-processamento 3 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + <i>Mean Center</i> ) e no modelo AE. (♦) andiroba; (■) cedro; (▲) cedrinho; (▼) curupixá; (★) mogno; (---) Limiar de discriminação. ....	46
<b>Figura 23)</b> Valores de classe estimados para o conjunto de validação para discriminação da espécie cedrinho em relação às demais espécies antes e após correção com EPO,	

respectivamente, na condição de pré-processamento 3 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + *Mean Center*) e no modelo AE. (♦) andiroba; (■) cedro; (▲) cedrinho; (▼) curupixá; (★) mogno; (---) Limiar de discriminação. .... 47

**Figura 24)** Valores de classe estimados para o conjunto de validação para discriminação da espécie curupixá em relação às demais espécies antes e após correção com EPO, respectivamente, na condição de pré-processamento 3 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + *Mean Center*) e no modelo AE. (♦) andiroba; (■) cedro; (▲) cedrinho; (▼) curupixá; (★) mogno; (---) Limiar de discriminação. .... 48

**Figura 25)** Valores de classe estimados para o conjunto de validação para discriminação da espécie mogno em relação às demais espécies antes e após correção com EPO, respectivamente, na condição de pré-processamento 3 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + *Mean Center*) e no modelo AE. (♦) andiroba; (■) cedro; (▲) cedrinho; (▼) curupixá; (★) mogno; (---) Limiar de discriminação. .... 49

**Figura 26)** Amostra de curupixá com fungo..... 50

**Figura 27)** Valores de classe estimados para o conjunto de validação para discriminação das espécies andiroba, cedro, cedrinho e mogno em relação às demais espécies na combinação 6. (♦) andiroba; (■) cedro; (▲) cedrinho; (★) mogno; (---) Limiar de discriminação..... 54

**Figura 28)** Valores de classe estimados para o conjunto de validação para discriminação das espécies andiroba, cedro, cedrinho e mogno em relação às demais espécies na combinação 8. (♦) andiroba; (■) cedro; (▲) cedrinho; (★) mogno; (---) Limiar de discriminação..... 55

## LISTA DE TABELAS

<b>Tabela 1)</b> Detalhamento dos pré-processamentos utilizados em cada condição. ....	25
<b>Tabela 2)</b> Combinações avaliadas para a validação de cada modelo de predição construído. ....	26
<b>Tabela 3)</b> Construção das matrizes de validação das amostras com maior teor de umidade. ....	31
<b>Tabela 4)</b> Combinações para a discriminação das amostras com maior teor de umidade. ....	32
<b>Tabela 5)</b> Média e desvio padrão (s) do teor de umidade das amostras por espécie inicial e após passar pelo processo de secagem. ....	33
<b>Tabela 6)</b> Atribuições de bandas características às vibrações de ligações O-H em espectros NIR de madeira <sup>83</sup> . ....	34
<b>Tabela 7)</b> Número de variáveis latentes aplicado em cada ciclo de PLS-DA para cada condição da matriz de treinamento. ....	38
<b>Tabela 8)</b> Resultados para os valores de taxa de eficiência, amostras indefinidas, amostras classificadas como <i>outliers</i> e como inconclusivas para todos os modelos de predição com pré-processamento 1 (EPO: bruto; PLS-DA: <i>Mean Center</i> ). ....	39
<b>Tabela 9)</b> Resultados para os valores de taxa de eficiência, amostras indefinidas, amostras classificadas como <i>outliers</i> e como inconclusivas para todos os modelos de predição com pré-processamento 2 (EPO: SNV; PLS-DA: SNV + <i>Mean Center</i> ). ....	40
<b>Tabela 10)</b> Resultados para os valores de taxa de eficiência, amostras indefinidas, amostras classificadas como <i>outliers</i> e como inconclusivas para todos os modelos de predição com pré-processamento 3 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + <i>Mean Center</i> ). ....	41
<b>Tabela 11)</b> Resultados para os valores de taxa de eficiência, de amostras indefinidas, amostras classificadas como <i>outliers</i> e como inconclusivas para todos os modelos de predição com pré-processamento 4 (EPO: 1ª derivada; PLS-DA: 1ª derivada + <i>Mean Center</i> ). ....	41

<b>Tabela 12)</b> Medida do teor de umidade inicial das amostras de quatro espécies florestais produtoras de madeira e do teor de umidade final após 55 dias de acondicionamento nas caixas com umidade controlada pelo sulfato de potássio .....	51
<b>Tabela 13)</b> Combinações para a discriminação das amostras com maior teor de umidade. ....	52
<b>Tabela 14)</b> Taxas de eficiência e de amostras indefinidas para todos os modelos de predição das amostras que passaram pela caixa umidificadora na condição 3 de pré-processamento 3 (EPO: SNV + 1 <sup>a</sup> derivada; PLS-DA: SNV + 1 <sup>a</sup> derivada + Mean Center). ....	53

## RESUMO

O comércio ilegal madeireiro é um problema de âmbito mundial cuja fiscalização carece de métodos de identificação de espécies de madeira que sejam confiáveis e de uso simples. Nesse contexto, a espectroscopia no infravermelho próximo (NIRS, do inglês *near infrared spectroscopy*) vem mostrando ser uma técnica bastante promissora. Se trata de uma técnica que permite a análise em campo e que, quando em conjunto com o modelo de mínimos quadrados parciais para análise discriminante (PLS-DA, do inglês *partial least squares for discriminant analysis*) teve sua eficácia demonstrada para identificação de diferentes espécies florestais produtoras de madeira em condições de umidade controladas. Entretanto, o fato da umidade das amostras utilizadas na fase de treinamento ter que ser aproximadamente a mesma das amostras que serão analisadas em rotina tem restringido a aplicabilidade do modelo. A ortogonalização dos parâmetros externos (EPO, do inglês *external parameter orthogonalisation*) é uma ferramenta matemática que vem se mostrando eficaz na correção do efeito da umidade em amostras de origem orgânica. Portanto, a presente dissertação teve como objetivo principal avaliar a utilização do EPO para correção da interferência da umidade nos espectros de madeira afim de minimizar esse problema em modelos de discriminação de amostras de cinco espécies florestais anatomicamente semelhantes (*Carapa guianensis* Aubl., *Cedrela odorata* L., *Erismia uncinatum* Warm., *Micropholis melinoniana* Pierre e *Swietenia macrophylla* King) com teores de umidade diferentes daquelas utilizadas para criar o modelo. Como resultado, teve-se que a correção dos espectros por EPO levou à minimização da diferença espectral causada pelo efeito da umidade. Com isso, as taxas de amostras com resultados indefinidos em modelos de discriminação diminuíram substancialmente. Em uma das condições analisadas, a taxa passou de 79,6 para 14,3% em amostras com teor de umidade menor que as utilizadas para criar o modelo, e de 71,4 para 10,2% em amostras com teor de umidade maior. Essa diminuição no número de *outliers* permite uma maior confiabilidade das taxas de eficiência e real capacidade de discriminação do modelo para amostras de madeira com teores de umidade diferentes. Além disso, a discriminação de amostras com teor de umidade mais elevado se provou eficaz quando teve seus espectros corrigidos por EPO. Tal resultado permite a discriminação confiável e eficaz de amostras de madeira sem a interferência da umidade e possibilita a implementação de uma ferramenta de fiscalização rápida, relativamente econômica e de fácil manuseio.

**Palavras-chave:** discriminação de madeira, NIRS, EPO, PLS-DA, umidade.

## ABSTRACT

Illegal timber trade is a worldwide problem whose control needs reliable and simple-to-use identification methods of species found in trade. In this context, Near Infrared Spectroscopy (NIRS) has proven to be a very promising technique. It allows field analysis and when combined with the partial least squares for discriminant analysis (PLS-DA) is effective in identifying different species of wood, although the samples have to be under controlled moisture conditions. However, the fact that the moisture of the samples used in the training phase has to be approximately the same of the ones analyzed in routine analysis has restricted the applicability of the model. External parameter orthogonalization (EPO) is a mathematical tool that has shown to be effective in the correction of the effect of moisture in organic samples. Therefore, this dissertation aimed to evaluate the use of EPO to correct the interference of moisture in wood spectra to minimize this problem. Furthermore, the correction was tested in samples of five wood species (*Carapa guianensis* Aubl., *Cedrela odorata* L., *Erisma uncinatum* Warm., *Micropholis melinoniana* Pierre e *Swietenia macrophylla* King) in different moisture levels than those used to create the model. As a result, it showed that spectra correction by EPO led to the minimization of the spectral difference caused by the effect of moisture. Consequently, sample rates with undefined results in discrimination models decreased substantially. Under one of the conditions analyzed, this rate decreased from 79.6 to 14.3% in samples with lower moisture content than those used to create the model. For samples with higher moisture content, it decreased from 71.4 to 10.2%. This decrease in the number of outliers gives the model a greater reliability in its efficiency rate and real capability for discrimination of wood samples in different moisture conditions. In addition, discrimination of samples with higher moisture content proved to be effective when their spectra were corrected by EPO. This result allows a reliable and effective discrimination of wood samples without the interference of humidity and enables the implementation of an inspection method that is fast, relatively cheap, and easy to handle.

**Keywords:** timber discrimination, NIRS, EPO, PLS-DA, moisture.

## 1. INTRODUÇÃO E OBJETIVOS

O comércio ilegal madeireiro é uma ameaça real à biodiversidade, contribuindo significativamente para o crescimento contínuo das taxas de desmatamento das florestas e colocando em perigo, inclusive, espécies já ameaçadas de extinção. Nesse sentido, é importante que o governo brasileiro possua agências capazes de realizar ações de fiscalização eficientes do comércio de madeira no país. Entretanto, uma das maiores dificuldades encontradas durante o processo de fiscalização é a identificação de forma rápida e segura da espécie florestal relacionada a uma madeira que está sendo inspecionada. Por conta disso, há uma demanda por métodos de identificação de madeira confiáveis, objetivos e que sejam de uso simples, não demandando um longo período de treinamento de agentes ambientais a cargo das ações de fiscalização (ex.: agentes do IBAMA e policiais rodoviários federais).

Uma técnica que vem se provando bastante relevante para a identificação de madeira é a espectroscopia no infravermelho próximo (NIRS, do inglês *Near Infrared Spectroscopy*). Ela se destaca por ter um curto tempo de análise, não ser destrutiva, necessitar de mínimo preparo da amostra, dispor de equipamentos portáteis comerciais de custo relativamente baixo e exibir o resultado da análise em tempo real, sendo ideal para análise de triagem em campo. Entretanto, mesmo com todas as qualidades da tecnologia NIRS, o efeito da umidade é um problema que vem sendo recorrente em estudos que a utilizam para realizar a identificação da espécie florestal de amostras de madeiras. A presença da água na madeira causa uma variação espectral que interfere no desempenho dos modelos de discriminação, restringindo sua aplicabilidade apenas para amostras que estejam na mesma condição a qual o modelo foi criado.

A ortogonalização dos parâmetros externos (EPO, do inglês *External Parameter Orthogonalisation*) é uma ferramenta matemática que vem mostrando resultados bastante positivos para correção do efeito da umidade em amostras de origem orgânica. Contudo, ainda são poucos os trabalhos com essa ferramenta, sendo que não há relatos na literatura de sua utilização em amostras de madeira, até o presente momento.

Estudos anteriores, frutos da parceria entre o Laboratório de Produtos Florestais (LPF) do Serviço Florestal Brasileiro (SFB) e o Laboratório de Automação, Quimiometria e Química Ambiental (AQQUA) do Instituto de Química da Universidade de Brasília (UnB) evidenciaram o potencial da tecnologia NIRS para a discriminação de

madeiras amazônicas, apresentando resultados bastante promissores, porém, também com problemas referentes à interferência da umidade. Com base nesse contexto, esta dissertação tem como objetivo principal dar continuidade ao projeto de identificação de madeiras com NIRS e Quimiometria, contribuindo para a finalização de um método que possa ser aplicado em campo para a fiscalização de madeiras tropicais brasileiras macroscopicamente semelhantes. Como objetivos específicos essa dissertação propõe: **(a)** avaliar a utilização do EPO para correção da interferência da umidade nos espectros NIR de madeira; caso comprovada, **(b)** aplicá-la de modo a criar modelos de discriminação que sejam eficazes independente da condição de umidade a qual a amostra se encontra; **(c)** propor e avaliar critérios mais flexíveis para classificação de amostras por PLS-DA e **(d)** testar tais modelos em amostras com diferentes teores de umidade, inclusive, daqueles das amostras utilizadas para criá-los. Assim, será avaliada a discriminação de cinco espécies florestais produtoras de madeira: *Carapa guianensis* Aubl. (andiroba), *Cedrela odorata* L. (cedro), *Erisma uncinatum* Warm. (cedrinho), *Micropholis melinoniana* Pierre (curupixá) e *Swietenia macrophylla* King (mogno) através da análise discriminante por mínimos quadrados.

## 2. FUNDAMENTAÇÃO TEÓRICA E REVISÃO BIBLIOGRÁFICA

### 2.1. Madeira

A extração e o comércio ilegal de madeira ameaçam não apenas as espécies exploradas, mas também o ecossistema o qual fazem parte e causam fortes danos ambientais, sociais e econômicos. É crescente a preocupação internacional com esse comércio ilegal, o que tem levado à restrição ou à proibição da importação de determinadas espécies florestais produtoras de madeira<sup>1</sup>. A exploração ilegal de espécies produtoras de madeiras é motivada principalmente pela alta demanda do mercado mundial. Os Estados Unidos e a China, juntos, foram responsáveis por 33% de todas as importações de madeira do ano de 2017<sup>2</sup>. Produtos feitos de madeira cortada ilegalmente causam desvalorização nos preços e impulsionam o mercado global, resultando no aumento das taxas de desmatamento e de extração ilegal nas principais regiões produtoras de madeira, como as florestas tropicais da bacia amazônica, África Central e sudeste asiático<sup>2,3</sup>.

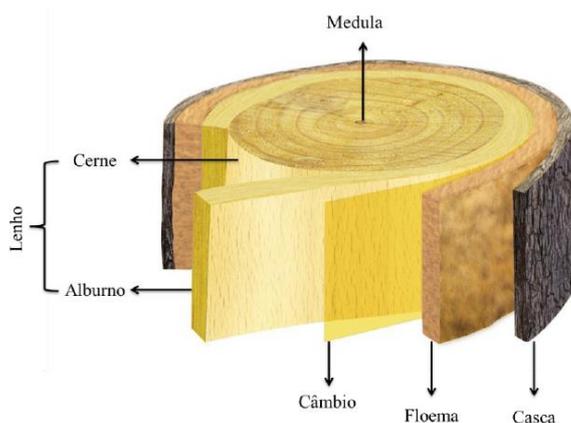
Buscando maior transparência no comércio de madeira, alguns países vêm adotando leis mais restritivas de importação, assim como medidas legislativas que possam conter o mercado provindo de exploração ilegal, como é o caso do Canadá, em 1992, dos Estados Unidos, em 2008, da União Europeia, em 2010, e da Austrália, em 2012<sup>4</sup>. Entretanto, são poucos os mecanismos em nível internacional para combater o comércio ilegal de madeira. A aplicação obrigatória da Convenção sobre o Comércio Internacional das Espécies da Fauna e da Flora Silvestres Ameaçadas de Extinção (CITES, do inglês *Convention on International Trade in Endangered Species of Wild Fauna and Flora*) por 183 países no mundo busca estabelecer o controle internacional no comércio de espécies de plantas e de animais, visando a garantia que não haverá ameaça à sobrevivência das espécies, principalmente daquelas listadas em seus apêndices como estando em risco<sup>5</sup>. Especificamente, para espécies florestais produtoras de madeira, essa legislação se aplica a toras, madeira serrada, madeira faqueada e móveis, sendo em alguns casos exigida a identificação ao nível de espécie<sup>6</sup>.

Na América do Sul, a exploração predatória e seletiva de *Swietenia macrophylla* King a fez ser incluída no Apêndice II da CITES<sup>5</sup>, que lista as espécies que não estão necessariamente ameaçadas de extinção, mas que podem se tornar, a menos que o comércio seja controlado de perto. Outra espécie nativa do Brasil, a *Dalbergia nigra*, está no Apêndice I, isto é, possui grau de proteção máximo por estar em risco de extinção.

Entretanto, um dos maiores desafios na implementação da CITES é a identificação das espécies encontradas no comércio, necessárias para fiscalização<sup>1</sup>.

### 2.1.1. Estrutura Anatômica

A madeira pode ser obtida de diversas partes da árvore: tronco, raízes e ramos. Entretanto, o tronco é a maior fonte devido às suas maiores proporções em relação às outras partes da árvore. O tronco é organizado em estruturas macroscópicas na seguinte ordem exterior/interior: casca, floema, câmbio, lenho e medula, como mostra a figura 1:



**Figura 1)** Estrutura macroscópica do tronco de uma árvore: casca, floema, câmbio, alburno, cerne e medula<sup>7</sup> (adaptado).

A casca é a estrutura mais externa do tronco e é responsável pela sua proteção; o floema é formado por tecidos superficiais que são responsáveis pelo transporte da seiva; o câmbio é uma camada de células situada entre o lenho e o floema; o lenho é a parte do tronco onde se extrai a madeira e é dividido em duas zonas: cerne e alburno; e a medula é estrutura mais interna, formada a partir das células que constituíram a zona de crescimento inicial que deu origem ao tronco<sup>8</sup>. Com o passar dos anos, o tronco aumenta em circunferência e o tecido interno que circunda a medula morre, ou seja, morrem os elementos vivos que compõem a parte periférica do lenho, chamada de alburno. Esse núcleo da madeira morta no centro do caule é conhecido como cerne, que fornece apenas suporte mecânico à árvore<sup>7</sup>.

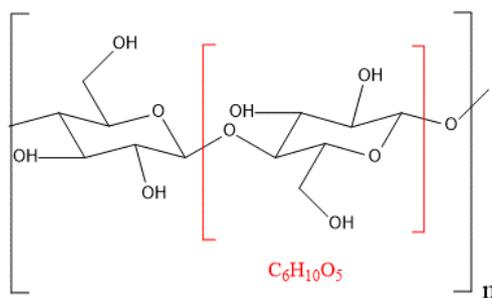
### 2.1.2. Composição Química

Todas as madeiras são derivadas de duas amplas categorias de plantas, as gimnospermas (sementes não-envolvidas pelo ovário da flor) e as angiospermas (sementes envolvidas pelo ovário da flor). No primeiro grupo, incluem-se as árvores que fornecem madeiras *softwoods* ou madeiras brancas (também conhecidas como coníferas),

enquanto no segundo grupo, estão as árvores que fornecem madeiras duras ou *hardwoods* (também chamadas de folhosas)<sup>9</sup>. Esses dois grupos de madeiras apresentam diferenças em sua composição química. No Brasil, a grande maioria das madeiras nativas pertence ao grupo das folhosas<sup>10</sup>.

A madeira é composta basicamente por três polímeros: a celulose, a hemicelulose e a lignina, em uma proporção aproximada de 50:25:25, distribuídas na parede celular. Além dos componentes principais, a madeira também possui os compostos de baixo peso molecular que são formados pelos extrativos e pelas substâncias minerais<sup>11</sup>.

A celulose é a base estrutural das células das plantas e, em termos socioeconômicos, é o componente mais importante da madeira, sendo matéria-prima para produzir polpa para papel e fibras naturais e derivados, como algodão, rayon, viscose e tencel<sup>12</sup>. Quimicamente, a celulose é um carboidrato, insolúvel em água, sem sabor, dificilmente isolado da madeira na forma pura, por estar associado com a lignina e com a hemicelulose<sup>13</sup>. A celulose é um polímero de cadeia longa de peso molecular variável, com fórmula geral  $(C_6H_{10}O_5)_n$ , com um valor mínimo de  $n=200$ <sup>14</sup>. Fisicamente, a celulose é um material sólido encontrado na forma cristalina ou amorfa. A madeira possui cerca de 50 a 70 % de regiões cristalinas que formam a estrutura da fibra da madeira<sup>15</sup>. A forma cristalina da celulose é particularmente resistente ao ataque químico e à degradação<sup>15</sup>. Estas propriedades devem-se, em parte, ao fato de seu peso molecular ser extremamente elevado. As cadeias longas, lineares e uniformes, estão dispostas lado a lado, em feixes, e são mantidas fixas, por meio de ligações de hidrogênio com os grupos OH adjacentes, como pode ser visto na figura 2. Estes feixes encontram-se enrolados sobre si mesmos em forma de estruturas entrelaçadas, as quais, por sua vez, estão agrupadas umas às outras, formando as fibrilas, que podem ser vistas por inspeção visual<sup>16</sup>. As madeiras do grupo das coníferas são compostas por, aproximadamente, 45-50% de celulose e as do grupo das folhosas por cerca de 40-50%<sup>15</sup>.



**Figura 2)** Estrutura da unidade repetidora da celulose

As hemiceluloses são materiais sólidos brancos, raramente cristalinos ou fibrosos, que formam um tipo de preenchimento da fibra que aumenta a resistência do papel e o rendimento da polpa<sup>15</sup>. Quimicamente, as hemiceluloses ou polioses são a fração polimérica de polissacarídeos, constituída de unidades de vários açúcares sintetizados na madeira e em outros tecidos das plantas. Diferem da composição da celulose pelas cadeias moleculares mais curtas e pelas ramificações da cadeia principal. As hemiceluloses possuem alta reatividade, sendo solúveis em álcalis e são facilmente hidrolisáveis por ácidos diluídos para formarem açúcares. Constituem cerca de 27% das madeiras do grupo das coníferas e 30% das do grupo das folhosas<sup>17</sup>.

Ligninas são polímeros formados de unidades fenilpropano (em estruturas tridimensionais), amorfos, altamente complexos e, essencialmente, aromáticos na madeira. O seu peso molecular é muito alto e não é facilmente mensurável<sup>15</sup>. As ligninas podem ser classificadas de acordo com os seus três elementos estruturais básicos: álcool p-coumaril, álcool coniferil e álcool sinapil. As madeiras do grupo das folhosas contêm dois deles, o álcool coniferil (50-75%) e o álcool sinapil (25-50%), e as do grupo das coníferas contêm somente o álcool coniferil. A quantidade de ligninas pode variar de 20 a 40% entre as espécies de madeira, ocorrendo, principalmente, em tecidos vasculares, porém sua distribuição não é uniforme nas diferentes partes da árvore. As ligninas estão também associadas às hemiceluloses e, em alguns casos, formam complexos lignina-carboidrato, resistentes à hidrólise<sup>13</sup>. Existem fortes evidências de que as ligninas estejam ligadas covalentemente às hemiceluloses, porém as investigações ainda permanecem não conclusivas quanto a este aspecto<sup>18</sup>. Depois da deposição das hemiceluloses e da celulose, a lignina é incorporada em quaisquer aberturas da parede celular, conferindo uma maior resistência mecânica à madeira, de forma a reduzir seu caráter higroscópico e a protegê-la contra a degradação microbiológica<sup>19</sup>.

Os extrativos são compostos orgânicos extraíveis da madeira. São separados dos compostos estruturais por meio de extração com solventes orgânicos neutros (etanol, acetona ou diclorometano) e água. Esta definição possui algumas restrições, pois as ligninas são parcialmente solúveis em solventes orgânicos polares e, nas madeiras ricas em materiais fenólicos, alguns extrativos altamente polimerizados são insolúveis tanto em água, como em solventes orgânicos neutros<sup>18</sup>. Estima-se que a quantidade de extrativos que ocorre em madeiras de clima temperado varie de 4 a 10% (m/m) do seu

peso seco. Já as espécies tropicais podem alcançar valores maiores que 20% (m/m)<sup>13</sup>. Os extrativos incluem um número muito grande de substâncias químicas radicalmente diferentes<sup>20</sup>. Devido a esta diversidade, é necessário tratá-las em subgrupos, por exemplo, os terpenóides, os compostos alifáticos e os fenólicos, conforme classificação sugerida por Sjöström<sup>21</sup>.

### **2.1.3. Umidade**

A madeira é um material higroscópico, ou seja, é susceptível a trocas de umidade com a atmosfera<sup>22-25</sup>. Essa troca de água depende de alguns fatores, como: umidade relativa, temperatura do ambiente e quantidade atual de água presente na madeira. A madeira proveniente de árvores recém-derrubadas é chamada de madeira verde e apresenta um alto teor de umidade. A tendência é que este teor diminua lenta e espontaneamente até atingir um ponto de equilíbrio. Porém, a velocidade dessa diminuição do teor de umidade varia de acordo com a espécie florestal, do corte aplicado à tora, da dimensão da peça cortada, das condições ambientais e do empilhamento aplicado<sup>26</sup>.

A primeira forma de água a ser evaporada da madeira verde é aquela que se encontra nas cavidades celulares (lúmens) das células e nos espaços intercelulares, chamada água livre (ou água de capilaridade). Ela é mantida por forças capilares fracas, necessitando de pouca energia para ser removida<sup>19,27</sup>. Quando toda a água livre é removida, as cavidades e paredes celulares da madeira ficam saturadas com água presa (também chamada de água higroscópica). Diferentemente da livre, a água presa se encontra adsorvida pelas moléculas de hemicelulose e celulose presentes na parede celular, atingindo um teor de umidade denominado ponto de saturação das fibras. Além da água presa, há também água na forma de vapor em partes do lúmen onde não se localiza mais água no estado líquido<sup>28</sup>.

Quando madeira com teor de umidade menor que a umidade relativa do ar é exposta ao meio ambiente, há a adsorção da água dispersa no ar e ela será identificada como água presa na madeira<sup>29</sup>. Nesta situação é atingido o teor de umidade de equilíbrio<sup>26</sup> com o ambiente, que é dependente da espécie de madeira e das condições ambientais circundantes.

A secagem da madeira se baseia na remoção da umidade da sua superfície ao mesmo momento em que acontece o movimento de água do interior para o exterior,

devido a capilaridade e a difusão. Diversos fatores influenciam na velocidade da secagem da madeira, assim como no seu teor de umidade de equilíbrio, tais como a temperatura, a umidade relativa do ar, a circulação do ar e características intrínsecas da madeira, como a espécie, a espessura da peça, a orientação do corte, a permeabilidade, o teor de umidade inicial, a densidade básica e a utilização de cerne e/ou alborno a serem analisados<sup>19,22,23,26,27,30</sup>.

Existem vários métodos de secagem de madeira descritos na literatura, desde aqueles usados em laboratório para fins experimentais até os utilizados na indústria, como a secagem natural (expondo a madeira ao ar livre), solar (por meio de uma estufa solar), convencional (por meio de estufa tradicional com controle de aquecimento) ou por desumidificação ou à vácuo<sup>25,26,30</sup>. Além destes, um outro método de climatização se fundamenta na utilização de soluções salinas saturadas em pequenos recipientes fechados. As soluções salinas, quando preparadas a uma concentração definida e mantidas a uma temperatura constante, permanecem em equilíbrio com uma determinada pressão parcial de vapor d'água, promovendo um meio com umidade relativa do ar controlada<sup>31,32</sup>. Sabendo, assim, que as soluções salinas saturadas dispostas em recipientes fechados podem adsorver ou liberar grandes quantidades de água sem modificar o teor de umidade de equilíbrio, esse método pode ser adotado não somente para fins de secagem, mas também para a umidificação da amostra.

Entender essas relações da umidade com a madeira, bem como métodos de secagem, possibilita uma maior compreensão do comportamento da madeira diante das mais adversas condições ambientais que possam vir a ocorrer.

#### **2.1.4. Métodos de Identificação de Madeira**

Um dos maiores desafios na implementação da CITES é a identificação das espécies encontradas no comércio, necessárias para fiscalização<sup>1</sup>. O Escritório das Nações Unidas para Drogas e Crime (UNODC, do inglês *United Nations Office on Drugs and Crime*) publicou um guia de boas práticas para a identificação de madeira para fins forenses que lista os principais métodos recomendados para essa análise, sendo estes: anatomia da madeira, dendrocronologia, espectrometria de massas (DART-MS, do inglês *Direct Analysis in Real Time Mass Spectrometry*), espectroscopia no infravermelho próximo (NIRS, do inglês *Near Infrared Spectroscopy*), análise por isótopos estáveis, radiocarbono e DNA<sup>1</sup>.

No que se diz respeito especificamente à identificação de madeira, o método mais difundido e empregado para identificação em campo é a análise por anatomia, que compara os caracteres anatômicos e morfológicos da madeira examinada com a de padrões depositados em coleções de madeiras ou xilotecas registradas pela IAWA (do inglês *International Association of Wood Anatomists*)<sup>33</sup>. Entretanto, esta forma de identificação requer um analista com alto grau de experiência, uma vez que existem espécies que apresentam características estruturais de madeira muito similares, facilmente confundíveis e subjetivas estando sujeitas ao erro humano<sup>6</sup>.

Essa análise anatômica ao nível de família ou gênero está sujeita a falhas na identificação em espécies com características semelhantes<sup>34</sup> e para identificar assinaturas de madeira ilícita recorre-se às metodologias de análise como a genômica (já consolidada) e as físico-químicas (em desenvolvimento), que utilizam características intrínsecas da amostra e permitem a sua identificação mais precisa e apurada. O DART-MS tem apresentado resultados bastante positivos para identificação de espécies florestais<sup>35</sup>. Este método visa caracterizar os fitoquímicos presentes no cerne, tornando possível gerar um perfil químico ou impressão digital para cada espécie de madeira. O laboratório forense americano do *U.S. Fish & Wildlife Services*, vem investindo neste método de análise e está fazendo uma ampla varredura em xilotecas ao redor do mundo, tornando público um banco de espectros<sup>36</sup>. Entretanto, a análise por espectrometria de massas envolve o uso de um equipamento de alto custo e que requer que a amostra seja levada a um laboratório, inviabilizando a análise em campo.

A tecnologia NIRS, que une a espectroscopia no infravermelho próximo com a análise multivariada, permite a aquisição direta de medidas de reflectância, que carregam informação dos grupos funcionais presentes nos compostos químicos que formam a madeira. Seu espectro também contém informações referentes às propriedades físicas e anatômicas, uma vez que a distribuição dos constituintes químicos na superfície da madeira se dá de forma heterogênea entre as espécies, porém repetindo o mesmo padrão dentro de uma espécie vegetal. Todas essas informações contidas no espectro NIRS, torna possível sua utilização para a caracterização da madeira e sua identificação a nível de gênero, espécies, origem, etc<sup>24,37,38</sup>. Além disso, essa técnica apresenta a vantagem de permitir o uso de equipamentos portáteis comerciais para realização de análises em campo e em tempo real, contribuindo para uma maior aplicação do método, quando comparada aos demais. Para análises em tempo real de diversas espécies florestais produtoras de

madeira, vários pesquisadores vêm utilizando equipamentos portáteis comerciais NIRS em seus estudos<sup>37-41</sup>. Os aparelhos portáteis são robustos, possuem baixo custo quando comparado com equipamento de bancada de laboratório FT-NIR (transformada de Fourier no infravermelho próximo) e, principalmente, não necessitam de longo período de treinamento para sua utilização<sup>38</sup>.

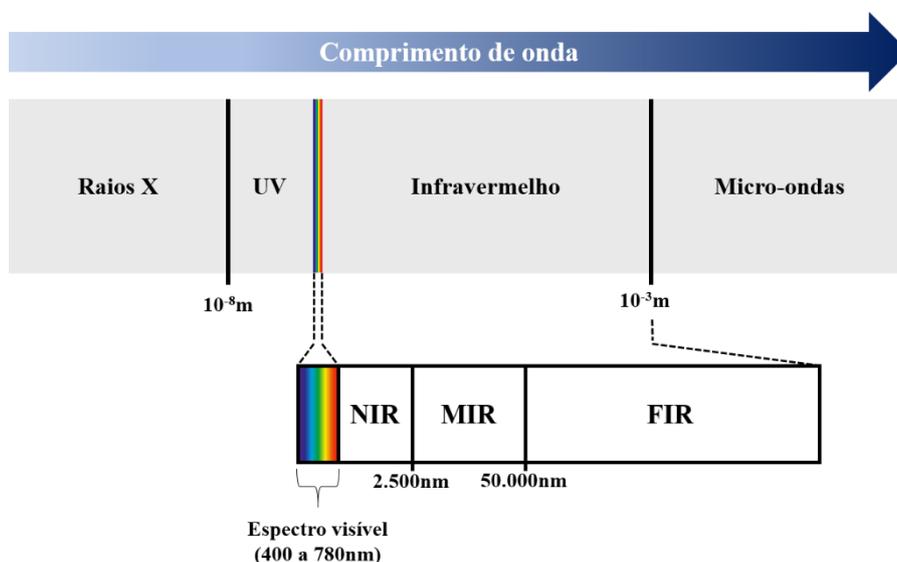
Devido sua eficiência como método de rastreamento de madeira, a tecnologia NIRS foi inserida como uma das cinco ferramentas disponíveis no documento publicado pela GTTN (do inglês *Global Timber Tracking Network*) em 2020<sup>41</sup>, que fornece uma visão geral das práticas atuais de análise para identificação de madeira. Diversos trabalhos publicados, inclusive do nosso próprio grupo de pesquisa, já demonstraram a utilização da tecnologia NIRS na discriminação de espécies do mesmo gênero<sup>37,42</sup>, de diferentes gêneros<sup>39,43</sup> e até mesmo da mesma espécie em diferentes procedências<sup>38,44</sup>, comprovando a eficácia do método. Entretanto, todos esses trabalhos têm em comum a característica de terem sido desenvolvidos com madeira seca e em condições controladas de umidade e temperatura. Deste modo, devido à alta sensibilidade do NIRS à presença de umidade, é comum que os modelos criados apresentem erros na predição quando as amostras preditas estão em uma condição de umidade diferente das amostras empregadas na fase de treinamento dos modelos. Um estudo anterior do nosso grupo de pesquisa<sup>8</sup> avaliou a possibilidade de corrigir este problema por meio da aplicação da ferramenta matemática de correção de sinal ortogonal (OSC, do inglês *Orthogonal Signal Correction*), porém, esta não gerou resultados satisfatórios. Mesmo depois da aplicação da correção por OSC, o modelo corrigido classificou 70% das amostras de madeira com maior teor de umidade como amostras anômalas (do inglês, *outliers*), evidenciando que a correção não havia sido efetiva. O mesmo estudo também sugeriu a secagem das amostras de forma experimental, entretanto, apesar de apresentar boas taxas de eficiência, o modelo criado apresentou um número elevado de *outliers* e o processo de secagem demandaria um preparo de amostra que não é ideal para análises de campo.

Uma ferramenta matemática que vem mostrando resultados bastante positivos para correção do efeito da umidade em amostras de origem orgânica é a ortogonalização dos parâmetros externos (EPO, do inglês *External Parameter Orthogonalisation*)<sup>45-47</sup>, que, em resumo, remove a informação espectral referente ao efeito que se deseja corrigir do conjunto de dados. Em consequência de seus resultados significativos corrigindo o

efeito da umidade em espectros de amostras de solo, este trabalho testará sua eficácia em amostras de madeira.

## 2.2. Espectroscopia NIR

A espectroscopia na região do infravermelho (NIRS, do inglês *near infrared spectroscopy*) é uma técnica de espectroscopia vibracional que se baseia na absorção de energia proveniente da radiação infravermelha, o que provoca mudanças nos estados vibracionais e rotacionais das moléculas<sup>48</sup>. Porém, isso só acontece quando os movimentos de vibração e rotação naturais da molécula causam variações no seu momento dipolar, resultando a geração de um campo elétrico oscilante<sup>49</sup>. Essa região espectral pode ser dividida em infravermelho (i) próximo, que está contido na faixa de 780 a 2.500 nm; (ii) médio, na faixa de 2.500 a 50.000 nm; e (iii) distante, contido entre 50.000 nm e 1 mm<sup>48</sup>, como exibido na figura 3.



**Figura 3)** Parte do espectro eletromagnético com destaque a região do visível ao infravermelho (adaptado<sup>50</sup>). (UV): ultravioleta; (NIR): infravermelho próximo; (MIR): infravermelho médio; (FIR): infravermelho distante.

A radiação eletromagnética referente à faixa do infravermelho próximo foi descoberta em 1800 por Sir William Herschel, constituindo a primeira parte do espectro eletromagnético não visível a ser identificada<sup>51</sup>. Devido à falta de seletividade dos espectros de absorção na região NIR, na qual os espectros são constituídos por inúmeras bandas de absorção fortemente sobrepostas, essa região espectral é considerada de difícil interpretação. Sua aplicação prática apenas se popularizou depois da realização dos

trabalhos de Karl Norris, que mostraram que este tipo de espectroscopia necessitava de ferramentas matemáticas multivariadas para contornar o problema da falta de seletividade da informação contida em seus espectros. Assim, a partir da década de 1980, se começou a introduzir nos trabalhos algumas técnicas quimiométricas de análise multivariada de dados<sup>52</sup>.

Como dito anteriormente, a absorção na região do infravermelho só ocorre quando há variação no momento de dipolo da molécula. Embora isso seja verdade considerando um modo de vibração fundamental, é importante notar que para bandas de combinação permitidas por anarmonicidade, seria necessário que apenas uma das vibrações combinadas estivesse ativa (causando mudança no momento dipolar). Esse recurso pode fazer com que algumas vibrações, que não podem ser observadas no infravermelho médio, sejam exibidas por um espectro NIR<sup>53</sup>.

A intensidade de uma determinada banda de absorção está associada à magnitude da mudança no momento dipolar durante o deslocamento de átomos em uma vibração e ao seu grau de anarmonicidade. Ambos os fenômenos estão presentes em grande extensão quando associados a ligações envolvendo o átomo de hidrogênio e alguns outros elementos mais pesados, como carbono, nitrogênio e enxofre. Assim, as ligações O-H, C-H, N-H e S-H tendem a apresentar alta anarmonicidade e alta energia de ligação com transições vibracionais fundamentais na região de 3.000 a 4.000nm. Portanto, permite prever a existência de sobretons e das combinações dessas vibrações fundamentais que ocorrem na região de energia associada aos fótons NIR. As ocorrências espectrais na região NIR são dominadas por estas bandas de absorção de sobretons e de combinações<sup>53</sup>. Como essas ocorrências são baseadas em efeitos relacionados à energia de ligação, ao grau de anarmonicidade e às possibilidades de combinação entre modos vibracionais, o NIR apresenta uma seletividade relativamente baixa, o que impossibilita o uso de técnicas analíticas quantitativas convencionais. Assim, o método somente pode ser explorado com o auxílio das ferramentas matemáticas e estatísticas desenvolvidas no âmbito da área de estudo da Quimiometria<sup>52</sup>.

A espectroscopia NIR vem sendo muito explorada para desenvolver métodos analíticos precisos, exatos, rápidos, não destrutivos, não invasivos, com alto poder de penetração do feixe de radiação, com mínima ou nenhuma preparação de amostras, além de não gerar resíduos químicos.

### 2.3. Quimiometria

Devido ao problema da baixa seletividade da informação analítica contida em espectros NIR, os sinais observados nessa região referentes a bandas de combinações e sobretons de ligações C-H, O-H, N-H e S-H, são relativamente pouco intensos e bastante sobrepostos, tornando o uso da Quimiometria absolutamente necessário.

A Quimiometria utiliza métodos matemáticos e estatísticos para definir as condições ótimas de medidas e experimentos e/ou para extrair o máximo de informações a partir da análise de dados químicos<sup>54</sup>. Do ponto de vista da análise de dados, sua utilização permite realizar reconhecimento de padrões, quantificação ou classificação sem a necessidade de resolução do sinal dos analitos de interesse. Assim, é possível substituir a separação física dos interferentes pela separação matemática dos seus sinais, usando a informação química contida em seus espectros. Desta forma, pode-se desenvolver métodos analíticos mais simples, rápidos, relativamente baratos e que tornam o processo de pré-tratamento das amostras mínimo ou até mesmo desnecessário, sendo, assim, ambientalmente amigáveis por geralmente não gerarem resíduos, nem consumirem reagentes ou solventes<sup>55</sup>.

Existem diversos métodos de análise multivariada que apresentam diferentes tipos de finalidade, logo, antes de sua escolha e aplicação, deve-se levar em conta os objetivos da análise em questão.

#### 2.3.1. Ortogonalização dos Parâmetros Externos – EPO

O método EPO (do inglês, *External Parameter Orthogonalisation*) é uma técnica de pré-processamento que visa remover ou, pelo menos, minimizar o efeito de parâmetros externos em espectros. O termo “parâmetro externo” se refere a um parâmetro que influencia os espectros NIR, mas que não tem relação com a propriedade de interesse que se deseja calibrar ou na separação das classes. Neste estudo, o parâmetro externo que se pretende remover é a deformação espectral causada pelo aumento do conteúdo de umidade na madeira, correspondente às vibrações O-H. Em resumo, o método encontra as regiões do espectro que são afetadas pelo aumento da concentração de água e, então, define um espaço vetorial que representa essa variação. Em seguida, é calculada uma projeção que seja ortogonal a esse espaço vetorial. A multiplicação de um novo espectro por essa projeção ortogonal, em teoria, removeria ou minimizaria as variações de umidade no espectro NIR<sup>46,47</sup>.

O algoritmo EPO assume que uma matriz  $\mathbf{X}$  (de tamanho  $n \times m$ ) pode ser decomposta em três partes, como ilustrado na equação 1<sup>46</sup>:

$$\mathbf{X} = \mathbf{XP} + \mathbf{XQ} + \mathbf{R} \quad (1)$$

onde  $\mathbf{P}$  é a matriz projeção (de tamanho  $m \times m$ ) da informação/variação espectral que se refere à propriedade de interesse que será modelada (idealmente ortogonal à informação do parâmetro externo):  $\mathbf{X}^* = \mathbf{XP}$ ;  $\mathbf{Q}$  é a matriz projeção (de tamanho  $m \times m$ ) da informação/variação espectral causada pelos parâmetros externos:  $\mathbf{X}^\# = \mathbf{XQ}$ ; e  $\mathbf{R}$  é uma matriz de resíduos (de tamanho  $n \times m$ ), idealmente independentes e aleatórios. Assumindo que a matriz de resíduos  $\mathbf{R}$  é composta apenas por variação aleatória, o objetivo do EPO é obter uma estimativa de  $\mathbf{X}^*$ , o qual geralmente é estimada a partir da subtração da informação devido o parâmetro externo ( $\mathbf{X}^\#$ ) dos dados originais  $\mathbf{X}$  através de uma projeção ortogonal.

Para este fim, primeiramente, se calcula a matriz  $\mathbf{D}$  (de tamanho  $n \times m$ ), que é a diferença entre espectros obtidos para as mesmas amostras em diferentes teores de umidade, como pode ser visto na equação 2<sup>46</sup>:

$$\mathbf{D} = \mathbf{X}_{\text{úmido}} - \mathbf{X}_{\text{seco}} \quad (2)$$

Em seguida, é construído um modelo PCA (análise das componentes principais, do inglês *Principal Component Analysis*) pela decomposição em valores singulares (SVD, do inglês *Singular Value Decomposition*) de  $\mathbf{D}^T\mathbf{D}$ , obtendo assim  $\mathbf{USV}^T$ , onde  $\mathbf{U}$  ( $n \times n$ ) e  $\mathbf{V}$  ( $m \times m$ ) são matrizes quadradas que contém os vetores singulares dos espaços das linhas e colunas de  $\mathbf{D}$ , respectivamente, e  $\mathbf{S}$  é uma matriz diagonal  $n \times m$  contendo os valores singulares de  $\mathbf{D}$ . Define-se então o número de componentes  $c$  a ser utilizado para modelar a informação da umidade em  $\mathbf{D}$ , resultando em  $\mathbf{V}_s$ , uma matriz composta das primeiras  $c$  colunas de  $\mathbf{V}$ <sup>46</sup>.

Através dessa decomposição de  $\mathbf{D}$ ,  $\mathbf{Q}$  é estimado por  $\mathbf{Q} = \mathbf{V}_s\mathbf{V}_s^T$ . Na sequência, a matriz projeção  $\mathbf{P}$  pode ser calculada por  $\mathbf{P} = \mathbf{I} - \mathbf{Q}$ , onde  $\mathbf{I}$  é uma matriz identidade do mesmo tamanho de  $\mathbf{Q}$ . Por fim, calcula-se o espectro sem a informação da umidade pela multiplicação da matriz de espectros originais pela matriz  $\mathbf{P}$ :  $\mathbf{X}^* = \mathbf{XP}$ <sup>46</sup>.

### 2.3.2. Mínimos Quadrados Parciais para Análise Discriminante – PLS-DA

A análise discriminante pelo método de mínimos quadrados parciais apresenta dois tipos de modelos que podem ser utilizados<sup>56</sup>. No modelo conhecido como PLS1-DA,

a matriz de dados  $\mathbf{X}$  é correlacionada com um vetor  $\mathbf{y}$ , no qual cada classe é discriminada em relação às outras em modelos distintos. Já no PLS2-DA, os vetores que discriminam cada classe em relação às outras são reunidos em uma matriz  $\mathbf{Y}$  e um único modelo de discriminação é construído<sup>57</sup>. Neste estudo, foi utilizado o modelo PLS2-DA.

O processo geral da análise consiste nas etapas de treinamento e validação. Na etapa de treinamento, são selecionadas as amostras para estabelecer uma relação entre a matrizes  $\mathbf{X}$  e  $\mathbf{Y}$  por meio da modelagem pelo algoritmo PLS. Na etapa de validação, um outro conjunto de amostras (que não faz parte da etapa de treinamento) é selecionado para ser testado e avaliar a eficiência do modelo. Assim, o modelo pode ser utilizado para discriminar novas amostras daquele mesmo material, geralmente nas mesmas condições das amostras utilizadas para a etapa de treinamento. Contudo, em situações em que as análises são realizadas em campo, é muito comum que as amostras analisadas estejam em uma condição diferente das amostras de treinamento. Por conta disso, a utilização do EPO é de suma importância para que as medidas em campo possam ser compatíveis com as medidas realizadas para as amostras nas condições de treinamento, já que diferentes valores de umidade podem levar a erros de classificação ou elevada ocorrência de *outliers*.

No que diz respeito às etapas de treinamento e validação, os conjuntos de amostras precisam ser escolhidos de modo que os resultados obtidos sejam confiáveis e levem em consideração a representatividade da variabilidade das amostras. Diversos algoritmos já foram propostos na literatura para essa finalidade, sendo os algoritmos de Kenard-Stone e Duplex os mais empregados<sup>58</sup>.

Na fase de treinamento, a matriz  $\mathbf{X}$  é formada pelos espectros das amostras cujas classes são conhecidas. A discriminação de cada classe é definida na matriz  $\mathbf{Y}$  pelas colunas contendo valores de 1 e 0, onde o valor 1 é atribuído às amostras pertencentes à classe que se pretende discriminar e o valor 0, às amostras pertencentes às outras classes. Na sequência, a decomposição dos dados é realizada pelas equações 3 e 4 a seguir<sup>59,60</sup>:

$$\mathbf{X} = \sum_{a=1}^A \mathbf{t}_a \mathbf{p}_a^T + \mathbf{E} \quad (3)$$

$$\mathbf{Y} = \sum_{a=1}^A \mathbf{t}_a \mathbf{q}_a^T + \mathbf{F} \quad (4)$$

onde  $\mathbf{t}_a$  é o vetor de escores;  $\mathbf{p}_a$  e  $\mathbf{q}_a$  são os pesos referentes à primeira variável latente; e  $\mathbf{E}$  e  $\mathbf{F}$  são as matrizes de erros de  $\mathbf{X}$  e  $\mathbf{Y}$ , respectivamente.

O número de variáveis latentes (VL) a ser utilizado no PLS-DA, geralmente, é determinado através do modelo que apresenta o menor valor de erro de classificação empregando validação cruzada (CVCE, do inglês *cross validation classification error*). Entretanto, em CVCE, se a amostra for corretamente classificada, não há penalização mesmo que esta apresente um erro elevado na estimativa do valor de classe ( $y$ ). Deve-se levar em consideração que um erro elevado pode ser um indicativo que a amostra apresenta características diferentes das demais amostras do conjunto de treinamento, ou que há a presença de uma amostra com um erro nos dados instrumentais, situações que caracterizam uma amostra anômala (em inglês, *outlier*). Entretanto, se o valor da raiz quadrada do erro médio quadrático de validação cruzada (RMSECV, do inglês *root mean square error of cross validation*) for utilizado, há uma minimização nesses erros de estimativa dos valores de classe, que tendem a proporcionar uma maior separação dos valores estimados para a classe discriminada ( $y=1$ ) em relação às outras classes ( $y=0$ ). Por conta disto, neste trabalho, o RMSECV foi o critério utilizado para determinar o número de variáveis latentes e, assim, otimizar os modelos PLS2-DA<sup>39</sup>.

Como já mencionado anteriormente, a presença de *outliers* (tanto na fase de treinamento, como na de validação ou previsão) pode levar a erros na discriminação e menor eficiência dos modelos. Assim, sua identificação e exclusão são bastante relevantes na otimização e aplicação do modelo. Diversos trabalhos da área de Quimiometria utilizam métodos para identificação de *outliers*, sendo os principais detalhados a seguir<sup>61-66</sup>:

- Amostras que apresentem valores dos parâmetros estatísticos de  $T^2$  de *Hotelling* e resíduo espectral  $Q$  acima do limite de 99% são identificadas como anômalas e removidas do conjunto de treinamento;
- Amostras que apresentem resíduos Student que excedem o valor crítico do teste  $t$  descrito na ASTM E1655-05<sup>62</sup> com correção de viés<sup>67</sup> no nível de 99% de confiança também são identificadas como anômalas e excluídas do conjunto de treinamento.

A soma quadrática dos resíduos espectrais, também conhecidos como resíduos  $Q$ , representam a parte não modelada do espectro de cada amostra pelo PLS-DA, sendo  $Q$  calculado utilizando a equação 5:

$$Q = \mathbf{e}_i \mathbf{e}_i^T \quad (5)$$

onde  $\mathbf{e}_i$  é a  $i$ -linha da matriz de resíduos  $\mathbf{E}$  dada na equação 3. Existem diversas maneiras de determinar os limites de confiança para os resíduos  $Q$ <sup>68</sup>. Para o presente estudo, o valor crítico será calculado de acordo com o método proposto por Jackson e Mudholkar<sup>69</sup>. Se um espectro de amostra pertencer a uma das classes incluídas no conjunto de treinamento (classes predefinidas) e não contiver nenhuma contaminação ou erro de medição, deve apresentar um valor  $Q$  compatível com os espectros das amostras utilizadas na etapa de treinamento. Caso contrário, se a amostra apresentar um valor  $Q$  significativamente superior aos observados no conjunto de treinamento (ex. acima do limite de 95 ou 99% de confiança), isso indica que o espectro da amostra não é compatível com aqueles das amostras e classes incluídas no conjunto de treinamento.

A medida de  $T^2$  de *Hotelling* está relacionada com a distância da amostra até o centro dos dados de treinamento, e seu cálculo é realizado para cada amostra empregando a equação 6:

$$T_{Hotelling}^2 = \frac{\mathbf{t}_i^T (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{t}_i}{I - 1} \quad (6)$$

onde  $\mathbf{T}$  é a matriz de escores das amostras do conjunto de treinamento com  $I$  amostras e calculado com  $VL$  variáveis latentes e  $\mathbf{t}_i$  é a  $i$ -linha da matriz  $\mathbf{T}$ <sup>70</sup>. Enquanto o resíduo  $Q$  é baseado na variância residual de cada amostra, o  $T^2$  de *Hotelling* se refere à variância modelada em cada amostra dentro do modelo<sup>64</sup>.

Os parâmetros  $T^2$  de *Hotelling* e resíduos  $Q$  são aplicados da mesma maneira para as amostras do conjunto de validação. Por outro lado, o teste  $t$  para resíduos de Student foi adaptado em um trabalho anterior de nosso grupo de pesquisa para estabelecer apenas o limite superior para as estimativas da classe 1 e o limite inferior para as estimativas da classe 0 de acordo com as equações 7 e 8 abaixo<sup>39</sup>:

$$Y_{\text{limite superior, classe 1}} = 1 + \text{vies}_{\text{classe 1}} + \left( t_{99,v} \text{RMSECV}_{\text{vies}} \sqrt{1 - \bar{h}_c} \right) \quad (7)$$

$$Y_{\text{limite inferior, classe 0}} = 0 + \text{vies}_{\text{classe 0}} - \left( t_{99,v} \text{RMSECV}_{\text{vies}} \sqrt{1 - \bar{h}_c} \right) \quad (8)$$

onde  $\bar{h}_c$  é o valor médio da influência (do inglês *leverage*) observada nas amostras de treinamento;  $RMSEC_{vies}$  é o RMSECV com correção de viés;  $vies_{classe0}$  e  $vies_{classe1}$  são os vieses estimados para as classes 0 e 1, respectivamente;  $t_{99,v}$  é o valor tabelado da distribuição de t-Student, com 99% de confiança e  $n-A-2$  graus de liberdade e  $n$  é o número de amostras de treinamento.

Devido a esta análise mais criteriosa para a identificação de *outliers*, o modelo PLS-DA adquire algumas características mais flexíveis, sendo capaz de identificar amostras que não pertencem a nenhuma das classes pré-definidas na fase de treinamento, criando uma nova classe de amostras indefinidas. Essa abordagem de classificação é chamada PLS-DA flexível (do inglês *soft PLS-DA*) e, ao contrário do PLS-DA rígido (do inglês *hard PLS-DA*), pode tanto atribuir uma amostra a várias classes, como detectar amostras que não são membros de quaisquer classes de treinamento. Isso nos permite reduzir o número de falsos positivos na presença de *outliers* nos dados. O PLS-DA flexível foi recentemente introduzido e, em alguns casos, pode ser uma ferramenta mais apropriada para discriminação<sup>71-73</sup>.

Assim, com essa característica de flexibilidade, há a possibilidade adotar critérios mais específicos para classificação da amostra além da utilização dos limiares de discriminação (do inglês, *discrimination threshold*). No presente trabalho, novos critérios foram propostos para classificação de amostras. Estes podem ser especialmente utilizados quando existem replicatas disponíveis, conforme será detalhado mais adiante na seção 3.1.5.

### 2.3.3. Figuras de Mérito

A validação dos modelos de discriminação foi avaliada de acordo com o cálculo das figuras de mérito<sup>74</sup>. Uma forma de avaliação da eficiência do modelo é dada pelo chamado coeficiente de correlação de Matthews<sup>75</sup>, que leva em consideração todos os quatro resultados possíveis:

- Verdadeiros positivos (VP): amostras pertencentes à classe sendo discriminada são corretamente classificadas como pertencentes;
- Verdadeiros negativos (VN): amostras não-pertencentes à classe sendo discriminada são corretamente classificadas como não-pertencentes;

- Falsos negativos (FN): amostras pertencentes à classe sendo discriminada são incorretamente classificadas como não-pertencentes;
- Falsos positivos (FP): amostras não-pertencentes à classe sendo discriminada são incorretamente classificadas como pertencentes.

Assim, a taxa de eficiência (TEF) dos modelos de discriminação pode ser representada pela equação 9:

$$TEF = \sqrt{\frac{VP \times VN}{(VP + FN) \times (VN + FP)}} \times 100 \quad (9)$$

Por fim, a taxa de amostras com resultado indefinido (TINDEF) é calculada como sendo a relação entre a soma dos *outliers* e das amostras com resultados de análise inconclusiva (*incon*), isto é, que não foram classificadas dentro das classes propostas, e o total de amostras testadas (*I*) multiplicada por 100, representada pela equação 10:

$$TINDEF = \frac{outliers + incon}{I} \times 100 \quad (10)$$

### 3. MATERIAIS E MÉTODOS

A parte experimental do presente trabalho foi realizada no Laboratório de Produtos Florestais (LPF) do Serviço Florestal Brasileiro, pertencente ao Ministério da Agricultura, Pecuária e Abastecimento, em colaboração com o Laboratório de Automação, Quimiometria e Química Ambiental (AQQUA) do Instituto de Química da Universidade de Brasília (UnB).

A parte experimental do trabalho se deu em duas etapas. A primeira tinha como finalidade criar um conjunto de dados com espectros NIR de amostras de madeira das cinco espécies florestais estudadas antes e depois de passar por um processo de secagem. Já na segunda etapa, o objetivo foi aplicar um tratamento para aumentar o teor de umidade de um grupo de amostras das cinco espécies. Assim como na parte I, as amostras passaram novamente pelo processo de secagem a fim de criar um conjunto de dados com seus espectros NIR.

#### 3.1. Parte I: Avaliação da eficiência do EPO e criação de modelos de predição

Nesta primeira parte do estudo, as amostras de madeira selecionadas tiveram seu teor de umidade e espectro NIR obtidos antes e depois de passar por um processo de secagem. Esses espectros foram utilizados para treinar e validar modelos de predição.

##### 3.1.1. Seleção e preparo de amostras

As amostras empregadas para a criação e validação do modelo já haviam sido utilizadas em estudos anteriores da parceria LPF/AQQUA e estavam armazenadas em uma sala com controle de umidade do LPF. Estas são pertencentes de cinco espécies florestais, selecionadas com base no livro “Madeiras similares ao mogno (*Swietenia macrophylla*): uma chave ilustrada para identificação anatômica em campo” editado pelo Serviço Florestal Brasileiro<sup>76</sup> e são elas: *Carapa guianensis* Aubl. (andiroba), *Cedrela odorata* L. (cedro), *Erismia uncinatum* Warm. (cedrinho), *Micropholis melinoniana* Pierre (curupixá) e *Swietenia macrophylla* King (mogno). Estas foram escolhidas por serem espécies nativas brasileiras de madeira anatomicamente similares e de difícil distinção até mesmo por anatomistas de madeira muito especializados.

O estudo foi realizado com um total de 30 amostras para quatro espécies (cedro, cedrinho, curupixá e mogno) e 26 amostras para a espécie andiroba. O número reduzido de amostras se deve ao fato de ter sido necessário selecionar entre as amostras existentes aquelas que puderam ser cortadas em tamanho apropriado para a realização do

condicionamento de umidade. Além disso, deve-se considerar que, com exceção das espécies cedrinho e curupixá, as demais apresentam restrições de comércio no Brasil<sup>77</sup> devido ao seu risco de extinção, sendo difíceis de serem encontradas no mercado madeireiro.

Todas as amostras foram cortadas em placas retangulares de tamanho (8x5x1cm), orientadas de forma que a superfície maior (8x5cm) estivesse nos sentidos de crescimento da madeira na tangencial ou radial. Em seguida, foram cuidadosamente lixadas com lixa para madeira nº 80 por cerca de 5 segundos, em sentido único, e limpas com um pincel. Tal granulação permite uma maior superfície de contato entre a madeira e a radiação IR incidente, mantém sua uniformidade granulométrica e remove a camada externa superficial oxidada. Além disso, evita a formação de brilho na superfície da madeira e, por consequência, diminui a componente de reflectância especular. Isto é, evita que a superfície da amostra reflita a radiação no mesmo ângulo de incidência e tenta garantir que os espectros sejam obtidos em reflectância difusa (quando a superfície da amostra reflete a luz incidente em diversos ângulos)<sup>39</sup>.

### 3.1.2. Obtenção do teor de umidade

Antes de obter os espectros NIR, a superfície lixada era limpa com um pincel para remover o pó de madeira formado durante o lixamento e se media a umidade da amostra com o aparelho medidor de umidade para madeiras DL 2000 da Digisystem™, exibido na figura 4 abaixo:



**Figura 4)** Aparelho medidor de umidade para madeiras DL 2000, Digisystem™.  
(Fonte: Digisystem<sup>78</sup>)

O aparelho é constituído por um eletrodo com quatro agulhas montadas em um cabo com condutor elétrico ligado ao painel medidor. O medidor lê a umidade na área imediatamente em contato com as agulhas do eletrodo quando essas são cravadas na madeira, indicando o ponto mais úmido encontrado na profundidade percorrida. Para as amostras utilizadas neste trabalho, foi utilizada uma profundidade de aproximadamente 3 milímetros. A leitura é instantânea e localizada, sendo um método não destrutivo.

### 3.1.3. Medidas NIR

Os espectros das amostras foram obtidos utilizando o espectrofotômetro microPHAZIR™ RX Analyzer, da ThermoFisher Scientific™, como exibido na figura 5. O equipamento era disposto em uma superfície e as amostras posicionadas na saída do feixe de radiação. Para cada amostra, foram obtidos três espectros em pontos distintos e aleatórios. Por conta da heterogeneidade natural presente nas amostras de madeira e visando ampliar a representatividade do conjunto de dados, não foi calculado o espectro médio das replicadas, sendo o conjunto de três espectros de cada amostra utilizado na criação e validação dos modelos.



**Figura 5)** Espectrofotômetro microPHAZIR™ RX Analyzer, da ThermoFisher Scientific™. (Fonte: ThermoFisher Scientific<sup>79</sup>)

### 3.1.4. Condicionamento

Os espectros das amostras de madeira foram obtidos em duas condições: com a umidade ambiente e após passar por um processo de secagem em estufa. Neste processo, as amostras eram dispostas em grupos de 10 unidades em uma grade que permitisse a circulação de ar em todas as faces da madeira, e secas por 20 minutos na estufa de secagem com circulação e renovação de ar, da marca Marconi™ (modelo MA035/5/10P), pré-aquecida a 40°C, como ilustrado na figura 6. Tal intervalo de tempo e temperatura foram escolhidos de forma que a secagem não danificasse a madeira formando rachaduras, fendas ou trincas.



**Figura 6)** Amostras de madeira na estufa para processo de secagem.

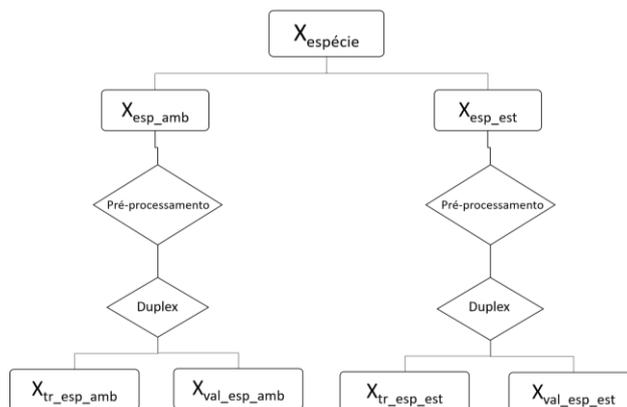
Ao serem retiradas da estufa, as amostras eram colocadas em pares dentro de placas de Petri grandes e resfriadas com o auxílio do jato de ar frio por 10 minutos. A placa de Petri foi utilizada para minimizar a reabsorção de umidade por exposição das amostras ao ambiente. Por fim, cada amostra tinha o teor de umidade medido e o espectro NIR obtido.

### 3.1.5. Análise multivariada

Todos os cálculos deste trabalho foram realizados no programa MATLAB versão 9.2.0.538062 (R2017a) com pacote PLS Toolbox 8.7 (R8.7) e programas desenvolvidos em nosso grupo de pesquisa para o cálculo do EPO, identificação de *outliers* e estimativa das figuras de mérito.

#### 3.1.5.1. Organização das matrizes de dados e seleção de amostras

Foram montadas matrizes para cada uma das espécies ( $X_{\text{espécie}}$ ), sendo que em cada espécie os espectros foram ainda separados em matrizes distintas referentes às amostras que estavam condicionadas à umidade ambiente ( $X_{\text{esp\_amb}}$ ) e às amostras após passarem pelo processo de secagem ( $X_{\text{esp\_est}}$ ). Essas novas matrizes foram submetidas ao pré-processamento definido e, em seguida, ao algoritmo Duplex para definir as amostras destinadas à matriz de treinamento para as duas condições de umidade ( $X_{\text{tr\_esp\_amb}}$  e  $X_{\text{tr\_esp\_est}}$ ) e as destinadas à de validação ( $X_{\text{val\_esp\_amb}}$  e  $X_{\text{val\_esp\_est}}$ ) num total de 20:10, respectivamente, para as espécies cedro, cedrinho, curupixá e mogno, e num total de 17:9 para a espécie andiroba. Essa separação de matrizes pode ser observada na figura 7, onde (esp) se refere a cada uma das espécies estudadas, (amb) se refere às amostras que estavam em umidade ambiente, (est) às amostras que passaram pelo processo de secagem na estufa, (tr) às amostras de treinamento e (val) às de validação.



**Figura 7)** Processo de separação de matrizes por espécie. (esp) cada uma das espécies estudadas; (amb) amostras que estavam em umidade ambiente; (est) amostras que passaram pelo processo de secagem na estufa; (tr) amostras de treinamento; (val) amostras de validação.

As matrizes de treinamento e validação para as duas condições de umidade  $\mathbf{X}_{tr\_amb}$ ,  $\mathbf{X}_{tr\_est}$ ,  $\mathbf{X}_{val\_amb}$  e  $\mathbf{X}_{val\_est}$ , respectivamente, foram formadas a partir da união das matrizes de cada espécie, como mostram as equações de 11 a 14 abaixo, onde (AN) se refere à espécie andiroba, (CE) à espécie cedro, (CH) à espécie cedrinho, (CX) à espécie curupixá e (M) à espécie mogno.

$$\mathbf{X}_{tr\_amb} = [\mathbf{X}_{tr\_AN\_amb} \quad \mathbf{X}_{tr\_CE\_amb} \quad \mathbf{X}_{tr\_CH\_amb} \quad \mathbf{X}_{tr\_CX\_amb} \quad \mathbf{X}_{tr\_M\_amb}] \quad (11)$$

$$\mathbf{X}_{tr\_est} = [\mathbf{X}_{tr\_AN\_est} \quad \mathbf{X}_{tr\_CE\_est} \quad \mathbf{X}_{tr\_CH\_est} \quad \mathbf{X}_{tr\_CX\_est} \quad \mathbf{X}_{tr\_M\_est}] \quad (12)$$

$$\mathbf{X}_{val\_amb} = [\mathbf{X}_{val\_AN\_amb} \quad \mathbf{X}_{val\_CE\_amb} \quad \mathbf{X}_{val\_CH\_amb} \quad \mathbf{X}_{val\_CX\_amb} \quad \mathbf{X}_{val\_M\_amb}] \quad (13)$$

$$\mathbf{X}_{val\_est} = [\mathbf{X}_{val\_AN\_est} \quad \mathbf{X}_{val\_CE\_est} \quad \mathbf{X}_{val\_CH\_est} \quad \mathbf{X}_{val\_CX\_est} \quad \mathbf{X}_{val\_M\_est}] \quad (14)$$

A partir disso, um conjunto para a criação de um modelo robusto (rob) foi criado contendo os espectros de ambas condições e novas matrizes de treinamento ( $\mathbf{X}_{tr\_rob}$ ) e validação ( $\mathbf{X}_{val\_rob}$ ). Esse processo está explicitado nas equações 15 e 16:

$$\mathbf{X}_{tr\_rob} = [\mathbf{X}_{tr\_amb} \quad \mathbf{X}_{tr\_est}] \quad (15)$$

$$\mathbf{X}_{val\_rob} = [\mathbf{X}_{val\_amb} \quad \mathbf{X}_{val\_est}] \quad (16)$$

Há, agora, três condições de umidade para serem analisadas denominadas: (i) **ambiente**, que contém os espectros das amostras que estão em equilíbrio com o teor de umidade ambiente; (ii) **estufa**, com os espectros das mesmas amostras após passarem pelo processo de secagem; e (iii) **robusto**, que nada mais é que a fusão das matrizes das duas condições anteriores.

### 3.1.5.2. Correção da umidade utilizando EPO

Conforme descrito da seção 2.3.1 da fundamentação teórica, para realização do EPO é necessária uma matriz diferença dos espectros na condição úmida e na condição seca. Para calcular essa matriz, foi usada a diferença entre as matrizes de treinamento dos espectros na condição ambiente e na condição estufa, conforme equação 17 abaixo, o que resulta numa matriz  $\mathbf{D}_0$ , que contém a informação referente à umidade.

$$\mathbf{D}_0 = \mathbf{X}_{tr\_amb} - \mathbf{X}_{tr\_est} \quad (17)$$

Uma análise PCA preliminar foi feita nesta matriz  $\mathbf{D}_0$  para identificação e exclusão de *outliers* pelos critérios de Q e T<sup>2</sup> de *Hotteling*. Os espectros remanescentes deram origem à matriz  $\mathbf{D}$ .

Para estimar o melhor número de componentes  $c$  para correção do EPO, os dados da matriz de treinamento na condição ambiente foram corrigidos com os pesos obtidos pela matriz  $\mathbf{D}$  e o erro do modelo foi calculado predizendo os dados da matriz de treinamento robusta, isso é, criada pela junção dos dados de treinamento das condições ambiente e estufa. Em seguida, foi feito um gráfico de curvas de nível de número de variáveis latentes do PLS-DA *versus* número de componentes do EPO pelo valor do erro do modelo. Na sequência, esse cálculo foi feito também construindo o modelo com os dados corrigidos por EPO da matriz de treinamento na condição estufa e predizendo os dados da matriz de treinamento robusta. A partir destes gráficos, foi estimado o número de componentes a ser utilizado no EPO.

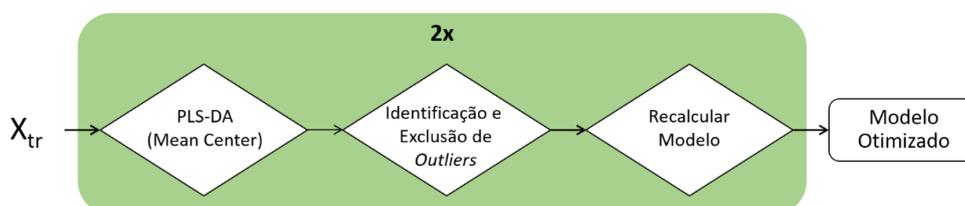
### 3.1.5.3. Criação dos modelos PLS-DA flexíveis

Os espectros obtidos foram submetidos a quatro condições de análise de dados. Todas as condições seguiram o mesmo procedimento, que será explicado nesta seção, apenas se diferenciando no pré-processamento dos espectros para utilização do EPO e para o PLS-DA, detalhado na tabela 1 abaixo. No desenvolvimento dos modelos foram avaliados os seguintes pré-processamentos: padronização normal de sinal (SNV, do inglês *standard normal variate*), primeira derivada pelo algoritmo Savitzky-Golay e centragem dos dados na média (*mean centering*).

**Tabela 1)** Detalhamento dos pré-processamentos utilizados em cada condição.

	<b>EPO</b>	<b>PLS-DA</b>
<b>Condição 1</b>	—	<i>Mean Center</i>
<b>Condição 2</b>	SNV	SNV + <i>Mean Center</i>
<b>Condição 3</b>	SNV + 1ª derivada	SNV + 1ª derivada + <i>Mean Center</i>
<b>Condição 4</b>	1ª derivada	1ª derivada + <i>Mean Center</i>

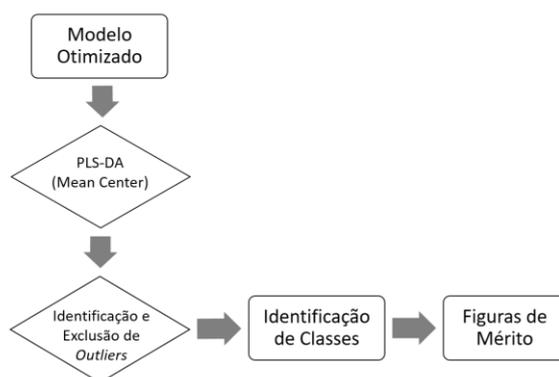
Para cada uma das condições, foi criado um modelo de treinamento a partir da análise PLS-DA, seguindo o esquema apresentado na figura 8:



**Figura 8)** Esquema para criação do modelo de treinamento, considerando a condição em que apenas o pré-processamento *Mean Center* foi utilizado.

Nesta etapa, após o PLS-DA ser aplicado na matriz de treinamento, o primeiro grupo de *outliers* foi identificado e excluído. Em seguida, o PLS-DA foi novamente aplicado na matriz, agora sem o primeiro grupo de *outliers*, e o segundo grupo foi identificado e excluído. Com a matriz de treinamento sem os dois grupos de *outliers*, o modelo PLS-DA otimizado foi criado.

A partir daí, foram feitas previsões em seis situações distintas, utilizando os três modelos de treinamento calculados anteriormente (condições “ambiente”, “estufa” e “robusto”) e as duas matrizes de validação (condições “ambiente” e “estufa”). O esquema para validação dos modelos, bem como as combinações para as previsões estão dispostos na figura 9 e na tabela 2, respectivamente:



**Figura 9)** Esquema para a análise dos conjuntos de validação no modelo de predição otimizado.

**Tabela 2)** Combinações avaliadas para a validação de cada modelo de predição construído.

<b>Condição do Modelo de Treinamento</b>	<b>Condição da Matriz de Validação</b>
Ambiente	Ambiente
Ambiente	Estufa
Estufa	Estufa
Estufa	Ambiente
Robusto	Ambiente
Robusto	Estufa

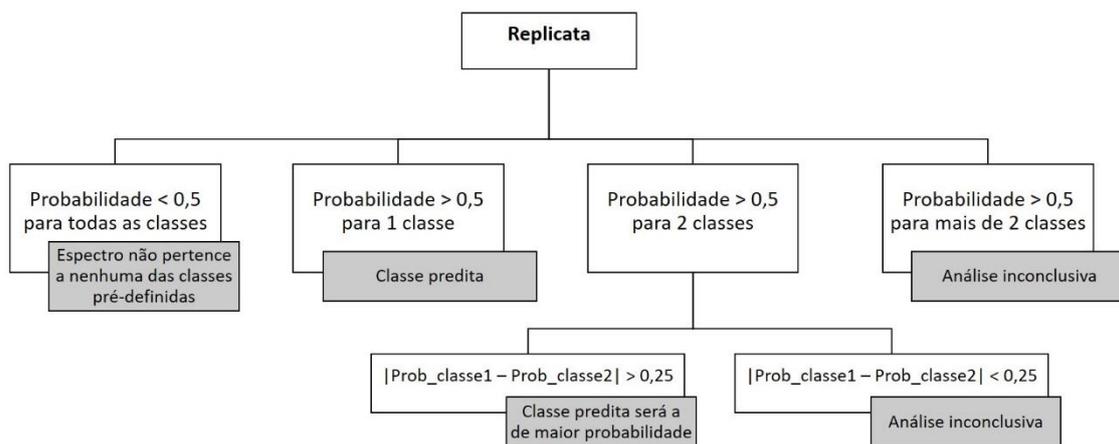
A identificação das classes foi feita através de um programa criado nessa dissertação que sistematiza os critérios para decidir a qual classe cada amostra pertence. Nesse programa, quando o modelo é criado, todos os espectros de cada amostra analisada

em triplicata têm sua probabilidade de pertencer a cada classe calculada. Esta probabilidade tem valores entre 0 e 1 e seguem o seguinte critério de avaliação:

- Probabilidade  $> 0,5$ : espectro pertencente à classe da espécie identificada;
- Probabilidade  $< 0,5$ : espectro pertence a outras espécies, diferentes daquela discriminada pelo modelo (análise inconclusiva).

Para o conjunto de validação, foram consideradas amostras anômalas em relação à estimativa dos valores de classe aquelas que apresentaram valor de  $y$  maior do que  $Y_{\text{limite superior, classe 1}}$  ou inferior que  $Y_{\text{limite inferior, classe 0}}$  (equações 7 e 8). Ou seja, uma amostra de validação  $i$  que apresentar valor de classe significativamente mais elevado do que os valores de  $y$  estimados para as amostras da classe 1 ou significativamente menor do que os valores estimados para classe 0 do conjunto de treinamento, tendo em conta o nível de confiança de 99%, serão consideradas *outliers*.<sup>39</sup>

A probabilidade de classe é analisada para cada uma das cinco espécies e pode apresentar diferentes resultados, como expostos na figura 10, que exibe um fluxograma para identificação da classe de cada replicata da amostra:



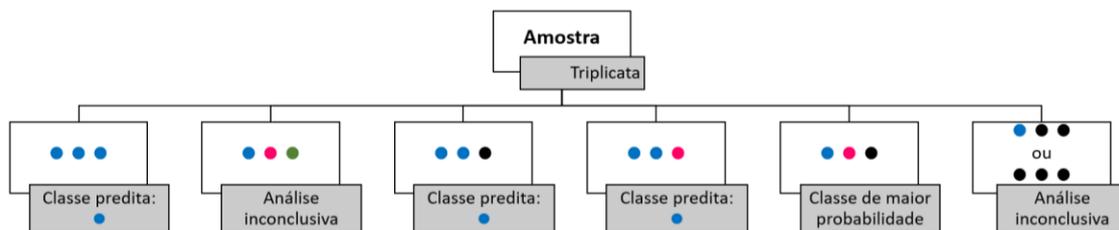
**Figura 10)** Fluxograma para identificação da classe de cada replicata.

Os critérios acima classificam cada espectro individual, porém, cada amostra será classificada de acordo com os resultados de suas triplicatas. Para isso, utilizou-se os seguintes critérios:

- Os três espectros apresentam a mesma classe predita: esta será a classe predita para a amostra;
- Os três espectros apresentam classes preditas diferentes entre si: análise inconclusiva;

- Dois espectros apresentam a mesma classe predita e um apresenta classe inconclusiva: a classe com dois resultados concordantes será a classe predita para a amostra;
- Dois espectros apresentam a mesma classe predita e um apresenta classe predita diferente: a classe com dois resultados concordantes será a classe predita para a amostra;
- Dois espectros apresentam classes previstas diferentes entre si e um apresenta classe inconclusiva: a classe predita será aquela que apresenta maior probabilidade.
- Dois espectros apresentam classe inconclusiva e um apresenta classe predita ou os três espectros da triplicata apresentam classe inconclusiva: análise inconclusiva.

A figura 11 exibe o esquema que ilustra os critérios descritos acima para identificação da classe da amostra:



**Figura 11)** Fluxograma para identificação da classe por amostra.

**Legenda:** (●) classe 1; (●) classe 2; (●) classe 3; (●) classe inconclusiva.

A partir destes resultados de predição, as figuras de mérito dos modelos foram calculadas de acordo com as equações 9 e 10, apresentadas na seção 2.3.3.

Todos os resultados foram obtidos seguindo o mesmo procedimento para os espectros antes e após passarem pela correção por EPO, como descrito na seção 2.3.1.

### 3.2. Parte II: Avaliação dos modelos de predição criados

Na segunda parte do estudo, uma fração das amostras de madeira selecionadas anteriormente recebeu um tratamento para aumentar seu teor de umidade e, em seguida, passou pelo mesmo processo descrito na parte I para medir o teor de umidade e obter os espectros antes e depois de passar por um processo de secagem. Esses espectros formaram

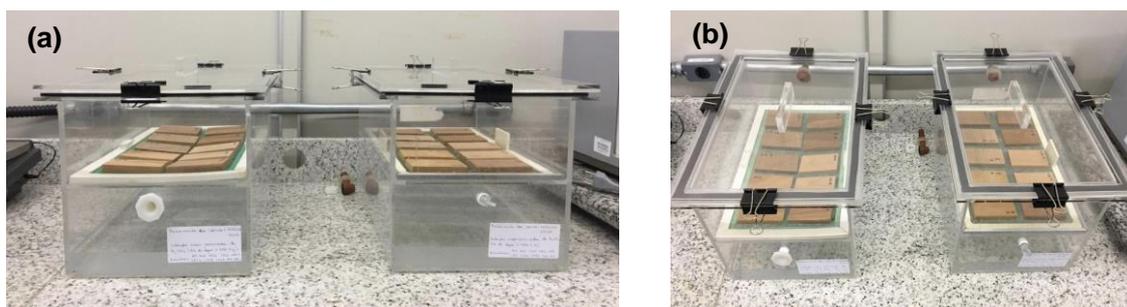
um conjunto de dados com amostras de teor de umidade mais elevados a fim de testar os modelos criados.

### 3.2.1. Seleção de amostras

Devido às mesmas dificuldades relacionadas à obtenção de amostras de madeira descritas na parte I do trabalho, foi utilizada uma fração das mesmas amostras utilizadas anteriormente. Além disso, apenas duas caixas umidificadoras estavam disponíveis, limitando o espaço para disposição das placas de madeira. Assim, inicialmente, o estudo foi realizado com um total de 4 amostras para cada espécie (andiroba, cedro, cedrinho, curupixá e mogno).

### 3.2.2. Tratamento umidificador

Para aumentar o teor de umidade das amostras de madeira, duas caixas umidificadoras foram montadas<sup>80</sup>. Tais caixas consistem em dois reservatórios acrílico retangulares com anteparos para disposição das amostras (de modo que permita a circulação de ar por toda a caixa) e tampas. Cada caixa foi preenchida com cerca de 3 litros de solução salina saturada de sulfato de potássio ( $K_2SO_4$ ), preparada reconhecendo que a solubilidade determinada de  $K_2SO_4$  a  $20^\circ C$  e com água como solvente é igual a  $111\text{ g/L}$ <sup>81</sup>. Como o mencionado na seção 2.1.3, o uso de soluções salinas saturadas em recipientes pequenos fechados propiciam um ambiente com umidade relativa do ar controlada, já que essas soluções permanecem em equilíbrio com uma determinada pressão parcial de vapor d'água, de maneira a adsorver ou liberar grandes quantidades de água sem alterar o teor de umidade de equilíbrio<sup>31,32,82</sup>. Seguindo esse princípio, a solução saturada de  $K_2SO_4$ , quando mantida em recipientes pequenos fechados a temperatura de  $25^\circ C$ , deve promover um ambiente controlado com umidade relativa do ar de  $97,5\%$ <sup>31</sup>. Em seguida, duas amostras de cada espécie foram postas no anteparo e a caixa foi vedada hermeticamente, como ilustrado na figura 12 (a) e (b).



**Figura 12) (a) e (b)** Amostras dispostas nas caixas umidificadoras com teor de umidade interno de  $97,5\%$ .

Para aferir o valor de umidade das amostras gravimetricamente, cada amostra foi pesada antes de ser introduzida na caixa e após a sua retirada. Além disso, também tiveram seu teor de umidade aferido pelo aparelho medidor de umidade e foram estes os valores que utilizados durante este trabalho.

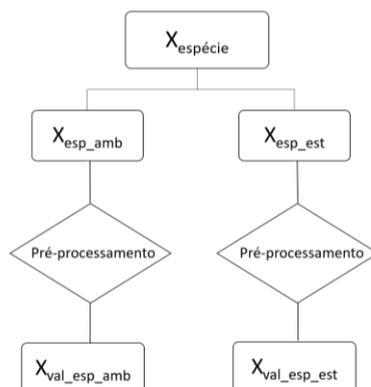
As caixas permaneceram fechadas pelo período de 35 dias, quando tiveram que ser abertas devido a proliferação de um fungo nas amostras de curupixá. Para as caixas serem limpas e a solução higienizada, as amostras foram armazenadas em sacos hermeticamente fechados por 24 horas e não houve diferença significativa em seu teor de umidade neste período. Devido ao risco de nova proliferação, as amostras de curupixá foram retiradas do experimento. As amostras das demais espécies voltaram à caixa por mais 19 dias e, quando retiradas, foram pesadas e tiveram seu teor de umidade aferido antes de continuar com o procedimento.

### 3.2.3. Teor de Umidade, Medidas NIR e Processo de Secagem

Assim como na parte I, as amostras de madeira tiveram seu teor de umidade e espectro NIR obtidos antes e depois de passar por um processo de secagem. Todos os processos foram realizados de maneira idêntica à detalhada nas seções 3.1.2. a 3.1.4.

### 3.2.4. Análise Multivariada

Os espectros obtidos antes e após o processo de secagem passaram pela etapa de separação de matrizes e pré-processamento de maneira semelhante à da parte I do trabalho. Como já estabelecido na seção 3.1.3, a condição de pré-processamento 3 foi aplicada (tabela 1, pág. 25), entretanto, visto que os modelos a serem estudados já haviam sido criados, todos os espectros foram utilizados apenas na etapa de validação dos modelos. Esse processo está exemplificado na figura 13:



**Figura 13)** Processo de separação de matrizes e pré-processamento por espécie. (*esp*) cada uma das espécies estudadas; (*amb*) amostras que estavam em umidade ambiente; (*est*) amostras que passaram pelo processo de secagem na estufa; (*val*) amostras de validação.

A partir daí, duas matrizes referentes às condições de umidade ambiente (amb) e estufa (est) foram formadas da união das matrizes de cada espécie, como mostram as equações de 18 e 19, onde (AN) se refere à espécie andiroba, (CE) à espécie cedro, (CH) à espécie cedrinho e (M) à espécie mogno.

$$\mathbf{X}_{\text{caixa\_amb}} = [\mathbf{X}_{\text{val\_AN\_amb}} \quad \mathbf{X}_{\text{val\_CE\_amb}} \quad \mathbf{X}_{\text{val\_CH\_amb}} \quad \mathbf{X}_{\text{val\_M\_amb}}] \quad (18)$$

$$\mathbf{X}_{\text{caixa\_est}} = [\mathbf{X}_{\text{val\_AN\_est}} \quad \mathbf{X}_{\text{val\_CE\_est}} \quad \mathbf{X}_{\text{val\_CH\_est}} \quad \mathbf{X}_{\text{val\_M\_est}}] \quad (19)$$

O EPO foi aplicado aos dois grupos de espectros e, então, quatro matrizes de validação foram criadas, como detalhado na tabela 3: (1) com os espectros das amostras da caixa antes do processo de secagem apenas com pré-processamento; (2) com os espectros das amostras da caixa após o processo de secagem também apenas com pré-processamento; (3) com os espectros das amostras da caixa antes do processo de secagem com pré-processamento e EPO; e (4) com os espectros das amostras da caixa após o processo de secagem também com pré-processamento e EPO.

**Tabela 3)** Construção das matrizes de validação das amostras com maior teor de umidade.

	<b>Amostras caixa sem secagem</b>	<b>Amostras caixa pós-secagem</b>
<b>Pré-processamento</b>	Matriz 1	Matriz 2
<b>Pré-processamento + EPO</b>	Matriz 3	Matriz 4

Por fim, foram feitas predições em oito situações distintas, utilizando as quatro matrizes criadas e quatro modelos de discriminação calculados na parte I. Esses modelos são os da condição ambiente e condição estufa calculados a partir do pré-processamento definido como sendo o mais eficiente, com e sem a aplicação do EPO. As combinações entre modelo e matriz para as discriminações estão dispostas na tabela 4 e foram calculadas de maneira análoga a detalhada na figura 9 da parte I.

**Tabela 4)** Combinações para a discriminação das amostras com maior teor de umidade.

<b>Condição do Modelo</b>	<b>Matriz de Validação</b>
Ambiente	Matriz 1
Ambiente	Matriz 2
Estufa	Matriz 1
Estufa	Matriz 2
Ambiente + EPO	Matriz 3
Ambiente + EPO	Matriz 4
Estufa + EPO	Matriz 3
Estufa + EPO	Matriz 4

A identificação das classes foi feita com o programa criado nessa dissertação (detalhado na seção 3.1.5.) e as figuras de mérito dos modelos de discriminação foram calculadas de maneira idêntica à da parte I.

## 4. RESULTADOS E DISCUSSÃO

Os resultados desse trabalho serão apresentados em duas partes. A primeira tinha como finalidade avaliar a utilização do EPO para correção da interferência da umidade nos espectros através da criação de um modelo de predição. Foram testadas diferentes condições de pré-processamento dos espectros e se estudou a melhor combinação entre os conjuntos de dados para os modelos de treinamento e validação. Isso foi feito a fim de encontrar a condição mais eficaz independente da umidade a qual a amostra se encontra. Já na segunda, foram utilizados os modelos determinados como sendo os mais eficientes na parte I para avaliar um novo conjunto de dados. Esse novo conjunto foi formado por uma fração das amostras utilizadas anteriormente, porém, após passar por um tratamento que aumentou seu teor de umidade.

### 4.1. Parte I: Avaliação da eficiência do EPO e criação de modelos de predição

#### 4.1.1. Avaliação preliminar via PCA da correção do efeito da umidade por EPO em amostras de madeira

As amostras utilizadas neste trabalho se encontravam em condição seca, com teor de umidade em equilíbrio com a umidade relativa do ar ambiente e em ambiente com temperatura controlada a 25°C. Após o processo de secagem descrito na seção 3.1.4, todas as espécies apresentaram uma diminuição significativa no teor de umidade, como está apresentado na tabela 5:

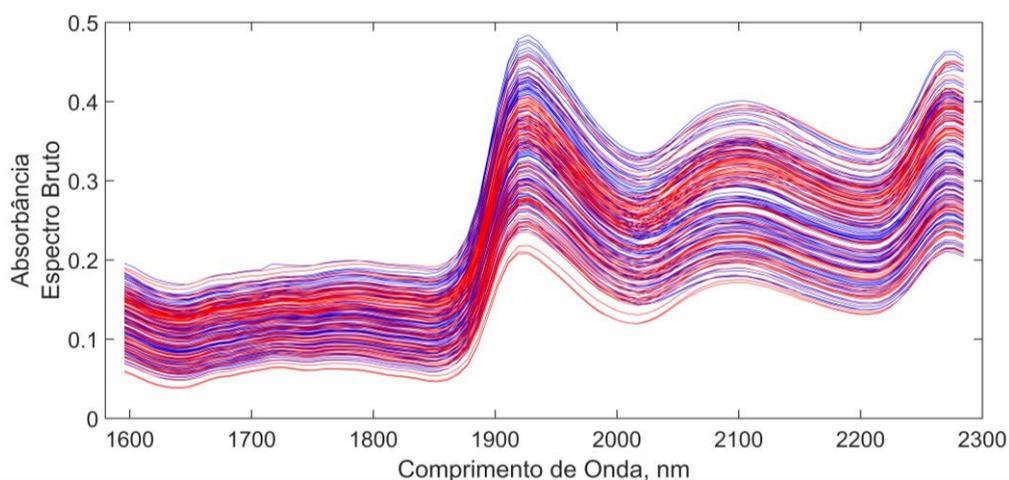
**Tabela 5)** Média e desvio padrão (s) do teor de umidade das amostras por espécie inicial e após passar pelo processo de secagem.

Espécie	Teor de umidade inicial (%)	$S_{inicial}$ (%)*	Teor de umidade final (%)	$S_{final}$ (%)*
Andiroba	13,3	0,7	10,4	0,6
Cedro	9,7	1,8	3,0	2,4
Cedrinho	12,0	1,3	6,4	1,1
Curupixá	13,1	1,4	5,8	3,1
Mogno	11,1	0,6	7,0	1,7

\*n=26 para a espécie andiroba e n=30 para as espécies cedro, cedrinho, curupixá e mogno.

É importante notar que o aparelho medidor de umidade para madeiras utilizado apresenta um limite de detecção de 6%, o que torna as medidas que foram obtidas com valor inferior a esse passíveis de erro. Tal fato pode ser notado no alto valor de desvio padrão das amostras de cedro e curupixá após o processo de secagem.

Analisando os espectros referentes às amostras de curupixá (*Micropholis melinoniana* Pierre) nas condições antes e após secagem (figura 14), é possível notar três bandas principais centradas em aproximadamente 1919, 2098 e 2270 nm.



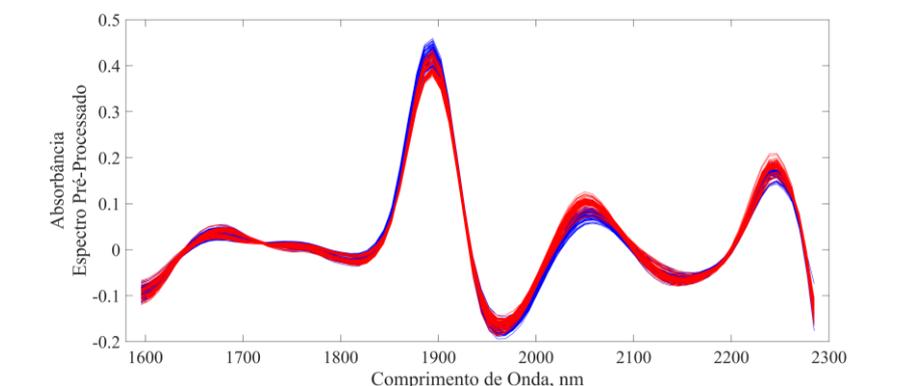
**Figura 14)** Espectros brutos referentes as amostras de curupixá (*Micropholis melinoniana* Pierre) antes (em azul) e após secagem (em vermelho).

Apesar do espectro NIR apresentar uma grande sobreposição de sinais e variação de linha de base, o que dificulta a atribuição precisa das bandas relacionadas a grupos de ligações específicos, as bandas mencionadas anteriormente são condizentes com as apresentadas na literatura<sup>83</sup> (relacionadas na tabela 6 a seguir). Elas são características de grupos OH presentes em compostos que fazem parte da constituição da madeira, podendo apresentar informação referente à umidade. A banda centrada em 1919 nm é referente à absorção do primeiro sobretom da ligação OH da água. Já a banda centrada em 2092 nm é referente a ligação OH proveniente de moléculas de celulose e a centrada em 2277 nm, se refere a ligação OH proveniente de moléculas de celulose ou hemicelulose, que formam a parede celular das células vegetais, logo, ambas não contêm informação referente à umidade.

**Tabela 6)** Atribuições de bandas características às vibrações de ligações O-H em espectros NIR de madeira<sup>83</sup>.

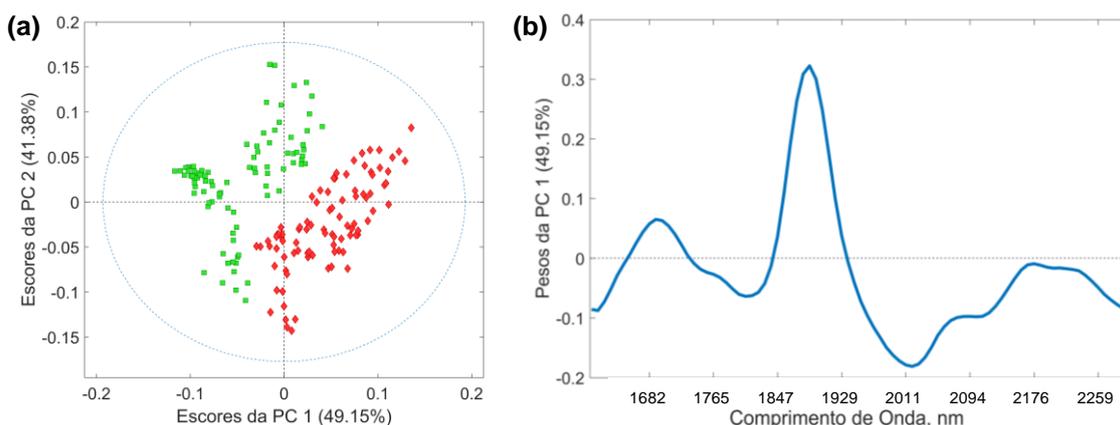
Comprimento de onda (nm)	Vibração
1916 – 1924	Ligações O-H de moléculas de água
2092	Ligações O-H provenientes de moléculas de celulose ou hemicelulose
2277	Ligações O-H provenientes de moléculas de celulose

Na figura 15, é possível observar que após o pré-processamento dos espectros por SNV e 1ª derivada, os espectros medidos na condição ambiente (em azul) apresentam uma maior intensidade que os na condição estufa (em vermelho) na banda centrada em 1919 nm. Como discutido, tal banda caracteriza a absorção das ligações OH de moléculas de água na madeira e contém informação espectral referente à umidade<sup>83</sup>.



**Figura 15)** Espectros pré-processados referentes as amostras de curupixá (*Micropholis melinoniana* Pierre) antes (em azul) e após secagem (em vermelho).

As vibrações de moléculas de água caracterizadas por essas bandas são responsáveis pela mudança no perfil do espectro NIR, mas também podem modificar propriedades físicas e mecânicas da madeira<sup>84</sup>. Uma análise PCA pode comprovar esta diferença espectral causada pela umidade, como pode ser visto na figura 16, referente aos espectros pré-processados do curupixá antes e depois do processo de secagem.

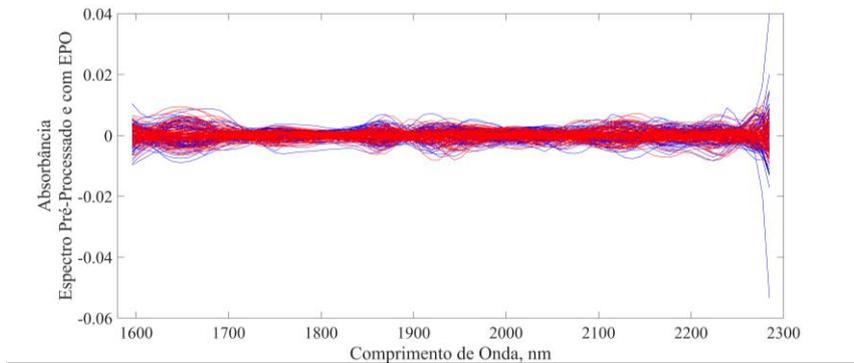


**Figura 16)** Gráficos dos (a) escores e (b) pesos com dados centrados na média para os espectros de curupixá pré-processados antes (♦) e após (■) secagem.

Na figura 16(a), observa-se que no gráfico do escores há uma divisão das amostras de uma mesma espécie em dois grupos distintos. Na figura 16(b), o gráfico dos pesos da PC1 mostra que a banda em 1919 nm é a principal responsável pela separação. A existência dessa separação entre amostras de uma mesma espécie dificulta o desenvolvimento de métodos para a identificação de espécies por NIR, uma vez que

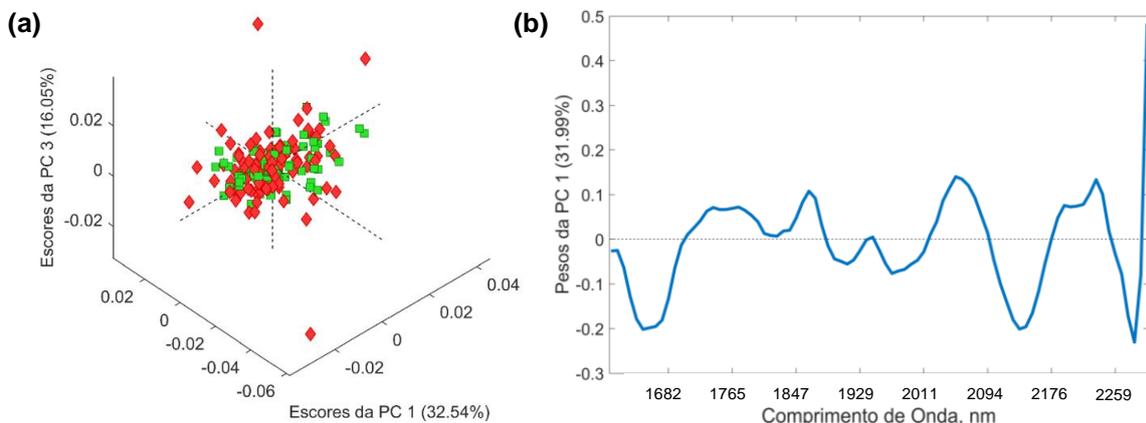
análises em condições de campo podem apresentar amostras com diferentes níveis de umidade.

Este trabalho visou corrigir esta diferença espectral por meio da ortogonalização parâmetros externos (EPO), como descrito na seção 2.3.1. Ao aplicar a correção, é possível observar uma clara mudança no perfil dos espectros pré-processados, como mostrado na figura 17:



**Figura 17)** Espectros pré-processados e corrigidos com EPO referentes as amostras de curupixá antes (em azul) e após secagem (em vermelho).

Entretanto, a principal diferença para os efeitos dessa correção é que não mais se nota a tendência de separação dos espectros das amostras com umidade ambiente das amostras que passaram pelo processo de secagem. Tal fato é ainda mais evidente ao analisar o gráfico dos escores desse novo conjunto de dados, como pode ser visto da figura 18 a seguir:



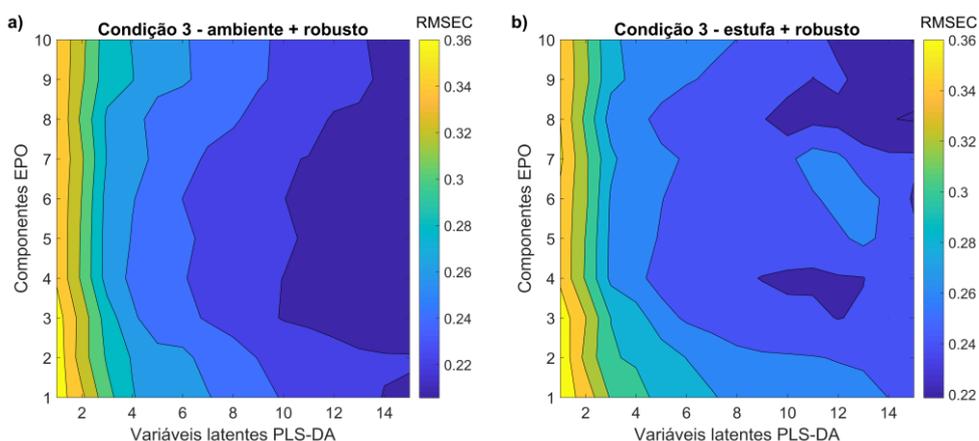
**Figura 18)** Gráficos dos (a) escores e (b) pesos com dados centrados na média para os espectros de curupixá pré-processados e corrigidos com EPO antes (♦) e após (■) secagem.

O gráfico dos escores e dos pesos dos espectros corrigidos por EPO mostra que não há mais diferença evidente entre os escores das amostras até, pelo menos, a terceira componente principal e que não há mais influência da banda em 1919 nm. Logo, este

estudo preliminar pode sugerir que a utilização do EPO é promissora para a correção do efeito da umidade em madeira, contornando o problema da diferenciação de espectros pertencentes à mesma espécie.

#### 4.1.2. Determinação dos parâmetros utilizados no EPO e no PLS-DA para o desenvolvimento do método de discriminação

A partir dos resultados promissores da correção por EPO obtidos pela análise preliminar por PCA, precisou-se, então, determinar o melhor número de componentes  $c$  a ser utilizado para modelar a informação referente à umidade nas matrizes de cada condição experimental. Este número de componentes tem que ser tal que, quando combinado ao número de variáveis latentes (VL) do PLS-DA, leve a um menor erro de predição da classe/espécie das amostras analisadas. Para isso, se utilizou os valores de RMSEC, isto é, raiz quadrada do erro quadrático médio quadrático de calibração (do inglês *Root Mean Square Error of Calibration*) como parâmetro para encontrar os melhores valores de  $c$  e VL. Quanto menor o RMSEC, menor o erro médio associado à predição do valor de classe. Para encontrar os melhores valores de  $c$  e VL, foram construídos vários modelos PLS-DA variando o valor de  $c$  entre 1 e 10, enquanto que os valores de VL variaram entre 1 e 15. Na figura 19, são apresentados os gráficos de curvas de nível de valores de RMSEC criado para escolher  $c$  e VL na condição de pré-processamento 3 da tabela 1, pág 25 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + *Mean Center*), a partir dos modelos calculados com os dados corrigidos por EPO da matriz de treinamento na condição (a) ambiente e (b) estufa, predizendo os dados da matriz de treinamento robusta (junção dos dados de treinamento nas condições ambiente e estufa):



**Figura 19)** Gráfico de curvas de nível de valores de RMSEC para modelos corrigidos por EPO da matriz de treinamento na condição (a) ambiente e (b) estufa, predizendo os dados da matriz de treinamento robusta na condição de pré-processamento 3 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + *Mean Center*).

É importante destacar que estes gráficos representam a média dos valores de RMSEC obtidos para as cinco espécies de madeira na condição de pré-processamento 3. Segundo a escala de cores do gráfico, quanto mais próxima a coloração for do azul escuro, menor é o erro de médio de estimativa do valor de classe (RMSEC).

Nos dois gráficos, é possível observar que os menores valores de RMSEC e, conseqüentemente, os menores erros de calibração, estão associados majoritariamente a valores de  $c$  maiores que 4 e a um número de variáveis latentes acima de 5. Visando obter uma condição que atendesse a todas as espécies simultaneamente, decidiu-se utilizar 4 componentes para o EPO. Quanto ao número de variáveis latentes, este foi estimado com maior detalhe por validação cruzada no momento da modelagem do PLS-DA com cada matriz de treinamento específica, buscando o melhor valor de variáveis latentes para as cinco espécies de madeira e estão dispostos na tabela 7. Conforme exemplificado pela figura 8 da seção 3.1.5.3, as matrizes de treinamento passaram por duas rodadas de identificação e exclusão de *outliers* pelos critérios de Q e T<sup>2</sup> de *Hotteling* para ter o modelo otimizado. Ao total, 3 modelos foram calculados, e, portanto, 3 valores de VL foram usados por cada matriz de treinamento. Entretanto, aquelas cujos espectros foram corrigidos por EPO apenas uma rodada de identificação e exclusão de *outliers* foi necessária, visto que após esta primeira rodada, todos os espectros reincidentes já se adequavam aos critérios de Q e T<sup>2</sup> de *Hotteling*.

**Tabela 7)** Número de variáveis latentes aplicado em cada ciclo de PLS-DA para cada condição da matriz de treinamento

<b>Condição da Matriz de Treinamento</b>	<b>VL do 1º ciclo PLS-DA</b>	<b>VL do 2º ciclo PLS-DA</b>	<b>VL do 3º ciclo PLS-DA</b>
<i>Ambiente</i>	16	13	13
<i>Estufa</i>	16	14	12
<i>Robusto</i>	12	12	12
<i>Ambiente (EPO)</i>	6	6	-
<i>Estufa (EPO)</i>	6	6	-
<i>Robusto (EPO)</i>	12	12	-

O gráfico de curvas de nível foi gerado em todas as quatro condições de pré-processamento e de umidade (ambiente, estufa e robusto) para assegurar que os parâmetros determinados estavam de acordo com cada condição estudada. Os gráficos das condições de pré-processamento 1, 2 e 4 estão dispostos na seção de Apêndices.

### 4.1.3. Escolha do pré-processamento aplicado aos espectros

As figuras de mérito foram calculadas para todos os modelos de predição propostos na tabela 2 (pág. 26) em cada uma das quatro condições de pré-processamento, relacionadas na tabela 1 (pág. 25). Os valores referentes às taxas de eficiência e de resultados indefinidos dos modelos estão dispostos nas tabelas 8 a 11, de acordo com cada condição de pré-processamento e validação.

**Tabela 8)** Resultados para os valores de taxa de eficiência, amostras indefinidas, amostras classificadas como *outliers* e como inconclusivas para todos os modelos de predição com pré-processamento 1 (EPO: bruto; PLS-DA: *Mean Center*).

Condição <i>I</i>	TEF (%)					TINDEF (%)	<i>outliers</i>	incon	emp
	AN	CE	CH	CX	M				
<i>AA</i>	100,0	100,0	100,0	100,0	100,0	22,4	11	0	0
<i>AA</i> (EPO)	100,0	100,0	98,4	100,0	100,0	16,3	8	0	0
<i>AE</i>	##	##	##	89,4	0,0	87,8	36	4	3
<i>AE</i> (EPO)	59,3	98,3	98,3	93,2	69,5	22,4	4	6	1
<i>EE</i>	100,0	100,0	100,0	100,0	100,0	14,3	6	0	1
<i>EE</i> (EPO)	100,0	100,0	100,0	100,0	100,0	8,2	2	2	0
<i>EA</i>	##	##	##	##	##	97,9	47	0	1
<i>EA</i> (EPO)	0,0	100,0	86,6	100,0	84,8	36,7	12	3	3
<i>robA</i>	84,2	0,0	0,0	0,0	0,0	44,9	13	6	3
<i>robA</i> (EPO)	100,0	100,0	100,0	100,0	100,0	24,5	11	0	1
<i>robE</i>	0,0	0,0	0,0	0,0	86,6	22,4	3	5	3
<i>robE</i> (EPO)	98,6	100,0	100,0	100,0	90,0	6,1	2	1	0

**Legenda:** AN: andiroba; CE: cedro; CH: cedrinho, CX: curupixá; M: mogno; TEF: taxa de eficiência; TINDEF: taxa de amostras indefinidas; *outliers*: nº de amostras anômalas; incon: nº de amostras com análise inconclusiva; emp: nº de amostras indefinidas com classes em empate; *AA*: ambiente + ambiente; *AE*: ambiente + estufa; *EE*: estufa + estufa; *EA*: estufa + ambiente; *robA*: robusto + ambiente; *robE*: robusto + estufa; (EPO): corrigido por EPO; ##: dados não disponíveis devido a exclusão de todas as amostras da classe como *outliers*.

A denominação disposta na primeira coluna da tabela se refere à *condição do modelo de treinamento + condição da matriz de validação* utilizados para criação e validação de cada modelo. Por exemplo, aquele que foi criado a partir do modelo de

treinamento de amostras em umidade ambiente e validado com amostras que passaram pela estufa será chamado AE. Os modelos de predição cujos dados foram corrigidos por EPO estão explicitados com a sigla entre parênteses. As colunas referentes à taxa de eficiência (TEF) de cada modelo de predição estão subdivididas entre as cinco espécies de madeira. A seguir, está disposta aquela referente à taxa de amostras indefinidas (TINDEF). As duas colunas seguintes dizem respeito ao número absoluto de amostras anômalas (*outliers*) e de amostras com classificação inconclusiva (*incon*), respectivamente, que como explicado na seção 2.3.3, dão origem à taxa de indefinidos. Como é necessário o resultado concordante de ao menos 2 replicatas para um resultado conclusivo, a última coluna diz respeito às amostras que são indefinidas, uma vez que uma de suas replicatas foi predita como pertencente a uma das classes de treinamento, outra como *outlier* e a última como inconclusiva. O mesmo se aplica às tabelas 9 a 11.

**Tabela 9)** Resultados para os valores de taxa de eficiência, amostras indefinidas, amostras classificadas como *outliers* e como inconclusivas para todos os modelos de predição com pré-processamento 2 (EPO: SNV; PLS-DA: SNV + *Mean Center*).

<b>Condição 2</b>	<b>TEF (%)</b>					<b>TINDEF (%)</b>	<i>outliers</i>	<i>incon</i>	<b>emp</b>
	<b>AN</b>	<b>CE</b>	<b>CH</b>	<b>CX</b>	<b>M</b>				
<b>AA</b>	100,0	100,0	95,9	88,2	100,0	30,6	14	0	1
<b>AA (EPO)</b>	91,3	100,0	98,3	94,3	98,5	18,4	5	4	0
<b>AE</b>	##	100,0	##	57,7	##	91,8	44	0	1
<b>AE (EPO)</b>	88,4	100,0	96,4	89,4	69,7	22,4	5	4	2
<b>EE</b>	100,0	100,0	100,0	100,0	100,0	34,7	16	0	1
<b>EE (EPO)</b>	100,0	100,0	96,8	89,4	100,0	14,3	6	1	0
<b>EA</b>	##	100,0	100,0	100,0	##	91,8	45	0	0
<b>EA (EPO)</b>	79,0	100,0	98,3	89,4	86,7	22,4	9	2	0
<b>robA</b>	84,5	0,0	0,0	0,0	0,0	34,7	14	3	0
<b>robA (EPO)</b>	100,0	100,0	95,9	89,4	100,0	32,6	16	0	0
<b>robE</b>	0,0	0,0	0,0	0,0	89,4	26,5	8	3	2
<b>robE (EPO)</b>	100,0	100,0	95,9	88,2	100,0	28,6	14	0	0

**Tabela 10)** Resultados para os valores de taxa de eficiência, amostras indefinidas, amostras classificadas como *outliers* e como inconclusivas para todos os modelos de predição com pré-processamento 3 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + *Mean Center*).

<b>Condição 3</b>	<b>TEF (%)</b>					<b>TINDEF (%)</b>	<b>outliers</b>	<b>incon</b>	<b>emp</b>
	<b>AN</b>	<b>CE</b>	<b>CH</b>	<b>CX</b>	<b>M</b>				
<b>AA</b>	100,0	100,0	100,0	100,0	100,0	30,6	14	0	1
<b>AA (EPO)</b>	77,8	98,4	100,0	94,9	97,2	18,4	6	1	2
<b>AE</b>	##	100,0	##	100,0	##	93,9	44	0	2
<b>AE (EPO)</b>	80,3	98,1	100,0	100,0	74,9	26,5	7	2	4
<b>EE</b>	100,0	100,0	100,0	100,0	100,0	18,4	9	0	0
<b>EE (EPO)</b>	79,1	98,3	100,0	100,0	80,4	20,4	6	0	4
<b>EA</b>	##	100,0	100,0	100,0	##	79,6	38	1	0
<b>EA (EPO)</b>	84,0	98,5	100,0	94,9	79,3	14,3	5	0	2
<b>robA</b>	83,4	0,0	0,0	0,0	0,0	46,9	11	8	4
<b>robA (EPO)</b>	89,7	100,0	100,0	100,0	85,2	28,6	12	1	1
<b>robE</b>	0,0	0,0	0,0	0,0	86,1	28,6	9	1	4
<b>robE (EPO)</b>	100,0	100,0	100,0	100,0	100,0	22,4	10	1	0

**Tabela 11)** Resultados para os valores de taxa de eficiência, de amostras indefinidas, amostras classificadas como *outliers* e como inconclusivas para todos os modelos de predição com pré-processamento 4 (EPO: 1ª derivada; PLS-DA: 1ª derivada + *Mean Center*).

<b>Condição 4</b>	<b>TEF (%)</b>					<b>TINDEF (%)</b>	<b>outliers</b>	<b>incon</b>	<b>emp</b>
	<b>AN</b>	<b>CE</b>	<b>CH</b>	<b>CX</b>	<b>M</b>				
<b>AA</b>	100,0	98,3	98,2	94,3	91,3	24,5	12	0	0
<b>AA (EPO)</b>	70,7	96,6	98,3	94,3	98,4	20,4	7	2	1
<b>AE</b>	##	100,0	##	81,6	0,0	87,8	41	1	1
<b>AE (EPO)</b>	91,1	100,0	98,3	94,3	89,9	20,4	6	3	1
<b>EE</b>	100,0	100,0	100,0	100,0	100,0	16,3	8	0	0
<b>EE (EPO)</b>	88,1	100,0	100,0	100,0	80,2	28,6	9	4	1
<b>EA</b>	##	##	100,0	100,0	##	93,9	45	1	0
<b>EA (EPO)</b>	81,6	97,7	100,0	100,0	98,3	36,7	16	2	0
<b>robA</b>	85,1	23,0	0,0	0,0	0,0	30,6	12	2	1
<b>robA (EPO)</b>	100,0	98,2	98,1	94,3	89,4	26,5	12	0	1
<b>robE</b>	0,0	0,0	0,0	0,0	91,8	20,4	4	6	0
<b>robE (EPO)</b>	98,5	98,4	100,0	100,0	84,5	18,4	7	2	0

Avaliando as tabelas de 8 a 11, nota-se que as taxas de eficiência (TEF) não sofrem grandes variações com o uso do EPO. Entretanto, o que pode ser claramente notado é a diminuição da taxa de amostras com resultado indefinidos (TINDEF), como no modelo AE da condição de pré-processamento 2, onde a taxa passou de 91,8% para 22,4% após a correção do modelo por EPO. Essa notável diminuição do TINDEF dos modelos após a correção por EPO indica uma redução significativa da interferência da umidade nos espectros NIR de amostras de madeira.

Como fundamentado anteriormente pela equação 10 da seção 2.3.3, o TINDEF é calculado através da relação entre a soma dos *outliers* e das amostras com análise inconclusiva (incon), isto é, que não foram classificadas dentro das classes predefinidas no conjunto de treinamento, e o total de amostras testadas (I) multiplicada por 100. Com o detalhamento de suas componentes, foi possível observar que o número de *outliers* é significativamente maior que o número de amostras inconclusivas, principalmente nos modelos que não passaram pela correção por EPO. Por exemplo, inicialmente, o modelo EA da condição de pré-processamento 4 apresentava um total de 46 amostras indefinidas num total de 49 amostras, das quais 45 eram *outliers*. Após a correção por EPO, o modelo passou a apresentar 16 *outliers* dentro das 18 amostras indefinidas. Tal fato indica a eficácia da predição dos modelos criados, o que denota a diferença espectral causada pela umidade como sendo a principal origem das amostras indefinidas.

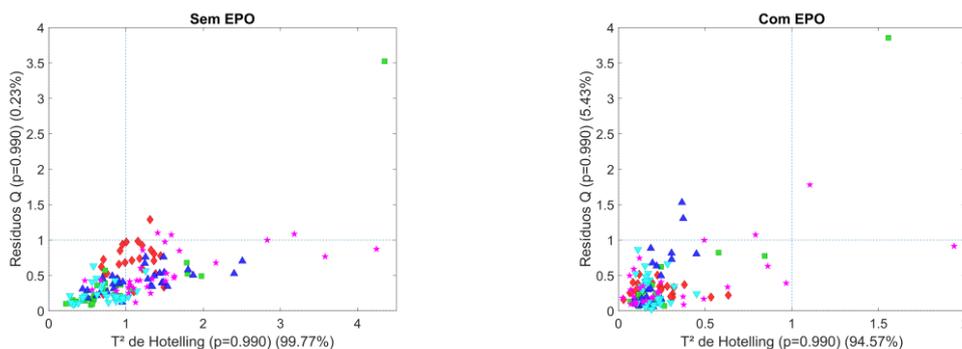
Assim, para determinar qual das quatro condições de pré-processamento apresentou um melhor resultado, os TINDEF dos modelos de cada condição foram analisados. Em especial, os resultados dos modelos AE e EA, visando a resolução do problema encontrado nos trabalhos descritos na revisão bibliográfica, que se dá na discriminação de amostras em condição de umidade diferente daquelas utilizadas na fase de treinamento. Apenas as condições 2 e 3 apresentaram valores de TINDEF abaixo dos 30% em todos os modelos após a correção por EPO, então, as condições 1 e 4 foram descartadas. A condição 3 foi preferível à condição 2 com base na melhora dos valores TINDEF dos modelos AE e EA após a correção por EPO, com valores de 26,5% e 14,3%, respectivamente, enquanto a outra apresentou valores de 22,4% nos dois modelos.

A análise dos dados foi realizada nas quatro condições de pré-processamento, descritas na tabela 1 (pág. 25), em todas as combinações entre modelo de treinamento e amostras de validação dispostos tabela 2 (pág. 26), configurando um total de 24 modelos de discriminação. Cada um deles foi testado para as cinco espécies de madeira,

totalizando 120 análises discriminatórias. Devido ao alto volume de figuras resultantes desse estudo, a seguir serão exibidos apenas os resultados provenientes da análise dos dados que passaram pelo pré-processamento da condição 3, conforme explicado anteriormente, no modelo de discriminação criado a partir das amostras de treinamento em umidade ambiente e validado com amostras que passaram pela estufa (AE). O modelo foi escolhido, como mencionado, buscando a resolução do erro na discriminação de amostras em condição de umidade diferente daquelas utilizadas na fase de treinamento.

#### 4.1.4. Resultados para o modelo de discriminação por PLS-DA utilizando conjunto de treinamento com amostras em umidade ambiente e validado com amostras que passaram pela estufa (AE)

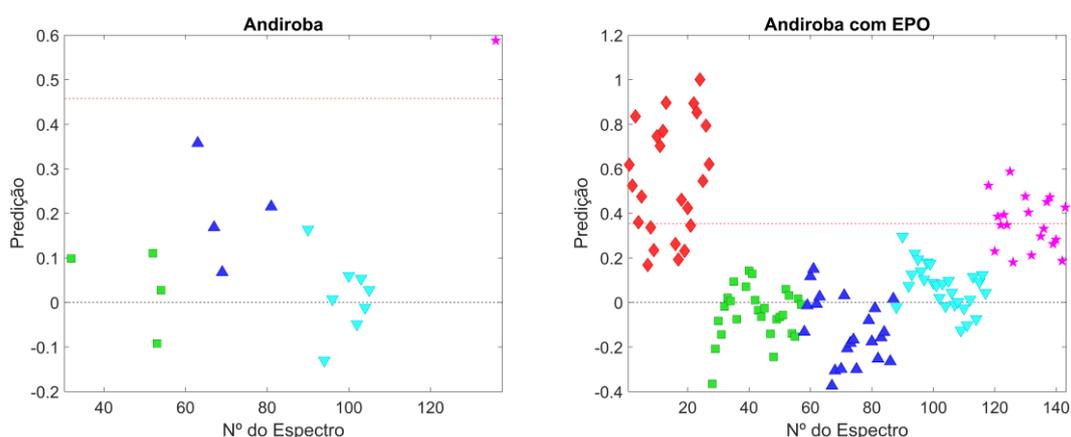
Como apresentado na seção anterior, a figura de mérito que apresentou a maior diferença na utilização do EPO como ferramenta de correção do efeito da umidade foi a taxa de amostras com resultados indefinidos. Com isso, pode-se dizer que os modelos de discriminação com EPO apresentam uma melhor representatividade que aqueles sem EPO, pois um número significativamente menor de espectros foi identificado como sendo amostras indefinidas e, portanto, excluídos. Também foi comprovado que os *outliers* são a principal origem das amostras indefinidas. O efeito da umidade modifica seu espectro em tal nível que estes não eram mais identificados como sendo compatíveis com aqueles utilizados na fase de treinamento ou pertencentes às suas respectivas espécies. Na figura 20, estão apresentados os gráficos de  $T^2$  versus resíduos Q dos modelos de discriminação referente aos espectros das amostras de mogno pré-processados pela condição 3 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + *Mean Center*) sem a correção por EPO e com a correção, respectivamente, criados pelo modelo treinado por espectros de amostras na condição ambiente e validado com a matriz de espectros de amostras na condição estufa (AE).



**Figura 20)** Gráficos  $T^2$  versus Q referentes às amostras de validação sem correção e com correção do efeito da umidade por EPO na condição de pré-processamento 3 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + *Mean Center*) no modelo AE. (♦) andiroba; (■) cedro; (▲) cedrinho; (▼) curupixá; (★) mogno.

Como se pode notar na figura 20, o gráfico referente ao modelo sem correção do efeito da umidade apresenta uma quantidade significativa de pontos com valores de  $T^2$  e de resíduos (Q) acima dos seus respectivos limites de 99% de confiança, sendo considerados *outliers* (seção 2.3.2). Já no gráfico do modelo que foi corrigido por EPO, tem-se uma diminuição significativa nesse número de amostras indefinidas.

Durante a fase de treinamento dos modelos e sua validação, é feita a identificação e exclusão dos *outliers*, de acordo com os critérios de  $T^2$  de *Hotelling*, resíduos Q e erros nos valores de classe, seguindo os esquemas das figuras 8 e 9. Entretanto, para casos semelhantes ao apresentado na figura 20 sem o uso de EPO, observou-se uma exclusão em massa de espectros, diminuindo consideravelmente o número destes na fase de validação do modelo. Como consequência dessa exclusão, pode-se observar uma taxa de eficiência aparentemente alta, mas que foi estimada com um número significativamente menor de espectros, o que não apresenta um conjunto de dados suficiente para considerá-lo representativo. Para ilustrar esse fenômeno, serão apresentadas a seguir os valores de classe estimados para o conjunto de validação sem e com a correção do efeito da umidade por EPO para a discriminação de cada espécie nas mesmas condições utilizadas anteriormente para criação dos gráficos da figura 20 (condição de pré-processamento 3, modelo AE).



**Figura 21)** Valores de classe estimados para o conjunto de validação para discriminação da espécie andiroba em relação às demais espécies antes e após correção com EPO, respectivamente, na condição de pré-processamento 3 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + *Mean Center*) e no modelo AE. (♦) andiroba; (■) cedro; (▲) cedrinho; (▼) curupixá; (★) mogno; (---) Limiar de discriminação.

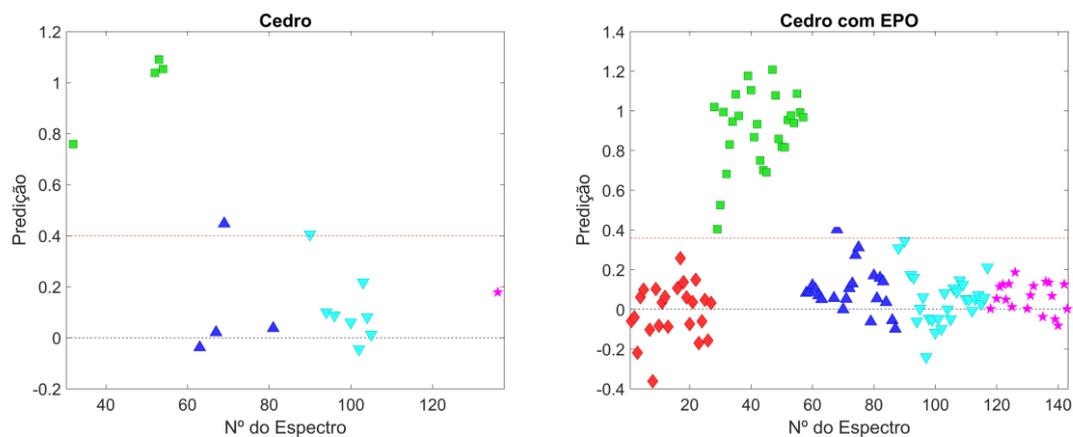
Na figura 21, estão representados os valores de classe estimados pelo modelo construído na condição de pré-processamento 3 com e sem correção do efeito da umidade por EPO. Os modelos foram criados com espectros de amostras na condição ambiente e

a matriz de validação por espectros de amostras na condição estufa. Nos gráficos, o valor de classe igual a 1 é atribuído para a espécie a ser discriminada, neste caso, a andiroba e o valor de classe igual a 0 para as demais espécies. Assim, o esperado é que apenas os valores de classe estimados para os espectros das amostras de andiroba, exibidos no gráfico pelo símbolo (♦), estejam acima do limite de discriminação, enquanto que para as demais espécies sejam obtidos valores abaixo do limiar de discriminação.

O gráfico referente ao modelo sem correção por EPO exemplifica bem o fenômeno de exclusão em massa de espectros mencionado anteriormente. Após a exclusão dos *outliers*, a classe andiroba foi completamente excluída, mostrando que o efeito da umidade tornava os espectros incompatíveis com os utilizados na fase de treinamento, a ponto de considerar as amostras com menor teor de umidade não-pertencentes a nenhuma das espécies analisadas. Assim, não restaram dados suficientes para o cálculo da taxa de eficiência do modelo para a discriminação da andiroba. Os poucos espectros que restaram das classes cedro, cedrinho e curupixá apresentaram verdadeiros negativos, espectros de amostras que não pertencem à classe andiroba, sendo classificados corretamente pelo modelo. Enquanto isso, a classe mogno apresentou para seu único espectro restante um erro falso positivo, ou seja, esse espectro da amostra não pertence à classe andiroba, mas o modelo o classifica como sendo pertencente. A taxa de resultados indefinidos do modelo foi de 93,9%, o que em números absolutos corresponde a 46 amostras classificadas como indefinidas num total de 49 amostras. Destas 46 indefinidas, 44 foram identificadas como *outliers* e 2 tiveram os espectros de sua triplicata empatados entre a classe de *outliers*, a classe de inconclusivas e uma das classes de treinamento, o que classifica a amostra como sendo indefinida, seguindo os critérios ilustrados no esquema da figura 11.

Já no gráfico referente ao modelo cujos espectros passaram pela correção por EPO, observa-se uma quantidade muito superior de espectros, indicando que estes foram considerados compatíveis com os utilizados na fase de treinamento, mesmo estando com um teor de umidade menor. Observa-se também uma discriminação aceitável das amostras de andiroba, ainda que haja uma certa mistura com a classe mogno, classificando alguns espectros desta classe como sendo andiroba. A diferenciação entre mogno e andiroba é complexa até mesmo para anatomistas experientes, já que que as semelhanças anatômicas entre eles são muitas. Também se observam falsos negativos, isto é, espectros de amostras que pertencem à classe andiroba classificados erroneamente como não

pertencentes à essa classe. Todas as amostras das outras classes foram classificadas corretamente, considerando os critérios de classificação descritos na seção 3.1.5. Com a aplicação do EPO foi obtido um modelo para andiroba nessa condição com taxa de eficiência de 80,3% e taxa de resultados indefinidos de 26,5%, com número absoluto de amostras indefinidas de 13 no total de 49 (tabela 10, pág 41). Destas 13 amostras indefinidas, 7 amostras foram classificadas como *outliers*, 2 amostras tiveram análise inconclusiva e 4 amostras são fruto de empate, mas ainda assim, classificadas como indefinidas.

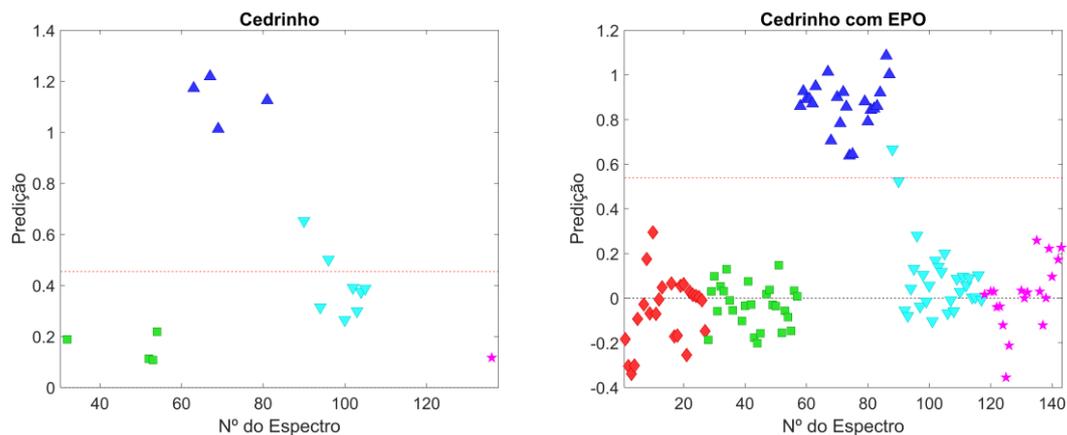


**Figura 22)** Valores de classe estimados para o conjunto de validação para discriminação da espécie cedro em relação às demais espécies antes e após correção com EPO, respectivamente, na condição de pré-processamento 3 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + *Mean Center*) e no modelo AE. (♦) andiroba; (■) cedro; (▲) cedrinho; (▼) curupixá; (★) mogno; (---) Limiar de discriminação.

O primeiro gráfico da figura 22 mostra o resultado do mesmo conjunto de validação apresentado na figura 21 (o número de *outliers* é o mesmo) sem correção por EPO, mas para discriminação da espécie cedro em relação às demais. Os quatro espectros restantes dos trinta iniciais referentes à própria classe cedro foram corretamente classificados, o que levaria à uma taxa de eficiência de 100,0%, mas novamente se observa que a grande exclusão de espectros inviabiliza o uso dos modelos sem correção.

Já o modelo de predição corrigido por EPO apresentou resultados mais adequados. Todas as amostras pertencentes à classe cedro foram corretamente classificadas, sendo, portanto, verdadeiros positivos. As classes andiroba e mogno também foram corretamente classificadas como não-pertencentes à classe cedro. Já as classes cedrinho e curupixá tiveram a maioria de seus espectros corretamente classificados como negativos à classe cedro, entretanto, a classe cedrinho apresentou um falso positivo e a classe curupixá exibiu um espectro no limite de classificação. Em resumo, corrigindo o efeito da umidade

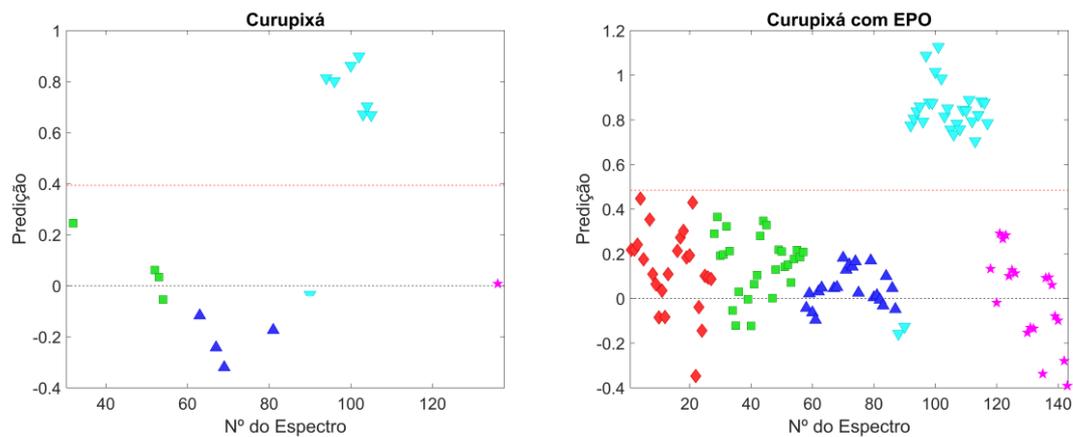
foi obtido um modelo para cedro nessa condição com taxa de eficiência de 98,1% e taxa de resultados indefinidos de 26,5%.



**Figura 23)** Valores de classe estimados para o conjunto de validação para discriminação da espécie cedrinho em relação às demais espécies antes e após correção com EPO, respectivamente, na condição de pré-processamento 3 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + *Mean Center*) e no modelo AE. (♦) andiroba; (■) cedro; (▲) cedrinho; (▼) curupixá; (★) mogno; (---) Limiar de discriminação.

De forma análoga ao apresentado na figura 22, a figura 23 mostra um resultado para o modelo de discriminação da espécie cedrinho. No modelo sem correção por EPO, os espectros pertencentes à própria classe cedrinho são corretamente classificados e os pertencentes às classes cedro e mogno também são classificados de maneira correta como não-pertencentes à classe em análise. Entretanto, a classe curupixá apresenta espectros como erro falso positivo. Lembrando que mesmo com um resultado razoável de discriminação, o alto número de *outliers* deletados influencia na real eficiência do modelo e inviabiliza o cálculo da taxa de eficiência.

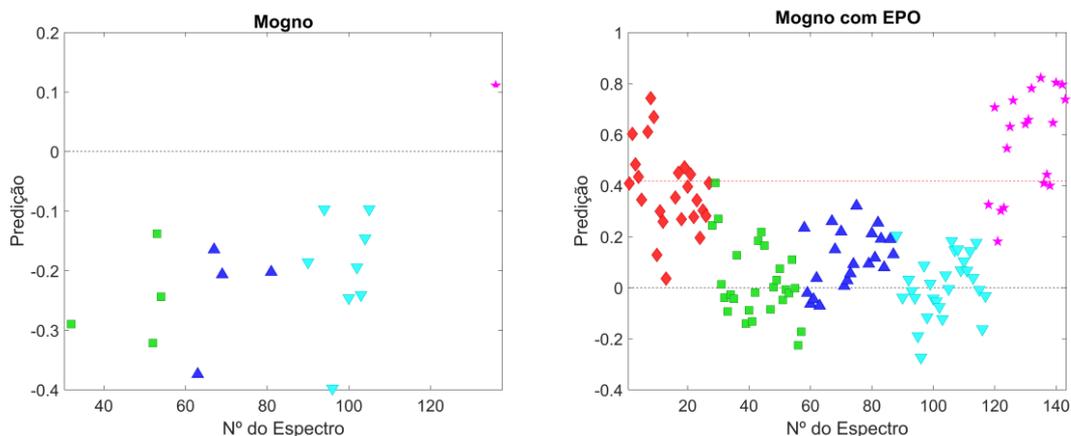
O gráfico que mostra os resultados com a correção por EPO apresentou ótimos resultados, com taxa de eficiência de 100% e taxa de resultados indefinidos de 26,5%. Nele, todos os espectros são corretamente classificados em suas devidas classes, com exceção de um espectro falso positivo pertencente à classe curupixá. Contudo, como o resultado da amostra é considerado pelo resultado conjunto das três replicatas, a amostra foi corretamente classificada.



**Figura 24)** Valores de classe estimados para o conjunto de validação para discriminação da espécie curupixá em relação às demais espécies antes e após correção com EPO, respectivamente, na condição de pré-processamento 3 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + *Mean Center*) e no modelo AE. (♦) andiroba; (■) cedro; (▲) cedrinho; (▼) curupixá; (★) mogno; (---) Limiar de discriminação.

A figura 24 mostra que os modelos de predição sem e com correção do efeito da umidade para discriminação do curupixá são, até agora, os mais satisfatórios. O pequeno número de espectros no modelo que não teve o efeito da umidade corrigido leva a classificação correta de quase todos eles. Apenas um espectro pertencente à própria classe curupixá é erroneamente classificado como falso negativo. Entretanto, é válido lembrar que mesmo com tais resultados e taxa de eficiência de 100,0%, o alto número de *outliers* deletados do modelo influenciam a eficiência do modelo.

De forma semelhante, o modelo com correção por EPO classifica corretamente a vasta maioria dos espectros analisados, havendo apenas dois espectros da própria classe curupixá classificados como falsos negativo e a aproximação de dois espectros pertencentes à classe andiroba do limite de classificação. Mesmo com esses dois falsos negativos, seus espectros devem formar a triplicata com outros dois espectros que foram corretamente classificados, o que leva à classificação correta da amostra, conforme os critérios de classificação ilustrados no esquema da figura 11 da seção 3.1.5. Para essa espécie, foi obtida uma taxa de eficiência de 100% e taxa de resultados indefinidos de 26,5%.



**Figura 25)** Valores de classe estimados para o conjunto de validação para discriminação da espécie mogno em relação às demais espécies antes e após correção com EPO, respectivamente, na condição de pré-processamento 3 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + *Mean Center*) e no modelo AE. (♦) andiroba; (■) cedro; (▲) cedrinho; (▼) curupixá; (★) mogno; (---) Limiar de discriminação.

Por fim, a figura 25 exhibe os resultados do modelo de discriminação da espécie mogno sem e com correção por EPO, respectivamente. No modelo cuja umidade não foi corrigida, o único espectro que restou após a exclusão dos *outliers* foi classificado corretamente. Contudo, o resultado para um único espectro não pode ser considerado como representativo, visto que uma amostra é composta por 3 espectros.

Já no modelo com correção do efeito da umidade por EPO, os espectros referentes às classes cedro, cedrinho e curupixá são classificados corretamente, sendo verdadeiros negativos, isto é, não-pertencentes à classe mogno. Entretanto, há uma mistura entre os espectros pertencentes às classes mogno e andiroba. Novamente, vale ressaltar o alto nível de semelhanças anatômicas entre essas duas espécies. Alguns espectros de mogno foram classificados como falsos negativos, e nos dos pertencentes à classe andiroba, classificados como falsos positivos. Para a discriminação de mogno se obteve taxa de eficiência igual a 74,9% e taxa de resultados indefinidos de 26,5%.

Os gráficos das figuras 21 a 25 são referentes ao modelo AE, escolhido para ilustrar a importância de se ter baixos valores de TINDEF por ser aquele com diferença mais significativa entre os modelos que passaram ou não pela correção por EPO. Entretanto, o modelo EA também apresentou excelentes valores de TINDEF (14,3%) e de TEF para a discriminação das cinco espécies de madeira, sendo eles: 84,0% para andiroba; 98,5% para cedro; 100,0% para cedrinho, 94,9% para curupixá; e 79,3% para mogno (tabela 10, pág. 41). O número absoluto de amostras indefinidas foi de 7 num total de 49 amostras e, destas 7, 5 foram identificadas como *outliers* e 2 foram fruto de empate

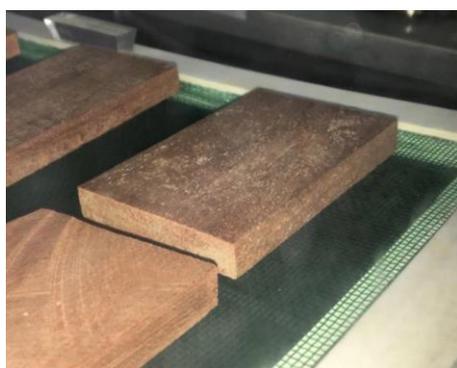
entre a classe de *outliers*, a classe de inconclusivas e uma das classes de treinamento, mas ainda assim, classificadas como indefinidas.

Os resultados obtidos nesta parte I do estudo não somente esclarecem o porquê das altas taxas de eficiência em modelos de discriminação criados com espectros que não passaram por correção do efeito da umidade, decorrentes do alto número de *outliers* excluídos, como também demonstram que corrigir os espectros por ortogonalização dos parâmetros externos (EPO) é um modo eficaz de remover as diferenças espectrais causadas pela umidade da madeira, que acabavam por comprometer a classificação das amostras.

## **4.2. Parte II: Avaliação dos modelos de predição criados**

### **4.2.1. Resultados do tratamento das amostras nas caixas umidificadoras**

As amostras selecionadas para esta parte do trabalho já se encontravam em condição seca e foram submetidas a um tratamento para aumentar seu teor de umidade. Como já comentado na seção 3.2.2, houve a proliferação de um fungo nas amostras de curupixá (figura 26), o que levou a exclusão da espécie na análise. Tais amostras foram imediatamente retiradas para que não houvesse contaminação nas amostras das demais espécies.



**Figura 26)** Amostra de curupixá com fungo.

Como já havia se passado 35 dias quando o fungo foi notado, a adição de novas amostras de curupixá às caixas atrasaria o cronograma da pesquisa, além da possibilidade de nova proliferação de fungos. Após o período de 55 dias, o acondicionamento das amostras nas caixas umidificadoras foi interrompido. O teor de umidade inicial das amostras e o teor obtido depois do tratamento são dispostos na tabela 12:

**Tabela 12)** Medida do teor de umidade inicial das amostras de quatro espécies florestais produtoras de madeira e do teor de umidade final após 55 dias de acondicionamento nas caixas com umidade controlada pelo sulfato de potássio

<b>Espécie</b>	<b>Amostra</b>	<b>Teor de umidade inicial (%)</b>	<b>Teor de umidade final (%)</b>
<b><i>Carapa guianensis</i> Aubl. (andiropa)</b>	<b>1</b>	6,4	17,8
	<b>2</b>	7,9	17,8
	<b>3</b>	6,8	17,6
	<b>4</b>	7,6	17,4
<b><i>Cedrela odorata</i> L. (cedro)</b>	<b>1</b>	5,7	17,6
	<b>2</b>	4,5	17,3
	<b>3</b>	6,5	18,2
	<b>4</b>	4,9	18,4
<b><i>Erismia uncinatum</i> Warm. (cedrinho)</b>	<b>1</b>	13,9	24,2
	<b>2</b>	14,3	23,1
	<b>3</b>	14,5	24,3
	<b>4</b>	12,4	24,0
<b><i>Swietenia macrophylla</i> King (mogno)</b>	<b>1</b>	2,6	14,2
	<b>2</b>	1,9	14,3
	<b>3</b>	1,6	15,2
	<b>4</b>	2,4	15,9

Analisando a tabela, é possível perceber claramente o efeito do tratamento umidificador. As amostras de andiroba, cedro, cedrinho e mogno tiveram um acréscimo médio aproximado em seu teor de umidade de 10%, 12%, 10% e 13%, respectivamente. Tais teores de umidade são consideravelmente maiores que os das amostras utilizadas para construção dos modelos de predição, expostos na tabela 5 (pág. 33). Assim, os espectros dessas amostras deram origem a um novo conjunto de dados para testar a eficácia dos modelos criados para discriminação das cinco espécies florestais produtoras de madeira em amostras com diferentes teores de umidade.

#### 4.2.2. Resultados das predições para a discriminação das amostras com maior teor de umidade

Como explanado na seção 3.2.4, quatro matrizes de validação foram criadas a partir dos espectros pré-processados das amostras da caixa: antes e após o processo de secagem, com e sem a correção por EPO (tabela 3, pág. 31). Essas matrizes foram testadas nos quatro modelos de discriminação calculados na parte I, resultando em oito combinações (tabela 4, pág. 32). Resumidamente, a tabela 13 dispõe tais combinações.

**Tabela 13)** Combinações para a discriminação das amostras com maior teor de umidade.

<b>Combinação</b>	<b>Condição do Modelo</b>	<b>Matriz de Validação</b>
<b>1</b>	Ambiente	Amostras caixa sem secagem
<b>2</b>	Ambiente	Amostras caixa pós-secagem
<b>3</b>	Estufa	Amostras caixa sem secagem
<b>4</b>	Estufa	Amostras caixa pós-secagem
<b>5</b>	Ambiente + EPO	Amostras caixa sem secagem + EPO
<b>6</b>	Ambiente + EPO	Amostras caixa pós-secagem + EPO
<b>7</b>	Estufa + EPO	Amostras caixa sem secagem + EPO
<b>8</b>	Estufa + EPO	Amostras caixa pós-secagem + EPO

Essas combinações entre modelo e matriz para as discriminações foram calculadas de maneira análoga a detalhada na figura 9 da parte I, a identificação das classes foi feita com o programa criado nessa dissertação (seção 3.1.5) e as figuras de mérito dos modelos de discriminação foram calculadas de maneira idêntica à da parte I (seção 2.3.3). Os valores referentes às taxas de eficiência e de resultados indefinidos estão dispostos na tabela 14.

**Tabela 14)** Taxas de eficiência e de amostras indefinidas para todos os modelos de predição das amostras que passaram pela caixa umidificadora na condição 3 de pré-processamento 3 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + *Mean Center*).

Combi- nação	TEF (%)				TINDEF (%)	<i>outliers</i>	incon	emp
	AN	CE	CH	M				
1	70,7	100,0	##	81,6	68,8	11	0	0
2	100,0	100,0	##	100,0	62,5	6	2	2
3	##	##	##	##	100,0	16	0	0
4	##	##	##	##	100,0	14	0	2
5	86,6	86,6	0,0	74,5	43,8	7	0	0
6	82,2	100,0	100,0	77,8	12,5	1	0	1
7	91,3	86,6	0,0	100,0	43,8	6	0	1
8	73,0	100,0	100,0	54,7	18,8	2	0	1

**Legenda:** AN: andiroba; CE: cedro; CH: cedrinho; M: mogno; TEF: taxa de eficiência; TINDEF: taxa de amostras indefinidas; *outliers*: nº de amostras anômalas; incon: nº de amostras com análise inconclusiva; emp: nº de amostras indefinidas com classes em empate; ##: dados não disponíveis devido a exclusão de todas as amostras da classe como *outliers*.

Avaliando a tabela, nota-se que a combinação 1 apresentou uma taxa de eficiência de 100,0% para a discriminação do cedro, de 81,6% para a discriminação do mogno e de 70,7% para a discriminação da andiroba. Já para a discriminação do cedrinho, não houveram dados disponíveis para calcular a taxa de eficiência devido a exclusão de todas as amostras da classe como *outliers*. A sua taxa de amostras indefinidas foi de 68,8%, um valor alto, ainda mais considerando que o tamanho do conjunto de dados, que conta com apenas 48 espectros. Dentre as amostras indefinidas, todas as 11 foram classificadas como *outliers*, indicando que o modelo está funcionando bem, mas que as amostras têm diferenças espectrais significativas, provavelmente causadas pela umidade. Portanto, como nos resultados da parte I, nota-se uma exclusão em massa de espectros e um resultado de predição pouco representativo.

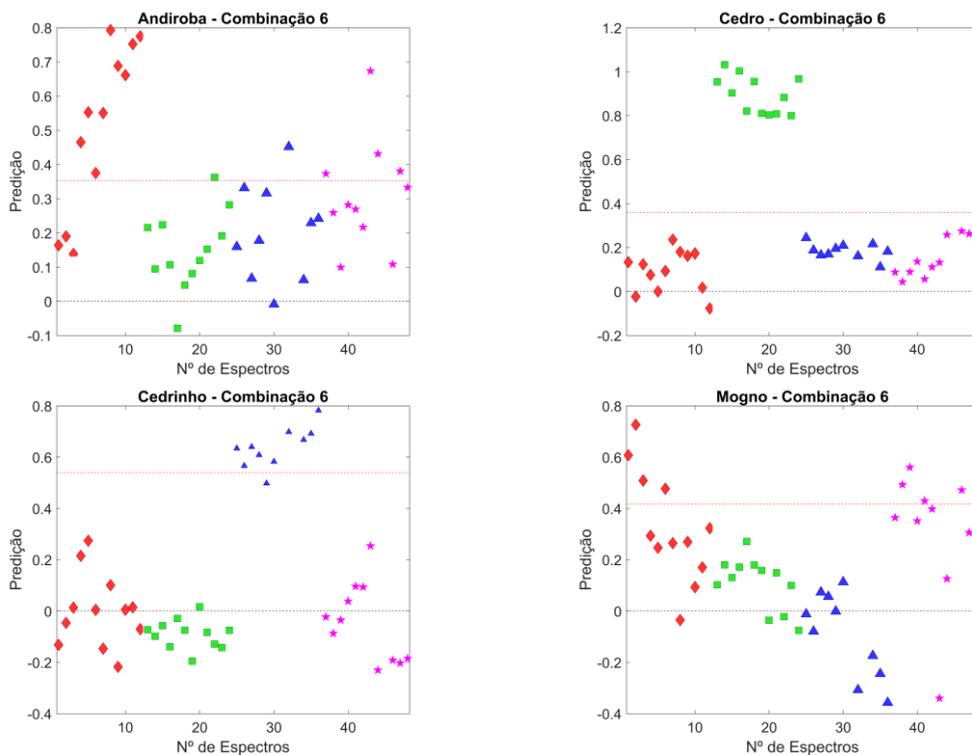
A combinação 2 apresentou taxas de eficiência de 100,0% para a discriminação da andiroba, do cedro e do mogno. Entretanto, assim como na combinação 1, houve a exclusão de todas as amostras de cedrinho como *outliers*. Sua alta taxa de indefinidos de 62,5% tem como origem 6 amostras classificadas como *outliers*, 2 com análise inconclusiva e 2 com classes em empate entre a classe de *outliers*, a classe de análises

inconclusivas e uma das classes de treinamento, o que a classifica como amostra indefinida.

As combinações 3 e 4 apresentaram taxas de amostras indefinidas de 100,0%, ou seja, todas as amostras do conjunto de dados foram excluídas como *outliers*.

Nas combinações de 5 a 8, passa a ser clara a atuação do EPO na melhora dos resultados. A combinação 5 apresentou taxa de eficiência de 86,6% para a discriminação da andiroba e cedro e de 74,5% para a discriminação do mogno. Entretanto, a discriminação do cedrinho foi falha, com zero de eficiência. Apesar da taxa de indefinidos ter diminuído com relação as das combinações cujos espectros não foram corrigidos por EPO, 43,8% ainda é consideravelmente alta e advém de 7 amostras classificadas como *outliers*.

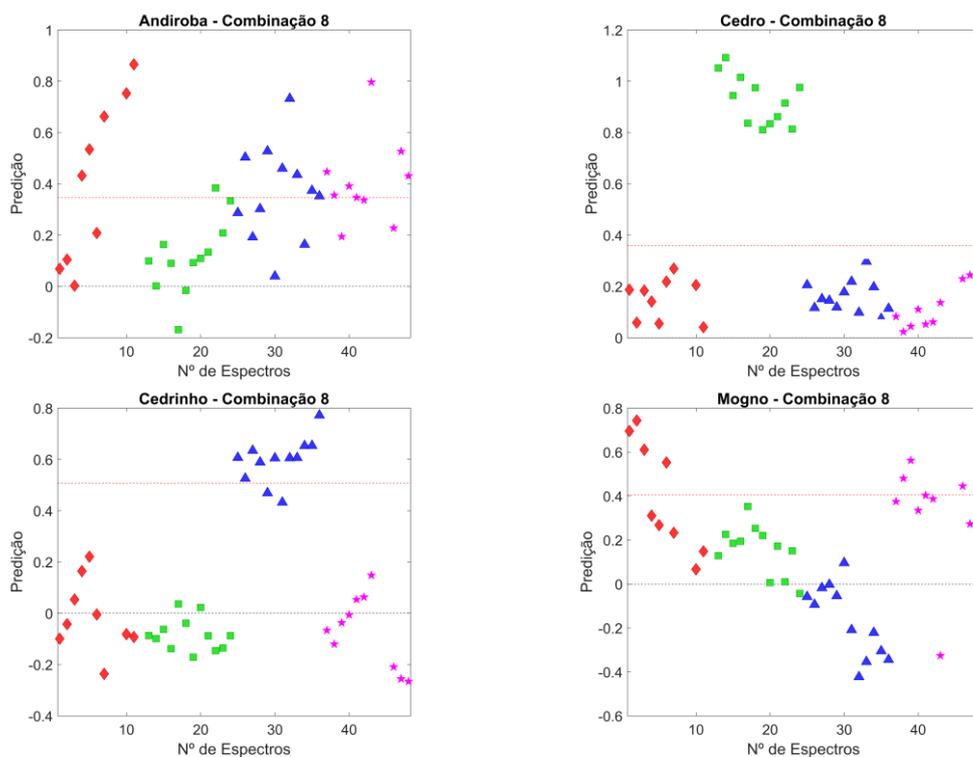
A combinação 6 foi aquela que apresentou os melhores resultados, com taxa de eficiência de 82,2% para a discriminação da andiroba, 100,0% para discriminação do cedro e do cedrinho, e de 77,8% para a discriminação do mogno, como apresentado nos gráficos da figura 27. Além disso, sua taxa de indefinidos foi de apenas 12,5%, formada por 1 amostra classificada como *outlier* e 1 fruto de empate entre classes.



**Figura 27)** Valores de classe estimados para o conjunto de validação para discriminação das espécies andiroba, cedro, cedrinho e mogno em relação às demais espécies na combinação 6. (♦) andiroba; (■) cedro; (▲) cedrinho; (▼) curupixá; (★) mogno; (---) Limiar de discriminação.

A combinação 7, assim como a 5, apresentou zero de eficiência na a discriminação do cedrinho. Já as taxas de eficiência para a discriminação da andiroba, do cedro e do mogno foram de 91,3%, 86,6% e 100,0%, respectivamente. Tal qual à combinação 5, apresentou taxa de indefinidos de 43,8%, formada por 6 amostras classificadas como *outliers* e 1 fruto de empate entre classes. As combinações 5 e 7 são aquelas cuja matriz de validação é o conjunto de espectros das amostras que passaram pelas caixas umidificadoras e apresentavam os maiores teores de umidade (tabela 12, pág. 51), indicando que este continua a ser um problema na discriminação. Entretanto, é importante notar que essas amostras apresentam teor de umidade superior ao das amostras que foram utilizadas para construir os modelos na parte I do trabalho (tabela 5, pág. 33). Esse fato é uma possível explicação para a correção por EPO não ter sido tão efetiva nessas duas combinações, principalmente observada na discriminação do cedrinho.

Já a combinação 8, apresentou taxas de eficiência de 73,0% para a discriminação da andiroba, 100,0% para discriminação do cedro e do cedrinho, e de 54,7% para a discriminação do mogno, como disposto nos gráficos da figura 28:



**Figura 28)** Valores de classe estimados para o conjunto de validação para discriminação das espécies andiroba, cedro, cedrinho e mogno em relação às demais espécies na combinação 8. (♦) andiroba; (■) cedro; (▲) cedrinho; (▼) curupixá; (★) mogno; (---) Limiar de discriminação.

Com uma taxa de indefinidos de 18,8% formada por 2 amostras classificadas como *outliers* e 1 fruto de empate entre classes, pode-se dizer que foi uma análise apropriada. Na combinação 8, a taxa de inconclusivos e *outliers* teve uma redução significativa, indicando que a correção por EPO foi efetiva. Contudo, infelizmente, não foram obtidos bons resultados de discriminação para as espécies andiroba e mogno. Porém, deve-se observar que o número de amostras que foi possível de ser realizado para esse estudo foi relativamente pequeno (apenas 4 amostras por espécie). E esse número pequeno de amostras acaba influenciando muito no cálculo das figuras de mérito (ex.: a mudança na classificação de uma amostra muda 25% da taxa de eficiência de uma espécie).

Levando em consideração as melhores taxas de eficiência e o pequeno índice de amostras inconclusivas e *outliers* nas combinações 6 e 8, pode-se confirmar a correção por EPO para minimizar a influência da umidade em espectros NIR de amostras de madeira. Dentre essas combinações, a de número 6 foi a que apresentou melhores resultados, a qual coincide com a condição escolhida na parte I (treinamento dos modelos na condição ambiente + aplicação do EPO). Uma hipótese para melhorar ainda mais os resultados seria aumentar a faixa do teor de umidade das amostras cujos espectros são responsáveis por construir a correção por EPO (matriz D, da equação 2, seção 2.3.1), o que em tese ampliaria a correção do efeito da umidade e poderia melhorar as discriminações das espécies andiroba, cedro, cedrinho, curupixá e mogno. Além disso, com o uso da caixa umidificadora e do processo de secagem empregados, pode-se criar um conjunto de dados em diferentes condições, mesmo utilizando um número limitado de amostras, apenas variando seu teor de umidade e obtendo mais espectros NIR. É plausível dizer que o método possa ser eficaz em mais espécies, não apenas as nesse trabalho estudadas, abrindo novas possibilidades para a pesquisa.

## 5. CONCLUSÕES

A aplicação da ferramenta matemática de ortogonalização parâmetros externos (EPO) na tentativa de eliminar ou reduzir a interferência da umidade nos espectros de amostras de madeira gerou resultados bastante promissores. A análise preliminar por componentes principais (PCA) foi útil para visualizar que seu uso minimiza a diferença espectral causada pelo efeito da umidade.

A condição de pré-processamento 3 (EPO: SNV + 1ª derivada; PLS-DA: SNV + 1ª derivada + *Mean Center*) apresentou a melhora mais significativa dos valores TINDEF dos modelos AE e EA após a correção por EPO, o que visa a resolução do problema encontrado nos trabalhos descritos na revisão bibliográfica, e apresentou valores de 26,5% e 14,3%, respectivamente.

Os modelos de discriminação que não foram corrigidos por EPO apresentaram boas taxas de eficiência, porém, essas são ilusórias. Durante sua criação, houve uma exclusão em massa de espectros considerados *outliers*, o que diminuiu consideravelmente o número de espectros para a avaliação do modelo e comprometeu sua representatividade.

A correção por EPO levou a uma diminuição significativa da taxa de amostras com resultado inconclusivo e *outliers* dos modelos de discriminação, inclusive naqueles criados combinando modelos de treinamento e amostras de validação em diferentes teores de umidade. Tal fato permitiu a obtenção de taxas de eficiência mais realistas e representativas dos dados e avaliar capacidade de discriminação do modelo para amostras de madeira com teores de umidade diferentes.

A discriminação de cinco madeiras foi feita utilizando o modelo de discriminação criado a partir do modelo de treinamento de amostras em umidade ambiente e validado com amostras que passaram pela estufa (AE). Como resultado, a taxa de amostras com resultado inconclusivo e *outliers* foi relativamente baixa (26,5%) e boas taxas de eficiência foram obtidas para a discriminação individual de cada uma das espécies (80,3%, 98,1%, 100%, 100% e 74,9% para discriminação de andiroba, cedro, cedrinho, curupixá e mogno, respectivamente). Os critérios mais flexíveis empregados para a classificação das amostras pelo soft-PLS-DA minimizaram erros falso positivos, negativos e possibilitam a melhor discriminação de amostras que a princípio eram identificadas como *outliers* ou apresentavam resultado inconclusivo.

Por fim, a correção por EPO se mostrou eficiente mesmo em amostras com teor de umidade elevado. A melhor combinação para as discriminações (combinação 6, com o modelo criado por amostras na condição de umidade ambiente e a matriz de validação formada pelas amostras que passaram pela caixa umidificadora e pelo processo de secagem, ambos corrigidos por EPO) apresentou altas taxas de eficiência (acima de 70%) para todas as espécies e baixos índices de *outliers* e resultados inconclusivos na discriminação de amostras com diferentes teores de umidade. Apenas amostras cuja umidade estava muito acima da faixa utilizada na construção do modelo demonstraram maior falha na discriminação.

Como perspectiva de trabalho futuro, o uso da caixa umidificadora e do processo de secagem empregados nesta dissertação possibilitam a criação de um conjunto de dados aumentando a faixa do teor de umidade das amostras. Assim, em tese, a matriz D seria mais abrangente e ampliaria a correção do efeito da umidade por EPO, o que levaria à melhora as discriminações das espécies estudadas. Além disso, o método pode ser testado em mais espécies florestais produtoras de madeira.

Em conjunto, a correção dos espectros de amostras de madeira por EPO e modelos de discriminação PLS-DA flexíveis por NIRS provaram ser uma combinação interessante para identificação das espécies de madeira andiroba, cedro, cedrinho, curupixá e mogno, sendo um potencial recurso na fiscalização do comércio ilegal madeireiro.

## 6. BIBLIOGRAFIA

- (1) United Nations Office on Drugs and Crime. *Best Practice Guide for Forensic Timber Identification*; Vienna, 2016.
- (2) Wiedenhoef, A. C.; Simeone, J.; Smith, A.; Parker-Forney, M.; Soares, R.; Fishman, A. Fraud and Misrepresentation in Retail Forest Products Exceeds U.S. Forensic Wood Science Capacity. *PLoS One* **2019**, *14* (7), e0219917. <https://doi.org/10.1371/journal.pone.0219917>.
- (3) Nellemann, C. *Green Carbon, Black Trade. A UNEP-INTERPOL Rapid Response Assessment*; 2012.
- (4) Dormontt, E. E.; Boner, M.; Braun, B.; Breulmann, G.; Degen, B.; Espinoza, E.; Gardner, S.; Guillery, P.; Hermanson, J. C.; Koch, G.; et al. Forensic Timber Identification: It's Time to Integrate Disciplines to Combat Illegal Logging. *Biol. Conserv.* **2015**, *191*, 790–798. <https://doi.org/10.1016/j.biocon.2015.06.038>.
- (5) United Nations Office on Drugs and Crime. Appendices I, II and III Valid from 28 August 2020. In *Best Practice Guide for Forensic Timber Identification*; 2020; p 74 p.
- (6) Gasson, P.; Baas, P.; Wheeler, E. Wood Anatomy of Cites-Listed Tree Species. *IAWA J.* **2011**, *32* (2), 155–198. <https://doi.org/10.1163/22941932-90000050>.
- (7) Amabis; Martho. *Biologia Dos Organismos, Volume 2, 2ª edição.*; Editora Moderna, 1996.
- (8) Soares, L. F. Avaliação Da Interferência Da Umidade Em Modelos de Discriminação de Madeira Por PLS-DA e de Métodos Para Sua Correção, UnB, 2018.
- (9) Hoadley, R. B. *Understanding Wood: A Craftsman's Guide to Wood Technology*; Taunton Press, 1980.
- (10) Coradin, V. T. R.; Camargos, J. A. A. *A Estrutura Anatômica Da Madeira e Princípios Para Sua Identificação*; Embrapa Amazônia Oriental (CPATU): Brasília, 2002.
- (11) Pastore, T. C. M. Estudos Do Efeito Da Radiação Ultravioleta Em Madeiras Por

- Espectroscopias Raman (FT-Raman), de Refletância Difusa No Infravermelho (DRIFT) e No Visível (CIE-L\*a\*b\*), Universidade de Brasília, 2004.
- (12) Schuchardt, U.; Ribeiro, M. L.; Gonçalves, A. R. A Indústria Petroquímica No Próximo Século: Como Substituir o Petróleo Como Matéria-Prima? *Quim. Nova* **2001**, *24* (2), 247–251. <https://doi.org/10.1590/S0100-40422001000200016>.
- (13) Pettersen, R. C. The Chemical Composition of Wood. In *The Chemistry of Solid Wood*; American Chemistry Society: Washington, DC, 1984; pp 57–126. <https://doi.org/10.1021/ba-1984-0207.ch002>.
- (14) Fengel, D.; Wegener, G. *Wood - Chemistry, Ultrastructure, Reactions*; Walter de Gruyter: Berlin, New York, 1984.
- (15) Biermann, C. *Handbook of Pulping and Papermaking*, 2nd editio.; Academic Press, 1996.
- (16) Morrison, R.; Boyd, R. *Química Orgânica*; Fundação Calouste Gulbenkian: Lisboa, 1970.
- (17) Dinwoodie, J. M.; OBE, J. M. D.; Establishment, B. R. *Timber: Its Nature and Behaviour, Second Edition*; Taylor & Francis, 2000.
- (18) Lewin, M.; Goldstein, I. S. *Wood Structure and Composition*; CRC Press, 1991.
- (19) Siau, J. F. *Transport Processes in Wood*; Springer Series in Wood Science; Springer Berlin Heidelberg: Berlin, Heidelberg, 1984; Vol. 2. <https://doi.org/10.1007/978-3-642-69213-0>.
- (20) Hillis, W. E. *Wood Extractives and Their Significance to the Pulp and Paper Industries*, 1st editio.; Academic Press, 1962.
- (21) Sjöström, E. *Wood Chemistry: Fundamentals and Applications*; Academic Press: New York, 1981.
- (22) Dinwoodie, J. Timber. In *Construction Materials: Their Nature and Behaviour, Fourth Edition*; 2010. <https://doi.org/10.4324/9780203927571>.
- (23) Ross, R. J.; USDA Forest Service., F. P. L. *Wood Handbook: Wood as an Engineering Material*; 2010. <https://doi.org/10.2737/FPL-GTR-190>.
- (24) Tsuchikawa, S. A Review of Recent Near Infrared Research for Wood and Paper.

*Appl. Spectrosc. Rev.* **2007**, *42* (1), 43–71.  
<https://doi.org/10.1080/05704920601036707>.

- (25) Jankowsky, I. P.; SANTOS, G. R. V.; ANDRADE, A. *Secagem Da Madeira Serrada de Eucalipto*. Piracicaba 2003, p 12 p.
- (26) MARTINS, V. A. *Secagem de Madeira Serrada*; IBDF/DPq - LPF: Brasília, 1988.
- (27) Simpson, W. *Dry Kiln Operator's Manual*; 1991.
- (28) El Kouali, M.; Vergnaud, J. M. Modeling the Process of Absorption and Desorption of Water above and below the Fiber Saturation Point. *Wood Sci. Technol.* **1991**, *25* (5), 327–339. <https://doi.org/10.1007/BF00226172>.
- (29) Galvao, A. P. M.; Jankowsky, I. P. *Secagem Racional Da Madeira*; Nobel, 1985.
- (30) U. S. Department of Agriculture, Forest Service, F. P. L. *Seasoning Mixed Species of Tropical Hardwoods*; Yokohama, 1988.
- (31) Winston, P. W.; Bates, D. H. Saturated Solutions For the Control of Humidity in Biological Research. *Ecology* **1960**, *41* (1), 232–237. <https://doi.org/10.2307/1931961>.
- (32) Young, J. F. Humidity Control in the Laboratory Using Salt Solutions-a Review. *J. Appl. Chem.* **1967**, *17* (9), 241–245. <https://doi.org/10.1002/jctb.5010170901>.
- (33) Souza, M. H. de; Magliano, M. M.; Camargos, J. A. A. *Madeiras Tropicais Brasileiras*, 1<sup>a</sup> ed.; Laboratorio de Produtos Florestais, IBAMA: Brasília, DF, 1997.
- (34) Michael Höltken, A.; Schröder, H.; Wischniewski, N.; Degen, B.; Magel, E.; Fladung, M. Development of DNA-Based Methods to Identify CITES-Protected Timber Species: A Case Study in the Meliaceae Family. *Holzforschung* **2012**, *66* (1). <https://doi.org/10.1515/HF.2011.142>.
- (35) Musah, R. A.; Espinoza, E. O.; Cody, R. B.; Lesiak, A. D.; Christensen, E. D.; Moore, H. E.; Maleknia, S.; Drijfhout, F. P. A High Throughput Ambient Mass Spectrometric Approach to Species Identification and Classification from Chemical Fingerprint Signatures. *Sci. Rep.* **2015**, *5* (1), 11520. <https://doi.org/10.1038/srep11520>.

- (36) Lancaster, C.; Espinoza, E. Analysis of Select Dalbergia and Trade Timber Using Direct Analysis in Real Time and Time-of-Flight Mass Spectrometry for CITES Enforcement. *Rapid Commun. Mass Spectrom.* **2012**, *26* (9), 1147–1156. <https://doi.org/10.1002/rcm.6215>.
- (37) Snel, F. A.; Braga, J. W. B.; da Silva, D.; Wiedenhoeft, A. C.; Costa, A.; Soares, R.; Coradin, V. T. R.; Pastore, T. C. M. Potential Field-Deployable NIRS Identification of Seven Dalbergia Species Listed by CITES. *Wood Sci. Technol.* **2018**, *52* (5), 1411–1427. <https://doi.org/10.1007/s00226-018-1027-9>.
- (38) Silva, D. C.; Pastore, T. C. M.; Soares, L. F.; de Barros, F. A. S.; Bergo, M. C. J.; Coradin, V. T. H.; Gontijo, A. B.; Sosa, M. H.; Chacón, C. B.; Braga, J. W. B. Determination of the Country of Origin of True Mahogany (*Swietenia Macrophylla* King) Wood in Five Latin American Countries Using Handheld NIR Devices and Multivariate Data Analysis. *Holzforschung* **2018**, *72* (7), 521–530. <https://doi.org/10.1515/hf-2017-0160>.
- (39) Soares, L. F.; da Silva, D. C.; Bergo, M. C. J.; Coradin, V. T. R.; Braga, J. W. B.; Pastore, T. C. M. Avaliação de Espectrômetro NIR Portátil e PLS-DA Para a Discriminação de Seis Espécies Similares de Madeira Amazônicas. *Quim. Nova* **2017**. <https://doi.org/10.21577/0100-4042.20170014>.
- (40) Meder, R.; Brawner, J. T.; Downes, G. M.; Ebdon, N. Towards the In-Forest Assessment of Kraft Pulp Yield: Comparing the Performance of Laboratory and Hand-Held Instruments and Their Value in Screening Breeding Trials. *J. Near Infrared Spectrosc.* **2011**, *19* (5), 421–429. <https://doi.org/10.1255/jnirs.954>.
- (41) Schmitz, N.; Beckman, H.; Cabezas, J. A.; Cervera, M. T.; Espinoza, E.; Fernandez-Golfin, J.; Gasson, P.; Hermanson, J.; Arteaga, M.; Koch, G.; et al. *The Timber Tracking Tool Infogram. Overview of Wood Identification Methods' Capacity.*; 2019. <https://doi.org/10.13140/RG.2.2.27920.25603>.
- (42) Braga, J. W. B.; Pastore, T. C. M.; Coradin, V. T. R.; Camargos, J. A. A.; da Silva, A. R. The Use of near Infrared Spectroscopy to Identify Solid Wood Specimens of *Swietenia Macrophylla* (Cites Appendix II). *IAWA J.* **2011**, *32* (2), 285–296. <https://doi.org/10.1163/22941932-90000058>.
- (43) Pastore, T. C. M.; Braga, J. W. B.; Coradin, V. T. R.; Magalhães, W. L. E.; Okino,

- E. Y. A.; Camargos, J. A. A.; de Muñiz, G. I. B.; Bressan, O. A.; Davrieux, F. Near Infrared Spectroscopy (NIRS) as a Potential Tool for Monitoring Trade of Similar Woods: Discrimination of True Mahogany, Cedar, Andiroba, and Curupixá. *Holzforschung* **2011**, *65* (1), 73–80. <https://doi.org/10.1515/hf.2011.010>.
- (44) Bergo, M. C. J.; Pastore, T. C. M.; Coradin, V. T. R.; Wiedenhoeft, A. C.; Braga, J. W. B. NIRS Identification of *Swietenia Macrophylla* Is Robust across Specimens from 27 Countries. *IAWA J.* **2016**, *37* (3), 420–430. <https://doi.org/10.1163/22941932-20160144>.
- (45) de Santana, F. B.; de Giuseppe, L. O.; de Souza, A. M.; Poppi, R. J. Removing the Moisture Effect in Soil Organic Matter Determination Using NIR Spectroscopy and PLSR with External Parameter Orthogonalization. *Microchem. J.* **2019**, *145*, 1094–1101. <https://doi.org/10.1016/j.microc.2018.12.027>.
- (46) Minasny, B.; McBratney, A. B.; Bellon-Maurel, V.; Roger, J.-M.; Gobrecht, A.; Ferrand, L.; Joalland, S. Removing the Effect of Soil Moisture from NIR Diffuse Reflectance Spectra for the Prediction of Soil Organic Carbon. *Geoderma* **2011**, *167–168*, 118–124. <https://doi.org/10.1016/j.geoderma.2011.09.008>.
- (47) Roger, J.-M.; Chauchard, F.; Bellon-Maurel, V. EPO–PLS External Parameter Orthogonalisation of PLS Application to Temperature-Independent Measurement of Sugar Content of Intact Fruits. *Chemom. Intell. Lab. Syst.* **2003**, *66* (2), 191–204. [https://doi.org/10.1016/S0169-7439\(03\)00051-0](https://doi.org/10.1016/S0169-7439(03)00051-0).
- (48) Skoog, D. A.; Holler, F. J.; Crouch, S. R.; Pasquini, C.; Da Unicamp, E. *Princípios de Análise Instrumental*; Bookman, 2009.
- (49) Ohlweiler, O. A. *Fundamentos de Análise Instrumental*; Livros Técnicos e Científicos: Rio de Janeiro, 1981.
- (50) Eskildsen, C. E. A. Prediction of Milk Quality Parameters Using Vibrational Spectroscopy and Chemometrics: Opportunities and Challenges in Milk Phenotyping, University of Copenhagen, 2016.
- (51) Herschel, W. XIV. Experiments on the Refrangibility of the Invisible Rays of the Sun. *Philos. Trans. R. Soc. London* **1800**, *90*, 284–292. <https://doi.org/10.1098/rstl.1800.0015>.

- (52) Pasquini, C. Princípios Da Espectroscopia No Infravermelho Próximo. In *Espectroscopia no Infravermelho Próximo para Avaliar Indicadores de Qualidade Tecnológica e Contaminantes em Grãos*; Embrapa: Brasília, DF, 2018; p 200.
- (53) Pasquini, C. Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. *J. Braz. Chem. Soc.* **2003**, *14* (2), 198–219. <https://doi.org/10.1590/S0103-50532003000200006>.
- (54) Bruns, R.E., F. J. F. G. QUIMIOMETRIA. *Quim. Nova* **1985**, v. 8, 84–89.
- (55) Sena, M. M. de; Almeida, M. R. de. Quimiometria Aplicada Aos Dados Espectrais No Infravermelho Próximo. In *Espectroscopia no Infravermelho Próximo para Avaliar Indicadores de Qualidade Tecnológica e Contaminantes em Grãos*; Embrapa: Brasília, DF, 2018; p 200.
- (56) Ferreira, M. M. C. *Quimiometria: Conceitos, Métodos e Aplicações*; 2015. <https://doi.org/10.7476/9788526814714>.
- (57) Brereton, R. G. Introduction to Multivariate Calibration in Analytical Chemistry. *Analyst* **2000**, *125* (11), 2125–2154. <https://doi.org/10.1039/b003805i>.
- (58) Westad, F.; Marini, F. Validation of Chemometric Models – A Tutorial. *Anal. Chim. Acta* **2015**, *893*, 14–24. <https://doi.org/10.1016/j.aca.2015.06.056>.
- (59) Barker, M.; Rayens, W. Partial Least Squares for Discrimination. *J. Chemom.* **2003**, *17* (3), 166–173. <https://doi.org/10.1002/cem.785>.
- (60) Brereton, R. G.; Lloyd, G. R. Partial Least Squares Discriminant Analysis: Taking the Magic Away. *J. Chemom.* **2014**, *28* (4), 213–225. <https://doi.org/10.1002/cem.2609>.
- (61) Borin, A.; Poppi, R. J. Multivariate Quality Control of Lubricating Oils Using Fourier Transform Infrared Spectroscopy. *J. Braz. Chem. Soc.* **2004**, *15* (4), 570–576. <https://doi.org/10.1590/S0103-50532004000400020>.
- (62) ASTM E1655-05. Standard Practices for Infrared Multivariate Quantitative Analysis. *ASTM Int.* **2012**. <https://doi.org/10.1520/E1655-05R12.2>.
- (63) Santana, F.; Souza, A.; Almeida, M.; Breikreitz, M.; Filgueiras, P.; Sena, M.; Poppi, R. Experimento Didático de Quimiometria Para Classificação de Óleos

Vegetais Comestíveis Por Espectroscopia No Infravermelho Médio Combinado Com Análise Discriminante Por Mínimos Quadrados Parciais: Um Tutorial, Parte V. *Quim. Nova* **2020**. <https://doi.org/10.21577/0100-4042.20170480>.

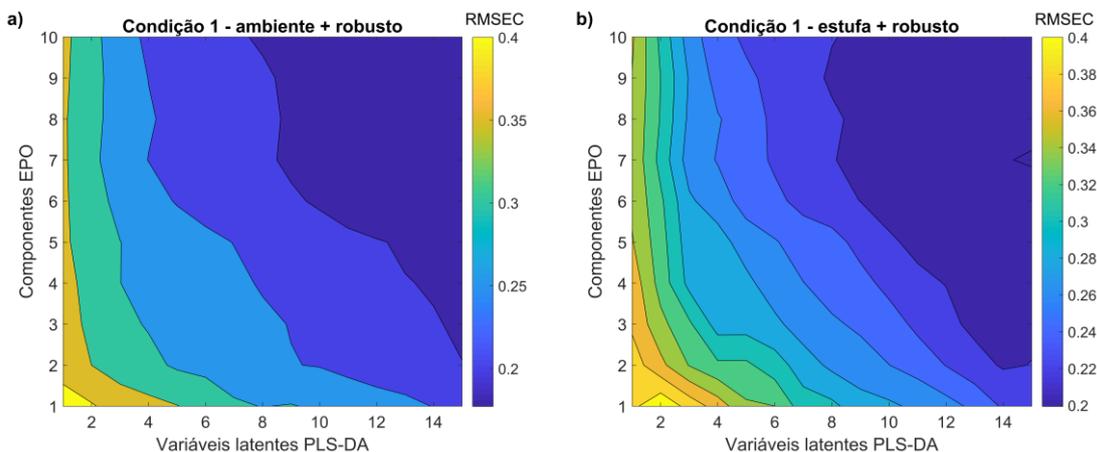
- (64) Martins, A. R.; Talhavini, M.; Vieira, M. L.; Zacca, J. J.; Braga, J. W. B. Discrimination of Whisky Brands and Counterfeit Identification by UV–Vis Spectroscopy and Multivariate Data Analysis. *Food Chem.* **2017**, *229*, 142–151. <https://doi.org/10.1016/j.foodchem.2017.02.024>.
- (65) Rousseeuw, P. J.; Debruyne, M.; Engelen, S.; Hubert, M. Robustness and Outlier Detection in Chemometrics. *Crit. Rev. Anal. Chem.* **2006**, *36* (3–4), 221–242. <https://doi.org/10.1080/10408340600969403>.
- (66) Walczak, B.; Massart, D. L. Multiple Outlier Detection Revisited. *Chemom. Intell. Lab. Syst.* **1998**, *41* (1), 1–15. [https://doi.org/10.1016/S0169-7439\(98\)00034-3](https://doi.org/10.1016/S0169-7439(98)00034-3).
- (67) da Silva, V. A. G.; Talhavini, M.; Peixoto, I. C. F.; Zacca, J. J.; Maldaner, A. O.; Braga, J. W. B. Non-Destructive Identification of Different Types and Brands of Blue Pen Inks in Cursive Handwriting by Visible Spectroscopy and PLS-DA for Forensic Analysis. *Microchem. J.* **2014**, *116*, 235–243. <https://doi.org/10.1016/j.microc.2014.05.013>.
- (68) Laursen, K.; Frederiksen, S. S.; Leuenhagen, C.; Bro, R. Chemometric Quality Control of Chromatographic Purity. *J. Chromatogr. A* **2010**, *1217* (42), 6503–6510. <https://doi.org/10.1016/j.chroma.2010.08.040>.
- (69) Jackson, J. E.; Mudholkar, G. S. Control Procedures for Residuals Associated With Principal Component Analysis. *Technometrics* **1979**, *21* (3), 341–349. <https://doi.org/10.1080/00401706.1979.10489779>.
- (70) Bro, R.; Smilde, A. K. Principal Component Analysis. *Anal. Methods* **2014**, *6* (9), 2812–2831. <https://doi.org/10.1039/C3AY41907J>.
- (71) Pomerantsev, A. L.; Rodionova, O. Y. Multiclass Partial Least Squares Discriminant Analysis: Taking the Right Way-A Critical Tutorial. *J. Chemom.* **2018**, *32* (8), e3030. <https://doi.org/10.1002/cem.3030>.
- (72) Nunes, K. M.; Andrade, M. V. O.; Almeida, M. R.; Sena, M. M. A Soft Discriminant Model Based on Mid-Infrared Spectra of Bovine Meat Purges to

- Detect Economic Motivated Adulteration by the Addition of Non-Meat Ingredients. *Food Anal. Methods* **2020**, *13* (9), 1699–1709. <https://doi.org/10.1007/s12161-020-01795-3>.
- (73) Zontov, Y. V.; Rodionova, O. Y.; Kucheryavskiy, S. V.; Pomerantsev, A. L. PLS-DA – A MATLAB GUI Tool for Hard and Soft Approaches to Partial Least Squares Discriminant Analysis. *Chemom. Intell. Lab. Syst.* **2020**, *203*, 104064. <https://doi.org/10.1016/j.chemolab.2020.104064>.
- (74) Botelho, B. G.; Reis, N.; Oliveira, L. S.; Sena, M. M. Development and Analytical Validation of a Screening Method for Simultaneous Detection of Five Adulterants in Raw Milk Using Mid-Infrared Spectroscopy and PLS-DA. *Food Chem.* **2015**, *181*, 31–37. <https://doi.org/10.1016/j.foodchem.2015.02.077>.
- (75) Oliveri, P.; Downey, G. Multivariate Class Modeling for the Verification of Food-Authenticity Claims. *TrAC Trends Anal. Chem.* **2012**, *35*, 74–86. <https://doi.org/10.1016/j.trac.2012.02.005>.
- (76) Coradin, V. T. R.; Camargos, J. A. A.; Marques, L. F.; Silva Junior, E. R. da S. *Madeiras Similares Ao Mogno (Swietenia Macrophylla King.): Chave Ilustrada Para Identificação Anatômica Em Campo*; Serviço Florestal Brasileiro, 2009.
- (77) Ministério do Meio Ambiente – MMA. Lista Nacional Oficial de Espécies Da Flora Brasileira Ameaçadas de Extinção; Portaria n. 443, de 17 de dezembro de 2014. Diário Oficial da União, 18/12/2014, Seção 1, 2014; pp 110–121.
- (78) Site da Digisystem  
<http://www.digisystemeletro.com.br/produtos/medidores/medidores.php>  
(acessado em 20 de Janeiro de 2020).
- (79) Site da Thermo Fisher  
<https://www.thermofisher.com/order/catalog/product/MICROPHAZIRRX>  
(acessado em 20 de Janeiro de 2020).
- (80) Lopes, R. E. C. Discriminação de Madeiras Similares Por NIRS e PLS-DA Considerando Variações de Temperatura e Umidade, Universidade de Brasília, 2015.
- (81) Merck Millipore. *Safety Data Sheet for Sulfato de Potássio*; 2018.

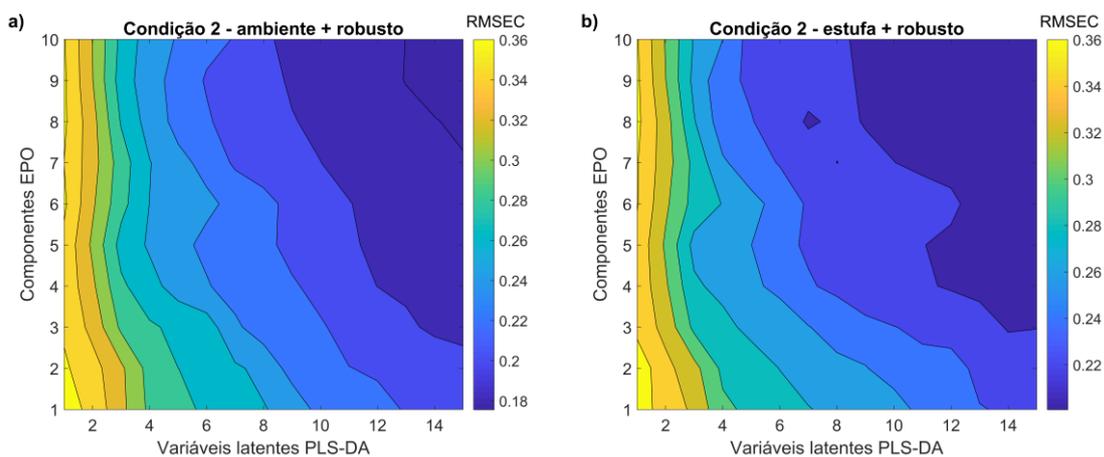
- (82) Brown, H. P.; Koehler, A.; Forsaith, C. C. *Textbook of Wood Technology, v.1*; McGraw-Hill Book Company: New York, 1949. <https://doi.org/10.2307/1137932>.
- (83) Schwanninger, M.; Rodrigues, J. C.; Fackler, K. A Review of Band Assignments in near Infrared Spectra of Wood and Wood Components. *J. Near Infrared Spectrosc.* **2011**, *19* (5), 287–308. <https://doi.org/10.1255/jnirs.955>.
- (84) Tsuchikawa, S.; Tsutsumi, S. Adsorptive and Capillary Condensed Water in Biological Material. *J. Mater. Sci. Lett.* **1998**, *17* (8), 661–663. <https://doi.org/10.1023/A:1006672324163>.

## 7. APÊNDICES

**7.1. APÊNDICE A - Gráfico de curvas de nível de valores de RMSEC para modelos corrigidos por EPO da matriz de treinamento na condição (a) ambiente e (b) estufa, predizendo os dados da matriz de treinamento robusta na condição de pré-processamento 1 (EPO: bruto; PLS-DA: *Mean Center*).**



**7.2. APÊNDICE B - Gráfico de curvas de nível de valores de RMSEC para modelos corrigidos por EPO da matriz de treinamento na condição (a) ambiente e (b) estufa, predizendo os dados da matriz de treinamento robusta na condição de pré-processamento 2 (EPO: SNV; PLS-DA: SNV + *Mean Center*).**



**7.3. APÊNDICE C - Gráfico de curvas de nível de valores de RMSEC para modelos corrigidos por EPO da matriz de treinamento na condição (a) ambiente e (b) estufa, predizendo os dados da matriz de treinamento robusta na condição de pré-processamento 4 (EPO: 1ª derivada; PLS-DA: 1ª derivada + *Mean Center*).**

