



UnB

Instituto de Química

Programa de Pós-Graduação em Química

DISSERTAÇÃO DE MESTRADO

Avaliação da interferência da umidade em modelos de discriminação de madeira por PLS-DA e de métodos para sua correção

Aluna: Liz Franco Soares

Orientador: Jez Willian Batista Braga

Coorientadora: Tereza Cristina Monteiro Pastore

Brasília, DF
2018

FOLHA DE APROVAÇÃO

Comunicamos a aprovação da Defesa de Dissertação do (a) aluno (a) **Liz Franco Soares**, matrícula nº **16/0163404**, intitulada “***Avaliação da interferência da umidade em modelos de discriminação de madeira por PLS - DA e de métodos para sua correção***”, apresentada no (a) PADCT da Universidade de Brasília (UnB) em 10 de agosto de 2018.

Prof. Dr. Jez Willian Batista Braga
Presidente de Banca (IQ/UnB)

Prof. Dr. Alexandre Fonseca
Membro Titular (IQ/UnB)

Dr. Paulo José Prudente de Fontes
Membro Titular (Serviço Florestal Brasileiro (SFB))

Dra. Fernanda Piccolo Pieruzzi
Membro Suplente (Serviço Florestal Brasileiro (SFB))

Em 10 de agosto de 2018.

DEDICATÓRIA

*Dedico este trabalho aos meus pais e meu
irmão.*

AGRADECIMENTOS

A Jeová Deus, por sua bondade imerecida.

Ao meu orientador Prof. Dr Jez Willian Batista Braga, pelos ensinamentos, acompanhamento e confiança.

À minha coorientadora Dra. Tereza Cristina Monteiro Pastore, pelo incentivo e motivação.

Ao Instituto de Química da Universidade de Brasília, pela oportunidade para realizar o mestrado.

Ao Laboratório de Produtos Florestais (LPF), pela disponibilização das amostras de madeira.

Ao Grupo de Estudos em Automação, Quimiometria e Química Ambiental (AQQUA), pelo ambiente de trabalho.

Ao Instituto Nacional de Ciência e Tecnologia em Bioanalítica (INCTBio) e CNPq, pela bolsa concedida.

Aos órgãos de fomento, CITES-ITTO, CNPq (processo 307223/2012-4) e INCTBio (processo 465389/2014-7), pelo apoio financeiro.

À Dra. Vera Teresa Rauber Coradin e ao Me. Alexandre Bahia Gontijo, pelas contribuições e identificação das amostras.

À Prof. Dra. Andressa Vasques, por ceder o desidratador de alimentos para realização dos experimentos.

À equipe de trabalho: Diego Carvalho, Filipe Snell, Nayara Cristina, Priscila Veras e Rafaela Dornelas, pela contribuição na realização e conclusão deste trabalho.

RESUMO

Um dos aspectos que tem dificultado a exploração sustentável de madeira é o fato de existirem muitas espécies florestais visualmente similares entre si, o que facilita o comércio ilegal. Além disso, não existem anatomistas em identificação de madeira suficientes para atender à crescente demanda. Por isso, requer-se o desenvolvimento de métodos instrumentais robustos para auxiliar a discriminação de espécies produtoras de madeira. Para contribuir nesse sentido, essa dissertação propõe o uso da espectroscopia no infravermelho próximo (NIRS) em conjunto com a quimiometria para identificar madeiras amazônicas em campo, com o uso do espectrômetro portátil MicroNirTM. Discriminaram-se as espécies semelhantes anatomicamente: andiroba (*Carapa guianensis*), cedrinho (*Erisma uncinatum*), cedro (*Cedrela odorata*), curupixá (*Micropholis melinoniana*) e mogno (*Swietenia macrophylla*). Na condição de madeira seca (umidade entre 10 e 15%), foram discriminadas 711 amostras diferentes das cinco espécies com taxa de eficiência (TEF) acima de 90%; na condição de umidade mais elevada (acima de 15%), duas estratégias foram estudadas para reduzir a interferência da umidade nos espectros NIR: (1) aplicação da ferramenta matemática: correção ortogonal de sinal (OSC) e (2) procedimento de secagem em estufa com circulação de ar. Na primeira estratégia, 70% dos espectros de amostras úmidas processados com OSC foram identificados como *outliers* na etapa de validação, indicando que a ferramenta não foi eficiente. Na segunda estratégia, amostras de cedrinho (*Erisma uncinatum*), cedro (*Cedrela odorata*) e mogno (*Swietenia macrophylla*) foram secas por 20 minutos à 80°C na estufa. Nesse caso, utilizando o algoritmo DUPLEX para selecionar os conjuntos de treinamento e validação, os modelos de discriminação apresentaram TEF acima de 90% e taxas de falso positivo e falso negativo abaixo de 5%. Além disso, o número de *outliers* foi compatível com o modelo desenvolvido com madeira seca. Os resultados demonstraram que os modelos de discriminação puderam ser aplicados para amostras com diferentes valores de umidade que passaram pelo mesmo procedimento de secagem adotado, o que minimizou a interferência da umidade nos espectros NIR da madeira.

ABSTRACT

One of the aspects that has hindered the sustainable wood exploitation is the fact that there are many similar visually species, which facilitates illegal trade. Besides, there are no wood identification experts in sufficient number to meet the growing demand. Because of that, it requires the development of robust instrumental methods to help discrimination of wood species. In this work, near infrared spectroscopy (NIRS) and chemometrics were successfully used to discriminate Amazonian wood species in the field, using portable spectrometer MicroNir™. Similar anatomical species crabwood (*Carapa guianensis*); cedrinho (*Erismia uncinatum*), cedar (*Cedrela odorata*), curupixá (*Micropholis melinoniana*) and mahogany (*Swietenia macrophylla*) were discriminated. In dry wood condition (humidity between 10 and 15%), 711 different samples were discriminated with efficiency rate above 90%; in the highest humidity condition (above 15%), two strategies were studied to reduce or eliminate the moisture interference in the NIR spectra: (1) application of the mathematical tool Orthogonal Signal Correction (OSC) and (2) drying procedure in a greenhouse with air circulation. In the first strategy, 70% of the spectra of samples, in highest humidity condition, processed with OSC were identified as outliers in the validation step, this indicates that OSC was not efficient. In the second strategy, samples of cedrinho (*Erismia uncinatum*), cedar (*Cedrela odorata*) and mahogany (*Swietenia macrophylla*) were dried for 20 minutes at 80°C in the greenhouse. In this case, using the DUPLEX algorithm to select training and validation sets, the discrimination models presented efficiency rate above 90% and false positive and false negative rate were below 5%. Moreover, the number of outliers was compatible with the model developed with samples in dry wood condition. The results showed that the discrimination models could be applied for samples with different humidity values that went through the same drying procedure adopted, which minimized the moisture interference in the NIR spectra of the wood.

ÍNDICE

| | |
|--|----|
| LISTA DE ABREVIATURAS E ACRÔNIMOS | |
| LISTA DE GRÁFICOS E TABELAS | |
| LISTA DE FIGURAS | |
| 1 INTRODUÇÃO E OBJETIVOS | 1 |
| 2 REVISÃO BIBLIOGRÁFICA | 3 |
| 2.1 MADEIRA | 3 |
| 2.1.1 Estrutura anatômica | 3 |
| 2.1.2 Composição química | 5 |
| 2.1.3 Umidade e propriedades físicas da madeira | 6 |
| 2.1.4 Métodos de identificação de madeira | 8 |
| 2.1.5 Espectroscopia no infravermelho próximo | 11 |
| 2.2 ANÁLISE MULTIVARIADA | 11 |
| 2.2.1 Seleção de amostras e desenvolvimento do modelo PLS-DA | 12 |
| 2.2.2 Determinação das figuras de mérito do modelo PLS-DA | 18 |
| 2.2.3 Correção Ortogonal de Sinal | 18 |
| 3 CORREÇÃO DA UMIDADE EM MODELOS DE DISCRIMINAÇÃO PLS-DA POR OSC | 22 |
| 3.1 MATERIAIS E MÉTODOS | 24 |
| 3.1.1 Obtenção e preparo das amostras | 24 |
| 3.1.2 Obtenção dos espectros NIR e medidas de umidade | 25 |
| 3.2 RESULTADOS E DISCUSSÃO | 28 |
| 3.2.1 Análise exploratória dos espectros das amostras secas e úmidas | 28 |
| 3.2.2 Modelo de discriminação por PLS-DA | 36 |
| 3.3 CONCLUSÕES PARCIAIS | 45 |

| | | |
|-------|---|----|
| 4 | CORREÇÃO DA UMIDADE POR SECAGEM | 46 |
| 4.1 | MATERIAIS E MÉTODOS | 48 |
| 4.1.1 | Obtenção e preparo das amostras | 48 |
| 4.1.2 | Avaliação do tempo de secagem | 49 |
| 4.1.3 | Desenvolvimento de novo modelo de discriminação | 51 |
| 4.2 | RESULTADOS E DISCUSSÃO | 53 |
| 4.2.1 | Avaliação do tempo de secagem | 53 |
| 4.2.2 | Desenvolvimento de modelo de discriminação | 66 |
| 4.3 | CONCLUSÕES PARCIAIS | 74 |
| 5 | CONCLUSÕES FINAIS | 75 |
| 6 | REFERÊNCIAS BIBLIOGRÁFICAS | 77 |
| 7 | ANEXO | 80 |

LISTA DE ABREVIATURAS E ACRÔNIMOS

| | |
|------------|--|
| AQUA | Laboratório de Automação, Quimiometria e Química Ambiental |
| CITES | Convenção sobre o Comércio Internacional de Espécies da Flora e Fauna Selvagens em Perigo de Extinção (do inglês, <i>Convention on International Trade in Endangered Species of Wild Flora and Fauna</i>) |
| DART-TOFMS | Análise direta em tempo real no espectrômetro de massas, do inglês <i>Direct Analysis in Real Time Mass Spectrometry</i> |
| FN | Falso Negativo |
| FP | Falso Positivo |
| GPWLFC | Programa Global de Combate a crimes contra a vida selvagem e florestas (do inglês, <i>Global Programme for Combating Wildlife and Forest Crime</i>) |
| IBAMA | Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis |
| IQ | Instituto de Química |
| LIM | Limite de Discriminação |
| LPF | Laboratório de Produtos Florestais |
| MMA | Ministério do Meio Ambiente |
| NIPALS | Quadrados mínimos iterativos não lineares (do inglês, <i>Non-linear iterative partial least squares</i>) |
| Nº | Número |
| NIR | Infravermelho Próximo (do inglês, <i>Near Infrared</i>) |
| NIRS | Espectroscopia no infravermelho próximo (do inglês <i>Near Infrared Spectroscopy</i>) |
| OSC | Correção de sinal ortogonal (do inglês, <i>Orthogonal Signal Correction</i>) |
| PC | Componente Principal |
| PC1 | Primeira Componente Principal |
| PC2 | Segunda Componente Principal |
| PCA | Análise de Componentes Principais (do inglês, <i>Principal Component Analysis</i>) |
| PLS | Mínimos Quadrados Parciais (do inglês, <i>Partial Least Squares</i>) |
| PLS-DA | Análise Discriminante por Mínimos Quadrados (do inglês, <i>Discriminant Analysis Partial Least Squares</i>) |
| TEF | Taxa de Eficiência |

| | |
|-------|---|
| TFN | Taxa de Falso Negativo |
| TFP | Taxa de Falso Positivo |
| TUMP | Teor de Umidade Máximo Possível |
| TU | Teor de Umidade |
| UnB | Universidade de Brasília |
| UNODC | Nações Unidas sobre Drogas e Crimes (do inglês, <i>United Nations Office on Drugs and Crime</i>) |
| VL | Variável Latente |
| VN | Verdadeiro Negativo |
| VP | Verdadeiro Positivo |
| WWF | Fundo Mundial da Natureza (do inglês, <i>World Wide Fund for Nature</i>) |

LISTA DE TABELAS

| | |
|--|----|
| Tabela 1. Atribuições de bandas características às vibrações de moléculas de água em espectros NIR de madeira (Schwanninger et al., 2011). | 31 |
| Tabela 2. Parâmetros do modelo de discriminação de mogno (<i>Swietenia macrophylla</i>) construído com amostras na condição seca e aplicado às amostras de madeira seca. | 37 |
| Tabela 3. Parâmetros do modelo de discriminação de mogno (<i>Swietenia macrophylla</i>) construído com amostras na condição seca e aplicado às amostras de madeira recém-cortada. | 40 |
| Tabela 4. Parâmetros do modelo de discriminação de mogno (<i>Swietenia macrophylla</i>) construído na condição seca após aplicação do método OSC e aplicado às amostras de madeira seca. | 42 |
| Tabela 5. Parâmetros do modelo de discriminação de mogno (<i>Swietenia macrophylla</i>) construído na condição seca após aplicação do método OSC e aplicado às amostras de madeira recém-cortada. | 44 |
| Tabela 6. Quantidade de amostras selecionadas para o conjunto de treinamento e validação. | 51 |
| Tabela 7. Quantidade de amostras selecionadas dos conjuntos de treinamento e validação para as condições: úmida e seca. | 51 |
| Tabela 8. Valores de umidade de uma amostra das cinco espécies nas condições: úmida e depois de 1, 5, 10, 15 e 20 minutos no desidratador. | 53 |
| Tabela 9. Valores de umidade das quatro amostras das cinco espécies nas condições: úmida e depois de 20 e 30 minutos no desidratador. | 60 |
| Tabela 10. Parâmetros do modelo de discriminação para cedrinho (<i>Erisma uncinatum</i>) | 68 |
| Tabela 11. Parâmetros do modelo de discriminação para cedro (<i>Cedrela odorata</i>). | 69 |
| Tabela 12. Parâmetros do modelo de discriminação para mogno (<i>Swietenia macrophylla</i>). | 69 |
| Tabela 13. Parâmetros do modelo de discriminação para cedrinho (<i>Erisma uncinatum</i>) com seleção de amostras por DUPLEX. | 72 |
| Tabela 14. Parâmetros do modelo de discriminação para cedro (<i>Cedrela odorata</i>) com seleção de amostras por DUPLEX. | 72 |
| Tabela 15. Parâmetros do modelo de discriminação para mogno (<i>Swietenia macrophylla</i>) com seleção de amostras por DUPLEX. | 73 |

LISTA DE FIGURAS

- Figura 1.** Estrutura macroscópica do tronco de uma árvore: casca, floema, câmbio, albúrnio, cerne e medula. Adaptado de: Amabis & Martho, 1996. 3
- Figura 2.** Distinção de cerne e albúrnio pela cor. (A) Distintos pela cor (*Peltogyne* sp). (B) Indistintos pela cor (*Aspidosperma pyrifolium* Mart.). (C) Pouco distintos pela cor (*Schefflera macrocarpa* (Cham. & Schltdl.) Frodin). Fonte: Coradin et al., 2010. 5
- Figura 3.** Procedimento para desenvolvimento e validação do modelo PLS-DA. Legenda: (X_{cal}) matriz de espectros de treinamento, (y_{cal}) vetor de classes de treinamento para cada espécie (an = andiroba (*Carapa guianensis*), cd = cedrinho (*Erisma uncinatum*), ce = cedro (*Cedrela odorata*), cx = curupixá (*Micropholis melinoniana*) e mo = mogno (*Swietenia macrophylla*)), (X_{val}) matriz de espectros de validação, (y_{val}) vetor de classes de validação. 14
- Figura 4.** Espectrômetro portátil MicroNir™ (a esquerda) e a realização de uma medida em madeira sólida com o auxílio de um notebook (a direita) 25
- Figura 5.** Higrômetro digital usado para as medidas de umidade. Fonte: www.bangggod.com. 27
- Figura 6.** Variação dos valores de umidade das amostras de mogno (*Swietenia macrophylla*) recém-cortadas procedentes da Fazenda Celeiro Sementes (Luziânia, GO) em função do tempo. 28
- Figura 7.** Espectros sem pré-processamento de mogno (*Swietenia macrophylla*). Legenda: (—) espectros de mogno (*Swietenia macrophylla*) de madeira recém-cortada ou com umidade acima de 15% e (—) espectros de mogno (*Swietenia macrophylla*) seco ou com umidade entre 10 e 15% 29
- Figura 8.** Variação dos espectros médios de mogno (*Swietenia macrophylla*) com indicação dos principais comprimentos de onda referentes à água. 30
- Figura 9.** Escores de PC1 e PC2 estimados com dados centrados na média para as amostras de mogno (*Swietenia macrophylla*). Legenda: (▼) amostras de mogno (*Swietenia macrophylla*) de madeira recém-cortada e (*) amostras de mogno (*Swietenia macrophylla*) seco. 32
- Figura 10.** Espectros de mogno (*Swietenia macrophylla*) pré-processados com primeira derivada Legenda: (—) espectros de mogno (*Swietenia macrophylla*) de madeira recém-cortada e (—) espectros de mogno (*Swietenia macrophylla*) seco. 33

Figura 11. Escores dos PC1 e PC2 estimados após a aplicação da 1ª derivada e dados centrados na média para amostras de mogno (*Swietenia macrophylla*). Legenda: (▼) amostras de mogno (*Swietenia macrophylla*) de madeira recém-cortada e (*) amostras de mogno (*Swietenia macrophylla*) seco. 33

Figura 12. Espectros de mogno (*Swietenia macrophylla*) corrigidos com aplicação de OSC. Legenda: (—) espectros de mogno (*Swietenia macrophylla*) de madeira recém-cortada e (—) espectros de mogno (*Swietenia macrophylla*) seco. 34

Figura 13. Escores de PC1 e PC2 estimados após aplicação do OSC para amostras de mogno (*Swietenia macrophylla*). Legenda: (▼) amostras de mogno (*Swietenia macrophylla*) de madeira recém-cortada e (*) amostras de mogno (*Swietenia macrophylla*) seco. 35

Figura 14. Valores de classe estimados para os conjuntos de treinamento e validação para a discriminação de mogno (*Swietenia macrophylla*) seco. Legenda: andiroba (*Carapa guianensis*) (▼), cedrinho (*Erisma uncinatum*) (*), cedro (*Cedrela odorata*) (■), curupixá (*Micropholis melinoniana*) (+) e mogno (*Swietenia macrophylla*) (◇) e (— —) limite de discriminação. 37

Figura 15. Gráfico dos valores de T^2 de Hotelling e resíduos Q da discriminação de mogno (*Swietenia macrophylla*) seco. 38

Figura 16. Gráfico dos valores de T^2 de Hotelling e resíduos Q da discriminação de mogno (*Swietenia macrophylla*) de madeira recém-cortada. 39

Figura 17. Valores de classe estimados para os conjuntos de treinamento e validação para a discriminação de mogno (*Swietenia macrophylla*) de madeira recém-cortada. Legenda: andiroba (*Carapa guianensis*) (▼), cedrinho (*Erisma uncinatum*) (*), cedro (*Cedrela odorata*) (■), curupixá (*Micropholis melinoniana*) (+) e mogno (*Swietenia macrophylla*) (◇) e (— —) limite de discriminação. 40

Figura 18. Gráfico dos valores de T^2 de Hotelling e resíduos Q da discriminação de mogno (*Swietenia macrophylla*) seco após aplicação de OSC. 41

Figura 19. Valores de classe estimados para os conjuntos de treinamento e validação para a discriminação de mogno (*Swietenia macrophylla*) seco após aplicação do OSC. Legenda: andiroba (*Carapa guianensis*) (▼), cedrinho (*Erisma uncinatum*) (*), cedro (*Cedrela odorata*) (■), curupixá (*Micropholis melinoniana*) (+) e mogno (*Swietenia macrophylla*) (◇) e (— —) limite de discriminação. 42

Figura 20. Gráfico dos valores de T^2 de Hotelling e resíduos Q da discriminação de mogno (*Swietenia macrophylla*) de madeira recém-cortada após aplicação de OSC. 43

- Figura 21.** Valores de classe estimados para os conjuntos de treinamento e validação para a discriminação de mogno (*Swietenia macrophylla*) de madeira recém-cortada após aplicação do OSC. Legenda: andiroba (*Carapa guianensis*) (▼), cedrinho (*Erismia uncinatum*) (*), cedro (*Cedrela odorata*) (■), curupixá (*Micropholis melinoniana*) (+) e mogno (*Swietenia macrophylla*) (◇) e (— —) limite de discriminação. 44
- Figura 22.** Procedimento de preparo das amostras e obtenção dos espectros. 48
- Figura 23.** Desidratador de alimentos usado para secagem das amostras de madeira. Fonte: pardal.com.br/desidratador-peg-60. 50
- Figura 24.** Estufa de secagem com circulação/renovação de ar usada para secagem das amostras. Fonte: marconi.com.br. 52
- Figura 25.** Espectro médio de (A) andiroba (*Carapa guianensis*), (B) cedrinho (*Erismia uncinatum*), (C) cedro (*Cedrela odorata*), (D) curupixá (*Micropholis melinoniana*) e (E) mogno (*Swietenia macrophylla*) nas condições: molhado, seco, 5 minutos, 10 minutos e 20 minutos 54
- Figura 26.** Gráfico dos valores de T^2 de Hotelling e resíduos Q de (A) andiroba (*Carapa guianensis*), (B) cedrinho (*Erismia uncinatum*), (C) cedro (*Cedrela odorata*), (D) curupixá (*Micropholis melinoniana*) e (E) mogno (*Swietenia macrophylla*) para uma amostra testada após 20 minutos no desidratador. Legenda: andiroba (*Carapa guianensis*) (▼), cedrinho (*Erismia uncinatum*) (*), cedro (*Cedrela odorata*) (■), curupixá v (+) mogno (*Swietenia macrophylla*) (◇) e amostra testada (•). 56
- Figura 27.** Valores de classe estimados para os conjuntos de treinamento e para a amostra testada de (A) andiroba (*Carapa guianensis*), (B) cedrinho (*Erismia uncinatum*), (C) cedro (*Cedrela odorata*), (D) curupixá e (E) mogno (*Swietenia macrophylla*) após 20 minutos no desidratador. Legenda: andiroba (*Carapa guianensis*) (▼), cedrinho (*Erismia uncinatum*) (*), cedro (*Cedrela odorata*) (■), curupixá (*Micropholis melinoniana*) (+) e mogno (*Swietenia macrophylla*) (◇), amostra testada (•) e (— —) limite de discriminação. 58
- Figura 28.** Gráfico dos valores de T^2 de Hotelling e resíduos Q de (A) andiroba (*Carapa guianensis*), (B) cedrinho (*Erismia uncinatum*), (C) cedro (*Cedrela odorata*), (D) curupixá (*Micropholis melinoniana*) e (E) mogno (*Swietenia macrophylla*) para 4 amostras testadas após 20 minutos no desidratador. Legenda: andiroba (*Carapa guianensis*) (▼), cedrinho (*Erismia uncinatum*) (*), cedro (*Cedrela odorata*) (■), curupixá (*Micropholis melinoniana*) (+) mogno (*Swietenia macrophylla*) (◇) e amostra testada (•). 61

Figura 29. Valores de classe estimados para os conjuntos de treinamento e para 4 amostras testadas de (A) andiroba (*Carapa guianensis*), (B) cedrinho (*Erisma uncinatum*), (C) cedro (*Cedrela odorata*), (D) curupixá e (E) mogno (*Swietenia macrophylla*) após 20 minutos no desidratador. Legenda: andiroba (*Carapa guianensis*) (▼), cedrinho (*Erisma uncinatum*) (*), cedro (*Cedrela odorata*) (■), curupixá (*Micropholis melinoniana*) (+) e mogno (*Swietenia macrophylla*) (◇), amostra testada (●) e (— —) limite de discriminação. 62

Figura 30. Gráfico dos valores de T^2 de Hotelling e resíduos Q de (A) andiroba (*Carapa guianensis*), (B) cedrinho (*Erisma uncinatum*), (C) cedro (*Cedrela odorata*), (D) curupixá (*Micropholis melinoniana*) e (E) mogno (*Swietenia macrophylla*) para 4 amostras testadas após 30 minutos no desidratador. Legenda: andiroba (*Carapa guianensis*) (▼), cedrinho (*Erisma uncinatum*) (*), cedro (*Cedrela odorata*) (■), curupixá (*Micropholis melinoniana*) (+) mogno (*Swietenia macrophylla*) (◇) e amostra testada (●). 63

Figura 31. Valores de classe estimados para os conjuntos de treinamento e para 4 amostras testadas de (A) andiroba (*Carapa guianensis*), (B) cedrinho (*Erisma uncinatum*), (C) cedro (*Cedrela odorata*), (D) curupixá (*Micropholis melinoniana*) e (E) mogno (*Swietenia macrophylla*) após 30 minutos no desidratador. Legenda: andiroba (*Carapa guianensis*) (▼), cedrinho (*Erisma uncinatum*) (*), cedro (*Cedrela odorata*) (■), curupixá (*Micropholis melinoniana*) (+) e mogno (*Swietenia macrophylla*) (◇), amostra testada (●) e (— —) limite de discriminação. 64

Figura 32. Gráfico dos valores de T^2 de Hotelling e resíduos Q de (A) cedrinho (*Erisma uncinatum*), (B) cedro (*Cedrela odorata*) e (C) mogno (*Swietenia macrophylla*). Legenda: cedrinho (*Erisma uncinatum*) (▼), cedro (*Cedrela odorata*) (*) e mogno (*Swietenia macrophylla*) (■). 66

Figura 33. Valores de classe estimados para os conjuntos de treinamento e validação de (A) cedrinho (*Erisma uncinatum*), (B) cedro (*Cedrela odorata*) e (C) mogno (*Swietenia macrophylla*). Legenda: cedrinho (*Erisma uncinatum*) (▼), cedro (*Cedrela odorata*) (*) e mogno (*Swietenia macrophylla*) (■) e (— —) limite de discriminação. 67

Figura 34. Gráfico dos valores de T^2 de Hotelling e resíduos Q de (A) cedrinho, (B) cedro (*Cedrela odorata*) e (C) mogno (*Swietenia macrophylla*) com seleção de amostras por DUPLEX. Legenda: cedrinho (*Erisma uncinatum*) (▼), cedro (*Cedrela odorata*) (*) e mogno (*Swietenia macrophylla*) (■). 70

Figura 35. Valores de classe estimados para os conjuntos de treinamento e validação de (A) cedrinho (*Erisma uncinatum*), (B) cedro (*Cedrela odorata*) e (C) mogno (*Swietenia*

macrophylla) com seleção de amostras por DUPLEX. Legenda: cedrinho (*Erismia uncinatum*) (▼), cedro (*Cedrela odorata*) (*) e mogno (*Swietenia macrophylla*) (■) e (—) limite de discriminação.

1 INTRODUÇÃO E OBJETIVOS

O Brasil é um país de proporções continentais que abriga diferentes zonas climáticas e biomas. A variedade de biomas reflete na enorme riqueza da flora, o que resulta em uma das maiores biodiversidades do planeta. No Brasil, pode-se encontrar uma extensa variedade de espécies florestais produtoras de madeira. No entanto, a exploração e comercialização de madeiras ilegais contribuem para o crescimento contínuo das taxas de desmatamento das florestas Amazônica e as demais do globo terrestre (Pastore et al., 2011).

Atualmente, há um esforço de várias instituições internacionais para combater a exportação de madeira ilegal respeitando a legislação existente em cada país. Tal esforço tem como finalidade controlar, proibir ou combater a exploração ilegal e não seletiva de espécies florestais produtoras de madeira (Dormontt et al., 2015).

Apesar de todo o empenho, ainda existem a dificuldade e a carência em identificar de forma rápida e de maneira confiável a espécie florestal que está sendo inspecionada. Surge, assim, a necessidade de desenvolver ferramentas eficientes que auxiliem a identificação de espécies florestais. Uma das técnicas que estão sendo estudadas e adaptadas para esse fim é a espectroscopia no infravermelho próximo (NIRS, do inglês *Near Infrared Spectroscopy*).

A Tecnologia NIRS, uso da espectroscopia no infravermelho próximo acompanhada da quimiometria, destaca-se por ser rápida, não destrutiva, reproduzível, precisa, requerer mínimo preparo da amostra, dispor de equipamentos portáteis comerciais e exibir resultado da análise em tempo real. Trabalhos anteriores da equipe de trabalho do Laboratório de Produtos Florestais (LPF) do Serviço Florestal Brasileiro (SFB) e Instituto de Química da Universidade de Brasília (IQ/UnB) evidenciam o potencial dessa tecnologia para a discriminação de madeiras amazônicas: mogno (*Swietenia macrophylla* King.), cedro (*Cedrela odorata* L.), andiroba (*Carapa guianensis* Aubl.) e curupixá (*Micropholis melinoniana* Pierre), demonstrando também sua aplicabilidade com amostras de diferentes países (Pastore et al., 2011; Braga et al., 2011; Bergo et al., 2016).

No entanto, a madeira, material proveniente de um organismo vivo, apresenta sua complexidade. A madeira é um material higroscópico com estrutura porosa e exibe uma característica de contração e expansão do material acompanhada com mudanças notáveis das propriedades mecânicas e físicas (por exemplo, elasticidade, constante dielétrica e

etc), quando absorve e desorbe água (Tsuchikawa & Tsutsumi, 1998). Logo, a água livre e absorvida presente na madeira influencia o espectro NIR e isso dificilmente pode ser evitado.

A presença de água na madeira implica na variação do espectro NIR no que diz respeito ao deslocamento de linha de base e à intensidade nas bandas de absorção referentes às ligações O-H (Igne et al., 2014). Essa variabilidade interfere no desempenho dos modelos de discriminação e compromete a validade do modelo de treinamento restringindo sua aplicabilidade. Por isso, algumas estratégias são necessárias para corrigir essa interferência e aperfeiçoar os modelos.

Tendo em vista a constatação da interferência da umidade nos modelos de discriminação, este trabalho tem como objetivo avaliar métodos para sua correção nos espectros NIR de madeira. Para isso, o trabalho foi dividido em duas etapas. Inicialmente, trabalhou-se com a hipótese de que uma correção matemática poderia compatibilizar os espectros obtidos por amostras secas e úmidas. Para testar essa hipótese, foi avaliada uma ferramenta quimiométrica chamada de Correção Ortogonal de Sinal (OSC, do inglês *Orthogonal Signal Correction*) em uma tentativa de remover a informação referente à umidade. Em segunda hipótese, avaliou-se a correção da umidade por secagem da madeira em laboratório antes da aquisição dos espectros NIR, usando uma estufa de circulação de ar.

Ambas as alternativas tinham como alvo eliminar ou reduzir a interferência da umidade nos espectros e fazer com que amostras com umidade diferente daquelas utilizadas na fase de treinamento dos modelos pudessem ser analisadas nos modelos de discriminação por Análise Discriminante por Mínimos Quadrados (PLS-DA, do inglês *Discriminant Analysis Partial Least Squares*) das cinco espécies florestais produtoras de madeira: andiroba (*Carapa guianensis* Aubl.), cedrinho (*Erismia uncinatum*), cedro (*Cedrela odorata*), curupixá (*Micropholis melinoniana*) e mogno (*Swietenia macrophylla*).

2 REVISÃO BIBLIOGRÁFICA

2.1 MADEIRA

2.1.1 Estrutura anatômica

A madeira sempre teve um papel importante na economia humana. Desde os primórdios da humanidade, ela tem sido usada como um material estrutural para construções civis, móveis, ferramentas, energéticos e armas e mais recentemente, a característica renovável da madeira para produção de combustíveis tem sido destacada e um interesse sobre isso tem sido despertado. Por exemplo, no Dia Internacional das Florestas, uma celebração global promovida pelas Nações Unidas (UN, do inglês *United Nations*), o combustível de madeiras como uma alternativa sustentável em relação aos combustíveis fósseis, como o carvão, o petróleo e o gás natural foi destacado. De modo geral, a madeira representa um dos recursos naturais renováveis mais importantes.

Além disso, a madeira é uma das partes mais nobre e cobiçada das árvores e pode ser obtida do tronco, raízes, ramos de árvores e até mesmo de alguns arbustos. Por sua vez, o tronco da árvore é formado basicamente de casca e madeira, sendo então a maior fonte de obtenção de madeira. Com exceção da casca, todo o restante do tronco é caracterizado como madeira. O tronco da árvore é organizado nas seguintes estruturas macroscópicas: casca, floema, câmbio, lenho e medula. As partes do tronco são ilustradas na Figura 1.

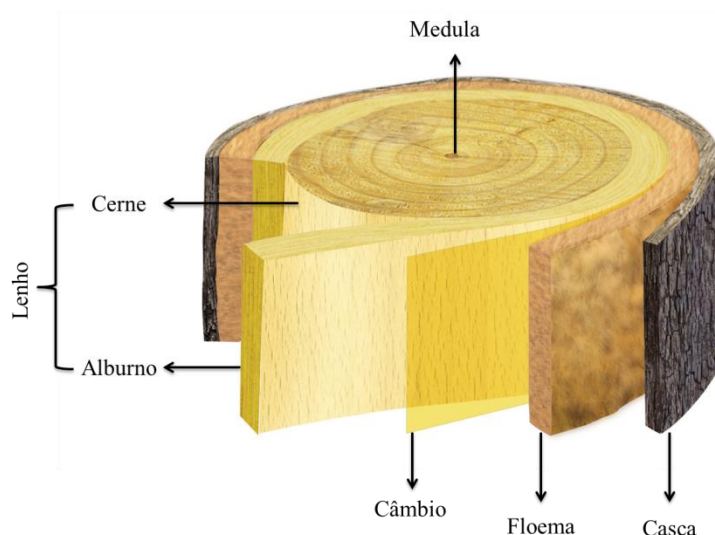


Figura 1. Estrutura macroscópica do tronco de uma árvore: casca, floema, câmbio, alburno, cerne e medula. Adaptado de: Amabis & Martho, 1996.

A casca é a parte exterior responsável pela proteção do tronco, o floema é formado pelos tecidos superficiais do tronco que são responsáveis pelo transporte da seiva. O câmbio é uma camada de células situada entre o xilema e o floema, e a medula é formada a partir das células que constituíram a zona de crescimento inicial que deu origem ao tronco. O lenho é a parte do tronco onde se extrai a madeira e é dividido em duas zonas: cerne e alburno.

A parte periférica do lenho, mais nova, é o alburno. Essa parte da madeira é fisiologicamente ativa e possui uma parte do tecido que está viva.

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 do lenho. Esse núcleo da madeira morta no centro do caule é conhecido como cerne, que fornece apenas suporte mecânico à árvore. Antes de morrerem, as células podem liberar substâncias corantes e resinas que aderem à parede das células. Tais substâncias são, muitas vezes, antissépticas e impedem a decomposição dessa parte do lenho (Amabis & Martho, 1996).

A iniciação da formação do cerne varia conforme a espécie e na mesma espécie, pode variar conforme o ambiente. Nas madeiras duras, em geral, o cerne é bem distinto do alburno, sendo este geralmente de coloração clara, quase bege, correspondente à parte viva da madeira. O cerne é a parte morta da madeira e varia de cor, pode ser marrom, negro, como o pau ferro (*Caesalpinia férrea* Mart. Ex Tul), violeta (*Dalbergia cearenses* Ducke), cor de brasa como o pau-Brasil (*Caesalpinia echinata* Lam.), dentre outras.

Nas madeiras duras ou moderadamente duras, depois que as células do alburno morrem, passam a integrar o cerne e no decorrer do tempo, o conteúdo das células sofre sucessivas alterações químicas, resultando uma gama de substâncias químicas diferentes quanto à natureza, responsáveis pelas tonalidades de coloração que o cerne apresenta (Richardson, 1993).

Nas madeiras leves, a distinção entre o cerne e o alburno é quase nula, devido à semelhança de coloração, em virtude da reduzida intensidade de alterações químicas que ocorrem no conteúdo celular nessas madeiras. A distinção de cerne e alburno pela cor é relativa, a Figura 2 ilustra diferentes madeiras com cerne e alburno muito e pouco distintos entre si.

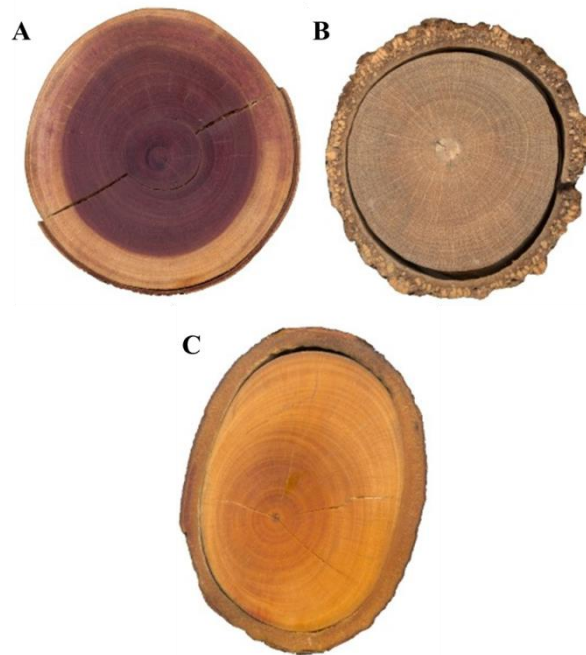


Figura 2. Distinção de cerne e alburno pela cor. **(A)** Distintos pela cor (*Peltogyne* sp ou roxinho). **(B)** Indistintos pela cor (*Aspidosperma pyrifolium* Mart ou peroba). **(C)** Pouco distintos pela cor (*Schefflera macrocarpa* (Cham. & Schltdl.) Frodin ou mandiocão).
Fonte: Coradin et al., 2010.

2.1.2 Composição química

A madeira, no que diz respeito a sua composição química, não apresenta diferenças consideráveis de uma espécie para outra. De modo geral, ela é composta por: carbono (50%), oxigênio (44%) e hidrogênio (6%). Podendo haver também pequenas quantidades de cálcio, potássio e magnésio (Pettersen, 1998).

A composição química da madeira é complexa. O tecido da madeira é constituído de vários componentes químicos que são distribuídos de maneira não uniforme em resultado de sua estrutura anatômica. Como consequência, o comportamento químico da madeira não pode simplesmente ser descrito pelas propriedades dos elementos que a constituem.

A grande parte da madeira é composta de materiais com alto peso molecular, como: polímeros de celulose, hemicelulose e lignina, podendo ser apropriadamente descrita como um emaranhado de vários polímeros. A separação e o isolamento desses polímeros sem que haja modificação significativa neles são desafiadores e pesquisas ainda precisam ser desenvolvidas para elucidar completamente suas naturezas e propriedades.

Segundo Browning et al. (1963), os componentes que estão normalmente presentes na madeira podem ser classificados quimicamente em seis grupos diferentes. O primeiro deles é o de carboidratos que são representados principalmente por polissacarídeos e incluem celulose, hemicelulose e amido. A celulose é o componente majoritário e corresponde aproximadamente à metade do peso da madeira, os açúcares podem estar presentes na seiva e nos tecidos, mas tem uma presença não significativa em tecidos mais maduros.

O segundo grupo é composto por substâncias fenólicas, as quais são materiais aromáticos caracterizados pela presença do grupo hidroxil do fenol. É um grupo diversificado de substâncias que compreende cerca de 20 a 30% dos tecidos e a lignina faz parte dessa classificação. No terceiro grupo, encontram-se os terpenos, esses incluem componentes voláteis. Estão presentes cerca de 5% em madeiras mais macias e quase inexistentes em madeiras mais duras. Por fim, o quarto grupo é de ácidos alifáticos; o quinto, de proteínas e o sexto, de constituintes inorgânicos.

2.1.3 Umidade e propriedades físicas da madeira

Assim como muitos materiais provenientes da natureza, madeira é um material higroscópico; ou seja, a madeira absorve a água presente no ambiente. Essa troca de água entre a madeira e o ar depende de alguns fatores, como: umidade relativa, temperatura do ambiente e quantidade atual de água presente na madeira. Essa relação de umidade influencia as propriedades das madeiras e também o seu desempenho. Muitos dos desafios do uso da madeira como material na engenharia estão relacionados com a variação de umidade e quantidade de água na madeira. Neste tópico, serão discutidas as propriedades físicas macroscópicas da madeira com ênfase no teor de umidade.

Muitas propriedades físicas e mecânicas da madeira dependem do teor de umidade da madeira. O teor de umidade (TU) é geralmente expresso pela porcentagem que pode ser calculada como na equação 1 abaixo:

$$TU = \frac{m_{\text{água}}}{m_{\text{madeira}}} \times 100\% \quad (1)$$

Onde, $m_{\text{água}}$ é a massa da quantidade de água na madeira e m_{madeira} é a massa da madeira total. Segundo a ABNT NBR 14929/2017, o teor de umidade da madeira corresponde à relação entre a massa da água nela contida e a massa da madeira seca, dado pela equação 2:

$$TU = \frac{m_i - m_s}{m_s} \times 100\% \quad (2)$$

Onde, m_i é a massa inicial da madeira, em gramas, e m_s é a massa da madeira seca, em gramas. Na fabricação dos corpos de prova, devem ser utilizadas ferramentas afiadas para evitar a chamada “queima” das suas faces, o que pode provocar uma perda de água imediata, prejudicial à determinação da real umidade da amostra. Para determinar a massa da madeira seca, o corpo de prova é colocado em uma câmara de secagem, com temperatura máxima de $103 \pm 2^\circ\text{C}$. Durante a secagem, a massa do corpo de prova deve ser medida a cada 6 h, até que ocorra uma variação, entre duas medidas consecutivas, menor ou igual a 0,5% da última massa medida. Essa massa será considerada como massa seca (m_s).

Uma madeira recém-serrada em que as paredes das células estão completamente saturadas com água normalmente é chamada de madeira verde. O teor de umidade de madeira verde pode variar cerca de 30% a mais de 200%. Em madeiras verdes macias, o teor de umidade do alburno é geralmente maior do que o do cerne e a diferença de umidade entre o cerne e o alburno depende da espécie.

A água pode existir na madeira na forma de água livre (água líquida ou vapor de água na lumina e nas cavidades das células) e como água de constituição (água presente nas paredes celulares por forças de atração intermolecular). Durante o processo de secagem, a água é liberada por evaporação e a primeira água a sair é a água livre que se encontra nos espaços vazios das células. A saída de água é maior nas seções transversais, ou seja, no topo das peças, do que nas superfícies laterais como cantos e faces da madeira.

Quando a água livre é perdida, pode-se medir o teor de umidade referente à quantidade de água presente nas paredes celulares, que é chamado de Ponto de Saturação das Fibras (PSF). Ele varia de espécie para espécie, sendo:

- a. 32 a 35% para espécies folhosas com porosidade difusa, sem cerne distinto;
- b. 22 a 24% para espécies coníferas e folhosas com porosidade em anel, ambas sem cerne distinto.

Na prática, o PSF é considerado como o teor de umidade acima do qual as propriedades físicas e mecânicas da madeira não mudam em função do teor de umidade e o PSF médio usado é de 30% (Glass & Zelinka, 2010).

Conceitualmente, o ponto de saturação de fibra distingue os dois tipos de água mantidos na madeira. No entanto, existem variações entre a quantidade de água de constituição e água livre numa mesma madeira. Por exemplo, em um mesmo pedaço de madeira, é possível que em determinada região, o lúmen não apresente moléculas de água

e a parede da célula esteja parcialmente seca, enquanto em uma outra região, as paredes celulares podem estar completamente saturadas e o lúmen parcialmente cheio de água.

Em contrapartida, o teor de umidade em que tanto o lúmen como as paredes das células vão estar completamente saturados de água é chamado de Teor de Umidade Máximo Possível (TUMP). Esse teor é determinado majoritariamente pela gravidade básica específica (G_b), determinada pela massa seca e o volume verde da madeira, pela equação 3:

$$TUMP = \frac{100 \times (1,54 - G_b)}{1,54G_b} \quad (3)$$

2.1.4 Métodos de identificação de madeira

Tendo em vista a importância do comércio madeireiro e a intensa exploração desse recurso florestal, existe uma preocupação internacional em controlar o comércio de madeira, principalmente, pela restrição ou proibição da importação de determinadas espécies florestais produtoras de madeira.

Restrições no comércio de madeira foram impostas, principalmente, através da Convenção sobre o Comércio Internacional de Espécies da Flora e Fauna Selvagens em Perigo de Extinção (CITES, do inglês, *Convention on International Trade in Endangered Species of Wild Flora and Fauna*) que classificou as espécies em três apêndices a depender do grau de proteção requerido.

O Apêndice I é o mais restritivo e proíbe o comércio internacional de espécies em ameaçadas de extinção, o comércio só é permitido em circunstâncias excepcionais. O Apêndice II lista as espécies que não estão ameaçadas de extinção atualmente, mas requerem um controle do comércio para evitar o aumento da exploração e futura extinção. Por fim, o Apêndice III lista as espécies que são controladas em pelo menos um país do mundo, aquele que requer supervisão do comércio por parte dos outros países signatários (Durmontt et al., 2015).

O método mais difundido e empregado para a identificação da madeira, desprovida de qualquer material botânico, é a anatomia da madeira que compara os caracteres anatômicos e morfológicos da madeira examinada com a madeira de padrões depositados em xilotecas registradas (Souza et al., 1997). A identificação da madeira via sua anatomia é o método mais antigo e por isso, possui um banco de dados com amostras de referência de aproximadamente 14.000 espécies, disponíveis nas várias coleções de madeira espalhadas pelo mundo. Desenvolvimento recente nessa área é o reconhecimento

de espécies utilizando coleções de imagens digitais como referência, por exemplo a *Vision Machine* (UNODC, 2016).

Contudo, ainda é necessária elevada experiência do analista para a aplicação do método com o nível de confiança necessário para realizar uma apreensão de carga ilegal, uma vez que as espécies frequentemente apresentam estrutura da madeira muito similares e facilmente confundíveis (Yu et al., 2017).

Outras metodologias estão sendo consideradas para identificar madeira. Por exemplo, a dendrocronologia foca no estudo dos anéis de crescimento, os quais contêm informações sobre o ambiente durante a fase de crescimento. A dendrologia tem potencial para dar informações específicas sobre a idade da árvore e sua localização geográfica. A identificação individual é uma possibilidade, uma vez que a informação encontrada nos anéis de crescimento pode coincidir com a informação de outras partes da madeira do mesmo indivíduo. No entanto, ainda é preciso avaliar a consistência desses padrões e a margem de erro é significativa. Métodos tradicionais de dendrocronologia podem ser aplicados somente em árvores que formam anéis de crescimento periódicos. A maioria das madeiras tropicais não produzem esses anéis (UNODC, 2016).

Ademais, a espectrometria de massas tem sido usada para identificação de espécies florestais. A avaliação de fitoquímicos estabelecidos no cerne pode ser feita por espectrometria de massas e pode ser analisada também a diminuição relativa de compostos químicos ao longo do tempo com o processamento e uso da madeira. Kite et al. (2010) usou cromatografia líquida acoplada a espectrometria de massas para diferenciar *Dalbergia nigra*, espécie listada pela CITES no Apêndice I, de 15 congêneros, e de forma similar, Lancaster e Espinoza (2012) e Espinoza et al. (2015) usaram análise direta em tempo real com espectrometria de massas por tempo de voo para diferenciar 12 espécies *Dalbergia* e 8 espécies similares. Recentemente, seis espécies do gênero *Cedrela*, três das quais relacionadas pela CITES (*C. fissilis*, *C. odorata* e *C. angustifolia*), foram coletadas em 11 localidades da Bolívia e identificadas usando DART-TOFMS (Análise direta em tempo real no espectrômetro de massas, do inglês *Direct Analysis in Real Time Mass Spectrometry*) (Paredes Villaneuva et al., 2018).

A análise de isótopos estáveis em madeira informa a procedência geográfica. À medida que os fotoquímicos são sintetizados, eles incorporam isótopos estáveis em função da disponibilidade no ambiente. Usando um ou mais isótopos, incluindo bioelementos, enxofre, estrôncio e outros, pode-se ter uma assinatura do isótopo daquela área. Os estudos para a aplicação do método na identificação de madeira ainda são

poucos, mas em 2011, WWF publicou um projeto que dava mais detalhes sobre a habilidade dos isótopos estáveis distinguirem a região geográfica de *Tectona grandis* e *Swietenia* (Förstel et al., 2011). Entre as desvantagens do método constam: não ser usada para identificar as espécies florestais, disponibilidade limitada de dados de referência, poucos laboratórios oferecem o serviço comercial, significando análise de preço elevado (Nature Economy and People Connected, 2017).

O método por DNA busca identificar a espécie de um indivíduo baseado na variação da região dos genes específicos. De modo geral, tenta-se imprimir um código de barras de DNA para cada indivíduo. O potencial desse método foi demonstrado usando uma região de genes para identificar espécies da família *Mahogany* (Muellner et al., 2011). No entanto, nem sempre o DNA extraído da madeira pode ser sequenciado em vários fragmentos; a idade, métodos de secagem, a presença de mofo, entre outros fatores, pode afetar a possibilidade de extrair um DNA usável. Além disso, o método é caro e é necessária uma biblioteca de referências atualizada.

As metodologias citadas apresentam vantagens, mas precisam de equipamentos caros para sua realização e é preciso também que a amostra seja encaminhada a um laboratório para análise. Em nenhum dos métodos citados surge a possibilidade de identificar a madeira em campo. Por isso, o estudo e aplicação da tecnologia NIRS tem ganhado espaço e visibilidade. Essa tecnologia permite o uso de equipamentos portáteis para realização da análise em campo e em tempo real.

A tecnologia NIRS, espectroscopia associada à análise multivariada dos espectros, permite a aquisição direta de medidas de reflectância que carregam informação dos grupos funcionais presentes nas moléculas de alto peso (celulose, hemicelulose e lignina) e de menor peso molecular (extrativos) da madeira (Tsuchikawa, 2007). Além disso, trazem informações físicas e anatômicas, referentes à distribuição desses constituintes químicos na superfície, tornando-se uma “impressão digital química” de cada espécie florestal produtora de madeira. Para aplicação em identificação de madeira, NIRS foi capaz de discriminar espécies de diferentes gêneros e do mesmo gênero (Pastore et al., 2011; Braga et al., 2011; Soares et al., 2017).

É importante ressaltar também que, devido à potencialidade demonstrada nos últimos anos pela NIRS aliada a métodos quimiométricos, essa tecnologia foi inserida como um dos métodos recomendados pelo guia de boas práticas para a identificação de madeira para fins forenses, publicado no ano de 2016 pelo escritório das Nações Unidas sobre Drogas e Crimes (UNODC, do inglês *United Nations Office on Drugs and Crime*)

do Programa Global de Combate a Crimes Contra a Vida Selvagem e Florestas (GPWLFC, do inglês *Global Programme for Combating Wildlife and Forest Crime*).

2.1.5 Espectroscopia no infravermelho próximo

A espectroscopia estuda as transições entre estados energéticos de átomos ou moléculas por meio da interação destes com a radiação eletromagnética. A interação entre matéria e radiação se estende ao longo de todo espectro eletromagnético e as transições observadas na região do infravermelho são as vibracionais, embora em alguns casos transições rotacionais sejam também observadas na região do infravermelho distante (Howarth, 2003). A região do infravermelho pode ser dividida em infravermelho próximo, médio e distante.

A espectroscopia de infravermelho próximo (NIRS, do inglês *Near Infrared Spectroscopy*) é um tipo de espectroscopia vibracional que utiliza a energia do fóton em um intervalo de $2,65 \times 10^{-19}$ a $7,96 \times 10^{-20}$ J, o que corresponde ao intervalo de comprimento de onda de 750 nm a 2500 nm (número de onda: 13.300 cm^{-1} a 4.000 cm^{-1}) (Pasquini, 2003). A técnica tem por objetivo analisar uma amostra a fim de adquirir informações qualitativas e/ou quantitativas provenientes da interação das ondas eletromagnéticas de infravermelho próximo com seus constituintes químicos.

O método analítico que resulta da utilização da região no infravermelho próximo reflete as suas características mais importantes: rápido, não destrutivo, não invasivo, com alto poder de penetração do feixe de radiação, aplicação quase universal (qualquer molécula contendo ligações C-H, N-H, S-H ou O-H), com a mínima preparação de amostras (Williams & Norris, 2001) e não gera resíduos químicos. No entanto, a técnica possui desvantagens que envolvem baixas seletividade e sensibilidade, dificuldade de atribuição do sinal a ligações específicas e a necessidade de grande quantidade de amostras para construção do modelo de calibração, dependendo da aplicação, o que conseqüentemente envolve custos, trabalho e tempo (Blanco & Villarroya, 2002). O uso do NIRS para análises químicas requer uma abordagem multivariada para a calibração.

2.2 ANÁLISE MULTIVARIADA

A obtenção dos resultados de um experimento químico, muitas vezes, envolve a análise de um grande número de variáveis. A extração da informação desses resultados deve levar em consideração o tipo de informação referente às variáveis. Por exemplo, pode ser que apenas um pequeno número das variáveis possua informações químicas

relevantes, enquanto a maioria das variáveis adiciona pouco ou nada à interpretação dos dados em termos químicos (Neto et al., 1997). Isso pode ser feito por meio da análise multivariada de dados.

A análise multivariada pode ser usada para estabelecer uma relação entre as medidas instrumentais e a propriedade de interesse da amostra. Neste caso, os termos espectro e espécie da madeira serão utilizados para se referir às medidas instrumentais e propriedade das amostras, respectivamente.

Os espectros podem ser obtidos em uma quantidade N de medidas e organizados em uma matriz \mathbf{X} , enquanto a informação sobre a espécie da madeira é organizada em um vetor \mathbf{y} ou uma matriz \mathbf{Y} . A seguir, vemos como a matriz \mathbf{X} é montada:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{1k} \\ x_{21} & x_{22} & x_{2k} \\ x_{n1} & x_{n2} & x_{nk} \end{bmatrix}$$

Os espectros obtidos, um para cada amostra, são organizados numa matriz $\mathbf{X}_{(n \times k)}$, onde n representa cada amostra e cada amostra é disposta em uma linha da matriz. Já as colunas representam os comprimentos de onda, que são as variáveis k independentes, ou seja, cada coluna é um comprimento de onda diferente que traz uma informação de absorvância diferente, que não está correlacionada com a dos outros comprimentos de onda. Uma variável é independente quando a medida não depende de nenhuma outra variável.

Ao analisar a matriz \mathbf{X} , podemos dizer que para a sétima amostra, x_{71} , x_{72} , x_{7k} são as absorvâncias registradas para os comprimentos de onda 1, 2 e k , respectivamente. Cada coluna de \mathbf{X} corresponde a um comprimento de onda específico. Portanto, x_{71} , x_{81} , x_{91} , são respostas no primeiro comprimento de onda para as amostras 7, 8 e 9, respectivamente.

Já a propriedade de interesse é um outro conjunto de dados organizado na matriz \mathbf{Y} ou pelo vetor \mathbf{y} . Caso haja mais de uma variável, ou seja, caso haja mais de uma propriedade que se deseja determinar das amostras, por exemplo: concentração de compostos fenólicos e lignina da madeira, os dados são plotados numa matriz. Em contrapartida, se há apenas uma única variável, tem-se o vetor \mathbf{y} e o total de elementos desse vetor é igual a n , isto é, o número de amostras (Ferreira et al., 1999). No caso de classificação de espécies de madeira, usa-se o vetor \mathbf{y} para organizar os dados, uma vez que existe uma única variável que se deseja prever: se uma amostra pertence ou não a uma determinada espécie.

Existem vários métodos de análise multivariada, com finalidade diferente entre si e por isso, os métodos são escolhidos de acordo com os objetivos da pesquisa. Neste trabalho, optou-se por utilizar apenas modelos de Análise Discriminante por Mínimos Quadrados Parciais (do inglês, *Discriminant Analysis Partial Least Squares* ou PLS-DA).

2.2.1 Seleção de amostras e desenvolvimento do modelo PLS-DA

Toda calibração multivariada utiliza modelos matemáticos para estabelecer uma relação entre uma propriedade que possa ser monitorada com alguma outra propriedade de interesse. O método de mínimos quadrados parciais é um modelo baseado em variáveis latentes (fatores), onde cada fator é definido como uma combinação linear das variáveis originais das matrizes \mathbf{X} e vetor \mathbf{y} .

O processo geral consiste nas etapas de treinamento e validação. Na etapa de treinamento, são selecionadas amostras para estabelecer uma relação entre \mathbf{X} e \mathbf{Y} por meio de modelagem matemática. 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. De modo geral, a etapa de treinamento constrói o modelo matemático que explica a relação existente entre as medidas instrumentais e a propriedade de interesse e a etapa de validação, como o próprio nome diz, valida o modelo matemático. Depois de concluir essas duas etapas, o modelo pode ser utilizado para previsão de novas amostras daquele mesmo material nas mesmas condições das amostras utilizadas para a etapa de treinamento.

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 dos dados. Diferentes critérios são propostos para essa seleção para que haja uma uniformidade no espaço amostral (Honorato et al., 2007). Nesse trabalho, um desses critérios foi testado e avaliado: o algoritmo Duplex. Esse algoritmo se assemelha ao algoritmo chamado de Kennard-Stone.

O algoritmo Kennard-Stone começa selecionando as duas amostras com maior distância Euclidiana entre si no espaço \mathbf{X} . Para cada uma das amostras restantes, calcula-se a distância mínima com respeito às amostras já selecionadas. Depois disso, a amostra com a maior distância mínima é retida, e o procedimento é repetido até que um determinado número de amostras seja selecionado (Kennard & Stone, 1969). A abordagem Kennard-Stone, dependendo da configuração dos dados, pode levar a um

sobreajuste nos dados. Por isso, uma modificação no algoritmo foi proposta pelo próprio Kennard, resultando em um novo algoritmo chamado de Duplex.

O Duplex começa como o algoritmo Kennard-Stone original computando a maior distância entre as amostras e selecionando as amostras mais distantes. A diferença agora é que o algoritmo continua adicionando outras duas amostras com a segunda maior distância no conjunto de validação. Assim, sucessivamente, as amostras mais diversas são adicionadas ou no conjunto de treinamento ou no conjunto de validação, de acordo com o critério *maximin* (distâncias máxima-mínimas), até o raio de divisão desejado e o número de amostras requisitado para cada conjunto ser obtido (Westad et al., 2015).

Assim que as amostras são selecionadas para os conjuntos das etapas de treinamento e validação, o modelo é desenvolvido e posteriormente validado. Para isso, é preciso seguir algumas etapas. As etapas foram descritas, com base na classificação de cinco espécies de madeira, no fluxograma a seguir na Figura 3:

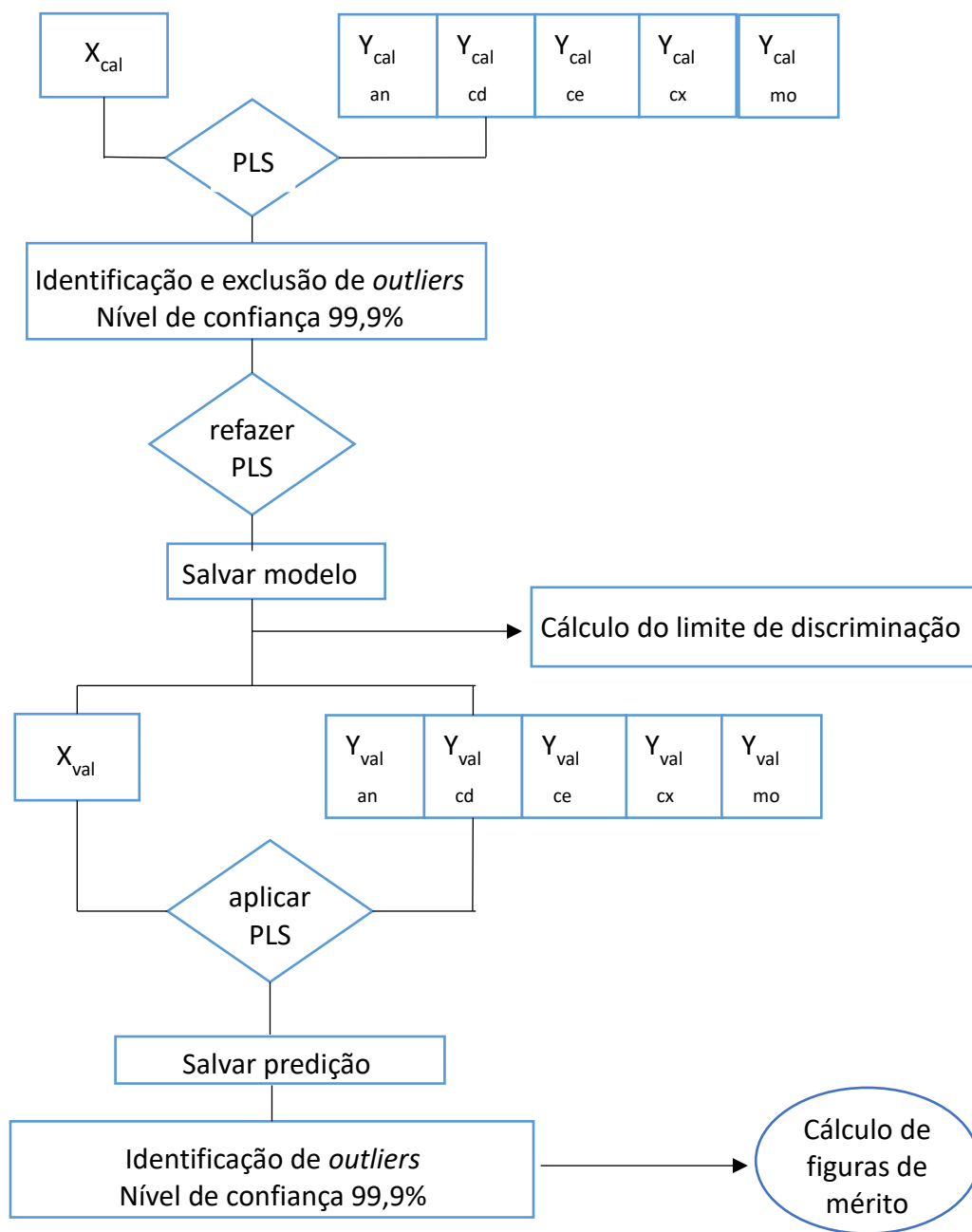


Figura 3. Procedimento para desenvolvimento e validação do modelo PLS-DA. Legenda: (X_{cal}) matriz de espectros de treinamento, (y_{cal}) vetor de classes de treinamento para cada espécie (an = andiroba (*Carapa guianensis*), cd = cedrinho (*Erisma uncinatum*), ce = cedro (*Cedrela odorata*), cx = curupixá (*Micropholis melinoniana*) e mo = mogno (*Swietenia macrophylla*)), (X_{val}) matriz de espectros de validação, (y_{val}) vetor de classes de validação.

No modelo PLS-DA, na fase de treinamento são utilizadas amostras cujas classes são conhecidas e, em cada modelo, o vetor y é composto por valores de 0 e 1, onde o

valor 1 é atribuído às amostras que pertencem a classe que se pretende discriminar e o valor 0, atribuído às amostras pertencentes as outras classes. Na sequência, a decomposição dos dados é realizada pelas equações 4 e 5 abaixo (da Silva et al., 2014):

$$\mathbf{X} = \sum_{a=1}^A \mathbf{t}_a \mathbf{p}'_a + \mathbf{E} \quad (4)$$

$$\mathbf{y} = \sum_{a=1}^A \mathbf{t}_a q'_a + \mathbf{f} \quad (5)$$

Onde, \mathbf{t}_a é o vetor de escores, p_a e 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, e o símbolo apóstrofo representa a operação de transposição. Para a otimização do modelo, diversos pré-processamentos podem ser utilizados, os quais consistem dos mesmos aplicados a problemas quantitativos com Regressão por Mínimos Quadrados Parciais (do inglês, *Partial Least Squares Regression* ou PLSR). O número de variáveis latentes a , assim como o melhor método de pré-processamento da matriz \mathbf{X} , é usualmente determinado em PLS-DA através do modelo que apresenta o menor valor de erro de classificação empregando validação cruzada (CVCE, do inglês *Cross Validation Classification Error*). Contudo, o CVCE não penaliza amostras que apresentam elevados erros na estimativa do valor classe \mathbf{y} , desde que as amostras sejam corretamente classificadas. No entanto, elevados erros nos valores de classe estimados podem ser uma indicação de que a amostra apresenta características diferentes das demais amostras do conjunto de treinamento ou da presença de uma amostra com um erro nos dados instrumentais, situações que caracterizam uma amostra anômala (do inglês, *outlier*).

Por outro lado, o valor da raiz quadrada do erro médio quadrático de validação cruzada (RMSECV, do inglês *Root Mean Square Error of Cross Validation*), frequentemente usado em PLSR, permite a otimização do modelo PLS-DA considerando a minimização dos erros de estimativa dos valores de classe, o qual, a princípio, tende a proporcionar uma maior separação dos valores estimados para a classe discriminada ($y = 1$) em relação às outras classes ($y = 0$). Portanto, considerando esse aspecto, neste trabalho foi empregado o RMSECV como critério para a otimização dos modelos PLS-DA.

Outro aspecto relevante na otimização dos modelos PLS-DA é a identificação e exclusão de amostras anômalas. Com base no trabalho de Borin & Poppi (2004) e da Silva et al. (2014), as amostras anômalas foram identificadas no conjunto de treinamento segundo os seguintes critérios:

- Amostras que apresentaram valores dos parâmetros estatísticos de T^2 de Hotelling e resíduo espectral Q acima dos limites de 99,9% foram identificadas como anômalas e removidas do conjunto de treinamento;

- Ao mesmo tempo, amostras que apresentem resíduos Student que excedem o valor crítico do test-t descrito na ASTM E1655-05 com correção de vies proposta por da Silva et al. (2014a) no nível de 99,9% de confiança, também serão identificadas como anômalas e excluídas do conjunto de treinamento (da Silva et al., 2014b; Annual Book of ASTM Standards, 2012).

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 é adaptado 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 6 e 7:

$$y_{limite\ superior, classe1} = 1 \pm vies_{class1} + \left(t_{99,v} RMSEC_{vies} \sqrt{1 - \hat{h}_c} \right) \quad (6)$$

$$y_{limite\ inferior, classe0} = 0 \pm vies_{class0} - \left(t_{99,v} RMSEC_{vies} \sqrt{1 - \hat{h}_c} \right) \quad (7)$$

Onde \hat{h}_c é o valor médio da influência (do inglês *leverage*) observada nas amostras de treinamento, $RMSEC_{vies}$ (do inglês *Root Mean Square Error of Calibration*) é a raiz quadrada do erro médio quadrático do conjunto de treinamento com correção de vies, $vies_{classe0}$, $vies_{classe1}$ são os vieses estimados para as classes 0 e 1, respectivamente, $t_{99,v}$ é o valor tabelado da distribuição de t-Student, com 99,9 % de confiança e $n-a-2$ graus de liberdade e n é o número de amostras de treinamento. Portanto, para o conjunto de validação, são consideradas amostras anômalas em relação à estimativa dos valores de classe aquelas que apresentaram valor de y maior do que $y_{limite\ superior, classe 1}$ ou inferior que $y_{limite\ inferior, classe 0}$. Em outras palavras, uma amostra de validação i que apresentar valor de classe (y_i) 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 99,9 %, serão anômalas.

Após otimização do modelo, um limite de discriminação é calculado com base na dispersão dos valores estimados de y para as amostras de treinamento de forma a minimizar a ocorrência de erros positivos falsos e negativos de acordo com o teorema de Bayes (Barker & Rayens, 2003; Botelho et al., 2015; Ferreira, 2015). Conforme definido anteriormente, é considerada a discriminação das amostras em duas classes, uma delas contendo as amostras da espécie que será discriminada, a qual será atribuído o valor de classe $y=1$ (classe A), e a outra contendo todas as amostras das demais espécies, as quais terão valor de classe $y=0$ (classe B). A partir dos valores estimados para o conjunto de

treinamento são estimadas as probabilidades *a priori* ($P(A)$ e $P(B)$) e as funções de densidade de probabilidade ($p(\hat{y}_i|A)$ e $p(\hat{y}_i|B)$) de cada classe. Considerando que as distribuições dos valores de classe estimados para as classes A e B se aproximam de uma distribuição normal, esses parâmetros podem ser definidos como (Ferreira, 2015):

$$P(A) = \frac{I_A}{I_A + I_B} \quad (8)$$

$$P(B) = \frac{I_B}{I_A + I_B} \quad (9)$$

$$p(\hat{y}_i|A) = \frac{1}{s_A\sqrt{2}} e^{\frac{-1}{2}\left(\frac{\hat{y}_i - \hat{y}_A}{s_A}\right)^2} \quad (10)$$

$$p(\hat{y}_i|B) = \frac{1}{s_B\sqrt{2}} e^{\frac{-1}{2}\left(\frac{\hat{y}_i - \hat{y}_B}{s_B}\right)^2} \quad (11)$$

A partir desses parâmetros, a probabilidade de uma amostra i pertencer às classes A ou B podem ser determinadas, respectivamente, por (Botelho et al., 2015; Ferreira, 2015):

$$P(\hat{y}_i|A) = \frac{p(\hat{y}_i|A)P(A)}{p(\hat{y}_i|A)P(A) + p(\hat{y}_i|B)P(B)} \quad (12)$$

$$P(\hat{y}_i|B) = \frac{p(\hat{y}_i|B)P(B)}{p(\hat{y}_i|A)P(A) + p(\hat{y}_i|B)P(B)} \quad (13)$$

De acordo com a regra de Bayes uma amostra é atribuída à classe A se $P(\hat{y}_i|A) > P(\hat{y}_i|B)$, caso contrário a amostra é atribuída à classe B. Outra maneira de tomar essa decisão é através da determinação do limite de discriminação, o qual é obtido através do valor de y no qual $P(\hat{y}_i|A) = P(\hat{y}_i|B)$. Considerando que o denominador das equações 9 e 10 são iguais, a determinação do limite se simplifica a (Ferreira, 2015):

$$p(\hat{y}_i|A)P(A) = p(\hat{y}_i|B)P(B) \quad (14)$$

O limite de discriminação é então obtido pela substituição das equações 8 a 11 na equação 14, substituição dos valores experimentais e determinação do valor de y . Detalhes dessas operações estão no anexo.

É importante destacar que se as probabilidades de ocorrência das classes A ou B forem iguais ($P(A)=P(B)$), as equações 12 e 13 passarão a depender apenas das funções de densidade de probabilidade de A e B. Na prática, em muitas situações o conjunto de treinamento não permite a obtenção de amostras que sejam boas aproximações das probabilidades de ocorrência das classes estudadas em amostras futuras. Nesses casos, é aconselhável a consideração de que $P(A)=P(B)$.

Portanto, uma amostra genérica i é identificada como pertencente à classe 1 caso seu valor de classe estimado (y_i) for maior que o valor do limite de discriminação do

modelo PLS-DA correspondente. Caso contrário, essa amostra será identificada como pertencendo à classe 0, que contém as amostras de todas as demais classes modeladas.

2.2.2 Determinação das figuras de mérito do modelo PLS-DA

A validação do modelo de discriminação é discutida com base no cálculo das figuras de mérito, descrito por Botelho et al. (2015) e definidas a seguir. A Taxa de Falso Positivo (TFP) representa o percentual de amostras que apresentaram erros falsos positivos e é calculada como a relação entre o número absoluto de falsos positivos (FP) e a soma do número absoluto de erros falsos positivos (FP) e verdadeiros negativos (VN) multiplicada por 100, representada pela equação 15:

$$TFP = \frac{FP}{FP+VN} 100 \quad (15)$$

Por outro lado, a Taxa de Falsos Negativos (TFN) representa o percentual de amostras que apresentou erros falsos negativos, sendo calculada como a relação entre o número absoluto de falsos negativos (FN) e a soma do número absoluto de erros falsos negativos (FN) e verdadeiros positivos (VP) multiplicada por 100, representada pela equação 16:

$$TFN = \frac{FN}{FN+VP} 100 \quad (16)$$

Por fim, a Taxa de Eficiência (TEF) do modelo de discriminação é obtida pela diferença entre o valor de 100% e a soma de TFN e TFP, representada pela equação:

$$TEF = 100 - (TFN + TFP) \quad (17)$$

2.2.3 Correção Ortogonal de Sinal

A Correção Ortogonal de Sinal (OSC, do inglês *Orthogonal Signal Correction*) é um dos métodos ortogonais existentes que são baseados na teoria de que a matriz de dados instrumentais \mathbf{X} é a soma de dois subespaços, entre os quais apenas um contém informação útil para o modelo. Existem duas maneiras para estimar o subespaço “não-informativo”: (1) através do espaço ortogonal ao vetor \mathbf{y} que contém a propriedade de interesse e (2) estimando o espaço onde os fatores externos tem uma influência significativa. No caso do OSC, calcula-se os parâmetros ortogonais de correção de sinal para o pré-processamento dos dados espectrais antes de desenvolver um modelo de regressão ou classificação, como por exemplo o PLS. Em situações em que o modelo PLS captura uma quantidade muito grande de variáveis na matriz \mathbf{X} , mas existem poucas

variáveis em no vetor \mathbf{y} ; então, o OSC pode ser útil para remover a variação externa de \mathbf{X} que não está relacionada com \mathbf{y} (Westerhuis et al., 2001).

A técnica remove da matriz \mathbf{X} variações sistemáticas de sinal que não estão correlacionados ao vetor \mathbf{y} , que podem estar relacionadas com a variação de condições experimentais que não podem ser controladas na aplicação do método em rotina. Dessa forma, o OSC tem como objetivo remover das variáveis independentes da matriz \mathbf{X} essa informação que não covaria com o vetor \mathbf{y} . Para realizar essa tarefa, é necessário medir os espectros em ao menos duas condições diferentes, que a princípio se referem às variações que não podem ser controladas em situações de rotina, e remover essa informação de \mathbf{X} por projeções ortogonais.

Segundo Wold et al. (1998), o algoritmo desenvolvido para o OSC pode ser descrito da seguinte maneira: primeiramente, calcula-se a primeira componente principal dos dados espectrais, \mathbf{X} , de acordo com o algoritmo NIPALS (do inglês, *Non-linear iterative partial least squares*):

$$\mathbf{X} = \mathbf{t}\mathbf{p}' + \mathbf{E} \quad (18)$$

Onde, \mathbf{t} é vetor de scores para a primeira componente principal, \mathbf{p} é o vetor de pesos (do inglês, “loadings”) para a primeira componente principal e \mathbf{E} , a matriz de resíduos. O vetor \mathbf{t} é ortogonalizado com relação a \mathbf{y} , resultando em \mathbf{t}^* :

$$\mathbf{t}^* = (\mathbf{I} - \mathbf{y}(\mathbf{y}'\mathbf{y})^{-1}\mathbf{y}')\mathbf{t} \quad (19)$$

Onde \mathbf{I} é uma matriz identidade de tamanho apropriado e \mathbf{y} é o vetor que contém a informação de interesse das amostras. Vemos que \mathbf{t}^* é ortogonal a \mathbf{y} por:

$$\mathbf{y}'\mathbf{t}^* = \mathbf{y}'(\mathbf{I} - \mathbf{y}(\mathbf{y}'\mathbf{y})^{-1}\mathbf{y}') \quad (20)$$

$$\mathbf{t} = (\mathbf{y}' - \mathbf{y}'\mathbf{y}(\mathbf{y}'\mathbf{y})^{-1}\mathbf{y}')\mathbf{t} = 0 \quad (21)$$

Em seguida, os pesos, \mathbf{w} , são calculados pelo o algoritmo PLS para estabelecer a direção de menor covariância entre \mathbf{X} e \mathbf{y} .

$$\mathbf{w} = \mathbf{X}^+\mathbf{t}^* \quad (22)$$

Onde \mathbf{X}^+ é a pseudoinversa de \mathbf{X} e \mathbf{t}^* corresponde ao vetor \mathbf{t} ortogonalizado.

Uma nova estimativa de \mathbf{t}^* , \mathbf{t}^{**} é calculada a partir da relação:

$$\mathbf{t}^{**} = \mathbf{X}\mathbf{w} \quad (23)$$

Os pesos \mathbf{w} são calculados para fazer com que $\mathbf{X}\mathbf{w}$ seja o mais próximo possível de \mathbf{t}^* . A estimativa \mathbf{t}^{**} é ortogonalizada contra o vetor \mathbf{y} , e o processo de iteração continua até a convergência de \mathbf{t}^{**} .

Calcula-se, então, o vetor de pesos, \mathbf{p} , dos dados espectrais \mathbf{X} a partir de \mathbf{t}^{**} :

$$\mathbf{p} = \mathbf{X}'\mathbf{t}^{**} / (\mathbf{t}^{**}'\mathbf{t}^{**}) \quad (24)$$

O espectro filtrado, após a remoção da primeira componente ortogonal a \mathbf{y} , é obtido pela seguinte relação:

$$\mathbf{X}_{osc} = \mathbf{X} \mathbf{t} \mathbf{t}' \mathbf{p} \mathbf{p}' \quad (25)$$

Para remover mais de um componente OSC, o procedimento é repetido utilizando a matriz \mathbf{X}_{osc} , já corrigida pelo método. Deve-se observar que a correção do sinal é expressa por uma subtração bilinear $\mathbf{t} \mathbf{p}'$ e os escores \mathbf{t} calculados durante o algoritmo do OSC podem ser recalculados para novos espectros como $\mathbf{t}_{new} = \mathbf{X}_{new} \mathbf{W}$. Ou seja, um novo escore é calculado a partir de uma nova matriz \mathbf{X} de dados e dos pesos \mathbf{w} .

Uma característica do OSC destacada por Wold é que quando o número de variáveis da matriz \mathbf{X} , representado por K , é maior do que o número de amostras de treinamento, representado por N , ou seja, $K > N$, existem infinitas soluções OSC. Dentre essas soluções, o algoritmo é ajustado para encontrar a solução que modela da melhor forma a matriz \mathbf{X} para cada componente. Logo, é possível encontrar uma solução OSC ortogonal. Em contrapartida, quando o número de variáveis é menor do que o número de amostras de treinamento ($K < N$), nem sempre é possível encontrar uma solução OSC ortogonal (Wold, 1998).

Um outro aspecto a ser considerado é com relação ao número de componentes usados para o algoritmo. Normalmente, duas componentes OSC são usadas para espectros de reflectância NIR, onde a aplicação da primeira componente se assemelha a uma correção de linha de base e a segunda, frequentemente, corrige efeitos multiplicativos nos espectros, mas isso é variável (Wold, 1998).

O número de componentes OSC necessário para uma aplicação pode ser determinado por meio do valor explicado de \mathbf{X} em cada componente. Quando o valor explicado por um componente corresponde a um autovalor abaixo de 2, isso indica que mais componentes não são necessários. Uma vez que cada componente removido da matriz \mathbf{X} elimina informações que não estão correlacionadas com \mathbf{y} , então teoricamente, chega-se um momento em que \mathbf{X} não contém nenhuma informação significativa sobre \mathbf{y} .

A aplicação de OSC e posterior desenvolvimento do modelo PLS pode convergir para uma regressão que se assemelha a regressão múltipla linear (MLR, do inglês *Multiple Linear Regression*). Como a MLR é conhecida por apresentar sobreajuste, isso mostra também que usar muitos componentes no OSC pode gerar sobreajuste em um modelo desenvolvido por PLS. Assim, o número de componentes deve ser estimado com cuidado. Além disso, recomenda-se também que os resultados com OSC sejam validados com amostras que não foram utilizadas na etapa de desenvolvimento do modelo.

Na literatura, o OSC já foi aplicado para transferência de calibração em espectros de infravermelho próximo. Sjöblom et al. (1998) investigou a eficiência do OSC em diminuir a variação dos espectros em função da mudança de equipamento. Com isso, a transferência de calibração se tornaria mais fácil entre instrumentos sem a necessidade de criar ou reanalisar todo o conjunto de amostras de treinamento. Nesse mesmo trabalho, o OSC foi aplicado em espectros NIR que foram usados para a calibração de teor de água em produto farmacêutico. O OSC foi capaz de remover parte da variação instrumental e os resíduos em y foram distribuídos aleatoriamente, não indicando uma tendência de distribuição dos resíduos da matriz y que continha a informação de teor de água.

3 CORREÇÃO DA VARIAÇÃO DA UMIDADE EM MODELOS DE DISCRIMINAÇÃO PLS-DA POR OSC

Em problemas de discriminação/classificação, como a identificação de espécies produtoras de madeira, a Tecnologia NIRS está relacionada com a construção de um banco de dados de espectros para o desenvolvimento dos modelos quimiométricos.

No entanto, a madeira, por ser um material complexo, tem o espectro constituído de um conjunto de bandas formadas pela sobreposição de várias transições vibracionais na região de sobretons e combinação de bandas (Pasquini, 2003). É difícil correlacionar as bandas com os grupos funcionais dos constituintes da madeira. Por isso, faz-se necessária a aplicação de pré-processamentos nos espectros e a correção de fatores externos variáveis nas amostras. Além disso, a madeira, como já mencionado na revisão bibliográfica, por ser higroscópica, apresenta variação nos valores de umidade, o que interfere o espectro NIR e prejudica o método para discriminação/classificação.

Para corrigir problemas em modelos de discriminação, pode-se tentar algumas estratégias, mesmo com um material complexo como a madeira. Pode-se incluir na fase de treinamento as fontes possíveis de variabilidade das amostras para garantir representatividade, remover a interferência do sinal indesejado, selecionar variáveis que não são afetadas pelas fontes de variabilidade ou fazer com que o modelo não seja sensível a essas interferências (Igne et al., 2014).

Sabendo dessas possibilidades, testou-se a hipótese para eliminar a variação relativa à umidade que está presente nos espectros, que é uma fonte de variabilidade que prejudica a identificação da espécie, através da aplicação da ferramenta matemática OSC. A transformação OSC foi proposta para melhorar a capacidade preditiva de modelos de treinamento.

Analisando os resultados Sjöblom et al. (1998), observou-se que o OSC obteve bom desempenho ao corrigir a variação de espectros NIR de dois equipamentos diferentes. Ou seja, mostrou-se capaz de diminuir a variação dos espectros com amostras em duas condições diferentes. Pensando nisso, aplicou-se o método OSC nas amostras secas (teor de umidade < 15%) e úmidas (umidade superior a 15%) para corrigir a variação existente entre elas, o método foi usado fundamentando-se na ideia de que a variação de umidade presente na matriz de dados tenha baixa correlação com a propriedade de discriminação das espécies e, em tese, pode ser removida através de uma transformação ortogonal. A hipótese foi testada, sendo desejável que essa variância seja

ortogonal à discriminação das classes presentes no conjunto de dados e a interferência possa ser eliminada, levando a um melhor desempenho do método.

3.1 MATERIAIS E MÉTODOS

A parte experimental foi realizada no Laboratório de Produtos Florestais do Serviço Florestal Brasileiro, pertencente ao Ministério do Meio Ambiente (MMA), em colaboração com o Laboratório de Automação, Quimiometria e Química Ambiental (AQQUA) do Instituto de Química da Universidade de Brasília.

3.1.1 Obtenção e preparo das amostras

As espécies foram selecionadas com base no livro “Madeiras similares ao mogno (*Swietenia macrophylla*): uma chave ilustrada para identificação anatômica em campo” editado pelo Serviço Florestal Brasileiro (Coradin et al., 2009). Das 15 espécies listadas, cinco foram escolhidas para o estudo: *Carapa guianensis* Aubl. (andiroba), *Erismia uncinatum* Warm. (cedrinho), *Cedrela odorata* L. (cedro), *Micropholis melinoniana* Pierre. (curupixá) e *Swietenia macrophylla* King (mogno). As amostras foram obtidas na xiloteca Harry Van der Sloten da Área de Anatomia e Morfologia do LPF, em saídas de campo no município de Manuel Urbano (Acre), Luziânia (Goiás) e Santa Cruz de la Sierra (Bolívia).

As amostras de Luziânia foram coletadas em 2016 no plantio florestal, com 14 anos, de responsabilidade da Fazenda Celeiro Sementes, entre as coordenadas geográficas 47°56'2'' W e 16°30'51'' S, a 30km de Luziânia – GO.

Foram obtidas amostras secas e de madeira recém-cortada. As amostras da xiloteca Harry Van der Sloten estavam secas, apresentando teor de umidade inferior a 15%, enquanto que as amostras obtidas nas saídas de campo eram de madeira recém-cortada. No Acre, as amostras coletadas eram de mogno (*Swietenia macrophylla*) e cedro (*Cedrela odorata*); em Luziânia, foram coletadas amostras de mogno (*Swietenia macrophylla*) de madeira recém-cortada e na Bolívia, as amostras analisadas eram de mogno (*Swietenia macrophylla*), cedrinho (*Erismia uncinatum*) e cedro (*Cedrela odorata*) que apresentaram umidade que variava de 15 a 25%.

As amostras foram identificadas por anatomista de madeira e foram analisadas 711 amostras diferentes, sendo 118 de andiroba (*Carapa guianensis*), 117 de cedro (*Cedrela odorata*), 151 de cedrinho (*Erismia uncinatum*), 81 de curupixá (*Micropholis melinoniana*) e 183 de mogno (*Swietenia macrophylla*), sendo provenientes majoritariamente da xiloteca. As amostras eram diferentes entre si porque correspondiam a partes diferentes da árvore, isso não quer dizer que todas as amostras eram de árvores

diferentes. De fato, existiam algumas amostras procedentes da mesma árvore, mas isso não comprometeu a variabilidade do conjunto de amostra porque o espectro era tirado em superfícies diferentes da mesma árvore.

Posteriormente, as amostras foram secas naturalmente ao serem deixadas no laboratório a temperatura ambiente (em torno de 25°C) e a superfície foi polida com lixas nº 80. Essa granulação foi escolhida por ser uma lixa mais grossa que aumenta a superfície de contato da madeira e evita a formação de brilho, o que diminui a componente de reflectância especular. Ou seja, evita que a superfície da amostra reflita a radiação no mesmo ângulo de incidência, como se fosse um espelho. Essa interferência é atenuada para que os espectros sejam obtidos em reflectância difusa, ou seja, a superfície da amostra reflete a luz incidente em diversos ângulos.

Além disso, esse procedimento é importante para manter a uniformidade granulométrica da superfície e remover a camada externa oxidada. Os espectros NIR foram obtidos logo após limpar a superfície com um pincel e para cada amostra, três espectros eram obtidos.

3.1.2 Obtenção dos espectros NIR e medidas de umidade

Os espectros de reflectância difusa das espécies estudadas foram obtidos com o auxílio do espectrômetro portátil MicroNir™ 1700 Spectrometer fabricado pela JDSU (Estados Unidos) com faixa espectral de 950 a 1.650 nm e resolução de 6,2 nm. Os seguintes parâmetros foram estabelecidos no software do equipamento: tempo de integração de 2.000 µs e 100 varreduras. A extremidade inicial do intervalo do comprimento de onda dos espectros foi removida para minimizar ruídos e variações não relacionadas às informações sobre as espécies. Portanto, a região espectral selecionada para a construção dos modelos de discriminação correspondeu ao intervalo de 1.000 a 1.650 nm. A Figura 4 ilustra o espectrômetro portátil utilizado:

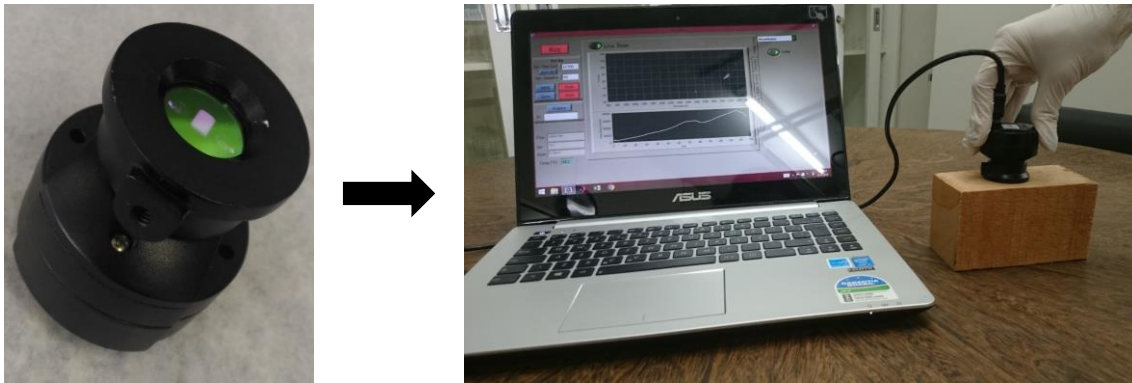


Figura 4. Espectrômetro portátil MicroNir™ (a esquerda) e a realização de uma medida em madeira sólida com o auxílio de um notebook (a direita).

- *Amostras obtidas na xiloteca Harry Van der Sloten e nas saídas de campo no Acre e na Bolívia :*

Para realização das medidas, os espectros foram obtidos a partir da superfície da madeira nas faces longitudinal tangencial ou transversal, sem que fosse feita distinção entre elas. Foram medidos três espectros em pontos distintos aleatórios sem seguir um padrão. Devido à heterogeneidade natural presente nas amostras de madeira, optou-se por não fazer médias dos espectros medidos em pontos diferentes de uma mesma amostra para ampliar a representatividade do conjunto de dados.

- *Amostras obtidas na saída de campo em Luziânia:*

Dentre as amostras coletadas, seis amostras, em formas de disco, de madeira de mogno (*Swietenia macrophylla*) recém-cortadas foram analisadas e 10 espectros foram obtidos em pontos distintos aleatórios em cada amostra na face transversal.

Como o objetivo era de se obter espectros à medida que a madeira recém-cortada secava por estar exposta a um ambiente relativamente aberto, foram medidos espectros durante os meses de abril e maio. Para isso, durante as quatro primeiras semanas, os espectros das amostras eram obtidos todos os dias da semana, exceto sábados e domingos. A partir da quinta semana, os espectros foram medidos a cada dois dias e nas últimas três semanas, os espectros eram tomados semanalmente.

Dessa forma, foi possível avaliar o perfil do espectro da madeira de mogno (*Swietenia macrophylla*) a cada semana. À medida que a umidade nas amostras era reduzida, aumentou-se o período entre a obtenção de um espectro e o outro.

Para obtenção dos valores de umidade das amostras, utilizou-se um medidor de umidade elétrico capacitivo: higrômetro digital KT-50B fabricado pela KLORTNER com faixa de medição entre 2 a 90% e precisão de 0,1%. Primeiramente, o equipamento foi configurado para fazer medidas para o Grupo de Madeira 03, que corresponde a madeiras com densidade específica na faixa de $0,6 \leq \rho < 0,7 \text{ g/cm}^3$. Essa configuração era ajustada no equipamento e isso foi feito considerando que a densidade específica do mogno (*Swietenia macrophylla*) varia de 0,5 a 0,7 g/cm³ e a densidade das outras espécies é próxima desse valor (Soares et al., 2017).

Em seguida, o higrômetro era posicionado sobre a superfície da madeira e o valor no display do equipamento correspondia ao valor de umidade presente na amostra. A Figura 5 a seguir ilustra o equipamento utilizado:



Figura 5. Higrômetro digital usado para as medidas de umidade. Fonte: www.banggood.com

3.2 RESULTADOS E DISCUSSÃO

3.2.1 Análise exploratória dos espectros das amostras secas e úmidas

A grande maioria das amostras que foi usada neste trabalho já se encontrava em condição seca com valor de umidade em equilíbrio com a umidade relativa do ar ambiente. O valor de umidade de equilíbrio das amostras de madeira foi de aproximadamente 12%, estando sempre inferior a 15%. No caso das amostras de madeira recém-cortada obtidas no Acre, realizou-se a secagem dessas amostras ao ambiente antes da obtenção dos espectros.

Já as amostras de madeira recém-cortada de Luziânia e as coletadas na Bolívia tinham seus valores de umidade variando entre 25 a 64%. As amostras da Bolívia tiveram seus espectros obtidos antes do processo de secagem, enquanto as amostras de Luziânia tiveram seus espectros registrados ao longo do processo de secagem durante os meses de abril e maio.

O teor de umidade das seis amostras de mogno (*Swietenia macrophylla*) recém-cortadas de Luziânia variou entre 14 a 64%. Essa variação pode ser observada por meio da Figura 6. Nota-se claramente uma queda no valor de umidade das amostras do primeiro para o segundo dia, diminuindo cerca de 20%. Depois disso, o decaimento é gradual e torna-se quase que constante a partir do vigésimo dia, entre 10 e 15%, indicando que as amostras entraram em equilíbrio com a condição de umidade do ambiente.

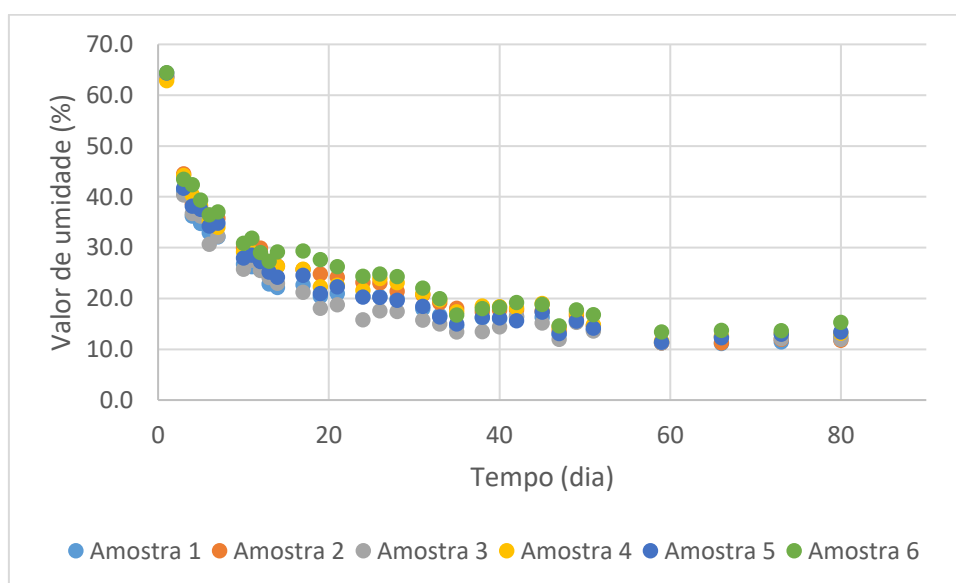


Figura 6. Variação de umidade das amostras de mogno (*Swietenia macrophylla*) recém-cortadas procedentes da Fazenda Celeiro Sementes (Luziânia, GO) em função do tempo.

Em seguida, foram avaliados os espectros NIR de madeira nas condições úmida (teor de umidade > 15%) e seca. Os espectros de mogno (*Swietenia macrophylla*) de amostras secas e de mogno (*Swietenia macrophylla*) recém-cortadas de Luziânia (Goiás) foram comparados e correspondem aos espectros medidos no período dos dois meses. Na Figura 7, estão apresentados esses espectros, sem qualquer tipo de processamento, representando em vermelho as amostras de mogno (*Swietenia macrophylla*) seco e em azul, as amostras de madeira recém-cortada.

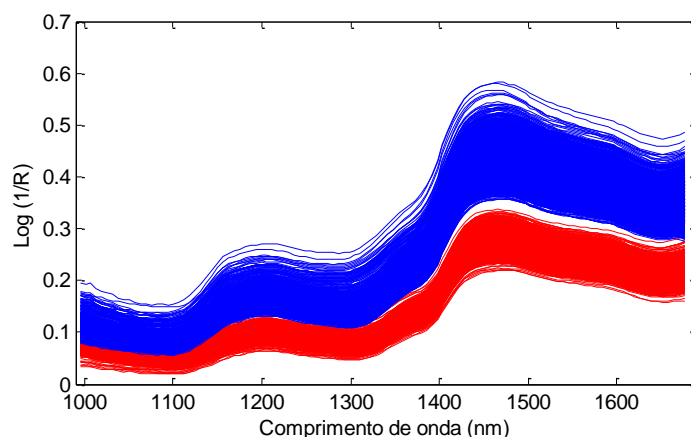


Figura 7. Espectros sem pré-processamento de mogno (*Swietenia macrophylla*). Legenda: (—) espectros de mogno (*Swietenia macrophylla*) de madeira recém-cortada ou com umidade acima de 15% e (—) espectros de mogno (*Swietenia macrophylla*) seco ou com umidade entre 10 e 15%.

Visualmente, os espectros apresentam um deslocamento de linha de base em relação ao eixo vertical representada por $\log(1/R)$. Os espectros em azul apresentam maiores valores de $\log(1/R)$ quando comparados aos espectros em vermelho, o que pode ser atribuído à maior presença de água nas amostras de madeira recém-cortada. No entanto, há uma diminuição dos valores de linha de base dos espectros das amostras de mogno (*Swietenia macrophylla*) de madeira recém-cortada. Observa-se também um conjunto de espectros de mogno (*Swietenia macrophylla*) de madeira recém cortada, em azul, que apresenta variação nos valores de $\log(1/R)$, essa variação aconteceu à medida que os espectros eram obtidos. Ou seja, os espectros com maiores valores de $\log(1/R)$ correspondem aos primeiros espectros obtidos, quando a madeira apresentava um maior teor de umidade e os espectros com menores valores de $\log(1/R)$ dizem respeito aos espectros obtidos quando as amostras estavam mais secas. A diminuição dos valores de intensidade de $\log(1/R)$ segue a ordem em que os espectros foram obtidos ao longo do

tempo à medida que as amostras eram secas a temperatura ambiente. Assim à medida que os espectros eram tomados ao longo das semanas, observa-se que as amostras eram secas ao ambiente e o perfil do espectro do mogno (*Swietenia macrophylla*) recém-cortado se aproximou do perfil das amostras secas. Esse comportamento refere-se à evaporação da água livre e, eventualmente, de outros compostos voláteis nas amostras recém-cortadas.

Posteriormente, observaram-se as bandas de absorção dos espectros. Para isso, calcularam-se os espectros médios das amostras e os principais comprimentos de onda foram indicados. O cálculo dos espectros médios foi feito no software MATLAB da seguinte forma: para cada coluna da matriz dos espectros, foi calculada uma média de valores. Então, um vetor_(1x111) foi criado contendo os valores da média em uma única linha e 111 colunas com os comprimentos de onda.

Na Figura 8, estão apresentados os espectros médios das amostras de madeira de mogno (*Swietenia macrophylla*) seco e recém-cortada procedentes de Luziânia, nas cores vermelha e azul, respectivamente.

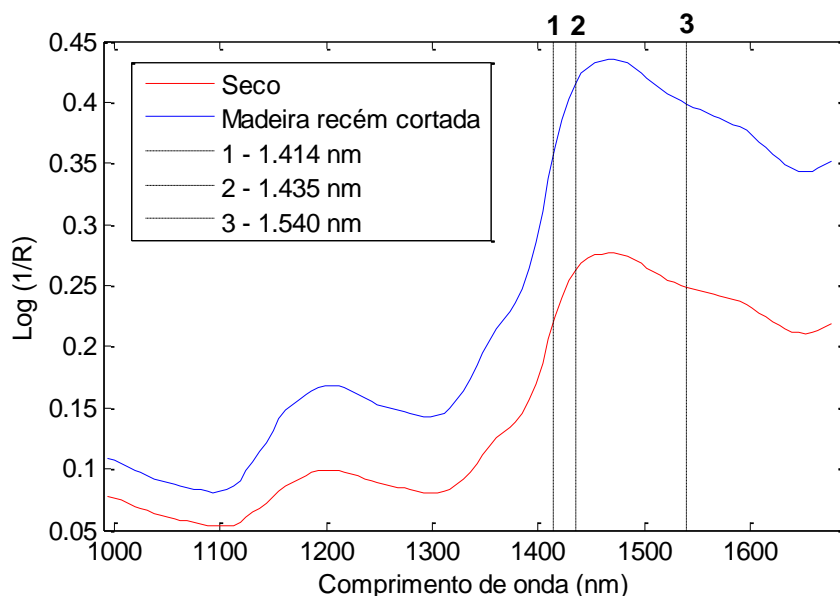


Figura 8. Variação dos espectros médios de mogno (*Swietenia macrophylla*) e indicação dos principais comprimentos de onda referentes à água.

As bandas de absorção entre 1.400 a 1.600 nm são mais intensas no conjunto de amostras com maior teor de umidade. Essa faixa de comprimento de onda caracteriza a absorção das ligações OH de moléculas de água. Logo, os picos são mais intensos porque existe uma maior quantidade de água presente na madeira.

Segundo Schwanninger et al. (2011), a região entre 1.400 a 1.600 nm apresenta as seguintes vibrações de moléculas de água na madeira, conforme Tabela 1:

Tabela 1. Atribuições de bandas características às vibrações de moléculas de água em espectros NIR de madeira (Schwanninger et al., 2011).

| Comprimento de onda (nm) | Vibração |
|-------------------------------------|--|
| 1414 | Ligações O-H intramoleculares de moléculas de água |
| 1435-1438 | Ligações O-H de moléculas de água |
| 1540 | Ligações O-H de água adsorvida por moléculas de celulose microcristalina |

Na Figura 8, pode-se observar também, próximo a 1200 nm, um aumento na intensidade de absorção das amostras com maior teor de umidade. Essa região de comprimento de onda caracteriza as vibrações de ligações C-H de lignina e celulose. Segundo Richardson (1993), cada unidade de glicose, presente nas cadeias de celulose, possui três grupos hidroxila que tem afinidade com água e podem formar ligações do tipo ligação de hidrogênio. Isso garante que as cadeias de celulose possuem a capacidade de reter facilmente a água disponível no meio. No que diz respeito às moléculas de lignina, essas têm como base estrutural o fenil-propano, tendo ligado ao anel benzênico um número variável de grupos hidroxílicos e metoxílicos capazes de se ligar à água. Provavelmente, a lignina e a celulose da madeira foram hidratadas pela água presente no conjunto de amostras com maior teor de umidade.

As vibrações identificadas são responsáveis pela mudança no perfil do espectro e podem também modificar propriedades físicas e mecânicas da madeira (Tsuchikawa & Tsutsumi, 1998). Essa mudança no perfil do espectro pode ainda ser comprovada pela análise PCA com dados centrados na média apresentada na Figura 9. As amostras foram divididas em dois grupos distintos, apesar de serem da mesma espécie, em função da diferença de umidade. A primeira componente principal da análise PCA explicou 99,94% da informação dos espectros.

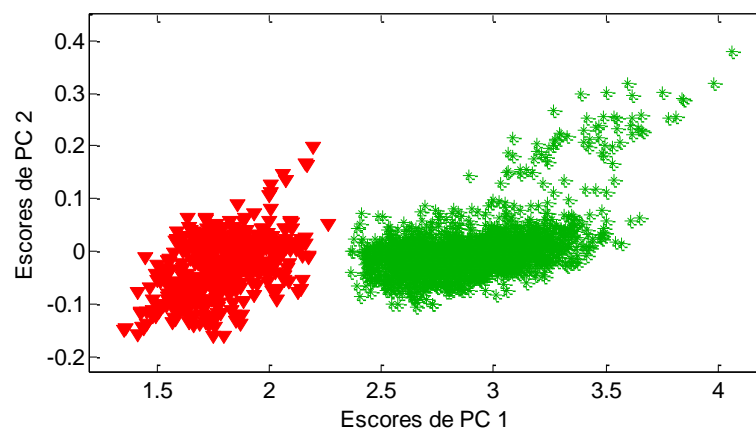


Figura 9. Escores de PC1 e PC2 estimados com dados centrados na média para as amostras de mogno (*Swietenia macrophylla*). Legenda: (▼) amostras de mogno (*Swietenia macrophylla*) de madeira recém-cortada e (*) amostras de mogno (*Swietenia macrophylla*) seco.

Uma alternativa para minimizar os efeitos dessa variação nos espectros é pelo uso de pré-processamentos. Neste trabalho, dois pré-processamentos foram usados:

- (1) Primeira derivada com alisamento pelo método de Savitzky-Golay: Calculou-se a 1ª derivada, mediante o uso de uma janela com 7 pontos para o alisamento de curva com polinômio de 2ª ordem.
- (2) Centrar na média: Calculou-se a média das intensidades para cada variável considerando todas as amostras de treinamento, resultando em um vetor de médias ou espectro médio. Em seguida, esse vetor médio foi subtraído de cada espectro bruto. Cabe destacar que o pré-processamento de centrar os dados na média é considerado padrão para a análise multivariada de dados espectroscópicos, razão pela qual foi utilizado para a obtenção da Figura 9.

O objetivo do uso do processamento de primeira derivada é reduzir as variações indesejáveis que não foram removidas durante a aquisição dos dados e que podem influenciar os resultados finais de desempenho do modelo PLS-DA.

Na Figura 7, observa-se um deslocamento constante na linha de base no eixo y, onde o espectro tende a valores positivos com relação ao zero de absorvância. Para corrigir isso, calculou-se a primeira derivada do espectro. Como a primeira derivada de uma constante é zero, no espectro resultante sua linha de base é deslocada para o zero de absorvância após a correção, como pode ser observado na Figura 10. Entretanto, pode-se observar que, mesmo com o uso da 1ª derivada, a variação da umidade nos espectros não

foi eliminada durante a análise, uma vez que ainda se observa uma diferença sistemática entre as amostras secas e úmidas.

Os resultados obtidos pela PCA com os dados pré-processados com 1ª derivada e dados centrados na média confirmam esse fato. É certo que os grupos de amostras foram aproximados, se observarmos a diminuição na variação do eixo de escores de PC1 da Figura 11 em relação à Figura 9; no entanto, ainda é presente a separação das amostras em dois grupos distintos. Nessa situação, a primeira componente principal explicou 86,04% da informação dos dados.

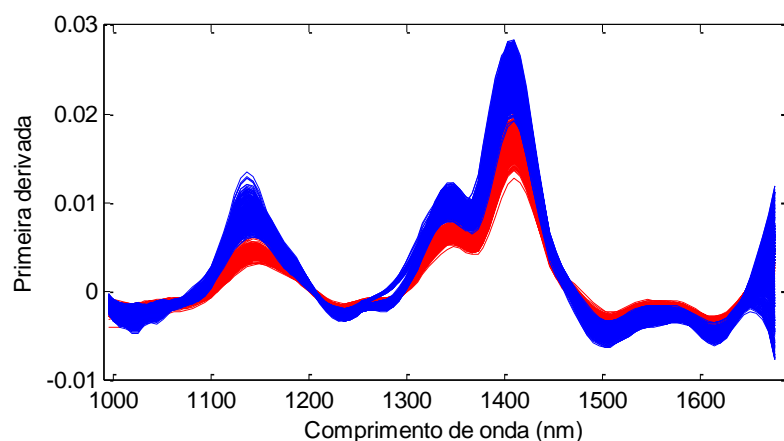


Figura 10. Espectros de mogno (*Swietenia macrophylla*) pré-processados com primeira derivada. Legenda: (—) espectros de mogno (*Swietenia macrophylla*) de madeira recém-cortada e (—) espectros de mogno (*Swietenia macrophylla*) seco.

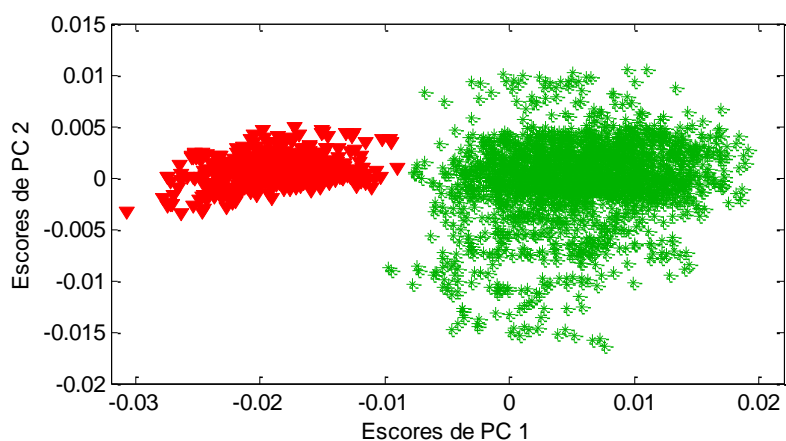


Figura 11. Escores dos PC1 e PC2 estimados após a aplicação da 1ª derivada e dados centrados na média para amostras de mogno (*Swietenia macrophylla*). Legenda: (▼) amostras de mogno (*Swietenia macrophylla*) de madeira recém-cortada e (*) amostras de mogno (*Swietenia macrophylla*) seco.

Na tentativa de corrigir os espectros, optou-se por aplicar a transformação de correção ortogonal de sinal. Essa transformação determina as componentes ortogonais a uma variação de sinal especificada, resolvendo um problema de autovalores, e procura corrigir os dados antes da construção dos modelos PLS-DA. Para realizar a correção foi criada uma matriz contendo 60 espectros das espécies mogno (*Swietenia macrophylla*), cedrinho (*Erismia uncinatum*) e cedro (*Cedrela odorata*), sendo que metade dos espectros de cada espécie estava na condição seca e a outra metade, com teor maior de umidade. Utilizando essa matriz foi feito o cálculo descrito na seção 2.2.3, no qual se estabeleceu o número de componentes OSC como sendo igual àquele que eliminasse a existência de dois grupos no gráfico de escores de PCA com os espectros corrigidos; neste caso, o número de componentes foi igual a 1. Os demais parâmetros do algoritmo, tais como o número máximo de iterações e nível de tolerância foram deixados nos valores padrão (“default”), iguais a zero e 99,9%, respectivamente.

A Figura 12 ilustra a aparência dos espectros após aplicação do método de correção OSC. A remoção de uma componente OSC dos espectros sem pré-processamento implica numa mudança no intervalo de variação do eixo y. Além disso, pode-se observar que os espectros não apresentam mais uma diferença de intensidades como observado na Figura 10. Essa melhor concordância entre os espectros de madeira seca e de madeira recém-cortada também pode ser confirmada pela PCA (Figura 12). Apesar da dispersão maior do conjunto de amostras de mogno (*Swietenia macrophylla*) recém-cortadas, o método pôde eliminar a separação existente entre os grupos vista nas Figuras 9 e 11. Na Figura 13, a visualização dos escores da PCA sugere que a variação espectral referente à diferença entre os dois conjuntos de espectros pôde ser compensada pela ferramenta matemática OSC.

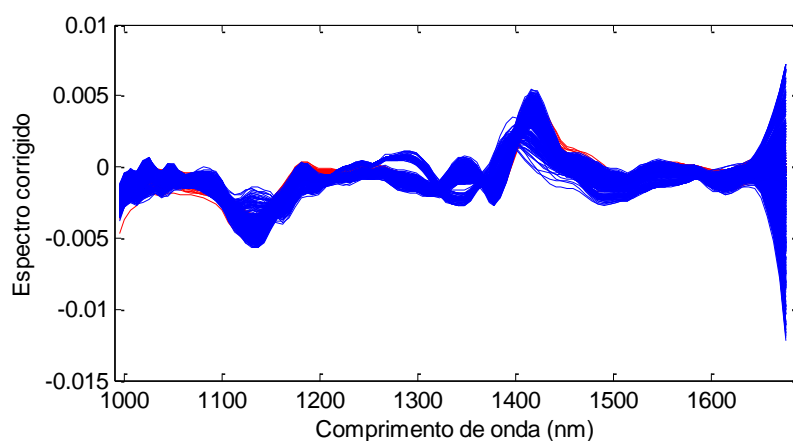


Figura 12. Espectros de mogno (*Swietenia macrophylla*) corrigidos com aplicação de OSC. Legenda: (—) espectros de mogno (*Swietenia macrophylla*) de madeira recém-cortada e (—) espectros de mogno (*Swietenia macrophylla*) seco.

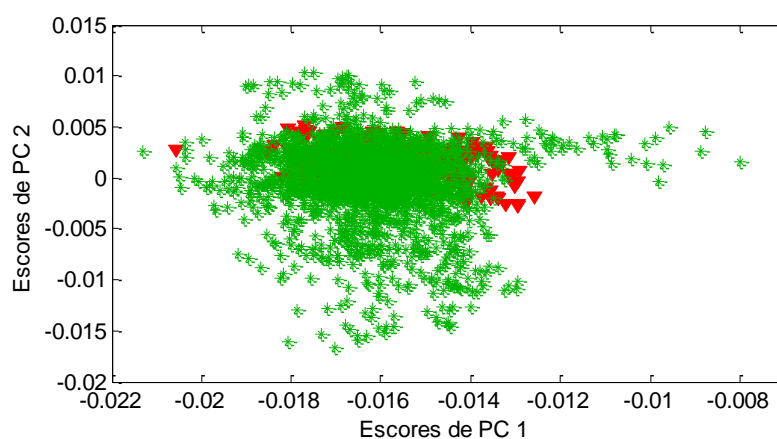


Figura 13. Escores de PC1 e PC2 estimados após aplicação do OSC para amostras de mogno (*Swietenia macrophylla*). Legenda: (▼) amostras de mogno (*Swietenia macrophylla*) de madeira recém-cortada e (*) amostras de mogno (*Swietenia macrophylla*) seco.

3.2.2 Modelo de discriminação por PLS-DA

A seguir, serão apresentados os resultados para o modelo de discriminação de mogno (*Swietenia macrophylla*) e espécies similares, usando PLS-DA e NIRS. A construção desse modelo dá continuidade ao estudo que vem sendo realizado para a discriminação dessas espécies no projeto de colaboração entre LPF-SFB e o laboratório AQQUA da UnB. Os resultados do modelo para a discriminação de amostras secas foi recentemente publicado na revista Química Nova (Soares et al., 2017), que apresentou resultados da discriminação de seis espécies brasileiras de madeira anatomicamente similares: *Carapa guianensis* Aubl.(andiroba), *Cedrela odorata* L. (cedro), *Erismia uncinatum* Warm. (cedrinho), *Micropholis melinoniana* Pierre (curupixá), *Hymenea coubaril* L. (jatobá) e *Swietenia macrophylla* King. (mogno).

Após a publicação desses resultados, decidiu-se retirar a espécie *Hymenea coubaril* L. (jatobá) do modelo de discriminação, tendo em vista que esta é a espécie que menos se assemelha visualmente ao mogno (*Swietenia macrophylla*), comparada com as demais, não está relacionada em lista de espécies ameaçadas no âmbito nacional ou internacional e, principalmente, porque a diminuição de uma classe no modelo reduz sua complexidade e pode representar um ganho de eficiência para as outras espécies modeladas.

Nessa seção serão apresentados apenas resultados do modelo que discrimina mogno (*Swietenia macrophylla*) em relação às outras quatro espécies, sendo que para mogno (*Swietenia macrophylla*) será atribuído o valor de classe 1 e para as demais espécies o valor de classe igual a 0. O número de variáveis latentes (VL), da Taxa de Falso Positivo (TFP), da Taxa de Falso Negativo (TFN), da Taxa de Eficiência (TEF), limite de discriminação (LIM) e a quantidade de espectros do conjunto de treinamento e validação do modelo estão dispostos na Tabela 2.

Tabela 2. Parâmetros do modelo de discriminação de mogno (*Swietenia macrophylla*) construído com amostras na condição seca e aplicado às amostras de madeira seca.

| Parâmetros | |
|--------------------------------|--------|
| Nº de espectros de treinamento | 1128 |
| Nº de espectros de validação | 563 |
| Nº de VL | 14 |
| LIM | 0,4764 |
| TFN | 1,80% |
| TFP | 2,15% |
| TEF | 96,03% |

Os valores de classe estimados para os conjuntos de treinamento e validação para a discriminação de mogno (*Swietenia macrophylla*) estão apresentados na Figura 14. Nota-se na Figura 14 que não há uma diferença significativa da dispersão dos espectros do conjunto de treinamento em comparação com aqueles do conjunto de validação do modelo, o que indica a ausência do sobreajuste no modelo.

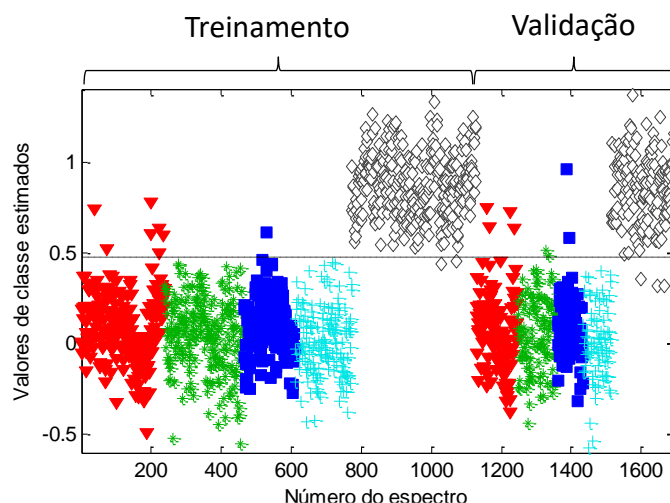


Figura 14. Valores de classe estimados para os conjuntos de treinamento e validação para a discriminação de mogno (*Swietenia macrophylla*) seco. Legenda: andiroba (*Carapa guianensis*) (▼), cedrinho (*Erisma uncinatum*) (*), cedro (*Cedrela odorata*) (■), curupixá (*Micropholis melinoniana*) (+) e mogno (*Swietenia macrophylla*) (◇) e (— —) limite de discriminação.

Além disso, identificou-se amostras anômalas na etapa de validação. Essas amostras apresentaram valores fora dos limites estabelecidos para os valores de classe estimados e também elevados valores de T^2 de Hotelling e resíduos Q. Os resíduos Q representam o quanto uma amostra não se adequa ao modelo PLS-DA construído e T^2 Hotelling representa o quanto cada amostra está distante das demais amostras no espaço definido pelas componentes do modelo. Se uma amostra apresenta altos valores para esses parâmetros, significa dizer que essa amostra apresenta características distintas das amostras usadas na fase de treinamento. Essas amostras podem ser identificadas na Figura 15, onde se localizam no quadrante do gráfico referente a elevados valores de T^2 e Q acima dos limites de confiança. Tendo em vista que apenas 6,7% das amostras de validação foram identificadas como *outliers*, pode-se considerar que o modelo construído apresenta capacidade de analisar mais de 90% das amostras em futuras análises, desde que estas estejam na condição de madeira seca.

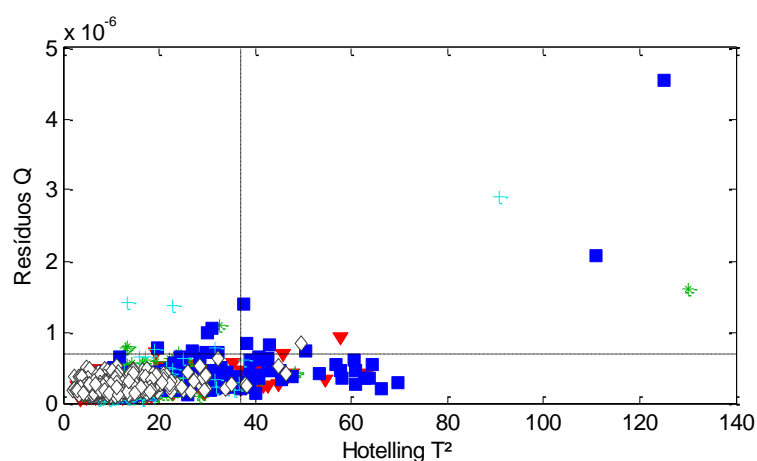


Figura 15. Gráfico dos valores de T^2 de Hotelling e resíduos Q da discriminação de mogno (*Swietenia macrophylla*) seco.

A aplicação do modelo construído em condição seca às amostras recém-cortadas mostra que estas claramente não preenchem esse requisito. A Figura 16 confirma esse fato pela grande quantidade de amostras fora dos limites de confiança de T^2 e Q. Para essas amostras, observou-se que a grande maioria apresenta valores de resíduos Q e T^2 de Hotelling acima de seus limites com 99,9 % de confiança. Isso indica que esse conjunto de amostras foi identificado como anômalo em relação ao conjunto de treinamento. Ou seja, quase todos os espectros foram identificados como muito diferentes dos utilizados

na fase de treinamento; sendo, portanto, o modelo inválido para analisar esse conjunto de amostras.

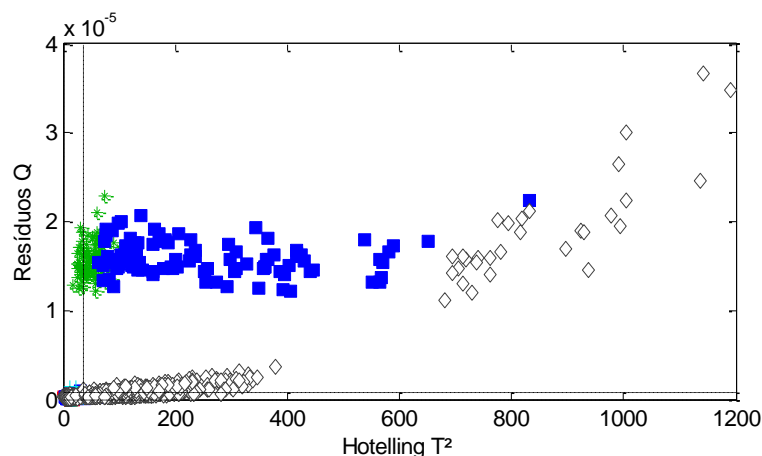


Figura 16. Gráfico dos valores de T^2 de Hotelling e resíduos Q da discriminação de mogno (*Swietenia macrophylla*) de madeira recém-cortada.

Ainda assim, verificou-se o desempenho do modelo para as amostras de mogno (*Swietenia macrophylla*) de madeira recém-cortada. A Figura 17 apresenta o gráfico de dispersão das amostras em relação aos valores de classe estimados. É evidente que há um aumento na dispersão das amostras. Além disso, as amostras de madeira recém-cortada, que pertencem a classe 1, estão deslocadas para valores de classe significativamente menores do que 1. As amostras de mogno (*Swietenia macrophylla*) de madeira recém-cortadas tiveram seus teores de umidade reduzidos ao longo do tempo, mas mesmo depois de estarem relativamente mais secas, o modelo continuou errando a sua classificação. Outro problema observado foi que muitas das amostras com maior teor de umidade das espécies cedrinho (*Erismia uncinatum*) e cedro (*Cedrela odorata*) foram classificadas como sendo de mogno (*Swietenia macrophylla*), uma vez que apresentaram valores de classe superiores ao limite de discriminação. Como consequência desses resultados, o modelo apresentou uma baixa TEF. Esse comportamento leva a crer que, além da umidade, outros compostos presentes na amostra provavelmente influenciaram os espectros NIR.

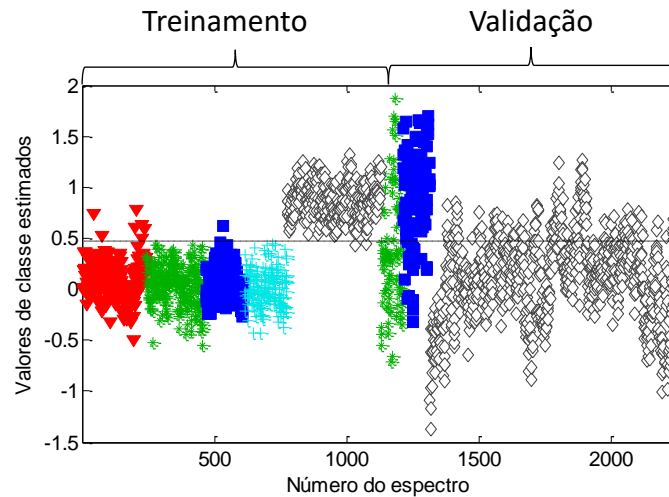


Figura 17. Valores de classe estimados para os conjuntos de treinamento e validação para a discriminação de mogno (*Swietenia macrophylla*) de madeira recém-cortada. Legenda: andiroba (*Carapa guianensis*) (▼), cedrinho (*Erisma uncinatum*) (*), cedro (*Cedrela odorata*) (■), curupixá (*Micropholis melinoniana*) (+) e mogno (*Swietenia macrophylla*) (◇) e (— —) limite de discriminação.

A Tabela 3 resume os resultados obtidos para a aplicação do modelo construído na condição seca às amostras recém-cortadas, confirmando a ineficiência do modelo para essas amostras. Mais de 90% das amostras de validação foram classificadas como *outliers* e a TEF apresentou um valor negativo. O valor negativo justifica-se pelos altos valores de TFN e TFP, que somados ultrapassam o valor de 100%, implicando em TEF negativos pela aplicação da equação 14.

Tabela 3. Parâmetros do modelo de discriminação de mogno (*Swietenia macrophylla*) construído com amostras na condição seca e aplicado às amostras de madeira recém-cortada.

| Parâmetros | |
|--|---------|
| Nº de espectros de validação das amostras recém-cortadas | 1116 |
| Nº de amostras identificadas como <i>outliers</i> | 1024 |
| TFN | 82,57% |
| TFP | 62,42% |
| TEF | -45,00% |

Na tentativa de melhorar o desempenho do modelo e diminuir os erros de classificação das amostras recém-cortadas, aplicou-se o método OSC em todos os conjuntos de dados para tentar remover a informação da umidade. Portanto, após os espectros terem sido pré-processados com OSC, o modelo foi reconstruído e validado com os espectros corrigidos do conjunto de treinamento e validação, respectivamente. Em seguida, o modelo foi aplicado aos espectros de mogno (*Swietenia macrophylla*) de madeira recém-cortada também corrigidos por OSC.

As Figuras 18 e 19 apresentam os gráficos de T^2 versus Q e de dispersão para as amostras de treinamento e validação após aplicação do OSC e a Tabela 4, os parâmetros e as figuras de mérito do modelo. Observa-se que não houve mudança significativa na dispersão das amostras e nos valores de classe estimados se comparados com o gráfico da Figura 14 quando o método OSC não tinha sido aplicado. Houve um aumento de 1,30% na TEF e um pequeno aumento na quantidade de amostras com valores de T^2 e Q elevados. No modelo anterior sem a correção OSC, havia 38 *outliers* na validação e neste modelo, há 43 *outliers*. Contudo, essas mudanças não foram consideradas significativas e o modelo continuou com a mesma eficiência para uso com amostras na condição seca.

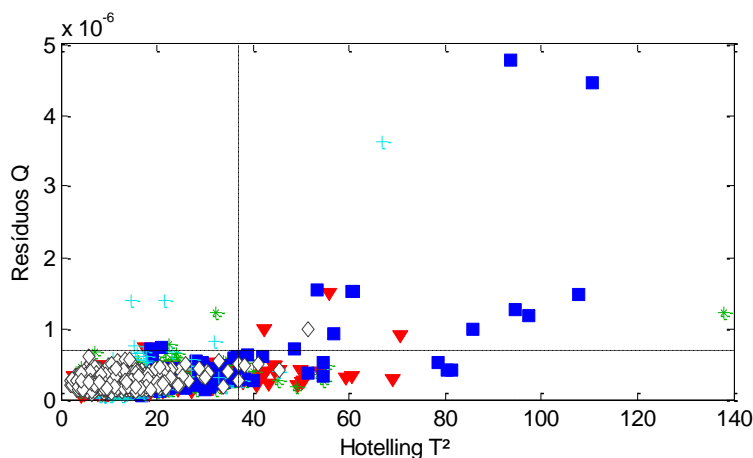


Figura 18. Gráfico dos valores de T^2 de Hotelling e resíduos Q para discriminação de mogno (*Swietenia macrophylla*) seco após aplicação de OSC.

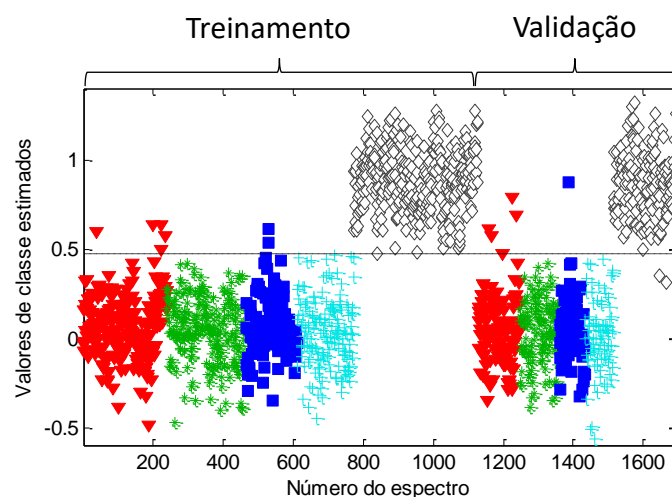


Figura 19. Valores de classe estimados para os conjuntos de treinamento e validação para a discriminação de mogno (*Swietenia macrophylla*) seco após aplicação do OSC. Legenda: andiroba (*Carapa guianensis*) (▼), cedrinho (*Erismia uncinatum*) (*), cedro (*Cedrela odorata*) (■), curupixá (*Micropholis melinoniana*) (+) e mogno (*Swietenia macrophylla*) (◇) e (—) limite de discriminação.

Tabela 4. Parâmetros do modelo de discriminação de mogno (*Swietenia macrophylla*) construído na condição seca após aplicação do método OSC e aplicado às amostras de madeira seca.

| Parâmetros | |
|--------------------------------|--------|
| Nº de espectros de treinamento | 1128 |
| Nº de espectros de validação | 563 |
| Nº de VL | 14 |
| Limite de discriminação | 0,4715 |
| TFN | 1,28% |
| TFP | 1,38% |
| TEF | 97,33% |

O desempenho do modelo após aplicação do método OSC com as amostras recém-cortadas foi analisado pelos resultados nas Figuras 20 e 21 e na Tabela 5. A Figura 20 mostra que, após a aplicação do método, os valores de T^2 e resíduos Q das amostras de mogno (*Swietenia macrophylla*) de madeira recém-cortada apresentaram uma diminuição em relação ao modelo anterior, mas ainda sim essas amostras são classificadas como

outliers. O número de *outliers* em comparação com a Figura 16 apenas apresenta uma diminuição de 20% em relação ao modelo anterior.

Na Figura 20, pode-se observar um aumento na dispersão das amostras de cedrinho (*Erisma uncinatum*) e cedro (*Cedrela odorata*) de madeira recém-cortada no conjunto de validação. As amostras estão deslocadas para valores menores do que zero e sua grande maioria está abaixo do limite de discriminação, diferente do gráfico na Figura 17. Existe ainda uma maior dispersão das amostras de mogno (*Swietenia macrophylla*) do conjunto de validação da Figura 20 em relação à Figura 17, sendo que a maior parte das amostras é classificada como não sendo mogno (*Swietenia macrophylla*).

Observa-se ainda que os primeiros 60 espectros das amostras de mogno (*Swietenia macrophylla*) desse conjunto apresentam menores valores. Nesse dia, as amostras estavam com o maior teor de umidade e os espectros foram classificados erroneamente, indicando que o OSC não conseguiu eliminar essa variação brusca de umidade. Em relação ao modelo anterior, observa-se que o único ganho foi o fato das amostras de cedrinho (*Erisma uncinatum*) e cedro (*Cedrela odorata*) com maior teor de umidade agora terem valores de classe abaixo do limite de discriminação, sendo classificadas corretamente como não sendo mogno (*Swietenia macrophylla*). Como resultado da melhora para essas amostras, a TEF teve um aumento para 18%, porém a aplicação do método não conseguiu compensar os efeitos da umidade nos espectros e por isso, o modelo não é válido para esse conjunto de amostras. Por fim, a Tabela 5 apresentou uma alta taxa de TFN, uma vez que uma grande parte das amostras de mogno (*Swietenia macrophylla*) ficou abaixo do limite de discriminação (Figura 21).

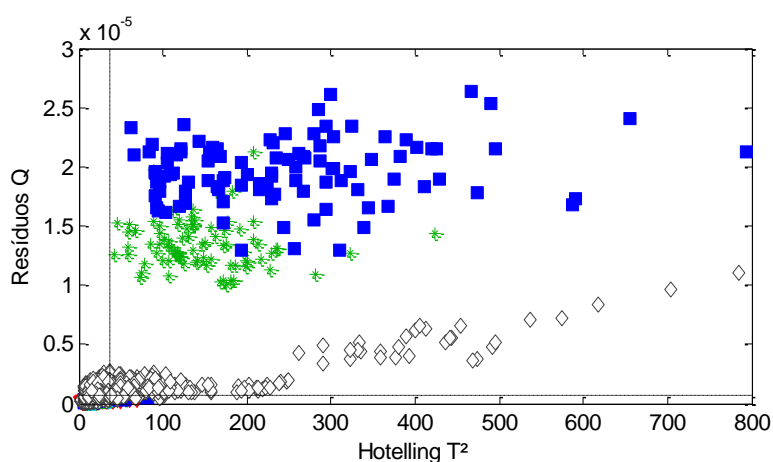


Figura 20. Gráfico dos valores de T^2 de Hotelling e resíduos Q da discriminação de mogno (*Swietenia macrophylla*) de madeira recém-cortada após aplicação de OSC.

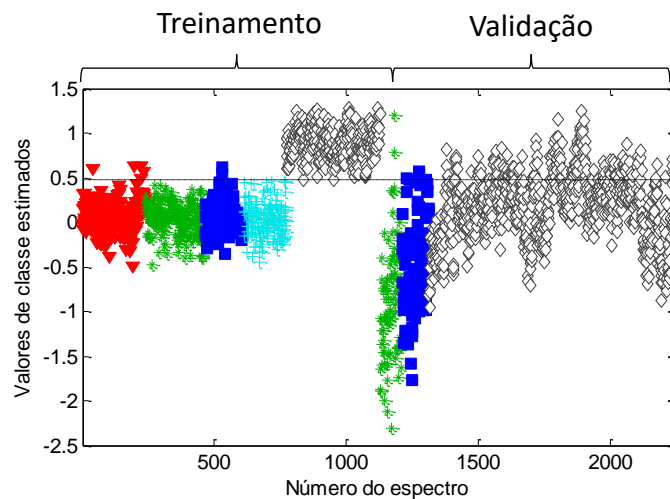


Figura 21. Valores de classe estimados para os conjuntos de treinamento e validação para a discriminação de mogno (*Swietenia macrophylla*) de madeira recém-cortada após aplicação do OSC. Legenda: andiroba (*Carapa guianensis*) (▼), cedrinho (*Erisma uncinatum*) (*), cedro (*Cedrela odorata*) (■), curupixá (*Micropholis melinoniana*) (+) e mogno (*Swietenia macrophylla*) (◇) e (— —) limite de discriminação.

Tabela 5. Parâmetros do modelo de discriminação de mogno (*Swietenia macrophylla*) construído na condição seca após aplicação do método OSC e aplicado às amostras de madeira recém-cortada.

| Parâmetros | |
|---|--------|
| Nº de espectros de validação amostras recém-cortadas | 1116 |
| Nº de amostras identificadas como <i>outliers</i> | 801 |
| TFN | 78,25% |
| TFP | 3,50% |
| TEF | 18,24% |

3.3 CONCLUSÕES PARCIAIS

A primeira etapa permitiu avaliar o comportamento das amostras quando submetidas a condições de umidade diferentes. O aumento na intensidade das bandas, principalmente na faixa de comprimento de onda de 1400 a 1600 nm, mudou o perfil do espectro e a PCA mostrou a separação das amostras em função da quantidade de água presente na madeira. A avaliação de PCA com dados pré-processados com OSC a nível de análise exploratória forneceu uma boa estimativa para corrigir essa variação nos espectros, uma vez que a separação entre os grupos de amostras foi eliminada.

Em seguida, avaliou-se o desempenho do modelo PLS-DA para amostras de madeira recém-cortadas, onde mais de 90% do conjunto de validação foi classificado como *outliers*, o que invalida o uso do modelo para amostras com maior teor de umidade. Logo, ficou claro que a mudança no perfil do espectro pela variação da umidade prejudica o bom desempenho do modelo e as amostras com teor de umidade maior apresentam altos valores de Hotelling T² e resíduos Q conforme o gráfico da Figura 12.

Na tentativa de resolução desse problema e levando em consideração os resultados da PCA dos dados corrigidos com OSC, aplicou-se essa ferramenta matemática tanto no conjunto de treinamento e validação de amostras secas, como no conjunto de validação de amostras de madeira recém-cortadas. A aplicação do OSC nos conjuntos de dados de amostras secas não implicou em mudança significativa nos resultados. Não houve grandes alterações nos valores de classe estimados e nem na TEF. Contudo, no que diz respeito ao conjunto de validação de amostras de madeira recém-cortada, esses espectros, mesmo corrigidos, foram classificados como *outliers*. Apesar da redução no número de *outliers* de validação, esse conjunto de amostras ainda não é compatível com as amostras secas utilizadas no conjunto de treinamento. Mais de 70% dos espectros foram classificados como *outliers* e mesmo que tenha havido o aumento da TEF após o modelo ser corrigido, ele não se torna válido.

Esses resultados permitem concluir que a correção ortogonal de sinal não foi efetiva na correção da variação de umidade nos espectros das amostras de madeira.

4 CORREÇÃO DA UMIDADE POR SECAGEM

Embora a umidade não possa ser considerada como uma característica intrínseca da madeira, a umidade é um parâmetro que afeta o espectro NIR. Por ser um material orgânico e de estrutura complexa e heterogênea, a madeira é higroscópica, retraindo-se e inchando de acordo com a umidade do ambiente. Por isso, a secagem da madeira é um processo importante para evitar mudanças físicas e mecânicas no material. Com a secagem da madeira, aumenta-se a resistência mecânica do material, reduz-se os riscos de ataques de fungos manchadores e apodrecedores e reduz-se o peso da madeira, o que diminui o custo do seu transporte (Galvão & Jankowsky, 1985).

A secagem é a perda da água da madeira, seja ela: água livre ou água de constituição química. Uma madeira pode ser considerada adequadamente seca quando tem um teor de umidade compatível com o uso pretendido e um mínimo de variação no teor de umidade (Lamb, 1994). Esse teor de umidade de equilíbrio é variável, existem fatores internos e externos que influenciam. Por exemplo, o teor de umidade (TU) depende do tipo de madeira que se estuda e até mesmo ocorre variação do TU em madeiras de uma mesma espécie quando cerne e alburno são comparados. Portanto, o TU não depende apenas da espécie, mas também de condições ambientais onde a madeira se encontra.

O método mais aplicado para a determinação do teor de umidade da madeira é o estabelecido na norma ABNT NBR 14929:2003, substituída por ABNT NBR 14929:2017, também chamado de método gravimétrico ou da estufa. Essa norma especifica o método para determinação da umidade (quantidade de água) de cavacos por secagem em estufa. A determinação do TU por esse método é feita por meio da diferença de massa das amostras antes e depois de passarem por um processo de secagem em estufa a $103 \pm 2^{\circ}\text{C}$, até massa constante.

O método é preciso, porém demorado e caro. Então, outros métodos são desenvolvidos para determinar o TU; entre eles, os medidores elétricos, tanto com princípio resistivo quanto capacitivo. Esses métodos fornecem a resposta imediata ao TU, porém não são tão precisos (Donato et al., 2015).

Uma vez que a abordagem anterior para eliminar a interferência da umidade com a ferramenta matemática OSC não foi bem-sucedida, a hipótese testada nesta etapa é para eliminar de forma experimental a água presente na madeira. As amostras de madeira sólida das espécies *Carapa guianensis* Aubl. (andiroba), *Erisma uncinatum* Warm.

(cedrinho), *Cedrela odorata* L. (cedro), *Micropholis melinoniana* Pierre. (curupixá) e *Swietenia macrophylla* King (mogno) foram submetidas a um processo de secagem, usando uma desidratadora de alimentos e uma estufa com circulação de ar. O objetivo era que os espectros obtidos posteriormente não apresentassem interferência significativa da umidade e pudessem ser classificados nos modelos de discriminação PLS-DA de maneira assertiva.

4.1 MATERIAIS E MÉTODOS

A parte experimental foi executada no Laboratório de Produtos Florestais do Serviço Florestal Brasileiro, pertencente ao Ministério do Meio Ambiente (MMA), em colaboração com o Laboratório de Automação, Quimiometria e Química Ambiental (AQQUA) do Instituto de Química da Universidade de Brasília.

4.1.1 Obtenção e preparo das amostras

As amostras foram obtidas e preparadas em duas etapas diferentes:

- *Avaliação do tempo de secagem:*

As espécies estudadas foram: *Carapa guianensis* Aubl. (andiropa), *Erismia uncinatum* Warm. (cedrinho), *Cedrela odorata* L. (cedro), *Micropholis melinoniana* Pierre. (curupixá) e *Swietenia macrophylla* King (mogno). Para o estudo, cinco amostras secas de cada espécie foram cedidas pela xiloteca Harry Van der Sloten da Área de Anatomia e Morfologia do LPF e posteriormente, cortadas nas dimensões 5,0x5,0x0,5cm. Nessa primeira etapa, avaliou-se o tempo necessário para secagem das amostras em um desidratador de alimentos com circulação de ar.

- *Desenvolvimento de novo modelo de discriminação:*

Nessa fase, 29 amostras de *Erismia uncinatum* Warm. (cedrinho), 20 amostras de *Cedrela odorata* L. (cedro) e 33 amostras de *Swietenia macrophylla* King (mogno) foram obtidas na xiloteca Harry Van der Stolen,, depois identificadas por anatomista de madeira e cortadas nas dimensões 5,0x5,0x0,5cm. As amostras foram secas em estufa com circulação de ar para construção de um novo modelo de discriminação.

As amostras em ambas as etapas tiveram sua superfície polida com lixas nº 80 e posteriormente, os espectros foram obtidos. A Figura 22 ilustra procedimento realizado:

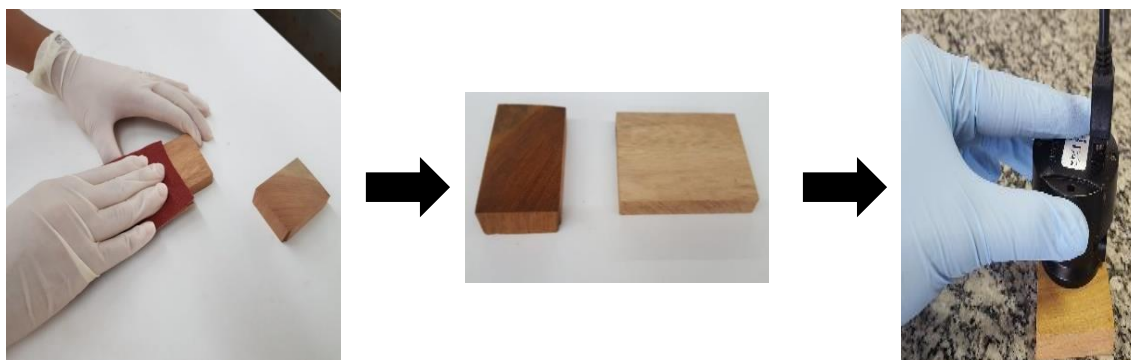


Figura 22. Procedimento de preparo das amostras e obtenção dos espectros.

4.1.2 Avaliação do tempo de secagem

Inicialmente, uma amostra de cada espécie foi selecionada e molhada debaixo de uma torneira de PVC com chuveirinho durante 1h30.

O uso do chuveirinho na torneira teve como objetivo simular uma condição de chuva nas amostras de madeira. Logo depois que as amostras foram molhadas, mediu-se a umidade delas com o auxílio do higrômetro digital KT-50B fabricado pela KLONTESER com faixa de medição entre 2 a 90% e precisão de 0,1%, e os espectros foram medidos. Deve-se ter em mente que o processo de molhar as amostras não tem intenção de saturar as mesmas com água e sim, simular uma amostra de madeira que está sendo transportada e que eventualmente pode estar sujeita a chuvas.

Em seguida, as amostras foram submetidas a diferentes tempos de secagem: 1 minuto, 5 minutos, 10 minutos, 15 minutos e 20 minutos em um desidratador de alimentos PARDAL PEG60 à temperatura de 75°C, como na Figura 23:



Figura 23. Desidratador de alimentos usado para secagem das amostras de madeira. Fonte: pardal.com.br/desidratador-peg-60.

O desidratador de alimentos possuía paredes duplas com isolamento térmico, lâmpada interna, bandejas perfuradas que permitiu a circulação de ar na amostra e sistema de ventilação forçada com o uso de uma ventoinha. Assim que as amostras eram secas, a umidade era medida e três espectros eram registrados em pontos distintos e aleatórios com o espectrômetro microNIRTM 1700. Esse procedimento foi repetido para os 5 diferentes tempos: 1 minuto, 5 minutos, 10 minutos, 15 minutos e 20 minutos.

Essa primeira etapa teve como objetivo avaliar o comportamento das amostras à medida que eram secas no desidratador e estimar um tempo ótimo para secagem da amostra. O objetivo era que o tempo de secagem fosse suficiente para diminuir os valores

de umidade das amostras, eliminar a interferência da umidade nos espectros e validar os espectros medidos para essas amostras no modelo de discriminação para amostras secas, construído anteriormente.

Com a avaliação do comportamento das amostras e análise dos espectros obtidos, determinou-se um tempo ótimo para secagem da madeira em 20 minutos. Assim que esse tempo foi estabelecido, o experimento foi realizado novamente, mas dessa vez, com replicatas. Quatro amostras de cada uma das espécies foram selecionadas, molhadas debaixo da torneira por 1h30 e secas no desidratador à 75°C por 20 minutos. Contudo, como será detalhado nos resultados, foi necessário repetir esse ensaio com 30 minutos de secagem. Para cada amostra, três espectros foram obtidos em pontos distintos e aleatórios e em seguida, testados no modelo de discriminação das espécies construído para amostras secas.

4.1.3 Desenvolvimento de novo modelo de discriminação:

Nesta etapa, foram selecionadas 82 novas amostras, sendo 29 de *Erismia uncinatum* Warm. (cedrinho), 20 de *Cedrela odorata* L. (cedro) e 33 de *Swietenia macrophylla* King (mogno) para desenvolvimento de novo modelo de discriminação, o qual considera as novas condições experimentais definidas na seção anterior. Nesta nova tentativa, procurou-se refazer o modelo com amostras secas (umidade < 15%) e que foram molhadas. Todas essas amostras passaram pelo mesmo procedimento de secagem, de forma a homogeneizar o preparo de todas as amostras na fase de treinamento e validação.

De modo aleatório, as amostras foram selecionadas para comporem o conjunto de treinamento e validação do modelo e a seleção de amostras resultou na seguinte divisão, conforme mostra a Tabela 6:

Tabela 6. Quantidade de amostras selecionadas para o conjunto de treinamento e validação.

| Espécie | Número de amostra | |
|--------------------------------------|-------------------|-----------|
| | Treinamento | Validação |
| <i>Erismia uncinatum</i> (cedrinho) | 17 | 12 |
| <i>Cedrela odorata</i> (cedro) | 10 | 10 |
| <i>Swietenia macrophylla</i> (mogno) | 15 | 18 |

Em seguida, as amostras dos conjuntos de treinamento e validação foram subdivididas em: amostras úmidas e secas, conforme mostra a Tabela 7.

Tabela 7. Quantidade de amostras selecionadas dos conjuntos de treinamento e validação para as condições: úmida e seca.

| Espécie | Número de amostra | | | |
|--------------------------------------|-------------------|-------|-----------|-------|
| | Treinamento | | Validação | |
| | Úmidas | Secas | Úmidas | Secas |
| <i>Erismia uncinatum</i> (cedrinho) | 7 | 10 | 6 | 6 |
| <i>Cedrela odorata</i> (cedro) | 5 | 5 | 5 | 5 |
| <i>Swietenia macrophylla</i> (mogno) | 7 | 8 | 9 | 9 |

As amostras da condição úmida foram submetidas ao seguinte procedimento: molhadas debaixo da torneira por 30 minutos e em seguida, secas em uma estufa com circulação/renovação de ar à 80°C por 20 minutos. As amostras da condição seca não foram molhadas, mas também foram colocadas direto para secagem na estufa à 80°C por 20 minutos. O procedimento de secagem também foi aplicado às amostras secas porque, no caso de uma amostra de madeira apreendida, não se terá disponível as condições exatas, as quais aquela amostra foi submetida. Portanto, todas as amostras passariam pelo mesmo procedimento. A estufa usada nessa etapa está ilustrada na Figura 24:



Figura 24. Estufa de secagem com circulação/renovação de ar usada para secagem das amostras. Fonte: marconi.com.br.

O desidratador de alimentos não foi usado nessa etapa porque não estava mais disponível para a realização desses procedimentos. Porém, ambos os equipamentos possuíam circulação de ar. No caso da estufa, essa possuía sistema de convecção mecânica por ar forçado no sentido horizontal, dispositivo mecânico para renovação de

ar e prateleiras gradeadas que permitiam a circulação do ar na amostra. Assim que as amostras eram retiradas da estufa, aguardava-se alguns minutos para que as amostras entrassem em equilíbrio com a temperatura ambiente ($\approx 25^{\circ}\text{C}$). Por fim, as amostras foram lixadas com lixa de granulação número 80, o pó era removido com pincel e os espectros eram obtidos em três pontos distintos e aleatórios para cada amostra.

4.2 RESULTADOS E DISCUSSÃO

4.2.1 Avaliação do tempo de secagem

Inicialmente, uma amostra de cada espécie foi submetida ao procedimento descrito em 4.1.2. A umidade das amostras foi medida assim que foram molhadas e também depois de 1, 5, 10, 15 e 20 minutos à 75°C no desidratador de alimentos. Os valores medidos estão dispostos na Tabela 8 a seguir:

Tabela 8. Valores de umidade de uma amostra das cinco espécies nas condições: úmida e depois de 1, 5, 10, 15 e 20 minutos no desidratador.

| Espécie | Umidade (%) | | | | | |
|---|-------------|----------|-----------|------------|------------|------------|
| | Úmida | 1 minuto | 5 minutos | 10 minutos | 15 minutos | 20 minutos |
| <i>Carapa guianensis</i> (andiroba) | * | * | 38,3 | 35,4 | 13,1 | 10,1 |
| <i>Erisma uncinatum</i> (cedrinho) | * | * | 22,9 | 11,6 | 11,5 | 11,4 |
| <i>Cedrela odorata</i> (cedro) | * | * | 56,9 | 20,3 | 9,2 | 9,2 |
| <i>Micropholis melinoniana</i> (curupixá) | * | * | 44,3 | 55,1 | 29,1 | 24,3 |
| <i>Swietenia macrophylla</i> (mogno) | * | * | 11,1 | 8,3 | 7,8 | 6,1 |

(*) indica condição saturada, onde o higrômetro não conseguiu mensurar a umidade da madeira.

Os valores da Tabela 8 indicam a diminuição do valor de umidade nas amostras à medida que passavam mais tempo no desidratador de alimentos. Dentre as espécies, o cedro (*Cedrela odorata*) apresentou uma variação maior no valor de umidade entre a condição úmida e após secagem por 20 minutos no desidratador. Essa variação pode ser explicada porque a espécie de cedro (*Cedrela odorata*) possui um valor de densidade (0,38 g/cm³) menor do que as demais espécies (Soares et al., 2017) e, por ser menos densa, possa ter absorvido e liberado mais facilmente a água na superfície.

Diferentemente do cedro (*Cedrela odorata*), a espécie de mogno (*Swietenia macrophylla*) varia pouco com relação aos valores de umidade. Depois de molhada, foi a espécie que menos absorveu água e os valores de umidade na condição úmida são os menores se comparados com os das demais espécies. O mogno (*Swietenia macrophylla*) possui uma densidade que varia de 0,5-0,72 g/cm³ (Soares et al., 2017), apresentando valor de densidade maior do que o das outras espécies, isso talvez possa explicar a baixa absorção da água pela madeira e como a madeira dificilmente absorve a água, dificilmente também a água é liberada.

Em seguida, avaliaram-se os espectros das amostras nas condições: molhado, 5, 10, 20 minutos no desidratador e seco. Cada amostra teve três espectros medidos em

diferentes posições e depois, a média entre os três espectros foi calculada. Os espectros médios estão plotados na Figura 25:

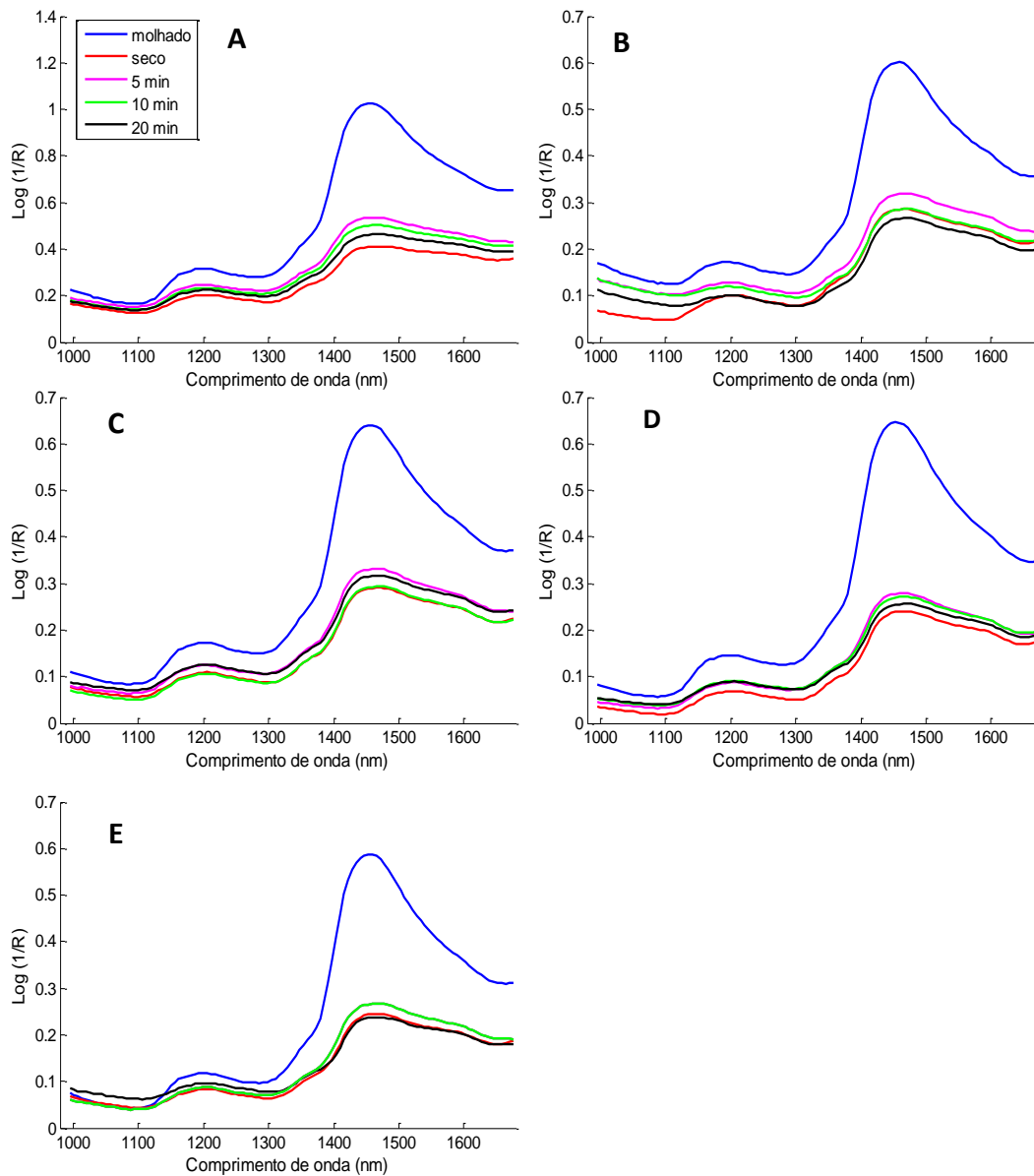


Figura 25. Espectro médio de (A) andiroba (*Carapa guianensis*), (B) cedrinho (*Erismia uncinatum*), (C) cedro (*Cedrela odorata*), (D) curupixá (*Micropholis melinoniana*) e (E) mogno (*Swietenia macrophylla*) nas condições: molhado, seco, 5, 10 e 20 minutos.

Na Figura 25, pode-se observar o aumento da intensidade em $\log(1/R)$ em função do aumento da umidade na madeira na região de comprimento de onda entre 1.400 e 1.500 nm e a diminuição dessa intensidade a medida que a madeira é seca no desidratador. Todas as espécies apresentam o mesmo comportamento. Na Figura 25 (A), é visível o decaimento da intensidade no eixo y ao longo do tempo, a escala no eixo varia entre 0 e 1.4, enquanto a dos demais gráficos varia de 0 e 0.7, indicando uma maior variação para a espécie de andiroba (*Carapa guianensis*). Observa-se também que o espectro da madeira, após 20 minutos no desidratador, possui valores de intensidade próximos aos do espectro da madeira seca. Por esse motivo, decidiu-se avaliar os espectros da madeira após 20 minutos de secagem no modelo de discriminação com madeira seca. Dessa forma, esperava-se que o espectro da madeira nesse momento fosse compatível com os demais das amostras secas do conjunto de treinamento do modelo de discriminação. A hipótese era que a interferência da umidade tivesse sido reduzida e o modelo de discriminação construído com amostras secas pudessem classificar as amostras de modo assertivo.

Para testar essa hipótese, os espectros da madeira, após 20 minutos de secagem no desidratador, foram validados no modelo de discriminação construído com madeira na condição seca e os gráficos de Hotelling T^2 e resíduos Q foram analisados para verificar a presença de *outliers* e se os espectros eram compatíveis com o conjunto de treinamento usado para a construção do modelo. Na Figura 26, tem-se os gráficos de Hotelling T^2 e resíduos Q para as cinco espécies:

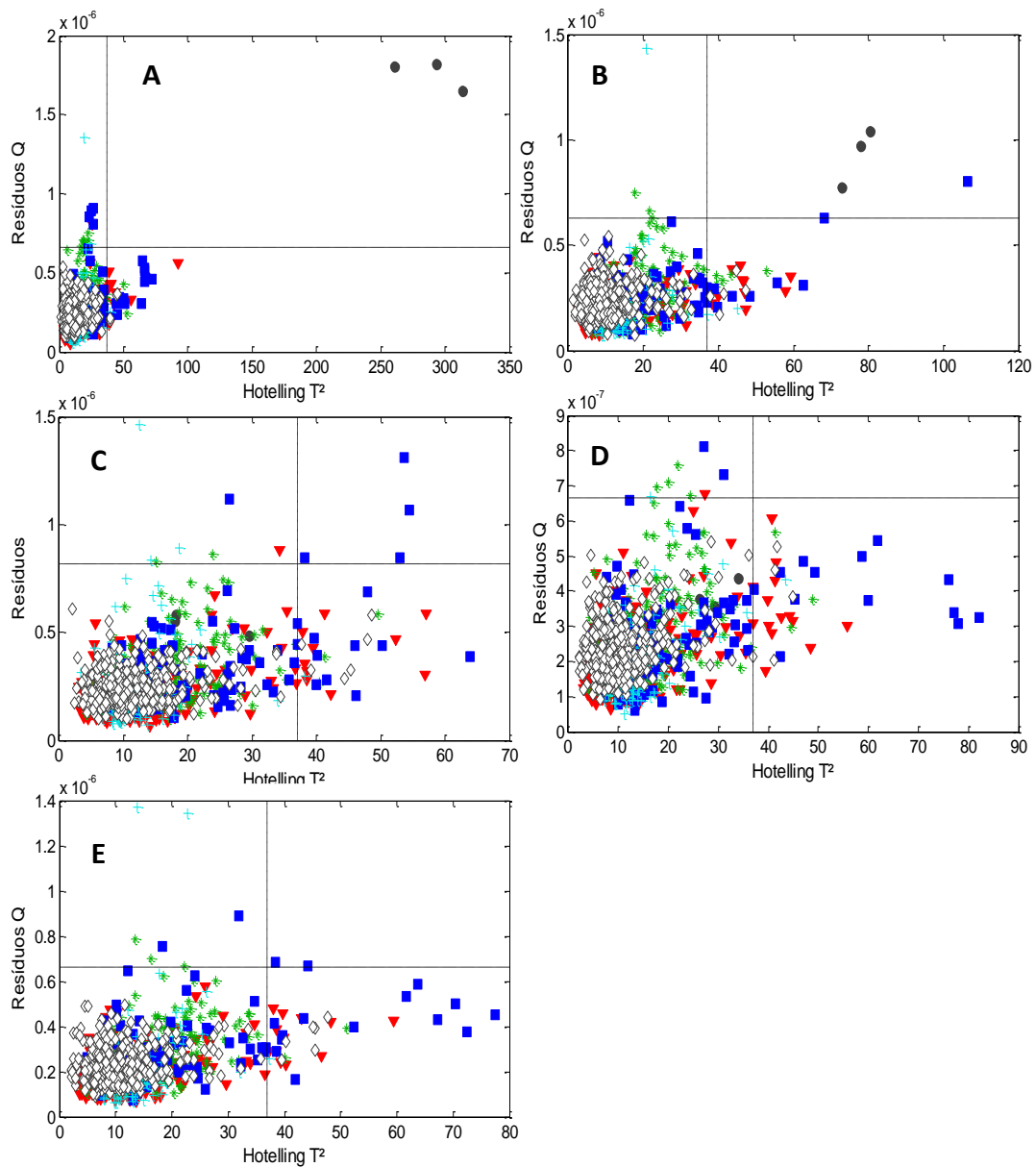


Figura 26. Gráfico dos valores de T^2 de Hotelling e resíduos Q de (A) andiroba (*Carapa guianensis*), (B) cedrinho (*Erisma uncinatum*), (C) cedro (*Cedrela odorata*), (D) curupixá (*Micropholis melinoniana*) e (E) mogno (*Swietenia macrophylla*) para uma amostra testada após 20 minutos no desidratador. Legenda: andiroba (*Carapa guianensis*) (\blacktriangledown), cedrinho (*Erisma uncinatum*) (*), cedro (*Cedrela odorata*) (\blacksquare), curupixá (*Micropholis melinoniana*) (\blacklozenge) e amostra testada (\bullet).

Na Figura 26, pode-se observar comportamentos diferentes entre as cinco espécies florestais. Em Figura 26 (A), a amostra de andiroba (*Carapa guianensis*) testada é classificada como *outlier* porque está situada no quadrante que apresenta valores de Hotelling T² e resíduos Q acima dos limites de confiança de 99,9%. Ou seja, essa amostra não é compatível com o conjunto de treinamento. Em Figura 26 (B), acontece o mesmo, a amostra testada é *outlier*, apesar de apresentar valores menores de Hotelling T² e resíduos Q se comparados a Figura 26 (A), mas ainda assim está fora dos limites de confiança.

Na Figura 26 (C), a amostra de cedro (*Cedrela odorata*) está dentro dos limites de confiança, o que indica compatibilidade com o conjunto de treinamento e a amostra não é classificada como *outlier*. Da mesma forma, a amostra de curupixá (*Micropholis melinoniana*) testada na Figura 26 (D) também apresenta valores de Hotelling T² e resíduos Q dentro dos limites de confiança e a amostra não é *outlier*. Por fim, na Figura 26 (E), a amostra de mogno (*Swietenia macrophylla*) testada também foi compatível com o resto das amostras pertencentes ao conjunto de treinamento do modelo.

Depois de avaliar os gráficos de Hotelling T² e resíduos Q, avaliou-se também o gráfico com os valores de classe estimados para as amostras testadas no modelo. A Figura 27 ilustra os gráficos de valores de classe estimados para as cinco espécies:

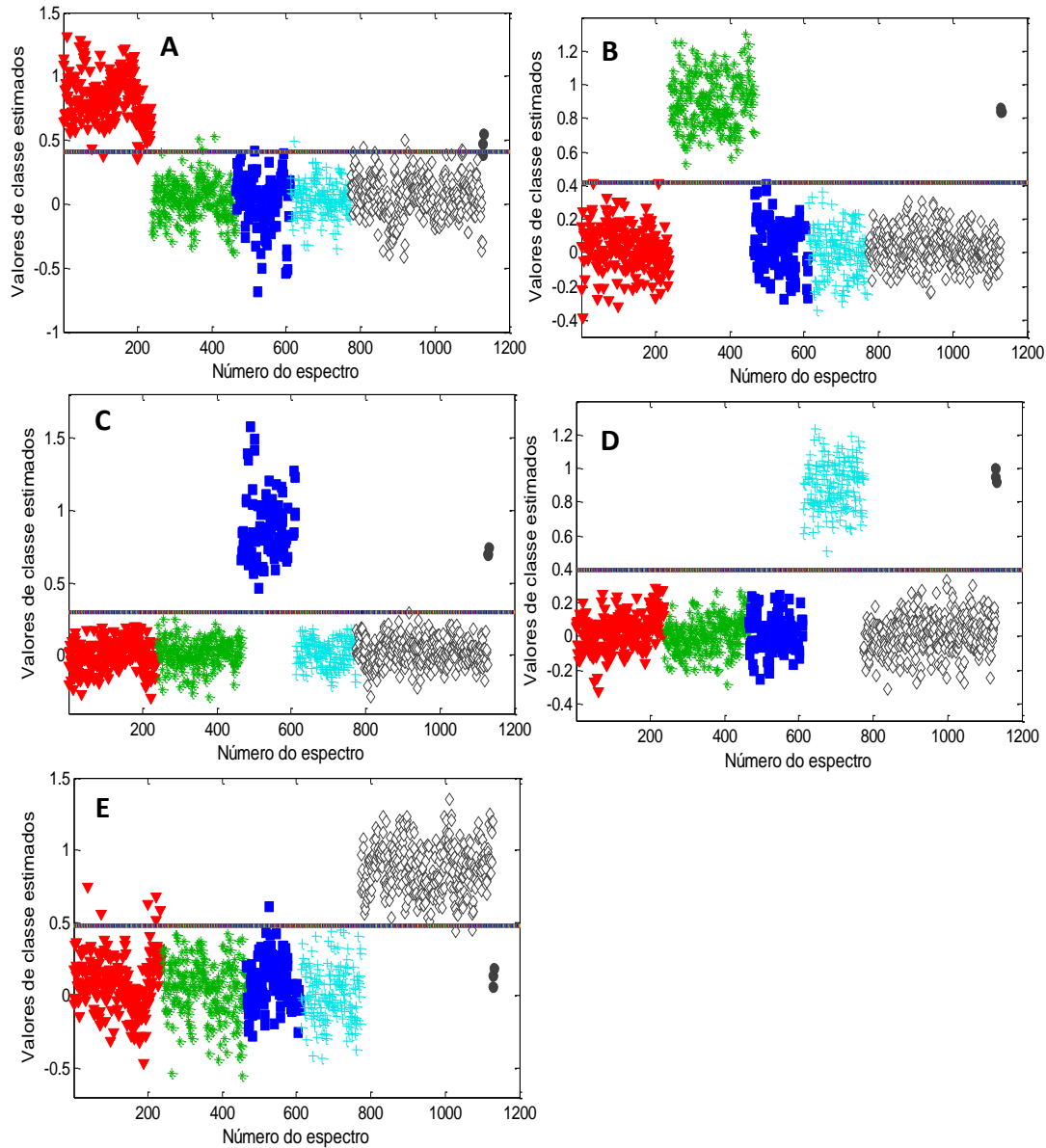


Figura 27. Valores de classe estimados para os conjuntos de treinamento e para a amostra testada de (A) andiroba (*Carapa guianensis*), (B) cedrinho (*Erisma uncinatum*), (C) cedro (*Cedrela odorata*), (D) curupixá (*Micropholis melinoniana*) e (E) mogno (*Swietenia macrophylla*) após 20 minutos no desidratador. Legenda: andiroba (*Carapa guianensis*) (▼), cedrinho (*Erisma uncinatum*) (*), cedro (*Cedrela odorata*) (■), curupixá (*Micropholis melinoniana*) (+), mogno (*Swietenia macrophylla*) (◇), (•) amostra testada e (—) limite de discriminação.

Na Figura 27, é analisado se o modelo classifica corretamente as amostras testadas. Em Figura 27 (A), um dos espectros está abaixo do limite de discriminação do modelo, enquanto os outros dois espectros são classificados corretamente. Seguindo o

critério de que o modelo acerta quando dois de três espectros de uma amostra é classificado corretamente, então pode-se concluir que o modelo classificou como andiroba (*Carapa guianensis*) a amostra testada. Em Figura 27 (B), ainda que o gráfico de Hotelling T² e resíduos Q tenha classificado a amostra como *outlier*, os espectros foram classificados como pertencentes à espécie de cedrinho (*Erisma uncinatum*) e o modelo apresentou 100% de acerto.

As amostras em Figura 27 (C) e Figura 27 (D) são classificadas de forma correta pelo modelo, onde todos os espectros testados estão acima do limite de discriminação. De acordo com os gráficos de Hotelling T² e resíduos Q, essas amostras são compatíveis com o conjunto de treinamento e não são identificadas como *outliers*. Os resultados mostram que os modelos de cedro (*Cedrela odorata*) e curupixá (*Micropholis melinoniana*) apresentaram taxa de acerto de 100% e o processo de secagem para as amostras testadas permitiu que elas fossem validadas no modelo de discriminação. Por fim, em Figura 27 (E), a amostra de mogno (*Swietenia macrophylla*), apesar de não ter apresentado valores de Hotelling T² e resíduos Q fora dos limites de confiança, foi classificada como pertencente a outras espécies, logo o modelo errou na classificação.

Com a observação dos gráficos, foi possível verificar que o tempo de 20 minutos permitiu a secagem das amostras e no caso das espécies de cedro (*Cedrela odorata*) e curupixá (*Micropholis melinoniana*), depois de secas, as amostras foram compatíveis com o conjunto de treinamento e o modelo as classificou de forma assertiva. Essa análise foi realizada apenas com a observação de uma amostra de cada espécie, porém para garantir a representatividade no conjunto de amostras e um resultado confiável, o mesmo procedimento foi realizado para outras 4 amostras de cada espécie e os resultados foram avaliados a seguir.

Inicialmente, mediu-se os valores de umidade das amostras nas condições: molhado, 20 e 30 minutos no desidratador. Incluiu-se uma nova condição de 30 minutos com o objetivo de abranger uma faixa maior de secagem das amostras. Os valores de umidade medidos estão organizados na Tabela 9:

Tabela 9. Valores de umidade das quatro amostras das cinco espécies nas condições: úmida e depois de 20 e 30 minutos no desidratador.

| Espécie | Umidade (%) | | | |
|--|-------------|-------|------------|------------|
| | Amostra | Úmido | 20 minutos | 30 minutos |
| <i>Carapa guianensis</i> (andiroba) | 1 | 41,7 | 20,4 | 16,1 |
| | 2 | 46,6 | 19,8 | 15,7 |
| | 3 | 47,0 | 20,0 | 16,1 |
| | 4 | 41,7 | 19,1 | 15,7 |
| <i>Erismia uncinatum</i> (cedrinho) | 1 | 38,8 | 18,4 | 18,0 |
| | 2 | 42,0 | 20,7 | 19,6 |
| | 3 | 37,3 | 19,8 | 19,1 |
| | 4 | 39,4 | 19,8 | 19,4 |
| <i>Cedrela odorata</i> (cedro) | 1 | 62,2 | 8,4 | 8,0 |
| | 2 | 66,7 | 8,5 | 8,4 |
| | 3 | 67,4 | 8,9 | 8,7 |
| | 4 | 68,2 | 8,8 | 8,3 |
| <i>Micropholis melinoniana</i> (curupixá) | 1 | 13,6 | 9,7 | 9,4 |
| | 2 | 14,3 | 9,0 | 8,5 |
| | 3 | 14,7 | 10,0 | 9,1 |
| | 4 | 13,6 | 10,3 | 9,1 |
| <i>Swietenia macrophylla</i> (mogno) | 1 | 10,3 | 8,0 | 7,6 |
| | 2 | 10,0 | 9,5 | 8,0 |
| | 3 | 11,0 | 8,7 | 7,9 |
| | 4 | 11,7 | 8,4 | 7,3 |

Na Tabela 9, as amostras apresentam o mesmo comportamento observado na Tabela 8. Observa-se a variação da umidade em função do tempo de secagem para cada espécie, onde o cedro (*Cedrela odorata*) apresenta maior variação nos valores de umidade e o mogno (*Swietenia macrophylla*), uma menor variação. De modo geral, os valores são coerentes ao que se espera, uma vez que diminuem ao passo que a amostra é seca no desidratador.

Em seguida, as novas amostras de teste foram validadas no modelo de discriminação construído com madeira na condição seca. Inicialmente, foram validados os espectros após 20 minutos de secagem e os gráficos de Hotelling T² e resíduos Q foram analisados. Os gráficos estão ilustrados na Figura 28:

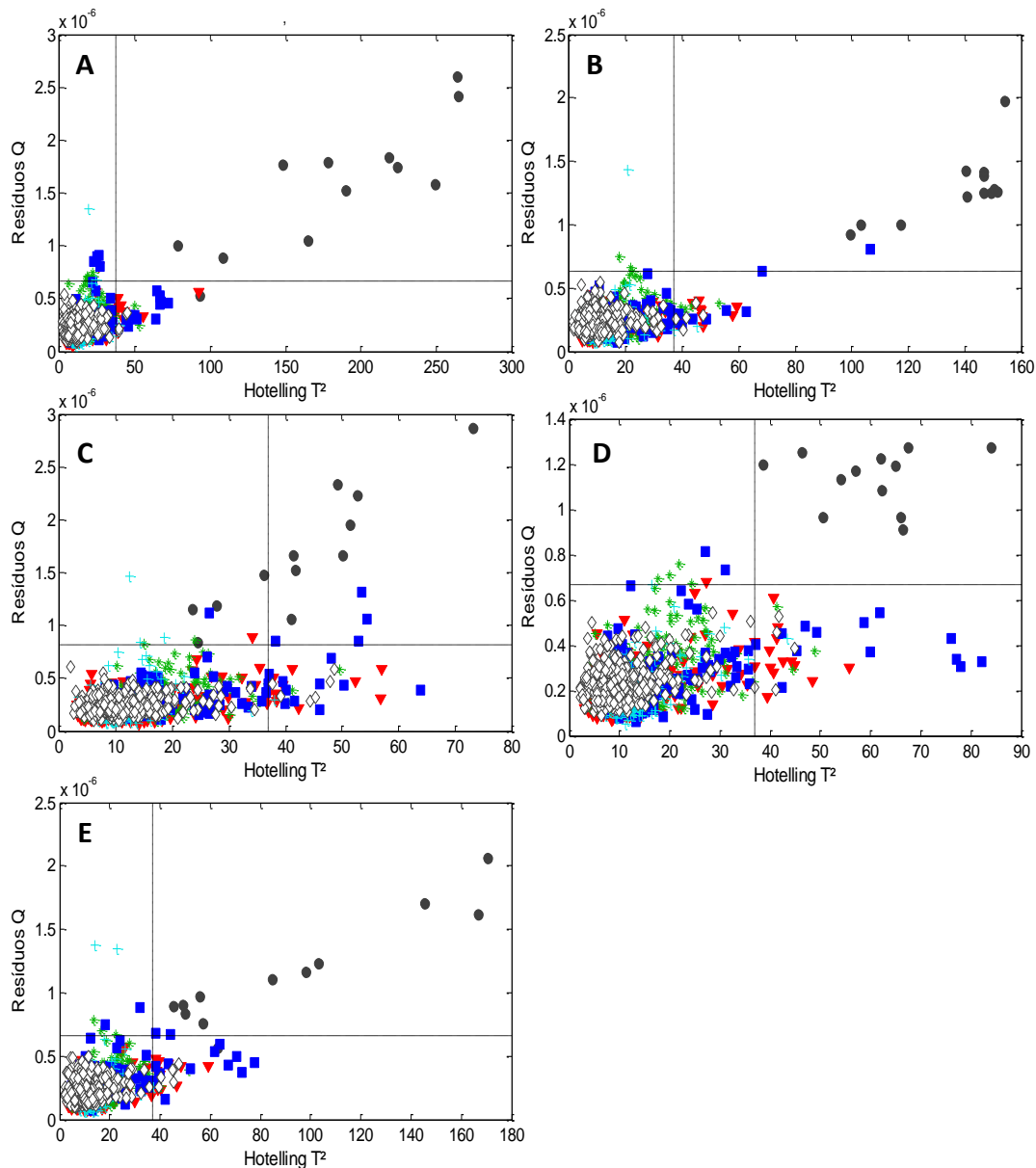


Figura 28. Gráfico dos valores de T^2 de Hotelling e resíduos Q de (A) andiroba (*Carapa guianensis*), (B) cedrinho (*Erismia uncinatum*), (C) cedro (*Cedrela odorata*), (D) curupixá (*Micropholis melinoniana*) e (E) mogno (*Swietenia macrophylla*) para 4 amostras testadas após 20 minutos no desidratador. Legenda: andiroba (*Carapa guianensis*) (\blacktriangledown), cedrinho (*Erismia uncinatum*) (*), cedro (*Cedrela odorata*) (\blacksquare), curupixá (*Micropholis melinoniana*) (+) mogno (*Swietenia macrophylla*) (\diamond) e amostra testada (\bullet).

Na Figura 28, para as espécies de andiroba (*Carapa guianensis*) na Figura 28 (A), cedrinho (*Erismia uncinatum*) na Figura 28 (B), curupixá (*Micropholis melinoniana*) na Figura 28 (D) e mogno (*Swietenia macrophylla*) na Figura 28 (E), as 4 amostras testadas foram classificadas como *outliers* porque apresentaram valores de Hotelling T^2 e resíduos em Q acima dos limites de confiança. Isso indica que as amostras não são compatíveis

com o conjunto de treinamento e o modelo não poderia ser aplicado para essas amostras. Para a espécie de cedro (*Cedrela odorata*) na Figura 28 (C), 70% dos espectros foram identificados como *outliers*, apresentando valores fora dos limites para Hotelling T² e resíduos Q, enquanto 30% estiveram dentro do limite de confiança para Hotelling T², mas fora do limite para resíduos Q. Ainda que as amostras tenham sido classificadas como *outliers*, foram analisados os gráficos de valores de classe estimados que estão apresentados na Figura 29:

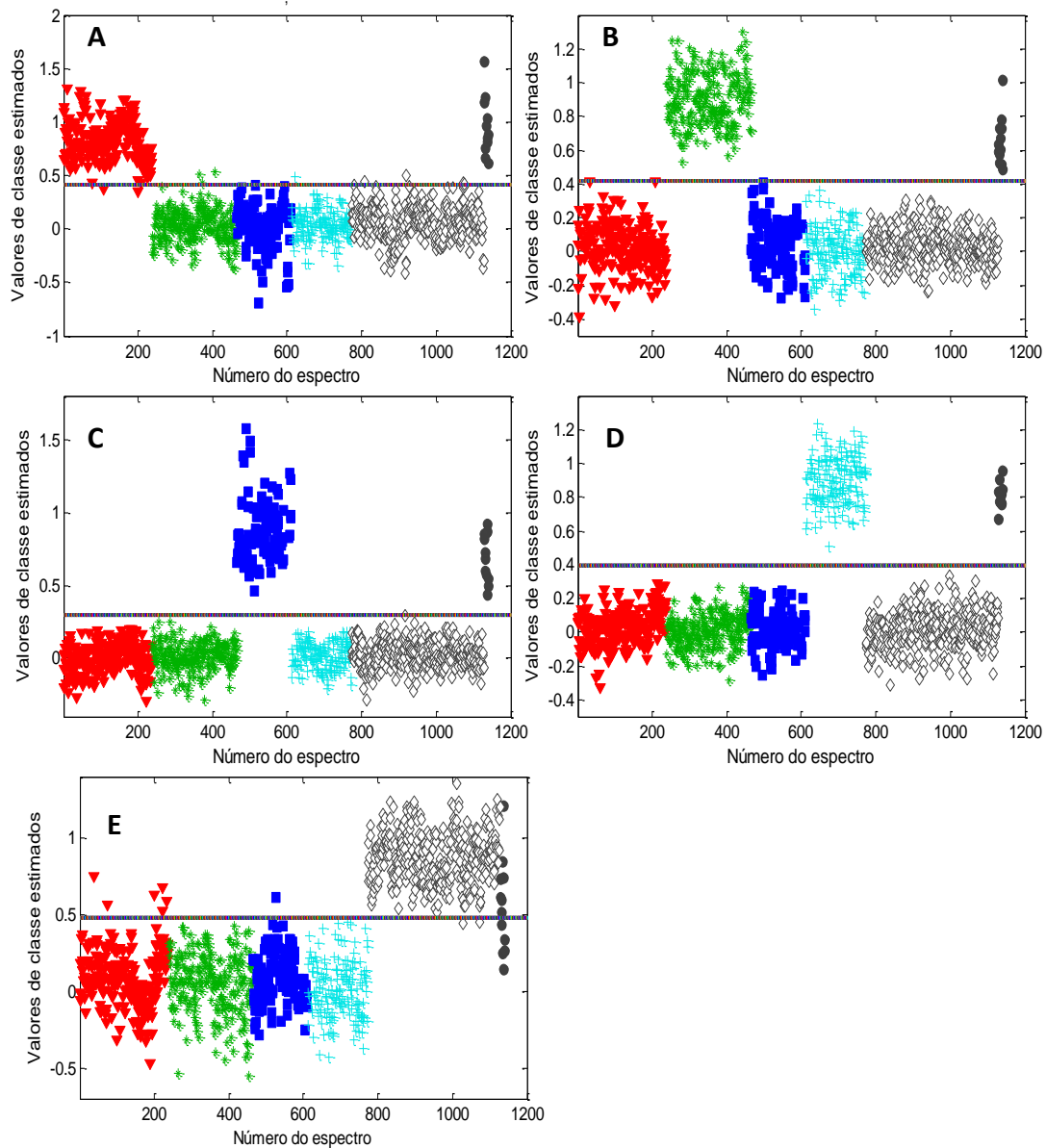


Figura 29. Valores de classe estimados para os conjuntos de treinamento e para a amostra testada de (A) andiroba (*Carapa guianensis*), (B) cedrinho (*Erisma uncinatum*), (C) cedro (*Cedrela odorata*), (D) curupixá (*Micropholis melinoniana*) e (E) mogno (*Swietenia macrophylla*) após 20 minutos no desidratador. Legenda: andiroba (*Carapa guianensis*) (▼), cedrinho (*Erisma uncinatum*) (*), cedro (*Cedrela odorata*) (■), curupixá

(*Micropholis melinoniana*) (+) e mogno (*Swietenia macrophylla*) (◇), (•) amostra testada e (— —) limite de discriminação.

É interessante notar que as espécies de andiroba (*Carapa guianensis*) na Figura 29 (A), cedrinho (*Erismia uncinatum*) na Figura 29 (B), cedro (*Cedrela odorata*) na Figura 29 (C), curupixá (*Micropholis melinoniana*) na Figura 29 (D) tiveram todas as suas amostras classificadas corretamente, mesmo sendo amostras identificadas como *outliers* e no que diz respeito ao mogno (*Swietenia macrophylla*) na Figura 29 (E), houve amostras classificadas como mogno (*Swietenia macrophylla*), mas também como de outras espécies, indicando um desempenho insatisfatório do modelo. A Figura 28 já indica que o modelo não é válido para esse tipo de amostras, porém é pertinente a análise da Figura 29 para avaliar o comportamento das amostras e a dispersão dos espectros dentro de cada classe. Essa análise pode dar mais informações para solucionar o problema de invalidez do método.

Uma vez que as amostras testadas após 20 minutos de secagem no desidratador não foram compatíveis com o conjunto de amostras usado na etapa de treinamento, verificou-se o desempenho dos modelos para amostras secas após 30 minutos no desidratador. Os gráficos de Hotelling T² e resíduos Q e valores de classe estimados estão dispostos a seguir nas Figuras 30 e 31:

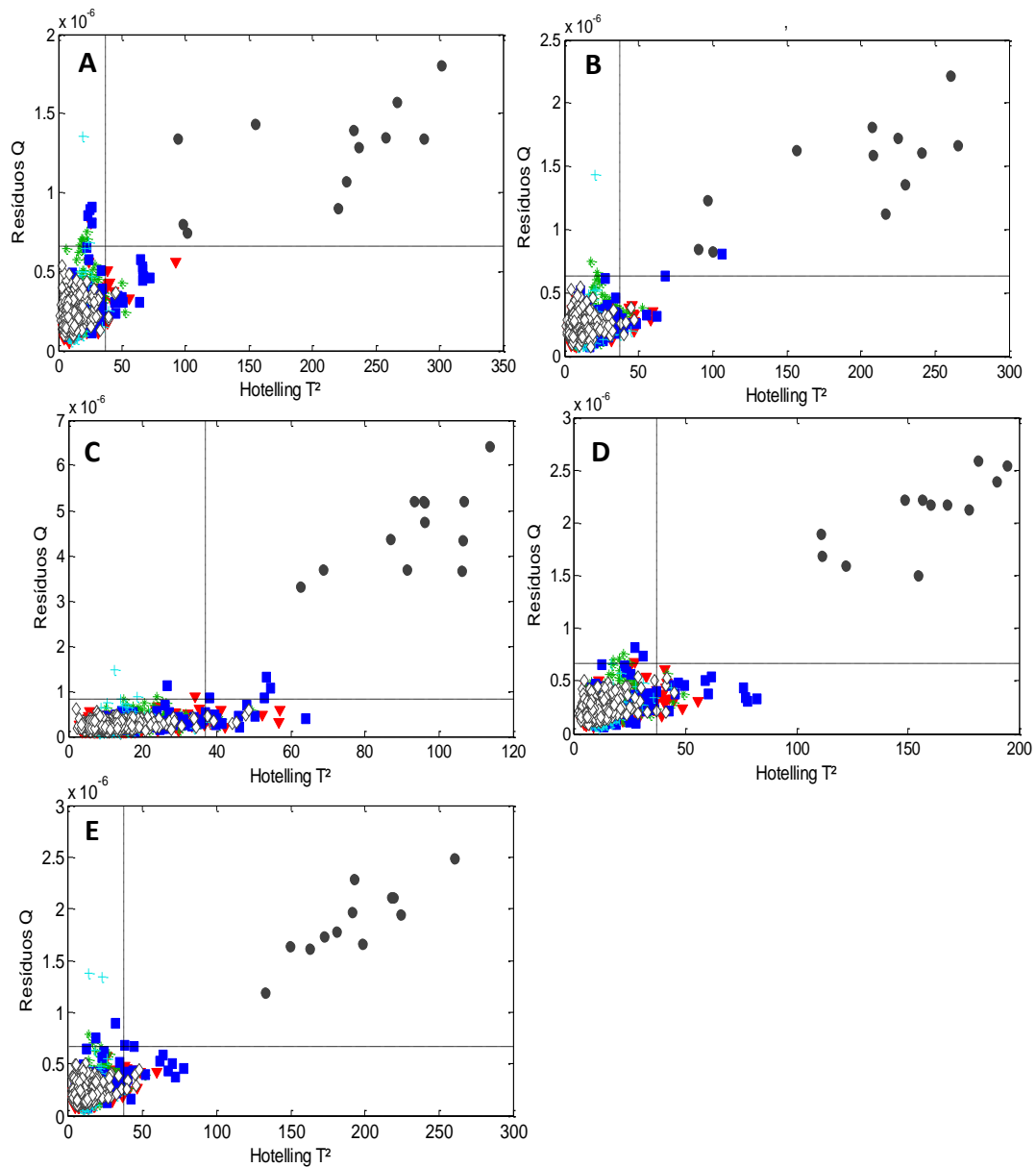


Figura 30. Gráfico dos valores de T² de Hotelling e resíduos Q de (A) andiroba (*Carapa guianensis*), (B) cedrinho (*Erisma uncinatum*), (C) cedro (*Cedrela odorata*), (D) curupixá (*Micropholis melinoniana*) e (E) mogno (*Swietenia macrophylla*) para 4 amostras testadas após 30 minutos no desidratador. Legenda: andiroba (*Carapa guianensis*) (▼), cedrinho (*Erisma uncinatum*) (*), cedro (*Cedrela odorata*) (■), curupixá (*Micropholis melinoniana*) (+) mogno (*Swietenia macrophylla*) (◇) e amostra testada (•).

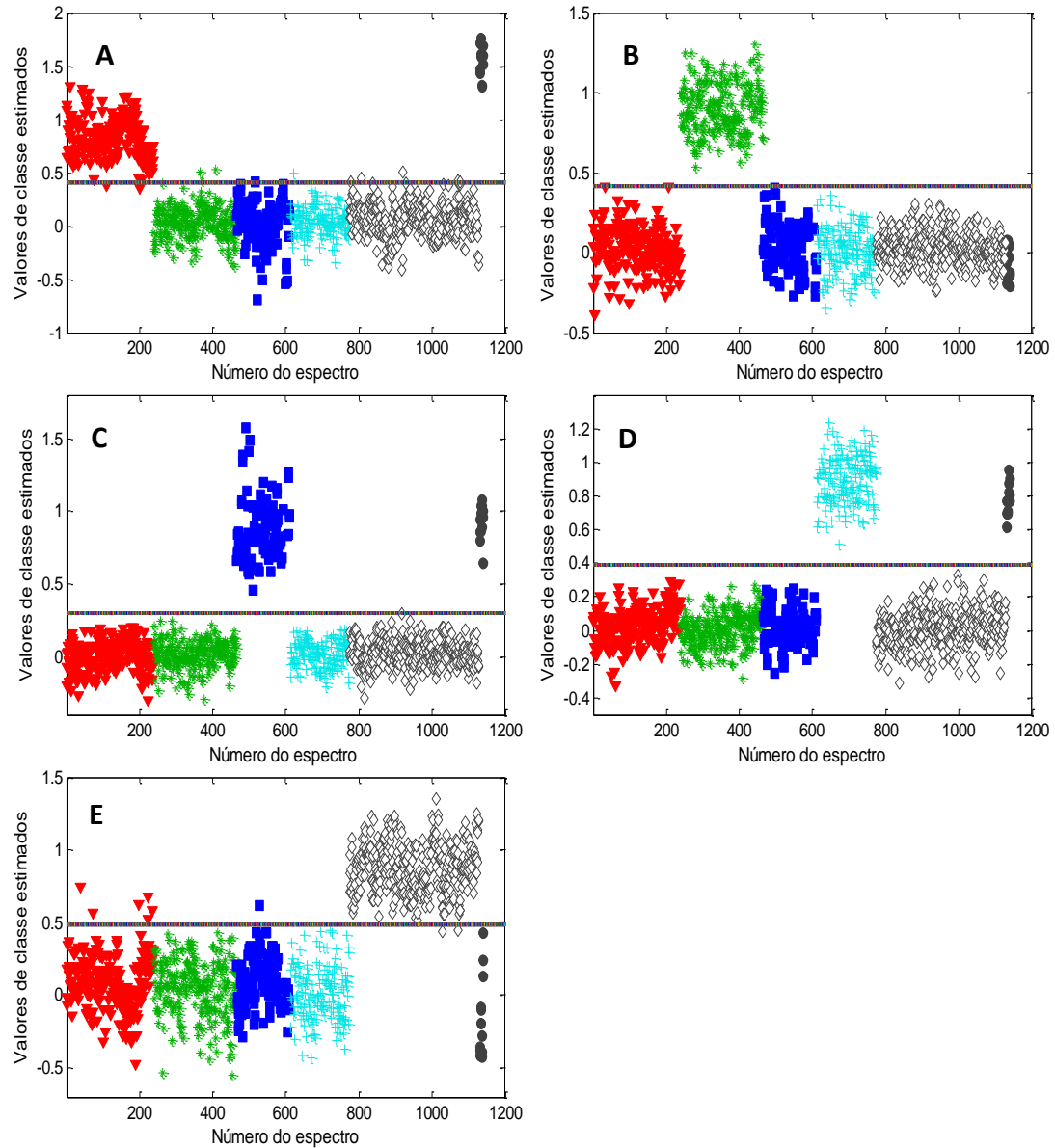


Figura 31. Valores de classe estimados para os conjuntos de treinamento e para a amostra testada de (A) andiroba (*Carapa guianensis*), (B) cedrinho (*Erismia uncinatum*), (C) cedro (*Cedrela odorata*), (D) curupixá (*Micropholis melinoniana*) e (E) mogno (*Swietenia macrophylla*) após 30 minutos no desidratador. Legenda: andiroba (*Carapa guianensis*) (\blacktriangledown), cedrinho (*Erismia uncinatum*) (*), cedro (*Cedrela odorata*) (\blacksquare), curupixá (*Micropholis melinoniana*) (+) e mogno (*Swietenia macrophylla*) (\diamond) e (— —) limite de discriminação.

Na Figura 30, observa-se que para todas as cinco espécies florestais, as amostras testadas foram identificadas como *outliers*, porque apresentaram valores de Hotelling T^2 e resíduos Q acima dos limites de confiança. É interessante notar isso porque o aumento

do tempo das amostras no desidratador piorou o desempenho das amostras no que diz respeito ao gráfico de identificação de *outliers*. Além disso, os gráficos de valores de classe estimados foram analisados e para cedrinho (*Erismia uncinatum*) na Figura 31 (B) e mogno (*Swietenia macrophylla*) na Figura 31 (E), os modelos erraram a classificação de todas as amostras testadas.

De modo geral, a secagem no desidratador não permitiu que as amostras testadas fossem compatíveis com o conjunto de amostras de treinamento, uma vez que muitos dos espectros das amostras foram identificados como *outliers* e os modelos de cedrinho (*Erismia uncinatum*) e mogno (*Swietenia macrophylla*) não classificaram de maneira assertiva as amostras validadas. Sendo assim, estabeleceu-se uma nova estratégia para diminuir a interferência da umidade nos espectros e garantir o acerto dos modelos de discriminação.

4.2.2 Desenvolvimento de novo modelo de discriminação

Nesta etapa, devido às limitações de tempo para a execução da parte experimental, foi desenvolvido um novo modelo de discriminação apenas considerando as espécies de cedrinho (*Erisma uncinatum*), cedro (*Cedrela odorata*) e mogno (*Swietenia macrophylla*), utilizando um conjunto reduzido de amostras detalhado na seção 4.1.3. Neste modelo, foram incluídas amostras nas condições úmida e seca e as amostras seguiram o procedimento descrito em 4.1.3. O objetivo foi estabelecer um mesmo procedimento para as amostras nas condições úmida e seca.

Desenvolveram-se os três modelos de discriminação: cedrinho (*Erisma uncinatum*), cedro (*Cedrela odorata*) e mogno (*Swietenia macrophylla*). Em seguida, analisou-se os gráficos de Hotelling T² e resíduos Q para as três espécies na Figura 32:

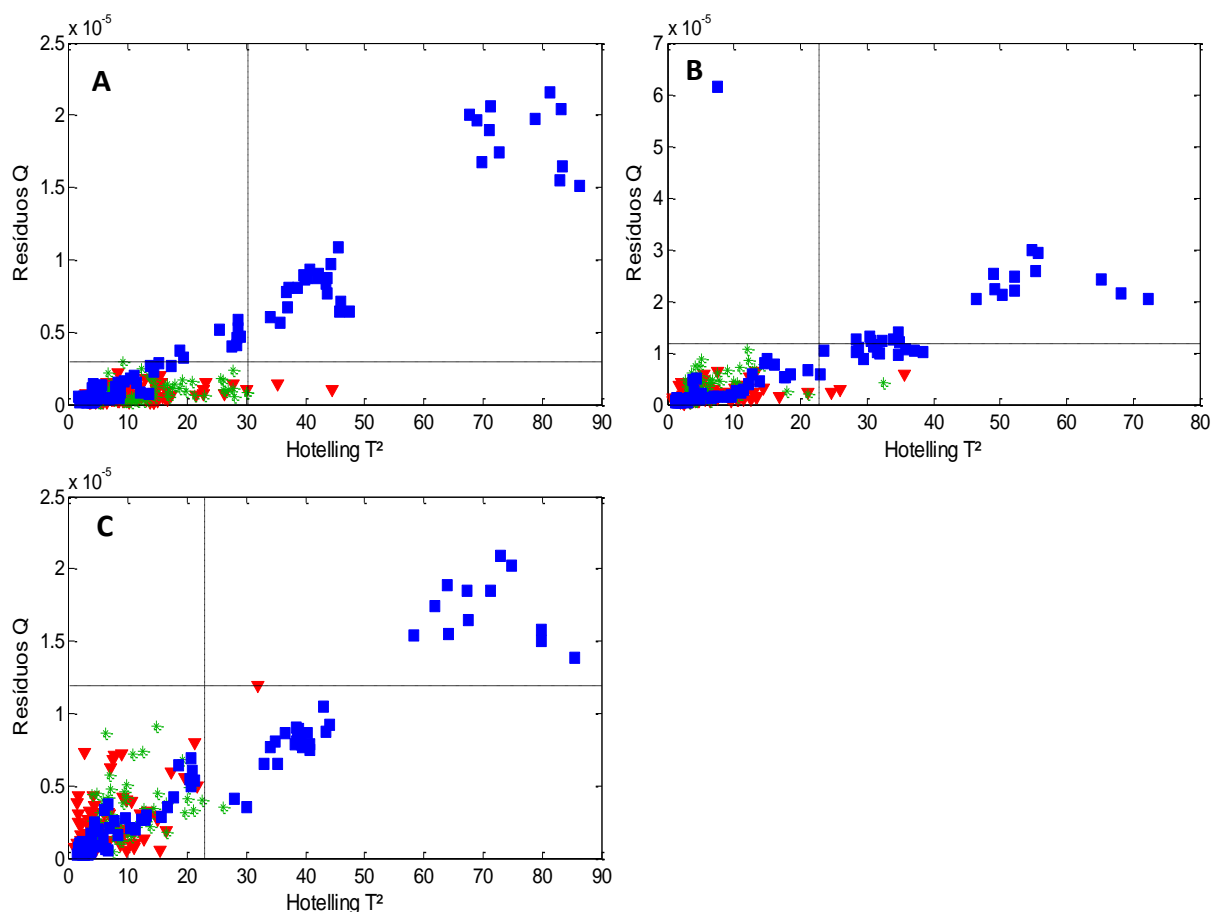


Figura 32. Gráfico dos valores de T² de Hotelling e resíduos Q de (A) cedrinho (*Erisma uncinatum*), (B) cedro (*Cedrela odorata*) e (C) mogno (*Swietenia macrophylla*). Legenda: cedrinho (*Erisma uncinatum*) (▼), cedro (*Cedrela odorata*) (*), mogno (*Swietenia macrophylla*) (■).

Nos gráficos das três espécies, existem espectros de mogno (*Swietenia macrophylla*) apresentando valores de Hotelling T^2 e resíduos Q fora dos limites de confiança. Esses espectros correspondem às amostras de mogno (*Swietenia macrophylla*) que foram molhadas e posteriormente, secas na estufa; isso quer dizer que essas amostras são *outliers* e não são compatíveis com as demais amostras do modelo. Em seguida, analisou-se os gráficos de valores de classe estimados na Figura 33:

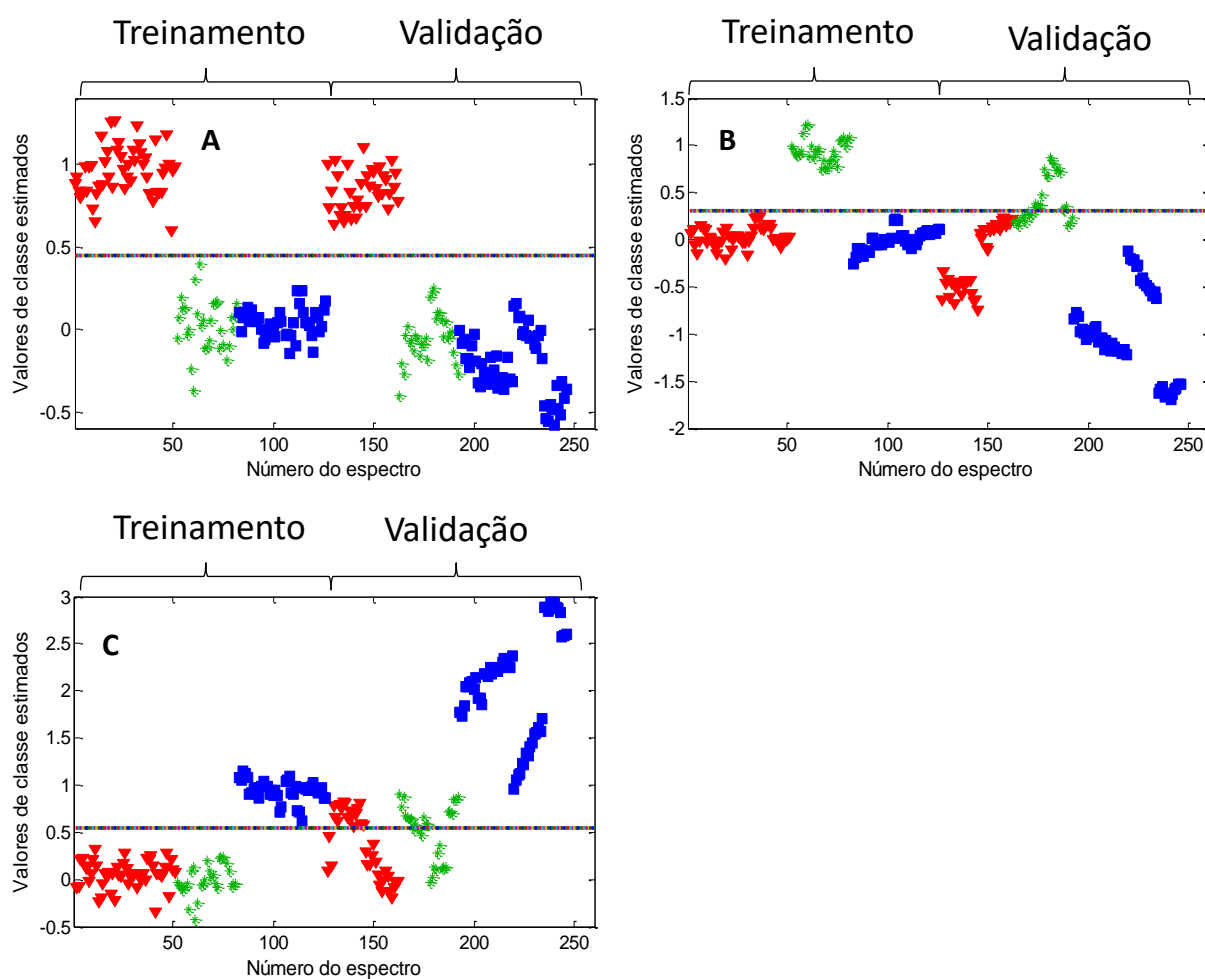


Figura 33. Valores de classe estimados para os conjuntos de treinamento e validação de (A) cedrinho (*Erismia uncinatum*), (B) cedro (*Cedrela odorata*) e (C) mogno (*Swietenia macrophylla*). Legenda: cedrinho (*Erismia uncinatum*) (\blacktriangledown), cedro (*Cedrela odorata*) (*), mogno (*Swietenia macrophylla*) (\blacksquare) e (—) limite de discriminação.

Na Figura 33 (A), o modelo classificou de maneira assertiva as amostras de cedrinho (*Erismia uncinatum*). Todas as amostras do conjunto de validação estão acima do limite de discriminação, o que indica que foram identificadas como pertencentes à espécie de cedrinho (*Erismia uncinatum*). Além disso, as amostras de cedrinho (*Erismia*

uncinatum) tanto do conjunto de calibração como de validação estão distribuídas em um único grupo e não existe dispersão significativa entre elas.

Na Figura 33 (B), temos uma maior distribuição das amostras do conjunto de validação. No conjunto de validação, pode-se observar que as amostras da espécie de cedrinho (*Erisma uncinatum*) estão organizadas em dois grupos diferentes, o primeiro grupo corresponde às amostras secas e o segundo, às amostras que foram molhadas. Para cedro (*Cedrela odorata*) e mogno (*Swietenia macrophylla*), as amostras estão divididas em três grupos, o primeiro grupo de amostra corresponde ao grupo de amostras secas enquanto os dois últimos grupos de cada espécie correspondem às amostras que foram molhadas e posteriormente secas na estufa. Esse comportamento prejudica o desempenho do modelo, uma vez que amostras pertencentes a uma mesma espécie deveriam estar distribuídas em um mesmo grupo.

Por fim, na Figura 33 (C), pode-se observar valores de classe estimados para as amostras do conjunto de validação de mogno (*Swietenia macrophylla*) maiores do que 1 e erro de classificação do modelo ao identificar amostras de outras espécies como pertencentes à espécie de mogno (*Swietenia macrophylla*).

Para o desenvolvimento dos três modelos, foram usados 126 espectros (43 amostras) para o conjunto de treinamento e 120 espectros (40 amostras) para o conjunto de validação e as Tabelas 10, 11 e 12 descrevem os parâmetros para cada um dos modelos:

Tabela 10. Parâmetros do modelo de discriminação para cedrinho (*Erisma uncinatum*).

| Parâmetros | |
|--|--------|
| Nº de VL | 8 |
| Limite de discriminação | 0,4514 |
| TFN | 0% |
| TFP | 0% |
| TEF | 100% |
| Nº de amostras identificadas como <i>outliers</i> | 17 |

Tabela 11. Parâmetros do modelo de discriminação para cedro (*Cedrela odorata*).

| Parâmetros | |
|--|--------|
| Nº de VL | 5 |
| Limite de discriminação | 0,3063 |
| TFN | 45,16% |
| TFP | 0% |
| TEF | 54,83% |
| Nº de amostras identificadas como <i>outliers</i> | 12 |

Tabela 12. Parâmetros do modelo de discriminação para mogno (*Swietenia macrophylla*).

| Parâmetros | |
|--|---------|
| Nº de VL | 5 |
| Limite de discriminação | 0,5501 |
| TFN | 0% |
| TFP | 51,51% |
| TEF | 48,48 % |
| Nº de amostras identificadas como <i>outliers</i> | 18 |

O modelo de cedrinho (*Erisma uncinatum*) apresentou taxa de acerto de 100% e TFP e TFN nulas, o que indica que o modelo não apresentou nenhum falso positivo e falso negativo. Apesar disso, 42,50% das amostras do conjunto de validação foram identificadas como *outliers*. O modelo de cedro (*Cedrela odorata*) teve uma taxa de acerto acima de 50%, e apresentou alta TFN (45,16%), ou seja, o modelo classificou as amostras de cedro (*Cedrela odorata*) como pertencentes às outras espécies. Por fim, o modelo de mogno (*Swietenia macrophylla*) teve taxa de acerto abaixo de 50% e alta TFP porque o modelo classificou amostras de outras espécies como pertencentes à espécie de mogno (*Swietenia macrophylla*).

Em todos os três modelos, muitos espectros de validação foram identificados como *outliers*, o que invalidou o uso dos modelos para essas amostras. Esse resultado

pode ser justificado de duas maneiras: ou o processo de secagem não foi eficiente para secar as amostras e os espectros das amostras se mostraram muito diferentes por causa da umidade ainda presente ou a seleção de amostras não garantiu a representatividade nos conjuntos de treinamento e validação, uma vez que foi feita de modo aleatório. Para tentar resolver esse problema, testou-se um algoritmo de seleção de amostras chamado DUPLEX para a escolha dos conjuntos de treinamento e validação. Em seguida, o modelo foi reconstruído e os gráficos de Hotelling T² e resíduos Q estão apresentados na Figura 34:

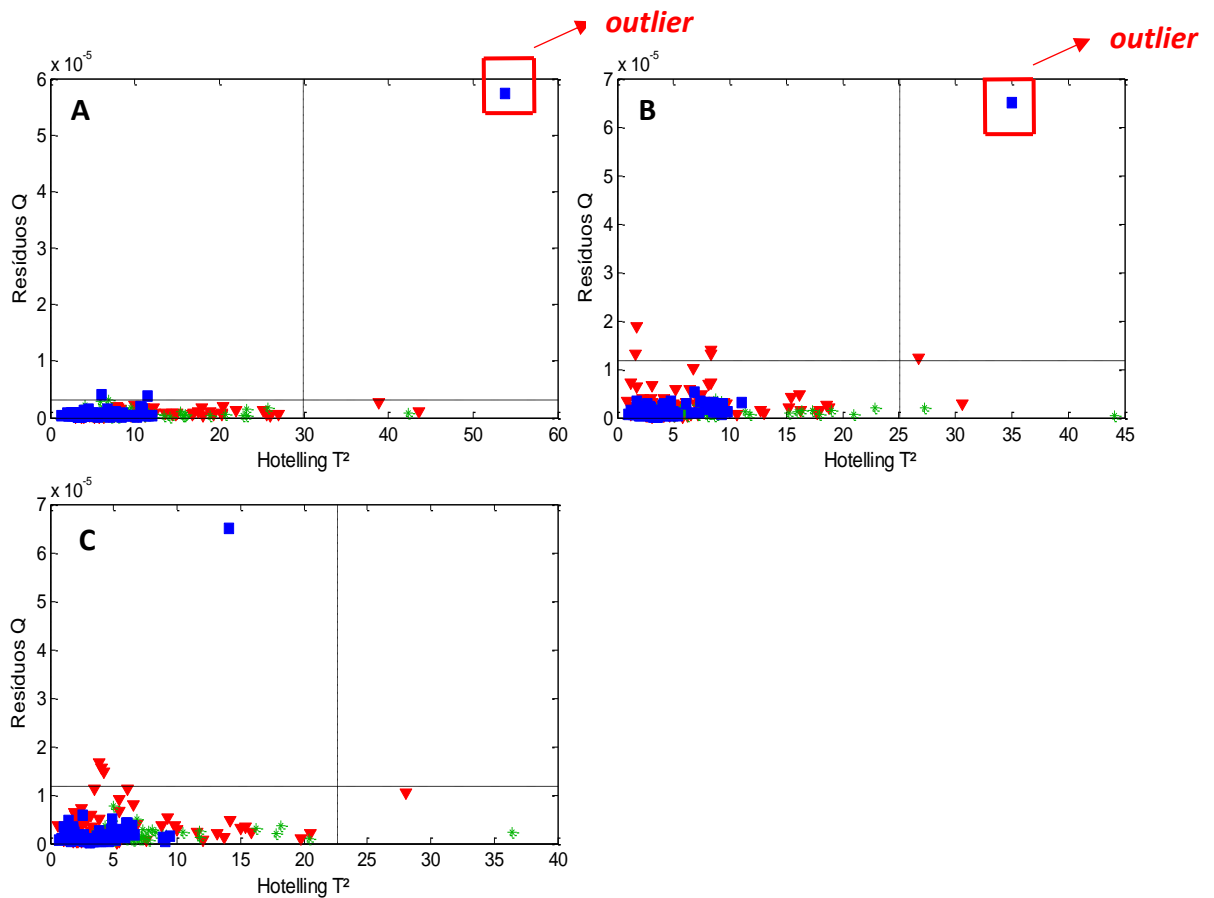


Figura 34. Gráfico dos valores de T² de Hotelling e resíduos Q de (A) cedrinho (*Erismia uncinatum*), (B) cedro (*Cedrela odorata*) e (C) mogno (*Swietenia macrophylla*) com seleção de amostras por DUPLEX. Legenda: cedrinho (*Erismia uncinatum*) (▼), cedro (*Cedrela odorata*) (*), mogno (*Swietenia macrophylla*) (■).

Os gráficos de Hotelling T² e resíduos Q na Figura 34 são bem diferentes dos gráficos apresentados anteriormente, isso porque agora os gráficos não apresentam um número grande de amostras identificadas como *outliers*. Na Figura 34 (A) e Figura 34 (B), existe apenas um espectro de uma amostra que não é compatível com as demais

usadas no conjunto de treinamento. Na Figura 34 (C), nenhum dos espectros apresentou valores fora dos limites de confiança. Isso indica que o modelo é adequado para classificar esse conjunto de validação, uma vez que ele não foi identificado como anômalo.

Em seguida, foram analisados os gráficos de valores de classe estimados para avaliar o desempenho do modelo na identificação das amostras. Os gráficos estão apresentados na Figura 35:

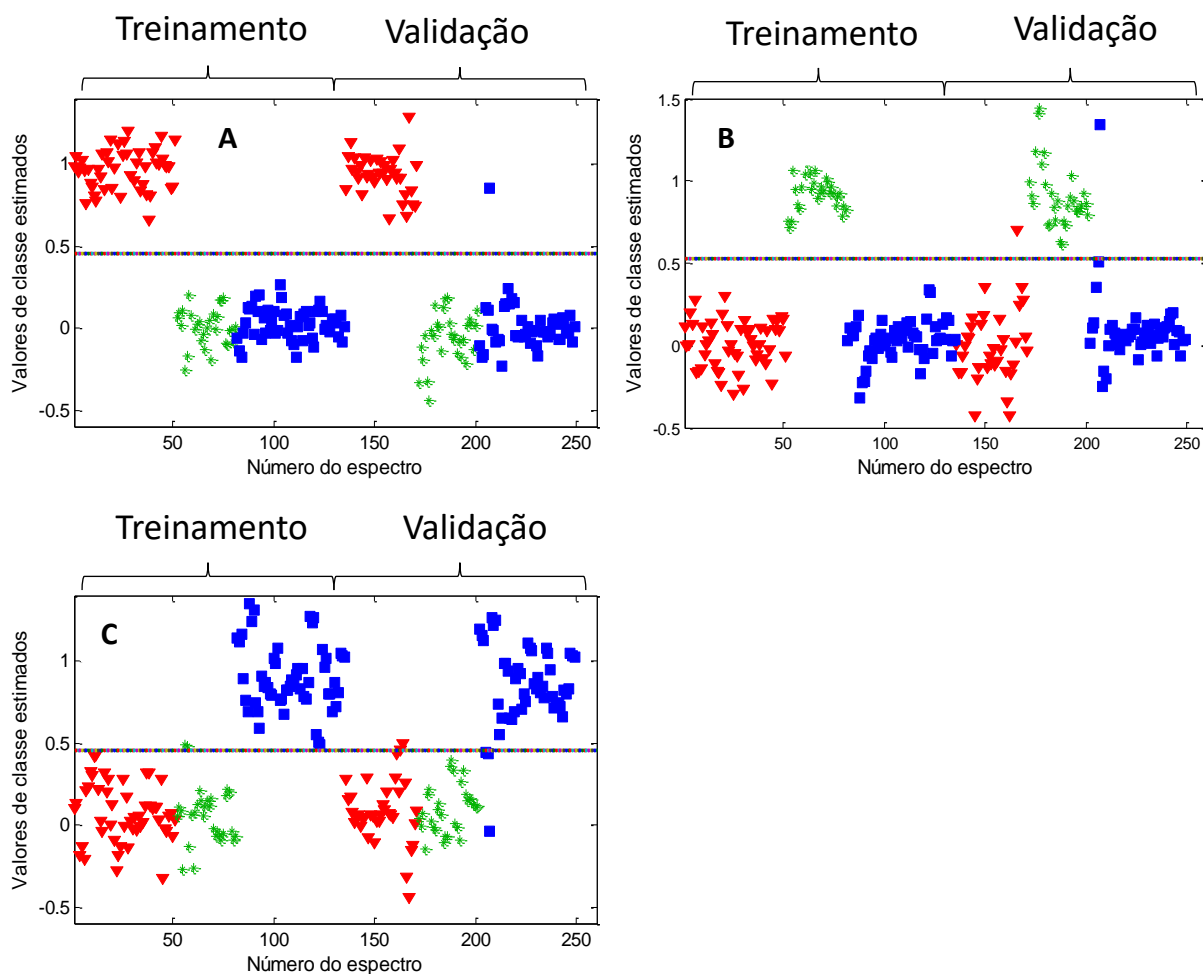


Figura 35. Valores de classe estimados para os conjuntos de treinamento e validação de (A) cedrinho (*Erisma uncinatum*), (B) cedro (*Cedrela odorata*) e (C) mogno (*Swietenia macrophylla*) com seleção de amostras por DUPLEX. Legenda: cedrinho (*Erisma uncinatum*) (▼), cedro (*Cedrela odorata*) (*), mogno (*Swietenia macrophylla*) (■) e (—) limite de discriminação.

Na Figura 35 (A), as amostras de cedrinho (*Erisma uncinatum*) do conjunto de validação foram classificadas de modo assertivo e apenas um dos espectros de mogno (*Swietenia macrophylla*) foi classificado como pertencente à classe de cedrinho (*Erisma*

uncinatum). Além disso, observa-se que as amostras de uma mesma espécie foram agrupadas em um único grupo. O modelo apresentou taxa de acerto de 100% e TFN e TFP nulas. Os parâmetros desse modelo estão descritos na Tabela 13:

Tabela 13. Parâmetros do modelo de discriminação para cedrinho (*Erisma uncinatum*) após seleção de amostras por DUPLEX.

| Parâmetros | |
|--------------------------------|--------|
| Nº de VL | 8 |
| Limite de discriminação | 0,4554 |
| TFN | 0% |
| TFP | 0% |
| TEF | 100% |

Na Figura 35 (B), o modelo de cedro (*Cedrela odorata*) teve também todas as amostras de cedro (*Cedrela odorata*) do conjunto de validação acima do valor do limite de discriminação, o que indica que o modelo as classificou de modo assertivo. O modelo apresentou poucos erros de classificação com alguns espectros de cedrinho (*Erisma uncinatum*) e mogno (*Swietenia macrophylla*) sendo indicados como pertencentes à espécie de cedro (*Cedrela odorata*), o que resultou numa TFP de 1,25%. O modelo apresentou bom desempenho com taxa de acerto acima de 98%, conforme descrito na Tabela 14:

Tabela 14. Parâmetros do modelo de discriminação para cedro (*Cedrela odorata*) após seleção de amostras por DUPLEX.

| Parâmetros | |
|--------------------------------|--------|
| Nº de VL | 6 |
| Limite de discriminação | 0,5280 |
| TFN | 0% |
| TFP | 1,25% |
| TEF | 98,75% |

Na Figura 35 (C), o modelo de mogno (*Swietenia macrophylla*) apresentou uma maior dispersão entre as amostras se comparado com os demais modelos, porém isso não

prejudicou o desempenho na identificação das amostras. A grande maioria das amostras de mogno (*Swietenia macrophylla*) foi classificada como pertencente à classe de mogno (*Swietenia macrophylla*) e o modelo apresentou baixas TFN e TFP. Por fim, o modelo apresentou taxa de acerto acima de 90%, o que indica um bom desempenho.

Tabela 15. Parâmetros do modelo de discriminação para mogno (*Swietenia macrophylla*) após seleção de amostras por DUPLEX.

| Parâmetros | |
|--------------------------------|--------|
| Nº de VL | 5 |
| Limite de discriminação | 0,4597 |
| TFN | 4,25% |
| TFP | 3,12% |
| TEF | 92,61% |

Com base nos resultados, pôde-se observar que depois da utilização do algoritmo DUPLEX para a seleção de amostras, houve uma diminuição no número de amostras identificadas como *outliers* e um aumento na taxa de acerto dos modelos de discriminação. Isso confirma a hipótese de que a seleção aleatória feita anteriormente não garantiu representatividade nos conjuntos de treinamento e validação e por isso, os modelos construídos anteriormente não apresentaram bom desempenho na identificação das amostras.

4.3 CONCLUSÕES PARCIAIS

A primeira etapa permitiu avaliar o comportamento e o tempo de secagem das amostras quando colocadas no desidratador de alimentos. Pôde-se observar que a medida que se aumentava o tempo de exposição no desidratador, as amostras apresentavam valores de umidade menores, o que era de se esperar. Por fim, ao analisar o perfil dos espectros ao longo do tempo, concluiu-se que o tempo de secagem de 20 minutos era suficiente para que as amostras fossem secas e tivessem seus espectros testados no modelo de discriminação.

No entanto, os espectros das amostras secas depois de 20 minutos no desidratador foram identificados como *outliers*, onde pode-se concluir que o modelo de discriminação não é compatível para essas amostras. Ainda assim, mais uma tentativa foi realizada deixando as amostras secando no desidratador por 30 minutos. Nesse caso, os espectros também apresentaram valores fora do limite de confiança de 99,9% para Hotelling T² e resíduos Q, sendo identificados como *outliers*.

Pode-se concluir que o procedimento de secagem no desidratador permitiu que as amostras fossem secas, uma vez que os valores de umidade medidos confirmaram isso, mas não possibilitou que as amostras tivessem seus espectros validados no modelo de discriminação. A identificação dos espectros registrados como *outliers* deixa claro que o conjunto de treinamento não é compatível para validar essas amostras.

Na segunda etapa, foi avaliado o desempenho de um novo modelo de discriminação construído com amostras secas e úmidas que foram submetidas a um mesmo procedimento experimental. Inicialmente, a seleção aleatória de amostras para o conjunto de treinamento e validação não funcionou, uma vez que os modelos identificaram as amostras úmidas como *outliers*. Além disso, os modelos de cedro (*Cedrela odorata*) e mogno (*Swietenia macrophylla*) apresentaram TEF abaixo de 50%, o que indica um mau desempenho.

Uma nova tentativa foi realizada aplicando o algoritmo DUPLEX para selecionar amostras. A aplicação pelo algoritmo DUPLEX permitiu um conjunto de treinamento e validação mais representativo, onde apenas um espectro de mogno (*Swietenia macrophylla*) foi identificado como *outlier*. E como consequência, os modelos de discriminação construídos contemplaram a representatividade dos conjuntos de amostras e apresentaram um bom desempenho, apresentando TEF acima de 90% para todas as espécies.

5 CONCLUSÕES FINAIS

A análise exploratória dos espectros NIR de madeira de mogno (*Swietenia macrophylla*) seco e recém-cortado confirmou a interferência da umidade nos espectros e a mudança do perfil do espectro nas regiões das bandas de absorção de água.

Ademais, a espectroscopia no infravermelho próximo (NIRS) combinada à análise discriminante por mínimos quadrados parciais (PLS-DA) mostrou-se uma metodologia eficiente para a discriminação de mogno (*Swietenia macrophylla*) em condição seca e com o equipamento portátil. O modelo de discriminação apresentou TEF acima de 90%.

No entanto, o modelo construído, não obteve bons resultados para um conjunto de validação com amostras de madeira recém-cortada, ou seja, fora da condição seca. Os resultados mostraram que o modelo identificou 90% das amostras como *outliers*.

A aplicação da ferramenta matemática de correção de sinal ortogonal (OSC), na tentativa de eliminar ou reduzir a interferência da umidade nos espectros, não gerou resultados satisfatórios. O modelo corrigido com OSC classificou 70% das amostras de madeira recém-cortada como *outliers*, amostras anômalas, e a TEF apresentou valor abaixo de 20%.

A princípio, a tentativa de secagem das amostras de forma experimental se mostrou uma boa estratégia uma vez que a medida que se aumentava o tempo de secagem das amostras no desidratador, os valores de umidade diminuía. No entanto, mais de 90% dos espectros dessas amostras, quando testados no modelo de discriminação construído apenas com amostras secas, foram identificados como *outliers*. Ou seja, o modelo não foi compatível para essas amostras.

Por fim, tentou-se o desenvolvimento de novo modelo de discriminação, onde amostras úmidas e secas, que seriam utilizadas na fase de treinamento e validação, foram submetidas ao mesmo procedimento antes da obtenção dos espectros. A seleção aleatória de amostras para os conjuntos de treinamento e validação não foi eficiente, uma vez que todas amostras úmidas de mogno (*Swietenia macrophylla*) foram identificadas como amostras anômalas.

Apesar disso, a aplicação do algoritmo DUPLEX para selecionar as amostras mostrou-se eficiente e garantiu a representatividade nos conjuntos de treinamento e validação. Os modelos de cedrinho (*Erismia uncinatum*), cedro (*Cedrela odorata*) e curupixá (*Micropholis melinoniana*) construídos nessa nova condição apresentaram TEF acima de 90% e TFN e TFP abaixo de 5%.

Isso nos leva a concluir que é necessário submeter as amostras de treinamento e amostras “reais” ao mesmo procedimento antes de serem testadas no modelo. Fazendo isso, garante-se a mesma condição para as amostras, independentemente de serem amostras úmidas ou secas. Logo, pode-se concluir que o método adotado, onde as amostras foram secas por 20 minutos a 80°C numa estufa com circulação de ar, atingiu o objetivo nesse sentido. Os resultados demonstraram que os modelos de discriminação puderam ser aplicados para amostras com diferentes valores de umidade que passaram pelo mesmo procedimento de secagem adotado, o que minimizou a interferência da umidade nos espectros NIR da madeira.

Como o objetivo a longo do prazo do projeto é aplicar os modelos de discriminação em postos de fiscalização para identificar espécies florestais transportadas e comercializadas, é preciso ainda incluir mais amostras no novo modelo construído para garantir um modelo robusto e representativo. Para o futuro, espera-se adquirir novas amostras para desenvolver os modelos para as espécies restantes: andiroba (*Carapa guianensis*) e curupixa e estabelecer esse procedimento como padrão para validação de amostras desconhecidas.

6 REFERÊNCIAS BIBLIOGRÁFICAS

- Amabis & Martho. *Biologia dos Organismos* volume 2. Editora Moderna, 2ª edição, 1996.
- Annual Book of ASTM Standards; Standards Practices for Infrared Multivariate Quantitative Analysis, E1655-05. ASTM International: West Conshohocken, 2012.
- Barker, M.; Rayens, W. J. *Chemom.* 2003, 17, 166.
- Bergo, M. C. J.; Pastore, T. C. M.; Coradin, V. T. R.; Wiedenhoeft, A. C.; Braga, J. W. B. *IAWA Journal*. **2016**, 37, 420.
- Borin, A.; Poppi, R. J. J. *Braz. Chem. Soc.* **2004**, 15, 570
- Botelho, B. G.; Reis, N.; Oliveira, L. S.; Sena, M. M.; *Food Chem.* **2015**, 181, 31.
- Braga, J. W. B.; Pastore, T. C. M.; Coradin, V. T. R.; Camargos, J. A. A.; da Silva, A. R. *IAWA Journal*. **2011**, 32, 285. 8.
- Browning, B. L. *The chemistry of wood*, Interscience Publishers, London, 1963.
- Coradin, V. T. R.; Camargos, J. A. A.; Marques, L. F.; Silva Jr., E. R.; *Madeiras similares ao mogno (Swietenia macrophylla King.): chave ilustrada para identificação anatômica em campo*, Serviço Florestal Brasileiro, Brasil, 2009.
- da Silva, V. A. G.; Talhavini, M.; Zacca, J. J.; Maldaner, A. O.; Peixoto, I. C. F.; Braga, J. W. B.; *Microchem, J. J. Braz. Chem. Soc.* **2014a**, 116, 235.
- da Silva, V. A. G.; Talhavini, M.; Zacca, J. J.; Trindade, B. R.; Braga, J. W. B.; *J. Braz. Chem. Soc.* **2014b**, 25, 1552.
- Donato, D. B.; Castro, R. V. O.; Carneiro, A. C. O.; Carvalho, A. M. M. L.; Vital, B. R.; Teixeira, R. U. *Sci. For.* **2015**, 43, 107, 703-712.
- Dormontt, E. E.; Boner, M.; Braun, B.; Breulmann, G.; Degen, B.; Espinoza, E.; Gardner, S.; Guillery, P.; Hermanson, J. C.; Koch, G.; Lee, S. L.; Kanashiro, M.; Rimbawanto, A.; Thomas, D.; Wiedenhoeft, A. C.; Yin, Y.; Zahnen, J.; Lowe, A. J. *Biol. Cons.* **2015**, 6, 38.
- Espinoza, E.; Wiemann, M.; Barajas-Morales, J.; Chavarria, G.D.; McClure, P.J. *IAWA J.* **2015**, 36 (3), 311–325.
- Ferreira, M M. C.; Antunes, A M.; Melgo, Marisa S.; Volpe, P L. O. *Quím. Nova.* **1999**, 22, 5, 724-731.
- Ferreira, M. M. C. *Quimiometria – Conceitos, Métodos e Aplicações*, 1ª ed., Editora Unicamp: Campinas, 2015

- Förstel, H.; Boner, M.; Höltken, A.M.; Fladung, M.; Degen, B.; Zahnen, J. Deutsche Bundesstiftung Umwelt, **2011**.
- Galvão, A.P.M.; Jankowsky, I.P. **Secagem racional da madeira**. São Paulo: Nobel, 1985. 111p.
- Glass, S.V.; Zelinka, A.L. Moisture relations and physical properties of wood. General Technical Report, 2010.
- Honorato, F. A.; Neto, B. B.; Martins, M. N.; Galvão, R. K. H.; Pimentel, M. F. Química Nova, **2007**, 30 (5).
- Igné, B.; Hossain, Md. N.; Drennen III, J. K.; Anderson, C. A. J. Near Infrared Spectrosc. **2014**, 22, 179-188.
- Kennard, R. W.; Stone, L. A.; Technometrics **1969**, 11.
- Kite, G.C.; Green, P.W.; Veitch, N.C.; Groves, M.C.; Gasson, P.E.; Simmonds, M.S. Phytochem. **2010**, 71, 1122–1131.
- Lamb, F. M. Targeting and achieving the final moisture content. In: Profitable Solutions for Quality Drying of Softwoods and Hardwoods. Charlotte, 1994.
- Lancaster, C. & Espinoza, E. Mass Spectrom. **2012**, 26, 1147–1156.
- Muellner, A.; Schaefer, H.; Lahaye, R. Mol. Ecol. Resour. **2011**, 11, 450–460.
- Nature Economy and People Connected. **Timber Testing Techniques**, NEPCOM, London, 2017.
- Neto, J.; Moita, G. Química Nova. **1997**, 21 (4).
- Paredes-Villaneuva, K.; Espinoza, E.; Ottenburghs, J.; Sterken, M. G.; Bongers, F.; Zuidema, P.A. Forestry: An International Journal of Forest Research. **2018**.
- Pasquini, C.; J. Braz. Chem. Soc. **2003**, 14, 198.
- Pastore, T. C. M.; Braga, J. W. B.; Coradin, V. T. R.; Magalhães, W. L. E.; Okino, E. Y. A.; Camargo, J. A. A.; Muñiz, G. I. B.; Bressan, O. A.; Davrieux, F. Holzforschung. **2011**, 65, 73. 7.
- Pettersen, R. C. J. Am. Chem. Soc. **1984**, 54-126.
- Richardson, B. A. Wood Preservation 2nd ed. London, 1993.
- Schwanninger, M.; Rodrigues, J.C.; Fackler, K. J. Near Infrared Spectrosc. **2011**, 19, 287-308.
- Sjöblom, J.; Svensson, O.; Josefson, M.; Wold, Svante. Chemometrics and Intelligent Laboratory Systems, **1998**, 44(1): 229-244.
- Soares, L. F.; Silva, D. C. da.; Bergo, M. C. J. Coradin, V. T. R.; Braga, J. W. B.; Pastore, T. C. M. Quím. Nova. **2017**, 40, 4.

- Souza, M. H.; Megliano, M. M.; Camargos, J. A. A.; Sousa, M. R. *Madeiras tropicais brasileiras*, 1ed., v. 1.; Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis, Laboratório de Produtos Florestais: Brasília, 1997.
- Tsuchikawa, S. *Appl. Spectrosc. Rev.* **2007**, 42, 43.
- Tsuchikawa, S.; Tsutsumi, S. *J. Mater. Sci. Lett.* **1998**, 17(8), 661.
- Tsuchikawa, S.; Tsutsumi, S. *J. Mater. Sci. Lett.* **1998**, 17(8), 661.
- United Nations Office on Drugs and Crime; Global Programme for Combating Wildlife and Forest Crime; Best Practice Guide for Forensic Timber Identification, UNODC, New York , 2016.
- Westad, F.; Marini, F. *Analytica Chimica Acta*, **2015**, 14-24.
- Westerhuis, J. A.; Jong, S. d.; Smilde, A. K. *J. Chem. Intell. Lab. Syst.* **2001**, 56, 13-25.
- Wold, S.; Antti, H.; Lindgren, F.; Öhman, J. *J. Chem. Intell. Lab. Syst.* **1998**, 44, 175-185.
- Yu, M.; Jiao, L.; Guo, J.; Yin, Yafang. *Planta*, **2017**, 246, 3.

7 ANEXO

O limite ou limiar de discriminação é determinado pelo valor de classe onde as probabilidades de uma amostra ser atribuída à classe A e B são iguais: $P(\hat{y}_i|A) = P(\hat{y}_i|B)$. O que é equivalente à equação a seguir:

$$p(Y|A).P(A) = p(Y|B).P(B) \quad (26)$$

Onde:

$$p(Y|A) = \frac{1}{s_A\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\bar{y}_A}{s_A}\right)^2} \quad (27)$$

$$p(Y|B) = \frac{1}{s_B\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\bar{y}_B}{s_B}\right)^2} \quad (28)$$

$$P(A) = \frac{I_A}{I_A+I_B} \quad (29)$$

$$P(B) = \frac{I_B}{I_A+I_B} \quad (30)$$

Onde

- $p(\hat{y}_i|A)$ e $p(\hat{y}_i|B)$ são as funções de densidade de probabilidade para as classes A e B, respectivamente;
- s_A e s_B os desvios padrões das estimativas obtidas no conjunto treinamento para as classes A e B, respectivamente;
- \bar{y}_A e \bar{y}_B as médias das estimativas de y obtidas no conjunto treinamento para as classes A e B, respectivamente;
- $P(A)$ e $P(B)$ são as probabilidades *a priori* obtidas no conjunto treinamento para as classes A e B, respectivamente;
- I_A e I_B o número de amostras do conjunto de treinamento para as classes A e B, respectivamente;

Para obter a equação que permite determinar o limite de discriminação, iniciamos substituindo as equações 22 e 23 em 21, obtendo:

$$\frac{1}{s_A\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\bar{y}_A}{s_A}\right)^2} . P(A) = \frac{1}{s_B\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\bar{y}_B}{s_B}\right)^2} . P(B) \quad (31)$$

Multiplicando os dois lados da equação por $\sqrt{2\pi}$, a expressão pode ser simplificada a:

$$\frac{1}{s_A} e^{-\frac{1}{2}\left(\frac{y-\bar{y}_A}{s_A}\right)^2} . P(A) = \frac{1}{s_B} e^{-\frac{1}{2}\left(\frac{y-\bar{y}_B}{s_B}\right)^2} . P(B) \quad (32)$$

Aplicando a propriedade $a^{-n} = \frac{1}{a^n}$, se obtém:

$$\frac{1}{S_A} \cdot \frac{1}{e^{\frac{1}{2}\left(\frac{y-\bar{y}_A}{S_A}\right)^2}} \cdot P(A) = \frac{1}{S_B} \cdot \frac{1}{e^{\frac{1}{2}\left(\frac{y-\bar{y}_B}{S_B}\right)^2}} \cdot P(B) \quad (33)$$

$$\frac{P(A)}{S_A} \cdot \frac{1}{e^{\frac{1}{2}\left(\frac{y-\bar{y}_A}{S_A}\right)^2}} = \frac{P(B)}{S_B} \cdot \frac{1}{e^{\frac{1}{2}\left(\frac{y-\bar{y}_B}{S_B}\right)^2}} \quad (34)$$

Aplicando a função de logaritmo natural (\ln) dos dois lados, temos que:

$$\ln \left[\frac{P(A)}{S_A} \right] + \ln(1) - \ln \left[e^{\frac{1}{2}\left(\frac{y-\bar{y}_A}{S_A}\right)^2} \right] = \ln \left[\frac{P(B)}{S_B} \right] + \ln(1) - \ln \left[e^{\frac{1}{2}\left(\frac{y-\bar{y}_B}{S_B}\right)^2} \right] \quad (35)$$

Sabendo que $\ln(1) = 0$ e $\ln(e^x) = x$ e rearranjando a equação temos que:

$$\ln \left[\frac{P(A)}{S_A} \right] - \frac{1}{2} \left(\frac{y-\bar{y}_A}{S_A} \right)^2 = \ln \left[\frac{P(B)}{S_B} \right] - \frac{1}{2} \left(\frac{y-\bar{y}_B}{S_B} \right)^2 \quad (36)$$

$$\frac{-1}{2} \left(\frac{y-\bar{y}_A}{S_A} \right)^2 + \frac{1}{2} \left(\frac{y-\bar{y}_B}{S_B} \right)^2 + \ln \left[\frac{P(A)}{S_A} \right] - \ln \left[\frac{P(B)}{S_B} \right] = 0 \quad (37)$$

$$\frac{-1}{2} \left(\frac{y-\bar{y}_A}{S_A} \right)^2 + \frac{1}{2} \left(\frac{y-\bar{y}_B}{S_B} \right)^2 + \ln \left[\frac{P(A)}{P(B)} \cdot \frac{S_B}{S_A} \right] = 0 \quad (38)$$

Multiplicando por (-2) e resolvendo os parênteses, temos que:

$$\left(\frac{y-\bar{y}_A}{S_A} \right)^2 - \left(\frac{y-\bar{y}_B}{S_B} \right)^2 - 2 \ln \left[\frac{P(A)}{P(B)} \cdot \frac{S_B}{S_A} \right] = 0 \quad (39)$$

$$\frac{(y^2 - 2y\bar{y}_A + \bar{y}_A^2)}{S_A^2} - \frac{(y^2 - 2y\bar{y}_B + \bar{y}_B^2)}{S_B^2} - 2 \ln \left[\frac{P(A)}{P(B)} \cdot \frac{S_B}{S_A} \right] = 0 \quad (40)$$

Fazendo o mínimo múltiplo comum, temos que:

$$S_B^2(y^2 - 2y\bar{y}_A + \bar{y}_A^2) - S_A^2(y^2 - 2y\bar{y}_B + \bar{y}_B^2) - 2S_A^2S_B^2 \ln \left[\frac{P(A)}{P(B)} \cdot \frac{S_B}{S_A} \right] = 0 \quad (41)$$

$$S_B^2y^2 - 2S_B^2y\bar{y}_A + S_B^2\bar{y}_A^2 - S_A^2y^2 + 2S_A^2y\bar{y}_B - S_A^2\bar{y}_B^2 - 2S_A^2S_B^2 \ln \left[\frac{P(A)}{P(B)} \cdot \frac{S_B}{S_A} \right] = 0 \quad (42)$$

Finalmente, reorganizando a equação, obtemos a seguinte solução:

$$(S_B^2 - S_A^2)y^2 + (2S_A^2\bar{y}_B - 2S_B^2\bar{y}_A)y + S_B^2\bar{y}_A^2 - S_A^2\bar{y}_B^2 - 2S_A^2S_B^2 \ln \left[\frac{P(A)}{P(B)} \cdot \frac{S_B}{S_A} \right] = 0 \quad (43)$$

Caso as probabilidades de ocorrência das classes A ou B forem iguais ($P(A)=P(B)$), a solução é simplificada a:

$$(S_B^2 - S_A^2)y^2 + (2S_A^2\bar{y}_B - 2S_B^2\bar{y}_A)y + S_B^2\bar{y}_A^2 - S_A^2\bar{y}_B^2 - 2S_A^2S_B^2 \ln \left[\frac{S_B}{S_A} \right] = 0 \quad (44)$$

Pode-se observar que as equações 38 e 39 são equações de segundo grau e que a única variável é y . Substituindo os valores experimentais e resolvendo a equação para y , o limiar ou limite de discriminação é determinado por uma das raízes da equação.