



Universidade de Brasília
Instituto de Química
Programa de Pós-Graduação em Química

DISSERTAÇÃO DE MESTRADO

Transferência de calibração na discriminação de mogno e espécies semelhantes utilizando NIRS e PLS-DA

Maria Cecília Jorge Bergo

Orientador

Prof. Dr. Jez Willian Batista Braga

Coorientadora

Dra. Tereza Cristina Monteiro Pastore

Brasília, 2014.



Universidade de Brasília
Instituto de Química
Programa de Pós-Graduação em Química

DISSERTAÇÃO DE MESTRADO

Transferência de calibração na discriminação de mogno e espécies semelhantes utilizando NIRS e PLS-DA

Maria Cecília Jorge Bergo

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

Orientador

Prof. Dr. Jez Willian Batista Braga

Coorientadora

Dra. Tereza Cristina Monteiro Pastore

Brasília, 2014.

Ficha catalográfica elaborada pela Biblioteca Central da Universidade de Brasília. Acervo 1017146.

B499t Bergo, Maria Cecília Jorge.
Transferência de calibração na discriminação de mogno e espécies semelhantes utilizando NIRS e PLS-DA / Maria Cecília Jorge Bergo. -- 2014.
xvii, 133 f. : il. ; 30 cm.

Dissertação (mestrado) - Universidade de Brasília, Instituto de Química, Programa de Pós-Graduação em Química, 2014.

Inclui bibliografia.

Orientação: Jez Willian Batista Braga ; Coorientação: Tereza Cristina Monteiro Pastore.

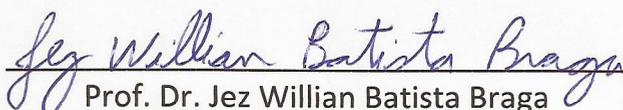
1. Espectroscopia de infravermelho. 2. Mogno. 3. Madeira - Identificação. I. Braga, Jez Willian Batista. II. Pastore, Tereza Cristina Monteiro. III. Título.

CDU 543.422.4

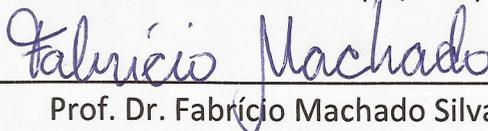


COMUNICADO

Comunicamos a aprovação da Defesa de Dissertação de Mestrado do (a) aluno (a) **Maria Cecília Jorge Bergo**, matrícula nº **12/0081521**, intitulada **“Transferência de Calibração na Discriminação de Mogno e Espécies Semelhantes Utilizando NIRS e PLS-DA”**, apresentada no (a) Auditório Azul do Instituto de Química (IQ) da Universidade de Brasília (UnB) em 18 de julho de 2014.



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



Prof. Dr. Fabrício Machado Silva
Membro Titular (IQ/UnB)



Dr. Dário Grattapaglia
Membro Titular (EMBRAPA / CENARGEN)

Prof. Dr. Fernando Fabríz Sodré
Membro Suplente (IQ/UnB)

Em 18 de julho de 2014.

DEDICATÓRIA

Dedico este trabalho aos meus queridos avós.

AGRADECIMENTOS

A Deus, pelo dom da vida e à Sua Mãe pelos cuidados.

Ao meu orientador Prof. Dr. Jez Willian Batista Braga pela oportunidade de realizar este trabalho, pelos ensinamentos, acompanhamento, confiança, incentivo, amizade e paciência, o meu sincero agradecimento e admiração.

À minha coorientadora Dra. Tereza Cristina Monteiro Pastore pelos ensinamentos, atenção, apoio, incentivo, motivação e orientação.

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

Ao Laboratório de Produtos Florestais (LPF) do Serviço Florestal Brasileiro (SFB), pela disponibilização das amostras de madeira em blocos e de um dos espectrômetros de bancada utilizados neste trabalho.

À Dra. Vera Teresinha Rauber Coradin do LPF do SFB pelo apoio, contribuições e identificação de amostras.

Ao Dr. Fabrice Davrieux do Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD) na França pela cessão do espectrômetro portátil para a realização das medidas.

Aos Prof. Dr. Alexandre Florian da Costa e Prof. Dr. Joaquim Carlos Gonzalez do Departamento de Engenharia Florestal da Universidade de Brasília pela disponibilização de um dos espectrômetros de bancada utilizados.

Ao Dr. Alex Wiedenholf do Forest Products Laboratory, Madison (USA) pela doação de amostras de mogno de diferentes países.

À Ma. Natasha Neiva Moura pela cessão dos espectros de mogno moído.

A Dr. José Natalino M. Silva, Dr. Paulo J. P. Fontes e Sr. Waldemar V. Lopes pelo apoio essencial para a obtenção das madeiras.

À Prof.^a Dra. Maria Fernanda Pimentel da Universidade Federal de Pernambuco (UFPE) pela disponibilização de material sobre transferência de calibração.

Às empresas Serraria Marajoara Ind. Com. e Exp., Selectas S.A. Ind. e Com.de Madeira Ltda, Madeireira Caingá Ltda e Empresa Juruá Florestal pela doação de espécimes de madeira.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa concedida.

Ao Instituto Nacional de Ciência e Tecnologia em Bioanalítica (INCTBio), pelo apoio financeiro.

À Fundação de Amparo a Pesquisa do Distrito Federal (FAPDF), através do projeto NEXTREE - Núcleo de Excelência em Genômica Florestal Aplicada.(processo FAPDF:193.000.570/2009).

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), através do projeto Utilização de espectroscopia no infravermelho próximo e quimiometria para a análise e discriminação de madeiras em campo (processos CNPq:473936/2013-5 e 307223/2012-4).

Aos meus pais Márcia e César pelo amor, compreensão, incentivo, dedicação e exemplo.

Ao meu namorado Igor pelo amor, companheirismo, paciência, compreensão, apoio e motivação.

Aos meus irmãos Felipe e Míriam e primos pelo apoio e amizade.

À minha querida amiga Caro pela ajuda, conselhos e orações.

Aos AQQUAmiguinhos pela convivência, amizade e descontração.

Aos meus avós e toda a minha família e amigos pelo carinho, motivação, incentivo, companhia e momentos inesquecíveis.

Aos professores, colegas e funcionários do Instituto de Química da UnB e a todos que colaboraram direta ou indiretamente ao longo deste trabalho, o meu muito obrigada.

RESUMO

O mogno (*Swietenia macrophylla* King) é uma das madeiras mais valiosas do mundo e, em 2003, foi adicionado ao Apêndice II da Convenção sobre o Comércio Internacional das Espécies de Flora e da Fauna Selvagens em Perigo de Extinção (CITES) como espécie ameaçada devido à sua ampla exploração. Existem de diversas espécies semelhantes ao mogno e sua identificação é crucial para fiscalização da exploração dessa espécie. O método geralmente empregado com essa finalidade é o anatômico visual, exigindo que a madeira seja examinada por especialistas, que são insuficientes para atender à demanda. A espectroscopia no infravermelho próximo (NIRS) é uma excelente alternativa para a identificação de madeira, pois possibilita análises rápidas e não destrutivas. Nesta dissertação, discriminaram-se as espécies semelhantes andiroba, cedro, curupixá e mogno utilizando equipamentos NIRS portátil e de bancada e análise discriminante por mínimos quadrados parciais (PLS-DA), que forneceram modelos com 100% de classificações corretas. Foram utilizadas diferentes estratégias de transferência de calibração para tornar possível o compartilhamento de dados entre os equipamentos de bancada com 100% de classificações corretas. A transferência entre equipamentos de bancada e portátil foi feita por métodos de reversos de padronização apresentando de 93% a 100% de classificações corretas. Também foi estudada a transferência para compatibilizar dados obtidos em amostras de mogno moídas e em blocos. Nesse caso, utilizando pré-processamento e seleção de intervalo espectral obteve-se 100% de classificações corretas. Utilizando esses mesmos modelos foi possível identificar por espécie 465 amostras de mogno moídas de 26 diferentes países com 97% de classificações corretas. Por último, entre as amostras moídas de mogno de Brasil, Honduras, México, Peru e Venezuela, foi possível discriminar quanto à procedência as oriundas dos três primeiros. Já as amostras do Peru e da Venezuela não tiveram boa separação entre si, embora tenham apresentado boa separação com relação aos demais países.

ABSTRACT

Mahogany (*Swietenia macrophylla* King) is the world's most valuable tropical timber and was included on Convention of International Trade in Endangered Species of Wild Fauna and Flora (CITES) Appendix II as an endangered species in 2003 due to its extensive logging. Some timbers are very similar to mahogany, so strategies for identification are crucial to supervise its exploitation. The method usually applied with these purposes is visual anatomical identification, requiring the wood to be examined by specialists, which are scarce. Near Infrared Spectroscopy (NIRS) is an excellent alternative for wood identification since it allows rapid and non-destructive analysis. In this work, the similar species andiroba, cedar, curupixá and mahogany were discriminated with 100% of correct classifications using handheld portable and high performance NIRS instruments and partial least squares discriminant analysis (PLS-DA). Some calibration transfer strategies were tested to allow data sharing between two high performance instruments with 100% of correct classifications. Reverse methods of standardization were applied to calibration transfer between high performance and handheld portable instruments with 93 to 100% of correct classifications. Calibration transfer was also used to make measurements of milled and block wood samples compatible. In this case, preprocessing and spectral range selecting resulted in models with 100% of correct classifications. Using these same models, 465 mahogany milled samples from 26 different countries were identified by species with 97% of correct classifications. Finally, among mahogany milled samples from Brazil, Honduras, Mexico, Peru and Venezuela, it was possible to discriminate those from the first three countries. Although it wasn't possible to separate Peru's samples from most of Venezuela's, the samples from both countries could be perfectly discriminated from the other three countries samples.

ÍNDICE

| | |
|--|------|
| LISTA DE ABREVIATURAS E ACRÔNIMOS | xi |
| LISTA DE FIGURAS | xiii |
| LISTA DE TABELAS | xvii |
| 1. INTRODUÇÃO E OBJETIVOS..... | 2 |
| 2. REVISÃO BIBLIOGRÁFICA..... | 6 |
| 3. FUNDAMENTAÇÃO TEÓRICA..... | 16 |
| 3.1 ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO | 16 |
| 3.1.1 Medidas em NIRS..... | 21 |
| 3.1.2 Instrumentação em NIRS..... | 23 |
| 3.1.2.1 Espectrofotômetros dispersivos..... | 23 |
| 3.1.2.2 Espectrofotômetros com transformada de Fourier | 24 |
| 3.1.2.3 Espectrofotômetros portáteis | 25 |
| 3.1.3 NIRS em análise de madeira | 27 |
| 3.2 QUIMIOMETRIA..... | 28 |
| 3.2.1 Calibração Multivariada..... | 28 |
| 3.2.1.1 PCA..... | 30 |
| 3.2.1.2 PLSR..... | 32 |
| 3.2.1.2.1 NIPALS | 33 |
| 3.2.1.2.2 SIMPLS..... | 34 |
| 3.2.1.2.3 PLS-DA..... | 37 |
| 3.2.1.2.4 Validação e validação cruzada..... | 39 |
| 3.2.1.2.5 Detecção de amostras anômalas | 41 |

| | |
|---|----|
| 3.2.1.2.6 Pré-processamentos | 43 |
| 3.2.2 TRANSFERÊNCIA DE CALIBRAÇÃO | 44 |
| 3.2.2.1 PRÉ-PROCESSAMENTOS PARA TRANSFERÊNCIA DE CALIBRAÇÃO | 46 |
| 3.2.2.2 PADRONIZAÇÃO DO EQUIPAMENTO..... | 47 |
| 3.2.2.3 MODELOS DE CALIBRAÇÃO ROBUSTOS | 47 |
| 3.2.2.3.1 Modelos globais | 48 |
| 3.2.2.3.2 Seleção de variáveis | 49 |
| 3.2.2.3.3 Atualização do modelo | 49 |
| 3.2.2.4 PADRONIZAÇÃO DOS ESPECTROS | 50 |
| 3.2.2.4.1 Padronização univariada..... | 50 |
| 3.2.2.4.2 Padronização direta | 51 |
| 3.2.2.4.3 Padronização direta por partes | 53 |
| 3.2.2.4.4 Diferenças espectrais por comprimento de onda | 54 |
| 3.2.2.5 PADRONIZAÇÃO DO MODELO DE CALIBRAÇÃO..... | 55 |
| 3.2.2.5.1 Padronização dos coeficientes de regressão | 55 |
| 3.2.2.5.2 Padronização dos valores previstos | 57 |
| 3.2.2.5.3. Padronizações reversas dos espectros | 58 |
| | |
| 4. DISCRIMINAÇÃO DE MOGNO E ESPÉCIES SEMELHANTES COM TRÊS EQUIPAMENTOS DE NIRS..... | 61 |
| | |
| 4.1 EXPERIMENTAL..... | 62 |
| 4.1.1 Amostras..... | 62 |
| 4.1.2 Espectros..... | 63 |
| 4.2 ANÁLISE DE DADOS..... | 64 |
| 4.3 RESULTADOS E DISCUSSÃO | 66 |
| 4.4 CONCLUSÕES DO CAPÍTULO | 76 |
| | |
| 5. TRANSFERÊNCIA DE CALIBRAÇÃO ENTRE EQUIPAMENTOS..... | 78 |
| | |
| 5.1 EXPERIMENTAL..... | 79 |
| 5.2. ANÁLISE DE DADOS..... | 79 |

| | |
|--|-----|
| 5.2.1. Transferência direta entre equipamentos de bancada..... | 79 |
| 5.2.2. Transferência reversa de equipamentos de bancada para equipamento portátil..... | 82 |
| 5.3 RESULTADOS E DISCUSSÃO..... | 84 |
| 5.3.1. Transferência direta entre equipamentos de bancada..... | 84 |
| 5.3.2. Transferência reversa entre equipamentos de bancada e portátil | 91 |
| 5.3.2.1. Transferência entre Antaris II e Phazir e recalibração com subconjunto | 91 |
| 5.3.2.2. Transferência entre Tensor 37 e Phazir..... | 99 |
| 5.4 CONCLUSÕES DO CAPÍTULO | 106 |
| 6. TRANSFERÊNCIA DE CALIBRAÇÃO ENTRE AMOSTRAS EM PÓ E EM BLOCOS | 109 |
| 6.1 EXPERIMENTAL..... | 109 |
| 6.2 ANÁLISE DE DADOS..... | 110 |
| 6.3 RESULTADOS E DISCUSSÃO..... | 111 |
| 6.4 CONCLUSÕES DO CAPÍTULO | 117 |
| 7. DISCRIMINAÇÃO DE MOGNO POR PAÍS DE PROCEDÊNCIA | 119 |
| 7.1 EXPERIMENTAL..... | 119 |
| 7.2 ANÁLISE DE DADOS..... | 120 |
| 7.3 RESULTADOS E DISCUSSÃO | 121 |
| 7.4 CONCLUSÕES DO CAPÍTULO | 124 |
| 8. CONCLUSÕES GERAIS | 126 |
| 9. REFERÊNCIAS BIBLIOGRÁFICAS..... | 128 |

LISTA DE ABREVIATURAS E ACRÔNIMOS

| | |
|--------|--|
| CITES | Convenção sobre o Comércio Internacional das Espécies da Flora e da Fauna Selvagens em Perigo de Extinção (<i>Convention on International Trade in Endangered Species of Wild Fauna and Flora</i>) |
| CV | Validação Cruzada (<i>Cross-Validation</i>) |
| DA | Análise discriminante (<i>Discriminant Analysis</i>) |
| DS | Padronização direta (<i>Direct Standardization</i>) |
| DTS | Espectroscopia com transformada digital (<i>Digital Transform Spectroscopy</i>) |
| FIR | Filtros de resposta finita a impulso (<i>Finite Impulse Response</i>) |
| FT-IR | Infravermelho com transformada de Fourier (<i>Fourier Transform Infrared</i>) |
| FT-NIR | Infravermelho próximo com transformada de Fourier (<i>Fourier Transform Near Infrared</i>) |
| LIBS | Espectroscopia de Emissão em Plasma Induzido por Laser (<i>Laser-induced Breakdown Spectroscopy</i>) |
| MEMS | Sistemas Microeletromecânicos (<i>Micro-electro-mechanical Systems</i>) |
| MS | Espectrometria de Massas (<i>Mass Spectrometry</i>) |
| MSC | Correção de espalhamento multiplicativo de sinal (<i>Multiplicative Signal Correction</i>) |
| NIPALS | Mínimos Quadrados Parciais iterativo não linear (<i>Nonlinear Iterative Partial Least Squares</i>) |
| NIRS | Espectroscopia no Infravermelho Próximo (<i>Near Infrared Spectroscopy</i>) |
| OMA | Analisador Ótico Multicanal (<i>Optical Multichannel Analyser</i>) |
| OSC | Correção de Sinal Ortogonal (<i>Orthogonal Signal Correction</i>) |
| PCA | Análise de Componentes Principais (<i>Principal Component Analysis</i>) |
| PDS | Padronização Direta por Partes (<i>Piecewise Direct Standardization</i>) |
| PLS | Mínimos Quadrados Parciais (<i>Partial Least Squares</i>) |
| PLS-DA | Análise Discriminante por Mínimos Quadrados Parciais (<i>Partial Least Squares Discriminant Analysis</i>) |

| | |
|----------------------|---|
| PLSR | Regressão por Mínimos Quadrados Parciais (<i>Partial Least Squares Regression</i>) |
| PRS | Padronização Reversa por Partes (<i>Piecewise Reverse Standardization</i>) |
| RMSEC | Raiz quadrada do erro médio quadrático de calibração (<i>Root Mean Squared Error of Calibration</i>) |
| RMSECV | Raiz quadrada do erro médio quadrático de validação cruzada (<i>Root Mean Squared Error of Cross-Validation</i>) |
| RMSEP | Raiz quadrada do erro médio quadrático de previsão (<i>Root Mean Squared Error of Prediction</i>) |
| RMSEV | Raiz quadrada do erro médio quadrático de validação (<i>Root Mean Squared Error of Validation</i>) |
| RS | Padronização Reversa (<i>Reverse Standardization</i>) |
| RSDW | SDW reversa (<i>Reverse SDW</i>) |
| SBC | Correção de inclinação e tendência (<i>Slope and Bias Correction</i>) |
| SDW | Diferença Espectral por Comprimento de Onda (<i>Spectral Difference by wavelengths</i>) |
| SIMCA | Modelagem Independente e Flexível por Analogia de Classes (<i>Soft Independent Modeling of Class Analogy</i>) |
| SIMPLS | Modificação Estatisticamente Inspirada do PLS (<i>Statistically Inspired Modification of PLS</i>) |
| SNV | Variação de Padrão Normal (<i>Standard Normal Variation</i>) |
| SVD | Decomposição de Valores Singulares (Singular Value Decomposition) |
| V-EASI | Auto-bombeamento via Efeito Venturi e Spray-sônico para Ionização Ambiente (<i>Venturi Easy Ambient Sonic-spray Ionization</i>) |
| V _L -EASI | Auto-bombeamento via Efeito Venturi e Spray-sônico para Ionização Ambiente em amostras líquidas |
| V _S -EASI | Auto-bombeamento via Efeito Venturi e Spray-sônico para Ionização Ambiente em amostras sólidas |
| VL | Variável Latente |

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1. <i>Swietenia Macrophylla</i> King. Foto por Paulo Ernani Ramalho Carvalho. ⁸ | 6 |
| Figura 2. Representação das regiões do tronco e orientação das faces seccionadas. . 7 | |
| Figura 3. Modos vibracionais em uma molécula, onde + e - indicam movimentos perpendiculares ao plano da página. ⁴⁶ | 17 |
| Figura 4. Diagrama de energia potencial de um oscilador harmônico (curva 1, (- - -)) e de um movimento anarmônico (curva 2, (—)). Figura extraída da referência 46..... | 18 |
| Figura 5. Medidas no infravermelho próximo por (a) transmitância; (b) transflectância; (c) reflectância difusa; (d) interatância; e (e) transmitância através de meio dispersivo. Onde <i>A</i> representa amostra, <i>I₀</i> a radiação da fonte e <i>I</i> a radiação após a passagem, reflexão e/ou interação com a amostra. Figura adaptada da referência 51. | 21 |
| Figura 6. Esquema de equipamento DTS portátil. Adaptado da referência 55. | 25 |
| Figura 7. Organização da matriz de dados X a partir de espectros de infravermelho. 30 | |
| Figura 8. Decomposição da matriz de dados X por PCA..... | 31 |
| Figura 9. Representação do PLS1-DA | 38 |
| Figura 10. Representação do PLS2-DA | 38 |
| Figura 11. Exemplo de gráfico de RMSECV por número de VL | 40 |
| Figura 12. Esquema das estratégias de transferência de calibração..... | 46 |
| Figura 13. Esquema da padronização de equipamentos. Adaptado da referência 74. | 47 |
| Figura 14. Esquema para modelos robustos. Adaptado da referência 74. | 48 |
| Figura 15. Esquema da atualização de modelo de calibração..... | 49 |
| Figura 16. Esquema da padronização de espectros. Adaptado da referência 74..... | 50 |
| Figura 17. Esquema de SBC. Adaptado da referência 74. | 58 |
| Figura 18. Esquema das padronizações reversas..... | 58 |
| Figura 19. Espectros dos equipamentos Tensor 37 (azul), Antaris II (vermelho) e Phazir (verde). | 66 |
| Figura 20. Espectros médios das amostras de andiroba (azul), cedro (vermelho), curupixá (verde) e mogno (roxo) medidas no equipamento Phazir. | 66 |
| Figura 21. Espectros dos equipamentos Tensor 37 (azul) e Antaris II (vermelho) pré-processados com derivada primeira e subtração de linha com a) região espectral completa e b) região de maior interesse ampliada. | 69 |
| Figura 22. Espectros do equipamento Phazir pré-processados com subtração de linha reta. | 69 |

- Figura 23.** Resultados do equipamento **Tensor 37**: Valores de classe e limites de discriminação (---) para a) andiroba (\diamond), b) cedro (\square), c) curupixá (Δ) e d) mogno (o). Símbolos vazios = calibração e preenchidos = previsão. Histogramas dos valores de classe para discriminação de e) andiroba (and), f) cedro (ced), g) curupixá (cur) e h) mogno (mog) das outras espécies (O.E) com gráficos de probabilidades (Prob) das amostras pertencerem à espécie em questão (—)ou às outras espécies (----). 72
- Figura 24.** Resultados do equipamento **Antaris II**: Valores de classe e limites de discriminação (---) para a) andiroba (\diamond), b) cedro (\square), c) curupixá (Δ) e d) mogno (o). Símbolos vazios = calibração e preenchidos = previsão. Histogramas dos valores de classe para discriminação de e) andiroba (and), f) cedro (ced), g) curupixá (cur) e h) mogno (mog) das outras espécies (O.E) com gráficos de probabilidades das amostras pertencerem à espécie em questão (—)ou às outras espécies (----). 73
- Figura 25.** Resultados do equipamento **Phazir**: Valores de classe e limites de discriminação (---) para a) andiroba (\diamond), b) cedro (\square), c) curupixá (Δ) e d) mogno (o). Símbolos vazios = calibração e preenchidos = previsão. Histogramas dos valores de classe para discriminação de e) andiroba (and), f) cedro (ced), g) curupixá (cur) e h) mogno (mog) das outras espécies (O.E) com gráficos de probabilidades das amostras pertencerem à espécie em questão (—)ou às outras espécies (—). 74
- Figura 26.** Fluxograma de etapas envolvidas na realização do método PDS e dos métodos de comparação sem transferência e de recalibração completa. 81
- Figura 27.** Fluxograma de etapas envolvidas na realização do método PDS e dos métodos de comparação sem transferência e de recalibração completa. 83
- Figura 28.** Derivada dos espectros dos equipamentos Tensor 37 e Antaris II e derivada dos espectros do Antaris II padronizada por PDS e regiões espectrais selecionadas para construção dos modelos (faixas sombreadas). 84
- Figura 29.** Valores de classe das amostras estimados na transferência entre os equipamentos **Tensor 37** e **Antaris II** pelos métodos recalibração (RC);sem transferência (ST); DS, PDS e SDW com 12 (SDW12) e 5 (SDW5) amostras de transferência (SDW5) para discriminação de **andiroba** (\diamond) das espécies cedro (\square), curupixá (Δ) e mogno (o) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão. 86
- Figura 30.** Valores de classe das amostras estimados na transferência entre os equipamentos **Tensor 37** e **Antaris II** pelos métodos recalibração (RC);sem transferência (ST); DS, PDS e SDW com 12 (SDW12) e 5 (SDW5) amostras de transferência (SDW5) para discriminação de **cedro** (\square) das espécies andiroba (\diamond), curupixá (Δ) e mogno (o) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão. 87
- Figura 31.** Valores de classe das amostras estimados na transferência entre os equipamentos **Tensor 37** e **Antaris II** pelos métodos recalibração (RC);sem transferência (ST); DS, PDS e SDW com 12 (SDW12) e 5 (SDW5) amostras de transferência (SDW5) para discriminação de **curupixá** (Δ) das espécies andiroba (\diamond), cedro (\square) e mogno (o) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão. 88

- Figura 32.** Valores de classe das amostras estimados na transferência entre os equipamentos **Tensor 37** e **Antaris II** pelos métodos recalibração (RC); sem transferência (ST); DS, PDS e SDW com 12 (SDW12) e 5 (SDW5) amostras de transferência (SDW5) para discriminação de **mogno** (o) das espécies andiroba (◊), cedro (□) e curupixá (Δ) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão. 89
- Figura 33.** Espectros dos equipamentos Antaris II e Phazir a) originais b) após interpolação e derivação e c) após interpolação, derivação e PRS. 92
- Figura 34.** Valores de classe das amostras estimados na transferência entre os equipamentos **Antaris II** e **Phazir** pelos métodos recalibração (RC); sem transferência (ST); RS, PRS, RSDW e recalibração com subconjunto (SubC) para discriminação de **andiroba** (◊) das espécies cedro (□), curupixá (Δ) e mogno (o) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão. 95
- Figura 35.** Valores de classe das amostras estimados na transferência entre os equipamentos **Antaris II** e **Phazir** pelos métodos recalibração (RC); sem transferência (ST); RS, PRS, RSDW e recalibração com subconjunto (SubC) para discriminação de **cedro** (□) das espécies andiroba (◊), curupixá (Δ) e mogno (o) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão. 96
- Figura 36.** Valores de classe das amostras estimados na transferência entre os equipamentos **Antaris II** e **Phazir** pelos métodos recalibração (RC); sem transferência (ST); RS, PRS, RSDW e recalibração com subconjunto (SubC) para discriminação de **curupixá** (Δ) das espécies andiroba (◊), cedro (□) e mogno (o) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão. 97
- Figura 37.** Valores de classe das amostras estimados na transferência entre os equipamentos **Antaris II** e **Phazir** pelos métodos recalibração (RC); sem transferência (ST); RS, PRS, RSDW e recalibração com subconjunto (SubC) para discriminação de **mogno** (o) das espécies andiroba (◊), cedro (□) e curupixá (Δ) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão. 98
- Figura 38.** Espectros dos equipamentos Tensor 37 e Phazir a) originais, b) após interpolação e derivação e c) após interpolação, derivação e RSDW com 10 amostras de transferência. 99
- Figura 39.** Valores de classe das amostras estimados na transferência entre os equipamentos **Tensor 37** e **Phazir** pelos métodos recalibração (RC); sem transferência (ST); RS, PRS e RSDW para discriminação de **andiroba** (◊) das espécies cedro (□), curupixá (Δ) e mogno (o) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão. 102
- Figura 40.** Valores de classe das amostras estimados na transferência entre os equipamentos **Tensor 37** e **Phazir** pelos métodos recalibração (RC); sem transferência (ST); RS, PRS e RSDW para discriminação de **cedro** (□) das espécies andiroba (◊), curupixá (Δ) e mogno (o) e limites discriminação (---). Símbolos vazios

| | |
|---|-----|
| representam amostras de calibração e preenchidos representam amostras de previsão..... | 103 |
| Figura 41. Valores de classe das amostras estimados na transferência entre os equipamentos Tensor 37 e Phazir pelos métodos recalibração (RC); sem transferência (ST); RS, PRS e RSDW para discriminação de curupixá (Δ) das espécies andiroba (\diamond), cedro (\square) e mogno (o) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão..... | 104 |
| Figura 42. Valores de classe das amostras estimados na transferência entre os equipamentos Tensor 37 e Phazir pelos métodos recalibração (RC); sem transferência (ST); RS, PRS e RSDW para discriminação de curupixá (Δ) das espécies andiroba (\diamond), cedro (\square) e mogno (o) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão..... | 105 |
| Figura 43. Espectros das amostras em pó e blocos a) originais e b) derivados e com região selecionada indicada (faixa sombreada). | 112 |
| Figura 44. Valores de classe estimados pelos modelos de discriminação de das espécies a) mogno (\bullet), b) andiroba (\diamond), c) cedro (\square) e d) curupixá (Δ), limites de discriminação (---) e dispersão dos valores estimados para as amostras moídas (\bullet). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão..... | 114 |
| Figura 45. Valores de classe das amostras moídas dos diferentes países estimados pelos modelos para discriminação de a) mogno, b) andiroba, c) cedro e d) curupixá, e limites de discriminação (---). | 115 |
| Figura 46. Mapa político mostrando a posição geográfica dos países Brasil, Honduras, México, Peru e Venezuela. Adaptado da referência | 120 |
| Figura 47. Valores de classe estimados para discriminação de mogno de Brasil (\square), Honduras (\diamond), México (Δ), Peru (\bullet) e Venezuela (\diamond) antes da exclusão dos <i>outliers</i> de previsão e limites de discriminação (---). Símbolos vazios correspondem a amostras de calibração e preenchidos de previsão..... | 123 |

LISTA DE TABELAS

| | |
|---|-----|
| Tabela 1. Nomes usuais e distribuição geográfica de quatro espécies tropicais similares com base em Pastore et al. ² , Camargos et al. ²⁰ e Coradin et al. ⁵ | 9 |
| Tabela 2. Pesquisas em aplicação de NIRS na discriminação de madeira descritas na literatura..... | 11 |
| Tabela 3. Correspondência entre alguns tipos de espectroscopia e tipos de transição energética. ⁴³ | 16 |
| Tabela 4. Regiões espectrais no infravermelho. ⁴⁶ | 17 |
| Tabela 5. Pré-processamentos utilizados nos equipamentos | 68 |
| Tabela 6. Erros e parâmetros de discriminação das quatro espécies nos três equipamentos | 71 |
| Tabela 7. Número de variáveis latentes (VL), valores de erros, limite de discriminação e porcentagem de acerto dos modelos de recalibração (RC), sem transferência (ST), DS, PDS e SDW entre os equipamentos Tensor 37 e Antaris II para discriminação das quatro espécies..... | 85 |
| Tabela 8. Número de variáveis latentes (VL), erros, limite de discriminação, porcentagem de acerto e número outliers de calibração (outc) e previsão (outp) dos modelos de recalibração (RC), sem transferência (ST), RS, PRS, RSDW entre os equipamentos Antaris II e Phazir e recalibração com subconjunto (SubC) para discriminação das quatro espécies. | 94 |
| Tabela 9. Número de variáveis latentes (VL), erros, limite de discriminação, porcentagem de acerto e número outliers de calibração (outc) e previsão (outp) dos modelos de recalibração (RC), sem transferência (ST), RS, PRS e RSDW entre os equipamentos Tensor 37 e Phazir | 101 |
| Tabela 10. Resultados dos modelos PLS-DA para discriminação quanto a espécie de amostras em diferentes formas de apresentação | 113 |
| Tabela 11. Número de variáveis latentes (VL), erros, outliers de calibração (outc) e previsão (outp), limites de discriminação e porcentagens de acerto (%) dos modelos para discriminação de mogno por país de procedência | 122 |

CAPÍTULO 1
Introdução e Objetivos

1. INTRODUÇÃO E OBJETIVOS

O mogno (*Swietenia macrophylla* King) é a uma das madeiras mais valiosas do mundo, apresentando praticamente todas as características desejáveis para a madeira.^{1,2,3,4} Sua ampla exploração levou à considerável redução das populações da espécie e fez com que fosse considerado como espécie ameaçada, sendo adicionada ao Apêndice II da Convenção sobre o Comércio Internacional das Espécies da Flora e da Fauna Selvagens em Perigo de Extinção (CITES, do inglês *Convention on International Trade in Endangered Species of Wild Fauna and Flora*) em 2003.^{5,6,7}

A existência de diversas espécies semelhantes ao mogno dificulta o controle de sua exploração e comércio. Assim, a identificação de espécies florestais produtoras de madeira é uma tarefa crucial para a fiscalização de seu comércio e exploração.

Outro aspecto importante a destacar é que a ocorrência natural do mogno se estende dos Estados Unidos à região da América do Sul que abrange Peru, Bolívia e Brasil e a legislação acerca de seu comércio e exploração varia entre os países.^{6,8,9,10} Portanto, além da identificação quanto à espécie, a identificação da procedência da madeira também é fundamental para a manutenção da exploração legalmente permitida em cada território e a preservação da espécie.

Numa situação de fiscalização ou apreensão, geralmente dispõe-se apenas da madeira em toras ou tábuas e o método geralmente utilizado para identificá-la é o anatômico visual, baseado na comparação de diversos caracteres da madeira analisada com aqueles esperados para uma determinada espécie.^{2,5,11,12,13} Contudo, a quantidade de pessoas especializadas capazes de fazer essa identificação é insuficiente para atender à demanda.^{14,15}

O desenvolvimento de métodos instrumentais para a identificação das espécies é uma alternativa que pode tornar a sua fiscalização mais eficiente. Dentre esses métodos destaca-se a espectroscopia no infravermelho próximo (NIRS, do inglês *Near Infrared Spectroscopy*), que possibilita análises rápidas e

não destrutivas, com simples preparo de amostra e pode ser aplicada a amostras sólidas.^{2,14,16}

Em estudos anteriores, NIRS combinada a análise discriminante por mínimos quadrados parciais (PLS-DA, do inglês *Partial Least Squares Discriminant Analysis*) foi utilizada com sucesso na discriminação de amostras em blocos e em pó de quatro espécies semelhantes - mogno, andiroba, cedro e curupixá - utilizando equipamento de bancada com alto desempenho.^{2,17} Nesse sentido, uma abordagem inovadora na identificação de mogno é a utilização de métodos instrumentais de campo que possibilitem uma supervisão da exploração mais prática e viável.

Entretanto, um aspecto que deve ser considerado na utilização de NIRS para a identificação de espécies de madeira é que a construção de modelos de calibração multivariada pode exigir um número elevado de amostras, implicando custo e tempo. Portanto, é interessante que tais modelos tenham ampla aplicabilidade. Algumas circunstâncias podem fazer com que um modelo perca sua validade, como a mudança de equipamento ou de condições das amostras.

A transferência de calibração oferece uma alternativa para a extensão da aplicabilidade desses modelos diante de mudanças nas condições da análise, evitando a necessidade de reconstruí-los nas novas condições.

A aplicação de transferência de modelos para análises discriminantes foi ainda pouco explorada, sendo que não foram encontrados trabalhos descritos na literatura que avaliam sua eficiência. Assim, o objetivo geral desta dissertação foi avaliar a transferência de modelos PLS-DA entre equipamentos NIRS portátil e de bancada e entre formas de apresentação das amostras (em pó e em blocos) medidas em equipamento de bancada para discriminação entre as espécies mogno, andiroba cedro e curupixá.

Esta dissertação foi dividida em capítulos, que apresentam a revisão bibliográfica, aspectos teóricos e resultados em diferentes situações.

O Capítulo 2 consiste numa breve revisão bibliográfica sobre mogno e identificação de espécies de madeira. No Capítulo 3, detalharam-se alguns conceitos fundamentais de NIRS, PLS-DA e transferência de calibração.

A primeira etapa experimental desta dissertação, descrita no Capítulo 4, consistiu no desenvolvimento de modelos PLS-DA para discriminar as espécies andiroba, cedro, curupixá e mogno utilizando equipamentos NIRS portátil e de bancada. O objetivo foi principalmente verificar a capacidade de discriminação dos modelos desenvolvidos no equipamento portátil e comparar seu desempenho com os dos demais.

No Capítulo 5, foram abordadas duas aplicações de transferência de calibração entre equipamentos. A primeira consistiu na transferência de calibração entre equipamentos de bancada utilizando métodos diretos de padronização, de forma a tornar possível o compartilhamento de dados entre laboratórios que possuem equipamentos diferentes.

A segunda aplicação descrita no Capítulo 5 consistiu na transferência de calibração de equipamentos de bancada com alta resolução para um equipamento portátil de baixa resolução utilizando métodos reversos de padronização. O objetivo foi possibilitar o aproveitamento dos dados dos equipamentos de bancada para construir modelos padronizados capazes de discriminar os espectros do equipamento portátil por espécie.

No Capítulo 6 foi abordada a transferência de calibração entre formas de apresentação das amostras para verificar se amostras moídas de mogno de diferentes países poderiam ser corretamente classificadas como mogno a partir de modelos PLS-DA para discriminação de mogno, andiroba, cedro e curupixá construídos a partir de amostras em blocos medidas no mesmo equipamento de bancada.

No Capítulo 7, NIRS combinada a PLS-DA foi aplicada para discriminar amostras de mogno provenientes de cinco países: Brasil, Honduras, México, Peru e Venezuela.

Por fim, o Capítulo 8 apresenta as conclusões gerais e o Capítulo 9 apresenta as referências bibliográficas desta dissertação.

CAPÍTULO 2
Revisão Bibliográfica

2. REVISÃO BIBLIOGRÁFICA

O mogno (*Swietenia macrophylla* King) é uma das madeiras mais valiosas do mundo, apresentando praticamente todas as características desejáveis para a madeira, como coloração e beleza do cerne, trabalhabilidade, estabilidade dimensional, durabilidade e resistência a pragas.^{1,2,3,4}



Figura 1. *Swietenia Macrophylla* King. Foto por Paulo Ernani Ramalho Carvalho.⁸

Swietenia macrophylla (Figura 1) é uma árvore tropical americana perenifólia, i.e., que permanece folhada ao longo do ano.^{8,18} As árvores dessa espécie medem cerca 30 metros de altura e 1,5 metros de diâmetro e as maiores podem chegar a 70 metros de altura e 3,5 metros de diâmetro.⁸ Seu tronco é alto, reto e sem galhos até aproximadamente a metade de sua altura.^{8,18} O cerne da madeira tem coloração marrom-rosado a marrom-avermelhado e o alburno, marrom-claro.⁵ A Figura 2 apresenta de forma esquematizada as diferentes estruturas que compõem o tronco da árvore e a distinção das faces com relação às três direções de orientação das células da madeira, utilizadas durante o seu corte. A cor e demais características da madeira podem também variar dependendo da origem, das condições do ambiente e do tempo passado desde o corte.^{8,18} A ocorrência natural da

Swietenia macrophylla se estende dos Estados Unidos, México, passando pela América Central até a região da América do Sul que abrange Peru, Bolívia e Brasil.^{6,8}

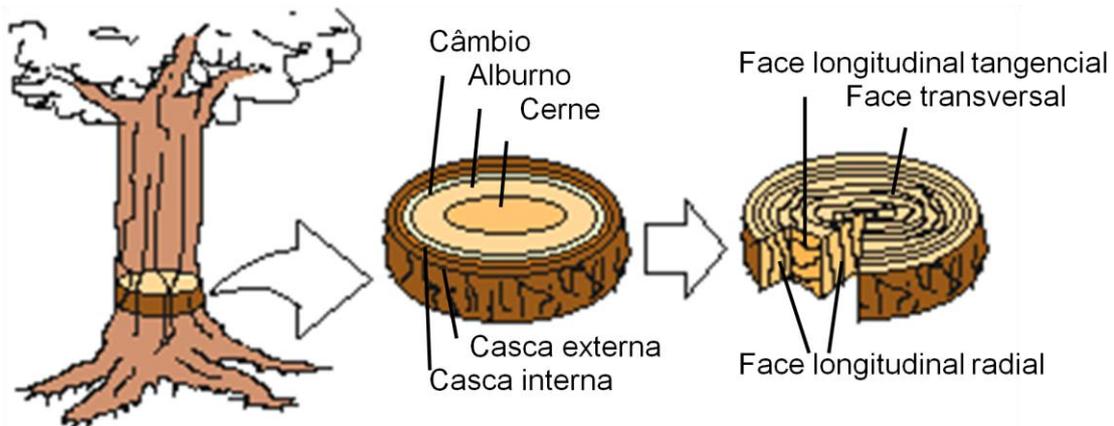


Figura 2. Representação das regiões do tronco e orientação das faces seccionadas.

O mogno já era utilizado pelas civilizações pré-colombianas e com a chegada dos espanhóis passou a ser usado para reparar os navios.¹⁸ Embora já viesse sendo utilizado, o primeiro uso documentado do mogno pelos espanhóis data de 1514, data cravada numa cruz rústica de mogno colocada na catedral de Santo Domingo de Guzmán, na República Dominicana, onde foram feitos também vários ornamentos em mogno.¹⁸ Não é sabido exatamente quando começou a ser exportado para a Europa, mas construções de mais de 400 anos na Espanha já eram constituídas de mogno e sua utilização bem como suas aplicações foram crescentes, abrangendo construção civil, construção de navios e mobílias.¹⁸ A extração de mogno no Brasil teve início com a colonização e a sua comercialização se iniciou na década de 30, no oeste do Acre e se intensificou a partir da década de 60 com a construção de estradas na Amazônia brasileira.^{4,9} No Peru a exploração comercial do mogno iniciou na década de 20 e na Bolívia, na década de 50. Entre as décadas de 70 e 90 o Brasil foi o principal fornecedor de mogno para o comércio internacional.⁹ A excessiva exploração de mogno foi motivada pelo seu alto valor de mercado, o que também suscitou a extração ilegal deste e levou à considerável redução das populações da espécie.^{5,6} Atualmente, o preço do metro cúbico de mogno peruano está entre US\$ 1570 e US\$ 1655.¹⁹

Em 15 de novembro de 2003 a *Swietenia macrophylla* foi incluída no Apêndice II da Convenção sobre o Comércio Internacional das Espécies da Flora e da Fauna Selvagens em Perigo de Extinção (CITES, do inglês *Convention on International Trade in Endangered Species of Wild Fauna and Flora*), ganhando maior proteção regulatória.⁷ São incluídas nesse apêndice as espécies que ainda não estão em perigo de extinção, mas podem ficar caso o comércio de seus espécimes não se submeta a regulação estrita para evitar exploração incompatível com a sobrevivência da espécie.

No Brasil, foram implementadas gradativas restrições à exploração, comercialização e exportação do mogno antes da inclusão da espécie na CITES, principalmente a partir de 1996, quando foi aumentada a área de reserva florestal nas propriedades de 50% para 80% e suspenderam-se as autorizações de exploração de mogno por dois anos.¹⁰ Em 2001 o governo brasileiro proibiu o corte, transporte e comercialização do mogno.⁹ Tudo isso contribuiu para o aumento das exportações do Peru.^{6,10} Após a inclusão na CITES, o Peru definiu cotas, reduzindo a exportação.⁶ A Bolívia também forneceu grande quantidade de mogno para o comércio no início da década de 90 até a quase exaustão dos estoques e uma nova regulamentação ambiental freou a exploração ilegal.⁹ O Peru se tornou o principal exportador de mogno devido à proibição da exploração no Brasil em 2001 e ao alto preço do mogno boliviano.⁹ As exportações do Peru aumentaram drasticamente no ano de 2002, e diminuíram gradativamente a partir de 2003, ano em que a espécie foi incluída no Apêndice II da CITES.⁶

Um fator que dificulta o controle da exploração e do comércio do mogno é que existem diversas espécies semelhantes ao mogno, de forma que são necessárias técnicas de identificação para evitar o comércio e a exploração ilegais dessa espécie e preservá-la.

A variedade de nomes populares para as espécies contribui para a dificuldade de identificação destas. Isso decorre não somente de regionalismos mas também da prática relativamente comum de utilizar nomes inadequados para burlar a fiscalização e da introdução no mercado de espécies menos conhecidas em substituição a outras já consagradas pelo uso utilizando o nome

destas.²⁰ Assim, existem dois problemas: i) várias espécies comercializadas com o mesmo nome e ii) uma espécie comercializada com nomes vários diferentes; e mesmo que alguns países tenham uma nomenclatura oficial padronizada, por se tratar de um mercado internacional, apenas a nomenclatura botânica formal evitaria dúvidas na identificação.¹¹ A Tabela 1 mostra os nomes usuais dados a quatro espécies tropicais, com os nomes mais comuns em negrito, e as regiões onde podem ser encontradas.

Tabela 1. Nomes usuais e distribuição geográfica de quatro espécies tropicais similares com base em Pastore et al.², Camargos et al.²⁰ e Coradin et al.⁵

| Espécie | Outros nomes usuais | Distribuição |
|---|--|--|
| <i>Carapa guianensis</i> Família: Meliaceae Nome comercial: Andiroba | Aboridã (PA), andiroba (AM, PA), andiroba-aruba, andiroba-branca (PA), andiroba-do-igapó (PA), andiroba-lisa, andiroba-saruda (PA), andiroba vermelha (PA), andirobeira, andirobeira-branca, andirobeira-vermelha, andirova (PA), angiroba (PA), angirova, audirova, camaçari (PA), carapa, carapá, carapinha (PA), caropá (PA), comaçari, gendiroba (PA), iandiroba (PA), iandirova, jandiroba (PA), jandirova, landirova, mandiroba (PA), nandirova, penaíba (PA), purga-de-santo-antônio, purga-de-santo-inácio (PA) | Belize, Brasil, Colômbia, Costa Rica, Cuba, República Dominicana, Equador, Guiana Francesa, Guiana, Honduras, Nicarágua, Panamá, Peru, Suriname, Venezuela, Trindade |
| <i>Cedrela odorata</i> Família: Meliaceae Nome comercial: Cedro | Acaju, acuju, capiúva, cedreilro, cedrilho, cedrinho, cedro , cedro-amargo, cedro-amargoso, cedro-aromático, cedro-batata, cedro-bordado, cedro branco, cedro-bravo, cedro-cheiroso (BA), cedro-de-mato-grosso, cedro-do-amazonas, cedro-do-brejo, cedro-do-paraguai, cedro-fêmea (BA), cedro-manso, cedro-mogno, cedro-pardo, cedro-rosa (RJ, SP), cedro-verdadeiro, cedro-vermelho (PA, BA) | Argentina, Belize, Bolívia, Colômbia, Costa Rica, Cuba, Equador, El Salvador, Guiana Francesa, Gana, Honduras, Jamaica, Madagascar, México, Nicarágua, Panamá, Paraguai, Peru, Suriname, Venezuela |
| <i>Micropholis melinoniana</i> Família: Sapotaceae Nome comercial: Curupixá | Pau-de-sapo (AM), curupixá | Brasil, Guiana Francesa, Guiana, Suriname |
| <i>Swietenia macrophylla</i> Família: Meliaceae Nome comercial: Mogno | Acaju, aguano , araputanga, caoba, cedro, cedro-aguano, cedro-mogno, cedroí, cedrorana (PA); mara, mara-vermelho (AM); mogno , mogno-de-marabá, mogno-aroeira, mogno-branco, mogno-brasileiro, , mogno-cinza, mogno-claro, mogno-do-Rio, Jurupari (AM) , mogno-escuro, mogno-peludo, mogno-rosa, mogno-róseo, mogno-vermelho | Bolívia, Brasil, Colômbia, México, Peru, Venezuela |

A repetição de nomes entre espécies que são visualmente semelhantes e a grande variedade de denominações para cada uma exemplificam a dificuldade de identificação destas.

A identificação de espécies florestais produtoras de madeira é uma tarefa crucial para a fiscalização de seu comércio e exploração. A identificação taxonômica de espécies é feita pela observação de seus frutos, flores e demais características nas diferentes fases do ciclo de vida. Contudo, numa situação de fiscalização ou apreensão, dispõe-se apenas da madeira em toras ou tábuas, sendo necessária uma metodologia alternativa para identificá-la. O método tradicionalmente utilizado com esse fim é a identificação anatômica visual. Ele baseia-se na comparação de diversos caracteres da madeira analisada com aqueles esperados para uma determinada espécie. Alguns desses caracteres são: densidade; cor; cheiro; brilho; textura; anéis de crescimento; presença, visibilidade e disposição dos vasos e fibras; forma, disposição e visibilidade das células; e porosidade.^{2,5,11,12,13}

Existem chaves de identificação que organizam os parâmetros para as características de várias espécies, contudo além desses parâmetros não serem precisamente objetivos, algumas vezes as características padrão de cada espécie diferem entre as fontes na literatura de forma que esse método requer que a madeira seja examinada por especialistas. Entretanto, a quantidade de pessoas especializadas é insuficiente para atender à demanda.^{14,15} Programas computacionais foram criados para facilitar a identificação dos caracteres da madeira principalmente pela comparação com imagens específicas para cada caractere, oferecendo também bancos de imagens de várias espécies para comparação^{21,22}. O desenvolvimento de métodos instrumentais para a identificação das espécies é uma alternativa que pode tornar a sua fiscalização mais eficiente.

Dentre esses métodos destaca-se a espectroscopia no infravermelho próximo (NIRS, do inglês *Near Infrared Spectroscopy*). Essa técnica oferece vantagens para o estudo da madeira, uma vez que possibilita análises rápidas e não destrutivas, com simples preparo de amostra e pode ser aplicada a amostras sólidas.^{2,14,16} A espectroscopia no infravermelho próximo foi utilizada

no estudo de diferentes propriedades da madeira com diversos propósitos, 23,24,25,26,27,28,29 e também para a discriminação e identificação de madeira. Na Tabela 2 estão relacionados alguns trabalhos onde NIRS foi utilizada para discriminar madeira.

Tabela 2. Pesquisas em aplicação de NIRS na discriminação de madeira descritas na literatura

| Autor | Objetivo | Nº/preparo de amostra | Método | Conclusão |
|---|---|---|--------------------|--|
| Adedipe <i>et al.</i> , ³⁰ 2008 | Diferenciar duas espécies de carvalho | 100 calibração, 50 teste/ Secagem em estufa por 24h | FT-NIR, SIMCA, PCA | Obtiveram menores erros para 5% de significância, com 98% de acerto na classificação |
| Braga <i>et al.</i> , ¹⁷ 2011 | Discriminar mogno, andiroba, cedro e curupixá em blocos | 66 calibração, 45 teste/ Secagem ao ar livre, lixamento | FT-NIR, PLS-DA | Obtiveram 100% de acerto na separação |
| Brunner <i>et al.</i> , ³¹ 1996 | Discriminar 12 espécies de madeira em duas séries de 6 | 2 a 5 por espécie sem preparo na a série 1 e 11 cubos serrados para cada espécie, três deles sem preparo e 8 fervidos em mistura glicerol-água e cortados com micrótomo na série 2. | FT-NIR, PCA | Concluíram ser possível discriminar espécies. Atribuíram não terem obtido 100% de acerto a terem poucas amostras, e concluíram que as amostras deveriam ter o mesmo preparo. |
| Gierlinger <i>et al.</i> , ³² 2004 | Discriminar três espécies do gênero <i>Larix</i> | 15 de cada espécie/ Sem preparo | FT-NIR, SIMCA, PCA | Obtiveram 100% de classificações corretas |
| Haartveit e Flaete, ³³ 2008 | Discriminar cerne seco e albarno de mesma espécie de pinheiro | 105 calibração e 54 teste/Serração imediatamente antes das medições | NIRS, PLS-DA | Obtiveram 96,3% de classificação correta (duas classificações erradas na validação) |
| | Discriminar madeiras secas de três espécies diferentes de abeto | 47 calibração e 36 teste/ Secagem ao ar | NIRS PLS-DA | Obtiveram 94% de classificação correta (duas classificações erradas na validação) |

| Autor | Objetivo | Nº/preparo de amostra | Método | Conclusão |
|---|--|--|---------------------------------|--|
| Pastore <i>et al.</i> , ² 2011 | Discriminar mogno, andiroba, cedro e curupixá em forma de pó | 88 calibração, 44 teste/Secagem ao ar, moagem. | FT-NIR, PLS-DA | Obtiveram 100% de acerto na separação |
| Sandak <i>et al.</i> , ³⁴ 2011 | Separar madeiras de abeto de mesma espécie provenientes da Finlândia, sul e norte da Polônia e diferentes regiões da Itália | 22, 26, 34, 57, 36, 43, 37 de sete localidades na Itália, as três últimas na mesma cidade, 32 da Finlândia, 78 do norte da Polônia e 74 do sul/Secagem em estufa, condicionamento em câmara climática (por semanas), lixamento, novo condicionamento por dois dias | FT-NIR, PCA | Obtiveram 100% de classificação tanto entre países como entre as localidades da Itália |
| Sandberg e Sterley, ³⁵ 2009 | Separar cerne e alburno de madeira de abeto seca de mesma espécie | 44 de calibração e 16 de teste/ Condicionamento em câmara climática, leve lixamento antes das medições | NIRS, PLS, OSC | Concluíram que a separação é possível(1 classificação errada na validação) |
| Tsuchikawa <i>et al.</i> , ¹⁴ 2003 | Discriminar 7 espécies de madeira e entre cerne e alburno de uma espécie, usando amostras em diferentes condições de secagem | 30 de calibração e 18 de teste para cada espécie em uma condição e 15 de calibração e 5 de validação em outra | NIRS e distância de Mahalanobis | Obtiveram porcentagens de acerto variadas para as condições testadas. Obtiveram 100% de acerto na melhor condição. |

Um ramo crescente nas pesquisas em identificação de madeira é a análise de imagens, onde a identificação dos caracteres da madeira é automatizada pelo cruzamento com uma base de dados e variadas tecnologias de processamento de imagem para “extrair” o caractere da imagem.³⁶ Khalid *et al.*³⁶ obtiveram 95% de classificações corretas para 20 espécies da Malásia. Esse método requer que as imagens sejam feitas com uma câmera especial com aumento de 10 vezes e depois analisadas em um computador. A desvantagem desse método é que requer extensas bases de dado que demandam memória elevada dos computadores e que se baseia em características lineares, sendo que as madeiras tropicais apresentam

caracteres não lineares.³⁷ Yusof *et al.*³⁷ propuseram um método para reduzir o tamanho da base de dados e aplicar a características não lineares e obtiveram 98,69% de classificações corretas para 53 espécies. O método proposto por eles possibilita análises em tempo real.

Na identificação da proveniência geográfica da madeira têm se utilizado a análise do DNA por meio da comparação genotípica das amostras com um banco de dados.^{38,39} Esse banco pode ser bastante extenso dependendo da aplicação, entretanto o DNA pode ser extraído de outras partes da planta, como as folhas, o que pode facilitar o processo.^{38,39} O método apresenta a dificuldade de extração do DNA completo.^{38,39} Outra abordagem possível é a análise do DNA dos microrganismos presentes na madeira, podendo ser fungos característicos da origem desta.⁴⁰

Uma abordagem interessante à identificação de madeira é o desenvolvimento de métodos de campo, que podem tornar mais práticas a supervisão da exploração e a preservação das espécies.

Cordeiro *et al.*⁴¹ discriminaram cedro, imbuia, perobinha e angelim utilizando duas técnicas diferentes: nariz eletrônico e espectroscopia de emissão em plasma induzido por laser (LIBS, do inglês *Laser-induced Breakdown Spectroscopy*). Eles obtiveram 100% de acerto na discriminação das espécies. Além de possibilitarem análises *in situ*, o nariz eletrônico criado por eles tem um custo baixo (aproximadamente US\$ 10). Foi utilizada uma única amostra de cada espécie medida várias vezes. Embora as técnicas não exijam necessariamente um vasto conjunto de amostras, convém destacar que é necessário utilizar um conjunto de amostras representativo da espécie. A LIBS mediu a emissão de dez elementos para a caracterização. Essa técnica tem as desvantagens de estar sujeita a variações do laser e do alto custo do equipamento.

Cabral *et al.*⁴² discriminaram *Cedrela odorata*, *Cariniana rubra*, *Microphilis venulosa*, *Hymenolobium ssp.*, *Dinizia excels* e *Hymenolobium petraeu* utilizando espectrometria de massas com auto bombeamento via efeito Venturi e spray-sônico para ionização ambiente em amostras líquidas (V_L -EASI-MS e V_L -EASI-MS/MS, do inglês *Venturi Easy Ambient Sonic-spray Ionization* e *Mass*

Spectrometry) e sólidas (V_S -EASI-MS e V_S -EASI-MS/MS). Eles não especificaram o número de amostras, mas o preparo da amostra líquida foi um simples extrato metanólico de um pequeno pedaço da madeira em um frasco, que pode ser feito *in situ* e as amostras sólidas foram apenas lixadas. Obtiveram perfil característico para o mogno, independente do tempo de extração e idade e das demais espécies possibilitando 100% de classificação correta. Identificaram coloração artificial utilizada em outras espécies para falsificar mogno. A técnica dispensa a construção de modelos de calibração, o V-EASI é portátil e a ionização empregada não requer aquecimento, radiação ou eletricidade. O espectrômetro utilizado não era portátil, mas há a perspectiva de realizar as medidas em campo. A desvantagem do método é o consumo de gás e solvente.

Na utilização de algumas técnicas, como NIRS, é necessário construir modelos de calibração que possam ser posteriormente aplicados a amostras reais. Uma regressão bastante utilizada na construção de modelos para discriminação de madeira é a regressão por mínimos quadrados parciais para análise discriminante (PLS-DA, do inglês *Partial Least Squares Discriminant Analysis*). A construção desses modelos requer que várias amostras sejam medidas, de forma que é interessante que estes tenham largas durabilidade e aplicabilidade. Para isso, é possível construir modelos robustos ou realizar operações de transferência do modelo de calibração para uma condição diferente daquela em que foi construído. A aplicação de transferência de calibração a análises discriminantes foi ainda pouco explorada. Conceitos e princípios fundamentais referentes a NIRS, PLS-DA e transferência de calibração serão detalhados no próximo capítulo.

CAPÍTULO 3
Fundamentação Teórica

3. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão detalhados alguns princípios fundamentais acerca de espectroscopia de infravermelho próximo (NIRS), análise discriminante por mínimos quadrados parciais (PLS-DA) e estratégias de transferência de calibração.

3.1 ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO

A espectroscopia estuda as transições entre estados energéticos de átomos ou moléculas por meio da interação destes com radiação eletromagnética.^{43,44} A interação entre matéria e radiação se estende ao longo de todo espectro eletromagnético.⁴⁵ Dependendo do tipo de níveis energéticos envolvidos, as transições são observadas em diferentes regiões espectrais, sendo estudadas pelas técnicas espectroscópicas específicas das faixas correspondentes, estando algumas dessas técnicas ilustradas na Tabela 3.⁴⁴

Tabela 3. Correspondência entre alguns tipos de espectroscopia e tipos de transição energética.^{43,46}

| Tipos de transição | Tipos de espectroscopia | Comprimento de onda usual (m) |
|-----------------------|--------------------------------|---|
| Nuclear | Raio γ | $5 \cdot 10^{-13} - 1,4 \cdot 10^{-10}$ |
| Eletrônica interna | Raio X | $1 \cdot 10^{-11} - 1 \cdot 10^{-8}$ |
| Eletrônica de ligação | Ultravioleta | $1 \cdot 10^{-8} - 1,8 \cdot 10^{-7}$ |
| | Ultravioleta-visível | $1,8 \cdot 10^{-7} - 7,8 \cdot 10^{-7}$ |
| Vibracional | Infravermelho e Raman | $7,8 \cdot 10^{-7} - 3 \cdot 10^{-5}$ |
| Rotacional | Micro-ondas | $7,5 \cdot 10^{-4} - 0,375$ |
| Orientação de spin | Ressonância de spin eletrônico | $3 \cdot 10^{-2}$ |
| | Ressonância Magnética Nuclear | 0,6 – 10 |

A radiação infravermelha foi descoberta em 1800 por Herschel, porém apenas cerca de 100 anos depois teve início sua aplicação à espectroscopia quando Coblentz obteve espectros de absorção no infravermelho para vários compostos orgânicos em diferentes estados físicos.⁴⁴

As transições observadas na região do infravermelho são as vibracionais, embora em alguns casos transições rotacionais sejam também observadas na

região do infravermelho distante.⁴⁴ A região do infravermelho pode ser dividida em infravermelho próximo, médio e distante, conforme disposto na Tabela 4.

Tabela 4. Regiões espectrais no infravermelho.⁴⁶

| Região | Comprimentos de onda (μm) | Números de onda (cm^{-1}) | Frequências (Hz) |
|----------|--|--------------------------------------|---|
| Próximo | 0,78 - 2,5 | 12.800 - 4.000 | $3,8 \cdot 10^{14}$ - $1,2 \cdot 10^{14}$ |
| Médio | 2,5 - 50 | 4.000 - 200 | $1,2 \cdot 10^{14}$ - $6,0 \cdot 10^{12}$ |
| Distante | 50 - 1000 | 200 - 10 | $6,0 \cdot 10^{12}$ - $3,0 \cdot 10^{11}$ |

Os movimentos que podem ocorrer em uma molécula são: i) translação, onde a molécula se move como um todo pelo espaço, deslocando-se o seu centro de gravidade; ii) rotação, onde a molécula se move em torno de seu centro de gravidade; e iii) vibração, onde cada átomo individualmente se move com relação aos outros átomos.⁴⁶ Os modos vibracionais de uma molécula estão representados na Figura 3.

Para que um modo vibracional seja ativo no infravermelho, deve haver uma alteração rítmica no momento de dipolo durante o movimento, e esse momento de dipolo é determinado pela diferença de carga e pela distância entre os centros de carga.^{44,46,47} Assim, o campo elétrico oscilante da radiação interage com o campo elétrico alternado produzido pela mudança de momento de dipolo da molécula e modifica a amplitude do movimento.^{46,47} Quando a frequência da radiação é igual à da vibração natural da molécula, ocorre absorção e variação na amplitude de vibração molecular.⁴⁶

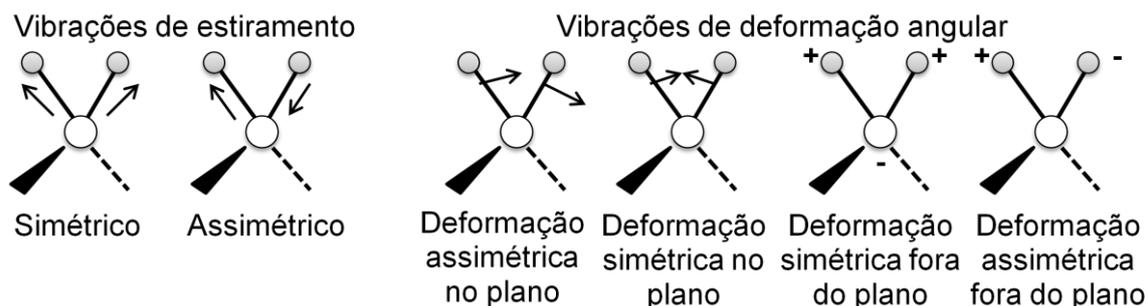


Figura 3. Modos vibracionais em uma molécula, onde + e - indicam movimentos perpendiculares ao plano da página.^{46,47}

Cada tipo de ligação possui uma frequência de vibração natural característica.⁴⁸ Contudo, ligações de mesmo tipo em compostos diferentes apresentam frequências vibracionais ligeiramente diferentes.⁴⁸ Logo, ainda que

algumas bandas possam coincidir, cada molécula tem um espectro de infravermelho único, comparável a uma impressão digital.⁴⁸

A vibração entre dois átomos pode ser aproximada como um movimento harmônico simples, cuja energia potencial está ilustrada pela linha tracejada (curva 1) na Figura 4.⁴⁶ Nesse sistema a energia potencial varia periodicamente com a distância entre as massas e a força entre elas obedece a lei de Hooke, sendo proporcional à compressão ou extensão da ligação com relação ao equilíbrio.^{43,46} Termos quânticos podem ser acrescentados para adaptar as equações que descrevem o oscilador harmônico segundo a mecânica clássica para considerar a natureza quantizada das energias vibracionais moleculares, onde estas só podem assumir alguns valores discretos.⁴⁶

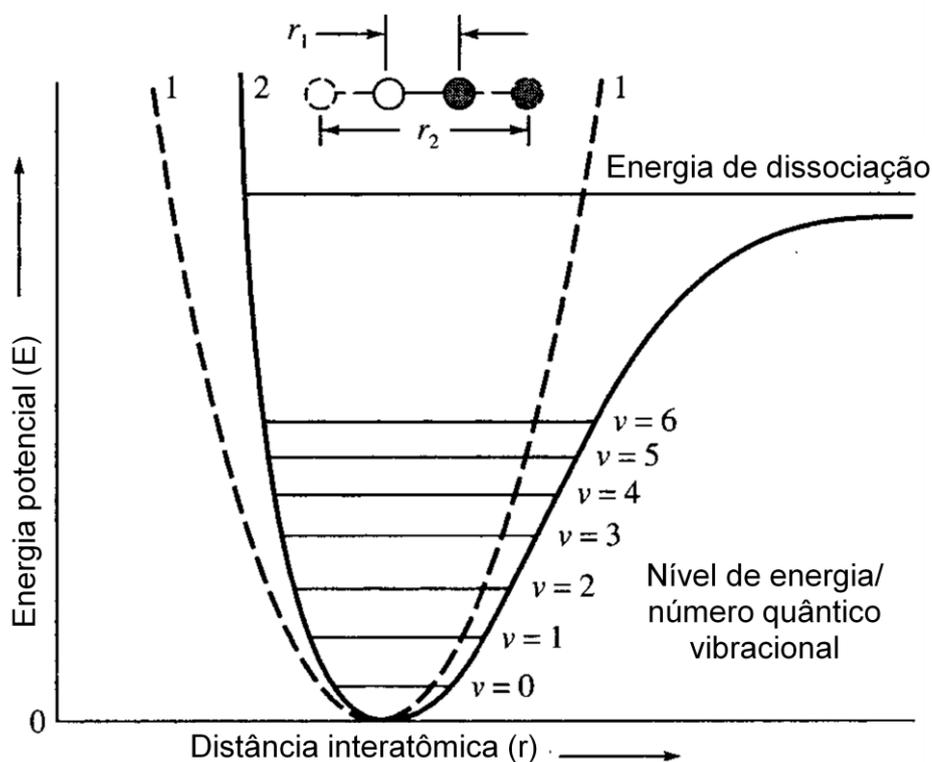


Figura 4. Diagrama de energia potencial de um oscilador harmônico (curva 1, (- - -)) e de um movimento anarmônico (curva 2, (—)). Figura extraída da referência 46.

No modelo do oscilador harmônico quântico, as transições vibracionais permitidas são aquelas onde a variação do número quântico é de uma unidade ($\Delta v = \pm 1$) e os níveis vibracionais são igualmente espaçados, sendo esperado um único pico de absorção para cada vibração.⁴⁶ Isso significa que a molécula

pode ganhar ou perder apenas um quantum de energia em cada transição vibracional.⁴³ Essas transições são chamadas transições fundamentais.

Entretanto, o oscilador harmônico não descreve perfeitamente a vibração molecular, uma vez que há outros fatores envolvidos no sistema além da força restauradora da ligação, como a força de repulsão coulombica entre os núcleos e o decréscimo da força restauradora quando a distância interatômica se aproxima daquela em que ocorre a dissociação dos átomos.⁴⁶

As equações de onda da mecânica quântica podem fornecer curvas de energia potencial mais acertadas, que possuem a forma anarmônica representada pela linha contínua (curva 2) na Figura 4, porém a elevada complexidade matemática restringe sua aplicação quantitativa a sistemas simples.⁴⁶ Essas curvas anarmônicas podem diferir do comportamento harmônico em diversos graus, mas observa-se que o comportamento das duas curvas é semelhante em energia potencial baixa.⁴⁶

No oscilador anarmônico a diferença de energia entre os níveis diminui à medida que o número quântico aumenta, como pode ser observado na Figura 4, e a diferença entre os níveis chega a zero no limite da dissociação dos átomos.⁴⁹ Transições com variação de duas ou três unidades de números quânticos ($\Delta v = \pm 2$ ou $\Delta v = \pm 3$) são observadas.^{43,46,49} Essas transições são chamadas sobretons (ou, em inglês, *overtones*) e apresentam frequências de aproximadamente duas ou três vezes a frequência da vibração fundamental e menores intensidades com relação a esta.⁴⁶

As várias vibrações que ocorrem numa molécula também podem interagir provocando acoplamentos que produzem bandas de combinação de baixa intensidade com frequências aproximadamente equivalentes à adição ou subtração de suas frequências fundamentais.⁴⁶ Acerca do acoplamento vibracional, podem ser feitas as seguintes considerações: i) as vibrações devem ter a mesma espécie de simetria; ii) deve haver um átomo comum às vibrações de estiramento para que haja um acoplamento forte entre estas; iii) pode ocorrer acoplamento entre vibrações de estiramento e de deformação angular se a ligação envolvida na primeira formar um lado do ângulo que varia na segunda; iv) deve haver uma ligação comum entre os grupos envolvidos nas

deformações angulares para que haja acoplamento entre estas; v) Quanto mais próximas as frequências de absorção dos grupos envolvidos, maior o acoplamento; e vi) quando os grupos estão separados por um ou mais átomos e as vibrações são perpendiculares entre si o acoplamento é pequeno ou inexistente.^{46,47}

A maior parte das transições vibracionais fundamentais situa-se na região do infravermelho médio, sendo a espectroscopia no infravermelho médio uma excelente ferramenta para identificação de compostos orgânicos.^{46,47,48}

A região do infravermelho próximo situa-se entre a região do infravermelho médio e a região do visível e corresponde ao intervalo de 4.000 cm^{-1} a aproximadamente 12.800 cm^{-1} .⁴⁶ Nessa região as bandas de absorção resultam de sobretons e combinações de vibrações fundamentais que ocorrem entre 3.000 e 1.700 cm^{-1} (infravermelho médio).^{45,46} Por isso, as absorvidades molares são baixas mas a radiação penetra mais profundamente na amostra possibilitando análise mais representativa e minimizando efeitos de impurezas.^{45,46} Dada a pequena massa molar do hidrogênio, este apresenta vibrações de maior amplitude e maior desvio do comportamento harmônico, de forma que as bandas observadas na região do infravermelho próximo geralmente são devidas a ligações entre hidrogênio e outros átomos leves.¹⁶ As bandas observadas no infravermelho próximo correspondem principalmente às vibrações das ligações C—H, O—H e N—H.^{45,46} Workman apresenta uma relação de comprimentos de onda para bandas correspondentes às principais ligações observadas nessa região.⁵⁰

As vantagens de NIRS são: i) possibilidade de análises não destrutivas; ii) preparo de amostra simples ou dispensável; iii) análises rápidas; iv) pouco treinamento necessário para utilização da técnica; v) determinação simultânea de vários analitos; vi) determinação de propriedades físicas; vii) robustez e baixo custo relativo dos equipamentos; e viii) possibilidade de análises em tempo real e de portabilidade.¹⁶

Já as desvantagens dessa técnica são: i) baixas seletividade e sensibilidade, o que pode ser minimizado pelo uso de quimiometria em alguns casos; ii) dificuldade de atribuição do sinal a ligações específicas, de forma que

em muitos casos os modelos de calibração são estritamente empíricos; iii) necessidade de grande quantidade de amostras para construção do modelo de calibração, dependendo da aplicação, o que conseqüentemente envolve custos, trabalho e tempo; iv) aplicabilidade pouco ampla dos modelos de calibração sendo necessário reconstruí-los com frequência; e v) assim como muitas técnicas instrumentais, requer valores de referência para construção dos modelos.¹⁶ Uma possibilidade para contornar desvantagens (iii) e (iv) é utilizar estratégias de transferência de calibração ou a construção de modelos robustos, tópicos que serão abordados na seção 3.2.2 deste capítulo.

3.1.1 Medidas em NIRS

As medidas em NIRS podem ser feitas nos modos de transmitância, transflectância, reflectância difusa, interatância e transmitância através de meio dispersivo, conforme esquematizado na Figura 5.⁵¹

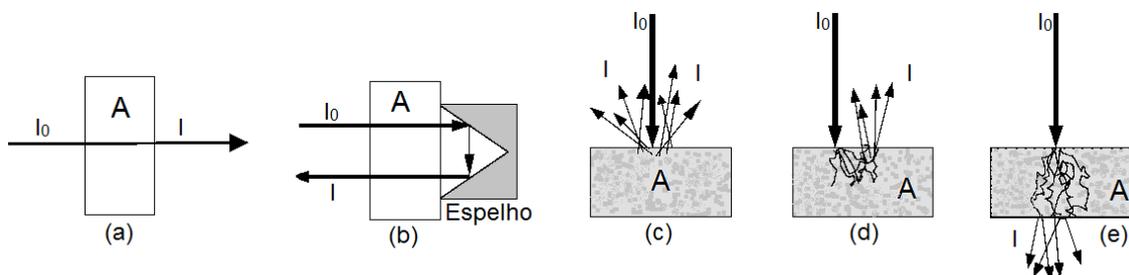


Figura 5. Medidas no infravermelho próximo por (a) transmitância; (b) transflectância; (c) reflectância difusa; (d) interatância; e (e) transmitância através de meio dispersivo. Onde A representa amostra, I_0 a radiação da fonte e I a radiação após a passagem, reflexão e/ou interação com a amostra. Figura adaptada da referência 51.

A transmitância pode ser obtida para amostras transparentes calculando-se a razão entre as potências feixe de radiação antes e depois de passar através da amostra e o valor de absorvância pode ser obtido calculando-se o logaritmo decimal da transmitância.⁴⁶ A absorvância por sua vez apresenta uma relação linear com a concentração e é proporcional ao caminho óptico, segundo a Lei de Beer.⁴⁶ A transflectância é uma forma diferente de medir a transmitância na qual o caminho óptico é duplicado, uma vez que a radiação passa pela amostra duas vezes.⁵¹ Normalmente a transmitância é utilizada para líquidos e a transflectância para emulsões e líquidos túrbidos.¹⁶

Nas medidas de interatância, a radiação incide sobre amostra e há uma maior probabilidade de interação entre elas, de forma que a radiação que emerge da amostra, coletada em um ponto diferente e relativamente distante do ponto de incidência, fornece resultados mais representativos acerca dos constituintes da amostra.⁵¹

A transmitância em sólidos apresenta um caminho óptico maior e fornece informações mais completas acerca da composição média da amostra, sendo apropriada para determinação de princípios ativos em comprimidos farmacêuticos, por exemplo.⁵¹

A medida de reflectância difusa é mais utilizada em NIRS, de um modo geral é aplicada a amostras sólidas e requer tempo e preparo de amostras reduzidos.^{16,46,51}

Sempre que um feixe de radiação encontra uma fronteira entre meios, ocorre a reflexão.⁴⁵ Quando a radiação incide sobre sólidos granulados ou fibrosos, o feixe encontra uma nova interface a cada pequena fração de milímetro, de forma que a reflexão é dispersa em várias direções.⁴⁵ A radiação incidente provoca a excitação dos modos vibracionais ao penetrar as superfícies da amostra, de forma que a radiação que emana da amostra fornece um espectro relacionado à sua composição.⁴⁶ Como a Lei de Beer só se aplica a materiais homogêneos e transparentes, um modelo que estabelece uma relação linear entre a concentração (C) e a reflectância difusa (R) foi proposto por Kubelka e Munk:⁵¹

$$f(C) = \frac{(1-R')^2}{2R'} \quad \text{Equação 1}$$

Onde R' é dada pela equação 2:⁵¹

$$R' = \frac{I_R}{I_{R0}} \quad \text{Equação 2}$$

sendo I_R a intensidade de radiação refletida pela amostra e I_{R0} a intensidade refletida por um material que não absorve em toda faixa espectral da medida.⁵¹

Um acessório comumente utilizado para obtenção de medidas de reflectância é a esfera integradora. Esta se trata de uma esfera revestida

internamente por um material altamente reflexivo, geralmente politetrafluoretileno (PTFE).⁵² A radiação entra na esfera por um orifício, é direcionada para a amostra e refletida por esta.⁵² Os feixes da radiação refletida difusamente pela amostra são também refletidos em todas as direções pela parede da esfera de forma que a luz é distribuída uniformemente e recebida pelo detector.⁵² Esse acessório confere portanto maior sensibilidade aos equipamentos de NIRS, uma vez que há um aproveitamento de toda a radiação refletida, que fica confinada no interior da esfera, e estabilidade devido à uniformidade da radiação percebida no detector, o que favorece a análise de materiais heterogêneos.⁵²

Entretanto, o modelo de Kubelka e Munk é com frequência substituído pela equação 3, que é mais prática por considerar apenas a reflectância (R) da amostra.⁵¹

$$f(C) = \log \frac{1}{R} \quad \text{Equação 3}$$

Pode-se assumir um comportamento linear da equação 3 com a concentração do analito quando a variação na reflectância é pequena, sendo essa relação também análoga à Lei de Beer.⁵¹

3.1.2 Instrumentação em NIRS

Três tipos de instrumentos para espectroscopia no infravermelho serão abordados adiante: espectrofotômetros dispersivos, espectrofotômetros com transformada de Fourier e espectrofotômetros portáteis.

3.1.2.1 Espectrofotômetros dispersivos

Até a década de 80, os espectrofotômetros dispersivos eram os mais utilizados para medidas no infravermelho.⁴⁶ Nesses equipamentos, um feixe de luz infravermelha passa pela amostra, a radiação transmitida é varrida por

dispersão numa rede de difração e o espectro é obtido pela rotação dessa rede.⁴⁷

3.1.2.2 Espectrofotômetros com transformada de Fourier

Os primeiros espectrofotômetros de infravermelho com transformada de Fourier (FT-IR, do inglês *Fourier Transform Near Infrared*) eram caros, volumosos e requeriam muita manutenção.⁴⁶ Por isso sua aplicação se limitava a situações onde sua velocidade, sensibilidade, alta resolução e precisão e exatidão de comprimento de onda eram indispensáveis.⁴⁶ Atualmente os equipamentos FT-IR são muito mais compactos, robustos e baratos, o que fez com que os instrumentos dispersivos fossem amplamente substituídos.⁴⁶

No espectrômetro FT-IR, a radiação contendo todos os comprimentos de onda é separada em dois feixes, um deles percorre uma distância fixa enquanto o outro percorre uma distância variável pela utilização de um espelho móvel.⁴⁷ Obtém-se um interferograma ao variar as distâncias percorridas gerando interferências construtivas e destrutivas, o que varia a intensidade da radiação detectada.⁴⁷ Realiza-se então a transformada de Fourier para converter o interferograma, que está no domínio do tempo, para um que esteja no domínio das frequências e ao fazer a transformada de Fourier em sucessivas posições do espelho móvel obtém-se o espectro de infravermelho da amostra.⁴⁷

Algumas vantagens de se utilizar um espectrômetro FT-IR são: i) melhor aproveitamento da energia radiante, uma vez que não há fendas que atenuam a radiação; ii) alta resolução e reprodutibilidade do comprimento de onda; iii) Rapidez na obtenção do espectro, uma vez que todos os elementos da fonte atingem o detector ao mesmo tempo; iv) melhor relação sinal/ruído.⁴⁶

3.1.2.3 Espectrofotômetros portáteis

Na última década, espectrofotômetros de infravermelho próximo portáteis têm sido de grande contribuição nas indústrias farmacêutica, de alimentos e em estudos forenses. Esses equipamentos compactos, leves e dotados de bateria possibilitam análises não destrutivas, rápidas, em campo, algumas vezes podem ser utilizados através das embalagens de produtos e, em alguns casos, apresentam baixos custos em comparação a espectrofotômetros FT-IR.⁵³

Há diferentes tecnologias aplicadas na confecção desses equipamentos e um tipo de espectrômetro portátil bastante utilizado com essas finalidades aplica baseia-se em espectroscopia com transformada digital (DTS, do inglês *Digital Transform Spectroscopy*) e sistemas micro eletromecânicos (MEMS, do inglês *Micro Electro-Mechanical Systems*).^{53,54} A Figura 6 mostra um esquema desse tipo de equipamento.⁵⁵

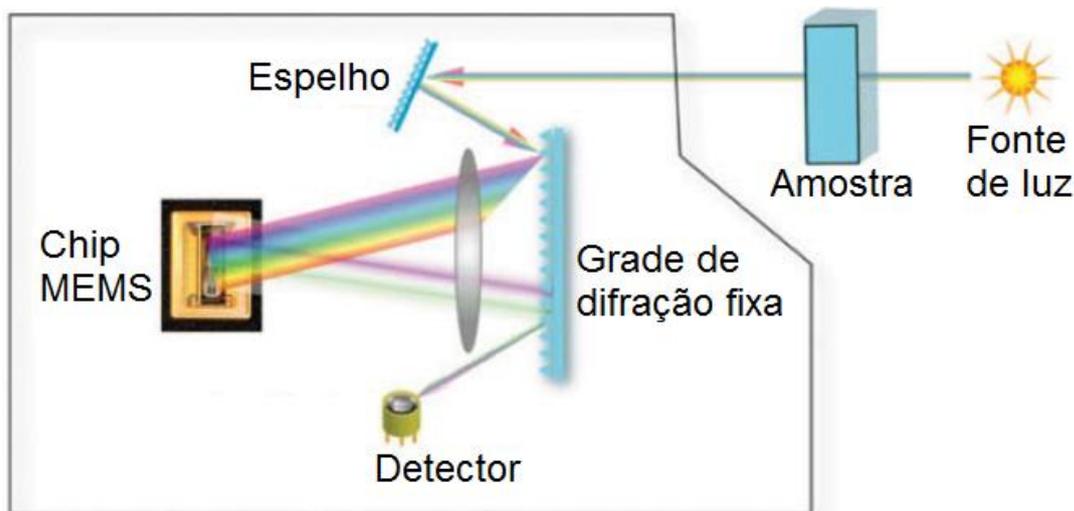


Figura 6. Esquema de equipamento DTS portátil. Adaptado da referência 55.

A tecnologia MEMS consiste na integração de sistemas mecânicos e eletrônicos em chips de silício fixos que funcionam como filtros para seleção comprimentos de onda.^{53,56} A transformada digital ajuda a melhorar a razão sinal ruído.^{53,54} Assim como a transformada de Fourier, a DTS utiliza uma modulação para transformar a codificação em domínio de tempo da intensidade da luz para o domínio das frequências.⁵⁴

A radiação transmitida ou refletida pela amostra é dispersa por uma grade de difração e direcionada para o chip MEMS, que contém um arranjo de refletores organizados em pixels.⁵⁴ Cada pixel é formado por elementos micro mecânicos que podem ser acionados em conjunto e mover-se quando submetidos a voltagem.^{54,55} Ao se aplicar uma determinada voltagem aos pixels, pode-se controlar quais pixels serão usados para dispersar a radiação e quais irão refletir na região espectral que lhes corresponde.⁵⁵ Assim, quando o pixel é acionado, ele se torna um elemento de difração programável, de forma que o chip MEMS combinado com a grade fixa de difração consistem em um refletor ótico com programação rápida e pixels de alto contraste.⁵⁴ A radiação refletida pelo MEMS é recombinada em um único feixe e direcionada ao detector InGaAs simples, que por constituir um único elemento, apresenta a vantagem de não haver variação entre elementos do detector e um custo reduzido em comparação com um detector de arranjo de diodos.⁵⁴

A programabilidade do sistema de conversão do espectro permite os seguintes modos: i) a simulação de um monocromador, quando os pixels são ativados um de cada vez; ii) simulação de infravermelho não dispersivo, onde uma sequência programada de filtros fixos possibilita observar linhas de absorção individualmente; e iii) transformada digital.⁵⁴ Além disso, o programa do equipamento possui uma ferramenta para análise de dados e construção de modelos de calibração multivariada por PLS.⁵⁴

No modo de transformada digital, o chip MEMS é programado de forma a ativar 50% dos pixels de cada vez criando “máscaras” espectrais. Uma matriz de transformação é construída pelo sequenciamento dessas máscaras em função do tempo. Em seguida é obtida uma transformação ao medir-se no detector a potência total para cada máscara. Essa transformação é multiplicada pela matriz de transformação para produzir o espectro completo.⁵⁵ Esse modo é o de maior eficiência para obtenção de um espectro completo nesse equipamento, uma vez que cada comprimento de onda atinge o detector em 50% do tempo, fornecendo uma melhora de 5 vezes na razão sinal-ruído com relação ao modo que simula um monocromador.⁵⁴

3.1.3 NIRS em análise de madeira

A espectroscopia no infravermelho próximo (NIRS) constitui uma excelente técnica para analisar madeira, uma vez que é relativamente simples, requer mínimo preparo de amostra, possibilita análises rápidas, não destrutivas e com boa reprodutibilidade e no caso dos equipamentos portáteis, possibilita análises em campo.^{25,57} A desvantagem principal da técnica é que a detecção para componentes minoritários é limitada a uma fração mínima de 0,1 e 0,5% em massa e depende dos grupos funcionais presentes, que determinam a intensidade da banda de absorção.⁵⁷

Medidas de NIRS em madeira podem ser coletadas nos modos de reflectância difusa ou transmitância, porém este último requer que a amostra seja moída e incorporada em pastilhas de KBr ou cortada em lâminas com microtomo.⁵⁷

A madeira é um material complexo composto majoritariamente pelos polímeros celulose, hemicelulose e lignina.²³ Uma revisão de diferentes aplicações de NIRS ao estudo de madeira foi feita por Tsuchikawa.²³ Essas aplicações incluem a determinação de propriedades físicas e químicas e a caracterização da madeira.

A frequência aproximada de absorção de grupos funcionais em moléculas simples pode ser calculada, entretanto em materiais complexos como a madeira, dadas as inúmeras interações, os valores calculados não correspondem à realidade.⁵⁷ As bandas na região do infravermelho são largas e apresentam sobreposição, de forma que é mais difícil estabelecer relação entre espectro e estrutura molecular.⁵⁷ Schwanninger et al.⁵⁷ realizaram uma extensa revisão de atribuição de bandas de absorção de NIRS em madeira e seus componentes (holocelulose, celulose, hemicelulose, lignina e extrativos).

A combinação com a quimiometria aumentou a aplicabilidade de NIRS para estudos da madeira, possibilitando a obtenção de mais informações a partir dos espectros. Alguns métodos quimiométricos serão detalhados na seção seguinte.

3.2 QUIMIOMETRIA

A Quimiometria pode ser definida como a aplicação de métodos estatísticos e matemáticos para auxiliar a interpretação e aquisição de dados químicos.⁵⁸

O termo quimiometria foi introduzido pelo químico sueco S. Wold em 1971.⁵⁸ Posteriormente ele colaborou com B.R. Kowalski, outro químico que vinha trabalhando em métodos matemáticos de identificação de padrões, e juntos fundaram a Sociedade Internacional de Quimiometria em 1974.⁵⁸

O crescimento da quimiometria ocorreu devido ao fato das análises univariadas não serem mais suficientes para descrever e modelar experimentos químicos, ao surgimento de instrumentos com respostas multivariadas para cada amostra analisada e à disponibilidade de computadores mais potentes.⁵⁹ O surgimento de programas computacionais com interface simplificada para realizar cálculos quimiométricos, que requerem do usuário menos conhecimentos em programação, estatística e matemática, também contribuiu para a difusão do uso da quimiometria.⁵⁸

Algumas das principais contribuições ou linhas de pesquisa da quimiometria consistem no desenvolvimento e na aplicação de técnicas de: i) planejamento de experimentos; ii) processamento de sinal; iii) resolução de curvas; iv) reconhecimento de padrões e classificação; e v) calibração multivariada.⁶⁰ Alguns modelos dessas duas últimas linhas de pesquisa foram aplicados nesta dissertação e serão abordados na próxima seção.

3.2.1 Calibração Multivariada

Calibração consiste em encontrar uma relação matemática entre a propriedade de interesse – geralmente a concentração de um analito – e as medidas instrumentais, para que seja possível prever a propriedade de interesse em uma nova amostra a partir de suas medidas experimentais.^{58,59,61,62,63}

Os dados experimentais obtidos para uma amostra podem ser univariados ou multivariados. Alguns instrumentos fornecem um único valor experimental para cada amostra e podem ser classificados como instrumentos de ordem zero, pois um número consiste em um escalar, que é um tensor de ordem zero.⁶³ Eletrodos íon-seletivos e fotômetros com filtro único são exemplos de instrumentos de ordem zero.⁶³ Dados de ordem zero são chamados univariados. Dados de primeira ordem em diante são chamados de multivariados.

Instrumentos de primeira ordem podem gerar várias medidas para uma única amostra em uma única leitura.⁶³ Exemplos de dados de primeira ordem são espectros, onde para cada amostra se obtêm valores de absorbância em diversos comprimentos de onda; e cromatogramas, onde para cada amostra se obtêm diferentes medidas do detector, que podem ser, por exemplo, a condutividade, em função do tempo. Esses dados podem ser organizados na forma de vetores, que consistem em tensores de primeira ordem.⁶³

Há também instrumentos de segunda ordem, que fornecem uma matriz de dados, que é um tensor de segunda ordem, para cada amostra.⁶³ Técnicas hífenadas como cromatografia acoplada a espectrometria de massas, ressonância magnética nuclear em duas dimensões são exemplos que utilizam dados de segunda ordem.⁶³

Não há limite máximo para a ordem dos dados que podem ser gerados.⁶³ Vantagens em se utilizar equipamentos que geram dados de maior ordem incluem o aumento da seletividade, menor necessidade de conhecimento prévio das espécies presentes na amostra, a possibilidade de análises em presença de interferentes e a possibilidade de se obter mais informações sobre o sistema em estudo a partir de cada medida; porém, a complexidade dos equipamentos e da matemática envolvida é proporcional à ordem dos dados.⁶³

Os modelos de calibração também podem ser classificados segundo a ordem. Essa classificação não necessariamente corresponde à ordem do equipamento utilizado, pois é possível selecionar variáveis para reduzir a ordem da calibração.⁶³ Pode-se, por exemplo, escolher apenas um comprimento de onda do espectro de cada amostra e utilizar uma calibração

univariada em vez de modelar o espectro inteiro com uma calibração de primeira ordem⁶³. Mas não é possível partir de dados de menor ordem para uma calibração de maior ordem.⁶³

Um dos métodos mais utilizados em calibração multivariada de primeira ordem é a regressão por mínimos quadrados parciais (PLSR, do inglês *partial least squares regression*).⁶⁶ Esse método, que será detalhado adiante, foi o método utilizado neste trabalho. Convém antes um breve esclarecimento acerca de análise de componentes principais (PCA, do inglês *principal components analysis*), que é a base para a PLSR.

3.2.1.1 PCA

O primeiro passo para realizar cálculos quimiométricos como dados de primeira ordem é organizá-los em uma matriz (\mathbf{X}), como mostra a Figura 7. Em \mathbf{X} , cada linha corresponde a uma amostra e cada coluna a uma variável. No caso da Figura 7, cada coluna corresponde a um número de onda do espectro, e os valores de absorvância de cada amostra nesse número de onda são dispostos ao longo da coluna.

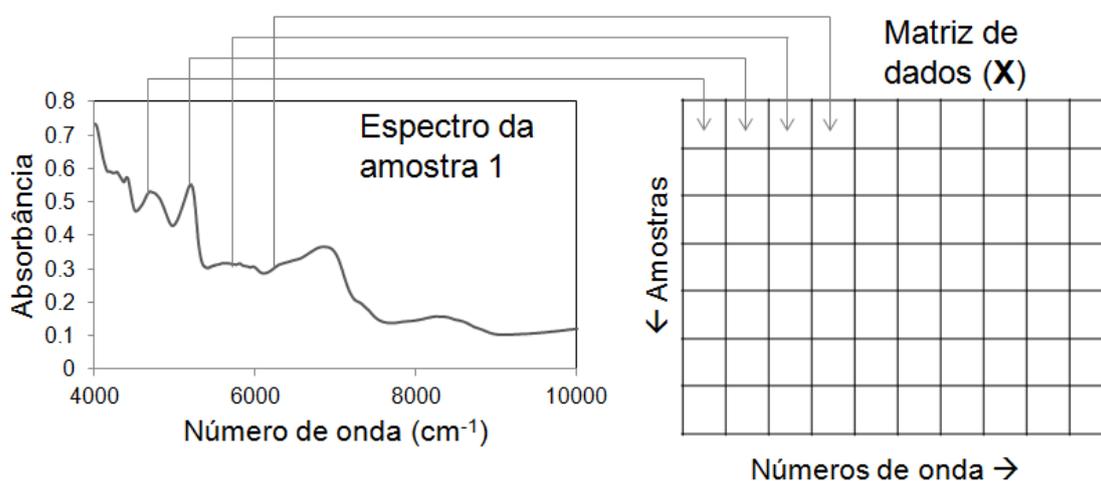


Figura 7. Organização da matriz de dados \mathbf{X} a partir de espectros de infravermelho.

O princípio inicial de PCA é que matrizes de dados contêm informações redundantes e podem ser reduzidas a termos de interpretação mais fácil.⁵⁹ Assim, o método procura combinações lineares dos valores de x , e as utiliza no

lugar das variáveis originais, descartando informações irrelevantes e instáveis de forma que apenas a parte mais relevante da variação em X é considerada.⁶⁴

Primeiramente, PCA decompõe a matriz X em duas matrizes menores:^{59, 60}

$$X = T \cdot P^T + E \quad \text{Equação 4}$$

sendo T a matriz de escores, P a matriz de pesos, P^T a transposta de P e E a matriz de erro. A matriz X é geralmente centrada na média ou escalada de alguma outra forma.⁵⁹ A Figura 8 ilustra essa decomposição, onde A representa o número de componentes principais.

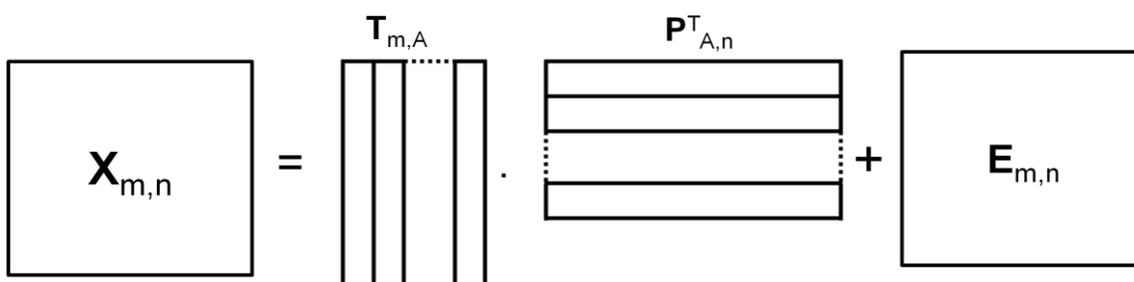


Figura 8. Decomposição da matriz de dados X por PCA.

PCA reduz as variáveis de X a um número menor de novas variáveis contendo as informações mais relevantes, que são os escores.⁶⁵ Esses escores são ponderados pelos pesos, que representam a relevância de cada variável.⁶⁵ A matriz E contém resíduos, ruídos e informações irrelevantes.

O número A de colunas na matriz de escores T e de linhas na matriz de pesos P^T corresponde ao número de fatores significativos nos dados.⁶⁶ Cada coluna de T e linha de P^T corresponde a uma componente principal.⁶⁶

O objetivo da PCA é explicar o máximo possível da variância de X com o mínimo de componentes principais.⁵⁹ Idealmente, o número de componentes principais deveria corresponder ao número de espécies químicas presentes na amostra.⁶⁰ Entretanto, isso pode ser afetado pela presença de ruído ou espécies em equilíbrio.⁶⁰ A escolha do número de componentes principais é muitas vezes subjetiva, e diversos métodos podem ser utilizados para realizar tal tarefa de forma mais objetiva, destacando-se a validação cruzada, que será discutida adiante.^{59,60}

3.2.1.2 PLSR

O método PLS foi desenvolvido por Hermann Wold por volta de 1975.⁶⁷ Nesse método, decompõem-se por PCA a matriz de resposta do equipamento \mathbf{X} e a matriz das propriedades de interesse \mathbf{Y} .⁶⁸

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad \text{Equação 5}$$

$$\mathbf{Y} = \mathbf{TQ}^T + \mathbf{F} \quad \text{Equação 6}$$

Onde \mathbf{T} é a matriz de escores, \mathbf{P} e \mathbf{Q} são as matrizes de pesos, \mathbf{P}^T e \mathbf{Q}^T suas transpostas e \mathbf{E} e \mathbf{F} são as matrizes de erros.

A equações matriciais 5 e 6 podem ser reescritas na forma de somatórias de produtos de vetores:

$$\mathbf{X} = \sum_{a=1}^A \mathbf{t}_a \mathbf{p}_a^T + \mathbf{E} \quad \text{Equação 7}$$

$$\mathbf{Y} = \sum_{a=1}^A \mathbf{t}_a \mathbf{q}_a^T + \mathbf{F} \quad \text{Equação 8}$$

Onde a representa o número da componente principal e A o número total de componentes principais.

A particularidade do PLS com relação a PCA está em que na decomposição de \mathbf{X} , a informação de \mathbf{Y} é incorporada de forma a modificar a componente principal para maximizar a covariância de \mathbf{X} e \mathbf{Y} .⁶⁴ As componentes principais em PLS recebem o nome de variáveis latentes (VL).

O nome mínimos quadrados parciais é devido ao fato de cada parâmetro do modelo ser estimado iterativamente como a inclinação de uma regressão bivariada simples pelo método dos mínimos quadrados e o termo “parciais” provém de um vetor ser considerado fixo a nova cada fase de iteração.⁶⁷

Um aspecto particularmente interessante do PLS é que considera erros tanto na variável independente quanto na variável dependente, enquanto métodos estatísticos mais tradicionais assumem que os erros estão apenas na variável dependente.⁶⁶

O método PLS possui duas abordagens diferentes. Uma chamada PLS1, onde há um vetor \mathbf{y} para cada propriedade de interesse – a concentração de

cada componente, por exemplo. Nesse caso, um modelo PLS distinto é desenvolvido para cada propriedade de interesse. Na outra abordagem, chamada PLS2, há uma matriz \mathbf{Y} com todas as propriedades de interesse. PLS2 é mais prático por possibilitar que todas as variáveis sejam modeladas conjuntamente e todas as propriedades de interesse sejam estimadas simultaneamente por um único modelo PLS. Contudo, embora PLS1 requeira que cada variável seja modelada separadamente, esse método permite otimizar os parâmetros individualmente para cada propriedade, possibilitando estimativas mais exatas e precisas. O método utilizado nessa dissertação foi o PLS1.

Existem diversos algoritmos para fazer a decomposição das matrizes \mathbf{X} e \mathbf{Y} . Dois algoritmos bastante utilizados são o NIPALS,^{64,67,68} acrônimo para mínimos quadrados parciais iterativo não linear (do inglês *Nonlinear Iterative Partial Least Squares*) e o SIMPLS,⁶⁹ acrônimo para modificação estatisticamente inspirada do PLS (do inglês *Statistically Inspired Modification of PLS*). Ambos foram utilizados no presente trabalho.

Os algoritmos NIPALS e SIMPLS fornecem estimativas idênticas para o PLS1.^{64,69} Já para o PLS2, as estimativas diferem, sendo que o SIMPLS fornece mais informações para esse método quanto a amostras anômalas e desvios da linearidade.⁶⁴

3.2.1.2.1 NIPALS

No algoritmo NIPALS, estima-se um vetor *peso*^a $\hat{\mathbf{w}}_1$, que se trata de um vetor unitário que determina a direção da primeira variável latente, obtido pelo critério da máxima covariância entre \mathbf{y} e funções lineares de \mathbf{X} , para o PLS1 ou entre funções lineares de \mathbf{X} e \mathbf{Y} , para o PLS2.⁶⁴ Nesta dissertação, o símbolo “^” sobre um termo indica que se trata de um termo estimado matematicamente. As matrizes \mathbf{X} e \mathbf{Y} devem estar centradas na média. Os escores nesse novo eixo são estimados como:⁶⁴

^a Esse *peso* (*loading weight* em inglês) não é o mesmo que peso em PCA (*loading* em inglês), embora a tradução seja a mesma.

$$\hat{\mathbf{t}}_1 = \mathbf{X}\hat{\mathbf{w}}_1 \quad \text{Equação 9}$$

Faz-se a regressão de todas as variáveis de \mathbf{X} e $\hat{\mathbf{t}}_1$ para obter-se o peso $\hat{\mathbf{p}}_1$. E faz-se a regressão de \mathbf{Y} em $\hat{\mathbf{t}}_1$ para obter o peso $\hat{\mathbf{q}}_1$.⁶⁴ Os produtos $\hat{\mathbf{t}}_1\hat{\mathbf{p}}_1$ e $\hat{\mathbf{t}}_1\hat{\mathbf{q}}_1$ são subtraídos de \mathbf{X} e \mathbf{Y} , respectivamente para encontrarem-se as matrizes de resíduos.⁶⁴ A próxima variável latente é obtida da mesma forma, utilizando, porém, as matrizes de resíduos em vez dos dados originais e o processo é repetido até que se tenha extraído o número desejado de variáveis latentes.⁶⁴

Os coeficientes de regressão para PLSR podem ser obtidos por:

$$\hat{\mathbf{B}} = \hat{\mathbf{W}}(\hat{\mathbf{P}}^T\hat{\mathbf{W}})^{-1}\hat{\mathbf{Q}} \quad \text{Equação 10}$$

Onde $\hat{\mathbf{W}}$ é a matriz de pesos.

As propriedades de interesse para uma amostra desconhecida podem ser obtidas por:

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}} \quad \text{Equação 11}$$

Um maior detalhamento matemático do NIPALS pode ser encontrado nas referências 67 e 68.

3.2.1.2.2 SIMPLS

No algoritmo SIMPLS, calcula-se uma matriz diferente de pesos \mathbf{R} diretamente sem decompor as matrizes \mathbf{X} e \mathbf{Y} . A discussão que segue acerca do algoritmo foi retirada da referência 69, onde seu desenvolvedor S. de Jong introduziu o SIMPLS.

Primeiramente, são calculados vetores peso \mathbf{r}_a e \mathbf{q}_a para a variando de 1 até o número máximo de variáveis latentes A :

$$\mathbf{t}_a = \mathbf{X}_0\mathbf{r}_a \quad (a = 1, 2, \dots, A) \quad \text{Equação 12}$$

$$\mathbf{u}_a = \mathbf{Y}_0\mathbf{q}_a \quad (a = 1, 2, \dots, A) \quad \text{Equação 13}$$

sendo \mathbf{X}_0 e \mathbf{Y}_0 as matrizes \mathbf{X} e \mathbf{Y} , respectivamente, centradas nas médias.

Os pesos são determinados pelo critério da máxima covariância entre \mathbf{t}_a e \mathbf{u}_a , sob as seguintes condições:

- 1) maximização da covariância: $\mathbf{u}_a^T \mathbf{t}_a = \mathbf{q}_a^T (\mathbf{Y}_0^T \mathbf{X}_0) \mathbf{r}_a = \text{máx}$
- 2) normalização dos pesos: $\mathbf{r}_a^T \mathbf{r}_a = 1$
- 3) normalização dos pesos: $\mathbf{q}_a^T \mathbf{q}_a = 1$
- 4) os escores \mathbf{t} devem ser ortogonais: $\mathbf{t}_b^T \mathbf{t}_a = 0$, para $a > b$

Com essa última condição, é necessário que

$$\mathbf{t}_b^T \mathbf{t}_a = \mathbf{t}_b^T \mathbf{X}_0 \mathbf{r}_a = (\mathbf{t}_b^T \mathbf{t}_b) \mathbf{p}_b^T \mathbf{r}_a = 0, \text{ para } a > b \quad \text{Equação 14}$$

onde \mathbf{p}_b é um vetor peso que expressa a relação entre as variáveis originais de \mathbf{X} e a b -ésima variável latente.

Segundo a equação 14, cada novo vetor peso \mathbf{r}_a ($a > 1$) deve ser ortogonal a todos os vetores peso que o antecedem (colunas de $\mathbf{P}_{a-1} \equiv [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{a-1}]$). Sendo \mathbf{P}_{a-1}^+ o projetor ortogonal requerido

$$\mathbf{P}_{a-1}^+ = \mathbf{I}_p - \mathbf{P}_{a-1} (\mathbf{P}_{a-1}^T \mathbf{P}_{a-1})^{-1} \mathbf{P}_{a-1}^T \quad \text{Equação 15}$$

onde \mathbf{I} é matriz identidade, exige-se que

$$\mathbf{r}_a = \mathbf{P}_{a-1}^+ \mathbf{r}_a, \text{ para } a > 1 \quad \text{Equação 16}$$

As equações 15 e 16 explicam concisamente a quarta condição de ortogonalidade. A solução para \mathbf{r}_a e \mathbf{q}_a é dada pelo primeiro par de vetores singulares obtido pela decomposição de valores singulares (SVD, do inglês *Singular Value Decomposition*) de \mathbf{S}_0 projetado em um subespaço ortogonal a \mathbf{P}_{a-1} , ou seja, a SVD de $\mathbf{P}_{a-1}^+ \mathbf{S}_0$, sendo $\mathbf{S}_0 = \mathbf{X}_0^T \mathbf{Y}_0$.

Pode-se determinar o produto cruzado \mathbf{S}_a após a projeção de a vetores peso por

$$\mathbf{S}_a \equiv \mathbf{P}_a^+ (\mathbf{X}_0^T \mathbf{Y}_0) = \mathbf{P}_a^+ \mathbf{S}_0, \text{ para } a \geq 1 \quad \text{Equação 17}$$

Para determinar \mathbf{S}_{a+1} a partir de \mathbf{S}_a , a projeção no espaço coluna de \mathbf{P}_a é feita como uma sequência de projeções ortogonais usando uma base

ortonormal de \mathbf{P}_a , $\mathbf{V}_a \equiv [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_a]$. \mathbf{V}_a pode ser obtida por uma ortonormalização de Gram-Schmidt de \mathbf{P}_a

$$\mathbf{v}_a \propto \mathbf{p}_a - \mathbf{V}_{a-1}(\mathbf{V}_{a-1}^T \mathbf{p}_a), \quad a = 2, 3, \dots, A \quad \text{Equação 18}$$

começando por $\mathbf{V}_1 = \mathbf{v}_1 \propto \mathbf{p}_1$, onde \propto indica proporcionalidade e normalização subsequente do vetor resultante.

No caso do PLS1, é possível fazer uma simplificação utilizando propriedades da ortogonalidade de \mathbf{P} segundo as quais $\mathbf{p}_b^T \mathbf{p}_a = 0$ para $b \leq a - 2$. Essas propriedades se estendem para os pesos ortonormalizados \mathbf{V} , de modo que $\mathbf{p}_b^T \mathbf{v}_a = 0$ para $b \leq a - 2$. Assim, a ortogonalidade de \mathbf{p}_a em relação a \mathbf{V}_{a-2} está garantida e a equação 18 pode ser simplificada para

$$\mathbf{v}_a \propto \mathbf{p}_a - \mathbf{v}_{a-1}(\mathbf{v}_{a-1}^T \mathbf{p}_a), \quad a > 1 \quad \text{Equação 19}$$

Assim, pela ortonormalidade de \mathbf{V} , as matrizes produto \mathbf{S}_a são obtidas projetando-se as direções perpendiculares \mathbf{v}_a :

$$\mathbf{S}_a = \mathbf{S}_{a-1} - \mathbf{v}_a(\mathbf{v}_a^T \mathbf{S}_{a-1}), \quad a > 1 \quad \text{Equação 20}$$

Diferentemente do NIPALS, o SIMPLS decompõe o produto cruzado em vez das matrizes originais, que são maiores, o que o torna mais rápido.

Calcula-se \mathbf{q}_a como o autovetor dominante da matriz $\mathbf{S}_{a-1}^T \mathbf{S}_{a-1}$ e encontra-se \mathbf{r}_a

$$\mathbf{r}_a \propto \mathbf{S}_{a-1} \mathbf{q}_a \quad \text{Equação 21}$$

No caso do PLS1 (onde $\mathbf{Y} = \mathbf{y}$), \mathbf{S}_{a-1} ($= \mathbf{s}_{a-1}$) é um vetor de covariâncias de \mathbf{y} com as variáveis de \mathbf{X} , portanto $\mathbf{S}_{a-1}^T \mathbf{S}_{a-1} = \mathbf{s}_{a-1}^T \mathbf{s}_{a-1}$ é escalar, $\mathbf{q}_a = 1$ e $\mathbf{r}_a \propto \mathbf{s}_{a-1}$.

No desenvolvimento da teoria do SIMPLS optou-se convenientemente pela normalização dos vetores peso \mathbf{r}_a . Entretanto essa escolha não é essencial, podendo-se adotar a normalização dos escores \mathbf{t}_a . Redefine-se então $\mathbf{r}_a = \mathbf{r}_a / |\mathbf{X}_0 \mathbf{r}_a|$ e $\mathbf{t}_a = \mathbf{t}_a / |\mathbf{t}_a|$, resultando em vetores de escores \mathbf{t}_a unitários e \mathbf{T} ortonormal tal que $\mathbf{T}^T \mathbf{T} = \mathbf{I}_A$.

As estimativas das propriedades de interesse para as amostras de calibração são obtidas por

$$\hat{Y}_0 = \mathbf{T}\mathbf{T}^T\mathbf{Y}_0 = \mathbf{X}_0\mathbf{R}\mathbf{R}^T\mathbf{X}_0^T\mathbf{Y}_0 = \mathbf{X}_0\mathbf{R}\mathbf{R}^T\mathbf{S}_0 \quad \text{Equação 22}$$

E o coeficiente de regressão é obtido por

$$\hat{\mathbf{B}} = \mathbf{R}(\mathbf{R}^T\mathbf{S}_0) = \mathbf{R}(\mathbf{T}^T\mathbf{Y}_0) = \mathbf{R}(\mathbf{T}^T\mathbf{Y}) = \mathbf{R}\mathbf{Q}^T \quad \text{Equação 23}$$

onde a matriz de pesos de \mathbf{Y} não normalizada $\mathbf{Q} \equiv \mathbf{Y}_0^T\mathbf{T}$.

As propriedades de interesse para uma amostra desconhecida podem ser obtidas por:

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}} \quad \text{Equação 24}$$

3.2.1.2.3 PLS-DA

A análise discriminante (DA, do inglês *discriminant analysis*), também chamada de classificação supervisionada, tem como objetivo criar um método para classificar amostras em grupos pré-definidos.⁶⁴ Ela difere dos métodos de classificação não supervisionada, também chamada de análise de agrupamentos (em inglês, *clusters*), onde não há informações prévias quanto a possíveis agrupamentos das amostras e o objetivo é encontrar ou identificar tendências nas amostras de se separarem em *clusters*, indicando a existência de subgrupos.⁶⁴

Uma forma de realizar uma DA é considerá-la uma calibração qualitativa, onde a propriedade de interesse a ser determinada consiste em se amostra pertence ou não a um determinado grupo.⁶⁴

A PLSR pode ser aplicada com esse propósito. Duas abordagens diferentes são possíveis.⁶⁴ Na primeira, cada classe é discriminada separadamente com relação a todas as outras, de forma que para cada classe há um vetor \mathbf{y} binário onde consta um valor para as amostras de calibração que pertencem à classe que está sendo modelada – por exemplo 1 – e outro valor para todas as outras amostras – por exemplo zero. Esse método, chamado

PLS1-DA, onde se constrói um modelo PLS1 para cada classe, está ilustrado na Figura 9.

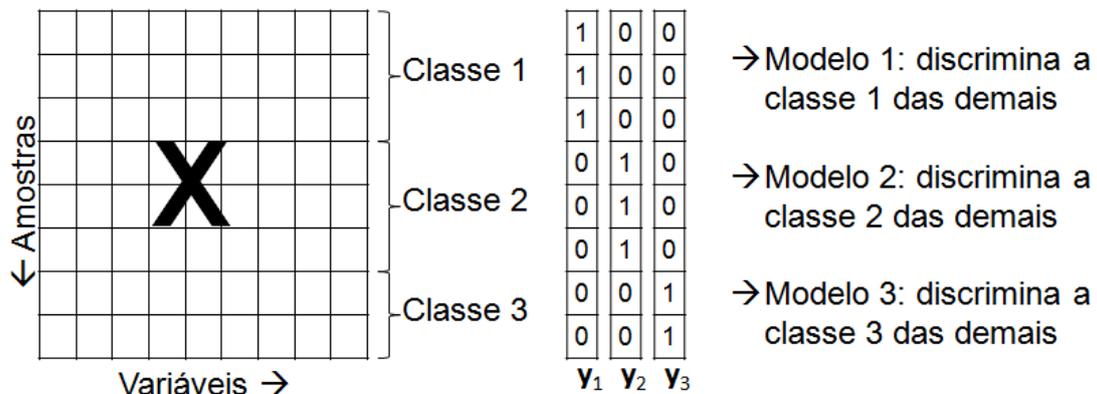


Figura 9. Representação do PLS1-DA

A segunda abordagem é chamada PLS2-DA e utiliza uma matriz Y com as informações de classes e todas as classes são modeladas conjuntamente, resultando em um único modelo, conforme ilustrado na Figura 10.

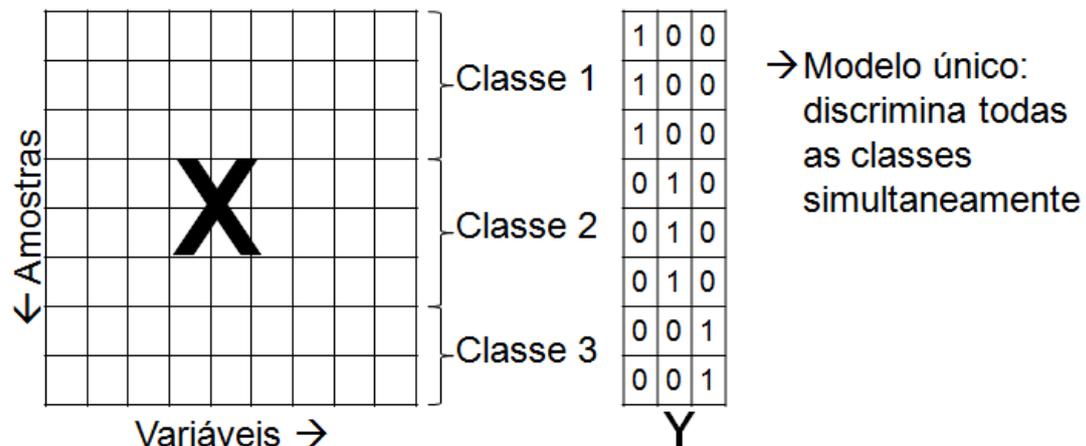


Figura 10. Representação do PLS2-DA

As vantagens e desvantagens de cada abordagem são as mesmas com relação ao PLS1 e PLS2: Embora PLS2-DA seja mais prática, PLS1-DA tende a fornecer melhores resultados, uma vez que os parâmetros como número de VL podem ser otimizados separadamente para o modelo de discriminação de cada classe. O método utilizado no presente trabalho foi o PLS1-DA.

3.2.1.2.4 Validação e validação cruzada

A validação de um modelo de calibração consiste em aplicá-lo a amostras que não pertencem ao conjunto de calibração e cujos valores das propriedades de interesse sejam conhecidos para avaliar sua capacidade de estimar a propriedade de interesse, comparando os valores estimados com os valores conhecidos.⁶¹ Nessa etapa é possível estimar erros e otimizar parâmetros do modelo para aumentar sua eficiência.

Esse processo pode ser feito pela aplicação do modelo a um conjunto de amostras independente. Os erros – no caso, a raiz quadrada do erro médio quadrático de validação (RMSEV, do inglês *root mean squared error of validation*) – para a previsão da propriedade de interesse nessas amostras é determinado utilizando os valores estimados e os valores conhecidos. Escolhe-se o menor número de PC ou VL que apresenta um valor mínimo de RMSEV, otimizam-se os outros parâmetros do modelo (como pré-processamentos e seleção de variáveis de x) e identificam-se e excluem-se amostras anômalas.

Outra forma de realizar a validação de um modelo é através da validação cruzada (CV, do inglês *cross-validation*). Esse é um dos métodos mais utilizados para escolha do número de VL dos modelos PLS e foi o método de validação utilizado no presente trabalho.

Na CV, há um revezamento entre as amostras de calibração de forma que algumas amostras são removidas do conjunto de calibração e utilizadas para validação enquanto as remanescentes são usadas para construção do modelo e ocorre uma rotatividade até que todas as amostras tenham sido utilizadas nas duas situações.

Uma possibilidade para realizar a CV é remover as amostras do conjunto de calibração uma a uma para serem previstas pelo modelo construído utilizando as demais, na chamada CV completa ou “deixa uma de fora” (do inglês *leave one out*).⁶⁴ Isso implica construir tantos modelos quanto o número de amostras de calibração.⁶⁶ Esse foi o método utilizado neste trabalho.

Outra alternativa é remover as amostras em grupos, contudo embora o processamento computacional demore menos dessa forma, os conjuntos de

calibração para construção de cada modelo ficam reduzidos aumentando a possibilidade de ocorrerem erros tendenciosos.⁶⁴ Quando há replicatas nos dados, dependendo do caso, é recomendado agrupar as replicatas de cada amostra para serem removidas conjuntamente, uma vez que uma CV *leave one out* poderia induzir a resultados erroneamente otimistas.⁶⁴

O erro estimado para a CV consiste na raiz quadrada do erro médio quadrático de validação cruzada (RMSECV, do inglês *root mean squared error of cross-validation*):^{60,64,66}

$$RMSECV = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}} \quad \text{Equação 25}$$

onde \hat{y}_i é o vetor das propriedades de interesse estimadas para a i -ésima amostra a partir do modelo construído com as demais amostras do conjunto de calibração, y_i é o vetor dos valores de referência das propriedades de interesse para a i -ésima amostra e N é o número total de amostras de calibração.

Os modelos são construídos com número de VL variando de 1 a A e um gráfico de RMSECV versus número de VL como o da Figura 11 pode ser utilizado para escolher o número ótimo de VL.

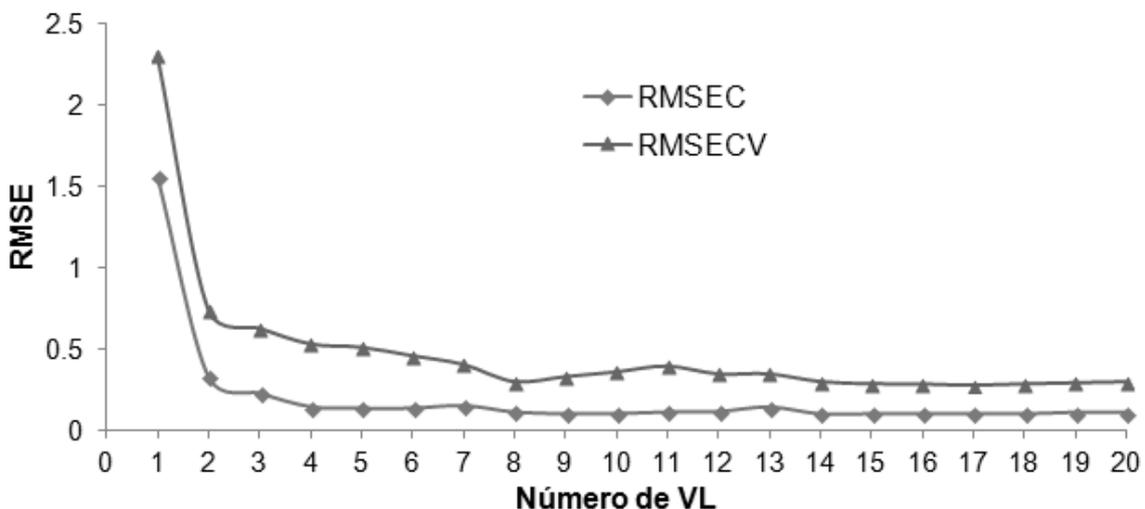


Figura 11. Exemplo de gráfico de RMSECV por número de VL

Escolhe-se o número de VL para o qual o RMSECV mínimo. No caso da Figura 11, o número de VL escolhido poderia ser 8. Entretanto, é possível que o comportamento do gráfico seja imprevisível, levando a uma escolha mais subjetiva.

O RMSEC corresponde à raiz quadrada do erro quadrático médio de calibração (do inglês, *root mean squared error of calibration*) que é calculado pela previsão das amostras de calibração por um modelo construído com essas mesmas amostras.

Se o gráfico do RMSECV estiver muito acima do gráfico do RMSEC, provavelmente há um sobreajuste do modelo. Isso significa que o modelo funciona muito bem para as amostras de calibração, mas não funciona bem para prever outras amostras. O sobreajuste é uma indicação de que o conjunto de calibração não é representativo do universo de amostras que se pretende analisar utilizando o modelo desenvolvido.

Depois da validação, o modelo resultante pode ser aplicado a outro conjunto independente de amostras – as amostras de previsão ou teste – cujos valores das propriedades de interesse também são conhecidos, para estimar a capacidade real de previsão do modelo, por meio da raiz quadrada do erro quadrático médio de previsão (RMSEP, do inglês, *root mean squared error of prediction*), que é um valor definitivo, uma vez que não devem ser feitas novas mudanças no modelo. Depois disso o modelo pode ser aplicado para prever amostras reais.

3.2.1.2.5 Detecção de amostras anômalas

Amostras anômalas ou, em inglês, *outliers* podem ser definidas como amostras que apresentam um comportamento diferente do restante dos dados.⁶⁴ A presença de *outliers* no conjunto de calibração pode influenciar os coeficientes de regressão prejudicando o desenvolvimento do modelo.⁶⁴ Já no conjunto de previsão, *outliers* não têm influência no modelo, entretanto as estimativas de y para essas amostras serão errôneas e, conseqüentemente os erros de previsão do modelo também serão.⁶⁴

Diversos métodos para detecção de *outliers* foram descritos na literatura.^{64,70,71,72,73}

Neste trabalho foram utilizados dois métodos para detectar *outliers*. O primeiro foi baseado no método descrito na norma da ASTM E1655-05,⁷⁰ e sugerido por Valderrama et al.⁷³ Esse método considera como outliers amostras com alta influência, resíduos não modelados em \mathbf{X} e, ainda, os resíduos não modelados em \mathbf{y} , utilizando intervalo de confiança de 99%.

A influência h_i , ou peso relativo ou ainda, em inglês, *leverage*, é uma medida do potencial que uma determinada amostra tem de influenciar no modelo.⁶⁴ Ela pode ser calculada da seguinte forma:

$$h_i = \sum_{a=1}^A \frac{t_{i,a}^2}{\hat{\mathbf{t}}_a^T \hat{\mathbf{t}}_a} \quad \text{Equação 26}$$

onde \mathbf{t}_a são os escores da calibração, $t_{i,a}$ é o escore da amostra i e a é o número da VL. Quanto maior for a influência de uma amostra, mais afastada ela estará do centro do conjunto de amostras.

A influência média \bar{h} pode ser calculada como

$$\bar{h} = \frac{A}{n_{\text{cal}}} \quad \text{Equação 27}$$

onde A é o número de VL do modelo e n_{cal} o número de amostras de calibração e representa o valor da influência que cada amostra teria se todas contribuíssem igualmente para a definição do modelo. De acordo com a norma E1655-05 da ASTM,⁷⁰ amostras onde $h_i > 3\bar{h}$ são consideradas *outliers*.

Os resíduos não modelados $\hat{\mathbf{e}}_i$ em \mathbf{X} para cada amostra são definidos por

$$\hat{\mathbf{e}}_i^T = \mathbf{x}_i^T - \hat{\mathbf{x}}_i^T = \mathbf{x}_i^T - \hat{\mathbf{t}}_i^T \hat{\mathbf{P}}^T \quad \text{Equação 28}$$

onde $\hat{\mathbf{t}}_i$ é o vetor de escores para a amostra i , $\hat{\mathbf{P}}$ é a matriz de pesos de \mathbf{X} e \mathbf{x}_i é o vetor de respostas instrumentais centrado na média para a amostra i .⁶⁴

Altos valores de resíduos indicam que a amostra contém constituintes diferentes e inesperados com relação às demais ou podem ser causados por ruídos instrumentais.⁶⁴ A soma quadrática residual de uma amostra pode ser comparada à soma quadrática residual total das amostras por um teste t e,

caso a primeira seja estatisticamente maior que a segunda, a amostra é considerada um *outlier*.

De forma semelhante, os resíduos não modelados em \mathbf{y} são dados pela diferença entre o valor esperado (ou de referência) e o valor calculado pelo modelo.

O outro método utilizado seguiu os critérios do programa OPUS 6,5 da Bruker Optics. Este utiliza também a influência e o resíduos não modelados em \mathbf{X} e \mathbf{y} . A amostra é considerada como *outlier* de acordo com a influência quando $h_i > 5\bar{h}$. Os resíduos não modelados em \mathbf{X} e \mathbf{y} são comparados por um teste F.

3.2.1.2.6 Pré-processamentos

Antes da regressão multivariada, é comum aplicar algum pré-processamento nos dados.⁶¹ Os pré-processamentos minimizam as variações nos dados que não estão relacionadas com a propriedade de interesse.⁷⁴ Alguns pré-processamentos utilizados nesta dissertação foram:

- **Centrar na média:** Calcula a média dos sinais de todas as amostras em uma variável fixa (um comprimento de onda, por exemplo), ou seja, a média das colunas de \mathbf{X} ou \mathbf{Y} . Subtrai essa média dos espectros de cada amostra em cada variável correspondente. Essa operação pode tornar o modelo mais robusto e reduzir um fator (VL) em seu tamanho.⁶¹
- **Subtração de linha reta** (em inglês, *straight line subtraction*): Ajusta uma linha reta ao espectro e a subtrai deste, corrigindo inclinações e desvios de linha de base.⁷⁵
- **Primeira Derivada:** Calcula a derivada do espectro por interpolação, utilizando uma janela de pontos de tamanho selecionável para alisamento da curva ajustada.⁷⁵ Esse processamento acentua os picos e os ruídos e corrige desvios aditivos de linha de base.⁷⁵

3.2.2 TRANSFERÊNCIA DE CALIBRAÇÃO

Frequentemente a construção de modelos de calibração multivariada requer a utilização de grande quantidade de amostras, algumas vezes coletadas ao longo de grandes períodos de tempo.⁷⁴ Diferente da calibração univariada, onde os instrumentos são recalibrados com frequência, é interessante que a validade dos modelos de calibração multivariada seja estendida, uma vez que o seu desenvolvimento exige maior esforço e tempo.⁷⁴

Existem basicamente quatro situações em que um modelo multivariado pode perder a capacidade de estimar a propriedade de interesse com a mesma eficiência de quando foi validado: i) quando se deseja aplicá-lo a medidas feitas em outro equipamento; ii) quando trocam-se componentes do equipamento no qual o modelo foi desenvolvido; iii) quando variam as condições em que as medidas são feitas em relação às condições iniciais; e iv) quando as amostras sofrem alterações.⁷⁶

Em alguns casos, situações como essas são inevitáveis. Em aplicações industriais, por exemplo, existem bancos de dados de milhares de amostras e, quando são adquiridos novos equipamentos, é desejável aproveitar esses dados em vez de reproduzir as medidas e calibrá-los individualmente.⁵⁶ Programas de monitoramento ambiental de longo prazo são outro exemplo, pois uma grande quantidade de amostras precisa ser medida ao longo de grandes períodos de tempo e os dados estão sujeitos a muitas variações.⁷⁷ Outro caso seria a obsolescência ou o término da vida útil do equipamento ou de algum de seus componentes.

A transferência de calibração é uma alternativa para evitar a necessidade de construir um novo modelo de calibração – a chamada recalibração – em situações como as supracitadas, poupando tempo e despesas.^{56,74,76,77,78,79,80}

Pode-se definir transferência de calibração como uma estratégia para que um modelo de calibração construído com respostas instrumentais obtidas em determinadas condições possa ser aplicado a respostas obtidas em condições diferentes. Essas mudanças de condições são as situações anteriormente citadas onde um modelo de calibração perde a sua validade.

Quando a transferência envolve equipamentos diferentes, o equipamento onde o modelo foi originalmente construído é comumente chamado de *primário* ou *mestre* e o equipamento para o qual o modelo é transferido é chamado de *secundário* ou *escravo*.

Neste trabalho adotaram-se os termos *primário* (P) e *secundário* (S). As situações exemplificadas para equipamentos diferentes estendem-se para as demais situações citadas onde cabe transferência de calibração, onde “equipamento primário” representa as condições iniciais de análise e “equipamento secundário” representa as novas condições.

Alguns dos principais métodos de transferência estão esquematizados na Figura 12 e serão discutidos a seguir, enfatizando-se aqueles utilizados neste trabalho: padronizações direta e direta por partes, reversa e reversa por partes e diferenças espectrais por comprimento de onda.

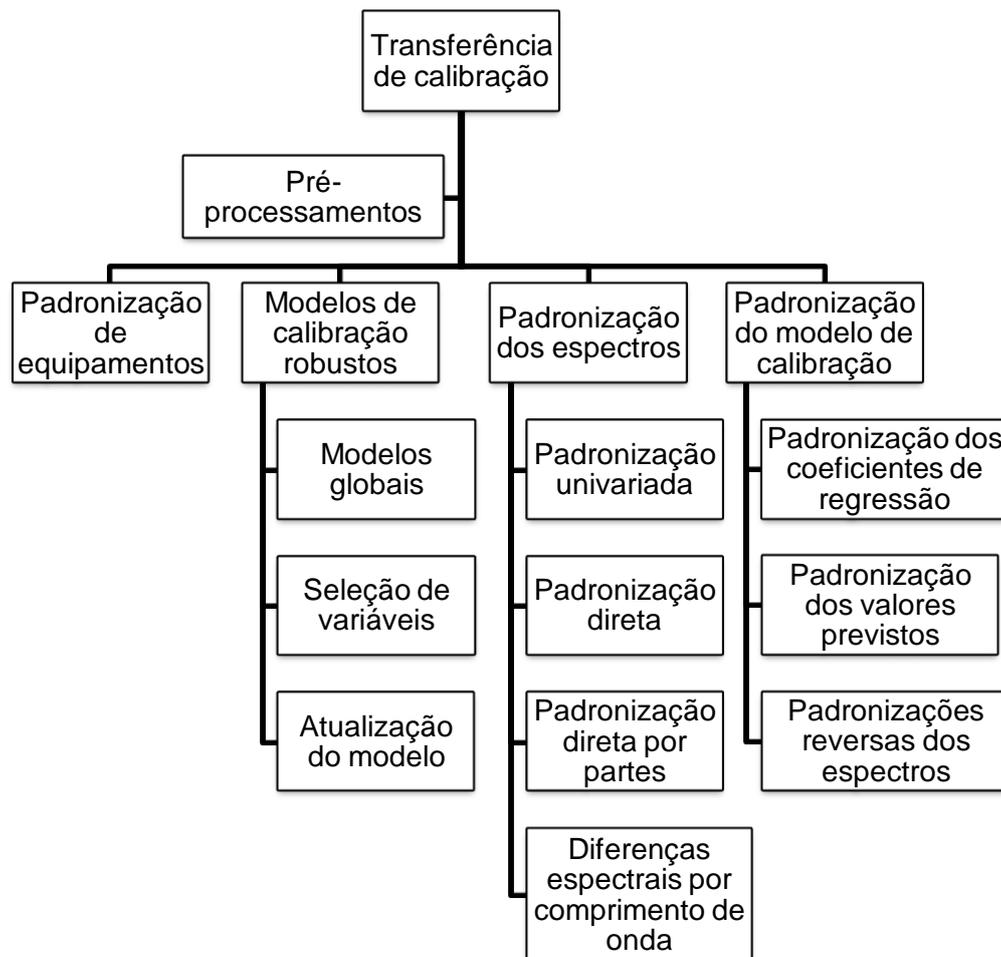


Figura 12. Esquema das estratégias de transferência de calibração

3.2.2.1 PRÉ-PROCESSAMENTOS PARA TRANSFERÊNCIA DE CALIBRAÇÃO

A utilização de pré-processamentos minimiza variações de intensidade de sinal que não estão relacionadas com a propriedade de interesse.⁷⁴ Essa correção por si só pode ser suficiente para amenizar as diferenças das respostas instrumentais obtidas nas distintas condições de análise a ponto de não ser necessário utilizar outra estratégia de transferência.^{74,77} Frequentemente o uso de pré-processamentos é combinado a outros métodos de transferência.⁷⁷

Alguns dos pré-processamentos mais utilizados são primeira e segunda derivadas, variação de padrão normal (SNV, do inglês *standard normal*

variation), correção de espalhamento multiplicativo de sinal (MSC, do inglês *multiplicative signal correction*), filtros de resposta finita a impulso (FIR, do inglês *finite impulse response*), correção ortogonal de sinal (OSC, do inglês *orthogonal signal correction*), transformada de Fourier e transformada Wavelet.^{77,78,79}

3.2.2.2 PADRONIZAÇÃO DO EQUIPAMENTO

Difícilmente equipamentos diferentes fornecem respostas idênticas e podem ser modelados conjuntamente sem necessidade de padronização, mesmo entre equipamentos com os mesmos modelo e fabricante, e isso frequentemente implica perda na eficiência da calibração.^{64,74}

A padronização de equipamentos – ou padronização de *hardware* – consiste em configurar todos os componentes mecânicos dos equipamentos de forma que eles forneçam medidas semelhantes o suficiente para serem modeladas conjuntamente.⁷⁴ Esse método está esquematizado na Figura 13.

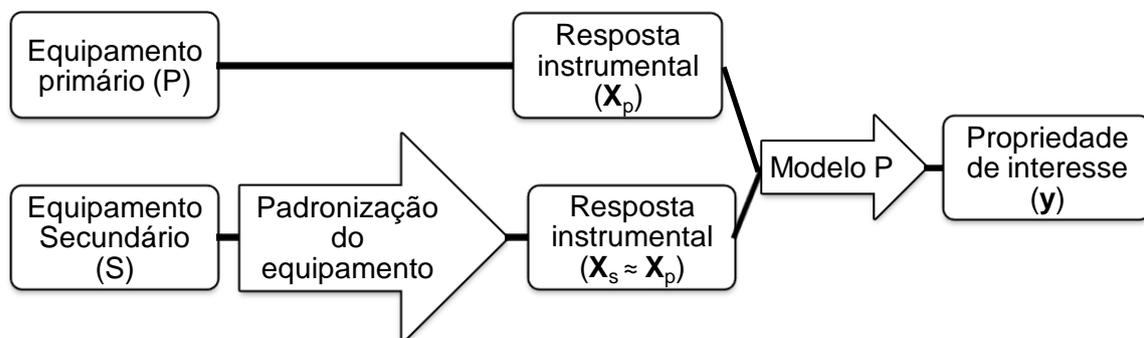


Figura 13. Esquema da padronização de equipamentos. Adaptado da referência 74.

3.2.2.3 MODELOS DE CALIBRAÇÃO ROBUSTOS

Modelos de calibração robustos são modelos que mantêm sua validade diante da variação de condições e, em alguns casos, de equipamentos.⁷⁴ Nesse contexto, a robustez significa que o erro de previsão do modelo mantém-se dentro de limites aceitáveis mesmo diante dessas variações.⁸¹ A Figura 14 mostra um esquema da construção e utilização de modelos robustos.

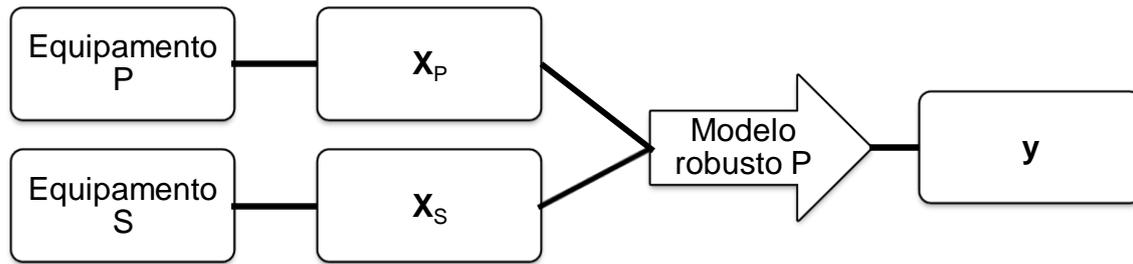


Figura 14. Esquema para modelos robustos. Adaptado da referência 74.

Serão comentadas a seguir algumas formas comuns para a construção de modelos robustos: a construção de modelos globais, a seleção de variáveis e a atualização do modelo.

3.2.2.3.1 Modelos globais

Um modelo de calibração global é um modelo com aplicabilidade tão ampla quanto possível. Para construir um modelo global, é preciso incluir todas as fontes de variação relevantes no modelo, para que este não seja sensível a elas.^{74,79}

Isso implica realizar as medidas das amostras de calibração nas diferentes condições que se deseja incluir no modelo, tais como equipamentos, intervalos de tempo e temperaturas diferentes, por exemplo, e utilizar todas essas medidas na construção do modelo.⁷⁴ Deve-se levar em conta a variação de uma gama de condições possíveis, para que o modelo resista a essas variações futuramente.⁷⁴ A realização de um planejamento fatorial pode auxiliar na identificação das fontes de variações relevantes e para desenvolver o modelo global.

Considerando a inclusão de tantas variáveis, modelos amplos tendem a ser menos precisos que modelos locais.⁷⁹ Outra limitação é a dificuldade de prever todas as fontes de variação significativas.⁷⁴ Além disso, a mudança de equipamento é uma das fontes de variação mais difíceis de incorporar ao modelo sem perda de eficiência e pode ocorrer que as respostas instrumentais

se agrupem em *clusters*, indicando que as medidas nas duas condições são muito distintas.^{74,79,80}

3.2.2.3.2 Seleção de variáveis

É possível aumentar a robustez e melhorar a capacidade de previsão de um modelo de calibração selecionando-se variáveis menos sensíveis a variações das condições experimentais.^{79,81} Existem diversos algoritmos que podem ser aplicados com essa finalidade, tais como o de projeções sucessivas e o de Kennard-Stone.⁷⁹

Através da seleção de variáveis pode-se, por exemplo, selecionar regiões do espectro onde há mais informação relacionada à propriedade de interesse e menos variações relacionadas às diferenças de condições de análise e utilizar apenas esses comprimentos de onda selecionados na construção de um modelo robusto.⁷⁹

3.2.2.3.3 Atualização do modelo

Quando a diferença entre os equipamentos não é muito grande, pode-se incluir a variação entre estes no modelo acrescentando algumas medidas de amostras do equipamento secundário ao conjunto de calibração do equipamento primário para construir um novo modelo.⁸⁰ Esse processo é chamado de atualização do modelo de calibração e está esquematizado na Figura 15.

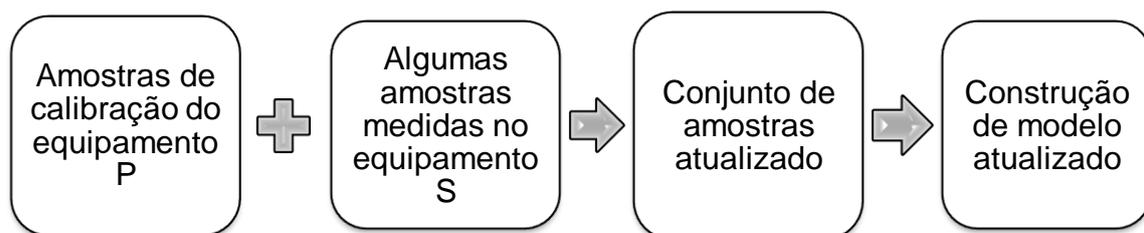


Figura 15. Esquema da atualização de modelo de calibração.

3.2.2.4 PADRONIZAÇÃO DOS ESPECTROS

A padronização dos espectros – ou padronização das respostas instrumentais – consiste em transformar as respostas do equipamento secundário para que se pareçam com as do primário, de forma que o modelo de calibração desenvolvido no primário possa ser aplicado inalterado às medidas transformadas do secundário.^{74,78,79,80,82} A Figura 16 esquematiza a padronização dos espectros.

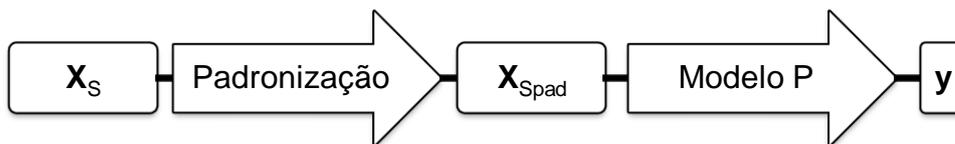


Figura 16. Esquema da padronização de espectros. Adaptado da referência 74.

Esses métodos requerem que amostras de transferência sejam medidas em todos os equipamentos envolvidos. As amostras de transferência devem ser um subconjunto das amostras de calibração, sendo que o número de amostras de transferência varia de acordo com a complexidade do sistema em estudo. Isso pode representar uma limitação, pois as amostras devem ser as mesmas e estar nas mesmas condições quando são medidas nos diferentes equipamentos.⁷⁸

Uma etapa inicial para a padronização de espectros é a compatibilização das regiões espectrais. É necessário que os comprimentos de onda das respostas instrumentais sejam correspondentes, o que pode ser alcançado por meio de correlações ou interpolações.

3.2.2.4.1 Padronização univariada

Nesse método proposto e patenteado Shenk e Westenhaus,⁸³ primeiramente, amostras de transferência são medidas nos equipamentos primário e secundário.

Para correção na escala de comprimento de onda, calcula-se a primeira derivada desses espectros e a cada ponto ($x_{p,i}$) dos espectros do equipamento primário é correlacionada uma janela de pelo menos 5 pontos dos espectros do

equipamento secundário em torno do comprimento de onda de $\mathbf{x}_{p,i}$ ($\mathbf{x}_{s,i-j}$, $\mathbf{x}_{s,i-j+1}$, ... , $\mathbf{x}_{s,i+k-1}$, $\mathbf{x}_{s,i+k}$), onde \mathbf{x} é um vetor contendo as respostas instrumentais para cada amostra num comprimento de onda fixo.^{78,82,83}

Os coeficientes de correlação são ajustados em uma função quadrática por mínimos quadrados, e o ponto máximo da curva corresponde ao ponto do espectro secundário $\mathbf{x}_{s,l}$ de maior correlação com o ponto $\mathbf{x}_{p,i}$ do primário.⁸³

Os respectivos comprimentos de onda λ_l e λ_i desses pontos são ajustados por regressão linear simples,⁸³

$$\lambda_l = a + b \lambda_i \quad \text{Equação 29}$$

onde os coeficientes a e b estabelecem o deslocamento de comprimento de onda entre os dois equipamentos. Provavelmente não há um ponto específico no espectro do equipamento secundário exatamente correspondente a λ_l , então $\mathbf{x}_{s,l}$ é calculado por interpolação linear.^{82,83}

Em seguida, calculam-se os coeficientes para correção na escala de absorbância por regressão linear simples

$$\mathbf{x}_{p,i} = c + d \mathbf{x}_{s,l} \quad \text{Equação 30}$$

onde os coeficientes a e b são estimados por regressão linear simples. Extremidades são calculadas a partir dos 4 pontos mais próximos por regressão linear múltipla.^{74,82,83}

Algumas variações desse método foram propostas na literatura.⁷⁸ A padronização univariada corrige deslocamentos simples nos comprimento de onda e diferenças lineares na intensidade, não corrigindo, interações entre as duas variáveis, como alargamento de bandas, que requerem abordagens multivariadas.^{74,82}

3.2.2.4.2 Padronização direta

A padronização direta (DS, do inglês *Direct Standardization*) foi desenvolvida por Y. Wang et al.⁸². Nesse método, as matrizes das respostas

das \mathbf{X}_P e \mathbf{X}_S dos equipamentos primário e secundário, respectivamente, são relacionadas por uma matriz de transformação \mathbf{F} :^{74,82}

$$\mathbf{X}_P = \mathbf{X}_S \mathbf{F} + \mathbf{E} \quad \text{Equação 31}$$

sendo \mathbf{F} quadrada com dimensão igual ao número de comprimentos de onda do espectro e \mathbf{E} a matriz de resíduos.

\mathbf{F} é calculada a partir das medidas das amostras de transferência nos dois equipamentos \mathbf{X}_{Pt} e \mathbf{X}_{St} :^{74,82}

$$\hat{\mathbf{F}} = \mathbf{X}_{St}^+ \mathbf{X}_{Pt} \quad \text{Equação 32}$$

onde \mathbf{X}_{St}^+ é a pseudo-inversa de \mathbf{X}_{St} . $\hat{\mathbf{F}}$ pode ser estimada por uma regressão multivariada como PLS2R e, em seguida, utilizada para transformar cada espectro \mathbf{x}_s do equipamento secundário em um espectro \mathbf{x}_{SPad} compatível com o equipamento primário:⁶⁴

$$\mathbf{x}_{SPad}^T = \mathbf{x}_s^T \hat{\mathbf{F}} \quad \text{Equação 33}$$

O modelo de calibração desenvolvido para o equipamento primário pode então ser aplicado aos espectros transformados \mathbf{x}_{SPad} para estimar a propriedade de interesse.

Frequentemente o número de amostras de transferência é inferior ao de variáveis e \mathbf{X}_S é retangular e não inversível.^{79,84} \mathbf{F} pode ser então determinada projetando-se as matrizes de dados instrumentais do conjunto de calibração (\mathbf{X}_{Pcal}) do equipamento primário e \mathbf{X}_{Pt} e \mathbf{X}_{St} no espaço coluna de \mathbf{X}_{cal} , calculado por SVD.⁸² Os escores obtidos pela projeção são utilizados no lugar das matrizes originais para a padronização e calibração.⁸² Essa projeção tem um efeito de filtro de ruído, mas também pode ocasionar perda de informação relevante caso os primeiros componentes para os espectros dos dois equipamentos descrevam fenômenos diferentes.⁷⁴

Originalmente a padronização direta e a padronização direta por partes, que será tratada adiante, são correções multiplicativas. Z. Wang et al.⁸⁵ propuseram a incorporação de um intercepto para correções aditivas e

obtiveram uma melhora nos resultados em relação à correção meramente multiplicativa.

Na DS, o número de amostras de transferência deve ser idealmente maior que o posto – o número de linhas ou colunas linearmente independentes – da matriz \mathbf{X}_{Pcal} .⁸² Além disso, todo o espectro do equipamento secundário é relacionado a cada ponto do equipamento primário, embora seja mais provável que cada ponto do primário se relacione a apenas a uma região próxima, em torno desse ponto no espectro do equipamento secundário.⁸² Isso levou ao desenvolvimento da padronização direta por partes.⁸²

3.2.2.4.3 Padronização direta por partes

A padronização direta por partes (PDS, do inglês *piecewise direct standardization*), também desenvolvido por Wang et al.⁸², é amplamente empregada e, inclusive, utilizada como método de referência na avaliação de novos métodos de transferência.⁷⁹

Na PDS, em vez de relacionar todo espectro do equipamento secundário a cada medida do equipamento primário, cada ponto $\mathbf{x}_{P,i}$ do espectro do equipamento primário no comprimento de onda λ_i é modelado pelos pontos do espectro do equipamento secundário contidos em uma janela em torno de λ_i , $\mathbf{X}_{2,i} = [\mathbf{x}_{2,i-k}, \dots, \mathbf{x}_{2,i}, \dots, \mathbf{x}_{2,i+j}]$. O tamanho da janela deve ser otimizado para cada aplicação. É feita então uma regressão linear multivariada

$$\mathbf{x}_{P,i} = \mathbf{X}_{S,i} \mathbf{b}_i + \mathbf{e}_i \quad \text{Equação 34}$$

para determinar o coeficiente \mathbf{b}_i . Essa regressão pode ser feita tanto por componentes principais quanto por PLS.^{64,82} Os pontos nas extremidades do espectro do equipamento primário não podem ser relacionados a uma janela ao redor destes e são simplesmente descartados ou é feita uma extrapolação.⁸²

Os coeficientes $\hat{\mathbf{b}}_i$ são posicionados como uma faixa diagonal da matriz de transformação \mathbf{F} , onde os demais elementos são zero.

$$\hat{\mathbf{F}} = \text{diag}(\mathbf{b}_1^T, \mathbf{b}_1^T, \dots, \mathbf{b}_i^T, \dots, \mathbf{b}_n^T) \quad \text{Equação 35}$$

sendo n número de variáveis (comprimentos de onda) de \mathbf{X} . Essa matriz pode ser utilizada para padronizar os espectros do equipamento secundário da mesma forma que na DS.

No artigo em que introduziram os métodos DS e PDS, Wang et al.⁸² compararam-nos aos métodos de padronização dos coeficientes de regressão, ao método patenteado por Shenk e Westerhaus⁸³ e à construção de um modelo de calibração novo a partir das amostras de transferência, aplicados a dados simulados e de espectroscopia no infravermelho próximo (NIRS) de amostras de gasolina. Nesse trabalho, os autores conseguiram demonstrar que o PDS apresentava um desempenho superior aos demais métodos avaliados.

3.2.2.4.4 Diferenças espectrais por comprimento de onda

Diferenças espectrais por comprimento de onda (SDW, do inglês *Spectral Difference by Wavelengths*) foi proposto por Fernández-Pierna et al.⁵⁶, embora não tenham utilizado esta denominação em sua publicação. Esse termo foi introduzido por Zamora-Rojas, que utilizou o método proposto por Fernández-Pierna, sendo também adotado nessa dissertação.⁵³

Medem-se algumas amostras de transferência nos dois equipamentos. Em seguida, interpolam-se os espectros para que os comprimentos de onda coincidam entre os equipamentos. A interpolação utilizada no trabalho original foi uma interpolação cúbica com preservação de forma (tradução própria e literal do inglês *shape-preserving piecewise cubic interpolation*) para preservar melhor a forma do espectro.

Em seguida, calcula-se a diferença em cada comprimento de onda entre os espectros médios das amostras de transferência dos equipamentos primário e secundário. Isso é feito subtraindo o espectro médio do equipamento secundário do espectro médio do primário. A diferença entre os espectros é somada aos espectros do equipamento secundário para torná-los semelhantes aos espectros do equipamento primário. O modelo de calibração do

equipamento primário pode, então, ser utilizado inalterado para prever amostras do equipamento secundário.

3.2.2.5 PADRONIZAÇÃO DO MODELO DE CALIBRAÇÃO

Na padronização do modelo de calibração é feita uma transformação no modelo para que ele seja compatível com o novo equipamento. Isso pode ser feito padronizando-se os coeficientes de regressão, os valores previstos da propriedade de interesse ou os espectros de calibração do equipamento primário. Assim como na padronização de espectros, esses métodos também requerem a utilização de amostras de transferência.

3.2.2.5.1 Padronização dos coeficientes de regressão

A padronização dos coeficientes de regressão foi proposta no mesmo trabalho em que Wang et al.⁸⁵ propuseram DS e PDS. Os resultados obtidos por eles não foram bons e esse método não é muito utilizado na literatura.

A padronização dos coeficientes de regressão pode ser feita para modelos de calibração clássico e inverso. Primeiramente, amostras de transferência são medidas nos dois equipamentos.

Segundo o modelo de calibração clássico (medidas instrumentais em função da concentração das espécies presentes) e supondo que todos os constituintes que contribuem para o espectro são conhecidos,^{74,82}

$$\mathbf{X}_{PCal} = \mathbf{Y}_{Cal} \mathbf{B}_P \quad \text{Equação 36}$$

onde \mathbf{X}_{PCal} corresponde às medidas das amostras de calibração no equipamento primário, \mathbf{B}_P aos coeficientes do modelo desse equipamento e \mathbf{Y}_{Cal} às propriedades de interesse dessas.⁸² De maneira análoga, pode-se escrever o modelo para o equipamento secundário como:⁸²

$$\mathbf{X}_{SCal} = \mathbf{Y}_{cal} \mathbf{B}_S = \mathbf{Y}_{Cal} (\mathbf{B}_P + \Delta \mathbf{B}) \quad \text{Equação 37}$$

onde $\Delta\mathbf{B}$ é a matriz de diferença entre as matrizes \mathbf{B}_P e \mathbf{B}_S . Equações análogas podem ser escritas para os conjuntos de amostras de transferência:

$$\mathbf{X}_{Pt} = \mathbf{Y}_t \mathbf{B}_P \quad \text{Equação 38}$$

$$\mathbf{X}_{St} = \mathbf{Y}_t \mathbf{B}_S = \mathbf{Y}_t (\mathbf{B}_P + \Delta\mathbf{B}) \quad \text{Equação 39}$$

Resolvendo as equações 38 e 39 para encontrar $\Delta\mathbf{B}$ tem-se que:⁸²

$$\Delta\mathbf{B} = \mathbf{Y}_t^+ (\mathbf{X}_{St} - \mathbf{X}_{Pt}) \quad \text{Equação 40}$$

onde \mathbf{Y}_t^+ é a pseudo-inversa de \mathbf{Y}_t . Substituindo $\Delta\mathbf{B}$ na equação 37 e utilizando a equação 36, calcula-se:⁸²

$$\mathbf{X}_{PCalPad} = \mathbf{X}_{PCal} + \mathbf{Y}_{Cal} \mathbf{Y}_t^+ (\mathbf{X}_{St} - \mathbf{X}_{Pt}) \quad \text{Equação 41}$$

onde $\mathbf{X}_{PCalPad}$ é uma matriz de calibração padronizada compatível com o equipamento S.

Utiliza-se $\mathbf{X}_{PCalPad}$ e \mathbf{Y}_{Cal} para construir um novo modelo, compatível com o equipamento secundário, utilizando alguma regressão multivariada.⁸²

Esse método requer linearidade nos dois equipamentos e propriedades de interesse conhecidas para todas as espécies que contribuem para o espectro, o que dificulta sua aplicação.^{74,82}

No modelo de calibração inverso para o equipamento primário, tem-se que:^{74,82}

$$\mathbf{y}_{Cal} = \mathbf{X}_{PCal} \mathbf{b}_P \quad \text{Equação 42}$$

$$\mathbf{y}_{Cal} = \mathbf{X}_{SCal} \mathbf{b}_S = \mathbf{X}_{SCal} (\mathbf{b}_{PCal} + \Delta\mathbf{b}) \quad \text{Equação 43}$$

E para as amostras de transferência

$$\mathbf{y}_t = \mathbf{X}_{Pt} \mathbf{b}_{Pt} + \mathbf{e} \quad \text{Equação 44}$$

$$\mathbf{y}_t = \mathbf{X}_{St} \mathbf{b}_{St} = \mathbf{X}_{St} (\mathbf{b}_{Pt} + \Delta\mathbf{b}) \quad \text{Equação 45}$$

Utilizando a equação 42 para calcular \mathbf{b}_{PCal} e combinando as equações 45 e 45 para calcular $\Delta\mathbf{b}$, calcula-se um vetor de regressão padronizado $\hat{\mathbf{b}}_S$.⁸²

$$\begin{aligned}\hat{\mathbf{b}}_S &= \mathbf{b}_P + \Delta\mathbf{b} = \mathbf{b}_P + (\mathbf{b}_{St} - \mathbf{b}_{Pt}) \\ &= \mathbf{X}_{PCal}^+ \mathbf{y}_{Cal} + (\mathbf{X}_{St}^+ - \mathbf{X}_{Pt}^+) \mathbf{y}_t\end{aligned}\quad \text{Equação 46}$$

Com o coeficiente $\hat{\mathbf{b}}_S$, que é uma estimativa de \mathbf{b}_S , pode-se prever a propriedade de interesse em amostras medidas no equipamento secundário.

A padronização de coeficientes no modelo inverso, diferentemente do modelo clássico, somente requer que a propriedade de interesse seja conhecida para o analito.⁸² Esse é o único método onde há uma transferência do modelo de calibração propriamente dita.⁷⁴ Os resultados obtidos com esse método originalmente foram insatisfatórios, provavelmente pela propagação de erros, uma vez que todos os erros são acumulados no vetor estimado para o equipamento secundário.^{74,82}

3.2.2.5.2 Padronização dos valores previstos

A padronização de valores previstos pode ser chamada de correção de inclinação e tendência (SBC, do inglês *slope and bias correction*). Esse método é simples, porém pode ser utilizado apenas para corrigir diferenças simples entre instrumentos que causem desvios de inclinação ou tendência.⁶⁴ Por isso, os equipamentos devem ser muito similares entre si.⁷⁸

São definidas algumas amostras de transferência, que são medidas nos dois equipamentos, e o modelo de calibração do equipamento primário é utilizado para estimar a propriedade de interesse a partir dos espectros dos dois equipamentos.^{78,79}

Com os valores previstos pelo modelo do equipamento primário para os equipamentos primário (\mathbf{y}_p) e secundário (\mathbf{y}_s) é feita uma regressão linear

$$\mathbf{y}_p = a + b \mathbf{y}_s \quad \text{Equação 47}$$

Onde o coeficiente angular “b” corrige a existência de erros sistemáticos proporcionais e o coeficiente linear “a” corrige erros sistemáticos constantes entre os resultados dos dois equipamentos.⁷⁹

Esses coeficientes são então utilizados para corrigir os valores previstos para novas amostras no equipamento secundário.⁶⁴ Esse método está esquematizado na Figura 17.

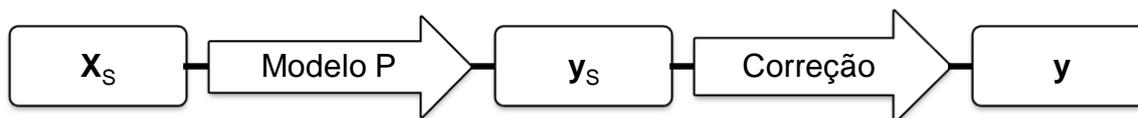


Figura 17. Esquema de SBC. Adaptado da referência 74.

3.2.2.5.3. Padronizações reversas dos espectros

Os métodos de padronização de espectros DS, PRS e SDW podem ser utilizados na formas “reversa” análogas, respectivamente chamadas padronização reversa (RS, do inglês *reverse standardization*), padronização reversa por partes (PRS, do inglês *piecewise reverse standardization*) e SDW reversa (RSDW, do inglês *reverse SDW*).

Nos métodos reversos, em vez de padronizarem-se as novas medidas do equipamento secundário, padroniza-se todo o conjunto de calibração do equipamento primário e utiliza-se esse conjunto para construir um novo modelo de calibração que possa ser aplicado a medidas inalteradas do equipamento secundário, conforma esquematizado na Figura 18.

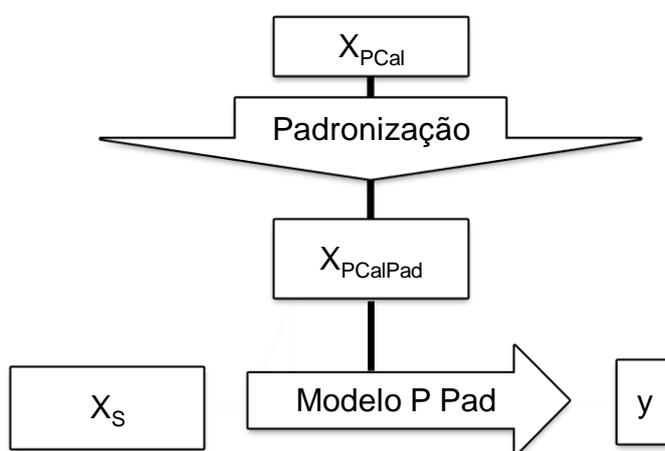


Figura 18. Esquema das padronizações reversas

Como o objetivo é transformar os espectros do equipamento primário para serem compatíveis com o equipamento secundário, os cálculos da matriz de transformação \mathbf{F} no caso da RS e da PRS é feito de forma análoga a DS e PDS, apenas trocando-se nos cálculos \mathbf{X}_P por \mathbf{X}_S e vice e versa. O mesmo raciocínio é válido para o cálculo das diferenças espectrais na RSDW com relação a SDW.

Apesar do maior trabalho de cálculo para padronizar todo o conjunto de calibração e calcular um novo modelo, esses métodos reversos são interessantes em ocasiões em que se queira realizar todas as novas medidas no equipamento secundário, uma vez que o modelo padronizado pode ser aplicado diretamente às suas medidas sem necessidade de novas padronizações.

CAPÍTULO 4

Discriminação de mogno e espécies semelhantes com três equipamentos de NIRS

4. DISCRIMINAÇÃO DE MOGNO E ESPÉCIES SEMELHANTES COM TRÊS EQUIPAMENTOS DE NIRS

Como foi destacado no capítulo 2, o mogno é uma das madeiras mais valiosas do mundo e, em 2003, foi incluído entre as espécies ameaçadas no Apêndice II CITES.⁷ Contudo, existem diversas espécies semelhantes ao mogno, fazendo necessárias técnicas de identificação para evitar o comércio e a exploração ilegal dessa espécie e preservá-la.

Atualmente a principal metodologia utilizada para identificar as espécies de madeira com fins de fiscalização é a anatômica visual. Entretanto, essa requer que a madeira, que geralmente já se encontra processada no momento da fiscalização, seja examinada por especialistas. O desenvolvimento de métodos instrumentais para realizar essa identificação pode contribuir para tornar a fiscalização mais eficiente.

A NIRS é uma excelente alternativa para o estudo da madeira, uma vez que possibilita análises rápidas e não destrutivas e com simples preparo de amostra, podendo ser aplicada a amostras sólidas.^{2,14,16} Em estudos anteriores, NIRS e PLS-DA foram aplicados com sucesso na discriminação da madeira de quatro espécies semelhantes - mogno, andiroba, cedro e curupixá - pela utilização de equipamento NIRS de alta resolução.^{2,17}

Uma abordagem inovadora na identificação de mogno é a utilização de métodos instrumentais portáteis, desenvolvidos para uso em campo que possibilitem uma supervisão da exploração mais prática e viável.

A primeira etapa do presente trabalho consistiu em desenvolver métodos de discriminação para as espécies andiroba (*Carapa guianensis* Aubl.), cedro (*Cedrela odorata* L.), curupixá (*Micropholis melinoniana* Pierre) e mogno (*Swietenia macrophylla* King) por PLS-DA em três equipamentos NIRS individualmente, sendo um portátil e dois de bancada. O objetivo era principalmente verificar a capacidade de discriminação dos modelos desenvolvidos no equipamento portátil e comparar seu desempenho com os demais.

4.1 EXPERIMENTAL

As espécies foram selecionadas com base nos trabalhos de Pastore et al.² e Braga et al.,¹⁷ onde foram escolhidas as três principais espécies semelhantes ao mogno de acordo com o livro "*Madeiras similares ao mogno (Swietenia macrophylla King.): chave ilustrada para identificação anatômica em campo.*",⁵ editado pelo Serviço Florestal Brasileiro (SFB).

4.1.1 Amostras

As amostras foram cedidas pelo Laboratório de Produtos Florestais (LPF) do SFB, já previamente identificadas por espécie, secas ao ar livre, cortadas e com as faces orientadas (de acordo com as seções transversal, radial e tangencial, conforme a Figura 2). As superfícies de todas as amostras foram lixadas com lixa número 80, mantendo-se a uniformidade granulométrica, que é um fator importante para a distribuição da radiação refletida.

As amostras de andiroba, cedro e curupixá foram obtidas de discos localizados às bases dos troncos das árvores. Cada amostra corresponde a uma árvore (indivíduo) diferente. Essas espécies foram coletadas em áreas de exploração autorizadas no estado do Pará. Os espécimes foram doados pelas empresas Serraria Marajoara Ind. Com. e Exp., Selectas S.A. Ind. e Com.de Madeira Ltda, Madeireira Caingá Ltda e Empresa Juruá Florestal.

As amostras de mogno foram obtidas de tábuas apreendidas oriundas do Mato Grosso do Sul que seriam exportadas.

Como forma de garantir a identificação das amostras por espécie, essas foram identificadas pela anatomista botânica Dra. Vera Teresinha Rauber Coradin do LPF, SFB.

Foram utilizadas 111 amostras na forma de blocos de aproximadamente 2x2x2cm, sendo 26 de andiroba, 28 de cedro, 29 de curupixá e 28 de mogno.

Não foi realizado nenhum preparo de amostras antes das medidas.

4.1.2 Espectros

Os equipamentos de alta resolução utilizados foram o espectrômetro FT-IR/NIR Tensor 37 (Bruker Optics) pertencente ao LPF e o espectrômetro FT-NIR Antaris II (Thermo Scientific) pertencente ao Departamento de Engenharia Florestal da Universidade de Brasília.

O espectrômetro NIR portátil utilizado foi o Phazir (Polychromix), pertencente ao Centro de Cooperação Internacional em Pesquisa Agronômica pelo Desenvolvimento (CIRAD, do francês *Centre de Coopération Internationale en Recherche Agronomique pour le Développement*), na França. Os espectros desse equipamento foram obtidos na unidade de Montpellier, na França, através de uma parceria entre LPF e CIRAD, e cedidos ao presente trabalho.

As configurações dos equipamentos utilizados foram as seguintes:

- **Tensor 37:** Resolução de 4 cm^{-1} , 64 varreduras por amostra, região espectral de 15.000 cm^{-1} a 3.500 cm^{-1} , medidas de reflectância difusa com esfera integradora e espectro resultante em escala de absorbância ($\log(1/R)$).
- **Antaris II:** Resolução de 4 cm^{-1} , 64 varreduras por amostra, região espectral de 10.000 cm^{-1} a 4.000 cm^{-1} , medidas de reflectância difusa com esfera integradora e espectro resultante em escala de absorbância ($\log(1/R)$).
- **Phazir:** Resolução não constante variando entre 102 cm^{-1} e $24,9\text{ cm}^{-1}$, 1 varredura por amostra e região espectral de 10.644 cm^{-1} a 5.5661 cm^{-1} , medidas de reflectância difusa com tecnologia de MEMS e DTS e espectro resultante em escala de absorbância ($\log(1/R)$).

Para cada amostra foram obtidos quatro espectros – dois em cada face radial, medidos em pontos distintos da face.

4.2 ANÁLISE DE DADOS

As análises de dados foram feitas no programa OPUS 6.5 da Bruker Optics.

Após uma rotina preliminar de otimização dos pré-processamentos para cada equipamento, escolheu-se derivar os espectros dos equipamentos Antaris II e Tensor 37. A primeira derivada foi feita utilizando filtro Savitzky–Golay com janela de 9 pontos e polinômio de segunda ordem. Os espectros do Phazir não foram derivados.

Foi calculada a média dos quatro espectros de cada amostra resultando em 111 espectros de cada equipamento. As análises dos dados dos três equipamentos foram feitas separadamente, porém foi seguido o mesmo procedimento detalhado a seguir.

Dois terços dos 111 espectros foram utilizados para a construção dos modelos de calibração e um terço para previsão.

Para construção dos modelos PLS1-DA, valores de classe de referência foram atribuídos a cada espécie do conjunto de calibração como sendo 1 para as amostras pertencentes à classe (espécie) em questão e 0 para as demais.

Em seguida, foi feita uma regressão PLS do vetor \mathbf{y} de cada classe e da matriz de dados \mathbf{X} correspondente aos espectros para obtenção de um modelo para cada classe.

Na rotina de otimização do OPUS, foi feita uma validação cruzada completa com otimização dos pré-processamentos e da região espectral. O número máximo de variáveis latentes (VL) foi fixado como 20. O programa selecionou os parâmetros com menores RMSECV, escolhendo automaticamente o número de VL.

Testaram-se para cada equipamento alguns dos melhores pré-processamentos apontados pela otimização levando-se em conta os menores valores de RMSECV e os menores números de VL e optou-se por utilizar o mesmo pré-processamento para as quatro espécies.

O modelo foi então construído com os espectros pré-processados e selecionou-se o número de VL que apresentou o menor RMSECV ou o número de VL a partir do qual não se observou diminuição significativa dos valores de RMSECV. Nessa etapa também foram excluídas as amostras anômalas (*outliers*) segundo os critérios do programa, que se baseia nos resíduos não modelados e influência das amostras.

A otimização do modelo com relação aos *outliers* foi feita em etapas. Um primeiro modelo foi construído e os *outliers* foram excluídos do conjunto de calibração. Em seguida construiu-se um novo modelo com as amostras restantes, os novos *outliers* (caso houvesse) foram excluídos e construiu-se outro modelo. Dado o número reduzido de amostras, esse procedimento de exclusão foi feito apenas duas vezes, ainda que no terceiro modelo houvesse novos *outliers*.

Após a construção do modelo de calibração, sua eficiência de previsão foi testada utilizando-se os espectros das amostras do conjunto de previsão. Os *outliers* de previsão foram excluídos. Os valores de RMSEP foram determinados e utilizados para avaliar o desempenho dos modelos, tomando-se o cuidado de utilizar o mesmo número de VL definido na fase de validação cruzada.

Os limites de discriminação dos modelos foram determinados por uma análise Bayesiana utilizando o programa Matlab. Essa análise assume a normalidade da distribuição das amostras de calibração e leva em consideração a sua dispersão.⁸⁶

4.3 RESULTADOS E DISCUSSÃO

A Figura 19 mostra os espectros dos três equipamentos não pré-processados.

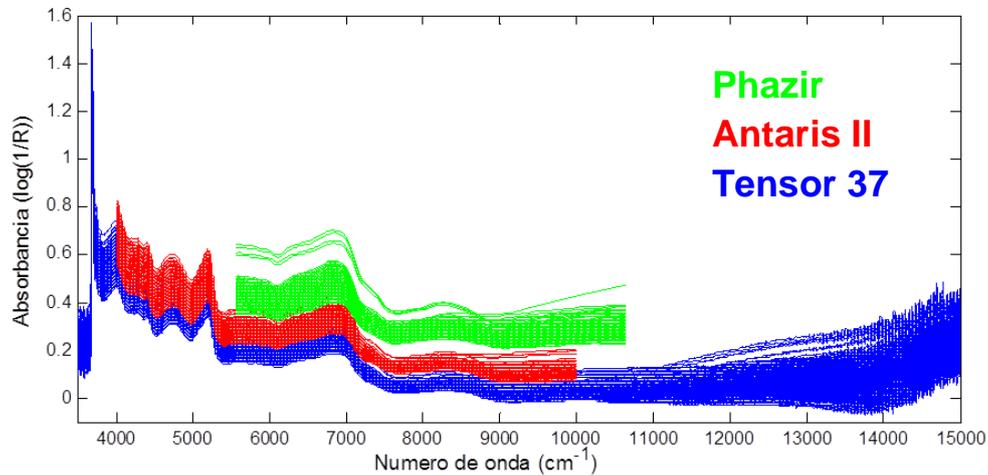


Figura 19. Espectros dos equipamentos Tensor 37 (azul), Antaris II (vermelho) e Phazir (verde).

A Figura 20 mostra os espectros do Phazir após a média das replicatas coloridos por espécie. Observa-se que os espectros das quatro espécies são muito semelhantes.

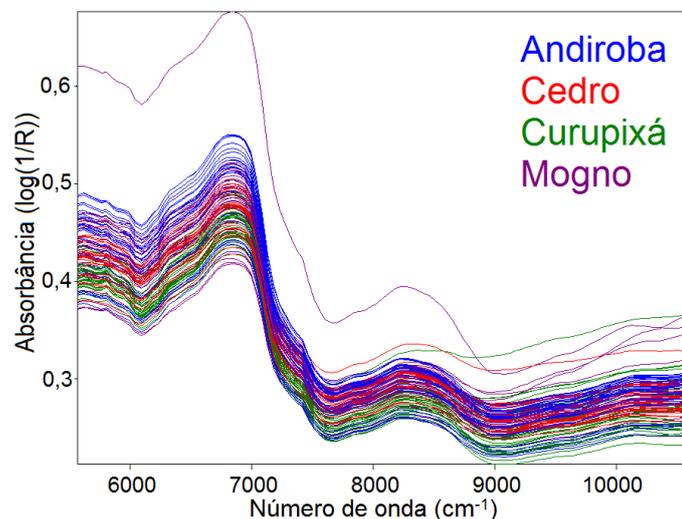


Figura 20. Espectros médios das amostras de andiroba (azul), cedro (vermelho), curupixá (verde) e mogno (roxo) medidas no equipamento Phazir.

Embora não seja necessário identificar as bandas dos espectros para realizar a discriminação por PLS-DA, é interessante conhecer as origens das principais bandas observadas na região. Por isso, segue uma breve discussão acerca da atribuição das bandas observadas nos espectros obtidos.

Em espectros de madeira no infravermelho próximo, as bandas na região espectral de 7.000 a 12.800 cm^{-1} são atribuídas aos primeiros e segundos *overtones* das vibrações de estiramento das ligações O-H e N-H, aos primeiros *overtones* das bandas de combinação C-H e aos segundos e terceiros *overtones* das vibrações de estiramento das ligações C-H alifáticas e aromáticas e N-H.⁵⁷ Essas ligações C-H e O-H estão presentes nos diversos componentes da madeira, como lignina, celulose, hemicelulose, extrativos.⁵⁷ As ligações N-H são encontradas em proteínas e aminas.

Pode-se observar nas Figuras 19 e 20 que não é possível identificar visualmente bandas acima de aproximadamente 9.000 cm^{-1} . Aproximadamente entre 7.500 e 9.000 cm^{-1} observa-se uma banda larga que se deve, possivelmente, à combinação de bandas dos segundos *overtones* das vibrações de estiramento das ligações C-H aromáticas e metílicas e estiramentos assimétricos das ligações C-H e HC=CH das moléculas de lignina com bandas dos segundos *overtones* das ligações C-H metílicas em grupos CH_3 em acetilas de ésteres presentes nas hemiceluloses e bandas dos segundos *overtones* de estiramentos C-H em celulose.⁵⁷

Em aproximadamente 7.400 cm^{-1} aparentemente há um pico de uma banda (que se funde a outras banda à esquerda) que pode ser atribuída ao primeiro *overtone* de bandas de combinação do estiramento C-H e das deformações angulares das ligações C-H em grupos CH_3 em acetila de ésteres presentes nas hemiceluloses.⁵⁷ Unindo-se a essa banda à esquerda estão as bandas dos primeiros *overtones* das bandas de combinação do estiramento C-H e das deformações angulares das ligações C-H da celulose, hemicelulose e outros componentes da madeira.⁵⁷

A região de 6.000 a 7.000 cm^{-1} apresenta as bandas do primeiro *overtone* da vibração de estiramento O-H, sendo estas influenciadas por ligações de hidrogênio presentes entre as moléculas.⁵⁷ Essas ligações estão presentes na celulose, hemicelulose, lignina, extrativos e também nas moléculas de água presentes na madeira, que é um material poroso e higroscópico.⁵⁷ Nas Figuras 19 e 20 observa-se uma banda larga e intensa nessa região devida à

4. DISCRIMINAÇÃO DE MOGNO E ESPÉCIES SEMELHANTES COM TRÊS EQUIPAMENTOS DE NIRS

combinação das bandas de vibrações de estiramento das ligações O-H presentes nos diversos componentes da madeira e moléculas de água.

A região entre 5.000 e 6.000 cm^{-1} apresenta as bandas do primeiro *overtone* das vibrações de estiramento das ligações C-H alifáticas e aromáticas presentes nas moléculas de celulose, hemicelulose e lignina e as bandas de combinação das ligações O-H presentes nas moléculas de celulose e água.⁵⁷ Ainda nessa região pode haver contribuição do segundo *overtone* da vibração de estiramento da ligação C=O de moléculas de hemicelulose.⁵⁷ Essa região é apenas parcialmente contemplada pelo espectro do equipamento Phazir. Na Figura 19 observa-se uma banda muito larga devida à união das bandas citadas anteriormente, sendo um pico mais intenso observado em aproximadamente 5.100 cm^{-1} provocado por bandas de combinação de vibrações da ligação O-H de moléculas de água.

A atribuição de bandas na região entre 4.000 e 5.000 cm^{-1} , que corresponde a bandas de combinação, é dificultada devido ao grande número de acoplamentos possíveis.⁵⁷

Os pré-processamentos selecionados para os três equipamentos estão organizados na Tabela 5.

Tabela 5. Pré-processamentos utilizados nos equipamentos

| Equipamento | Pré-processamentos | | |
|--------------------|---------------------------|------------------|-------------------------|
| Tensor 37 | 1ª Derivada | Centrar na média | Subtração de linha reta |
| Antaris II | 1ª Derivada | Centrar na média | Subtração de linha reta |
| Phazir | - | Centrar na média | Subtração de linha reta |

As Figuras 21 e 22 mostram os espectros pré-processados, porém não centrados na média.

4. DISCRIMINAÇÃO DE MOGNO E ESPÉCIES SEMELHANTES COM TRÊS EQUIPAMENTOS DE NIRS

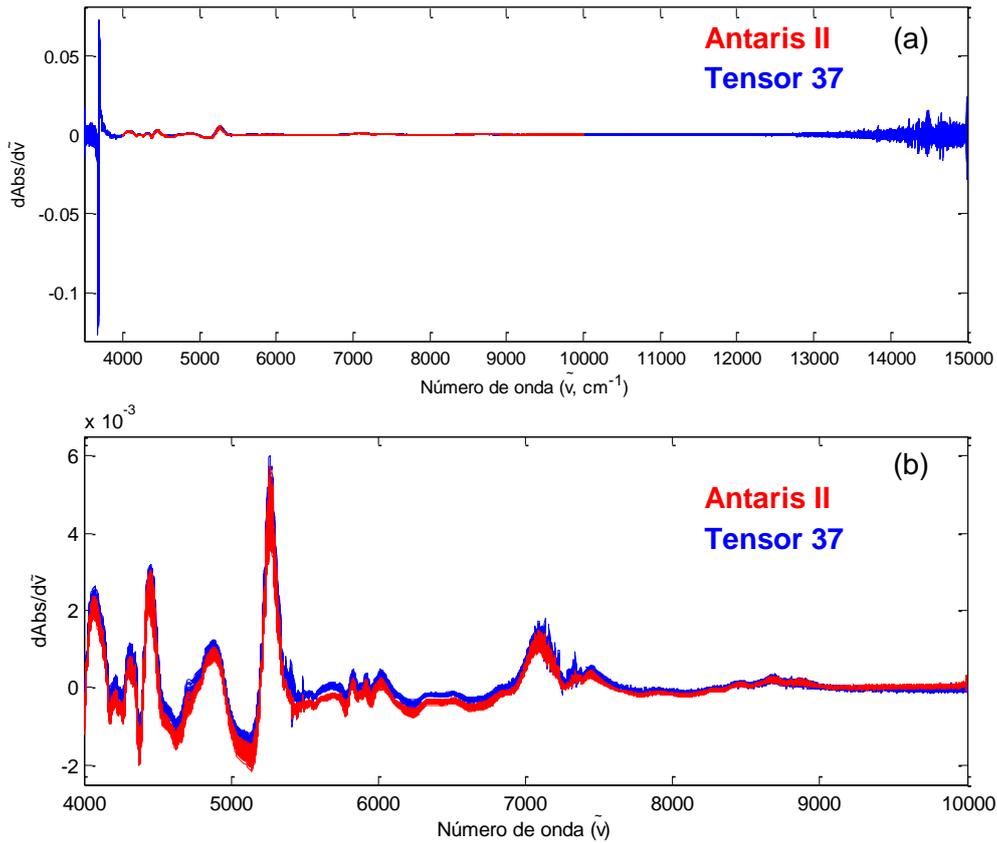


Figura 21. Espectros dos equipamentos Tensor 37 (azul) e Antaris II (vermelho) pré-processados com derivada primeira e subtração de linha com a) região espectral completa e b) região de maior interesse ampliada.

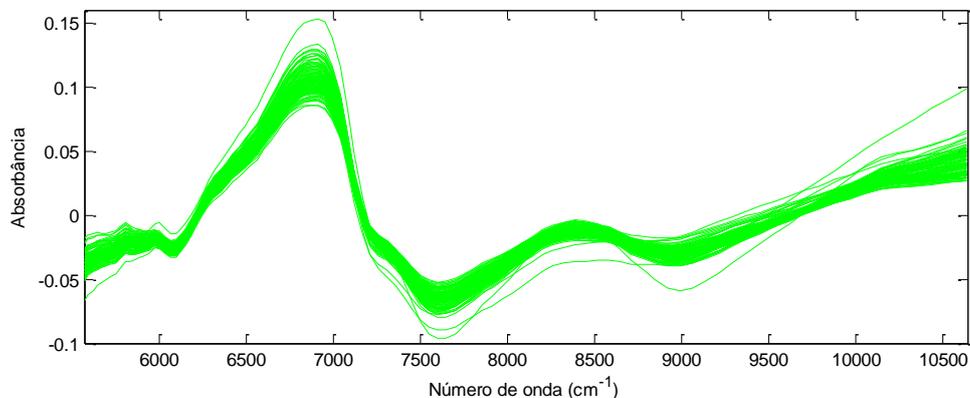


Figura 22. Espectros do equipamento Phazir pré-processados com subtração de linha reta.

Os valores das raízes dos erros quadráticos médios de previsão (RMSEP), calibração (RMSEC) e validação cruzada (RMSECV), parâmetros de discriminação e as regiões espectrais dos modelos desenvolvidos nos três equipamentos para as quatro espécies estão dispostos na Tabela 6. As Figuras 23, 24 e 25 mostram as dispersões das amostras quanto ao valor de classe estimado e o limite de discriminação (a-d), os histogramas dos valores de

classe e gráficos de probabilidade (e-h) para as quatro espécies nos equipamentos Tensor 37, Antaris II e Phazir respectivamente.

O equipamento de campo Phazir apresentou erros cerca de duas a três vezes maiores do que os demais. Entretanto, esses erros foram suficientemente baixos para permitir a separação entre as espécies sem erros de classificação, como se pode observar na Figura 25.

Quando há erros de classificação, duas situações são possíveis: i) amostras que não pertencem à espécie que se quer discriminar das demais apresentarem valor de classe estimado superior ao limite de discriminação ou ii) amostras que pertencem à espécie que se quer discriminar das demais apresentarem valor de classe estimado inferior ao limite de discriminação.

O limite de discriminação é um valor intermediário entre 0 e 1 calculado levando-se em conta a dispersão das amostras que pertencem ou não à espécie que está sendo discriminada e, por isso, costuma ser próximo porém diferente de 0,5. O valor do limite de discriminação corresponde ao valor de classe onde a curva da probabilidade de uma amostra pertencer à classe que está sendo discriminada intersecta a curva da probabilidade de uma amostra pertencer às demais classes (Figuras 23-25 (f-h)).

Os histogramas das Figuras 23 a 25, f a h apresentam um eixo horizontal correspondente ao valor de classe estimado para as amostras e dois eixos verticais. O eixo vertical da esquerda se refere ao número de amostras que estão dentro de cada faixa de valores de classe, representadas pelas barras do histograma. O eixo vertical da direita corresponde às curvas de probabilidade de uma amostra com um determinado número de classe estimado pertencer (linha contínua nas figuras) ou não (linha tracejada nas figuras) à classe que está sendo discriminada. Por isso, espera-se uma distribuição das amostras que não pertencem à classe discriminada à esquerda do limite de discriminação (encontro das curvas de probabilidade) e centrada próximo ao valor zero. E espera-se uma distribuição das amostras que pertencem à classe discriminada à direita do limite de discriminação e centrada próximo a 1.

4. DISCRIMINAÇÃO DE MOGNO E ESPÉCIES SEMELHANTES COM TRÊS EQUIPAMENTOS DE NIRS

Tabela 6. Erros e parâmetros de discriminação das quatro espécies nos três equipamentos

| Espécie | Andiroba | | | Cedro | | | Curupixá | | | Mogno | | |
|----------------------------|------------------|-----------------|-----------------|------------------|-----------------|------------------|------------------|-----------------|-------------------|-------------------|-----------------|------------------|
| | Tensor 37 | Antaris II | Phazir | Tensor 37 | Antaris II | Phazir | Tensor 37 | Antaris II | Phazir | Tensor 37 | Antaris II | Phazir |
| NIRS | | | | | | | | | | | | |
| Nº de variáveis latentes | 8 | 7 | 13 | 7 | 8 | 10 | 8 | 7 | 9 | 8 | 7 | 13 |
| RMSEC | 0,034 | 0,039 | 0,095 | 0,024 | 0,041 | 0,083 | 0,026 | 0,039 | 0,078 | 0,034 | 0,034 | 0,073 |
| RMSECV | 0,070 | 0,075 | 0,156 | 0,056 | 0,082 | 0,116 | 0,050 | 0,074 | 0,104 | 0,072 | 0,064 | 0,116 |
| RMSEP | 0,065 | 0,068 | 0,151 | 0,069 | 0,060 | 0,155 | 0,054 | 0,058 | 0,140 | 0,061 | 0,061 | 0,179 |
| Limite de discriminação | 0,481 | 0,518 | 0,516 | 0,432 | 0,421 | 0,492 | 0,438 | 0,468 | 0,388 | 0,399 | 0,451 | 0,530 |
| Outliers calibração | 1 | 1 | - | 2 | - | 2 | 2 | 1 | 1 | 2 | 1 | - |
| Outliers previsão | 1 | - | - | - | 2 | 1 | 1 | - | - | 1 | - | - |
| Região (cm ⁻¹) | 10.001 -6.099 | 9.999- 7.500 | 9.093- 5.881 | 10.001 -7.725 | 9.999- 7.500 | 10.644 -7.645 | 10.001 -6.099 | 9.999- 7.500 | 10.644 -9.028; | 10.001 -7.725; | 9.999- 7.500 | 10.644 -6.666 |
| | | | | | | | | | 10.644 -9.028; | 10.001 -7.725; | 9.999- 7.500 | 10.644 -6.666 |

4. DISCRIMINAÇÃO DE MOGNO E ESPÉCIES SEMELHANTES COM TRÊS EQUIPAMENTOS DE NIRS

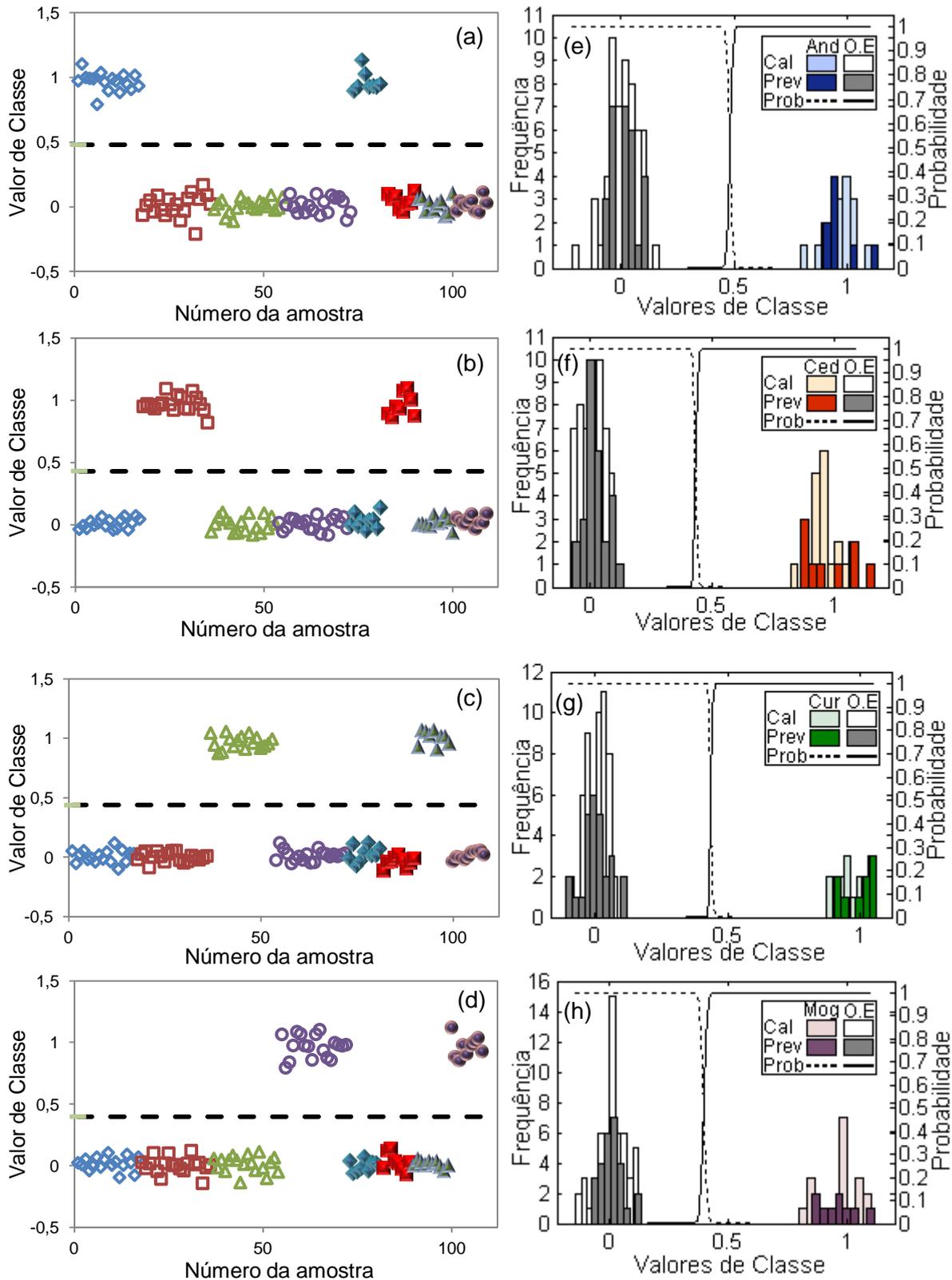


Figura 23. Resultados do equipamento **Tensor 37**: Valores de classe e limites de discriminação (---) para a) andiroba (\diamond), b) cedro (\square), c) curupixá (\triangle) e d) mogno (\circ). Símbolos vazios = calibração e preenchidos = previsão. Histogramas dos valores de classe para discriminação de e) andiroba (and), f) cedro (ced), g) curupixá (cur) e h) mogno (mog) das outras espécies (O.E) com gráficos de probabilidades (Prob) das amostras pertencerem à espécie em questão (—)ou às outras espécies (----).

4. DISCRIMINAÇÃO DE MOGNO E ESPÉCIES SEMELHANTES COM TRÊS EQUIPAMENTOS DE NIRS

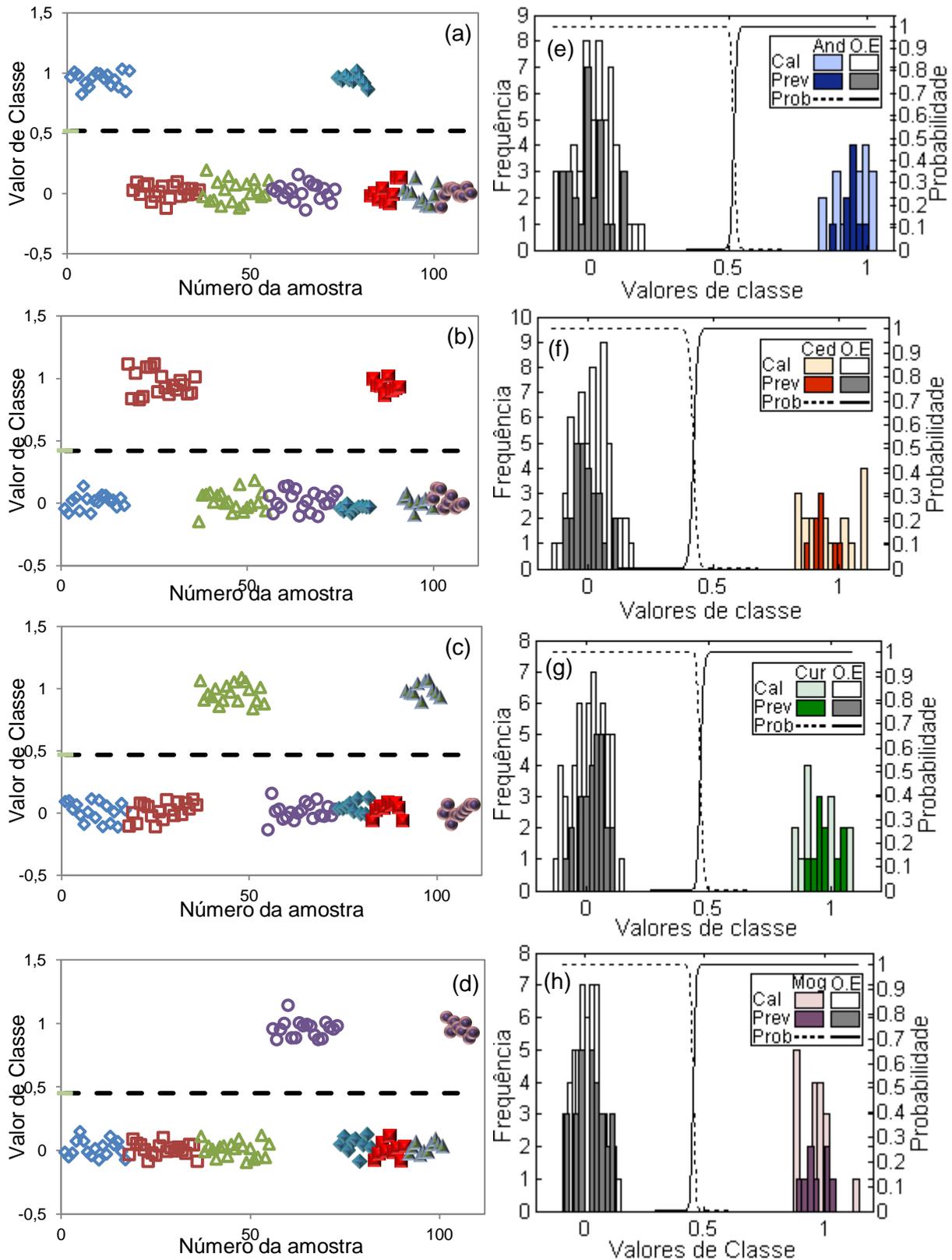


Figura 24. Resultados do equipamento **Antaris II**: Valores de classe e limites de discriminação (---) para a) andiroba (\diamond), b) cedro (\square), c) curupixá (Δ) e (d) mogno (\circ). Símbolos vazios = calibração e preenchidos = previsão. Histogramas dos valores de classe para discriminação de e) andiroba (and), f) cedro (ced), g) curupixá (cur) e h) mogno (mog) das outras espécies (O.E) com gráficos de probabilidades das amostras pertencerem à espécie em questão (—)ou às outras espécies (----)

4. DISCRIMINAÇÃO DE MOGNO E ESPÉCIES SEMELHANTES COM TRÊS EQUIPAMENTOS DE NIRS

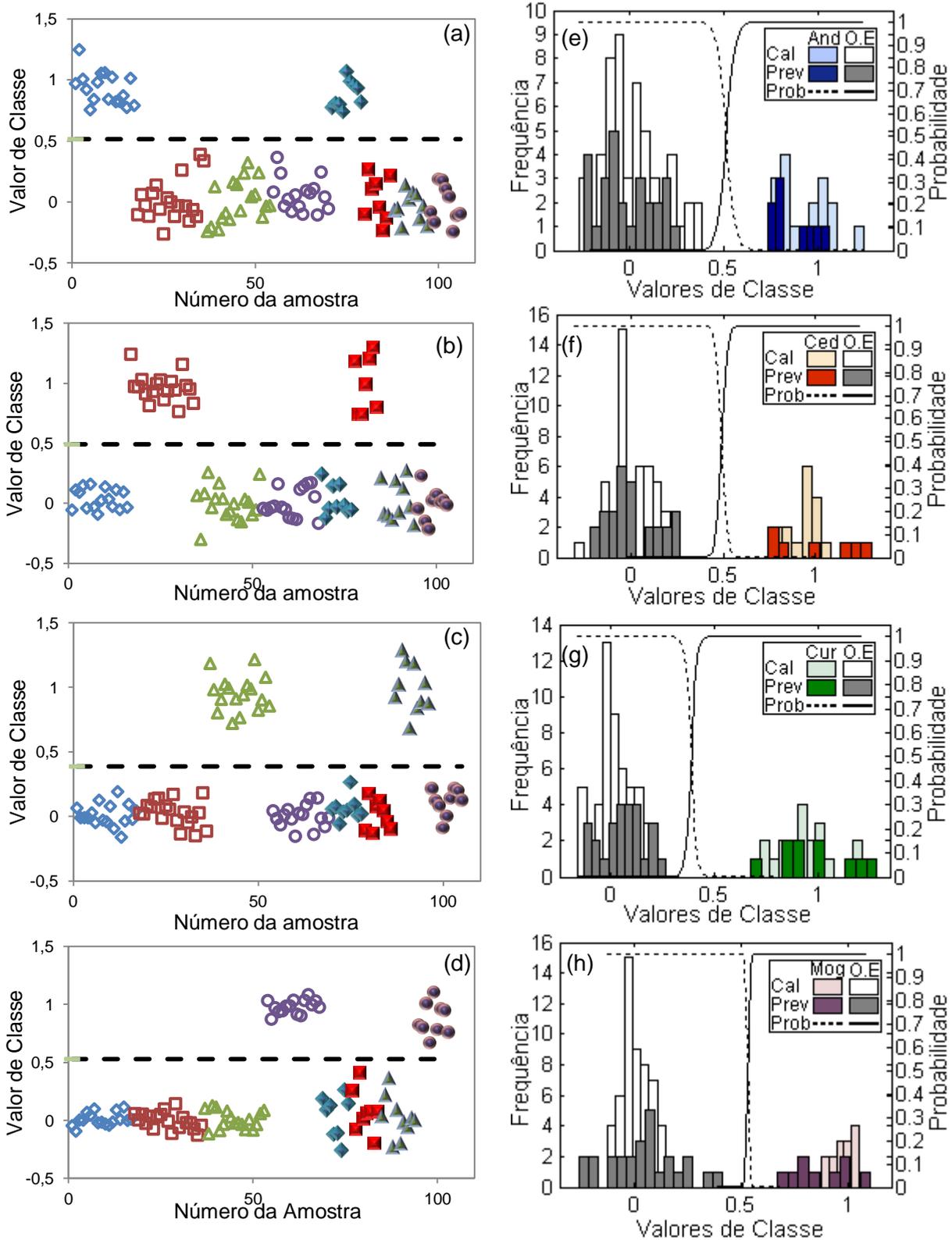


Figura 25. Resultados do equipamento **Phazir**: Valores de classe e limites de discriminação (---) para a) andiroba (\diamond), b) cedro (\square), c) curupixá (Δ) e d) mogno (\circ). Símbolos vazios = calibração e preenchidos = previsão. Histogramas dos valores de classe para discriminação de e) andiroba (and), f) cedro (ced), g) curupixá (cur) e h) mogno (mog) das outras espécies (O.E) com gráficos de probabilidades das amostras pertencerem à espécie em questão (—) ou às outras espécies (—).

O equipamento Tensor 37 mostrou-se superior na discriminação de andiroba, cedro e curupixá enquanto o equipamento Antaris II forneceu melhores resultados para mogno. A eficiência do equipamento na discriminação é verificada pelos menores erros de previsão, ausência de sobreajuste e menor número de *outliers*. Os erros estão associados à dispersão das amostras e ao desvio do valor de classe estimado com relação ao valor esperado. Outro fator importante para verificar a eficiência da discriminação é a distância entre os valores de classe estimados e os limites de discriminação.

Observa-se nas Figuras 23 a 25 que as amostras de calibração do Tensor 37 estão menos dispersas com relação aos demais equipamentos, notando-se que os histogramas estão mais compactos e com picos próximos aos valores de referência. O modelo para discriminação de mogno é uma exceção, pois se observa que o histograma do Antaris II para esta espécie apresenta melhor configuração e as amostras estão menos dispersas. Os resultados do Phazir estão consideravelmente mais dispersos, se aproximando do limite de discriminação. Entretanto, como foi dito anteriormente, os erros foram suficientemente baixos para possibilitar a discriminação com 100% de classificações corretas. Isso mostra que esse equipamento pode ser aplicado com essa finalidade, considerando a praticidade que o equipamento oferece.

Os números de variáveis latentes (VL) foram maiores para o equipamento portátil Phazir, o que se deve, provavelmente à maior variação do equipamento e menor sensibilidade com relação aos equipamentos de bancada.

Em alguns casos a diferença relativamente grande entre RMSEC e RMSEP indica a possível ocorrência de sobreajuste, onde o modelo funciona muito bem para prever as amostras de calibração, com as quais foi desenvolvido, mas não funciona tão bem para novas amostras, como o conjunto independente de previsão. Esse efeito foi mais acentuado nos resultados do equipamento portátil Phazir para a discriminação de mogno das demais espécies e nos resultados da discriminação de cedro com o equipamento Tensor 37. Nesses casos, o RMSEP é maior que o dobro do RMSEC. Observa-se que nos gráficos e histogramas da Figura 24 as amostras de previsão estão relativamente mais

dispersas com relação às de calibração, ilustrando a diferença entre os erros. Também foi possível observar nos histogramas a distribuição das amostras em alguns casos não se assemelha visualmente a uma distribuição normal. Entretanto, pelo teste de Jarque-Bera,⁸⁷ pode-se assumir a normalidade de todas as distribuições com 95% de confiança exceto para as amostras de andiroba do conjunto de teste do equipamento Tensor 37 na discriminação dessa espécie.

A presença de sobreajuste pode estar relacionada ao número reduzido de amostras. Contudo a aquisição de espectros de mais amostras com o equipamento Phazir estava impossibilitada no presente estudo.

Uma estratégia possível para se obter melhores resultados, adotada nos capítulos seguintes, é não realizar a média das replicatas, considerando que a madeira é um material complexo e heterogêneo. Assim, cada uma das quatro medidas feitas em pontos distintos da amostra seria considerada como uma amostra.

4.4 CONCLUSÕES DO CAPÍTULO

A discriminação da madeira de mogno das espécies semelhantes andiroba, cedro e curupixá por NIRS associada a PLS-DA mostrou-se eficiente nos três equipamentos estudados. Apesar de apresentarem erros de duas a três vezes maiores, as análises com o equipamento portátil de baixa resolução não apresentaram erros de classificação, indicando a possibilidade de utilização deste para fins de fiscalização da exploração e comércio de madeira. Observaram-se indícios de sobreajuste, especialmente nos modelos construídos utilizando o equipamento portátil. Contudo, mesmo nos casos de possível sobreajuste, não foram observados erros de discriminação.

CAPÍTULO 5
Transferência de calibração entre equipamentos

5. TRANSFERÊNCIA DE CALIBRAÇÃO ENTRE EQUIPAMENTOS

Conforme foi argumentado na seção 3.2.2 do capítulo 3, a construção de modelos de calibração multivariada pode exigir que seja medido um número elevado de amostras, o que implica certo custo e tempo, de forma que é interessante que tais modelos apresentem ampla aplicabilidade. Entretanto, existem diversas circunstâncias que fazem com que um modelo perca sua validade, ou seja, sua capacidade de estimar a propriedade de interesse para novas amostras.

A transferência de calibração pode oferecer uma alternativa viável para a extensão da aplicabilidade desses modelos diante de mudanças nas condições da análise, evitando a necessidade de reconstruí-los nas novas condições.

Nesse capítulo serão abordadas duas situações onde a transferência de calibração entre equipamentos distintos foi aplicada para a discriminação de mogno, andiroba, cedro e curupixá.

A primeira situação consiste na transferência de calibração entre dois equipamentos de bancada utilizando os métodos diretos DS, PDS e SDW. Essa transferência possibilita o compartilhamento de dados entre laboratórios que possuem equipamentos diferentes. Nos métodos diretos de transferência, os espectros obtidos no equipamento secundário são padronizados para se assemelharem aos espectros do equipamento primário. Assim, o modelo construído no equipamento primário pode ser aplicado aos espectros padronizados do equipamento secundário.

A segunda situação consiste na transferência de calibração de equipamentos de bancada com alta resolução para um equipamento portátil de baixa resolução utilizando os métodos reversos RS, PRS e RSDW. Nesses métodos, os dados dos equipamentos de bancada são padronizados para que se assemelhem aos dados do equipamento portátil. Esses dados padronizados são utilizados para construção de novos modelos, que são compatíveis com as medidas do equipamento portátil. Dessa forma, é possível aproveitar bancos de

dados dos equipamentos de bancada e a propriedade de interesse pode ser estimada em novas medidas do equipamento portátil sem necessidade de padronização. Essa transferência possibilita implementar novos equipamentos sem perder bancos de dados espectrais antigos.

5.1 EXPERIMENTAL

Os espectros utilizados neste capítulo foram os mesmos descritos na seção 4.1 do capítulo anterior.

5.2. ANÁLISE DE DADOS

A análise de dados foi conduzida no programa Matlab 7.12.0 (R2011a) com o pacote PLS Toolbox 6.5. A numeração das medidas das amostras nos três equipamentos utilizados é perfeitamente correspondente, de forma que os conjuntos de calibração e previsão de cada equipamento são constituídos pelas mesmas amostras. Das 111 amostras, dois terços foram selecionados para construção dos modelos de calibração e um terço para validação dos modelos. Os procedimentos para as padronizações diretas e reversas serão descritos a seguir.

5.2.1. Transferência direta entre equipamentos de bancada

O equipamento Tensor 37 foi designado como equipamento primário (P) e o equipamento Antaris II como secundário (S).

Primeiramente, foi feita uma interpolação linear para que os pontos do espectro do Antaris II correspondessem aos do Tensor 37, sendo cortadas as extremidades do espectro do Tensor 37 que não tinham correspondência nos espectros do Antaris II.

Começando com a discriminação de mogno com relação às demais espécies, construíram-se modelos PLS1-DA no Tensor 37 empregando diversos pré-processamentos.

Selecionou-se o pré-processamento que gerou o modelo PLS-DA com menor valor de RMSECV e construiu-se um modelo no Antaris II utilizando as mesmas condições.

Plotando-se os espectros pré-processados, optou-se por utilizar duas faixas de comprimento de onda que apresentavam menos ruído e construíram-se novos modelos nos dois equipamentos utilizando apenas essas faixas.

Os *outliers* de calibração foram detectados com base no método descrito na norma da ASTM E1655-05,⁷⁰ e sugerido por Valderrama et al.,⁷³ considerando como *outliers* amostras com altos resíduos não modelados em y , utilizando intervalo de confiança de 99%. Excluíram-se os *outliers* de ambos os modelos dos conjuntos de calibração dos dois equipamentos.

Construíram-se novos modelos sem os outliers de calibração e aplicaram-se estes aos conjuntos de previsão dos respectivos equipamentos. Detectaram-se os *outliers* de previsão e excluíram-nos de ambos os conjuntos de teste.

O resultado da recalibração foi obtido pela aplicação do modelo construído no Antaris II às amostras de previsão do mesmo equipamento.

O resultado da não transferência foi avaliado pela aplicação do modelo construído no equipamento Tensor 37 ao conjunto de previsão do Antaris II

As amostras de transferência foram selecionadas utilizando o algoritmo Kennard-Stone (KS),⁸⁸ a partir do conjunto de calibração do Tensor 37. Esse algoritmo encontra as duas amostras com maior distância Euclidiana entre si e, em seguida, seleciona a amostra com maior distância com relação à amostra selecionada mais próxima, repetindo esse último processo até que o número desejado de amostras seja selecionado. Tomaram-se as quatro replicatas de cada amostra selecionada como amostra de transferência. Caso alguma replicata estivesse entre os outliers excluídos, selecionaram-se apenas as replicatas restantes.

Realizou-se a DS, conforme descrita na seção 3.2.2.4.2, variando o número de amostras de transferência entre cinco (menor número necessário

para o algoritmo selecionar pelo menos uma amostra de cada espécie) e doze. Para escolher o melhor número de amostras de transferência, aplicou-se o modelo P do equipamento primário a um conjunto independente de amostras padronizado do equipamento secundário. Uma vez determinado o número de amostras de transferência, aplicou-se o modelo P ao conjunto de previsão padronizado do equipamento S.

A PDS foi conduzida da mesma forma e também se utilizou o conjunto padronizado e independente de amostras do equipamento S para otimizar o tamanho da janela de pontos empregada. A Figura 26 esquematiza a PDS e a DS pode ser entendida analogamente.

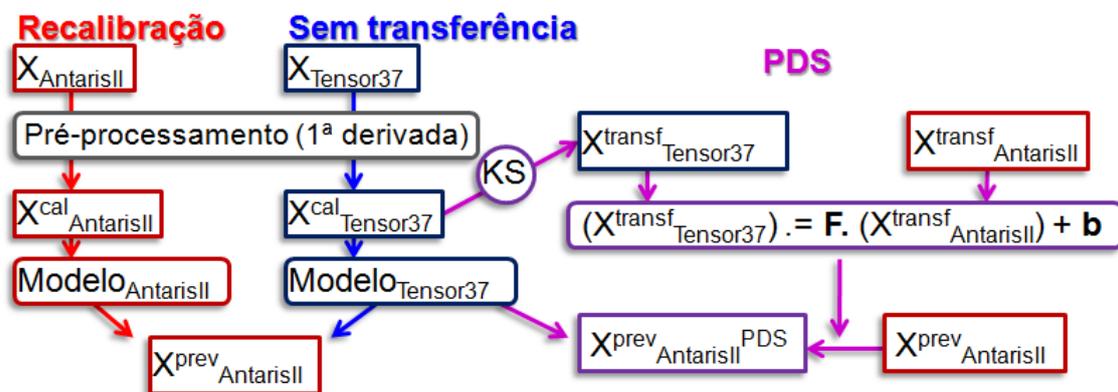


Figura 26. Fluxograma de etapas envolvidas na realização do método PDS e dos métodos de comparação sem transferência e de recalibração completa.

Para calcular a SDW, utilizou-se o mesmo número de amostras de transferência selecionado para os demais métodos. Calculou-se a média dos espectros das amostras de transferência em cada equipamento. Em seguida subtraiu-se a média do equipamento S da média do equipamento P. Essa diferença foi somada a cada espectro do conjunto de previsão do equipamento S, que foi previsto pelo modelo P.

O mesmo procedimento foi realizado para as demais espécies, andiroba cedro e curupixá, mantendo-se o mesmo conjunto de amostras de transferência selecionado para o mogno, excluindo-se os *outliers* de cada modelo caso coincidisse com alguma replicata das amostras de transferência.

5.2.2. Transferência reversa de equipamentos de bancada para equipamento portátil

Primeiramente, foi feita uma interpolação linear para que os pontos do espectros do Antaris II e do Tensor 37 correspondessem aos do Phazir, sendo cortadas as extremidades dos espectros dos equipamentos de bancada que não tinham correspondência nos espectros do portátil. A transferência entre cada equipamento de bancada e o portátil foram conduzidas separadamente e seguiram o mesmo procedimento.

Começando pela discriminação de mogno das demais espécies, os modelos PLS1-DA foram construídos, o pré-processamento foi selecionado, os *outliers* de calibração foram excluídos, os novos modelos foram construídos e os *outliers* de previsão foram excluídos da mesma forma descrita na seção anterior. Nesse caso, utilizou-se toda a região espectral do equipamento Phazir para a construção dos modelos.

O resultado da recalibração foi obtido pela aplicação do modelo construído no Phazir às amostras de previsão do mesmo equipamento.

O resultado da não transferência foi avaliado pela aplicação do modelo construído no equipamento de bancada ao conjunto de previsão do Phazir.

As amostras de transferência foram selecionadas utilizando o algoritmo KS, a partir do conjunto de calibração do equipamento de bancada, da mesma forma descrita na seção anterior.

Realizaram-se as padronizações RS e PRS otimizando o número de amostras de transferência entre 4 e 14 e tamanho da janela utilizando o conjunto de calibração do Phazir.

Uma vez construídos os modelos com parâmetros de transferência ótimos, estes foram aplicados às amostras de previsão do Phazir. A Figura 27 esquematiza o procedimento da PRS e a RS pode ser entendida de forma análoga.

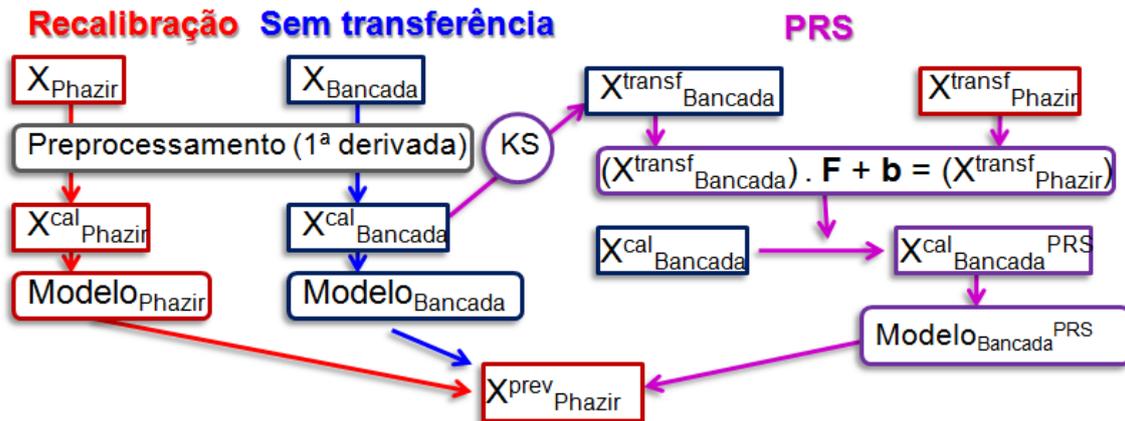


Figura 27. Fluxograma de etapas envolvidas na realização do método PDS e dos métodos de comparação sem transferência e de recalibração completa.

A RSDW foi calculada utilizando o mesmo número de amostras de transferência selecionado para os demais métodos. Calculou-se a média dos espectros das amostras de transferência em cada equipamento. Em seguida subtraiu-se a média do equipamento de bancada da média do Phazir. Essa diferença foi somada a cada espectro do conjunto de calibração do equipamento de bancada e o conjunto foi utilizado para construir um modelo de calibração que foi aplicado às amostras de previsão do Phazir.

O mesmo procedimento foi realizado para as demais espécies, andiroba cedro e curupixá, mantendo-se o mesmo conjunto de amostras de transferência selecionado para o mogno, excluindo-se os outliers de cada modelo.

Por fim construíram-se modelos de recalibração a partir de subconjuntos de calibração do Phazir para avaliar o mínimo de amostras necessárias de cada espécie para construir modelos de calibração com resultados comparáveis aos de transferência. Para isso, geraram-se números aleatórios no programa Microsoft Excel para selecionar amostras entre as 18, 20, 18 e 18 amostras de andiroba, cedro, curupixá e mogno, respectivamente, que compunham o conjunto de calibração. A ordem das amostras selecionadas pelo programa foi respeitada. Em seguida, construíram-se os modelos utilizando de 3 a 9 amostras de cada espécie, tomando as quatro replicatas de cada amostra.

5.3 RESULTADOS E DISCUSSÃO

Os resultados para a transferência entre equipamentos de bancada serão abordados na seção a seguir, e os resultados para a transferência envolvendo equipamentos de bancada e portátil serão abordados na seção subsequente.

5.3.1. Transferência direta entre equipamentos de bancada

O pré-processamento escolhido foi a primeira derivada e os dados foram centrados na média. A Figura 28 mostra os espectros dos equipamentos Tensor 37 e Antaris II derivados após a interpolação e o espectro derivado e padronizado do Antaris II.

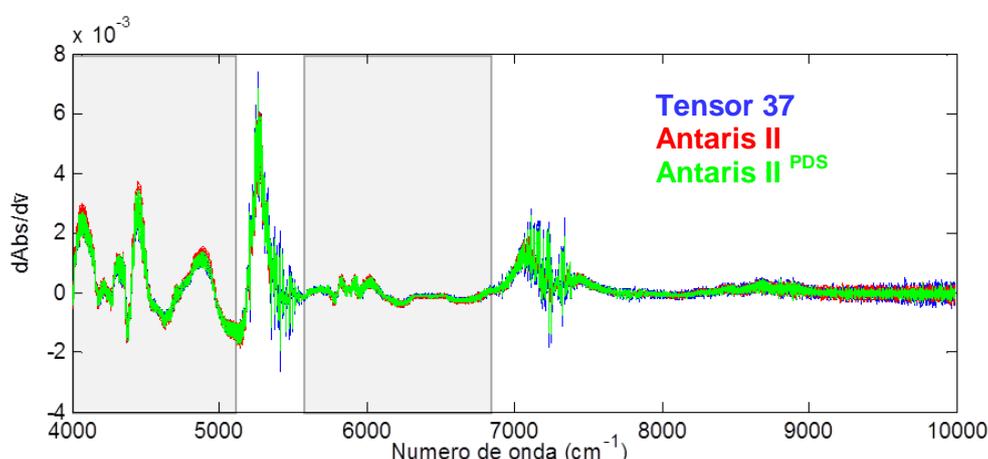


Figura 28. Derivada dos espectros dos equipamentos Tensor 37 e Antaris II e derivada dos espectros do Antaris II padronizada por PDS e regiões espectrais selecionadas para construção dos modelos (faixas sombreadas).

A transferência de calibração utilizando toda a região espectral não apresentou bom desempenho. Por isso, selecionaram-se visualmente as faixas de $4.000,4 \text{ cm}^{-1}$ a $5022,7 \text{ cm}^{-1}$ e de 5.559 cm^{-1} a $6.828,1 \text{ cm}^{-1}$, que apresentavam menores variação e ruído e construiu-se cada modelo utilizando apenas essas duas faixas.

As Tabelas 7 a 10 mostram os números VL, RMSEC e RMSECV dos modelos construídos no Antaris II, que foi utilizado para a recalibração e no Tensor 37 que foi aplicado ao conjunto de previsão do Antaris II não padronizado e padronizado por DS, PDS e SDW. As Figuras 29 a 32 mostram os valores de estimados para cada espécie e método de transferência.

5. TRANSFERÊNCIA DE CALIBRAÇÃO ENTRE EQUIPAMENTOS

Nos modelos para discriminação de andiroba, cedro, curupixá e mogno foram excluídos, respectivamente, 10, 8, 9 e 9 *outliers* de 296 medidas de 74 amostras de calibração e 22, 3, 19 e 15 de 148 medidas de 37 amostras de previsão. É importante destacar que a etapa de exclusão de *outliers* antecede a etapa de transferência, sendo que nenhuma amostra foi excluída após a realização da transferência.

Tabela 7. Número de variáveis latentes (VL), valores de erros, limite de discriminação e porcentagem de acerto dos modelos de recalibração (RC), sem transferência (ST), DS, PDS e SDW entre os equipamentos **Tensor 37** e **Antaris II** para discriminação das quatro espécies.

| Andiroba | Nº de VL | RMSEC | RMSECV | RMSEP | limite | % |
|----------------------------------|-----------------|--------------|---------------|--------------|---------------|----------|
| RC | 8 | 0,095 | 0,117 | 0,118 | 0,460 | 100 |
| ST | | | | 0,177 | | 100 |
| DS (n12)^a | | | | 0,217 | | 98,4 |
| PDS (n5;j3)^{a,b} | 7 | 0,087 | 0,111 | 0,168 | 0,501 | 100 |
| SDW(n5)^a | | | | 0,164 | | 100 |
| SDW(n12)^a | | | | 0,165 | | 100 |
| Cedro | Nº de VL | RMSEC | RMSECV | RMSEP | limite | % |
| RC | 5 | 0,087 | 0,094 | 0,101 | 0,388 | 100 |
| ST | | | | 0,109 | | 100 |
| DS (n12)^a | | | | 0,131 | | 100 |
| PDS (n5;j1)^{a,b} | 5 | 0,087 | 0,096 | 0,115 | 0,425 | 100 |
| SDW(n5)^a | | | | 0,115 | | 100 |
| SDW(n12)^a | | | | 0,116 | | 100 |
| Curupixá | Nº de VL | RMSEC | RMSECV | RMSEP | limite | % |
| RC | 8 | 0,045 | 0,054 | 0,053 | 0,400 | 100 |
| ST | | | | 0,101 | | 100 |
| DS (n12)^a | | | | 0,104 | | 100 |
| PDS (n5;j1)^{a,b} | 8 | 0,041 | 0,055 | 0,079 | 0,415 | 100 |
| SDW(n5)^a | | | | 0,065 | | 100 |
| SDW(n12)^a | | | | 0,067 | | 100 |
| Mogno | Nº de VL | RMSEC | RMSECV | RMSEP | limite | % |
| RC | 5 | 0,123 | 0,130 | 0,126 | 0,491 | 100 |
| ST | | | | 0,146 | | 100 |
| DS (n12)^a | | | | 0,144 | | 100 |
| PDS (n5;j1)^{a,b} | 5 | 0,115 | 0,125 | 0,122 | 0,472 | 100 |
| SDW(n5)^a | | | | 0,132 | | 100 |
| SDW(n12)^a | | | | 0,130 | | 100 |

^a n corresponde ao número de amostras de transferência

^b j corresponde ao número de pontos da janela de PDS

5. TRANSFERÊNCIA DE CALIBRAÇÃO ENTRE EQUIPAMENTOS

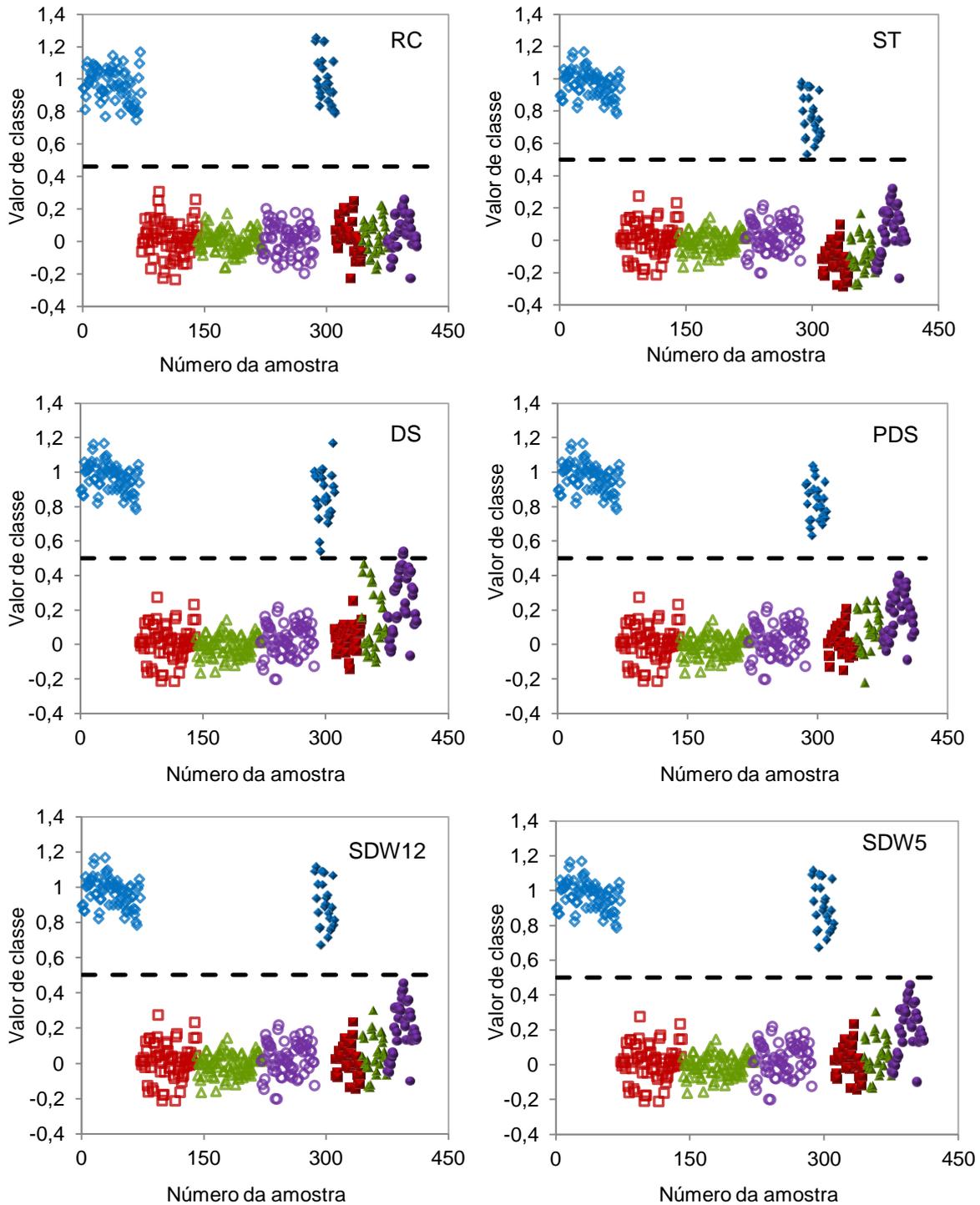


Figura 29. Valores de classe das amostras estimados na transferência entre os equipamentos **Tensor 37** e **Antaris II** pelos métodos recalibração (RC); sem transferência (ST); DS, PDS e SDW com 12 (SDW12) e 5 (SDW5) amostras de transferência (SDW5) para discriminação de **andiroba** (\diamond) das espécies cedro (\square), curupixá (Δ) e mogno (\circ) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão.

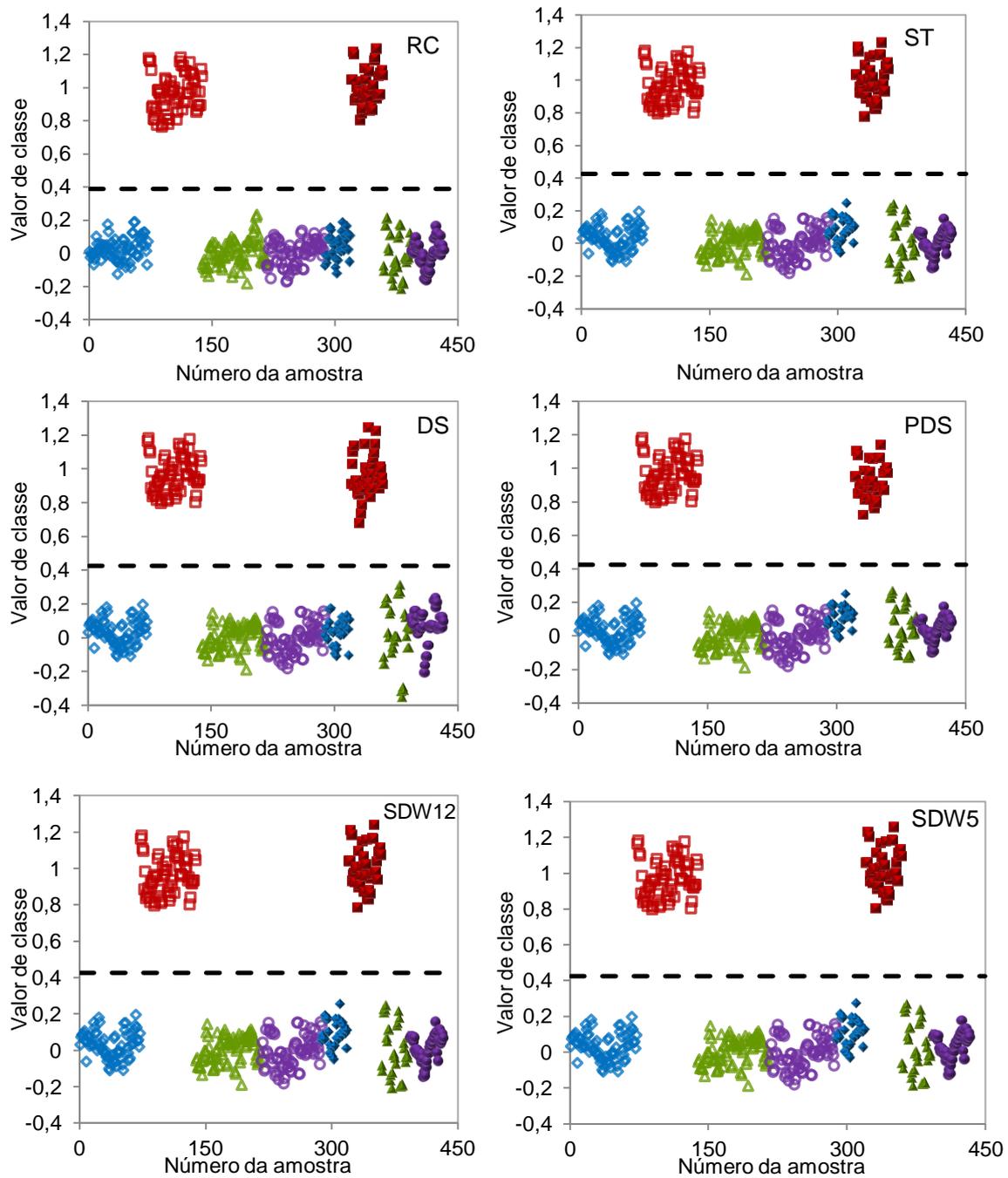


Figura 30. Valores de classe das amostras estimados na transferência entre os equipamentos **Tensor 37** e **Antaris II** pelos métodos recalibração (RC); sem transferência (ST); DS, PDS e DS com 12 (SDW12) e 5 (SDW5) amostras de transferência (SDW5) para discriminação de **cedro** (□) das espécies andiroba (◇), curupixá (△) e mogno (○) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão.

5. TRANSFERÊNCIA DE CALIBRAÇÃO ENTRE EQUIPAMENTOS

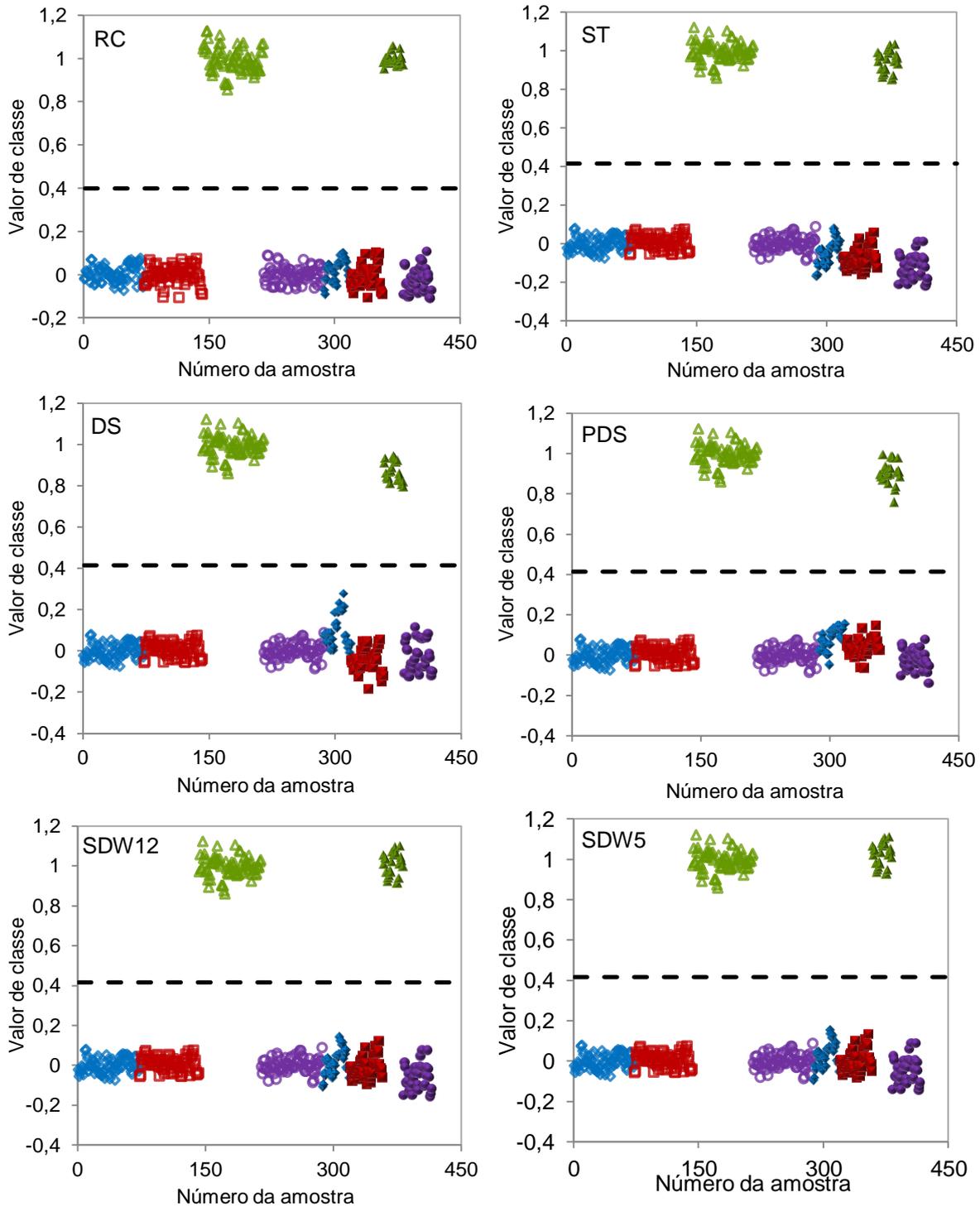


Figura 31. Valores de classe das amostras estimados na transferência entre os equipamentos **Tensor 37** e **Antaris II** pelos métodos recalibração (RC); sem transferência (ST); DS, PDS e SDW com 12 (SDW12) e 5 (SDW5) amostras de transferência (SDW5) para discriminação de **curupixá** (Δ) das espécies andiroba (\diamond), cedro (\square) e mogno (\circ) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão.

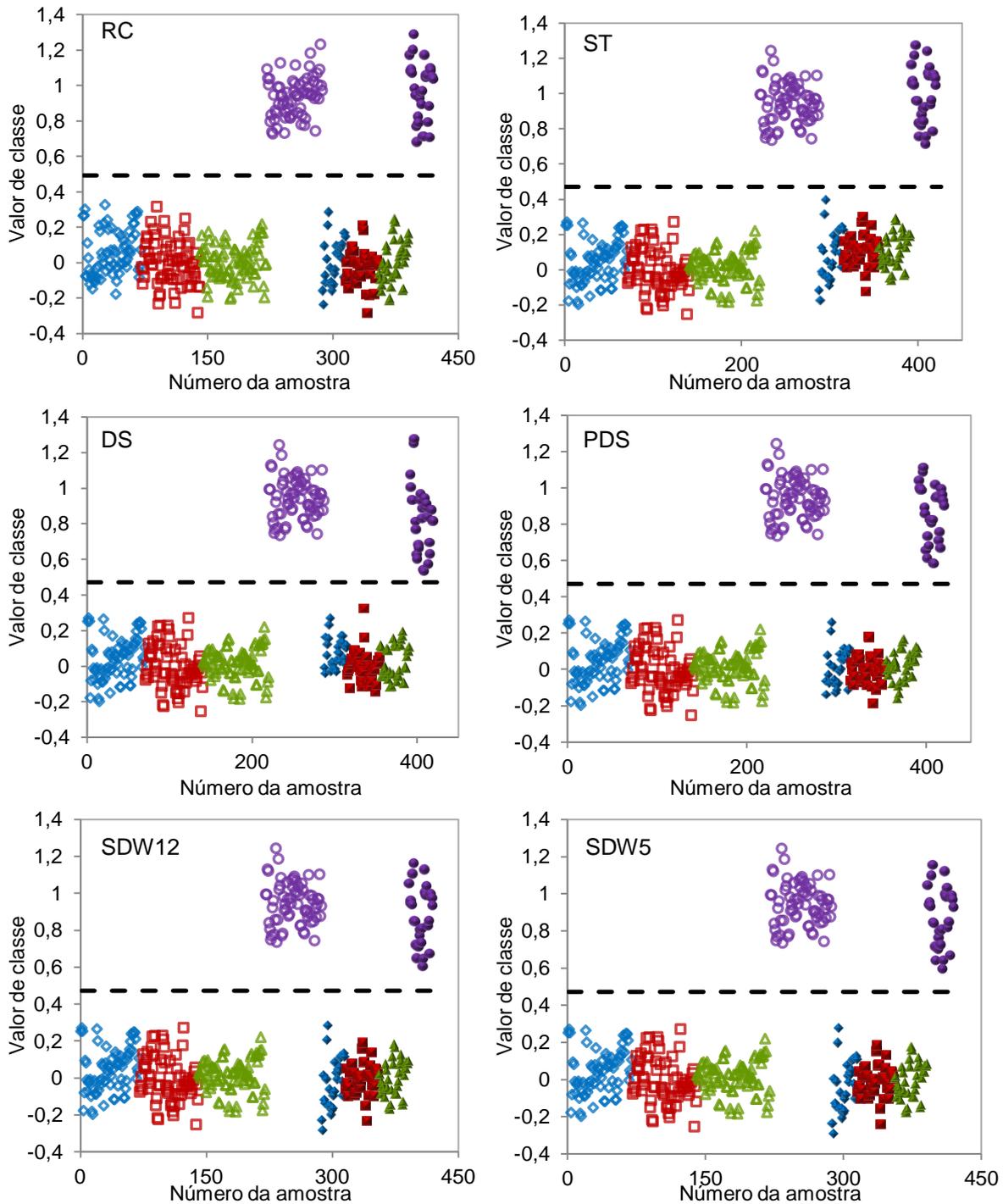


Figura 32. Valores de classe das amostras estimados na transferência entre os equipamentos **Tensor 37** e **Antaris II** pelos métodos recalibração (RC); sem transferência (ST); DS, PDS e SDW com 12 (SDW12) e 5 (SDW5) amostras de transferência (SDW5) para discriminação de **mogno** (○) das espécies **andiroba** (◇), **cedro** (□) e **curupixá** (△) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão.

O número de amostras de transferência utilizadas na PDS foi 5, que corresponde ao mínimo necessário para que o algoritmo selecionasse uma amostra de cada espécie. As janelas foram de 3 pontos para andiroba, curupixá e mogno e 1 ponto para cedro.

O método DS não apresentou bons resultados. Mesmo com 12 amostras de transferência, os valores RMSEP para DS foram superiores aos da não transferência para todas as espécies, à exceção do mogno.

O método de SDW foi realizado com o mesmo número de amostras de transferência que as padronizações DS e PDS para possibilitar a comparação. À exceção do mogno, SDW com 5 amostras de transferência teve melhores resultados que SDW com 12 amostras de transferência. Seria esperado uma melhora de resultado com o aumento do número de amostras de transferência, uma vez que haveria uma maior representatividade dos dados do equipamento S pelas amostras de transferência. Entretanto, é possível que as 7 amostras a mais no conjunto de 12 não contribuam com mais informações relevantes referentes à diferença dos equipamentos. Esse resultado ressalta a importância do critério de escolha das amostras de transferência.

Como se pode observar nas Tabela 7 e nas Figuras 29 a 32, a aplicação direta do modelo do Tensor 37 ao conjunto de previsão do Antaris II sem transferência resultou em 100% de classificações corretas em todos os casos, apresentando inclusive menor valor de RMSEP para a discriminação de cedro com relação aos métodos de transferência. Esse resultado indica que a resposta dos dois equipamentos é bastante semelhante nas faixas espectrais selecionadas, sendo essa seleção suficiente para a aplicação do modelo primário aos espectros secundários não padronizados.

Na discriminação de andiroba na Figura 29 observa-se um viés na dispersão quanto ao valor de classe estimado para as amostras de previsão sem transferência (ST). Enquanto as amostras de calibração de andiroba estão dispersas em torno do valor de classe 1, as amostras de previsão dessa espécie estão deslocadas para baixo, em torno de aproximadamente 0,7. E enquanto as amostras de calibração das demais espécies estão dispersas em torno do valor de classe 0, as amostras de previsão de cedro e curupixá estão

dispersas em torno de -0,1. Os métodos de transferência, principalmente por SDW, corrigem esse viés, entretanto observa-se que após a transferência as amostras de previsão de mogno ficam mais deslocadas para cima.

Na discriminação de mogno, Figura 32, a dispersão das amostras de previsão sem transferência (ST) de mogno não apresentam viés, entretanto as amostras de previsão das demais espécies encontram-se deslocadas para cima (em torno de aproximadamente 0,1). Os métodos de transferência corrigem o viés das amostras das demais espécies, porém deslocam para baixo também as amostras de mogno, efeito mais pronunciado na transferência por DS.

O método PDS de forma geral apresentou os melhores resultados, possibilitando a redução do RMSEP e uma melhora na dispersão das amostras com relação à situação sem transferência para todas as espécies à exceção do cedro, sendo superado pelo método SDW apenas para o curupixá. Observa-se também que todos os métodos resultaram em 100% de classificações corretas exceto a padronização DS para cedro.

5.3.2. Transferência reversa entre equipamentos de bancada e portátil

O pré-processamento selecionado para todos os métodos que serão discutidos a seguir foi a primeira derivada. Esse pré-processamento ajuda a corrigir desvios aditivos da linha de base presente nos espectros de cada equipamento e também diminui a diferença dos espectros entre equipamentos, como pode ser observado nas Figuras 33 e 38. Os resultados de cada transferência serão mostrados nas seções seguintes.

5.3.2.1. Transferência entre Antaris II e Phazir e recalibração com subconjunto

A Figura 33 mostra os espectros dos dois equipamentos antes e depois da interpolação e derivação.

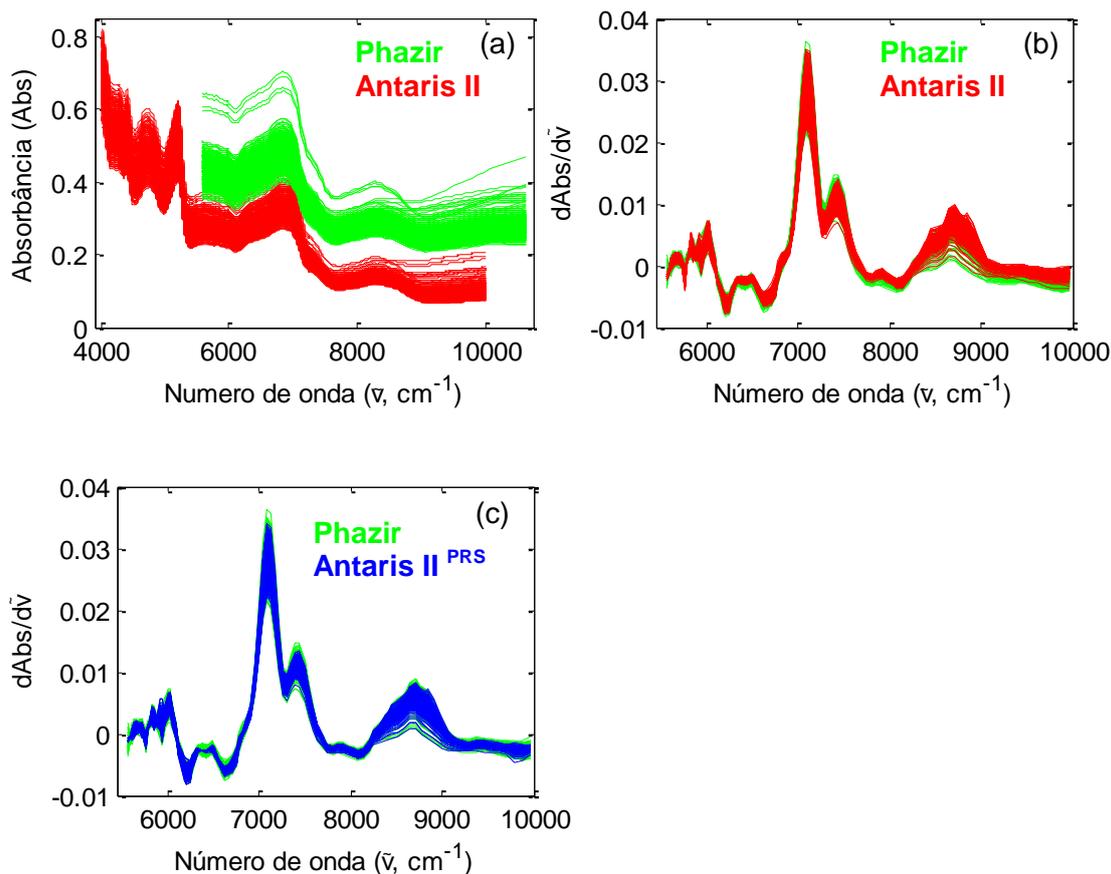


Figura 33. Espectros dos equipamentos Antaris II e Phazir a) originais b) após interpolação e derivação e c) após interpolação, derivação e PRS.

A derivada corrige os desvios aditivos de linha de base até aproximadamente 8000 cm^{-1} , a partir de onde os espectros apresentam maior diferença de natureza multiplicativa, ou seja, de inclinação. Entretanto, os resultados obtidos com esse processamento foram melhores que os obtidos com outros pré-processamentos que corrigem desvios multiplicativos dos sinais como MSC e SNV.

Os resultados para a transferência entre os equipamentos Antaris II e Phazir e para a recalibração utilizando um subconjunto das amostras de calibração estão organizados na Tabela 8 e as Figuras 33 a 36 mostram os valores de classe estimados para as amostras a partir dos modelos obtidos pelos diferentes métodos.

Observa-se que todos os métodos de transferência possibilitaram a melhora dos resultados quanto ao valor de RMSEP e a porcentagem de classificações corretas com relação à situação sem transferência, onde o

modelo não padronizado do equipamento Antaris II é aplicado diretamente às amostras do Phazir.

A recalibração apresentou 100% de classificações corretas para todas as amostras, assim como a recalibração a partir do subconjunto de amostras, embora os erros associados a esse segundo método sejam ligeiramente maiores que os do primeiro.

Foi necessário um subconjunto de no mínimo 28 amostras para que o desempenho do modelo fosse comparável aos modelos transferidos, uma vez que ao utilizar menos amostras de cada espécie os erros médios de previsão foram elevados e cerca de 20 a 60% das amostras de previsão foram consideradas *outliers* indicando a falta de capacidade do modelo em estimar corretamente o valor de classe de novas amostras.

O método PRS foi, de forma geral, o melhor para a transferência entre os dois equipamentos. PRS apresentou menores valores de RMSEP para andiroba, curupixá. Porém RSDW e RS superaram a porcentagem de PRS para cedro, apresentando 100% de classificações corretas. Embora o valor de RMSEP do método RS tenha sido menor que o do PRS para mogno, a porcentagem de classificações corretas do PRS foi superior. O método RSDW forneceu melhores resultados em comparação a RS, pois apresentou porcentagens de classificações corretas superiores ou igual, no caso do cedro, embora seus valores de RMSEP tenham sido maiores para curupixá e mogno.

Embora a recalibração com subconjunto tenha obtido melhores resultados, com menores valores de RMSEP e 100% de classificações corretas para todas as espécies, a diferença entre o número de amostras necessárias para esse método (28 amostras) e para a transferência (8 amostras) pode fazer com que a transferência seja preferível em situações onde não é possível ou viável medir muitas amostras do conjunto de calibração original no novo equipamento.

5. TRANSFERÊNCIA DE CALIBRAÇÃO ENTRE EQUIPAMENTOS

Tabela 8. Número de variáveis latentes (VL), erros, limite de discriminação, porcentagem de acerto e número outliers de calibração (outc) e previsão (outp) dos modelos de recalibração (RC), sem transferência (ST), RS, PRS, RSDW entre os equipamentos **Antaris II** e **Phazir** e recalibração com subconjunto (SubC) para discriminação das quatro espécies.

| Andiroba | VL | RMSEC | RMSECV | RMSEP | limite | % | outc | outp |
|----------------------------------|-----------|--------------|---------------|--------------|---------------|----------|-------------|-------------|
| RC | 12 | 0,142 | 0,181 | 0,193 | 0,475 | 100% | 6 | 20 |
| ST | 11 | 0,120 | 0,133 | 0,434 | 0,493 | 82,8% | 6 | 20 |
| RS (n8)^a | 6 | 0,151 | 0,162 | 0,344 | 0,457 | 86,7% | 6 | 20 |
| PRS (n8;j9)^{a,b} | 7 | 0,135 | 0,148 | 0,285 | 0,472 | 93,8% | 6 | 20 |
| RSDW(n8)^a | 8 | 0,142 | 0,155 | 0,277 | 0,497 | 93,0% | 6 | 20 |
| SubC(n28)^c | 9 | 0,154 | 0,187 | 0,221 | 0,511 | 100% | 0 | 17 |
| Cedro | VL | RMSEC | RMSECV | RMSEP | limite | % | outc | outp |
| RC | 6 | 0,142 | 0,151 | 0,156 | 0,419 | 100% | 3 | 15 |
| ST | 10 | 0,088 | 0,096 | 0,360 | 0,430 | 63,2% | 3 | 15 |
| RS (n8)^a | 7 | 0,103 | 0,112 | 0,200 | 0,419 | 100% | 3 | 15 |
| PRS (n8;j3)^{a,b} | 4 | 0,148 | 0,153 | 0,168 | 0,407 | 99,2% | 3 | 15 |
| RSDW(n8)^a | 10 | 0,088 | 0,096 | 0,186 | 0,430 | 100% | 3 | 15 |
| SubC(n28)^c | 5 | 0,143 | 0,163 | 0,175 | 0,357 | 100% | 1 | 2 |
| Curupixá | VL | RMSEC | RMSECV | RMSEP | limite | % | outc | outp |
| RC | 7 | 0,117 | 0,132 | 0,118 | 0,414 | 100,0% | 9 | 19 |
| ST | 11 | 0,061 | 0,067 | 0,344 | 0,391 | 76,2% | 9 | 19 |
| RS (n8)^a | 8 | 0,079 | 0,088 | 0,220 | 0,375 | 90,0% | 9 | 19 |
| PRS (n8;j1)^{a,b} | 9 | 0,068 | 0,075 | 0,196 | 0,382 | 95,4% | 9 | 19 |
| RSDW(n8)^a | 9 | 0,065 | 0,071 | 0,234 | 0,392 | 93,1% | 9 | 19 |
| SubC(n28)^c | 10 | 0,091 | 0,134 | 0,126 | 0,452 | 100% | 1 | 18 |
| Mogno | VL | RMSEC | RMSECV | RMSEP | limite | % | outc | outp |
| RC | 8 | 0,145 | 0,167 | 0,147 | 0,441 | 100,0% | 6 | 7 |
| ST | 8 | 0,125 | 0,133 | 0,484 | 0,511 | 85,1% | 6 | 7 |
| RS (n8)^a | 6 | 0,135 | 0,143 | 0,261 | 0,481 | 94,3% | 6 | 7 |
| PRS (n8;j1)^{a,b} | 8 | 0,135 | 0,145 | 0,277 | 0,486 | 97,9% | 6 | 7 |
| RSDW(n8)^a | 7 | 0,132 | 0,142 | 0,288 | 0,495 | 95,7% | 6 | 7 |
| SubC(n28)^c | 9 | 0,113 | 0,158 | 0,151 | 0,472 | 100% | 2 | 15 |

^a n corresponde ao número de amostras de transferência

^b j corresponde ao número de pontos da janela da PRS

^c n corresponde ao número de amostras do subconjunto de calibração

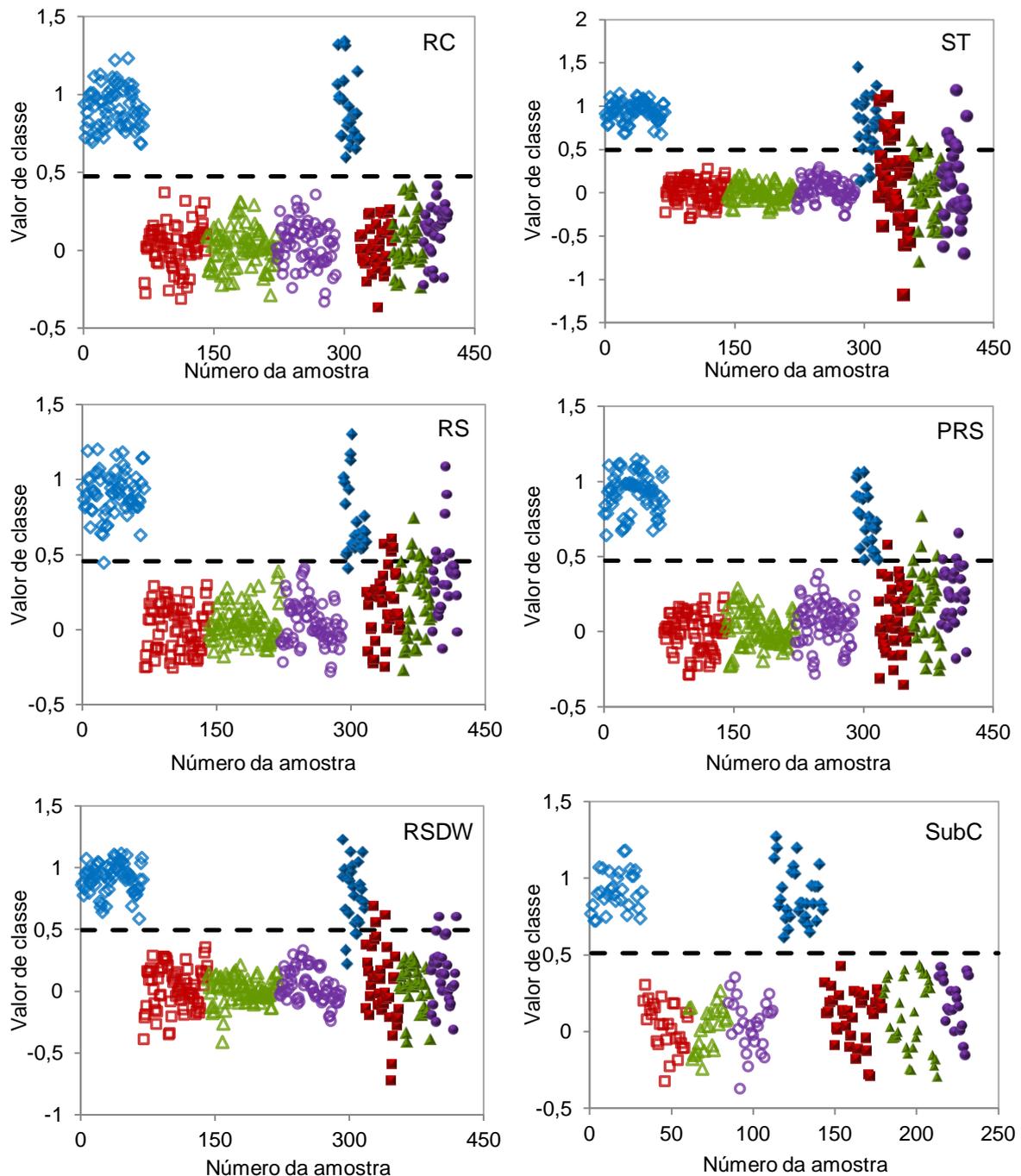


Figura 34. Valores de classe das amostras estimados na transferência entre os equipamentos **Antaris II** e **Phazir** pelos métodos recalibração (RC); sem transferência (ST); RS, PRS, RSDW e recalibração com subconjunto (SubC) para discriminação de **andiroba** (\diamond) das espécies cedro (\square), curupixá (\triangle) e mogno (\circ) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão.

5. TRANSFERÊNCIA DE CALIBRAÇÃO ENTRE EQUIPAMENTOS

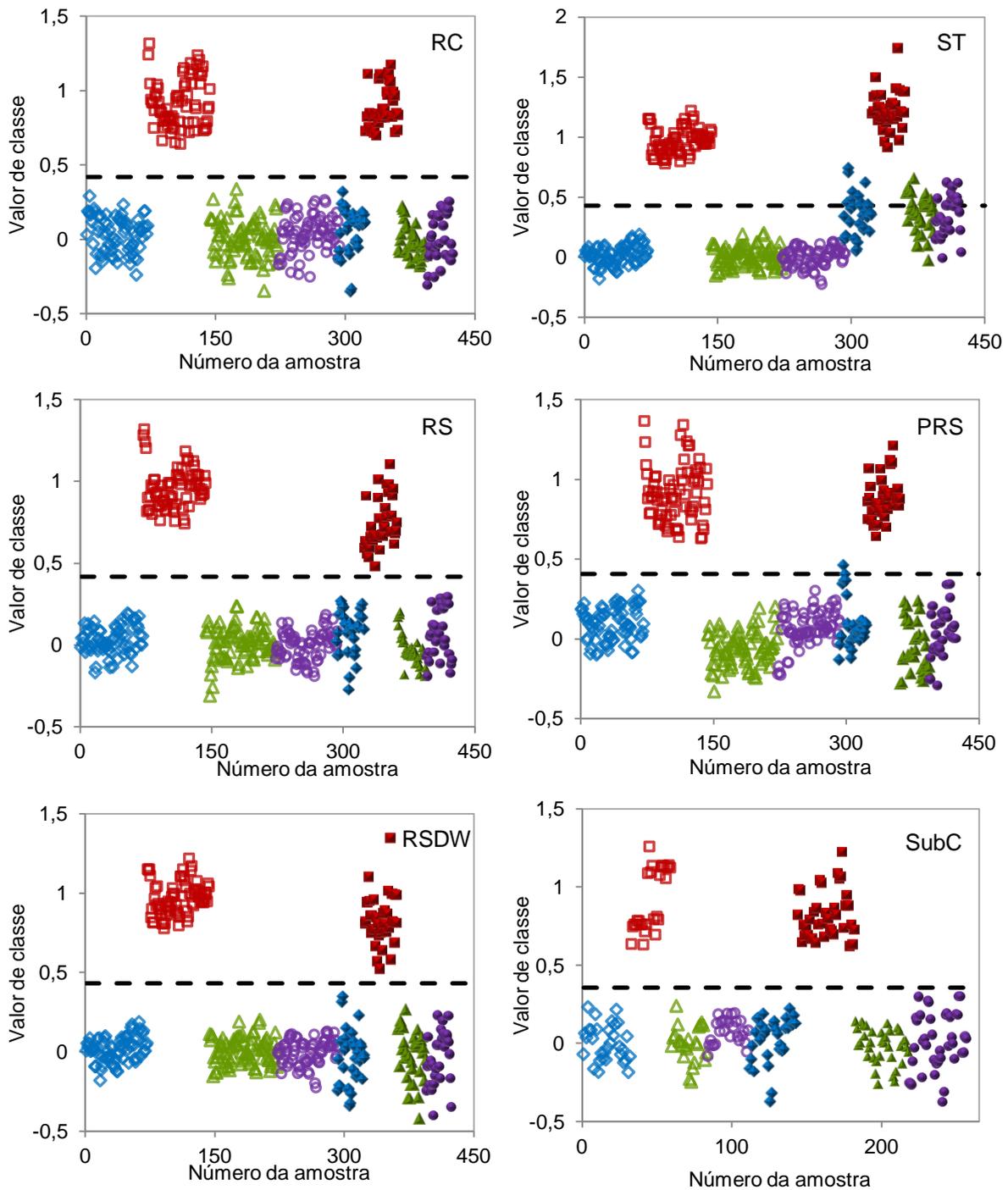


Figura 35. Valores de classe das amostras estimados na transferência entre os equipamentos **Antaris II** e **Phazir** pelos métodos recalibração (RC); sem transferência (ST); RS, PRS, RSDW e recalibração com subconjunto (SubC) para discriminação de **cedro** (□) das espécies andiroba (◇), curupixá (△) e mogno (○) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão.

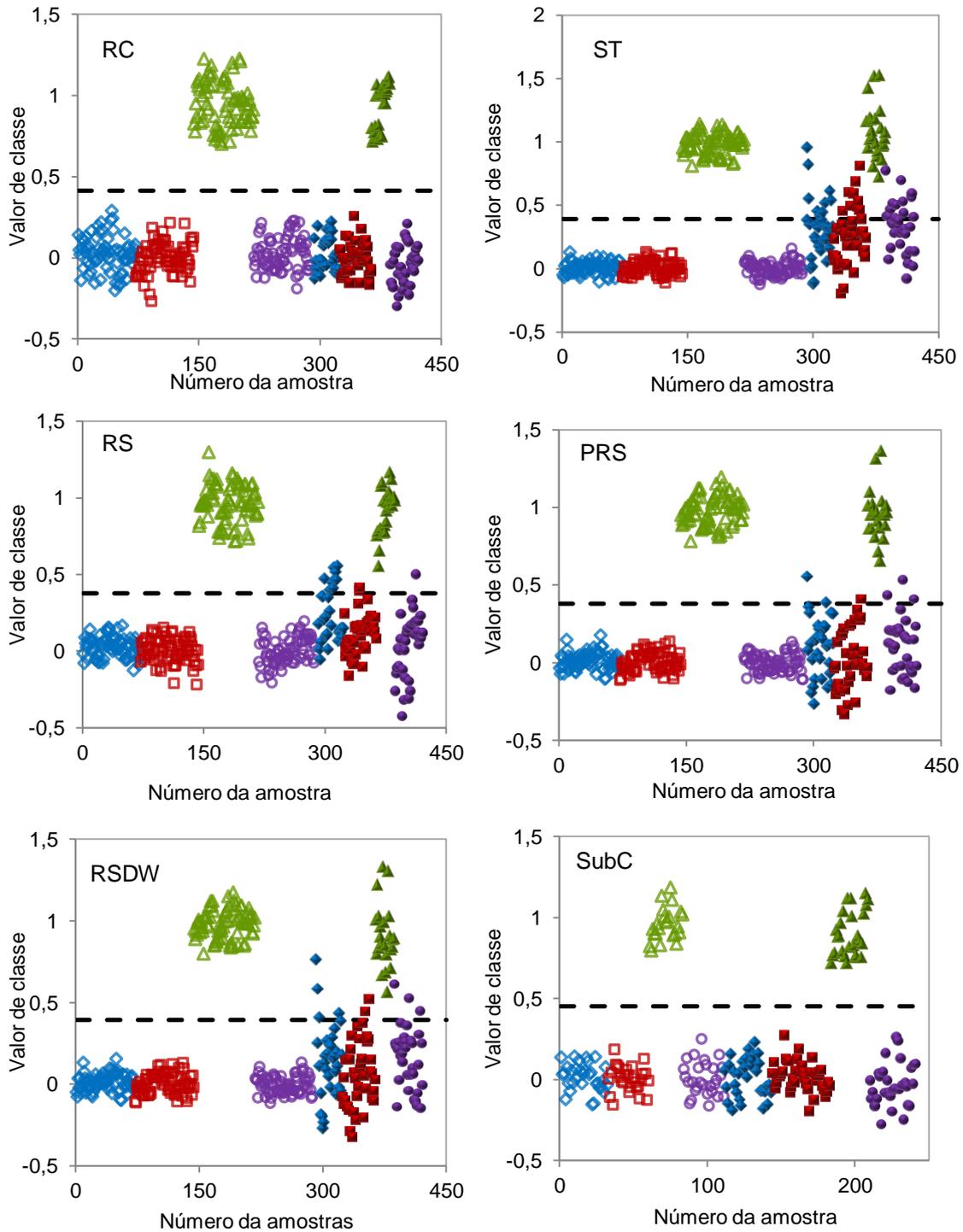


Figura 36. Valores de classe das amostras estimados na transferência entre os equipamentos **Antaris II** e **Phazir** pelos métodos recalibração (RC); sem transferência (ST); RS, PRS, RSDW e recalibração com subconjunto (SubC) para discriminação de **curupixá** (Δ) das espécies andiroba (\diamond), cedro (\square) e mogno (\circ) e limites discriminação (- -). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão.

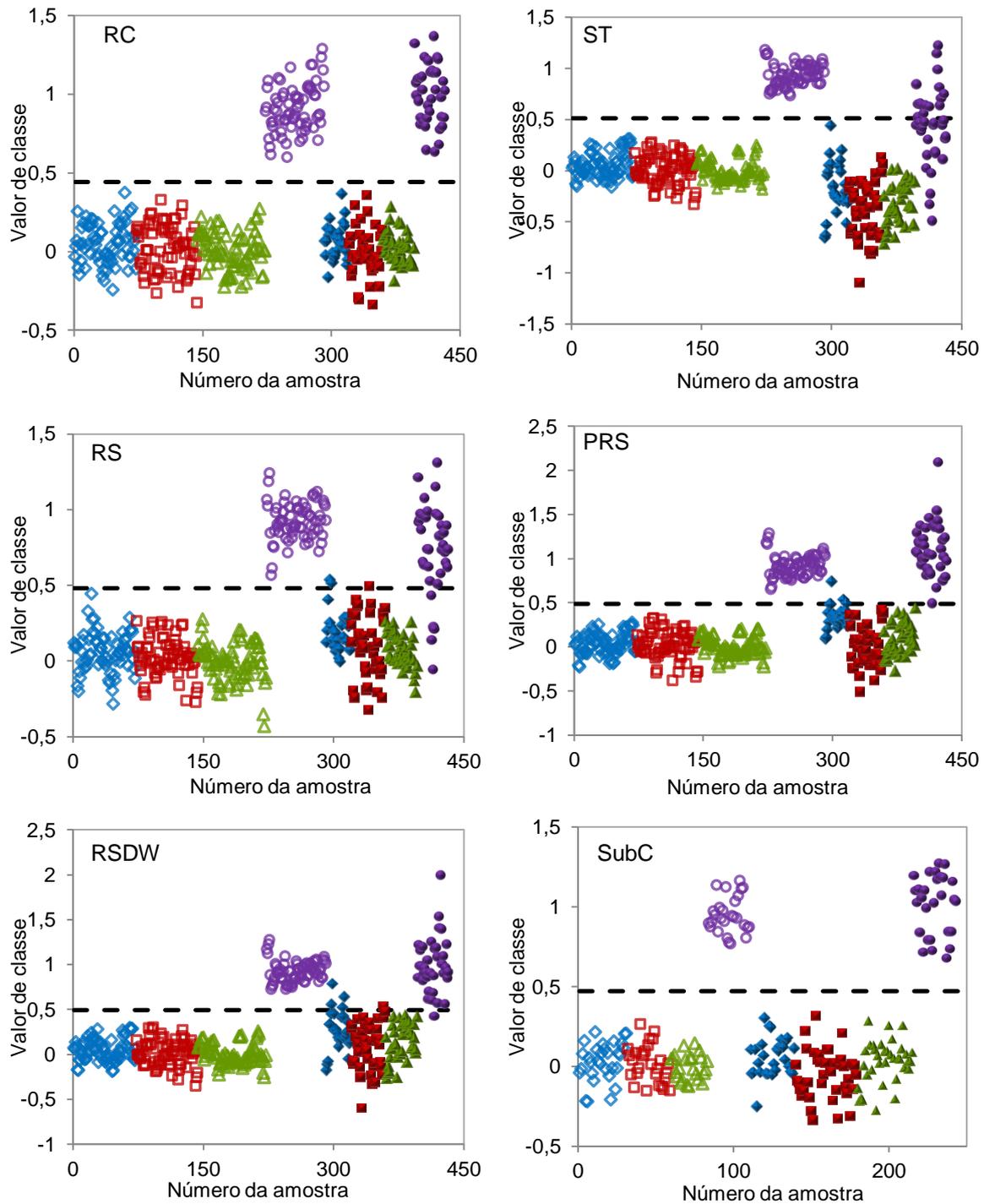


Figura 37. Valores de classe das amostras estimados na transferência entre os equipamentos **Antaris II** e **Phazir** pelos métodos recalibração (RC); sem transferência (ST); RS, PRS, RSDW e recalibração com subconjunto (SubC) para discriminação de **mogno** (○) das espécies andiroba (◇), cedro (□) e curupixá (△) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão.

5.3.2.2. Transferência entre Tensor 37 e Phazir

A Figura 38 mostra os espectros dos equipamentos Tensor 37 e Phazir antes e após a interpolação e derivação.

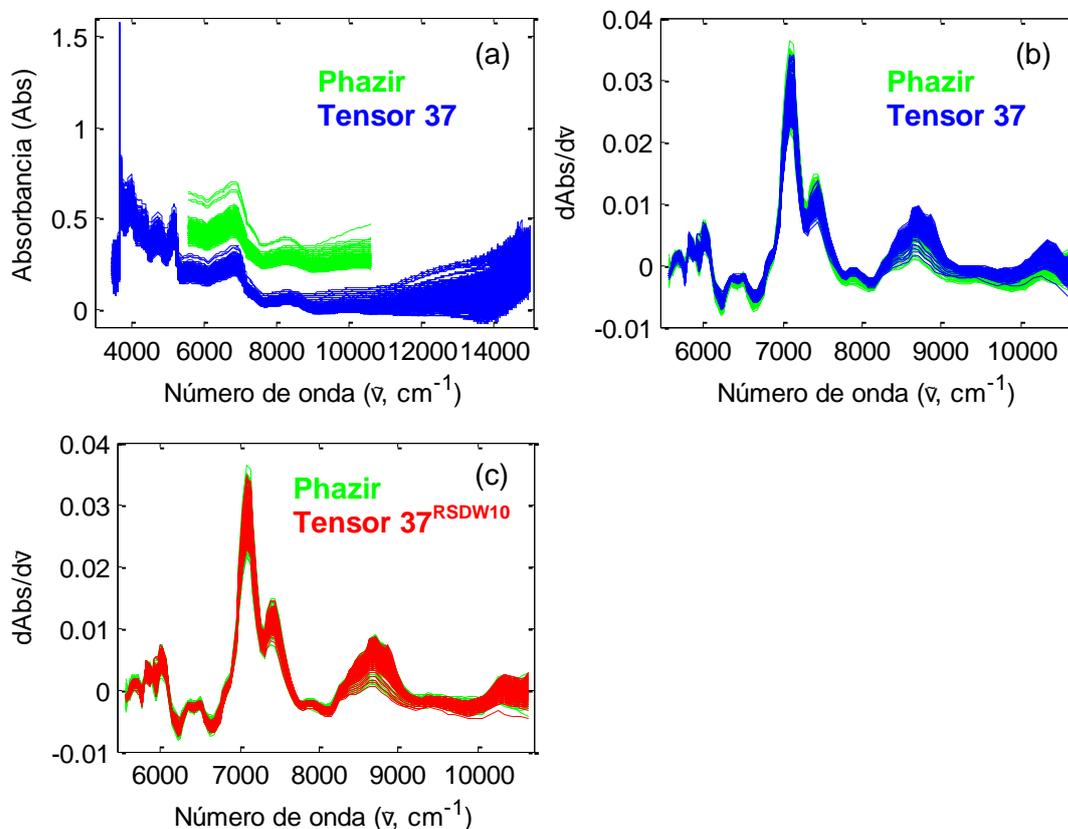


Figura 38. Espectros dos equipamentos Tensor 37 e Phazir a) originais, b) após interpolação e derivação e c) após interpolação, derivação e RSDW com 10 amostras de transferência.

Dada a faixa mais reduzida do equipamento de campo, a região compatível entre os dois equipamentos exclui as extremidades do espectro do equipamento de bancada que apresentam mais variações e ruído. A derivada corrige os desvios aditivos, e, embora novamente os desvios multiplicativos, sobretudo a partir de 8.000 cm^{-1} , aproximadamente, não sejam corrigidos, os resultados obtidos com esse processamento foram melhores que os obtidos com outros pré-processamentos como MSC e SNV.

Os resultados para a transferência entre os equipamentos Tensor 37 e Phazir estão organizados na Tabela 9 e as Figuras 39 a 42 mostram os valores de classe estimados para as amostras a partir dos modelos obtidos pelos diferentes métodos.

Observa-se uma melhora dos resultados quanto ao valor de RMSEP e a porcentagem de classificações corretas dos métodos de transferência com relação à situação sem transferência, onde o modelo não padronizado do equipamento Tensor 37 é aplicado diretamente às amostras do Phazir. A única exceção é o método PRS para andiroba, cuja porcentagem de classificações corretas é igual à da situação sem transferência.

A recalibração apresentou 100% de classificações corretas para todas as amostras.

O método de transferência que apresentou os melhores resultados variou para cada espécie. No caso da andiroba, RS teve os melhores resultados com relação ao valor de RMSEP e porcentagem de classificações corretas. Para o cedro, todos os métodos de transferência obtiveram 100% de classificações corretas e PRS apresentou o menor valor de RMSEP. Para curupixá, PRS e SDW apresentaram 100% de classificações corretas e PRS novamente apresentou o menor valor de RMSEP. Por último, para mogno, embora RS tenha apresentado o menor valor de RMSEP, SDW com 10 amostras de transferência apresentou a maior porcentagem de classificações corretas.

A diferença entre os modelos e resultados de SDW com 10 e 14 amostras de transferência é bastante pequena, como se pode observar na Tabela 9 e nas Figuras 39 a 42. Essa semelhança indica que a diferença da média das 10 amostras de transferência medidas em cada equipamento é muito próxima daquela calculada a partir de 14 amostras de transferência e novamente mostra que a identidade das amostras escolhidas possivelmente tem maior importância que a quantidade de amostras. SDW com 14 amostras de transferência apresentou resultados melhores para andiroba e cedro, equiparáveis para curupixá e piores para mogno.

Os resultados de transferência podem ser comparados também aos resultados de recalibração a partir do subconjunto expostos na seção anterior. Novamente, embora a recalibração com subconjunto tenha apresentado menores valores de RMSEP e 100% de classificações corretas para todas as espécies, a diferença entre o número de amostras necessárias para esse

5. TRANSFERÊNCIA DE CALIBRAÇÃO ENTRE EQUIPAMENTOS

método (28 amostras) e para a transferência (10 ou 14 amostras) pode fazer com que a transferência seja preferível em certas situações.

Tabela 9. Número de variáveis latentes (VL), erros, limite de discriminação, porcentagem de acerto e número outliers de calibração (outc) e previsão (outp) dos modelos de recalibração (RC), sem transferência (ST), RS, PRS e RSDW entre os equipamentos **Tensor 37** e **Phazir**.

| Andiroba | VL | RMSEC | RMSECV | RMSEP | limite | % | outc | outp |
|------------------------------------|-----------|--------------|---------------|--------------|---------------|----------|-------------|-------------|
| RC | 6 | 0,182 | 0,197 | 0,218 | 0,480 | 100,0% | 7 | 19 |
| ST | 12 | 0,105 | 0,122 | 0,561 | 0,495 | 82,2% | 7 | 19 |
| RS (n14)^a | 8 | 0,167 | 0,183 | 0,263 | 0,510 | 94,6% | 7 | 19 |
| PRS (n10;j9)^{a,b} | 6 | 0,153 | 0,163 | 0,346 | 0,435 | 82,2% | 7 | 19 |
| RSDW(n10)^a | 10 | 0,115 | 0,129 | 0,326 | 0,476 | 88,4% | 7 | 19 |
| RSDW(n14)^a | 10 | 0,115 | 0,129 | 0,315 | 0,476 | 89,1% | 7 | 19 |
| Cedro | VL | RMSEC | RMSECV | RMSEP | limite | % | outc | outp |
| RC | 6 | 0,142 | 0,153 | 0,159 | 0,418 | 100,0% | 2 | 11 |
| ST | 8 | 0,113 | 0,122 | 0,365 | 0,397 | 68,6% | 2 | 11 |
| RS (n14)^a | 7 | 0,118 | 0,128 | 0,220 | 0,438 | 100,0% | 2 | 11 |
| PRS (n10;j11)^{a,b} | 6 | 0,120 | 0,128 | 0,155 | 0,444 | 100,0% | 2 | 11 |
| RSDW(n10)^a | 8 | 0,113 | 0,122 | 0,164 | 0,397 | 100,0% | 2 | 11 |
| RSDW(n14)^a | 8 | 0,113 | 0,122 | 0,163 | 0,397 | 100,0% | 2 | 11 |
| Curupixá | VL | RMSEC | RMSECV | RMSEP | limite | % | outc | outp |
| RC | 9 | 0,1100 | 0,127 | 0,130 | 0,408 | 100,0% | 5 | 15 |
| ST | 11 | 0,0668 | 0,077 | 0,202 | 0,444 | 96,2% | 5 | 15 |
| RS (n14)^a | 12 | 0,0791 | 0,092 | 0,189 | 0,409 | 99,2% | 5 | 15 |
| PRS (n10;j7)^{a,b} | 7 | 0,1025 | 0,112 | 0,156 | 0,446 | 100,0% | 5 | 15 |
| RSDW(n10)^a | 6 | 0,1185 | 0,127 | 0,161 | 0,421 | 100,0% | 5 | 15 |
| RSDW(n14)^a | 6 | 0,1185 | 0,127 | 0,161 | 0,421 | 100,0% | 5 | 15 |
| Mogno | VL | RMSEC | RMSECV | RMSEP | limite | % | outc | outp |
| RC | 12 | 0,125 | 0,156 | 0,136 | 0,444 | 100,0% | 5 | 13 |
| ST | 12 | 0,092 | 0,108 | 0,546 | 0,496 | 87,4% | 5 | 13 |
| RS (n14)^a | 15 | 0,104 | 0,121 | 0,212 | 0,483 | 97,0% | 5 | 13 |
| PRS (n10;j5)^{a,b} | 6 | 0,152 | 0,162 | 0,229 | 0,434 | 97,0% | 5 | 13 |
| RSDW(n10)^a | 7 | 0,149 | 0,161 | 0,222 | 0,468 | 98,5% | 5 | 13 |
| RSDW(n14)^a | 7 | 0,149 | 0,161 | 0,227 | 0,468 | 97,0% | 5 | 13 |

^a n corresponde ao número de amostras de transferência

^b j corresponde ao número de pontos da janela da PRS

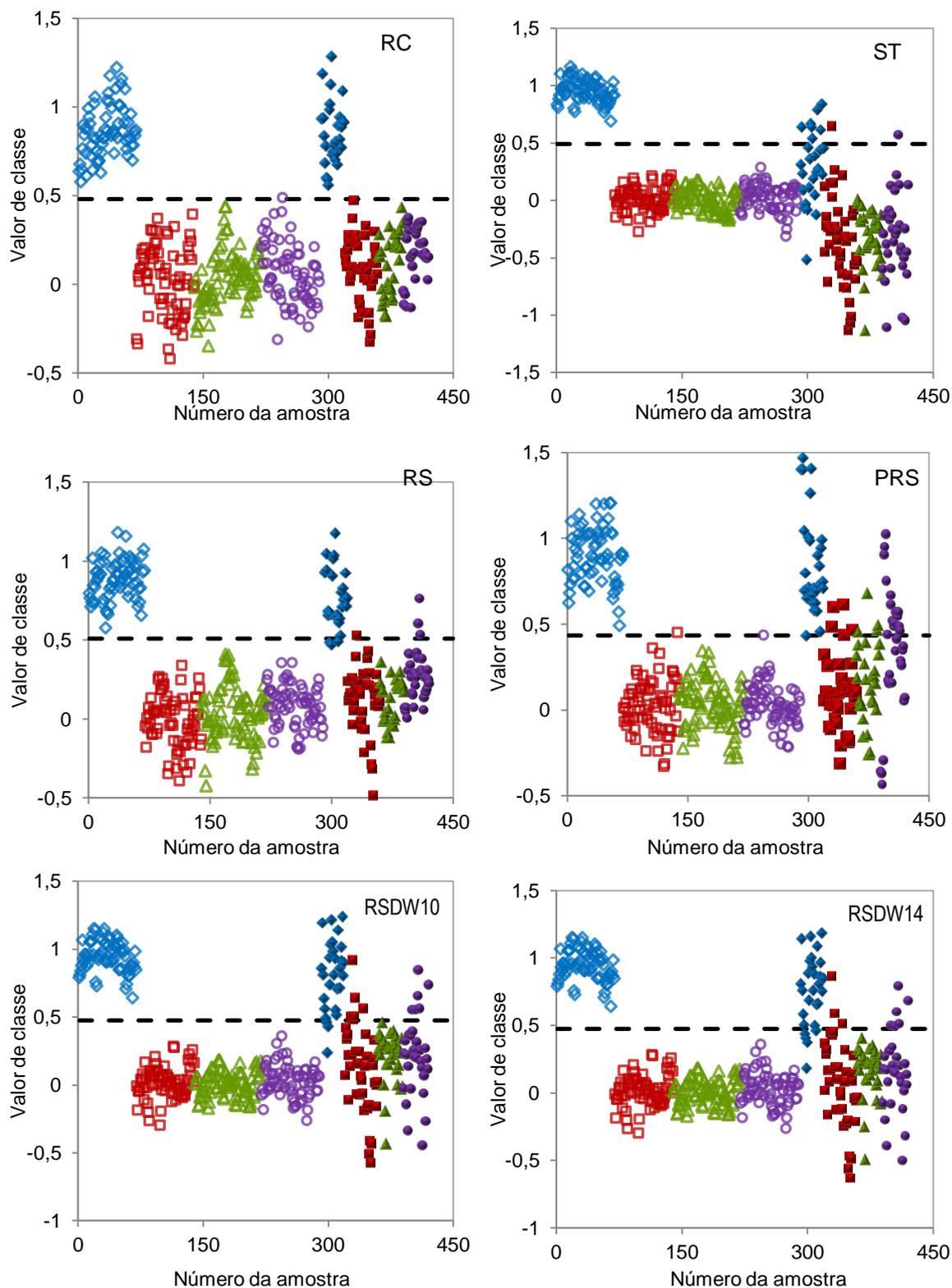


Figura 39. Valores de classe das amostras estimados na transferência entre os equipamentos **Tensor 37** e **Phazir** pelos métodos recalibração (RC); sem transferência (ST); RS, PRS e RSDW para discriminação de **andiroba** (\diamond) das espécies cedro (\square), curupixá (\triangle) e mogno (\circ) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão.

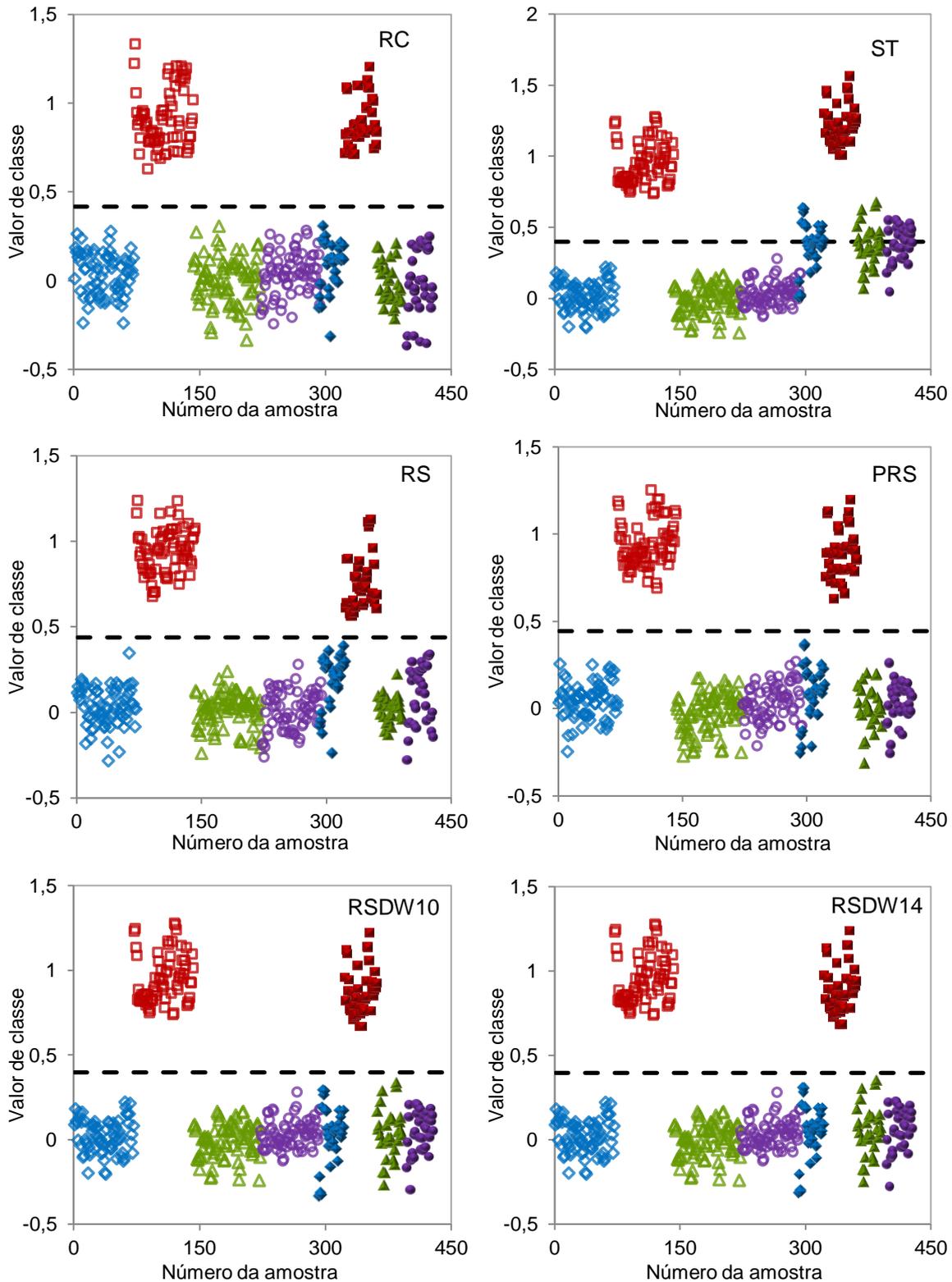


Figura 40. Valores de classe das amostras estimados na transferência entre os equipamentos **Tensor 37** e **Phazir** pelos métodos recalibração (RC); sem transferência (ST); RS, PRS e RSDW para discriminação de **cedro** (□) das espécies andiroba (◇), curupixá (△) e mogno (○) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão.

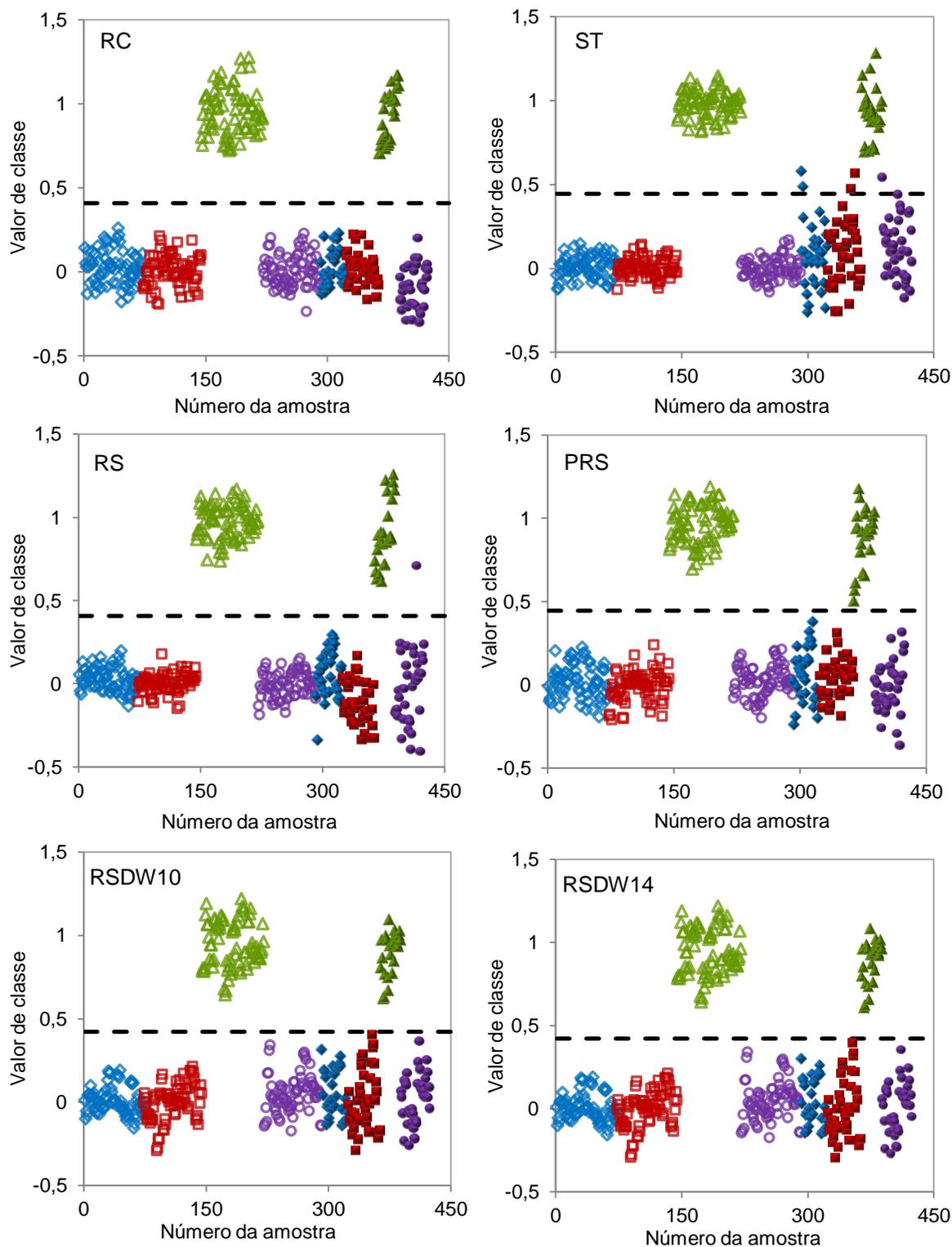


Figura 41. Valores de classe das amostras estimados na transferência entre os equipamentos **Tensor 37** e **Phazir** pelos métodos recalibração (RC); sem transferência (ST); RS, PRS e RSDW para discriminação de **curupixá** (Δ) das espécies andiroba (\diamond), cedro (\square) e mogno (\circ) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão.

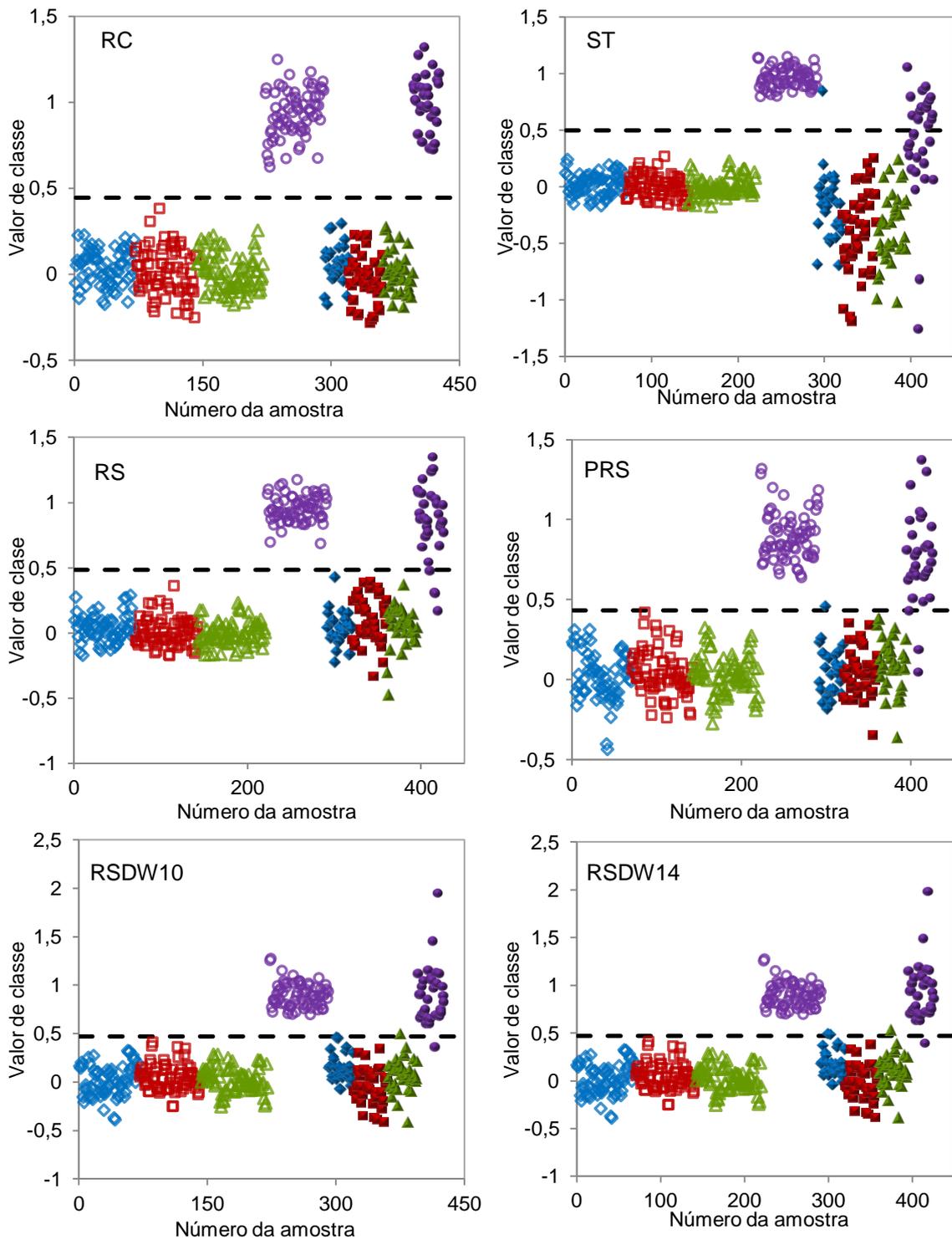


Figura 42. Valores de classe das amostras estimados na transferência entre os equipamentos **Tensor 37** e **Phazir** pelos métodos recalibração (RC); sem transferência (ST); RS, PRS e RSDW para discriminação de **curupixá** (Δ) das espécies andiroba (\diamond), cedro (\square) e mogno (\circ) e limites discriminação (---). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão.

5.4 CONCLUSÕES DO CAPÍTULO

A transferência de calibração tornou possível o compartilhamento de dados entre equipamentos de bancada diferentes. No caso estudado a seleção de faixas espectrais com menos variações foi suficiente para que os modelos do equipamento primário apresentassem 100% de classificações corretas para as amostras de previsão do equipamento secundário. A padronização pelos métodos PDS e SDW reduziram os valores de RMSEP com relação à situação sem transferência exceto para o cedro, enquanto padronização por DS não apresentou bom desempenho. Esses resultados comprovam que a disponibilização ou aproveitamento de bancos de dados entre equipamentos de bancada é perfeitamente viável, desde que sejam escolhidos o pré-processamento e a região espectral adequada.

Foi possível também aproveitar bancos de dados de equipamentos de bancada para construir um modelo padronizado aplicável às medidas do equipamento de campo utilizando padronizações reversas.

O método PRS apresentou os melhores resultados para a transferência entre os equipamentos Antaris II e Phazir, com porcentagens de classificações corretas de 93,8%, 100%, 95,4% e 97,9% para andiroba, cedro, curupixá e mogno, respectivamente.

Na transferência entre os equipamentos Tensor 37 e Phazir, os melhores métodos foram RS com 94,6% de classificações corretas para andiroba, PRS com 100% de classificações corretas para cedro e curupixá e RSDW com 98,5% para mogno.

Numa situação de substituição de equipamento, outra alternativa à recalibração completa além da transferência de calibração é a recalibração utilizando conjunto de calibração menor, que pode ser um subconjunto do conjunto de calibração original. No caso estudado foi possível construir modelos de calibração com um subconjunto de 28 amostras aleatórias retiradas do conjunto de calibração original de 74 amostras. Os modelos construídos a partir desse subconjunto apresentaram 100% de classificações corretas para todas as espécies. Portanto, estando disponíveis as amostras

para a realização das medidas, essa estratégia pode ser a mais apropriada, uma vez que as medidas espectrais são adquiridas de forma relativamente rápida. Contudo, se as amostras não estiverem mais disponíveis, os métodos de transferência de calibração demonstraram ser eficientes para aproveitar dados oriundos de outro equipamento.

CAPÍTULO 6

Transferência de Calibração entre amostras em pó e em blocos

6. TRANSFERÊNCIA DE CALIBRAÇÃO ENTRE AMOSTRAS EM PÓ E EM BLOCOS

A discriminação das quatro madeiras tropicais andiroba, cedro, curupixá e mogno por NIRS e PLS-DA pode ser realizada utilizando amostras sólidas em blocos ou moídas.^{2,17}

Como foi destacado na seção 3.1.1 do capítulo 3, a superfície da amostra onde a radiação incide influencia o resultado da interação entre essa radiação e a amostra, de forma que esperam-se diferenças na forma como a radiação é refletida por amostras na forma de pó e de blocos.⁴⁵ A variação entre as medidas de amostras em diferentes formas de apresentação pode fazer com que o modelo de calibração construído a partir de amostras não processadas não seja aplicável a amostras moídas e vice e versa.

Neste capítulo utilizaram-se pré-processamento e seleção de intervalos espectrais para verificar se amostras moídas de mogno de diferentes países poderiam ser corretamente classificadas como mogno a partir de modelos PLS-DA para discriminação das quatro espécies citadas construídos a partir de amostras em blocos medidas no mesmo equipamento.

6.1 EXPERIMENTAL

Os espectros das amostras na forma de blocos foram medidos no equipamento Tensor 37, descritos na seção 4.1 do capítulo 4.

Os espectros das amostras moídas utilizadas para a transferência foram cedidos ao projeto e medidos em amostras obtidas a partir do mesmo conjunto que originou as amostras utilizadas no capítulo 4. Embora se trate do mesmo conjunto, a identificação das amostras moídas era incompatível com a das amostras em blocos, o que impossibilitou a obtenção de amostras de transferência.

As 155 amostras moídas de diferentes países foram doadas pelo Forest Products Laboratory, Madison, EUA e consistiam em: 2 amostras de Barbados (BRB), 4 de Belize (BLZ), 3 da Bolívia (BOL), 7 do Brasil (BRA), 2 da China (CHN), 2 da Colômbia (COL), 4 de Costa Rica (CRI), 10 de Cuba (CUB), 7 de Curaçao (CUW), 1 de Dominica (DMA), 1 do Equador (ECU), 11 dos Estados Unidos (USA), sendo 3 da Flórida (USA(FL)) e 1 do Havaí (USA(HI)), 3 das Filipinas (PHL), 4 da Guatemala (GTM), 24 de Honduras (HND), 1 da Índia (IND), 4 da Jamaica (JAM), 1 da Malásia (MYS), 18 do México (MEX), 5 da Nicarágua (NIC), 5 do Panamá (PAN), 14 do Peru (PER), 5 da ilha de São Domingo onde se localizam a República Dominicana (DOM) e o Haiti (HTI), 1 da Sri Lanka (LKA), 1 da ilha de Trinidad e Tobago (TTO) e 15 da Venezuela (VEN). As siglas dos países utilizadas são as constantes na norma ISO 3166-1.

As amostras foram cedidas previamente identificadas como mogno por meio de identificação botânica e moídas a tamanho de partículas de 420 a 250 μm (partículas que passaram na peneira de 40 mesh e foram retidas na de 60 mesh).

As medidas foram feitas em triplicata no equipamento FT-IR Tensor 37 (Bruker Optics) com resolução de 4 cm^{-1} , 64 varreduras por amostra, região espectral de 12.000 cm^{-1} a 3.500 cm^{-1} , medidas de reflectância difusa com esfera integradora e espectro resultante em escala de absorbância ($\log(1/R)$).

6.2 ANÁLISE DE DADOS

A análise de dados foi feita no programa Matlab 7.12.0 (R2011a).

Os espectros foram derivados utilizando 13 pontos de suavização com filtro Savitzky-Golay. Não foi feita a média das replicatas.

Uma vez que a incompatibilidade na identificação das amostras impossibilitou a existência de amostras de transferência, neste capítulo procurou-se minimizar as diferenças espectrais divididas à forma de apresentação das amostras através do uso de pré-processamentos e seleção

de intervalos espectrais. Por isso, de forma semelhante à transferência entre equipamentos de bancada (capítulo 5), selecionou-se visualmente uma região espectral que apresentava menos ruído e menor diferença entre os espectros pré-processados das amostras em pó e em blocos.

Das 444 medidas das amostras em blocos, dois terços das amostras foram utilizados para a construção dos modelos de calibração e um terço para previsão.

Construiu-se um modelo PLS1-DA para discriminação de mogno em relação às demais espécies a partir das amostras em blocos utilizando a região selecionada.

Os outliers de calibração foram excluídos com base no método descrito na norma da ASTM E1655-05,70 e sugerido por Valderrama et al.,⁷³ considerando como outliers amostras com elevados resíduos não modelados em y , utilizando intervalo de confiança de 99%.

Construíram-se novos modelos sem os outliers de calibração e aplicaram-se estes aos conjuntos de previsão das amostras em blocos. Excluíram-se os outliers de previsão.

O modelo foi aplicado às amostras na forma de pó do conjunto de transferência e dos países e não foi excluída nenhuma amostra. Os resultados foram registrados e o limite de discriminação e as porcentagens de classificações corretas foram calculados.

Por último, construíram-se da mesma forma modelos para a discriminação de andiroba, cedro e curupixá a partir do conjunto de calibração das amostras em blocos e aplicaram-se esses modelos às amostras em blocos de previsão e às amostras de mogno em pó.

6.3 RESULTADOS E DISCUSSÃO

A Figura 43 mostra os espectros não pré-processados das amostras em blocos e das amostras moídas de mogno e os espectros pré-processados com

6. TRANSFERÊNCIA DE CALIBRAÇÃO ENTRE AMOSTRAS EM PÓ E EM BLOCOS

a região espectral selecionada indicada. O pré-processamento utilizado foi a primeira derivada com filtro Savitzky-Golay de 13 pontos e polinômio de segunda ordem. A região espectral selecionada foi de 5.547 a 6.897 cm^{-1} . A seleção do limite superior da região selecionada está em concordância com a influência do tamanho de partículas reportada em regiões espectrais acima de 7000 cm^{-1} .⁵⁷

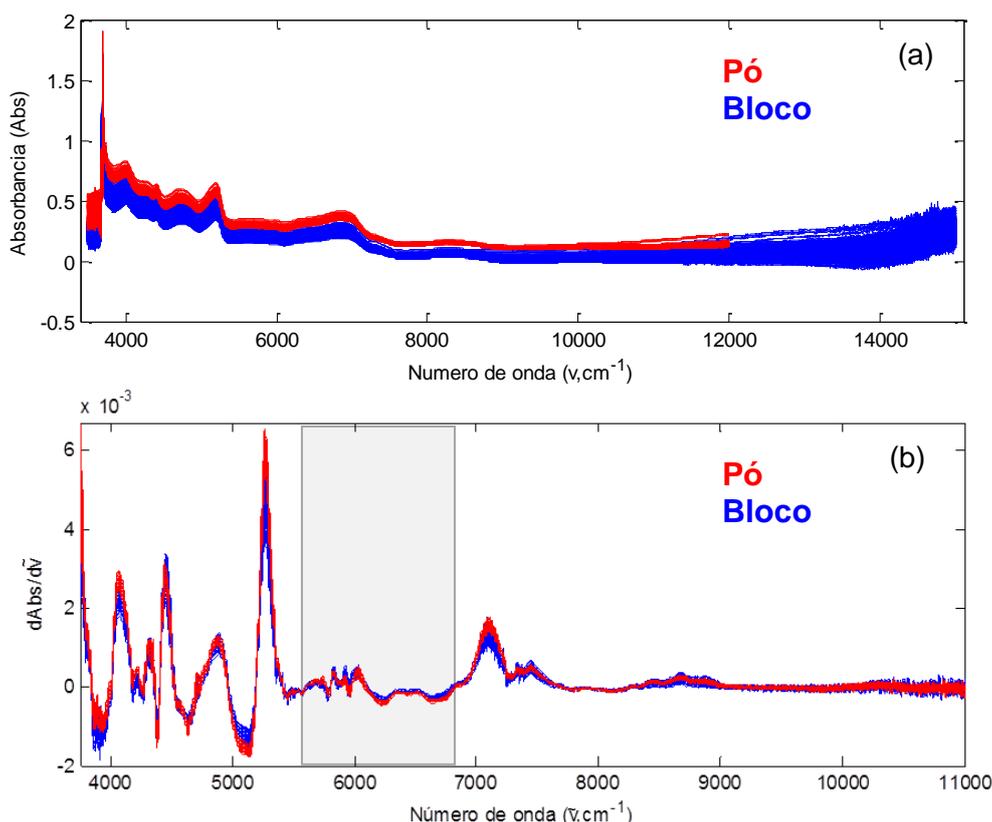


Figura 43. Espectros das amostras em pó e blocos a) originais e b) derivados e com região selecionada indicada (faixa sombreada).

A Tabela 10 mostra parâmetros dos modelos e resultados para discriminação das amostras em blocos e em pó do conjunto de transferência e dos países. A Figura 44 mostra os valores de classe estimados pelos modelos para discriminação de mogno, andiroba, cedro e curupixá aplicados às amostras de mogno moídas.

Observa-se que o modelo para discriminação de mogno classifica as amostras moídas como mogno com 100% de acerto, embora o valor de RMSEP seja maior para essas amostras com relação as amostras em blocos de previsão.

6. TRANSFERÊNCIA DE CALIBRAÇÃO ENTRE AMOSTRAS EM PÓ E EM BLOCOS

Tabela 10. Resultados dos modelos PLS-DA para discriminação quanto a espécie de amostras em diferentes formas de apresentação

| Modelo | Amostras | VL | RMSEC | RMSECV | RMSEP | outc | outp | limite | % |
|-----------------|----------|----|-------|--------|-------|------|------|--------|-------|
| Mogno | Bloco | | | | 0,139 | | 3 | | 100% |
| | Pó | 7 | 0,113 | 0,123 | 0,226 | 3 | - | 0,449 | 100% |
| | Países | | | | 0,311 | | - | | 98,5% |
| Andiroba | Bloco | | | | 0,143 | | 14 | | 100% |
| | Pó | 7 | 0,113 | 0,125 | 0,507 | 6 | - | 0,477 | 45,2% |
| | Países | | | | 0,279 | | - | | 97,4% |
| Cedro | Bloco | | | | 0,105 | | 3 | | 100% |
| | Pó | 8 | 0,078 | 0,089 | 0,113 | 3 | - | 0,455 | 100% |
| | Países | | | | 0,180 | | - | | 98,7% |
| Curupixá | Bloco | | | | 0,081 | | 12 | | 100% |
| | Pó | 7 | 0,074 | 0,082 | 0,416 | 6 | - | 0,356 | 100% |
| | Países | | | | 0,171 | | - | | 98,1% |

VL= número de variáveis latentes. Outc=outliers de calibração. Outp=outliers de previsão.
%=porcentagem de classificações corretas.

Os modelos para discriminação de cedro e curupixá também apresentam 100% de classificações corretas para as amostras moídas. O alto valor de RMSEP (0,416) para previsão das amostras moídas pelo modelo de discriminação de curupixá não interfere na correta classificação das amostras, uma vez que o a dispersão está abaixo de zero, longe do limite de discriminação.

Já o modelo para discriminação de andiroba apresentou um grande número de classificações incorretas e a dispersão das amostras moídas ficou em torno de aproximadamente 0,5. Entretanto, deve-se observar que a aplicação prática desses modelos consiste na análise de uma mesma amostra no modelo de cada espécie, sendo esperado que cada amostra seja considerada como pertencente a uma única espécie. Considerando que a dispersão dessas amostras estimadas pelo modelo para discriminação de mogno ficou mais próxima de 1, pode-se concluir que há maior probabilidade das amostras serem de mogno e não de andiroba.

A Figura 45 mostra os valores de classe estimados pelos modelos das quatro espécies para as amostras moídas dos diferentes países. Observa-se na Figura 45 e na Tabela 10 que os modelos apresentaram porcentagens de classificações corretas superiores a 97%.

6. TRANSFERÊNCIA DE CALIBRAÇÃO ENTRE AMOSTRAS EM PÓ E EM BLOCOS

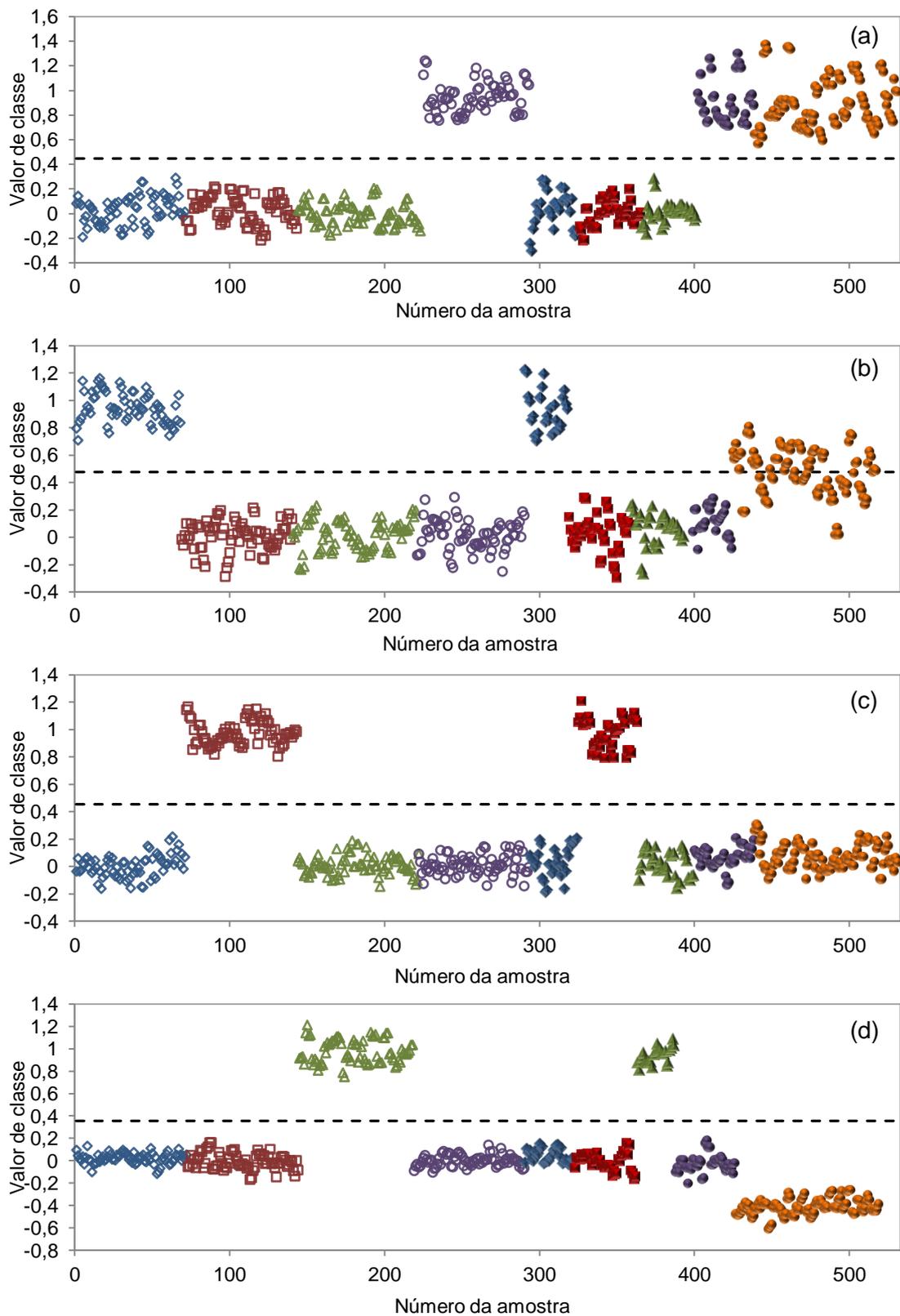


Figura 44. Valores de classe estimados pelos modelos de discriminação de das espécies a) mogno (◇), b) andiroba (◇), c) cedro (□) e d) curupixá (△), limites de discriminação (---) e dispersão dos valores estimados para as amostras moídas (●). Símbolos vazios representam amostras de calibração e preenchidos representam amostras de previsão.

6. TRANSFERÊNCIA DE CALIBRAÇÃO ENTRE AMOSTRAS EM PÓ E EM BLOCOS

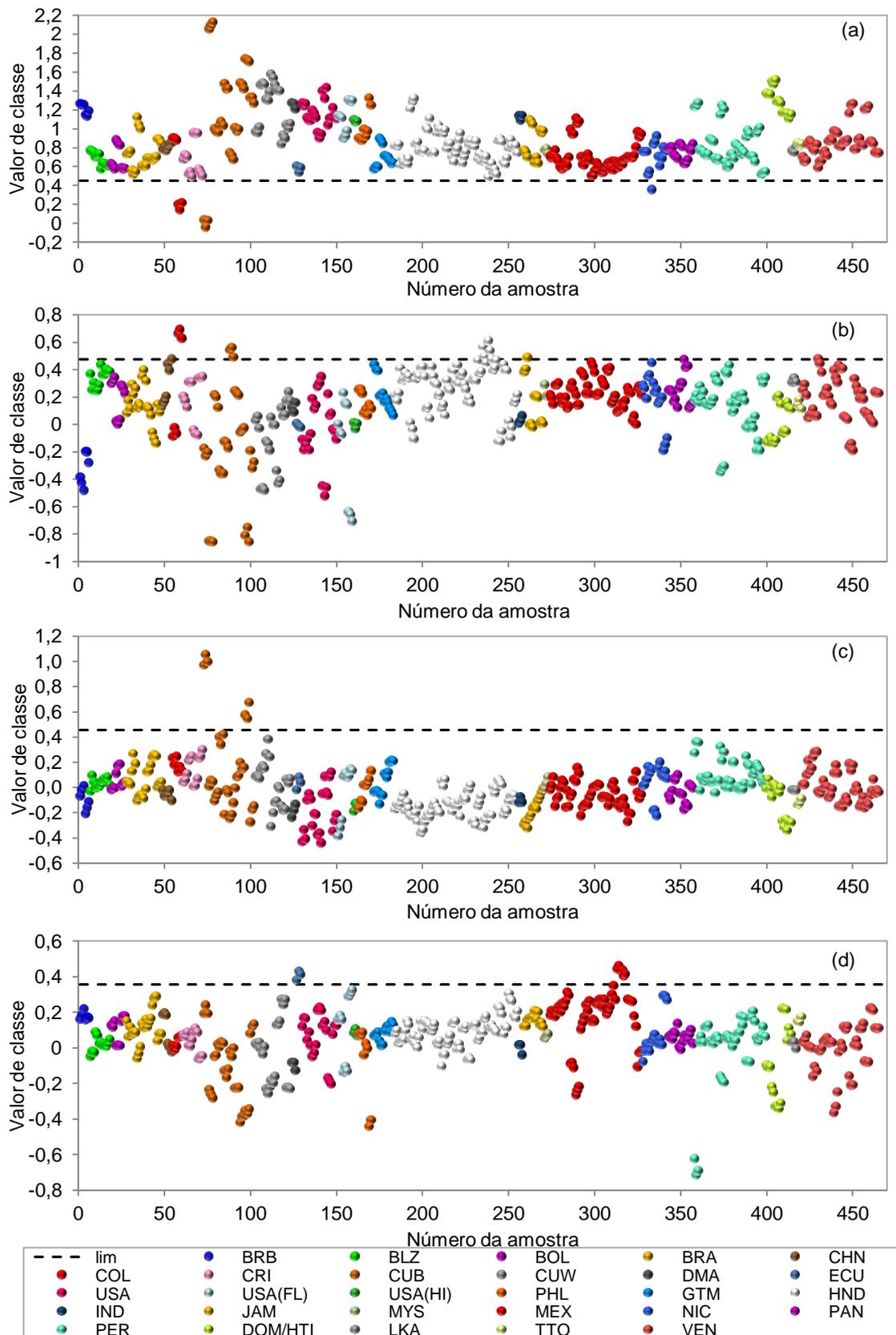


Figura 45. Valores de classe das amostras moídas dos diferentes países estimados pelos modelos para discriminação de a) mogno, b) andiroba, c) cedro e d) curupixá, e limites de discriminação (---).

6. TRANSFERÊNCIA DE CALIBRAÇÃO ENTRE AMOSTRAS EM PÓ E EM BLOCOS

A dispersão das amostras quanto aos valores de classe estimados pelo modelo de discriminação de mogno (Figura 45 a) de forma geral apresentou um ligeiro deslocamento para baixo com relação ao valor 1. As amostras classificadas erroneamente foram 1 amostra da Colômbia (COL) (3 replicatas), 1 amostra de Cuba (CUB) (3 replicatas) e 1 replicata da Nicarágua (NIC).

Comparando-se as dispersões para os modelos das quatro espécies (Figura 45 a-d) observa-se que a amostra da Colômbia (COL) que não foi classificada como mogno foi classificada como andiroba. Já a amostra de Cuba (CUB) que não foi classificada como mogno foi classificada como curupixá. A replicata da Nicarágua (NIC) que não foi classificada como mogno também não foi classificada como nenhuma das outras três espécies. Algumas amostras classificadas como mogno foram também classificadas como outras espécies, conforme será detalhado a seguir.

A dispersão das amostras quanto aos valores de classe estimados pelo modelo de discriminação de andiroba (Figura 45 b) apresentou de forma geral um ligeiro deslocamento para cima com relação ao zero. As amostras classificadas erroneamente foram 1 amostra da Colômbia (COL) (3 replicatas), 1 amostra de Cuba (CUB) (3 replicatas), 2 replicatas de uma amostra de Honduras (HND) e outras 2 replicatas de outra amostra do mesmo país, 1 replicata da Jamaica (JAM) e 1 replicata da Venezuela (VEN).

A dispersão dos valores de classe das amostras estimados pelo modelo para discriminação de cedro (Figura 45 c) não apresentaram viés de forma geral, apresentando um ligeiro deslocamento sistemático maior ou menor dependendo de cada país. As classificações errôneas foram 2 amostras (6 replicatas) de Cuba (CUB). Os valores de classe para as amostras dos diferentes países estimados pelo modelo de discriminação de curupixá (Figura 45 d) também não apresentaram um viés de forma geral, apresentando apenas um ligeiro viés maior ou menor dependendo do país. As classificações errôneas consistiram em 1 amostra (3 replicatas) do Equador (ECU) e 2 amostras (6 replicatas) do México (MEX).

Observa-se também na Figura 45 que as dispersões das amostras de Cuba (CUB) apresentaram grande variação em todos os modelos, sendo

menos pronunciada no modelo para discriminação de curupixá, único modelo em que nenhuma amostra de Cuba (CUB) foi classificada erroneamente.

Os resultados mostrados na Tabela 10 e nas Figuras 44 e 45 indicam ainda que a espécie que se parece mais com o mogno pelos métodos utilizados dentre as três demais espécies é a andiroba, o que é coerente com a maior similaridade anatômica existente entre as duas espécies.

6.4 CONCLUSÕES DO CAPÍTULO

Foi possível identificar amostras moídas de mogno utilizando modelo construído com amostras em forma de blocos com 100% de classificações corretas utilizando pré-processamento e seleção de faixa espectral para minimizar as variações dos espectros relativas à diferença de forma de apresentação das amostras. Também foi possível identificar as amostras moídas de mogno como não pertencentes às espécies cedro e curupixá. Já o modelo de andiroba apresentou grande número de classificações errôneas para essas amostras, embora a comparação com o resultado obtido pelo modelo de discriminação de mogno evidencie que essas amostras pertencem a essa classe.

Também foi possível identificar como mogno, com porcentagens de classificação correta superiores a 97%, 465 medidas de 155 amostras moídas de mogno de 26 países diferentes. Isso demonstra que a variação espectral devida à diferença entre espécies é maior do que a variação entre países.

CAPÍTULO 7

Discriminação de mogno por país de procedência

7. DISCRIMINAÇÃO DE MOGNO POR PAÍS DE PROCEDÊNCIA

Como foi destacado no capítulo 2, o mogno é uma das madeiras mais valiosas do mundo e sua ampla exploração fez com que fosse considerado como espécie ameaçada.⁷

A ocorrência natural do mogno se estende dos Estados Unidos, México, passando pela América Central até a região da América do Sul que abrange Peru, Bolívia e Brasil e a legislação quanto à sua exploração e comercialização varia entre os países.^{6,8,9,10} Portanto, a identificação da procedência da madeira é fundamental para a manutenção da exploração legalmente permitida em cada território e a preservação da espécie.

Como demonstrado nos capítulos anteriores, a espectroscopia no infravermelho próximo é uma boa alternativa para realizar essa identificação, uma vez que possibilita análises rápidas e não destrutiva.

Neste capítulo NIRS combinada a PLS-DA foi aplicada para discriminar amostras de mogno provenientes de cinco países: Brasil, Honduras, México, Peru e Venezuela. A Figura 46 mostra um mapa indicando a posição geográfica dos cinco países.

7.1 EXPERIMENTAL

Os cinco países foram escolhidos dentre os países abordados no capítulo 6 por apresentarem maior número de amostras disponíveis. Foram utilizadas 7 amostras do Brasil, 24 de Honduras, 18 do México, 14 do Peru e 15 da Venezuela. Essas amostras foram doadas pelo Forest Products Laboratory, Madison, EUA. Ao conjunto de amostras do Brasil acrescentaram-se 31 amostras cedidas pelo Laboratório de Produtos Florestais do Serviço Florestal Brasileiro. As amostras foram moídas a tamanho de partículas de 420 a 250 µm (partículas que passaram na peneira de 40 mesh e foram retidas na de 60 mesh) e medidas em triplicata.



Figura 46. Mapa político mostrando a posição geográfica dos países Brasil, Honduras, México, Peru e Venezuela. Adaptado da referência 89.

As medidas foram feitas no equipamento FT-IR Tensor 37 (Bruker Optics) com resolução de 4 cm^{-1} , 64 varreduras por amostra, região espectral de 12.000 cm^{-1} a 3.500 cm^{-1} , medidas de reflectância difusa com esfera integradora e espectro resultante em escala de absorbância ($\log(1/R)$).

7.2 ANÁLISE DE DADOS

A análise dos dados foi feita programa OPUS 6.5 da Bruker Optics. Não foi feita a média das replicatas. Dois terços das amostras foram utilizados para a construção dos modelos de calibração e um terço para previsão.

Para construção dos modelos PLS1-DA, valores de classe de referência foram atribuídos a cada país do conjunto de calibração como sendo 1 para as amostras pertencentes ao país em questão e 0 para as demais. Em seguida,

foi feita a regressão PLS do vetor \mathbf{y} de cada país e da matriz de dados \mathbf{X} correspondente aos espectros para obtenção de um modelo para cada país. Na rotina de otimização do OPUS, foi feita uma validação cruzada completa com otimização dos pré-processamentos e da região espectral.

Testaram-se alguns dos melhores pré-processamentos apontados pela otimização levando em conta o valor de RMSECV e o número de VL e optou-se por utilizar o mesmo pré-processamento para todos os países.

O modelo de cada país foi construído utilizando o pré-processamento e a região selecionados. Selecionou-se o número de VL que apresentou o menor RMSECV ou o número de VL a partir do qual o valor de RMSECV se estabilizou relativamente.

Um primeiro modelo foi construído e os *outliers* foram excluídos do conjunto de calibração segundo os critérios do programa, que se baseia nos resíduos não modelados e influência das amostras. Em seguida construiu-se um novo modelo com as amostras restantes.

A eficiência de previsão do modelo foi testada utilizando-se os espectros das amostras do conjunto de previsão. Os *outliers* de previsão foram excluídos. Os valores de RMSEP foram determinados antes e após a exclusão dos *outliers* de previsão e utilizados para avaliar o desempenho dos modelos, tomando-se o cuidado de utilizar o mesmo número de VL definido na fase de validação cruzada.

Os limites de discriminação dos modelos foram determinados por uma análise Bayesiana utilizando o programa Matlab.

7.3 RESULTADOS E DISCUSSÃO

O pré-processamento selecionado foi subtração de linha reta e a região espectral selecionada foi de 7505,2 a 9306,7 cm^{-1} .

A Tabela 11 mostra número de VL, valores de RMSEC, RMSECV e RMSEP antes e depois da exclusão de *outliers* de previsão, número outliers,

limites de discriminação e porcentagens de acerto dos modelos para discriminação de mogno de cada país. A Figura 47 mostra os valores de classe estimados na discriminação de cada país antes da exclusão dos outliers de previsão e os limites de discriminação.

Os modelos para discriminar as amostras do Brasil, de Honduras e do México com relação às dos demais países apresentaram bom desempenho, com valores reduzidos de erro e 100% de classificações corretas mesmo antes da exclusão dos poucos *outliers* de previsão.

Já os modelos para discriminação de amostras do Peru e da Venezuela apresentaram 4 de 5 amostras da Venezuela (12 replicatas) classificadas de forma errada, o que indica que o método não foi capaz de identificar as amostras desse país. Por isso, antes da exclusão dessas amostras, os valores de RMSEP foram elevados e a porcentagem de classificações corretas foi de 89,5%. Os modelos para discriminação desses dois países apresentaram também um número elevado de amostras da Venezuela consideradas *outliers* de calibração. Em ambos os casos, 6 de 10 amostras de calibração da Venezuela (18 replicatas) foram consideradas *outliers*.

Observou-se que as amostras de Venezuela que foram consideradas *outliers* aproximavam-se mais das amostras do Peru do que das demais amostras da Venezuela.

Tabela 11. Número de variáveis latentes (VL), erros, outliers de calibração (outc) e previsão (outp), limites de discriminação e porcentagens de acerto (%) dos modelos para discriminação de mogno por país de procedência

| | VL | RMSEC | RMSECV | outc | limite | RMSEP ^a | % ^a | outp | RMSEP ^b | % ^b |
|------------|----|-------|--------|------------------|--------|--------------------|----------------|---------------|--------------------|----------------|
| BRA | 9 | 0,058 | 0,094 | 2 BRA | 0,502 | 0,090 | 100% | 1 MEX | 0,085 | 100% |
| HND | 9 | 0,052 | 0,075 | 2 HND, 1BRA | 0,495 | 0,067 | 100% | 1BRA, 1MEX | 0,063 | 100% |
| MEX | 8 | 0,051 | 0,065 | 1 HND | 0,438 | 0,071 | 100% | - | 0,071 | 100% |
| PER | 9 | 0,049 | 0,077 | 18 VEN | 0,567 | 0,316 | 89,5% | 12 VEN | 0,071 | 100% |
| VEN | 8 | 0,057 | 0,075 | 18 VEN, 1 PER | 0,567 | 0,315 | 89,5% | 12 VEN | 0,069 | 100% |

^aAntes da exclusão dos outp.

^bApós exclusão dos outp

7. DISCRIMINAÇÃO DE MOGNO POR PAÍS DE PROCEDÊNCIA

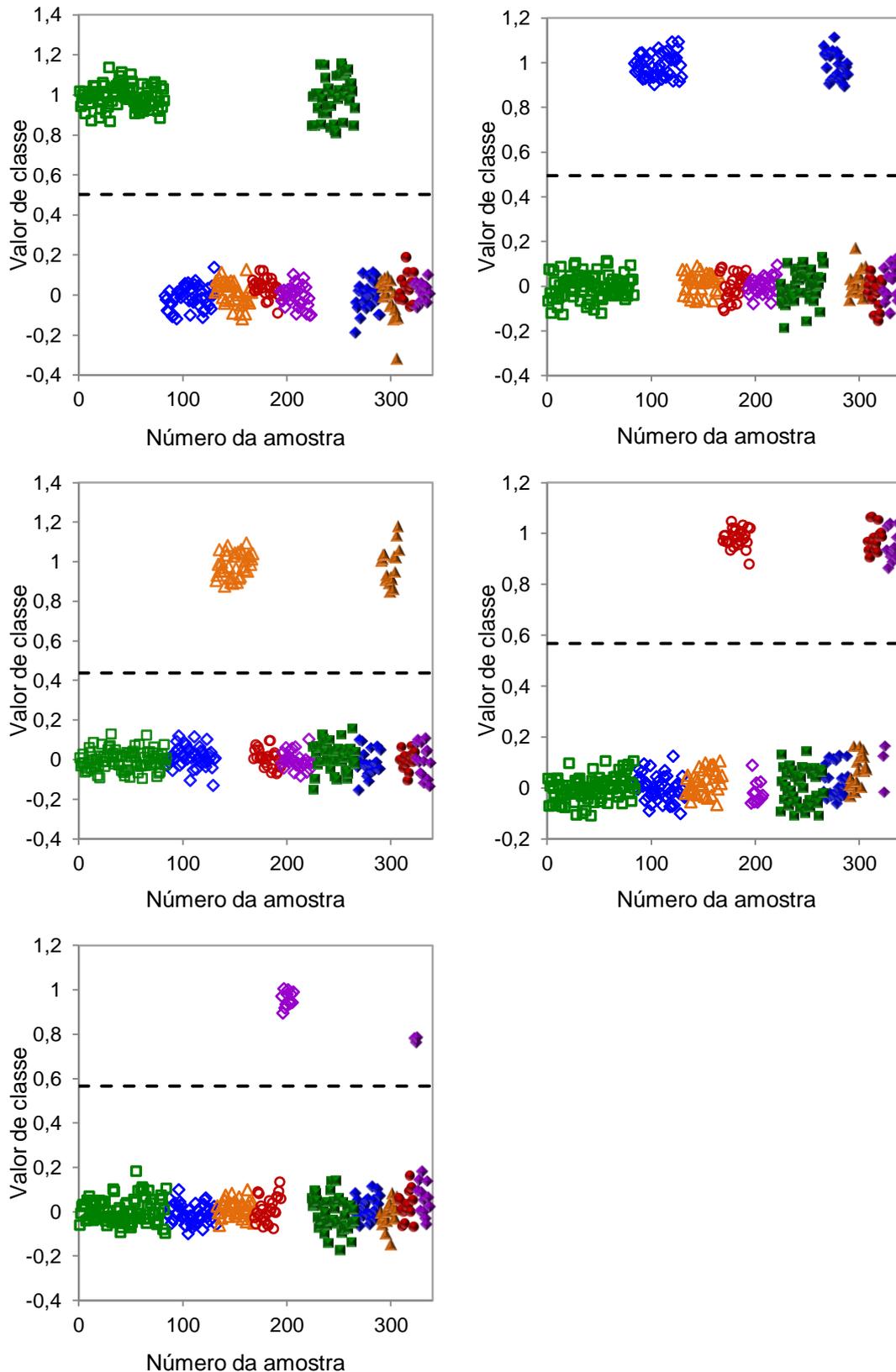


Figura 47. Valores de classe estimados para discriminação de mogno de Brasil (\square), Honduras (\diamond), México (\triangle), Peru (\circ) e Venezuela (\diamond) antes da exclusão dos *outliers* de previsão e limites de discriminação (---). Símbolos vazios correspondem a amostras de calibração e preenchidos de previsão.

Isso pode ser devido a semelhanças entre madeiras de mogno produzidas nos dois países, o que dificulta a discriminação entre elas. Para se obter mais informações, seria necessário um estudo com um número maior de amostras desses dois países, considerando que poucas medidas de amostras da Venezuela puderam ser discriminadas corretamente (uma amostra de previsão e as amostras de calibração que não foram consideradas *outliers*). É possível observar um viés na amostra de previsão da Venezuela cuja dispersão está em torno de 0,8 quando deveria estar em torno de 1.

Observou-se também que as amostras do Brasil fornecidas pelo Forest Products Laboratory apresentaram comportamento semelhante às fornecidas pelo SFB, reforçando eficiência do método para discriminação das amostras brasileiras entre as amostras dos outros quatro países.

7.4 CONCLUSÕES DO CAPÍTULO

Entre amostras de mogno de cinco países: Brasil, Honduras, México, Peru e Venezuela, foi possível discriminar perfeitamente as oriundas dos três primeiros utilizando NIRS e PLS-DA. Os modelos para esses três países apresentaram baixos erros, poucos *outliers* e 100% de classificações corretas mesmo antes da exclusão dos *outliers* de previsão.

Já as amostras do Peru e da Venezuela não tiveram boa separação entre si, embora tenham apresentado boa separação com relação aos demais países. Foi necessário excluir a maioria das amostras da Venezuela tanto nos conjuntos de calibração quanto de previsão para ser possível a discriminação com as amostras remanescentes desses dois países.

CAPÍTULO 8
Conclusões gerais

8. CONCLUSÕES GERAIS

A espectroscopia no infravermelho próximo (NIRS) associada a análise discriminante por mínimos quadrados parciais (PLS-DA) mostrou-se eficiente para discriminar amostras de mogno, andiroba, cedro e curupixá utilizando equipamentos de campo e de bancada.

Apesar de apresentarem maiores valores de erro absoluto (RMSEP), as análises com o equipamento portátil de baixa resolução apresentaram 100% de classificações corretas. Esse resultado corrobora a possibilidade de aplicação desse tipo de equipamento na fiscalização da exploração e comércio de madeira.

Com a transferência de calibração dos modelos PLS-DA foi possível compartilhar dados entre diferentes equipamentos. Entre equipamentos de bancada, a seleção de faixas espectrais e o uso de pré-processamento foram suficientes para que os modelos do equipamento primário apresentassem 100% de classificações corretas para as amostras de previsão do equipamento secundário.

A padronização pelos métodos PDS e SDW reduziu os valores de RMSEP com relação à situação sem transferência para andiroba, curupixá e mogno e a padronização por DS não apresentou bom desempenho. Esses resultados comprovam que a disponibilização ou aproveitamento de bancos de dados entre equipamentos de bancada é viável, desde que se escolham pré-processamento e região espectral adequados.

Também foi possível aproveitar dados de equipamentos de bancada para construir um modelo padronizado aplicável às medidas do equipamento de campo utilizando padronizações reversas. O método PRS apresentou os melhores resultados para a transferência entre os equipamentos Antaris II e Phazir, com porcentagens de classificações corretas de 93,8%, 100%, 95,4% e 97,9% para andiroba, cedro, curupixá e mogno, respectivamente. Entre os equipamentos Tensor 37 e Phazir, os melhores métodos foram RS com 94,6% de classificações corretas para andiroba, PRS com 100% de classificações corretas para cedro e curupixá e RSDW com 98,5% para mogno.

Modelos construídos a partir de recalibração com um subconjunto de 28 amostras aleatórias retiradas do conjunto de calibração original de 74 amostras apresentaram 100% de classificações corretas para todas as espécies. Portanto, uma vez que haja amostras disponíveis, essa estratégia pode ser a mais apropriada, considerando que as medidas espectrais são adquiridas de forma relativamente rápida. Contudo, se as amostras não estiverem mais disponíveis, os métodos de transferência de calibração mostraram-se eficientes para aproveitar dados oriundos de diferentes equipamentos.

Também foi possível o compartilhamento de dados entre medidas feitas a partir de amostras em diferentes formas de apresentação. Utilizando-se pré-processamento e seleção de faixa espectral, o modelo construído a partir das amostras em blocos permitiu a discriminação de amostras moídas de mogno com 100% de classificações corretas.

Também foi possível identificar as amostras moídas de mogno como não pertencentes às espécies cedro e curupixá. Já o modelo para andiroba apresentou grande número de classificações errôneas para essas amostras. Entretanto, a comparação com o resultado obtido pelo modelo de discriminação de mogno indica que as amostras pertencem a essa espécie. Esse resultado indica uma maior semelhança entre as espécies andiroba e mogno, o que condiz com a maior similaridade anatômica existente entre as duas espécies.

Foi possível, pelos mesmos modelos, identificar como mogno, com porcentagens de classificação correta superiores a 97%, 465 amostras moídas de mogno de 26 países diferentes. Isso demonstra que as diferenças espectrais entre espécies são mais pronunciadas que as diferenças espectrais entre países.

Por último, entre as amostras moídas de mogno de Brasil, Honduras, México, Peru e Venezuela, foi possível discriminar perfeitamente as oriundas dos três primeiros utilizando NIRS e PLS-DA com baixos valores de erros, poucos *outliers* e 100% de classificações corretas. Já as amostras do Peru e da Venezuela não tiveram boa separação entre si, embora tenham apresentado boa separação com relação aos demais países.

9. REFERÊNCIAS BIBLIOGRÁFICAS

- ¹ H. Krisnawati, M. Kallio, M. Kanninen. *Swietenia macrophylla King: Ecology, silviculture and productivity*. Center for International Forestry Research (CIFOR), Bogor, **2011**. 14 p. Disponível em <http://www.cifor.org/publications/pdf_files/Books/BKrisnawati1104.pdf>. Acesso em 16. out. 2013
- ² T.C.M. Pastore, J.W.B. Braga, V.T.R. Coradin, W.L.E. Magalhaes, E.Y.A. Okino, J.A.A. Camargos, G.I.B. de Muniz, O.A. Bressan, F. Davrieux. *Near infrared spectroscopy (NIRS) as a potential tool for monitoring trade of similar woods: Discrimination of true mahogany, cedar, andiroba, and curupixa*. *Holzforschung*. **2011**, 65, 73.
- ³ J.E. de Melo, G.M. de Carvalho, V.A. Martins. *Espécies de madeira substitutas do mogno*. IBAMA, Brasília, **1989**. 16p.
- ⁴ A. Veríssimo, J. Grogan. *Síntese da situação do mogno em nível internacional*. Ministério do meio ambiente, Brasília, **2002**. 39p.
- ⁵ V.T.R. Coradin, J.A.A. Camargos, L.F. Marques, E.R. da Silva Jr. *Madeiras similares ao mogno (Swietenia macrophylla King.): Chave ilustrada para identificação anatômica em campo*. 2 ed. Serviço Florestal Brasileiro/LPF, Brasília, **2009**. 28p.
- ⁶ I. Tomaselli, S.R. Hirakuri. *Converting mahogany*. ITTO Tropical forest update. **2008**, 18, 12.
- ⁷ J. Grogan, P. Barreto. *Big-leaf mahogany on CITES Appendix II: Big challenge, big opportunity*. *Conserv. Biol.* **2005**, 19, 973.
- ⁸ P.E.R. Carvalho. *Mogno – Swietenia Macrophylla*. Circular técnica 140. Embrapa Florestas, Colombo, PR, 2007. 12p. Disponível em <<http://www.infoteca.cnptia.embrapa.br/handle/doc/313906>>. Acesso em 1º nov. 2013.
- ⁹ E. Mejía, X. Buitrón, M. Peña-Claros, J. Grogan. *Big-leaf mahogany (Swietenia macrophylla) in Peru, Bolivia and Brazil*. In: NDF Workshop Case Studies, WG 1- Trees, Case Study 4. **2008**. 36p.
- ¹⁰ E.F.M. Terezo. *Políticas e manejo do mogno*. Ministério do meio ambiente, Brasília, **2002**. 21p.
- ¹¹ V.T.R. Coradin, J.A.A. Camargos. *A estrutura anatômica da madeira e princípios para sua identificação*. LPF, Brasília, **2002**. 28p.
- ¹² E.A. Wheeler, P. Baas. *Wood identification – a review*. *IAWA J.* **1998**, 19, 241.
- ¹³ P. Gasson, P. Baas, E. Wheeler. *Wood anatomy of CITES-listed tree species*. *IAWA J.* **2011**, 32, 155.
- ¹⁴ S. Tsuchikawa, K. Inoue, J. Noma, K. Hayashi. *Application of near-infrared spectroscopy to wood discrimination*. *J. Wood Sci.* **2003**, 49, 29.
- ¹⁵ R.C.V. Martins-da-Silva, M.G. Hopkins, I.S. Thompson. *Identificação Botânica na Amazônia: Situação Atual e Perspectivas*. 1ª ed. Embrapa Amazônia Oriental, Belém,

2003. 81p. Disponível em: <<http://www.infoteca.cnptia.embrapa.br/handle/doc/407112>>
Acesso em: 16 nov. 2012.

- ¹⁶ M. Blanco, I. Villarroya. *NIR spectroscopy: a rapid-response analytical tool*. *Trends Anal. Chem.* **2002**, 21b 240.
- ¹⁷ J.W.B. Braga, T.C.M. Pastore, V.T.R. Coradin, J.A.A. Camargos, and A.R. da Silva, *The use of near infrared spectroscopy to identify solid wood specimens of Swietenia macrophylla (CITES Appendix II)*. *I J.* **2011**, 32, 285.
- ¹⁸ F.B. Lamb. *Utilization, distribution and management of tropical American mahogany*. Tese de doutorado. University of Michigan, Estados Unidos, **1953**.
- ¹⁹ ITTO TTM Report, v.17, 23, 1-15 Outubro **2013**. p.8
- ²⁰ J.A.A. Camargos. V.T.R. Coradin, C.M. Czanerski, D. Oliveira, I. Meguerditchian. *Catálogo de árvores do Brasil*. 2 ed. Ibama, Brasília, **2001**. 896p.
- ²¹ G. Koch, H.-G. Richter, U. Schmitt. *Design and application of CITESwoodID computer-aided identification and description of CITES-protected timbers*. *IAWA J.* **2011**, 32, 213.
- ²² C. Sarmiento, P. Détienne, C. Heinz, J.-F. Molino, P. Grard. *PI@ntWood: a computer-assisted identification tool for 110 species of Amazon trees based on wood anatomical features*. *IAWA J.* **2011**, 32, 221.
- ²³ S. Tsuchikawa. *A review of recent near infrared research for wood and paper*. *Appl. Spectrosc. Rev.* **2007**, 42, 43.
- ²⁴ M. Schwanninger, B. Hinterstoisser, N. Gierlinger, R. Wimmer, J. Hanger. *Application of Fourier transform near infrared Spectroscopy (FT-NIR) to thermally modified wood*. *Holz. Roh. Werkst.* **2004**, 62, 483.
- ²⁵ S.S. Kelley, T.G. Rials, R. Snell, L.H. Groom, A. Sluiter. *Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood*. *Wood Sci. Technol.* **2004**, 38, 257.
- ²⁶ R. Rana, R. Langenfeld-Heyser, R. Finkeldey, A. Polle. *FTIR spectroscopy, chemical and histochemical characterisation of wood and lignin of five tropical timber wood species of the family of Dipterocarpaceae*. *Wood Sci. Technol.* **2010**, 44, 225.
- ²⁷ J.B. Hauksson, G. Bergqvist, U. Bergsten, M. Sjoström, U. Edlund. *Prediction of basic wood properties for Norway spruce. Interpretation of Near Infrared Spectroscopy data using partial least squares regression*. *Wood Sci. Technol.* **2001**, 35, 475.
- ²⁸ A.D.G. Gouvea, P.F. Trugilho, J.L. Gomide, J.R.M. da Silva, C.R. Andrade, I.C.N. Alves. *Determinação da densidade básica da madeiras de eucalyptus por diferentes métodos não destrutivos*. *Rev. Árvore.* **2011**, 35 349.
- ²⁹ A. Terdwongworakul, V. Punsuwan, W. Thanapase, S. Tsuchikawa. *Rapid assessment of wood chemical properties and pulp yield of Eucalyptus camaldulensis in Thailand tree plantations by near infrared spectroscopy for improving wood selection for high quality pulp*. *J. Wood Sci.* **2005**, 51, 167.

- ³⁰ O.E. Adedipe, B. Dawson-Andoh, J. Slahor, L. Osborn. *Classification of red oak (Quercus rubra) and white oak (Quercus alba) wood using a near infrared spectrometer and soft indrprndrnt modrlng of class analogies*. J. Near Infrared Spectrosc. **2008**, 16, 49.
- ³¹ M. Brunner, R. Eugster, E. Trenka, L. Bergamin-Strotz. *FT-NIR spectroscopy and wood identification*. Holzforschung, **1996**, 50, 130.
- ³² N. Gierlinger, M. Schwanninger, R. Wimmer. *Characteristics and classification of Fourier-transform near infrared spectra of the heartwood of different larch species (Larix sp.)*. J. Near Infrared Spectrosc. **2004**, 12, 113.
- ³³ E.Y. Haartveit, e P.O. Flæte. *Near infrared spectroscopy (NIRS) as a tool for effective classification of wood*. In: *Proceedings of the 51st International Convention of Society of Wood Science and Technology*. Concepción, Chile, **2008**, 9p.
- ³⁴ A. Sandak, J. Sandak, M. Negri. *Relationship between near-infrared (NIR) spectra and the geographical provenance of the timber*. Wood Sci. Technol. **2011**, 45, 35.
- ³⁵ K. Sandberg, M. Sterley. *Separating Norway spruce heartwood and sapwood in dried condition with near-infrared spectroscopy and multivariate data analysis*. Eur. J. Forest. Res. **2009**, 128, 475.
- ³⁶ M. Khalid, E.L.Y. Lee, R. Yusof, M. Nadaraj. *Design of an intelligent wood species recognition system*. IJSSST. **2008**, 9,9.
- ³⁷ R. Yusof, M. Khalid, A. Salwa, M. Khairuddin. *Application of kernel-genetic algorithm as nonlinear feature selection in tropical wood species recognition system*. Comput. Electron. Agr. **2013**, 93, 68.
- ³⁸ M.F. Deguilloux, M.H. Pemonge, L. Bertel, A. Kremer, J. Petit. *Checking the geographical origin of oak wood: molecular and statistical tools*. Mol. Ecol. **2003**, 12, 1629.
- ³⁹ B. Degen, S.E. Ward, M.R. Lemes, C. Navarro, S. Cavers, A.M. Sebbenn. *Verifying the geographic origin of mahogany (Swietenia macrophylla King) with DNA-fingerprints*. Forensic Sci. Int.-Gen. **2013**, 7, 55.
- ⁴⁰ A.F. El Sheikha, C. Chalier, A. Zaremski, D. Montet. *Novel molecular fingerprinting for geographical traceability of timber*. J. Trop. For. Sci. **2013**, 25, 387.
- ⁴¹ J.R. Cordeiro, M.I.V. Martinez, R.W.C. Li, A.P. Cardoso, L.C. Nunes, F.J. Krug, T.R.L.C. Paixão, C.S. Nomura, J. Gruber. *Identification of Four Wood Species by an Electronic Nose and by LIBS*. Int. J. Electrochem. **2012**, 2012, 1.
- ⁴² E.C. Cabral, R.C. Simas, V.G. Santos, C.L. Queiroga, V.S. Cunha, G.F. Sá, R.J. Daroda, M.N. Eberlin. *Wood typification by Venturi easy ambient sonic spray ionization mass spectrometry: the case of the endangered Mahogany tree*. J. Mass Spectrom. **2012**, 47, 1.
- ⁴³ O. Howarth. *Theory of Specyctoscopy: An Elementary Introduction*. Thomas Nelson & Sons, Londres, **1973**. 214p.

-
- ⁴⁴ O. Sala. *Fundamentos da Espectroscopia Raman e no Infravermelho*. Ed. Unesp, São Paulo, **1996**. 223p.
- ⁴⁵ H.H. Willard, L.L. Merritt Jr., J.A. Dean, F.A. Settle Jr. *Instrumental methods of analysis*. Wadsworth, Inc., Belmont, California, **1988**. 895p.
- ⁴⁶ F.J. Holler, D.A. Skoog, S.R. Crouch. *Princípios de Análise Instrumental*. 6 ed. Bookman, Porto Alegre, **2009**, 1056p.
- ⁴⁷ R.M. Silverstein, F.X. Webster. *Identificação Espectrométrica de Compostos Orgânicos*. 6 ed. LTC Ed., Rio de Janeiro. **2000**. 460p.
- ⁴⁸ D.L. Pavia, G.M. Lampman, G.S. Kriz. *Introduction to spectroscopy: a guide for students of organic chemistry*. 3 ed. Brooks/Cole. **2001**. 680p.
- ⁴⁹ J. M. Hollas. *Modern Spectroscopy*. 4 ed. John Wiley & Sons, Chichester, **2004**. 452p.
- ⁵⁰ J.J. Workman Jr. *Interpretative Spectroscopy for Near Infrared*. Appl Spectrosc. Rev. **1996**, 31, 251.
- ⁵¹ C. Pasquini. *Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications*. J. Braz. Chem. Soc. **2003**, 12, 198.
- ⁵² L.M. Hanssen, K.A. Snail. Integrating Spheres for Mid- and Near Infrared Reflection Spectroscopy. In: *Handbook of Vibrational Spectroscopy*; J.M. Chalmers, P.R. Griffiths, Eds.; John Wiley & Sons, Ltd: New York, **2002**; 2, 1175 - 1192.
- ⁵³ E. Zamora-Rojas, D. Pérez-Marín, E. De Pedro-Sanz, J.E. Guerrero-Ginel, A. Garrido-Varo. *Handheld NIRS analysis for routine meat quality control: Database transfer from at-line instruments*. Chemom. Intell. Lab Syst. **2012**, 114, 30.
- ⁵⁴ D.R. Day, M.A. Butler, M.C. Smith, A. McAllister, E.R. Deutsch, K. Zafiriou, S.D. Senturia. Diffractive-MEMS implementation of Hadamard near-infrared spectrometer. In: *Solid-State Sensors, Actuators and Microsystems, 2005. Digest of Technical Papers. TRANSDUCERS'05. The 13th International Conference on*. **2005**, 2, 1246.
- ⁵⁵ Y. Geller. *A new approach to NIR spectroscopy allowing remote analysis*. Lab Plus international. **2006**, 20, 13.
- ⁵⁶ J.A. Fernández Pierna, P. Vermeulen, B. Lecler, V. Baeten, P. Dardenne. *Calibration transfer from dispersive instruments to handheld spectrometers*. Appl. Spectrosc. **2010**, 64, 644.
- ⁵⁷ M. Schwanninger, J.C. Rodrigues, K. Fackler. *A review of band assignments in near infrared spectra of wood and wood components*. J. Near Infrared Spectrosc. **2011**, 19, 287.
- ⁵⁸ R.G. Brereton. *Chemometrics: applications of mathematics and statistics to laboratory systems*. Ellis Horwood Limited. **1990**. 307 p.
- ⁵⁹ P. Geladi. *Chemometrics in spectroscopy. Part 1. Classical chemometrics*. Spectrochim. Acta B. **2003**, 58, 767.

- ⁶⁰ R.G. Brereton. *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. John Wiley & Sons. **2003**. 489 p.
- ⁶¹ E.V. Thomas. *A primer on multivariate calibration*. Anal. Chem. **1994**, 66, 795A.
- ⁶² M.F. Pimentel. *Calibração: uma revisão para químicos analíticos*. Quim. Nova. **1996**, 19, 268.
- ⁶³ K.S. Booksh, B.R. Kowalski. *Theory of Analytical Chemistry*. Anal. Chem. **1994**, 66, 782A.
- ⁶⁴ T. Næs, T. Isaksson, T. Fearn, T. Davies. *A user-friendly guide to Multivariate Calibration and Classification*. NIR Publications, Chichester, **2004**. 344 p.
- ⁶⁵ J. Trygg. *Chemometrics made easy*. Homepage of chemometrics, Editorial de setembro, **2003**.
- ⁶⁶ R.G. Brereton. *Introduction to multivariate calibration in analytical chemistry*. Analyst. **2000**, 125, 2125.
- ⁶⁷ S. Wold, M. Sjöström, L. Eriksson. *PLS-regression: a basic tool of chemometrics*. Chemom. Intell. Lab. Syst. **2001**, 58, 109.
- ⁶⁸ P. Geladi, B. Kowalski. *Partial least-squares regression: a tutorial*. Anal. Chim. Acta. **1986**, 185, 1.
- ⁶⁹ S. de Jong. *SIMPLS: an alternative approach to partial least squares regression*. Chemom. Intell. Lab. Syst. **1993**, 18, 251.
- ⁷⁰ The American Society for Testing and Materials (ASTM). E1655-05: *Standards Practices for Infrared Multivariate Quantitative Analysis*, ASTM Annual Book of Standards, West Conshohocken, USA, **2005**.
- ⁷¹ H. Martens, T. Naes. *Multivariate Calibration*. John Wiley & Sons, New York, 1989.
- ⁷² B. Walczak, D.L. Massart. *Multiple outlier detection revisited*. Chemom. Intell. Lab. Syst. **1998**, 41, 1.
- ⁷³ P. Valderrama, J.W.B. Braga, R.J. Poppi. *Variable selection, outlier detection and figures of merit estimation in a PLSR multivariate calibration model. A case study for the determination of quality parameters in alcohol industry by near infrared spectroscopy*. J. Agric. Food Chem. 2007, 52, 8331.
- ⁷⁴ O.E. de Noord. *Multivariate Standardization*. Chemom. Intell. Lab. Syst. **1994**, 25, 85.
- ⁷⁵ Bruker Optik GmbH. *OPUS Spectroscopy Software, version 6, User Manual, QUANT*. **2006**.
- ⁷⁶ F.S.G. Lima, L.E.P. Borges. *Evaluation of standardization methods of near infrared calibration models*. J. Near Infrared Spectrosc. **2002**, 10, 269.
- ⁷⁷ P. Geladi, H. Bårring, E. Dåbakk, J. Trygg, H. Antti, S. Wold, B. Karlberg. *Calibration transfer for predicting lake-water pH from near infrared spectra of lake sediments*. J. Near Infrared Spectrosc. **1999**, 7, 251.

-
- ⁷⁸ T. Fearn. *Standardisation in calibration transfer for near infrared instruments: a review*. J. Near Infrared Spectrosc. **2001**, 9, 229.
- ⁷⁹ F.A. Honorato, B.B. Neto, M.N. Martins, R.K.H. Galvão, M.F. Pimentel. *Transferência de calibração em métodos multivariados*. Quim. Nova. **2007**, 5, 1301.
- ⁸⁰ R.N. Feudale, N.A. Woody, H. Tan, A.J. Myles, S.D. Brown, J. Ferré. *Transfer of multivariate alibration models: a review*. Chemom. Intell. Lab. Syst. **2002**, 64, 181.
- ⁸¹ H. Swierenga, A.P. de Weijer, R.J. van Wijk, L.M.C. Buydens. *Strategy for constructing robust multivariate calibration models*. Chemom. Intell. Lab. Syst. **1999**, 49, 1.
- ⁸² Y. D. Wang, D. J. Velkamp, B. R. Kowalski. *Multivariate Instrument Standardization*. Anal. Chem. **1991**, 63, 2750.
- ⁸³ J.S. Shenk, M.O. Westerhaus. *Optical instrument calibration system*. US Patent. 4,866,644 **1989**.
- ⁸⁴ F.S.G. Lima, L.E.P. Borges. *Evaluation of standardisation methods of near infrared calibration models*. J. Near Infrared Spectrosc. **2002**, 10, 269.
- ⁸⁵ Z. Wang, T. Dean, B.R. Kowalski. *Additive Background Correction in Multivariate Instrument Standardization*. Anal. Chem. **1995**, 67, 2379.
- ⁸⁶ N. Armstrong, D.B. Hibbert. *An introduction to Bayesian methods for analyzing chemistry data Part 1: An introduction to Bayesian theory and methods*. Chemom. Intell. Lab. Syst. **2009**, 97, 194.
- ⁸⁷ C.M. Jarque, A.K. Bera. *A test for normality of observations and regression residuals*. Int. Stat. Rev, **1987**, 55, 163.
- ⁸⁸ R.W. Kennard, L.A. Stone. *Computer aided design of experiments*. Technometrics. **1969**, 11, 137.
- ⁸⁹ <<http://www.conceptdraw.com/How-To-Guide/picture/geo-map-south-america-argentina/Geo-map-south-america-countour.png>>. Acesso em 13 jun. 2014.