



UnB

Instituto de Química
Programa de Pós-Graduação em Química

DISSERTAÇÃO DE MESTRADO

**Determinação da origem de madeira mogno (*Swietenia macrophylla*)
proveniente de cinco países da América Latina e discriminação entre as
espécies de mogno *S. macrophylla* e *S. humilis* usando equipamentos NIR
portáteis e análise multivariada**

Diego Carvalho da Silva

Orientador: Jez Willian Batista Braga

Coorientadora: Tereza Cristina Monteiro Pastore

Brasília, DF

2018

FOLHA DE APROVAÇÃO

Comunicamos a aprovação da Defesa de Dissertação do (a) aluno (a) **Diego Carvalho da Silva**, matrícula nº **17/0118355**, intitulada “***Determinação da origem de madeira mogno (Swietenia macrophylla) proveniente de cinco países da América Latina e discriminação entre as espécies de mogno S. macrophylla e S. humilis usando equipamentos NIR portáteis e análise multivariada.***”, apresentada no (a) Auditório 1 da Faculdade de Ciências da Saúde (FS) da Universidade de Brasília (UnB) em 6 de dezembro de 2018.

Prof. Dr. Jez Willian Batista Braga

Presidente de Banca

Prof. Dr. Floriano Pastore Júnior

Membro Titular

Dra. Fernanda Piccolo Pieruzzi

Membro Titular Serviço Florestal Brasileiro (SFB)

Prof. Dr. Alexandre Florian da Costa

Membro Suplente

Em 6 de dezembro de 2018.

DEDICATÓRIA

*Dedico esse trabalho aos meus pais
e meus amigos.*

AGRADECIMENTOS

A Deus, pela sua bondade e por me garantir força, saúde e sabedoria.

À minha família e amigos por me apoiarem nessa jornada.

Ao meu orientador Prof. Dr. Jez Willian Batista Braga, pelos ensinamentos, paciência, confiança e atenção. Minha sincera admiração.

À minha coorientadora Dra. Tereza Cristina Monteiro Pastore, pelo incentivo, pelos ensinamentos, pela orientação, pela amizade e pelo carinho e cuidados que me foram passados durante esses anos, desde o estágio até o mestrado, meu sincero e imenso obrigado.

Ao Instituto de Química da Universidade de Brasília, pela oportunidade para realizar o mestrado. E ao Grupo de Estudos em Automação, Quimiometria e Química Ambiental (AQQUA), pelo ambiente de trabalho.

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

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

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

À equipe de trabalho: Liz Franco, Filipe Snell, Nayara Cristina, pela contribuição na realização e conclusão deste trabalho.

À Prof. Myrna Herrera (USAC) e Eng. César Beltetón (CONAP) pela ajuda na viagem de campo na Guatemala.

Ao Eng. Jorge E. Avila da Cámara Forestal de Bolivia, Eng. Manuel Chavesta Custodio da Universidad Nacional Agraria La Molina e Eng. Luiz Rogério de Oliveira da Geflor pela ajuda e apoio.

Ao Sr. Rui Pedro Ribeiro da empresa Agrocortex pela ajuda na viagem de campo ao Acre. Ao Mc. Eric R. A. Diaz Maldonado (INIFAP) e ao Sr. Paulo J. P. Fontes e ao Sr. Waldemar Vieira Lopes pelo apoio e consulta.

Ao Prof. Dr. Ivan Tomaselli e à Sra. Sofia R. Hiraku pelo apoio integral ao projeto em especial à realização das missões em outros países.

Resumo

A identificação rápida e confiável do país de origem de madeiras em risco de extinção e protegidas por lei e a diferenciação entre espécies produtoras de madeiras dentro de um mesmo gênero são formas combate direto à extração e comércio ilegais de insumos florestais e constituem contribuição indireta à preservação do meio ambiente. Este trabalho apresenta resultados dessas duas abordagens. A primeira, dedicada a verificação da viabilidade de identificar geograficamente as madeiras de *S. macrophylla* oriundas de cinco países da América Latina - Bolívia, Brasil, Guatemala, México e Peru - utilizando dois espectrômetros NIR portáteis distintos e análise multivariada. A segunda parte consistiu em discriminar duas espécies consideradas mogno genuíno: *S. macrophylla* e *S. humilis*, usando essa mesma metodologia. Essas duas espécies são consideradas botanicamente como mogno genuíno, apesar de ocorrerem em regiões diferentes. Ambas estão incluídas no Anexo II da Convenção (CITES), logo sua identificação e discriminação se torna importante.

Com base na análise do espectro registrado na madeira, a espectroscopia no infravermelho próximo (NIRS) tem provado ser uma técnica promissora tanto para a identificação rápida e em campo de espécies florestais desprovidas de material botânico auxiliar, assim como para identificação da origem geográfica. Dois espectrômetros NIR portáteis foram utilizados e a análise dos dados foi realizada com uso dos algoritmos de seleção de amostras, Kennard-Stone (KS) e Duplex, para posterior construção de modelos de discriminação (PLS-DA, *Partial Least Square Determinant Analysis*) e modelagem (SIMCA, *Soft Independent Modeling of Class Analogy*). Analisando as Taxas de Eficiência (TEFs) dos modelos construídos, o modelo da Bolívia obteve melhores resultados com o equipamento MicroPHAZIR, tendo obtido TEF de 100% com ambos, PLS-DA e SIMCA. Já o modelo para o Brasil apresentou melhor resultado com MicroPHAZIR e PLSDA (TEF=100%). Por outro lado, para a Guatemala, México e Peru, o espectrômetro MicroNIR com PLS-DA apresentou melhor desempenho, tendo obtido TEFs iguais a 100; 94,9 e 99,4%, respectivamente. Para análise das espécies produtoras de madeiras *S. macrophylla* e *S. humilis*, obteve-se o valor 100% de TEF para o modelo PLS-DA e 98,5% para o modelo SIMCA, mostrando que é possível a discriminação entre essas espécies pertencentes a um mesmo gênero.

Palavras-chave: *procedência, comércio ilegal, madeira tropical, PLS-DA, SIMCA.*

Abstract

The rapid and reliable identification of the country of origin of timber in danger of extinction and protected by law and the differentiation between timber species within the same genus are forms of direct combat to the illegal extraction and trade of forest inputs and constitute an indirect contribution to the preservation the environment. This paper presents results from these two approaches. The first one, dedicated to verify the feasibility of geographically identifying *S. macrophylla* wood from five Latin American countries - Bolivia, Brazil, Guatemala, Mexico and Peru - using two distinct portable NIR spectrometers and multivariate analysis. The second part consisted in discriminating two species considered genuine mahogany: *S. macrophylla* and *S. humilis*, using this same methodology. These two species are considered botanically as genuine mahogany, although they occur in different regions. Both are included in Annex II of the Convention (CITES) so their identification and discrimination become important.

Based on the analysis of the spectrum recorded in the wood, near infrared spectroscopy (NIRS) has proved to be a promising technique for the rapid and field identification of forest species lacking auxiliary botanical material, as well as for identification of geographic origin. Two portable NIR spectrometers were used and the data analysis was performed using the Kennard-Stone (KS) and Duplex sample selection algorithms, for later construction of Partial Least Square Determinant Analysis (PLS-DA) and modeling (SIMCA, Soft Independent Modeling of Class Analogy). Analyzing the Efficiency Rates (EFRs) of the constructed models, the Bolivian model obtained better results with the MicroPHAZIR equipment, obtaining 100% EFR with both PLS-DA and SIMCA. The model for Brazil presented better results with MicroPHAZIR and PLSDA (EFR = 100%). On the other hand, for the Guatemala, Mexico and Peru, the MicroNIR spectrometer with PLS-DA presented better performance, obtaining EFRs equal to 100; 94.9 and 99.4%, respectively. For the analysis of *S. macrophylla* and *S. humilis* wood species, 100% EFR was obtained for the PLS-DA model and 98.5% for the SIMCA model, showing that it is possible to discriminate between these species belonging to the same genre.

Keywords: origin, illegal logging trade, tropical wood, PLS-DA, SIMCA.

ÍNDICE

LISTA DE ABREVIATURAS E ACRÔNIMOS

LISTA DE TABELAS

LISTA DE FIGURAS

1	INTRODUÇÃO E OBJETIVO	1
2	FUNDAMENTAÇÃO TEÓRICA	3
2.1	Exploração e comércio de madeira mogno	3
2.2	Componentes químicos da Madeira	4
2.3	Tecnologia NIR	7
2.4	Equipamentos NIRS	8
2.5	Análise multivariada	10
2.5.1	Métodos de seleção de amostras	10
2.5.1.1	Algoritmo Kennard-Stone	11
2.5.1.2	Algoritmo Duplex	12
2.5.2	<i>Partial Least Square Determinant Analysis</i> (PLS-DA)	13
2.5.3	<i>Soft Independent Modeling of Class Analogy</i> (SIMCA)	16
2.5.4	Figuras de mérito	18
3	MATERIAIS E MÉTODOS	20
3.1	Obtenção e preparação de amostras	20
3.1.1	Amostras de <i>S.macrophylla</i> dos países da América Latina	20
3.1.2	Amostras de <i>S.macrophylla</i> e <i>S.humilis</i> coletadas na Guatemala	21
3.2	Aquisição dos espectros de infravermelho (IR)	21
3.3	Processamento dos espectros e análises de dados	22
3.4	Seleção de amostras para treinamento e validação	23
3.4.1	Algoritmo Kennard-Stone	23
3.4.2	Algoritmo Duplex	24
3.5	Modelos quimiométricos	25
3.5.1	<i>Partial Least Square Determinant Analysis</i> (PLS-DA)	25
3.5.2	<i>Soft Independent Modeling of Class Analogy</i> (SIMCA)	26
4	RESULTADOS E DISCUSSÃO	28
4.1	Determinação da origem de amostras de madeira mogno (<i>S. macrophylla</i>) dos cinco países	28

4.1.1 Espectros NIR das amostras de <i>Swietenia macrophylla</i> dos cinco países	28
4.1.2 Análise quimiométrica	32
4.1.2.1 PLS-DA	32
4.1.2.1.1 Algoritmo Kennard-Stone	32
4.1.2.1.2 Algoritmo Duplex	40
4.1.2.2 SIMCA	48
4.1.2.2.1 Algoritmo Kennard-Stone	48
4.1.2.2.2 Algoritmo Duplex	50
4.1.2.3 Comparação entre os modelos para identificação de origem	52
4.1.2.4 Conclusões parciais	53
4.2 Discriminação das espécies <i>S. macrophylla</i> e <i>S. humilis</i>	54
4.2.1 Espectros NIR das madeiras de <i>S. macrophylla</i> e <i>S. humilis</i>	54
4.2.2 Análise Quimiométrica	56
4.2.2.1 PLS-DA	56
4.2.2.2 SIMCA	59
5 CONCLUSÕES GERAIS	60
REFERÊNCIAS BIBLIOGRÁFICAS	61

LISTA DE ABREVIATURAS E ACRÔNIMOS

AQUA	Laboratório de Automação, Quimiometria e Química Ambiental
CITES	Convenção sobre o Comércio Internacional das Espécies da Flora e Fauna Selvagens em Perigo de Extinção (<i>Convention on International Trade in Endangered Species of Wild Fauna and Flora</i>)
CoC	Cadeira de Custódia
CV	Validação Cruzada (<i>Cross-Validation</i>)
CVCE	<i>Cross Validation Classification Error</i>
FN	Falso Negativo
FP	Falso Positivo
IBAMA	Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis
InGaAs	Índio-Gálio-Arsênio
IQ/UnB	Instituto de Química da Universidade de Brasília
IR	Infravermelho (<i>Infrared</i>)
KS	Kennard-Stone
LPF/SFB	Laboratório de Produtos Florestais do Serviço Florestal Brasileiro
MOEMS	<i>Micro-Optoelectronic-Mechanical-System</i>
MMA	Ministério do Meio Ambiente
NIR	Infravermelho Próximo (<i>Near Infrared</i>)
NIRS	Espectroscopia no Infravermelho Próximo (<i>Near Infrared Spectroscopy</i>)
Out_Trein	<i>Outlier</i> do conjunto de treinamento
Out_Val	<i>Outlier</i> do conjunto de validação
PCA	<i>Principal Component Analysis</i>
PC	<i>Principal Components</i>
PLS	<i>Partial Least Squares</i>
PLS-DA	<i>Partial Least Squares Discriminant Analysis</i>
RMSEC	<i>Root Mean Square Error</i>
RMSECV	<i>Root Mean Square Error of Cross Validation</i>
TEF	Taxa de Eficiência
TFN	Taxa de Falso Negativo
TFP	Taxa de Falso Positivo
UnB	Universidade de Brasília
VL	Variáveis Latentes
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

LISTA DE TABELAS

Tabela 1. Localidade, região e quantidade de amostras das espécies <i>S.macrophylla</i> e <i>S.humilis</i> coletadas na Guatemala.	21
Tabela 2. Número de espectros usados nos conjuntos de treinamento e validação dos cinco países da América Latina para os equipamentos portáteis MicroPHAZIR™ e MicroNIR™, utilizando-se o algoritmo Kennard-Stone.	24
Tabela 3. Número de espectros dos conjuntos de treinamento e validação dos cinco países da América Latina para os equipamentos portáteis MicroPHAZIR™ e MicroNIR™, utilizando-se o algoritmo Duplex.	24
Tabela 4. Parâmetros dos modelos PLS-DA e número de outliers identificados para o MicroNIR e MicroPHAZIR, utilizando-se o algoritmo Kennard-Stone.	32
Tabela 5. Figuras de mérito dos modelos PLS-DA construídos utilizando-se o algoritmo-Kennard-Stone e calculadas a partir do conjunto de validação.	33
Tabela 6. Resultados dos parâmetros dos modelos PLS-DA e número de <i>outliers</i> identificados para o MicroNIR e MicroPHAZIR, respectivamente, utilizando-se o algoritmo Duplex.	41
Tabela 7. Figuras de mérito para os modelos PLS-DA calculadas a partir do conjunto de validação, utilizando-se o algoritmo Duplex.	41
Tabela 8. Figuras de mérito para os modelos de SIMCA calculadas a partir do conjunto de validação, utilizando-se o algoritmo Kennard-Stone.	48
Tabela 9. Correlação entre o número de amostras identificadas na etapa de validação pelo modelo SIMCA de cada país de acordo com os dados de cada espectrômetro, utilizando-se o algoritmo Kennard-Stone.	49
Tabela 10. Figuras de mérito para os modelos de SIMCA calculadas a partir do conjunto de validação, utilizando-se o algoritmo Duplex.	50

Tabela 11. Correlação entre o número de amostras identificadas na etapa de validação pelo modelo SIMCA de cada país de acordo com os dados de cada espectrômetro, utilizando-se o algoritmo Duplex.	51
Tabela 12. Conjunto de taxas de eficiências dos modelos PLS-DA e SIMCA para os dados dos dois espectrômetros portáteis (MicroNIR™ e MicroPHAZIR™) das amostras dos países a partir dos dois algoritmos de seleção de amostra (Kennard-Stone e Duplex).	52
Tabela 13. Resultados dos parâmetros dos modelos PLS-DA e número de <i>outliers</i> identificados para a discriminação das espécies <i>S. macrophylla</i> e <i>S. humilis</i> .	56
Tabela 14. Figuras de mérito para os modelos PLS-DA das espécies <i>S. macrophylla</i> e <i>S. humilis</i> calculadas a partir do conjunto de validação.	57
Tabela 15. Figuras de mérito para os modelos PLS-DA das espécies <i>S. macrophylla</i> e <i>S. humilis</i> calculadas a partir do conjunto de validação.	59
Tabela 16. Correlação entre o número de amostras identificadas na etapa de validação pelo modelo SIMCA de cada espécie.	59

LISTAS DE FIGURAS

- Figura 1.** Representação do monômero que origina a estrutura química da celulose (Siau, 1984). 6
- Figura 2.** Estrutura da lignina. (Haider et al, 1964). 6
- Figura 3.** Representação das estruturas químicas dos elementos básicos formadores da lignina (Pastore, 2004). 7
- Figura 4.** Representação das estruturas químicas de açúcares componentes da hemicelulose (Pastore 2004). 7
- Figura 5.** Espectrômetros NIR portáteis: **a)** MicroNIR™, **b)** MicroPHAZIR™. 9
- Figura 6.** **a)** duas classes delimitadas por regiões distintas bem definidas, exemplo de modelagem. **b)** um limite separa duas classes sem distinção de regiões, exemplo de discriminação (Radionova et. al 2016). 10
- Figura 7.** Ilustração de seleção de amostra utilizando o algoritmo Kennard-Stone a partir das distâncias euclidianas e critério de *maxmin*. 12
- Figura 8.** Representação esquemática de PLS1-DA e PLS2-DA (Lopes, 2015). 13
- Figura 9.** Gráfico de RMSECV por número de variáveis latentes (VL). Método para escolha do número de VL para construções dos modelos PLS-DA. 15
- Figura 10.** Ilustração do gráfico de T^2 de Hotelling versus Q residuais, mecanismo que possibilita a identificação e exclusão de *outliers*. 18
- Figura 11.** Esquema geral do processo de aquisição dos espectros NIR para amostras da madeira mogno. 22
- Figura 12.** Passo a passo para o desenvolvimento e validação do modelo PLS1-DA. X_{cal} : matriz de espectros do conjunto de calibração/treinamento; Y_{cal} : vetor de classes da calibração para cada país; X_{val} : matriz do conjunto de validação; Y_{val} : vetor de classe de validação 26
- Figura 13.** Média dos espectros brutos de NIR das amostras da Bolívia, Brasil, Guatemala, México e Peru obtidos com MicroNIR™ **(a)** e MicroPHAZIR™ **(b)**. 29

- Figure 14.** Média dos espectros das amostras da madeira mogno de cada país após o pré-processamento com primeira derivada SG (Savitzky-Golay). Espectros pré-processados do MicroPHAZIR™ **(a)** e MicroNIR™ **(b)**. As bandas espectroscópicas mais significativas estão numeradas. 31
- Figura 15.** Gráfico dos valores de T^2 de Hotelling e resíduos Q das amostras de **(a)** Bolívia, **(b)** Brasil, **(c)** Guatemala, **(d)** México e **(e)** Peru obtidos com PLS-DA e algoritmo KS a partir dos dados do MicroNIR™. Legenda: (▼) Brasil, (*) Guatemala, (■) México, (+) Peru e (◇) Bolívia. 35
- Figura 16.** Gráfico dos valores de T^2 de Hotelling e resíduos Q das amostras de **(a)** Bolívia, **(b)** Brasil, **(c)** Guatemala, **(d)** México e **(e)** Peru obtidos com PLS-DA e algoritmo KS a partir dos dados do MicroPHAZIR™. Legenda: (▼) Brasil, (*) Guatemala, (■) México, (+) Peru e (◇) Bolívia, (--) Limites de confiança (99,9%). 36
- Figura 17.** Distribuição dos valores de classe estimados para os conjuntos de treinamento e validação obtidos com PLS-DA e algoritmo KS a partir dos dados do MicroNIR™. **(a)** Bolívia, **(b)** Brasil, **(c)** Guatemala, **(d)** México e **(e)** Peru. Legenda: (▼) Brasil, (*) Guatemala, (■) México, (+) Peru e (◇) Bolívia, (--) Limite de discriminação. 38
- Figura 18.** Distribuição dos valores de classe estimados para os conjuntos de treinamento e validação obtidos com PLS-DA e algoritmo KS a partir dos dados do MicroPHAZIR™. **(a)** Bolívia, **(b)** Brasil, **(c)** Guatemala, **(d)** México e **(e)** Peru. Legenda: (▼) Brasil, (*) Guatemala, (■) México, (+) Peru e (◇) Bolívia, (--) Limite de discriminação. 39
- Figura 19.** Gráfico dos valores de T^2 de Hotelling e resíduos Q das amostras de **(a)** Bolívia, **(b)** Brasil, **(c)** Guatemala, **(d)** México e **(e)** Peru obtidos com PLS-DA e algoritmo Duplex a partir dos dados do MicroNIR™. Legenda: (▼) Bolívia, (*) Brasil, (■) Guatemala, (+) México e (◇) Peru, (--) Limites de confiança (99,9%). 43
- Figura 20.** Gráfico dos valores de T^2 de Hotelling e resíduos Q das amostras de **(a)** Bolívia, **(b)** Brasil, **(c)** Guatemala, **(d)** México e **(e)** Peru obtidos com PLS-DA e algoritmo Duplex a partir dos dados do MicroPHAZIR™. Legenda: (▼) Bolívia, (*) Brasil, (■) Guatemala, (+) México e (◇) Peru, (--) Limites de confiança (99,9%). 44

- Figura 21.** Distribuição dos valores de classe estimados para os conjuntos de treinamento e validação obtidos com PLS-DA e algoritmo Duplex a partir dos dados do MicroNIRTM. **(a)** Bolívia, **(b)** Brasil, **(c)** Guatemala, **(d)** México e **(e)** Peru. Legenda: (▼) Bolívia, (*) Brasil, (■) Guatemala, (+) México e (◇) Peru, (--) Limite de discriminação. 46
- Figura 22.** Distribuição dos valores de classe estimados para os conjuntos de treinamento e validação obtidos com PLS-DA e algoritmo Duplex a partir dos dados do MicroNIRTM. **(a)** Bolívia, **(b)** Brasil, **(c)** Guatemala, **(d)** México e **(e)** Peru. Legenda: (▼) Bolívia, (*) Brasil, (■) Guatemala, (+) México e (◇) Peru, (-) Limite de discriminação. 47
- Figura 23.** Média dos espectros NIR das madeiras das espécies *S. macrophylla* e *S. humilis* ou mogno antes do pré-processamento. 54
- Figure 24.** Média dos espectros NIR das madeiras das espécies *S. macrophylla* e *S. humilis* ou mogno após o pré-processamento com primeira derivada SG (Savitzky-Golay). 55
- Figura 25.** Gráfico dos valores de T^2 de Hotelling e resíduos Q das amostras de **(a)** *S. macrophylla* e **(b)** *S. humilis* obtidos com PLS-DA e algoritmo Duplex a partir dos dados do MicroNIRTM. Legenda: (▼) *S. macrophylla*, (*) *S. humilis*, (--) Limites de confiança (99,9%). 58
- Figura 26.** Distribuição dos valores de classe estimados para os conjuntos de treinamento e validação obtidos com PLS-DA a partir dos dados do MicroNIRTM. **(a)** *S. humilis* e **(b)** *S. macrophylla*. Legenda: (▼) *S. macrophylla* e (*) *S. humilis*, (--) Limite de discriminação. 58

1. INTRODUÇÃO E OBJETIVOS

Swietenia macrophylla, *Swietenia humilis* e *Swietenia mahogany* são espécies botanicamente consideradas mogno genuíno. A ocorrência da espécie *S. macrophylla* é bastante vasta, distribuindo-se por grande parte da floresta tropical nas Américas, começando no México e terminando na Bolívia e com um braço de extensão que percorre o sul da Amazônia brasileira (Pastore et al, 2011; Bergo et al, 2016). Já a ocorrência da espécie *S. humilis* restringe-se à América Central, desde do México até a Costa Rica (Salazar, 1998). Devido à sua marrom avermelhada, boa trabalhabilidade e moderada resistência ao intemperismo e à ação biológica (Coradin et al, 2009), o mogno é utilizado em diversas áreas como instrumentos musicais, carpintaria, mobília, revestimento de interiores entre outros (Silva et al, 2013). Assim, tornou-se uma madeira valiosa e mundialmente comercializada.

Como resultado do extrativismo extensivo e seletivo, as três espécies consideradas mogno, tornaram-se espécies em risco de extinção (Pastore et al, 2011 e Salazar, 1998) e foram inclusas no Apêndice II da Convenção Internacional de Comércio de Espécie da Flora e Fauna em Perigo de Extinção – CITES (*Convention on International Trade in Endangered Species of Wild Fauna and Flora*) a nível internacional. No Brasil, com a finalidade de assegurar a sobrevivência da espécie de mogno nativa, a *S. macrophylla* foi classificada como em risco de extinção, na Lista Oficial das Espécies da Flora Brasileira Ameaçadas de Extinção, pelo decreto nº 6.101 da portaria 443/2014 do Ministério do Meio Ambiente (Brasil, 2014; Degen et al, 2013; Dormontt et al, 2015; Nuoniah et al, 2016).

A inclusão do mogno no apêndice II da CITES não proíbe o comércio internacional dessas espécies, mas requer que o volume comercializado e a situação de conservação da espécie sejam documentados. Adicionalmente, requer que o comércio seja efetuado apenas após o governo determinar que o comércio não será feito em detrimento da sobrevivência das populações selvagens. Essa é uma medida para regular o recurso existente e controlar o comércio ilegal, além de incentivar a utilização sustentável e cientificamente manejada do mogno (Bruce et al, 2000).

Instituições internacionais e nacionais, como o Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA) se esforçam para prevenir e combater a extração e o comércio ilegais de madeira e de seus produtos derivados, que é a causa de muitos problemas econômicos e socioambientais.

Uma das formas de combate à extração e comércio ilegais da madeira é o rastreamento da origem da espécie madeireira. Com a identificação da procedência, o importador terá uma garantia da produção legal do recurso florestal. Assim, são necessárias ferramentas práticas de rastreamento, como, por exemplo, a certificação florestal, sistema de certificação (por exemplo o DOF, Documento de Origem Florestal) que verifica a legalidade da madeira na cadeia de custódia (CoC) (Nuroniah et al, 2016). Contudo, em vários pontos do trajeto comercial podem ocorrer falsificações dos certificados, como falsificações de chips eletrônicos e documentação via papel (Degen et al, 2013; Höltken et al, 2012). Logo, é necessário desenvolver ferramentas científicas que podem validar e verificar a origem da extração madeireira para legitimar o comércio.

No Brasil, ocorre apenas uma única espécie de mogno, *Swietenia macrophylla*, enquanto na América Central e Caribe ocorrem as três espécies consideradas mogno genuíno, já citadas anteriormente. Atualmente quase todo mogno comercializado é originário de florestas primárias, e os estoques e recursos do gênero *Swietenia* têm declinado, particularmente na América Central e Caribe. Assim, a distinção entre as espécies consideradas mogno genuíno é importante, para distribuir o peso e a proporção do extrativismo entre as espécies e não gerar a extinção de uma espécie em detrimento da outra.

Os dados da espectroscopia de infravermelho próximo (NIRS - *Near Infrared Spectroscopy*) variam seu intervalo espectral de 800 a 2500 nm, fornecendo informações relacionadas à composição química da madeira. As diferenças observáveis em um espectro de NIR (*Near Infrared*) são bem sutis, assim a avaliação do espectro é realizada por meio de análise multivariada (Pastore et al, 2011).

Assim, os objetivos desse trabalho foram: a determinação via NIRS e análise multivariada da origem da madeira mogno da espécie *Swietenia macrophylla* proveniente de cinco países da América Latina - Bolívia, Brasil, Guatemala, México e Peru, com a finalidade de obter um método que possa ser aplicado no rastreamento da origem de espécies madeireiras no âmbito de fiscalização ambiental. Outro aspecto importante é a identificação de espécies produtoras de madeira anatomicamente semelhantes. A separação entre duas espécies de mogno consideradas mogno genuíno (*S. macrophylla* e *S. humilis*) ainda não foi reportada na literatura. Logo, também será avaliada a identificação/discriminação dessas espécies empregando a mesma abordagem utilizada na determinação da origem.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. Exploração e comércio da madeira mogno

O comércio ilegal de espécies florestais produtoras de madeira protegidas é um sério problema ecológico (Degen et al, 2012; Dormontt et al, 2015; Soares et al, 2017). Acredita-se que a extração ilícita de madeira seja uma das principais causas do desflorestamento e que o comércio ilegal de madeira e seus derivados criam desvantagens econômicas em relação aos produtos de fontes legais e ecologicamente sustentáveis (Degen et al, 2012; Dormontt et al, 2012).

Nesse contexto, países da União Europeia, EUA, Canadá, Austrália e Japão, principais compradores do mogno (Ravidran et al, 2018), implementaram regras sobre o comércio de madeira, em que o importador é obrigado a procurar garantias sobre a produção legal da madeira, além de certificação da sustentabilidade dos produtos madeireiros, isto é, certificação que os produtos madeireiros foram extraídos de maneira sustentável e consoante com as leis dos seus países de origem (Lowe et al, 2016). Essas regulamentações aumentam a pressão em cima das concessões florestais e dos comerciantes para verificar a fonte de origem das madeiras.

Para atingir o objetivo de rastreamento da origem das espécies madeireiras, são necessárias ferramentas práticas, como, por exemplo, a certificação florestal, projetada como resposta a um manejo florestal sustentável e é um dos sistemas de certificação que verifica a legalidade da madeira na cadeia de custódia (CoC). A CoC é um processo de registro metódico cronológico e detalhado sobre controle e caminho que o insumo madeireiro percorre desde sua extração até a importação. A certificação florestal se apropria de informações filogeográficas para documentação da origem de determinada espécie (Nuroniah et al, 2016). Contudo, nos pontos de importação podem ocorrer falsificações dos certificados, como falsificações de chips eletrônicos e documentação via papel (Degen et al, 2013; Höltkén et al, 2012).

Os mognos verdadeiros são uma das madeiras mais importantes no comércio internacional e são obtidos de árvores do gênero *Swietenia* (*S. humilis*, *S. mahogany* e *S. macrophylla*). O comércio de mogno teve início há quase 500 anos com *S. mahogany* (mogno caribenho) e se estendeu para *S. humilis*, mas a exploração exaustiva dessas duas espécies reduziu dramaticamente seus estoques naturais. Como resposta, os madeireiros passaram a explorar a *S. macrophylla*, uma espécie que ocorre na América Central e ao longo de um arco ao sul da Amazônia brasileira, boliviana e peruana.

As características do mogno contribuíram para o desenvolvimento de um mercado altamente lucrativo e para uma taxa de exploração que alarma grupos

ambientalistas e vários governos que se preocupam com a extinção da espécie. A fim de responder à percepção de ameaça de extinção do mogno, desde o início da década de 1990, o governo brasileiro tem adotado uma série de ações regulamentadoras deste mercado (Grogan et. al 2002). Como ponto de partida, a extração de madeira no Brasil tinha que ser certificada pelo governo federal através do Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA). Antes da proibição da extração do mogno por meio decreto nº4.722 do Ministério do Meio Ambiente (Brasil, 2003) a maior parte da espécie era explorada através de planos de manejo florestal autorizados pelo governo. Além disso, o governo brasileiro determinava uma quota anual de exportação para ajudar a proteger o estoque do mogno na floresta.

Em novembro de 2003, apesar de forte oposição do governo brasileiro, o mogno foi incluído no apêndice II da CITES, a qual dita regras restritas para o comércio internacional, mas não determina a proibição. Mesmo assim o Brasil continuou proibindo a extração do mogno. Em 2016, através do decreto nº 34/2016, o governo brasileiro instituiu o Regulamento Sobre o Comércio Internacional de Espécies de Fauna e Flora Silvestres Ameaçadas de Extinção (Brasil, 2016) com a finalidade de assegurar a aplicação das disposições previstas na CITES.

2.2. Componentes Químicos da Madeira

A parede celular das células da madeira é constituída, principalmente, por celulose, hemicelulose e lignina, além dos extrativos (Siau et al, 1984; Dinwoodie et al, 2000). A celulose (Figura1) é um polímero linear de alto peso molecular formado unicamente por monômeros de glucose e presente em maior quantidade na madeira, cerca de 50% em massa (Dinwoodie et al, 2000). Para a madeira mogno, o teor de celulose é cerca de 63,3% (Rutiaga-Quiñones et al, 2001).

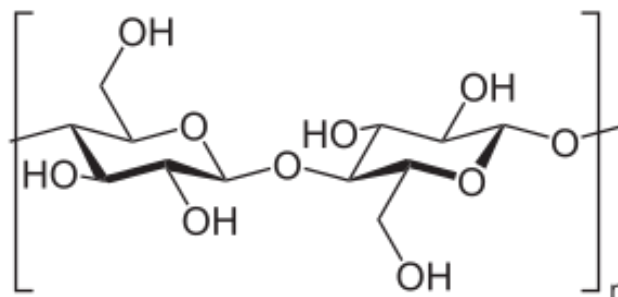


Figura 1. Representação do monômero que origina a estrutura química da celulose (Siau, 1984).

A lignina (Figura 2), uma macromolécula derivada da polimerização de três elementos estruturais básicos: álcool p-cumarílico, álcool coniferílico e álcool sinapílico (Figura 3) (Siau et al, 1984). Esse polímero confere uma maior resistência mecânica à madeira, de forma a reduzir seu caráter higroscópico e a protegê-la contra a deterioração microbiológica (Siau et al, 1984). Essa macromolécula compreende, aproximadamente, 30% a 25% em massa. Para o mogno, essa porcentagem corresponde a 31,1% (Rutiaga-Quiñones et al, 2001).

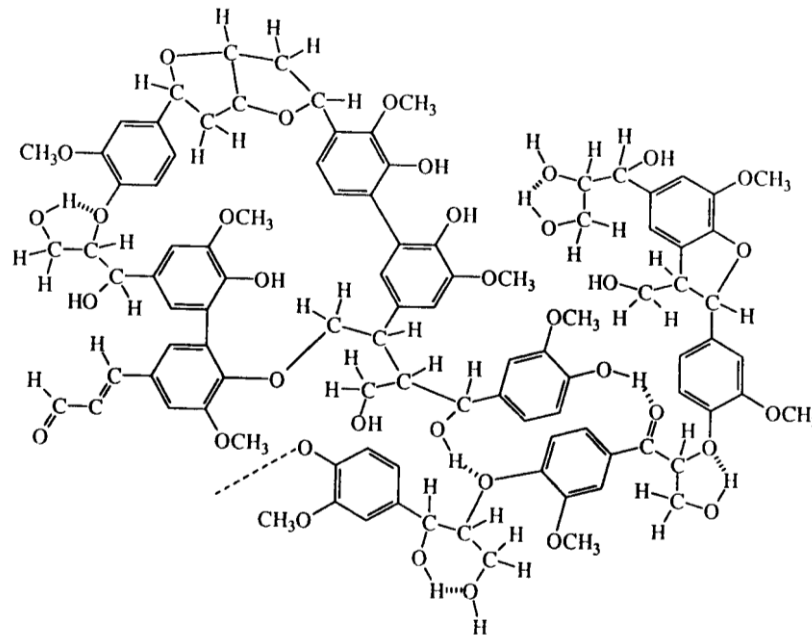


Figura 2. Estrutura da lignina. (Haider et al, 1964).

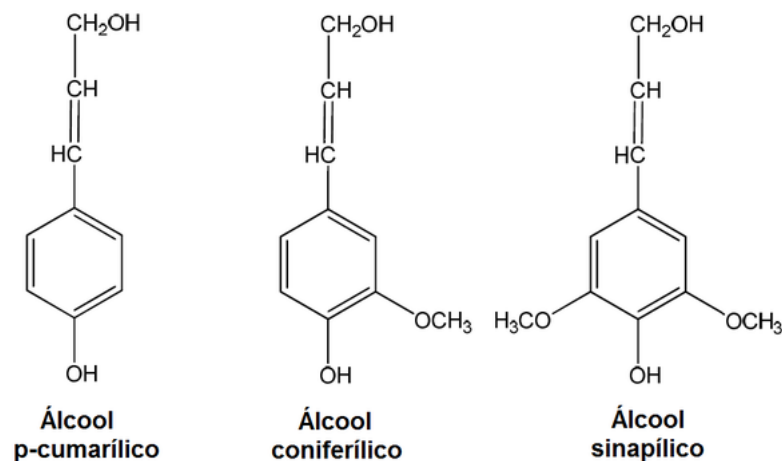


Figura 3. Representação das estruturas químicas dos elementos básicos formadores da lignina (Pastore, 2004).

Diferentemente da celulose, as hemiceluloses são polímeros de cadeias menores constituídos por diversas unidades de açúcares, incluindo as pentoses e hexoses (Figura 4), e compõem entre 27% a 35%, dependendo da espécie de madeira (Dinwoodie et. al 2000).

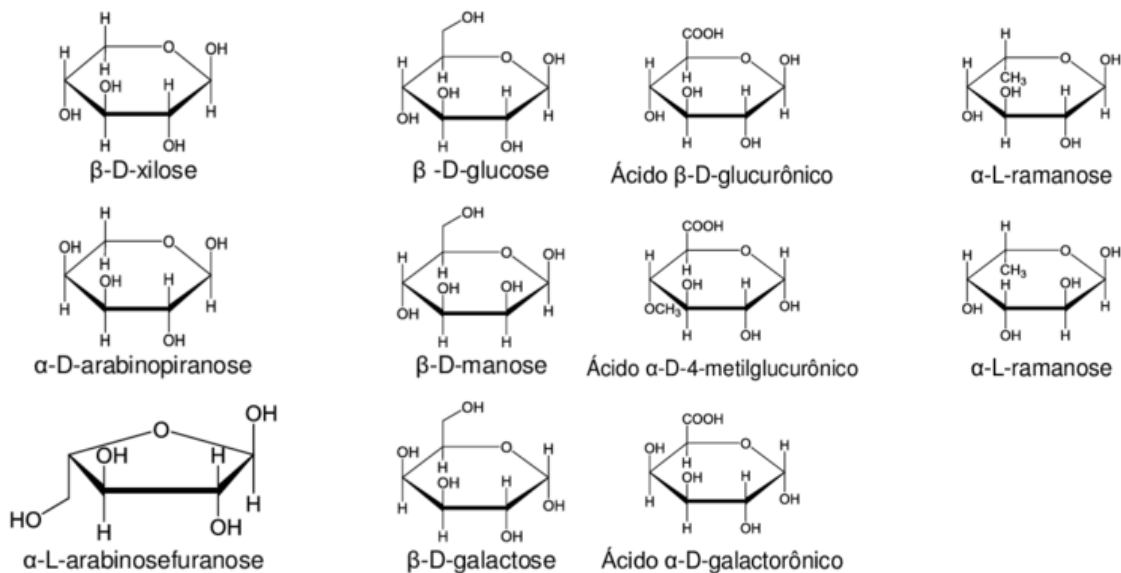


Figura 4. Representação das estruturas químicas de açúcares componentes da hemicelulose (Pastore, 2004).

Na estrutura celular também são encontrados outros componentes não participantes da formação da parede celular em uma menor quantidade, geralmente em concentrações menores que 10% em massa (Tsuchikawa, 2007). No entanto, essa porcentagem pode aumentar para madeiras tropicais (Siau et al, 1984). Os extrativos são compostos orgânicos de baixo peso molecular, abrangendo diversas substâncias como resinas, taninos, flavonoides, quinonas e óleos essenciais (Tsuchikawa, 2007) frequentemente presentes no cerne da madeira, influenciando na coloração e atribuindo uma maior resistência ao ataque de fungos e insetos (Dinwoodie et al, 2000). Além desses compostos orgânicos, sais minerais e ácidos (metanóico e etanóico) também contribuem para a composição química da madeira (Pastore, 2004). De acordo com Silva et al (2013), o teor de extrativos na madeira de mogno é de aproximadamente 8,38%.

2.3. Tecnologia NIRS

As madeiras das árvores que crescem em várias partes do mundo são influenciadas pelo clima local, tipo de solo, relevo, densidade florestal, práticas de silvicultura, vida animal, doenças e outros fatores, além da variação genética das árvores em questão. Esses parâmetros e características têm influência no conteúdo de extrativos, celulose, lignina e hemicelulose (Sandak et al, 2011; Sandak et al, 2015), bem como na densidade dos anéis de crescimento anuais das madeiras. Ou seja, uma espécie madeireira que cresceu em uma região determinada carrega informações físico-químicas de todas as condições daquele local específico. A ciência que lida com a identificação de madeira por meio de análise química é conhecida como quimiotaxonomia, podendo ser realizada por análises químicas tradicionais (via úmida) (Sandermann, 1962) ou por métodos espectroscópicos (Rana et al, 2008; Wang et al, 2017).

Em função das diferenças na composição química da madeira, amostras de espécies produtoras de madeira como o mogno podem apresentar espectros de infravermelho próximo distintos de acordo com diferentes regiões ou países. NIRS (*Near Infrared Spectroscopy*) é uma técnica instrumental de análise, na qual se obtém informações da interação da radiação eletromagnética com a matéria, fornecendo dados espectrais da região entre 800 a 2.500 nm. Os dados de um espectro contêm informações sobre ligações químicas dos componentes presentes na amostra de maneira e como esses componentes estão rearranjados na superfície sólida analisada. A olho nu, as diferenças de um espectro NIR de uma amostra para a outra é bastante sutil, logo é necessário o uso de modelos quimiométricos para analisar o espectro. A combinação do NIRS com quimiometria fornece métodos e modelos significantes para solução de problemas sendo referenciado como **tecnologia NIRS** (Pastore et al, 2011).

A identificação da madeira por meio da anatomia da madeira ou qualquer outro método é uma etapa necessária no uso da tecnologia NIR para a construção de modelos corretos. A tecnologia NIRS é método indireto e se apoia na análise de espectros em condições pré-estabelecidas de amostras, tais como umidade e rugosidade da superfície da amostra, além de outros parâmetros, uma vez que tais fatores afetam as medidas espectroscópicas e conseqüentemente a acurácia do modelo quimiométrico a ser desenvolvido (Sandak et al, 2011; Tsuchikawa et al, 2007).

A tecnologia NIRS permite uma análise rápida e não-destrutiva da amostra, pode ser aplicada em amostras sólidas, requer o pouco preparo da amostra e na maioria das vezes não gera resíduos químicos. Além disso, o procedimento

operacional de um equipamento de NIR portátil requer mínimo treino. Essa técnica permite a determinação de vários componentes químicos ao mesmo tempo com alta precisão e acurácia a um custo relativamente baixo (Sandak et al, 2011, Tsuchiwaka et al, 2007, 2015).

O Laboratório de Produtos Florestais (LPF) do Serviço Florestal Brasileiro (SFB) em parceria com o Instituto de Química da Universidade de Brasília (UnB) vem trabalhando em vários estudos associando a tecnologia NIR e identificação de madeira. Os espectros NIR, do cerne de madeira, em combinação com o PLS-DA (do inglês, *Partial Least Squares for Discriminant Analysis* – Análise Discriminante por Mínimos Quadrados Parciais) para discriminação de espécies produtoras de madeira, como a separação de *S. macrophylla* de outras três espécies similares *Carapa guianensis* (andiroba), *Cedrela odorata* (cedro) e *Micropholis melinoniana* (curupixá), foi demonstrada com bastante êxito (Pastore et al, 2011; Braga et al, 2014). Através desse método, Bergo et al (2016) foram capazes de identificar amostras de *S. macrophylla* de 27 países distintos usando NIRS e PLS-DA.

Um aprimoramento à tecnologia NIR foi o advento dos espectrômetros NIR portáteis, uma ferramenta prática que permite a identificação da madeira, o monitoramento e rastreamento *in situ* de toras ou tábuas de madeira e para validar áreas exploradas. Muitos espectrômetros NIR portáteis estão sendo usados nas indústrias agrícola e alimentícia (Ribeiro et al, 2016). Meder et al (2011) comparou a performance de um equipamento NIR de bancada típico de laboratório (FTNIR) e dois equipamentos portáteis para a previsão do rendimento de pasta Kraft e do teor de celulose em árvores em pé e encontrou resultados equivalentes entre os espectrômetros. Soares et al (2017) obtiveram sucesso ao discriminar seis espécies produtoras de madeiras amazônicas anatomicamente semelhantes ao mogno usando um instrumento NIRS portátil. Contudo, ainda não apareceram estudos para determinar a origem da madeira em um contexto de combate ao comércio madeireiro ilegal. Alguns estudos já foram desenvolvidos na área de identificação de espécies produtoras de madeira, porém a separação entre espécies de mogno (*S. macrophylla* e *S. humilis*) ainda não foi estudada.

2.4 Equipamentos NIRS

Um espectrômetro NIR portátil é composto por uma fonte de luz, um sistema óptico e um detector. A fonte de luz pode ser uma lâmpada de halogênio ou LED's. Normalmente é usada uma lâmpada composta por um filamento de tungstênio selada em um bulbo com vapor de um elemento halogêneo, visto que é pequena e robusta.

O sistema óptico pode ser composto por filtros ópticos, monocromadores ou outros dispositivos que decompõem o feixe de luz policromático em comprimentos de onda discretos. Este deve proporcionar uma largura de banda estreita em relação à banda de absorção a ser medida e uma elevada intensidade em toda a gama de comprimento de onda analisada (Aenugu, 2011).

Os detectores normalmente usados incluem os de semicondutor, sulfureto de chumbo (PbS) e índio-gálio-arsénio (InGaAs). Os detectores de semicondutor são rápidos, silenciosos, pequenos e altamente sensíveis da região visível até 1100 nm. Os detectores de PbS são mais lentos, mas muito populares visto que têm alta durabilidade, baixo preço e são sensíveis desde 1100 nm até 2500 nm. Os detectores de InGaAs são os mais caros, pois combinam as características de velocidade e tamanho do detector de silício com o alcance do comprimento de onda do detector de PbS (Reich, 2005).

O MicroNIR™ (Figura 5.a) apresenta fonte de luz com duas lâmpadas de tungstênio a vácuo integrais e detector de InGaAs de arranjo de fotodiodo. Enquanto, o MicroPHAZIR™ (Figura 5.b) apresenta fonte de luz de lâmpada de bulbo de tungstênio e detector MOEMS (*Micro-Optoelectronic-Mechanical-Systems*) integral com fotodiodo de InGaAs.

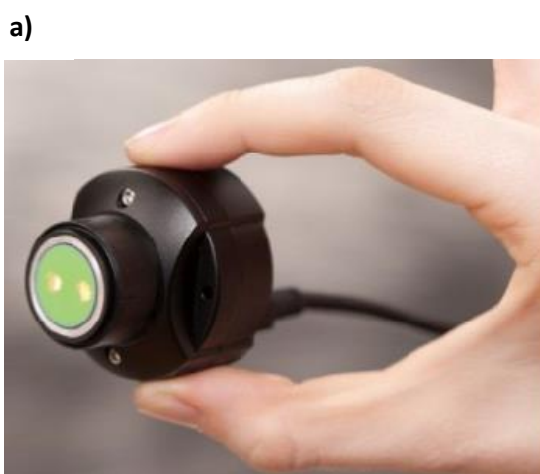


Figura 5. Espectrômetros NIR portáteis: **a)** MicroNIR™, **b)** MicroPHAZIR™.

Espectrômetros de NIR portáteis possuem vantagens normalmente mencionadas para a técnica NIRS, o uso de equipamentos portáteis apresenta vantagens como robustez, custo relativamente acessível, simplicidade das análises e portabilidade. Entretanto, um questionamento válido é a reprodutibilidade dos

resultados obtidos em condições de campo (Santos et al, 2013). A faixa espectral limitada e restrições no software ainda podem ser consideradas desvantagens desses aparelhos portáteis (Barton, 2016). Mais informações sobre os espectrômetros portáteis podem ser encontradas em Tibola et al (2018).

2.5 Análise multivariada

Com o surgimento de instrumentos capazes de medir em uma única amostra diversas variáveis, gerando uma grande quantidade de dados químicos, necessitou-se do desenvolvimento e aplicação de métodos de análise multivariada. Neste estudo, foram desenvolvidos modelos de discriminação e classificação baseado na modelagem de classe.

Na modelagem a determinação de classes para determinação de amostras é feita através da delimitação de um espaço para cada classe, tendo em conta a análise de componentes principais e um limite de confiança pré-definido (Figura 6.a). Para modelos de discriminação, a identificação das amostras é feita definindo-se um limiar de separação que determina se uma amostra pertence a uma classe ou qualquer uma das outras inseridas na fase treinamento, sem delimitar uma região específica (Figura 6.b).

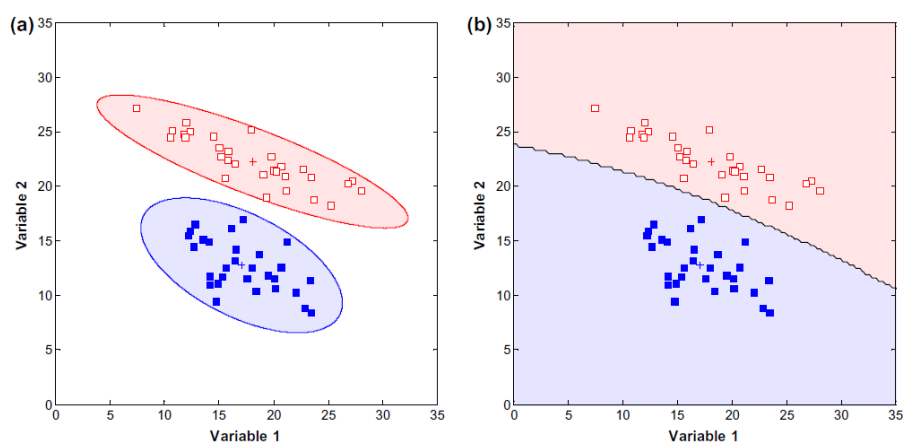


Figura 6. a) Exemplo de modelos baseados em modelagem de classe, onde duas classes são delimitadas por regiões distintas bem definidas. **b)** Exemplo de modelo de discriminação, onde um limite separa duas classes sem distinção de regiões, (Radionova et. al 2016).

2.5.1 Métodos de seleção de amostras

Um problema a ser considerado na seleção de um conjunto de amostras para desenvolvimento de modelos de classificação ou discriminação é que muitas vezes

as amostras selecionadas não representam com precisão a variabilidade dos indivíduos a serem avaliados (Sousa et al, 2011).

Quando se realiza a separação em conjuntos de treinamento e validação do conjunto de dados original (do inglês, *dataset*), torna-se necessária uma seleção inteligente de amostras em cada conjunto, a fim de que os dois conjuntos sejam representativos do universo amostral do problema abordado. Diferentes critérios foram propostos na literatura com a finalidade de operar um modo perspicaz para a separação das amostras disponíveis entre os conjuntos, e todos compartilham a mesma ideia, isto é, distribuir as amostras ao longo dos conjuntos da maneira mais uniforme e representativa possível (Westad e Marini, 2015). Foram abordados dois tipos de algoritmo de seleção para as análises: os algoritmos **Kennard-Stone** (KS) e **Duplex**.

Esses algoritmos têm por objetivo a seleção de um subconjunto de amostras **X** a partir de um conjunto mais amplo de amostras **Z**, de modo que as amostras selecionadas para formar o subconjunto **X** possam representar o máximo de variabilidade de **Z** (Kennard e Stone, 1969).

2.5.1.1 Algoritmo Kennard-Stone

Historicamente, o primeiro algoritmo funcional a separar as amostras disponíveis em um subconjunto representativo em um espaço mais uniforme possível foi proposto por Kennard and Stone (Kennard e Stone 1969). Em resumo, dado um conjunto de amostras candidatas, o algoritmo objetiva escolher as amostras mais diferentes entre si, dentro do conjunto, para serem incluídas no conjunto de treinamento, a partir da distância euclidiana entre as amostras e de acordo com o critério de *maximin*.

Para assegurar uma distribuição uniforme do subconjunto de amostras representadas pelo espaço de dados baseado na resposta instrumental, o algoritmo Kennard-Stone segue um procedimento orientado, no qual novas seleções de amostras são realizadas em regiões do espaço distante das amostras já selecionadas. Emprega-se a distância euclidiana entre os vetores **x** para cada par (p,q) de amostras, calculadas de acordo com a Equação (1):

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

Os dados espectrais $x_p(j)$ e $x_q(j)$ são as respostas instrumentais nos j-ésimos comprimentos de onda para as amostras *p* e *q*, respectivamente, e *J* representa o

número de comprimentos de ondas no espectro. A seleção se inicia escolhendo o par (p_1, p_2) de amostras para os quais a distância $D_x(p_1, p_2)$ seja maior. Em cada interação subsequente, o algoritmo seleciona a amostra que apresentar a maior distância em relação a alguma amostra já selecionada. Ou seja, sucessivamente, para cada amostra candidata restante, a distância mínima para toda amostra já selecionada é computada. Assim, aquela mostrando o valor máximo da mínima distância (critério de *maxmin*) se torna uma amostra selecionada para ser incluída no conjunto de treinamento. Todo o procedimento é repetido até se obter o número desejado de amostras de treinamento (Westad e Marini, 2015; Sousa et al, 2011). A Figura 7 representa uma ilustração da seleção de amostras com base na distância euclidiana D_x e no critério de *maxmin*. As amostras 1, 2 e 4 foram selecionadas nessa ordem por serem consideradas as representativas (Sousa et al, 2011).

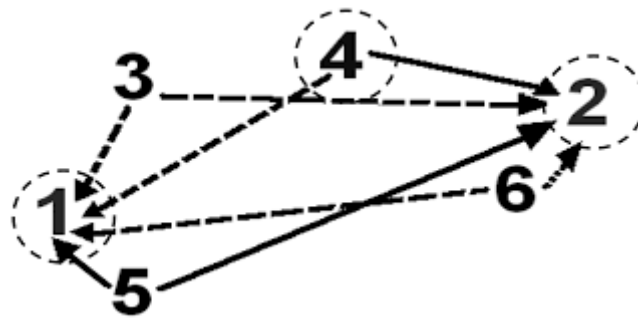


Figura 7. Ilustração de seleção de amostra utilizando o algoritmo Kennard-Stone a partir das distâncias euclidianas e critério de *maxmin*.

2.5.1.2 Algoritmo Duplex

Como a abordagem do algoritmo Kennard-Stone tenta concentrar o máximo da diversidade do conjunto de dados original no conjunto de treinamento de amostras, dependendo da estrutura dos dados, pode-se chegar a resultados superestimados, isto é, em um grupo de amostras com elevada heterogeneidade a divisão entre os subconjuntos pode não ser a mais representativa. Baseando-se nessas considerações, uma modificação do algoritmo almejou manter uma variedade comparável entre ambos os conjuntos (treinamento e validação), proposto pelo próprio Kennard, chamado *Duplex* (Snee, 1977).

O algoritmo é similar ao Kennard-Stone, computando todas as distâncias euclidianas a partes das respostas instrumentos entre as amostras e selecionando as duas amostras mais distantes para serem adicionadas no conjunto de treinamento. Contudo, outras duas amostras com a segunda maior distância são adicionadas ao conjunto de validação. Assim, sucessivamente, alternando entre os

conjuntos de treinamento e validação, logo, as amostras mais diversas são por sua vez adicionadas ou no treinamento ou na validação, de acordo com o critério *maximin*, já mencionado, até o raio de divisão desejado e o número de amostras requisitado para cada conjunto ser obtido (Westad e Marini, 2015).

2.5.2 Partial Least Square Determinant Analysis (PLS-DA)

O modelo de mínimos quadrados parciais (PLS) é um modelo de regressão multivariada amplamente empregado em análises quantitativas em diversas áreas. Em análises qualitativas, PLS é usado para fins de discriminação como PLS-DA (*Partial Least Square Determinant Analysis*), sendo um tipo de modelo de discriminação. Nesse caso, o vetor que introduz uma informação da propriedade de interesse é substituído por um vetor contendo a identificação das classes pré-definidas (Braga et al, 2014; Oliveri et al, 2014).

Nos modelos PLS-DA a matriz de dados (X) pode ser correlacionada com um vetor y , onde cada classe é discriminada em relação às outras em modelos distintos, o qual é conhecido como PLS1-DA. Outra variação do modelo é quando os vetores que discriminam cada classe em relação às outras são reunidos em uma matriz Y e um único modelo de discriminação é construído, sendo essa variação referida como PLS2-DA na literatura (Brereton, 2000). A Figura 8, ilustra a construção do PLS1-DA e PLS2-DA.

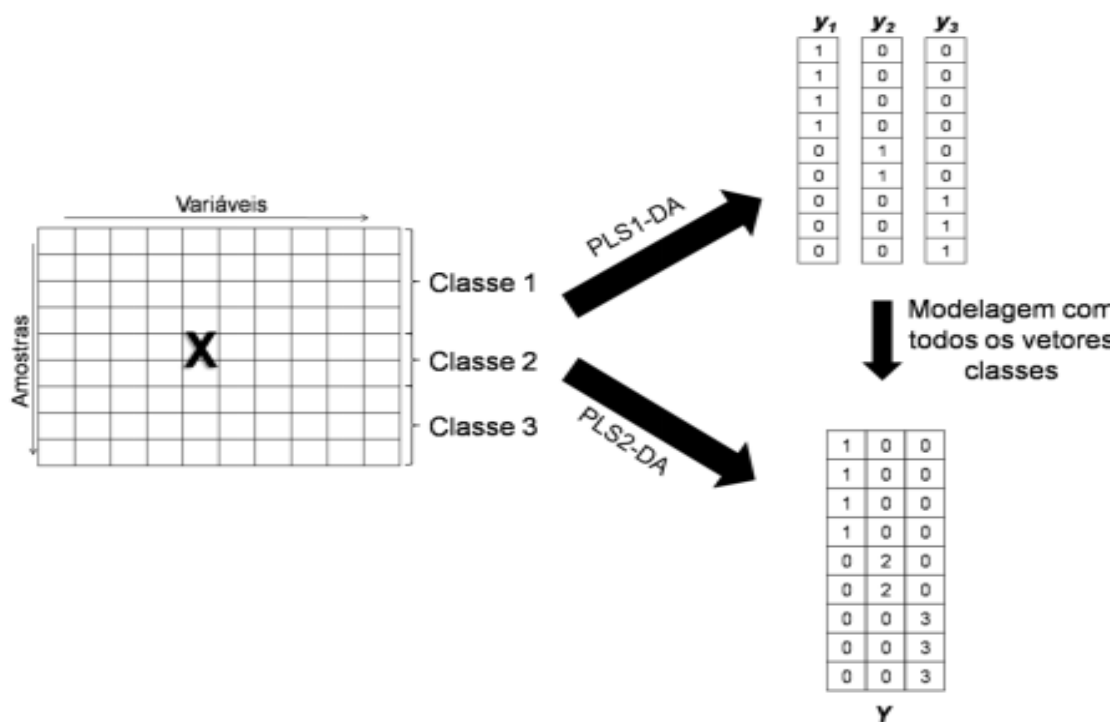


Figura 8. Representação esquemática de PLS1-DA e PLS2-DA (Lopes, 2015).

No PLS1-DA, uma matriz de dados \mathbf{X} é formada pelos dados instrumentais das amostras de treinamento de todas as classes pré-definidas, enquanto que o vetor \mathbf{y} é composto por um vetor binário contendo o valor 1 para amostras que pertencem a classe a ser discriminada e 0 para amostras que pertencem a outras classes (Breteron et al. 2000). A decomposição dos dados em Variáveis Latentes (combinação linear das variáveis originais) pode ser representada por meio das seguintes Equações 2 e 3 (Barker e Rayens, 2003; Brereton e Lloyd, 2014; Soares et al, 2017):

$$\mathbf{X} = \sum_{a=1}^A \mathbf{t}_a \mathbf{p}_a^T + \mathbf{E} \quad (\text{Eq. 2})$$

$$\mathbf{y} = \sum_{a=1}^A \mathbf{t}_a \mathbf{q}_a^T + \mathbf{f} \quad (\text{Eq. 3})$$

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

Para otimização do modelo, o número de variáveis (VL) A e o melhor método de seleção de variáveis da matriz \mathbf{X} é usualmente determinado em PLS-DA através do modelo que apresenta o menor valor de erro de classificação (CVCE, *Cross Validation Classification Error*) empregando validação cruzada (CV - *Cross Validation*) (Brereton, 2003). Contudo, é importante notar que o CVCE não penaliza amostras que apresentam elevados erros na estimativa do valor de classe \mathbf{y} , desde que as amostras sejam corretamente classificadas. No entanto, elevados erros nos valores de classe estimados podem ser uma indicação de que a amostra apresenta características diferentes das demais amostras do conjunto de treinamento ou caracterizam amostras anômalas (*outliers*).

Por outro lado, o valor da raiz quadrada do erro médio quadrático de validação cruzada (RMSECV, *Root Mean Square Error of Cross Validation*) para cada variável latente (Eq. 4), frequentemente usado em PLS, permite a otimização do modelo PLS-DA considerando a minimização dos erros de estimativa dos valores estimados de classe para a classe discriminada ($y=1$) em relação às outras classes ($y=0$). (Martins et a. 2017 e Soares et al. 2017).

$$RMSECV = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (\text{Eq. 4})$$

onde N é o número total de amostras de calibração utilizadas. O y_i é o vetor dos valores de referência das propriedades de interesse e o \hat{y}_i é o vetor das propriedades de interesse estimadas, ambos para a i -ésima amostra (Ferreira et. al 1999; Brereton 2003).

Um modelo é construído sob diferentes números de VL e este número é determinado pelo RMSECV. Normalmente, a partir de um gráfico RMSECV versus número de VL, como o mostrado na Figura 9, seleciona-se o valor mínimo ou o valor a partir do qual se observa uma estabilização do RMSECV, pois indicará o menor número de VL que atingiu o menor erro estimado para a CV (Brereton, 2000, 2003).

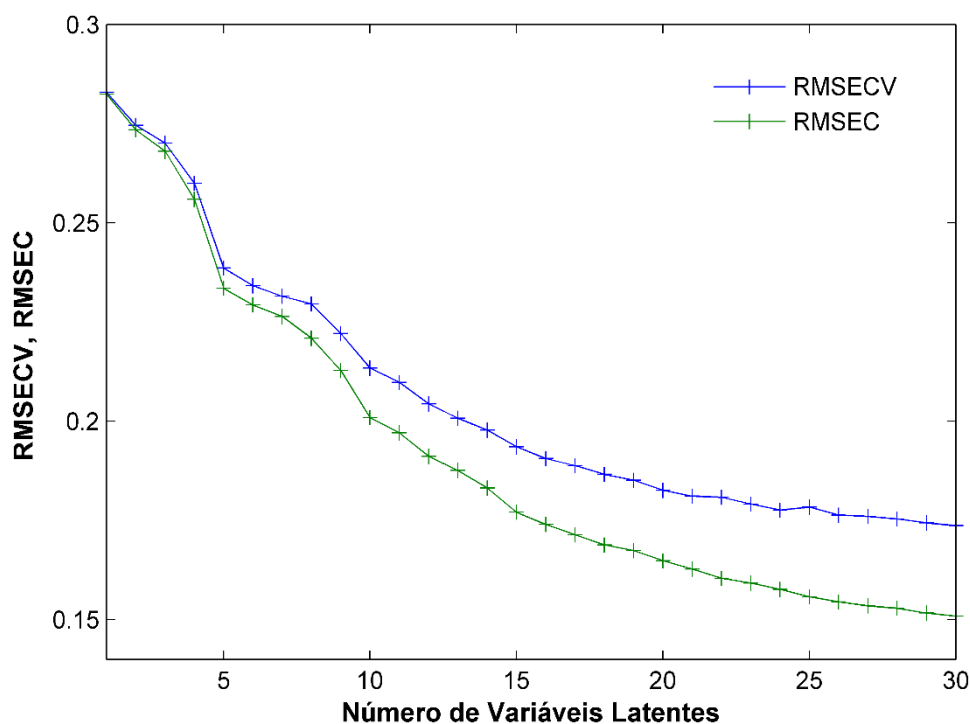


Figura 9. Gráfico de RMSECV por número de variáveis latentes (VL). Método para escolha do número de VL para construções dos modelos PLS-DA.

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

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

- Ao mesmo tempo, amostras que apresentaram resíduos Student que excederam o valor crítico do teste t com correção de viés proposta por da Silva et al (2014) no nível de 99,9% de confiança também foram identificadas como anômalas e excluídas do conjunto de treinamento.

A aplicabilidade de modelos discriminantes para autenticação de alimentos, por exemplo, entrou em discussão em função da probabilidade de ocorrência de erros de identificação quando a amostra, que não pertence a nenhuma das classes predefinidas na etapa de treinamento do modelo, é analisada (Rodionova et al, 2016). De acordo com Amigo et al (2015), um modelo de discriminação bem estruturado irá perfeitamente classificar uma nova amostra apenas se essa amostra é membro de uma das classes pré-definidas. Entretanto, Martins et al (2017) mostrou que, para análises de whisky, o uso dos parâmetros Hotelling T^2 , Q residuais e limites de confiança para os valores de classes estimados permitiu a correta classificação para amostras que pertencem a uma classe, mesmo que essas não tenham sido incluídas no conjunto de treinamento.

Modelos de modelagem de classes representam uma abordagem quimiométrica distinta que tem a capacidade de lidar melhor com amostras pertencentes a uma classe não modelada. Faz parte dessa classe de modelos o SIMCA, que determina uma região de confiança ao redor da classe alvo (Rodionova et al, 2016).

2.5.3 *Soft Independent Modeling of Class Analogy (SIMCA)*

SIMCA é uma ferramenta de modelagem supervisionada baseada em classificação por distância em um espaço definido por componentes principais. Ela está fundamentada na Análise de Componentes Principais (*Principal Component Analysis – PCA*), um método estatístico de análise multivariada que efetua a decomposição da matriz de dados calculando combinações lineares das variáveis originais no intuito de criar novas variáveis ortogonais denominadas componentes principais (Jolliffe, 2002; Oliveri e Downey, 2012; Shou et al, 2014).

O método assume que os valores medidos para amostras semelhantes apresentem distribuição uniforme e modelável. Um modelo baseado em SIMCA deve apresentar uma distribuição probabilística, de acordo com o teorema de Bayes, no qual a decisão de Bayes é aplicada para atribuição de classes (Ferreira, 2015).

A decisão de Bayes funciona da seguinte maneira: uma amostra i possui certa probabilidade de pertencer à classe C (Eq. 5) e essa mesma amostra (i) possui outra probabilidade de pertencer a classe D (Eq. 6).

$$P(C|\hat{y}_i) = \frac{p(\hat{y}_i|C) \times P(C)}{p(\hat{y}_i|C) \times P(C) + p(\hat{y}_i|D) \times P(D)} \quad (\text{Eq. 5})$$

$$P(D|\hat{y}_i) = \frac{p(\hat{y}_i|D) \times P(D)}{p(\hat{y}_i|C) \times P(C) + p(\hat{y}_i|D) \times P(D)} \quad (\text{Eq. 6})$$

Se $P(C|\hat{y}_i) > P(D|\hat{y}_i)$, a amostra i é da classe C. Caso contrário, amostra i é da classe D. De acordo com a Eq. 7.

$$p(\hat{y}_i|C) \times P(C) > p(\hat{y}_i|D) \times P(D) \quad (\text{Eq. 7})$$

Gera-se um modelo PCA para cada classe pré-definida (w) a ser modelada, de acordo com (Lu et al, 2007; Shou et al, 2014):

$$\mathbf{X}_w = \sum_{a=1}^A \mathbf{t}_{a,w} \mathbf{p}_{a,w}^T + \mathbf{E}_w \quad (\text{Eq. 8})$$

onde \mathbf{X}_w é a matriz de dados de treinamento contendo as amostras da classe “ w ”; $\mathbf{T}_{a,w}$ é a matriz de escores contendo os vetores de escores $\mathbf{t}_{a,w}$; $\mathbf{P}_{A,w}$ é a matriz de pesos contendo $\mathbf{p}_{a,w}$ para toda componente principal A e \mathbf{E}_w é a matriz de erros que contém a informação não modelada.

Assim, para o SIMCA, é necessário construir um modelo de componentes principais para cada classe de interesse, e assim usar o modelo para prever qual classe a nova amostra desconhecida pertencerá. Nota-se, entretanto, que as informações de outras classes não são incluídas na construção do modelo estabelecido de uma certa classe, limitando o poder discriminante do método (Yang et al, 2016).

O número de Componentes Principais (PC – *Principal Components*) é escolhido da mesma forma que o número de variáveis latentes (VL) para os modelos PLS-DA. Através do método de validação cruzada (CV) e seguinte análise do gráfico de RMSECV versus PCs. Assim, escolhendo o número de PCs que apresenta estabilização da curva no gráfico de RMSECV e PCs.

A exclusão de *outliers* é feita manualmente através do gráfico de T^2 de Hotelling versus Q residuais, sendo que as amostras que caírem na região do terceiro quadrante (Figura 10) serão consideradas *outliers* sendo excluídas, pois podem gerar classes com áreas discriminantes maiores, diminuindo a acurácia do modelo.

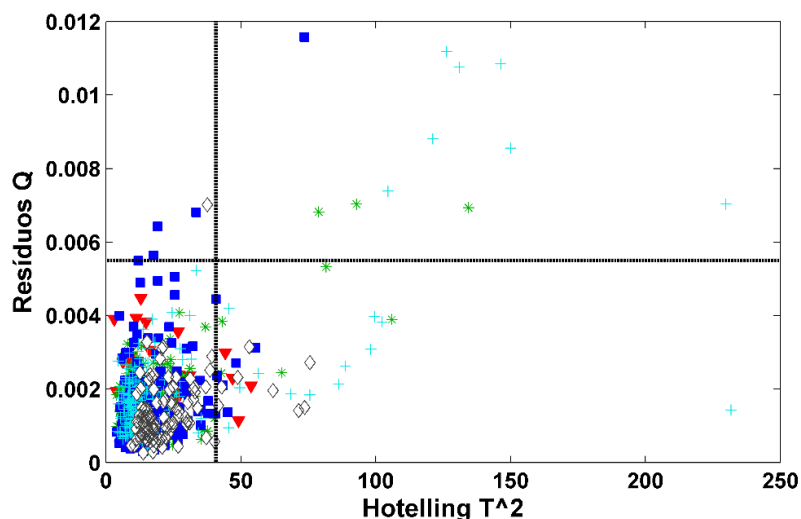


Figura 10. Ilustração do gráfico de T^2 de Hotelling versus Q residuais, mecanismo que possibilita a identificação e exclusão de *outliers*.

2.5.4 Figuras de Mérito

Com a finalidade de avaliar a eficiência dos modelos SIMCA e PLS-DA são estipuladas as figuras de mérito dos métodos quimiométricos (Botelho et al, 2015). O cálculo de todas as figuras considera apenas as amostras de validação não identificadas como *outliers*.

Considerando uma amostra i pertencente a uma classe w , quando a amostra é corretamente classificada como pertencente a classe w , chama-se *verdadeiro positivo (VP)*. Em contrapartida, se uma amostra i pertencente à classe w for classificada como outra classe haverá um *falso negativo (FN)*. Se uma amostra u pertencente a classe h for classificada como classe w , tem-se um *falso positivo (FP)*. Consequentemente, quando uma amostra u pertencente a classe h é analisada no modelo da classe w e é classificada como pertencente a classe h tem-se um *verdadeiro negativo (VN)* (Botelho et al, 2015).

Assim, algumas figuras de mérito podem ser calculadas com base nas relações citadas acima, como a taxa de falso positivo (TFP) - (Eq. 9) e a taxa de falso negativo (TFN) - (Eq. 10), que representam a porcentagem de amostras identificadas como erros de falso positivo e negativo, respectivamente (Soares et al, 2017).

$$TFP = \frac{FP}{FP+VN} \times 100 \quad (\text{Eq. 9})$$

$$\text{TFN} = \frac{\text{FN}}{\text{FN} + \text{VP}} \times 100 \quad (\text{Eq. 10})$$

Outra figura de mérito, chamada taxa de eficiência (TEF) – (Eq. 11), expressa a eficiência global dos modelos, sendo obtida da seguinte forma:

$$\text{TEF} = 100 - (\text{TFN} + \text{TFP}) \quad (\text{Eq. 11})$$

As figuras de mérito apresentadas servirão para avaliar a performance os modelos quimiométricos e fornecer uma perspectiva de qual método é melhor para cada tipo de análise.

3. MATERIAIS E MÉTODOS

3.1. Obtenção e preparação de amostras

3.1.1. Amostras de *S. macrophylla* dos países da América Latina

No total, 326 amostras de madeira mogno (*Swietenia macrophylla* King) de cinco países foram utilizadas para as medidas espectrais: 102 amostras do Brasil, 30 da Bolívia, 80 da Guatemala, 78 do México e 36 do Peru. Cada amostra de madeira era originária de uma tábua de madeira individual. A maioria das amostras foram obtidas a partir de madeira já serrada. Assim, os detalhes da localidade das árvores não estão disponíveis. Boa parte das amostras (principalmente México e Peru) foram obtidas sem expedições, anteriores ao início do projeto Programa ITTO-CITES.

As amostras do Brasil eram originárias de diversas regiões da Floresta Amazônica brasileira, algumas amostras chegaram ao Laboratório de Produtos Florestais (LPF), Brasília (DF) vindas de Rondônia e outras vindas de Curitiba (PR), obtidas a partir de apreensões no Porto de Paranaguá. As amostras bolivianas foram obtidas na serraria Indústria Forestal Colser Ltda, localizada na cidade de Santa Cruz de La Sierra. As amostras do Peru foram obtidas a partir de diferentes locais e expedições a campo, uma parte das amostras do Peru vieram de Porto Maldonado, sendo que a serraria Orotongo providenciou amostras de cerne coletadas de árvores ao longo de dois anos. Similarmente, as amostras mexicanas foram originárias de quatro serrarias localizadas na cidade de Chemutal, próximo a floresta tropical da Península do Yucatán, sul do México. Outras 78 amostras do México foram coletadas sem se ter conhecimento do número das árvores que elas representam. Todas as amostras foram identificadas como pertencendo ao cerne de *S. macrophylla* por anatomistas de madeira do LPF/SFB. As amostras do Brasil, México e algumas da Bolívia estão alocadas no Laboratório de Produtos Florestais. Já as amostras de Bolívia e Guatemala foram analisadas em expedições de campo nos seus respectivos países.

A manutenção das condições da superfície das amostras é muito importante para a reprodutibilidade dos espectros de NIR (Shout et al, 2014; Hein et al, 2010). Por meio de uma lixa de papel (grã 80) se procedeu à homogeneização da superfície das amostras. Após serem lixadas, removeu-se o pó da área lixada com um pincel. Amostras do Brasil, Peru e México alocadas no LPF foram acondicionadas a 20 - 25°C e umidade relativa de 10 -13% e as amostras da Guatemala, Bolívia e algumas do Brasil foram analisadas em condições de campo, em temperaturas que variavam de 15-35°C e umidade relativa de 10-25%.

3.1.2 Amostras de *S. macrophylla* e *S. humilis* coletadas na Guatemala

Foram obtidas 125 amostras das duas espécies de mogno sendo 45 amostras de *Swietenia humilis* e 80 amostras de *Swietenia macrophylla*. Essas amostras pertencem a duas regiões distintas da Guatemala: Escuintla e Péten.

As amostras de *S. humilis* de árvores nativas ou cultivadas da região de Escuintla e as amostras de *S. macrophylla* foram originárias de mata nativa da região de Péten. A Tabela 1 apresenta informações referente à localidade específica e quantidade de amostras por local.

Tabela 1. Localidade, região e quantidade de amostras das espécies *S. macrophylla* e *S. humilis* coletadas na Guatemala.

Espécie	Serraria	Região	Número de Amostras
<i>Swietenia macrophylla</i>	Asociación Floretal Integral de San Andrés (AFISAP)	Petén	42
	Custodios de la Selva (CUSTOSEL)	Petén	38
<i>Swietenia humilis</i>	Finca Bondurria	Escuintla	21
	Finca San Julian Suchitepequez	Escuintla	20
	Forestry Industry el Tejar	Escuintla	4

3.2. Aquisição dos espectros de Infravermelho Próximo (IR):

Os espectros de NIR foram coletados usando dois espectrômetros portáteis: MicroPHAZIR™ RX Analyser (Thermo) e MicroNIR™ 1700 Spectrometer (JDSU), possuindo 8.7 nm e 6.2 nm de resolução e faixa espectral de 1.595 a 2.396 nm e 950 a 1.650 nm, respectivamente. O espectro de fundo ou branco de ambos instrumentos foi obtido a partir de padrões de politetrafluoretileno fornecidos pelos fabricantes dos equipamentos. Para as medidas espectrais usando o MicroNIR™ foram utilizados os parâmetros de 2.000 µs para o tempo de integração e cada espectro resultou da média de 100 varreduras configurados no software do equipamento. Os espectros das amostras de *S. macrophylla* dos cinco países foram tomados usando ambos espectrômetros portáteis, enquanto os espectros das espécies de mogno (*S. macrophylla* e *S. humilis*) foram medidos apenas com o instrumento MicroNIR™.

Espectros de reflectância difusa foram coletados nas faces tangenciais e radiais das amostras, ou seja, os espectros foram tomados nessas faces de forma aleatória. Três espectros foram obtidos por amostra da espécie *S. macrophylla* dos cinco países utilizando cada instrumento portátil NIR, totalizando 923 espectros com o MicroPHAZIR™ e 930 espectros com o MicroNIR™, respectivamente. Para as amostras das duas espécies de mogno foram tomados um total de 375 espectros utilizando apenas o equipamento MicroNIR™. A Figura 11 mostra um esquema geral da obtenção dos espectros NIR tanto para as amostras dos cinco países, quanto para as espécies de mogno.



Figura 11. Esquema geral do processo de aquisição dos espectros NIR para amostras da madeira mogno.

3.3. Processamento dos espectros e análises de dados

Todas as análises dos dados e processamento dos espectros foram conduzidas no software Matlab7.14, versão R2012a com o pacote PLS_Toolbox, versão 7.03.

Todos os espectros de NIR coletados possuem informações referentes à composição química e a estrutura física de cada amostra. Entretanto, possuem informações referentes a variações de linha de base e outras flutuações

instrumentais que podem afetar a acurácia dos modelos, sendo necessário o pré-processamento dos dados (Scwanninger et al, 2011., Braga et al, 2014).

Nesse estudo, tanto para as análises das amostras dos países como para as duas espécies, foi utilizada a primeira derivada, baseada em alisamento Savitzky-Golay (11 pontos e polinômio de 2ª ordem), normalização vetorial (Variável Normal Padrão, do inglês Standard Normal Variate) e dados centrados na média.

Para espectros obtidos com o MicroNIR™ das amostras de *S.macrophylla* dos cinco, foi visualmente selecionada uma faixa espectral entre 1.043 a 1.650 nm para minimizar os ruídos espectrais e mudanças na linha de base observada na região entre 950 a 1.042 nm. Já, para o MicroPHAZIR™, toda a região espectral foi selecionada para a análise, pois o pré-processamento foi suficiente para anular a variação de linha de base.

Devido à heterogeneidade natural da madeira e para obter-se uma maior quantidade de dados representativos, a média dos espectros não foi utilizada para ambas as análises (Soares et al, 2017).

3.4. Seleção de amostras para treinamento e validação

Quando se realiza a separação do conjunto total de dados dos conjuntos de treinamento e validação, torna-se necessária uma seleção inteligente de amostras em cada conjunto, a fim de que os dois conjuntos sejam representativos do universo e amostras do problema abordado. Como já citado, foram utilizados dois algoritmos de seleção de amostras: Kennard-Stone e Duplex.

Para a análise com as amostras dos cinco países foram utilizados os dois algoritmos de seleção de amostras nos dois tipos de modelo quimiométrico estudados (PLS-DA e SIMCA). Enquanto que, na análise que envolve as duas espécies de mogno, os modelos foram desenvolvidos com algoritmo Kennard-Stone.

3.4.1. Algoritmo Kennard-Stone

Para o desenvolvimento dos modelos foram usados dois terços das amostras do conjunto total de dados para treinamento e um terço para validação para os dados de ambos equipamentos. A Tabela 2 apresenta o número de espectros pertencentes a cada conjunto de treinamento e validação para os cinco países com ambos espectrômetros NIR portáteis. Para assegurar a correta validação, quando uma amostra foi designada para um desses conjuntos, todas as replicatas da amostra foram incluídas nesse conjunto específico.

Tabela 2. Número de espectros usados nos conjuntos de treinamento e validação dos cinco países da América Latina para os equipamentos portáteis MicroPHAZIR™ e MicroNIR™, utilizando-se o algoritmo Kennard-Stone.

País	Número de espectros			
	Conjunto de treinamento		Conjunto de validação	
	<i>MicroPHAZIR™</i>	<i>MicroNIR™</i>	<i>MicroPHAZIR™</i>	<i>MicroNIR™</i>
Bolívia	60	72	30	30
Brasil	205	203	101	103
Guatemala	156	159	84	81
México	161	156	78	78
Peru	74	76	34	32

3.4.2. Algoritmo Duplex

Assim como para construção dos modelos com algoritmo Kennard-Stone, foram utilizados dois terços das amostras para construção do conjunto de treinamento e um terço para o conjunto de validação. A Tabela 3 exibe os números de espectros pertencentes a cada conjunto por país utilizando-se o algoritmo Duplex.

Tabela 3. Número de espectros dos conjuntos de treinamento e validação dos cinco países da América Latina para os equipamentos portáteis MicroPHAZIR™ e MicroNIR™, utilizando-se o algoritmo Duplex.

País	Número de espectros			
	Conjunto de treinamento		Conjunto de validação	
	<i>MicroPHAZIR™</i>	<i>MicroNIR™</i>	<i>MicroPHAZIR™</i>	<i>MicroNIR™</i>
Bolívia	60	69	30	31
Brasil	207	202	102	104
Guatemala	158	152	82	88
México	159	168	78	69
Peru	76	77	32	31

3.5. Modelos Quimiométricos

3.5.1. *Partial Least Square Determinant Analysis (PLS-DA)*

Foram desenvolvidos 5 modelos PLS-DA, um para cada país estudado com os espectros das suas respectivas amostras. Nesse trabalho optou-se por utilizar modelos PLS1-DA. Portanto, na fase de treinamento foram utilizadas amostras cujas classes foram conhecidas e, em cada modelo, o vetor \mathbf{y} foi composto por valores 1 e 0, onde 1 foi atribuído às amostras que pertencem a classe que se pretende discriminar e o valor 0 atribuído às amostras pertencentes as outras classes.

A identificação de outliers no conjunto de treinamento foi estabelecida com base nos valores de T^2 e Q residuais e nos limites de confiança para a classe estimada (Soares et al. 2017, Martins et al. 2017). Todas as amostras que apresentam valores para os parâmetros T^2 e Q espectro-residuais maiores que 99.9% foram identificadas como *outliers* e removidas do conjunto de treinamento (Borin et al, 2004; Silva e Talhavini et al, 2014). Ademais, amostras que revelam valores de classe estimados, após a correção de viés, fora dos limites para classe discriminada ou para outras classes, considerando o limite de confiança de 99.9%, foram identificadas também como *outliers* e excluídas do conjunto de treinamento.

Os passos para o desenvolvimento do modelo e subsequente análise de validação foram ilustrados na Figura 12. Inicialmente, modelos preliminares de PLS-DA foram construídos para cada classe que foi classificada e identificando-se os *outliers*, seguida da sua exclusão do conjunto de treinamento. Em seguida, os modelos foram recalculados com as amostras restantes.

Após a otimização do modelo, um limite de discriminação (LD) (do inglês, *Discrimination Threshold*) foi determinado com base na dispersão dos valores de classes estimados, obtidos para o conjunto de treinamento, com finalidade de minimizar a ocorrência de erros de falsos positivos (FP) e falsos negativos (FN), de acordo com a teorema de Bayes (Barker et al. 2003; Botelho et al. 2015; Ferreira et al. 2015). Uma descrição detalhada da determinação do LD pode ser encontrada em Soares et al (2017).

É importante notar que a identificação de uma amostra a uma classe na fase de validação ou aplicação em rotina é determinada apenas após a análise todos os modelos PLS-DA. Portanto, uma amostra hipotética i é classificada como classe 1 apenas se a amostra não for identificada como um outlier e seus valores de classe estimados forem maiores que o LD quando a amostra é inserida do modelo PLS-DA da classe 1 e simultaneamente possuir valores de classe estimados menores que a LD dos modelos para outras classes pré-definidas.

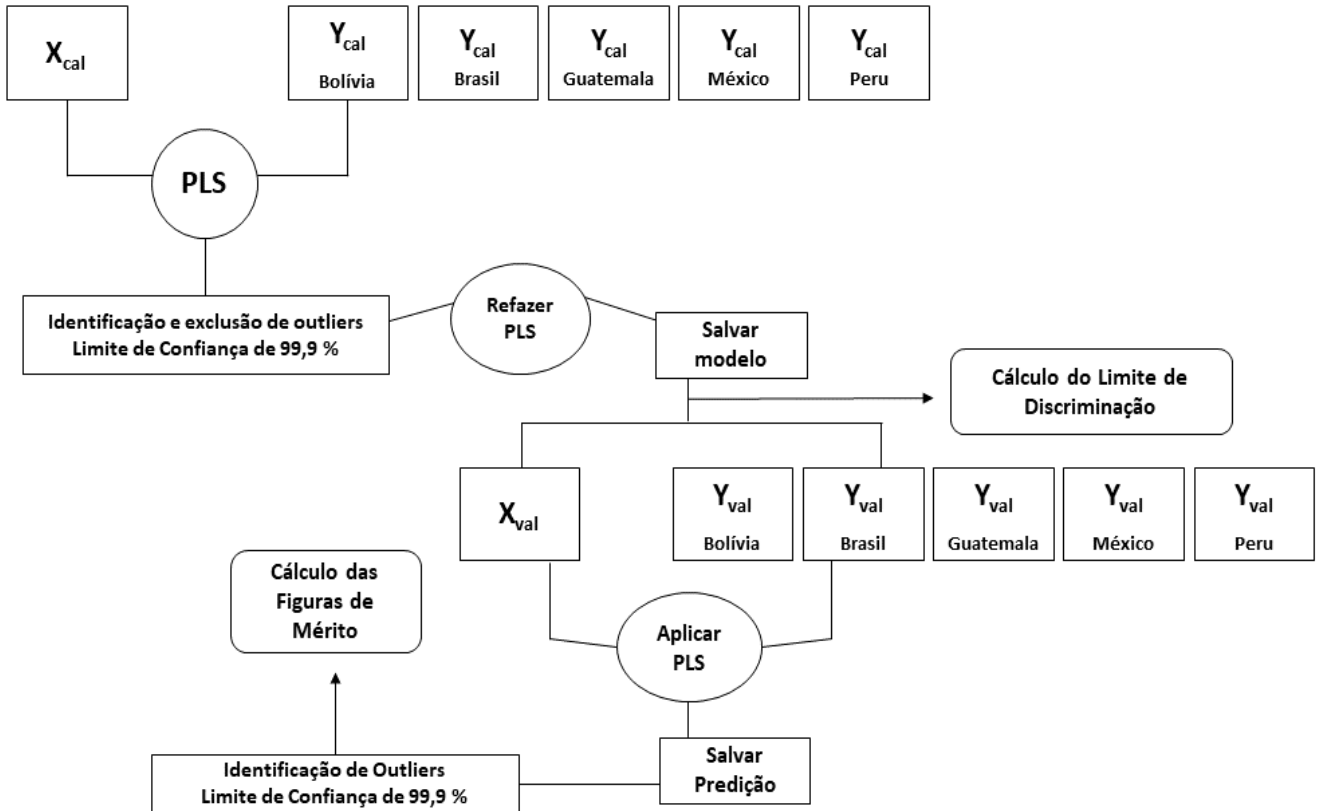


Figura 12. Passo a passo para o desenvolvimento e validação do modelo PLS1-DA. X_{cal} : matriz de espectros do conjunto de calibração/treinamento; Y_{cal} : vetor de classes da calibração para cada país; X_{val} : matriz do conjunto de validação; Y_{val} : vetor de classe de validação.

3.5.2. Soft Independent Modeling of Class Analogy (SIMCA)

Para classificação usando SIMCA, inicialmente o conjunto de treinamento foi utilizado para construir um modelo PCA específico para cada classe pré-definida. O número de PCs (do inglês, *Principal Components*) foi determinado por validação cruzada a fim de minimizar o número de erros de classificação no conjunto de treinamento. Após a modelagem PCA da classe pré-definida, a classe de uma amostra desconhecida foi determinada como sendo aquela que fornece a menor distância (d) em relação ao modelo PCA das classes pré-definidas, onde $d_{i,w}$ referente a uma amostra i e uma classe w é determinada de acordo com a Equação 12.

$$d_{i,w} = \sqrt{\left(\frac{T_{i,w}^2}{T_{0,999,w}^2}\right)^2 + \left(\frac{Q_{i,w}}{Q_{0,999,w}}\right)^2} \quad (\text{Eq. 12})$$

sendo $T^2_{i,w}$ é o fator estatístico de Hotelling, $Q_{i,w}$ é a soma dos quadrados dos resíduos do PCA, $T^2_{0,999,w}$ e $Q_{0,999,w}$ são os correspondentes limiares de nível de confiança de 99,9% de T^2 e Q para a classe w , respectivamente. O fator estatístico $T^2_{i,w}$ considera a variância modelada pelo modelo PCA para a classe w e é determinada pelos escores conforme a Equação 13.

$$T^2_{i,w} = \frac{\mathbf{t}_{i,w}^T \mathbf{t}_{i,w}}{(\mathbf{T}_w^T \mathbf{T}_w) / (I-1)} \quad (\text{Eq. 13})$$

Em contrapartida, os valores de $Q_{i,w}$ explicam a variância não modelada dos dados do conjunto de treinamento na classe w nos modelos PCA, sendo determinados conforma a Equação 14.

$$Q_{i,w} = \mathbf{e}_{i,w} \mathbf{e}_{i,w}^T \quad (\text{Eq. 14})$$

No presente estudo, os valores críticos para ambos T^2 e Q foram determinados com 99.9% de nível de confiança e de acordo com o método proposto por Jackson e Mudholkar (1979).

4. RESULTADOS E DISCUSSÃO

4.1. Determinação da origem de amostras de madeira mogno (*S. macrophylla*) dos cinco países

4.1.1. Espectros NIR das amostras de *Swietenia macrophylla* dos cinco países

As Figuras 13(a) e 13(b) mostram a média dos espectros de NIR das amostras de mogno, por país estudado, obtidos com os espectrômetros MicroNIR™ e MicroPHAZIR™, respectivamente. É perceptível que os intervalos espectrais varridos por cada equipamento são complementares. O MicroNIR™ abrange o intervalo espectral de 950 a 1.650 nm, cobrindo as regiões dos 2º e 3º sobretons, porém, nos espectros da Fig.13(a), um intervalo espectral foi cortado, uma vez que suas extremidades apresentaram ruídos que poderiam prejudicar a análise posterior (Bergo et al, 2016). Já MicroPHAZIR™ cobre a região espectral de 1.596 a 2.396 nm, correspondendo ao 1º sobreton e às combinações de banda (Jiang et al, 2006).

Em ambas as Figuras (13.a e 13.b), somente uma avaliação visual não permite a identificação dos países, considerando as variações espectrais muito pequenas e a variação da linha base.

Figuras 14(a) e 14(b) apresentam a média dos espectros das amostras de mogno após o pré-processamento com primeira derivada para os dois equipamentos. Todavia, mesmo com a correção das variações de linha de base e acentuação das diferenças espectrais, promovidas pela 1ª derivada, ainda não é possível identificar visualmente regiões distintas que permitem determinar a procedência de uma amostra de mogno de outra, sendo necessário procedimento de análise multivariada.

A atribuição de bandas espectrais aos componentes químicos da madeira, como celulose, hemicelulose, lignina e extrativos, é uma tarefa difícil devido às estruturas moleculares complexas dos componentes e ao ambiente molecular variado que é a madeira em si (Schwanninger et al, 2011). A partir das atribuições propostas por Schwanninger et al (2011); Esteves e Pereira (2008); Bachle et al (2010), Soares et al (2017) tornou-se possível encontrar as principais bandas dos espectros NIR em ambos instrumentos portáteis.

Para o intervalo espectral do MicroNIR™, entre 1.000 e 1.466 nm, manifestam-se os primeiros sobretons e segundos sobretons para estiramentos dos grupos O-H e N-H, as combinações de banda dos primeiros sobretons do grupo C-H e segundos e terceiros sobretons do C-H, estiramentos dos grupos C_{Aromático}-H e N-H. Já, entre 1.486 e 1.667 nm se evidencia o primeiro sobreton dos estiramentos

do O-H de carboidratos. (Esteves e Pereira, 2008; Bachle et al, 2010; Schwanninger et al, 2011).

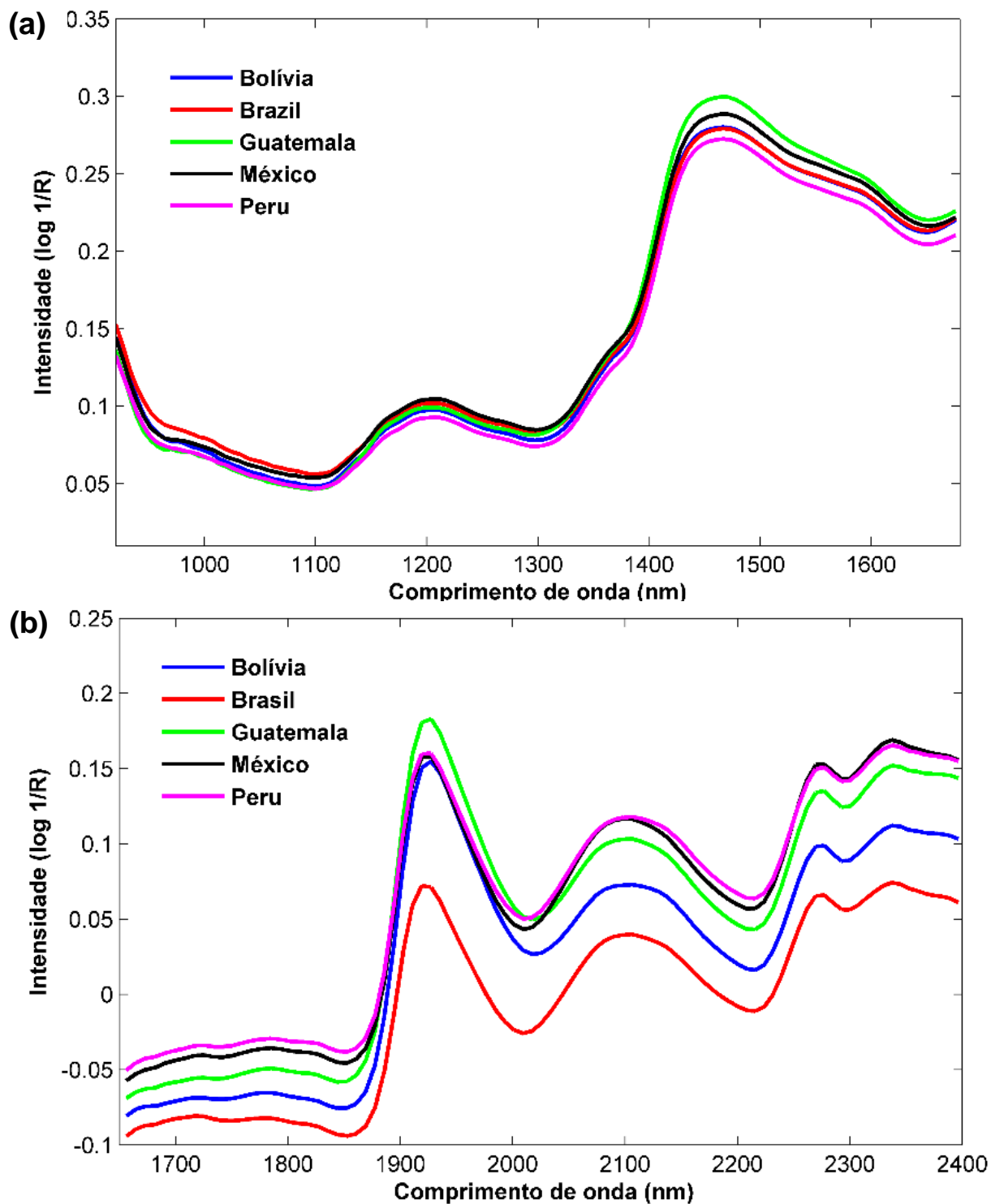


Figura 13. Média dos espectros NIR das amostras da Bolívia, Brasil, Guatemala, México e Peru sem qualquer pré-processamento, obtidos com MicroNIR™ (a) e MicroPHAZIR™ (b).

Na Figura 14(a) podem-se observar as principais bandas obtidas com o MicroNIR™ (1.000 e 1.466 nm) são:

(1) 1.100 a 1.200 nm, associada ao estiramento assimétrico do C-H e HC=CH da lignina, segundo sobretom do estiramento do C-H dos grupos de CH₃ do conjunto de acetil éster nas hemiceluloses e segundos sobretoms dos estiramentos de C-H das moléculas de celulose e lignina;

(2) 1.350 a 1.380 nm relacionada aos primeiros sobretoms do estiramento do C-H e deformação angular em celulose e hemicelulose;

(3) 1.390 a 1.450 nm atribui-se aos primeiros sobretoms do estiramento do O-H das hidroxilas fenólicas (extrativos), lignina, água e celulose (Soares et al. 2017; Schwanninger et al. 2011).

A região espectral do MicroPHAZIR™ (1.596 a 2.396 nm) relaciona-se aos 1º sobretoms de estiramentos de C-H alifáticos e aromáticos e combinações de banda do O-H (Schwanninger et al, 2011). Assim, na Figura 14(b) as principais atribuições de bandas são:

(1) 1.850 a 1.950 nm associada ao estiramento do grupo O-H e ao segundo sobretom do estiramento de C-O da celulose, segundos sobretoms do estiramento do C=O da hemicelulose, estiramento assimétrico do O-H e deformação angular da molécula de H₂O (água);

(2) 2.050 a 2.110 nm relacionada ao estiramento mais deformação angular do O-H e deformação do C-H da celulose, estiramentos C_{Ar}-H e C=C de lignina e extrativos, estiramentos C-H e C=C dos grupos acetil da hemicelulose;

(3) 2.240 a 2.280 nm se atribui os estiramentos dos grupos O-H e C-O da lignina e celulose, estiramentos e deformação de C-H₂ da celulose e hemicelulose, estiramentos do O-H e C-C e deformação do C-H da celulose, estiramento e deformação do C-H da metila da hemicelulose;

(4) 2.300 a 2.340 nm relacionado a estiramento e deformação do C-H, segundos sobretoms da deformação do C-H da celulose e hemicelulose (Schwanninger et al, 2011).

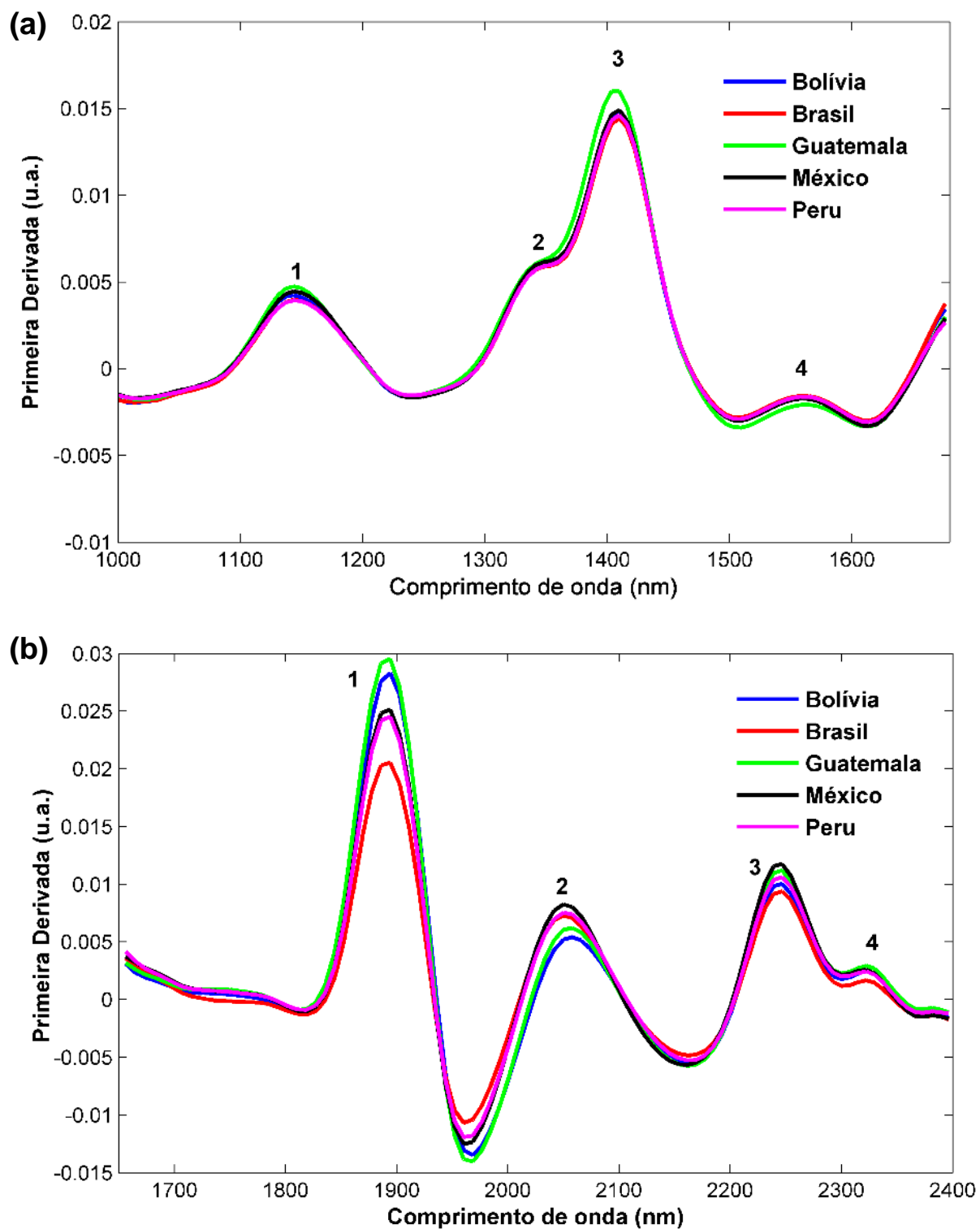


Figure 14. Média dos espectros das amostras da madeira mogno de cada país após o pré-processamento com primeira derivada SG (Savitzky-Golay). Espectros pré-processados do MicroNIR™ (a) e MicroPHAZIR™ (b). As bandas espectroscópicas mais significativas estão numeradas.

4.1.2 Análise quimiométrica

4.1.2.1 PLS-DA

4.1.2.1.1 Algoritmo Kennard-Stone

A Tabela 4 reúne os parâmetros obtidos a partir dos modelos PLS-DA desenvolvidos para discriminar as madeiras de mogno dos cinco países a partir da seleção de amostras realizada segundo o algoritmo Kennard Stone (KS) e dos dados de ambos equipamentos NIR.

Tabela 4. Parâmetros dos modelos PLS-DA e número de *outliers* identificados para os modelos do MicroNIR e MicroPHAZIR, utilizando-se o algoritmo Kennard-Stone.

<i>Países</i>	MicroNIR™				MicroPHAZIR™			
	VL	LD	OUT_ trein	OUT_ val	VL	LD	OUT_ trein	OUT_ val
<i>Bolívia</i>	20	0,384	18	29	18	0,436	18	21
<i>Brasil</i>	18	0,440	13	5	18	0,415	16	21
<i>Guatemala</i>	18	0,454	23	25	18	0,465	24	18
<i>México</i>	20	0,392	18	23	18	0,465	17	22
<i>Peru</i>	20	0,383	17	21	18	0,310	24	24

VL: número de variáveis latentes; LD: limite de discriminação; OUT_trein: número de *outliers* excluídos no conjunto de treinamento; OUT_val: número de *outliers* excluídos no conjunto de validação.

É notável que para a construção dos modelos de cada país, em ambos espectrômetros, foi necessário um número similar de variáveis latentes em função do grau elevado de variabilidade do conjunto de dados fornecidos pela amostragem e em função da alta heterogeneidade das amostras de madeira. Na maioria dos modelos construídos, a quantidade de *outliers* foi ligeiramente maior para os dados de MicroPHAZIR™ que para os dados de MicroNIR™, sendo exceções os modelos para México e Bolívia. A porcentagem de *outliers* mais elevada dos modelos para MicroNIR™ foi obtida pela Guatemala (~3%), enquanto para MicroPHAZIR™, o modelo para Peru obteve a maior taxa de exclusão (~3,8%).

A quantidade de *outliers* identificada para ambos equipamentos se deve provavelmente aos espectros tomados em condições de campo na Bolívia, Guatemala e parcialmente no Brasil que influenciaram os espectros das amostras. Essas medições estavam suscetíveis a vários fatores externos, como variação de temperatura, manuseio inadequado do instrumento portátil, variantes na rugosidade

da superfície da amostra, umidade relativa do local da medida e o teor de umidade da amostra, uma vez que os espectros de NIR são influenciados pela presença de moléculas de água. Não obstante, a taxa de exclusão de *outliers* abaixo de 5% foi considerada aceitável.

Os resultados para o limite de discriminação (LD) para cada modelo também estão listados na Tabela 4. Os valores de LD próximos de 0,5 demonstram que as distribuições das classes estão semelhantes, já que o PLS-DA separa as classes entre 0 e 1, distribuição binária.

Valores abaixo de 0,5 indicam uma grande dispersão dos dados para classe 1 (classe discriminada), isto é, são normalmente associados a uma tendência negativa nos valores de classe estimados para a classe 1. A maioria dos modelos apresentaram LD maiores que 0,4, considerado aceitável (Silva et al. 2014). O valor mais baixo de LD foi observado para discriminação das amostras do Peru com dados de MicroPHAZIR™ (LD = 0,310). Ainda assim, bons resultados foram obtidos observando as taxas de eficiência (TEF) exibidos na Tabela 5.

Tabela 5. Figuras de mérito dos modelos PLS-DA construídos utilizando-se o algoritmo Kennard-Stone e calculadas a partir do conjunto de validação.

Modelo/ País	MicroNIR™			MicroPHAZIR™		
	TFN (%)	TFP (%)	TEF (%)	TFN (%)	TFP (%)	TEF (%)
<i>Bolívia</i>	0,0	0,7	99,3	0,0	0,4	99,6
<i>Brasil</i>	4,2	5,7	90,1	0,5	1,1	98,4
<i>Guatemala</i>	0,0	0,0	100,0	0,0	3,5	96,5
<i>México</i>	1,5	3,6	94,9	6,6	3,8	89,6
<i>Peru</i>	0,0	0,6	99,4	0,0	8,8	91,2

TFN – taxa de falso negativo; TFP – taxa de falso positivo; TEF – taxa de eficiência.

As figuras de mérito da Tabela 5 revelam que as Taxas de Eficiência (TEFs) dos modelos de MicroNIR™ foram todas acima de 90%. Já os modelos da Bolívia e Peru apresentam valores próximos de 100% e o modelo da Guatemala consegue discriminar com 100% de eficiência as amostras guatemaltecas em relação as dos outros países. O menor valor da TEF, dentre os dados do MicroNIR™, foi atribuído ao modelo das amostras brasileiras e isso provavelmente se deve à dispersão dos dados em função da grande variabilidade das condições de amostragem em

diferentes regiões geográficas do país. Um estudo poderá ser feito para determinar como essa dispersão de dados dentro das amostras do próprio país pode afetar os modelos de discriminação.

Observa-se ainda na Tabela 5 que as taxas de falso negativo (FN) e falso positivo (FP) são próximas entre si, revelando que não há tendências em direções específicas em relação aos erros de classificação. As TEFs dos espectros obtidos com o MicroPHAZIR™ variaram entre 89,6 a 99,6%, obtendo também uma excelente eficiência de discriminação. A menor TEF obtida foi para o modelo das amostras mexicanas (89,6%).

Todos os modelos apresentaram TEFs acima de 89,6%, assim a taxa de classificação correta para ambos espectrômetros foi elevada. A diferença mais marcante entre os modelos para um mesmo país com espectros de instrumentos diferentes foi entre as amostras classificadas do Brasil e do Peru, que apresentarem variação de 8% de um espectrômetro para o outro. Um fato a ser observado, mesmo apresentando uma diferença significativa no número de amostras da Guatemala em relação aos outros países, dentre os modelos de espectrômetros diferentes, os resultados diferiram pouco.

Os gráficos de Hotelling T^2 e resíduos Q foram analisados para verificar a presença de *outliers* e se os espectros de validação eram compatíveis com o conjunto de treinamento usado para construção dos modelos. Cabe lembrar que uma amostra é identificada como anômala se tiver ambos os valores de T^2 e Q acima dos limites de confiança. Na Figura 15, tem-se os gráficos de Hotelling T^2 e resíduos de Q para os cinco países a partir dos dados obtidos com MicroNIR™. Ainda nessa Figura, observa-se comportamentos distintos entre as amostras dos cinco países. Apesar de terem sido identificadas amostras anômalas em todos os modelos, isto é, estão localizadas fora dos limites de confiança de 99,9%, a grande maioria das amostras se encontra dentro desses limites. Portanto, o modelo se mostrou válido para mais que 95% das amostras de validação.

Comparando os gráficos de Hotelling T^2 e resíduos Q das Figuras 15 e 16, ou seja, comparando os gráficos obtidos de ambos espectrômetros portáteis, observa-se que os modelos construídos com os dados do MicroPHAZIR™ apresentam uma quantidade de espectros de amostras considerados *outliers* menor que os modelos construídos com o MicroNIR™.

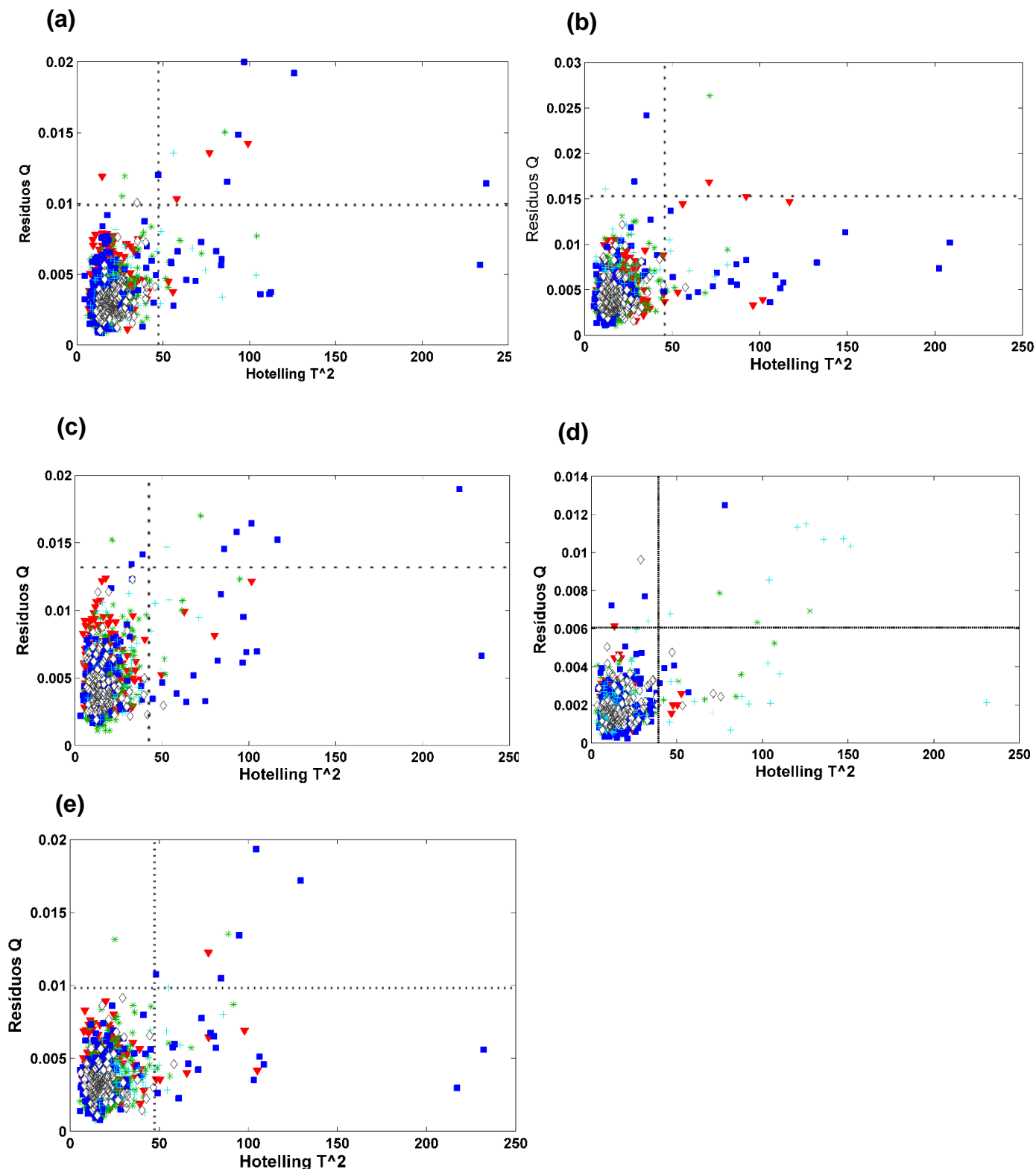


Figura 15. Gráfico dos valores de T^2 de Hotelling e resíduos Q das amostras de (a) Bolívia, (b) Brasil, (c) Guatemala, (d) México e (e) Peru obtidos com PLS-DA e algoritmo KS a partir dos dados do MicroNIR™. Legenda: (\blacktriangledown) Brasil, ($*$) Guatemala, (\blacksquare) México, ($+$) Peru e (\diamond) Bolívia, (--) Limite de confiança (99,9%).

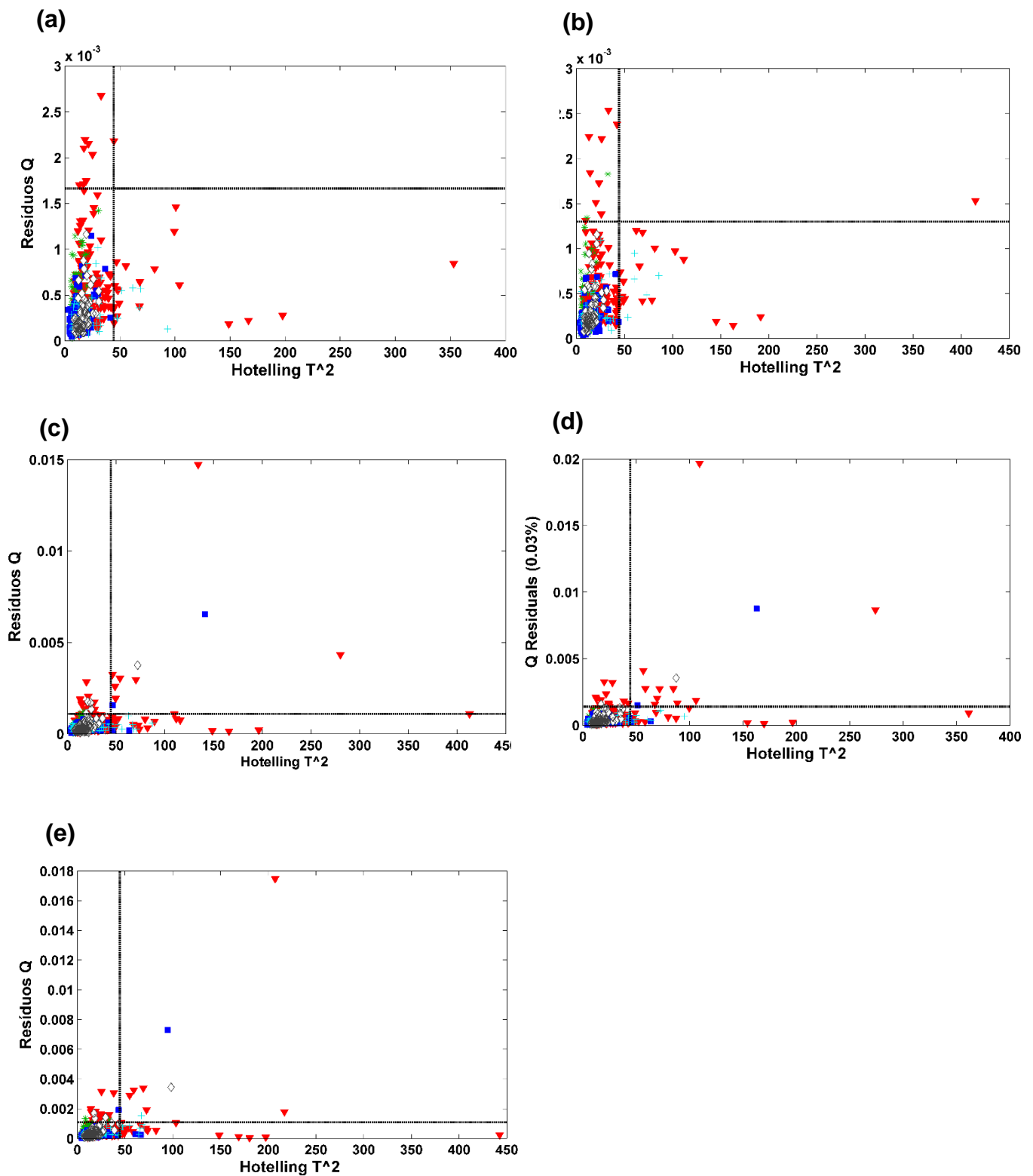


Figura 16. Gráfico dos valores de T^2 de Hotelling e resíduos Q das amostras de (a) Bolívia, (b) Brasil, (c) Guatemala, (d) México e (e) Peru obtidos com PLS-DA e algoritmo KS a partir dos dados do MicroPHAZIR™. Legenda: (▼) Brasil, (*) Guatemala, (■) México, (+) Peru e (◇) Bolívia, (--) Limites de confiança (99,9%).

A Figura 17 e 18 apresentam os gráficos de distribuição de valores de classe estimados para cada modelo PLS-DA construído por meio das amostras da Bolívia: (Fig. 17.a e 18.a), Brasil (Fig. 17.b e 18.b), Guatemala (Fig. 17.c e 18.c), México (Fig. 17.d e 18.d) e Peru (Fig. 17.e e 18.e), respectivamente, com os dados de MicroNIR™ e MicroPHAZIR™, respectivamente, a partir do algoritmo de KS.

Por meio dos gráficos foi possível notar visualmente que os espectros das amostras utilizadas na fase de treinamento tiveram uma dispersão nos valores de classe semelhante à observada nas amostras de validação. Além disso, percebeu-se quais espectros foram falsos positivos ou negativos. Cada país possuiu um gráfico de distribuição de valores de classes estimados que representa o modelo construído para as amostras daquele país em questão. Quanto maior a distância das dispersões entre as classes, melhor a discriminação. Uma análise dos gráficos (Figura 17) dos modelos de dados do MicroNIR™ atingiu as seguintes qualidades de classificação: Guatemala > Peru > Bolívia > México > Brasil.

Analisando os gráficos de distribuição de valores de classes estimados, para os dados de MicroPHAZIR™ (Fig. 18); as amostras da Guatemala possuíram os melhores valores de discriminação obtidos, sendo amostras do Brasil melhores discriminadas que as amostras peruanas. Uma distribuição semelhante foi obtida para Bolívia e México, tanto para o MicroNIR™ como para o MicroPHAZIR™, respectivamente (Fig. 17.a e Fig. 18.d).

As dispersões dos valores de classe estimados de todos os modelos dentre os conjuntos de treinamento e validação não foram significativamente diferentes entre si, indicando a ausência de sobreajuste. Pela análise, foi possível inferir que ambos espectrômetros forneceram bons resultados em termos de identificação da origem das amostras de mogno. Entretanto, os dados do MicroNIR™, em combinação com PLS-DA, apresentaram resultados ligeiramente melhores.

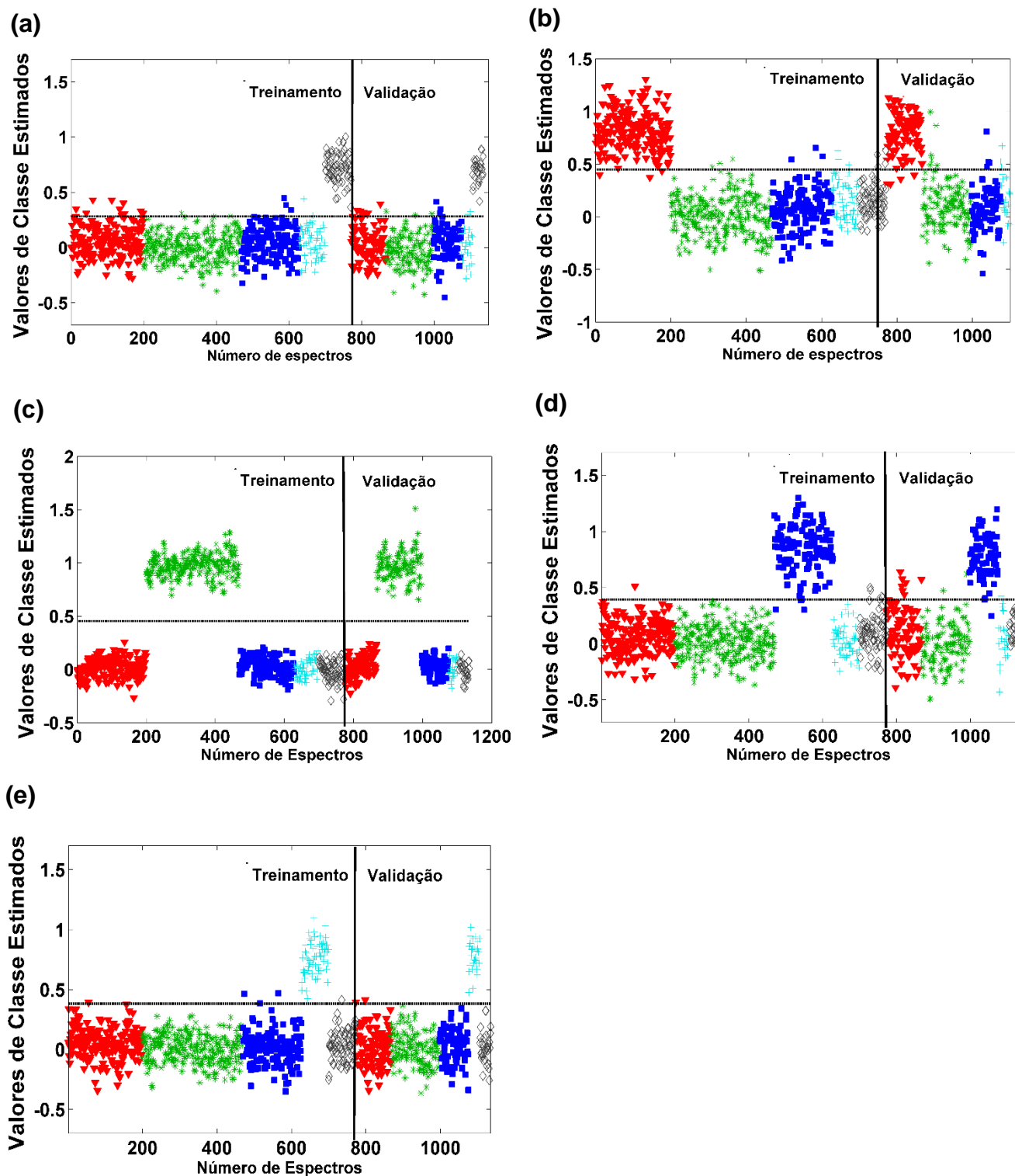


Figura 17. Distribuição dos valores de classe estimados para os conjuntos de treinamento e validação obtidos com PLS-DA e algoritmo KS a partir dos dados do MicroNIR™. **(a)** Bolívia, **(b)** Brasil, **(c)** Guatemala, **(d)** México e **(e)** Peru. Legenda: (▼) Brasil, (*) Guatemala, (■) México, (+) Peru e (◇) Bolívia, (--) Limite de discriminação.

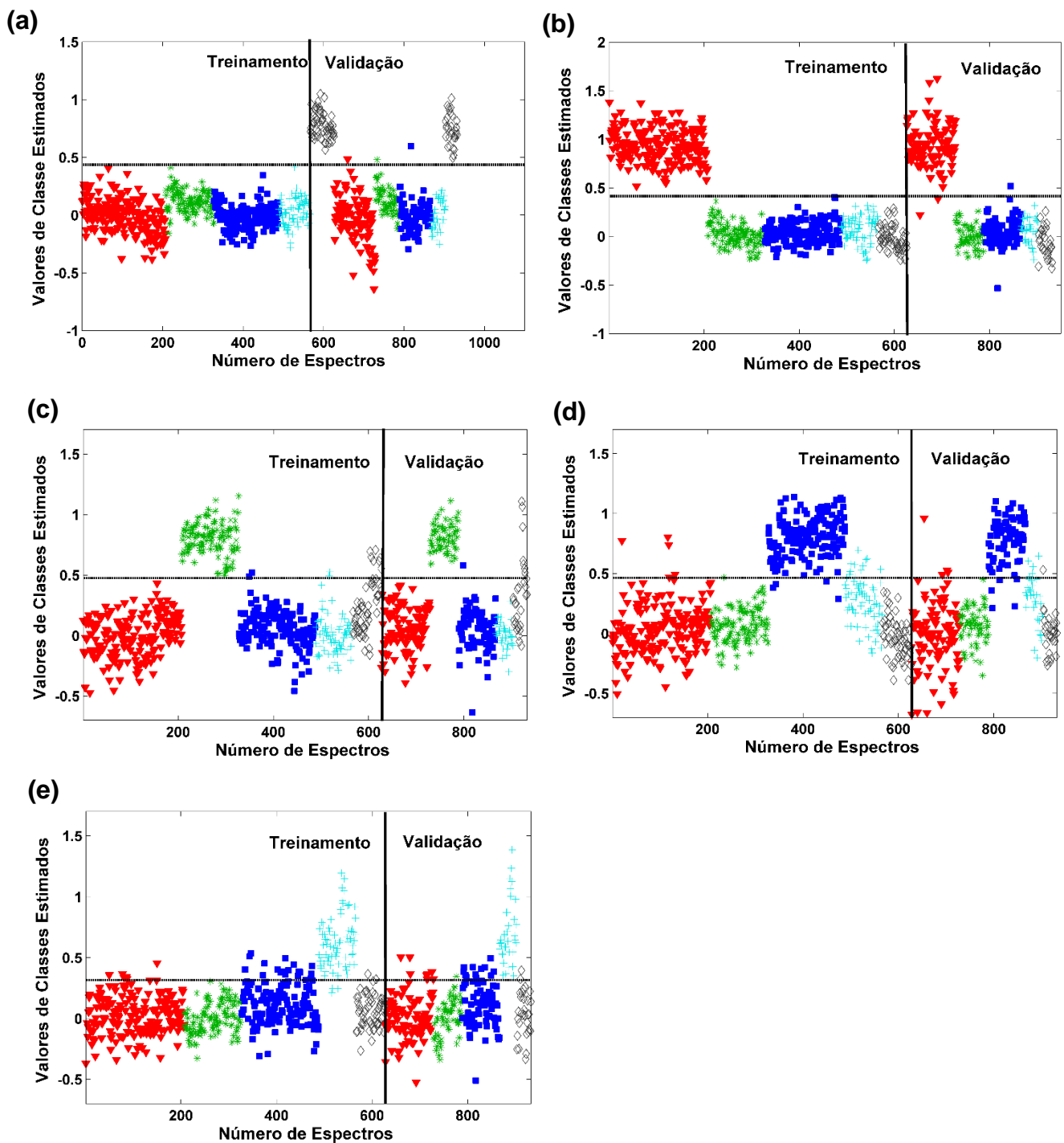


Figura 18. Distribuição dos valores de classe estimados para os conjuntos de treinamento e validação obtidos com PLS-DA e algoritmo KS a partir dos dados do MicroPHAZIR™. **(a)** Bolívia, **(b)** Brasil, **(c)** Guatemala, **(d)** México e **(e)** Peru. Legenda: (\blacktriangledown) Brasil, ($*$) Guatemala, (\blacksquare) México, ($+$) Peru e (\diamond) Bolívia, (--) Limite de discriminação.

4.1.2.1.2. Algoritmo Duplex

A Tabela 6 apresenta os parâmetros dos modelos PLS-DA utilizando-se o algoritmo Duplex. Ao se comparar, por exemplo, a quantidade de Variáveis Latentes (VL) para ambos algoritmos, os valores não diferiram significativamente para o MicroPHAZIR™ sabendo que a média da quantidade de VL foi 18 para o algoritmo Kennard-Stone (KS), (ver Tabela 4). Já para o MicroNIR™, ocorre uma pequena variação entre a quantidade de variáveis, havendo uma diminuição do número de VL do KS para o Duplex, uma vez que a média da quantidade de VL usada com algoritmo KS foi de 19.

Essa variação provavelmente mostra que o método de seleção de amostras por Duplex consegue distribuir bem as amostras entre os conjuntos de treinamento e validação, não necessitando de um número alto de VL e tendo uma dispersão mais homogênea dos dados entre os conjuntos.

Analisando o Limite de Discriminação (LD) na Tabela 6, os valores oscilaram em torno de 0,350 a 0,442 para MicroNIR™ e de 0,350 a 0,471 para MicroPHAZIR™, mantendo valores próximos aos resultados com o Kennard-Stone (KS) expostos na Tabela 4, que variou entre 0,383 a 0,454 e 0,310 a 0,465, respectivamente, para dados do MicroNIR™ e MicroPHAZIR™.

A principal diferença encontrada entre os resultados dos modelos desenvolvidos com os algoritmos foi o aumento da quantidade de *outliers* de validação, do algoritmo KS para o Duplex. A média dos *outliers* de validação para o KS foi de 21 tanto para os dados do MicroNIR™ quanto para os do MicroPHAZIR™, enquanto que com o algoritmo Duplex a média dos *outliers* de validação foi de 25 para MicroNIR™ e 28 para MicroPHAZIR™. O número de *outliers* de treinamento permaneceu próximo para ambos os algoritmos com os dados dos dois espectrômetros.

Tabela 6. Resultados dos parâmetros dos modelos PLS-DA e número de *outliers* identificados para o MicroNIR e MicroPHAZIR, respectivamente, utilizando-se o algoritmo Duplex.

<i>Países</i>	MicroNIR™				MicroPHAZIR™			
	VL	LD	OUT_trein	OUT_val	VL	LD	OUT_trein	OUT_val
<i>Bolívia</i>	16	0,355	16	23	16	0,471	22	28
<i>Brasil</i>	16	0,426	16	19	17	0,463	16	25
<i>Guatemala</i>	15	0,442	20	30	17	0,449	26	27
<i>México</i>	15	0,428	23	26	18	0,449	17	28
<i>Peru</i>	16	0,351	15	25	16	0,350	16	33

VL: número de variáveis latentes; LD: limite de discriminação; OUT_trein: número de *outliers* excluídos no conjunto de treinamento; OUT_val: número de *outliers* excluídos no conjunto de validação.

Analisando as figuras de mérito para os modelos PLS-DA desenvolvidos com o algoritmo Duplex, descritos na Tabela 7, para os resultados do MicroNIR™, a Guatemala apresentou a maior TEF no valor de 100,0% e o México apresentou a menor no valor de 77,4%. No entanto, para o MicroPHAZIR™, a Bolívia e o Brasil apresentaram as maiores TEFs com 100,0% e o Peru a menor com 91,7%.

Tabela 7. Figuras de mérito para os modelos PLS-DA calculadas a partir do conjunto de validação, utilizando-se o algoritmo Duplex.

Modelo/ País	MicroNIR™			MicroPHAZIR™		
	TFN (%)	TFP (%)	TEF (%)	TFN (%)	TFP (%)	TEF (%)
PLS-DA						
<i>Bolívia</i>	2,8	2,4	94,7	0,0	0,0	100,0
<i>Brasil</i>	5,3	10,9	83,7	0,0	0,0	100,0
<i>Guatemala</i>	0,0	0,0	100,0	0,0	2,2	97,7
<i>México</i>	18,9	3,6	77,4	2,7	2,4	94,8
<i>Peru</i>	2,9	2,4	94,6	6,6	1,6	91,7

TFN – taxa de falso negativo; TFP – taxa de falso positivo; TEF – taxa de eficiência.

Os gráficos de Hotelling T² e resíduos Q (Figuras 19 e 20) foram analisados para verificar a presença de *outliers* e se os espectros eram compatíveis com o conjunto de treinamento usado para construção do modelo. Na Figura 19 mostra os

gráficos T^2 e Q para os modelos construídos com os dados do MicroNIR™. Na Figura 19.a, espectros de amostras dos cinco países analisados apresentaram *outliers* dentro do modelo da Bolívia. Nas Figuras 19.b, 19.d e 19.e, foi possível observar que, para os três modelos, *outliers* de amostras do Brasil, Guatemala e México, ou seja, essas amostras não foram compatíveis com o conjunto de treinamento utilizado para construção dos modelos. Já na Figura 19.c, os espectros das amostras consideradas *outliers* foram oriundas das amostras do México, Brasil e Guatemala. Contudo, assim como nos modelos construídos com o algoritmo Kennard-Stone, mais que 95% das amostras de validação foram compatíveis com as de treinamento, confirmando a validade dos modelos.

A Figura 20 apresenta os gráficos de Hotelling T^2 e resíduos Q para os modelos construídos a partir dos dados do espectrômetro MicroPHAZIR™. Percebe-se que na Figura 20.a, não houve presença de *outliers*, uma vez que todos espectros de amostras estão localizados dentro dos limites de confiança de 99,9%. As Figuras 20.b e 20.e apresentam ambas poucos *outliers* de espectros de amostras brasileiras, peruanas e mexicanas. Já nas Figuras 20.c e 20.d há em comum *outliers* de amostras do Brasil, Peru e México e diferindo em função de Figura 20.c possuir *outliers* da Bolívia e em Fig. 20.d possuir *outliers* da Guatemala.

Comparando os gráficos de T^2 e Q (Figuras 19 e 20) para ambos instrumentos portáteis, a maior quantidade de *outliers* observada foi para os modelos do MicroNIR™, da mesma forma que a análise utilizando o algoritmo Kennard-Stone.

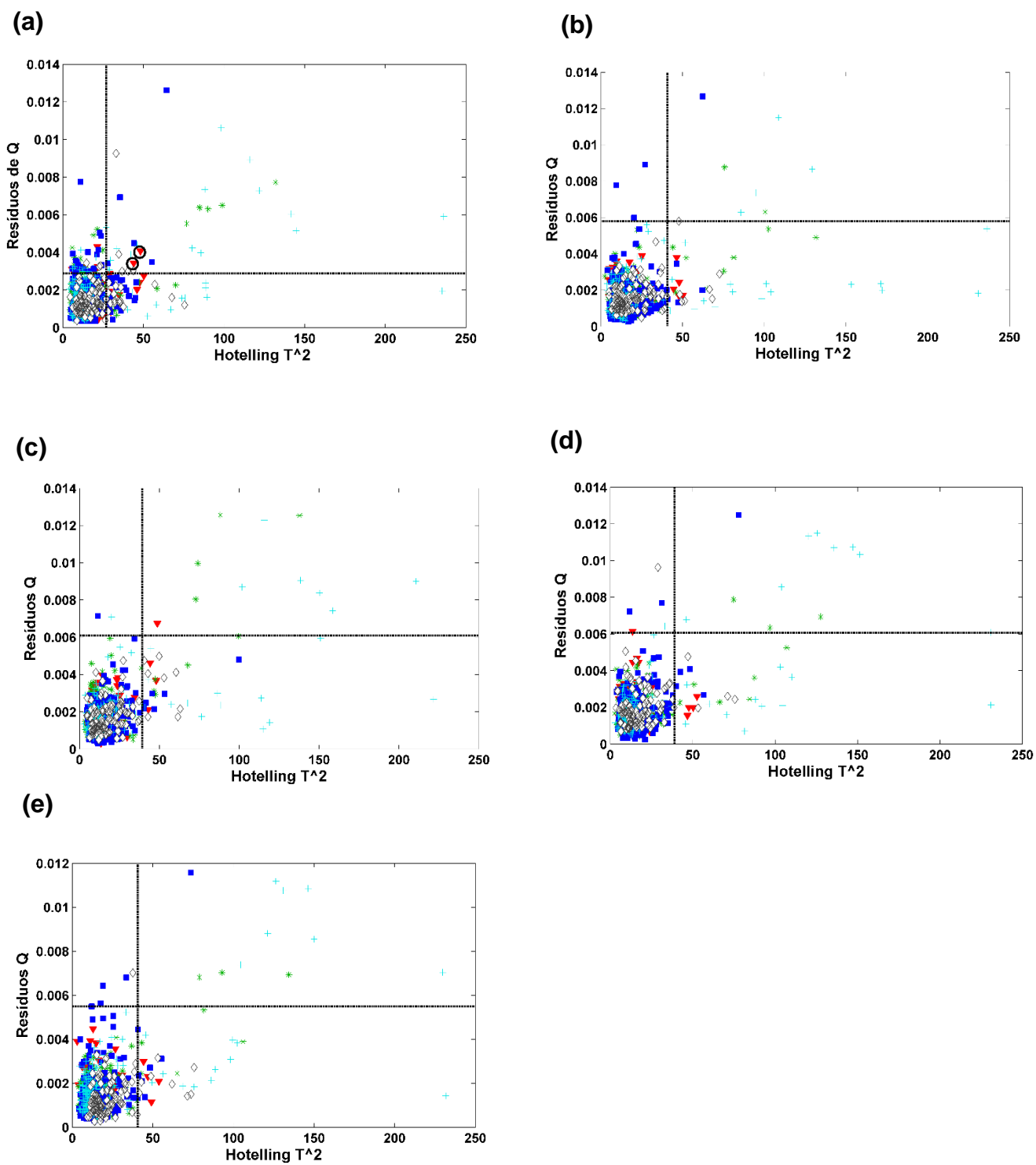


Figura 19. Gráfico dos valores de T^2 de Hotelling e resíduos Q das amostras de (a) Bolívia, (b) Brasil, (c) Guatemala, (d) México e (e) Peru obtidos com PLS-DA e algoritmo Duplex a partir dos dados do MicroNIR™. Legenda: (▼) Bolívia, (*) Brasil, (■) Guatemala, (+) México e (◇) Peru, (--) Limites de confiança (99,9%).

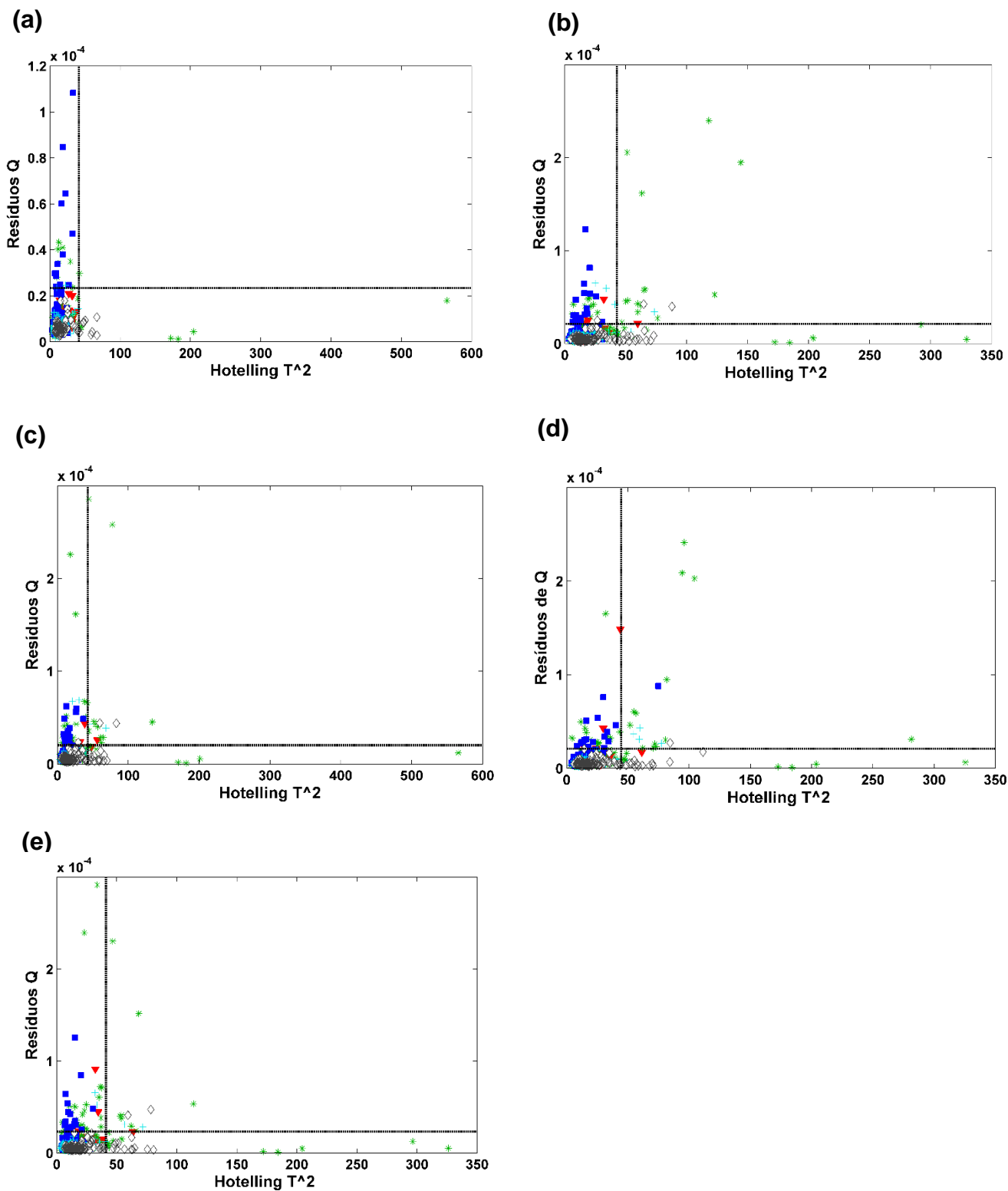


Figura 20. Gráfico dos valores de T^2 de Hotelling e resíduos Q das amostras de (a) Bolívia, (b) Brasil, (c) Guatemala, (d) México e (e) Peru obtidos com PLS-DA e algoritmo Duplex a partir dos dados do MicroPHAZIR™. Legenda: (\blacktriangledown) Bolívia, ($*$) Brasil, (\blacksquare) Guatemala, ($+$) México e (\diamond) Peru, (--) Limites de confiança (99,9%).

As Figuras 21 e 22 apresentam os gráficos de distribuição dos valores de classes estimados para os modelos PLS-DA que utilizaram algoritmo de seleção de amostras Duplex, a partir dos dados espectrais do MicroNIR™ e MicroPHAZIR™, respectivamente.

Analisando a Figura 21, o modelo da Guatemala (Fig. 21.c) foi aquele com a dispersão dos valores de classes estimados mais isolada e definida em relação às outras classes, enquanto o restante dos modelos apresenta uma dispersão desses valores menor e valores mais próximos, tornando-as sujeitas a falsos positivos e falsos negativos. Visualmente, os modelos do Brasil (Fig. 21.b) e do México (Fig. 21.d) possuem a maior quantidade de falsos positivos, concordando com as Taxas de Falso Positivo (TFP) de 10,9% e 3,6%, respectivamente, mostradas na Tabela 7. Foi ainda possível observar nos gráficos desses modelos que esses possuem as mais altas Taxas de Falso Negativo (TFN).

Os gráficos da Figura 22 construídos a partir dos dados do MicroPHAZIR™ mostram visualmente uma redução na quantidade de falsos positivos e falsos negativos, em comparação com a Figura 21. Essa redução pode também ser verificada a partir da observação das TFP e TFN (Tabela 7) dos modelos desenvolvidos com cada instrumento. As Fig. 22.c e 22.d ilustram os modelos da Guatemala e México com as maiores quantidades de falsos positivos – 2,2% e 2,4%, respectivamente.

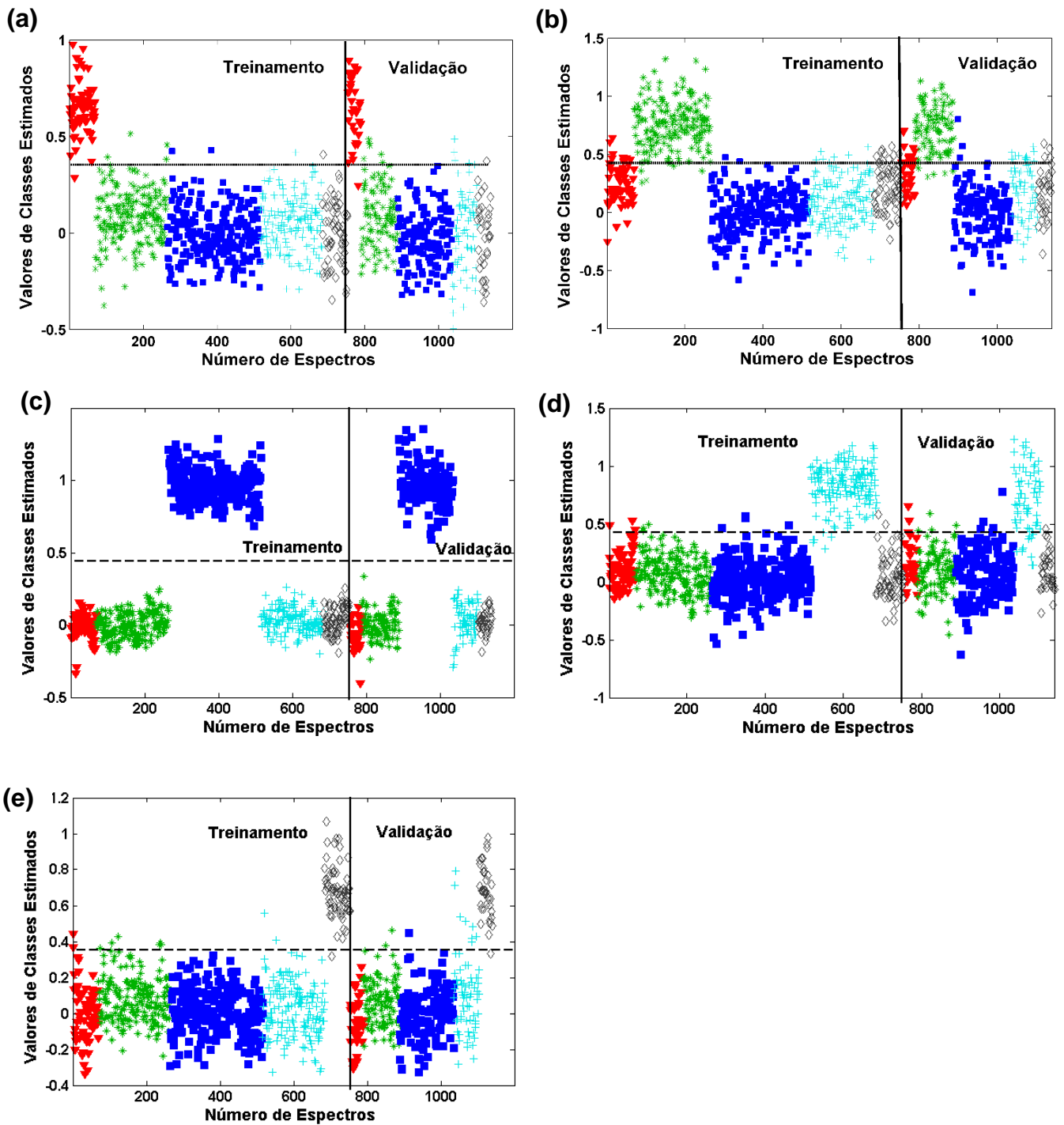


Figura 21. Distribuição dos valores de classe estimados para os conjuntos de treinamento e validação obtidos com PLS-DA e algoritmo Duplex a partir dos dados do MicroNIR™. **(a)** Bolívia, **(b)** Brasil, **(c)** Guatemala, **(d)** México e **(e)** Peru. Legenda: (▼) Bolívia, (*) Brasil, (■) Guatemala, (+) México e (◇) Peru, (--) Limite de discriminação.

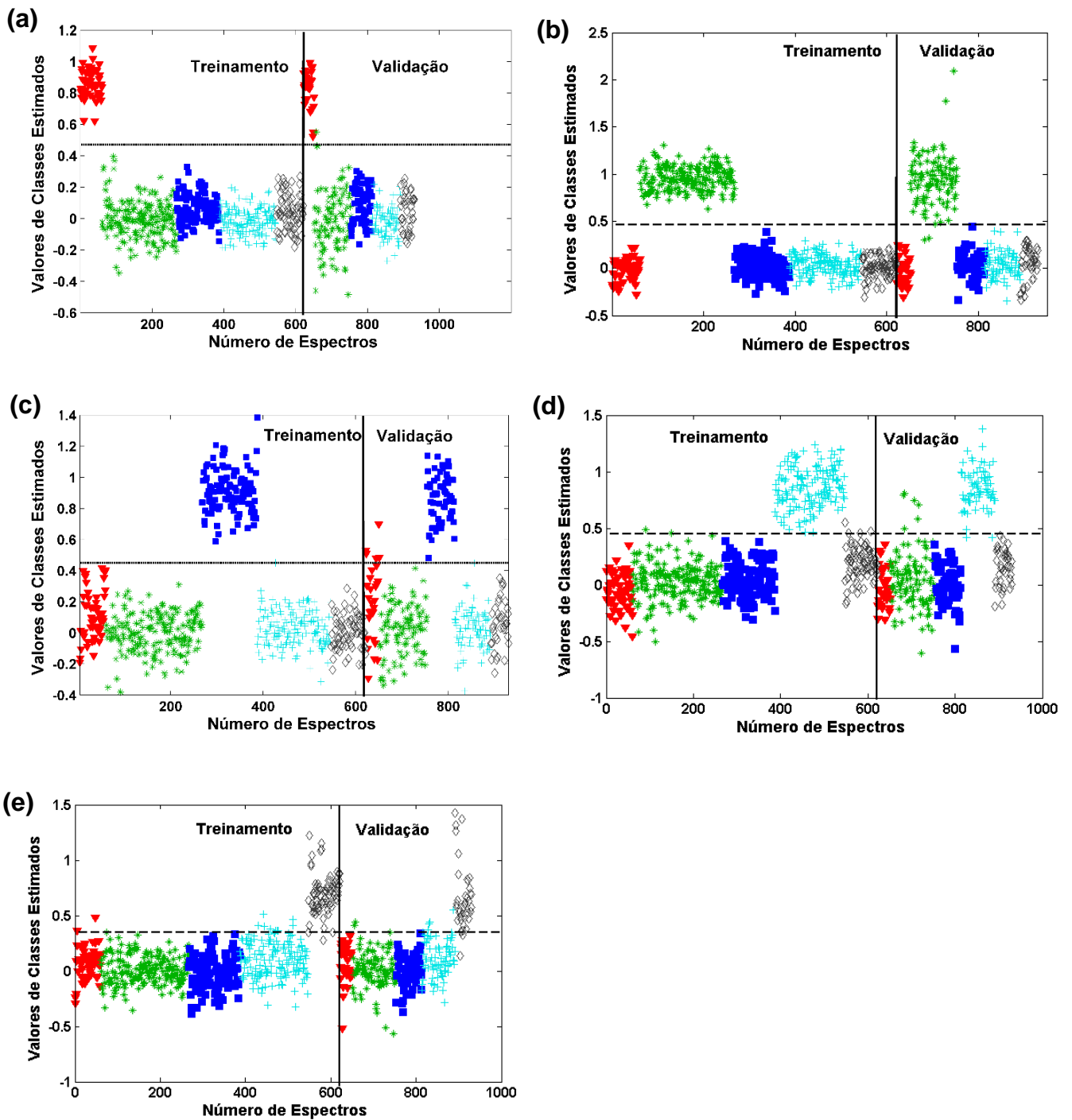


Figura 22. Distribuição dos valores de classe estimados para os conjuntos de treinamento e validação obtidos com PLS-DA e algoritmo Duplex a partir dos dados do MicroNIR™. **(a)** Bolívia, **(b)** Brasil, **(c)** Guatemala, **(d)** México e **(e)** Peru. Legenda: (▼) Bolívia, (*) Brasil, (■) Guatemala, (+) México e (◇) Peru, (--) Limite de discriminação.

4.1.2.2. SIMCA

4.1.2.2.1. Algoritmo Kennard-Stone

A análise SIMCA dos dados espectrais das amostras dos cinco países recebeu uma abordagem distinta daquela nos modelos PLS-DA, uma vez que o SIMCA é um método de modelagem. O número de componentes principais (CP) escolhido para o modelo SIMCA com os dados do MicroNIR™, utilizando o algoritmo de seleção de amostra Kennard-Stone (KS), foram 14, 12, 11, 11 e 11 para Bolívia, Brasil, Guatemala, México e Peru, respectivamente. Enquanto, para os dados do MicroPHAZIR™ foram 16, 14, 15, 16 e 15, para os modelos de Bolívia, Brasil, Guatemala, México e Peru, respectivamente. Uma média de 12 CPs para MicroNIR™ e 15 CPs para MicroPHAZIR™. Esse número relativamente alto de componentes principais (CP) utilizado nos modelos tornou-se necessário para explicar o maior número de informações válidas possível, devido às variações significativas ao longo da amostragem e à heterogeneidade das amostras.

A Tabela 8 apresenta as figuras de mérito obtidas dos modelos SIMCA usando o algoritmo KS. Observa-se que as Taxas de Eficiências (TEFs) variam entre 54,2 a 98,4% para o MicroPHAZIR™, e entre 58,3 a 100,0% para o MicroNIR™.

Tabela 8. Figuras de mérito para os modelos de SIMCA calculadas a partir do conjunto de validação, utilizando-se o algoritmo Kennard-Stone.

Modelo País	MicroNIR™			MicroPHAZIR™		
	TFN (%)	TFP (%)	TEF (%)	TFN (%)	TFP (%)	TEF (%)
SIMCA						
<i>Bolívia</i>	53,3	0,0	46,7	0,0	3,3	96,7
<i>Brasil</i>	8,3	4,4	87,3	0,0	13,2	86,8
<i>Guatemala</i>	0,0	0,0	100,0	1,6	0,0	98,4
<i>México</i>	10,2	9,0	80,8	24,3	2,6	73,1
<i>Peru</i>	40,0	1,7	58,3	44,4	1,4	54,2

TFN – taxa de falso negativo; TFP – taxa de falso positivo; TEF – taxa de eficiência.

Com base nos valores de TEF de alguns modelos de países, como Peru e Bolívia com dados do MicroNIR™, e México e Peru com dados do MicroPHAZIR™, darem altas taxas de falso negativo (TFN), evidenciando uma aproximação entre as classes em relação ao espaço das CPs. A performance dos modelos das amostras peruanas e mexicanas foram um pouco melhor para os dados do MicroNIR™ que

para os dados do MicroPHAZIR™; porém, para o modelo da Bolívia, houve uma inversão, onde o MicroPHAZIR™ apresentou TEF maior que o MicroNIR™. No geral, os modelos com dados do MicroNIR™ foram cerca de 20% melhor que o MicroPHAZIR™.

A correlação de dados está presente na Tabela 9. Essa correlação serviu para os cálculos de definição das figuras de mérito, uma vez que relaciona as amostras que foram preditas no modelo e a qual país pertencem realmente.

Tabela 9. Correlação entre o número de espectros de amostras identificados na etapa de validação pelo modelo SIMCA de cada país, de acordo com os dados de cada espectrômetro, utilizando-se o algoritmo Kennard-Stone.

Instrumento, <i>País predito</i>	País em questão				
	<i>Bolívia</i>	<i>Brasil</i>	<i>Guatemala</i>	<i>México</i>	<i>Peru</i>
MicroPHAZIR™					
<i>Bolívia</i>	29	0	0	0	0
<i>Brasil</i>	1	102	1	15	10
<i>Guatemala</i>	0	0	129	0	0
<i>México</i>	0	0	0	59	6
<i>Peru</i>	0	0	0	4	20
MicroNIR™					
<i>Bolívia</i>	14	0	0	0	0
<i>Brasil</i>	5	94	0	4	7
<i>Guatemala</i>	0	0	130	0	0
<i>México</i>	11	6	0	70	9
<i>Peru</i>	0	2	0	4	20

Os resultados do modelo SIMCA das amostras brasileiras usando dados do MicroPHAZIR™ e o algoritmo Kennard-Stone (MicroPHAZIR-SIMCA-KS) mostram que a maioria das amostras classificadas erroneamente como amostras brasileiras pertencem ao México. As amostras da Guatemala e Bolívia mostraram os menores erros de classificação para os modelos MicroPHAZIR™, enquanto apenas o modelo da Guatemala com os dados do MicroNIR™ apresentou os menores erros de classificação. Os resultados dos modelos MicroNIR™-SIMCA-KS revelam que apenas as amostras guatemaltecas possuem o grupamento mais isolado. Os grupamentos dos outros países estão se sobrepondo, dificultando a discriminação pelo modelo SIMCA.

4.1.2.2.2 Algoritmo Duplex

Os números de Componentes Principais escolhidos usando o algoritmo Duplex na modelagem dos dados espectrais das amostras dos países por SIMCA com dados do MicroNIR™ foram 14, 16, 10, 17 e 15; enquanto que, para o MicroPHAZIR™ foram 14, 16, 15, 14 e 12, para Bolívia, Brasil, Guatemala, México e Peru, respectivamente. Foi observada uma proximidade entre o número de CPs de cada espectrômetro, tendo uma média de 14 CPs por cada equipamento. Assim como na modelagem utilizando o Kennard-Stone (KS), utilizou-se um número alto de componentes principais, sendo que a média de CPs para ambos instrumentos com Duplex, 14, foi maior que a média de CPs para o MicroNIR™ com KS (12) e menor que a média para o MicroPHAZIR™ com KS (15).

As Tabelas 8 e 9 apresentam as figuras de mérito para os modelos SIMCA para o algoritmo KS, mostram que as TEFs variando entre 54,2 a 98,4% para o MicroPHAZIR™, e entre 58,3 a 100,0% para o MicroNIR™. E para o algoritmo Duplex, as TEFs variando entre 56,7 a 81,3% utilizando o MicroNIR™ e 72,2 a 100% para MicroPHAZIR™.

A Tabela 10 apresenta as figuras de mérito para os modelos SIMCA utilizando o algoritmo Duplex. Foi observado Taxas de Eficiência (TEFs) variando entre 56,7 a 81,3% para MicroNIR™ e 72,2 a 100% para MicroPHAZIR™. Observou-se que os valores de TEFs para os dados do MicroNIR™ foram menores para os modelos utilizando o Duplex, em comparação com o KS. Em contraste, as TEFs dos modelos do MicroPHAZIR™ apresentaram melhores resultados com o Duplex do que com KS. Foi observado ainda que, para o Duplex todos os modelos de dados MicroPHAZIR™ apresentaram melhores resultados do que os modelos MicroNIR™.

Tabela 10. Figuras de mérito para os modelos de SIMCA calculadas a partir do conjunto de validação, utilizando-se o algoritmo Duplex.

Modelo País	MicroNIR™			MicroPHAZIR™		
	TFN (%)	TFP (%)	TEF (%)	TFN (%)	TFP (%)	TEF (%)
SIMCA						
<i>Bolívia</i>	16,6	3,1	80,3	0,0	0,0	100,0
<i>Brasil</i>	11,1	7,6	81,3	4,3	0,9	94,8
<i>Guatemala</i>	0,0	1,2	98,8	0,4	0,0	99,6
<i>México</i>	23,1	1,8	75,1	16,6	3,4	80,0
<i>Peru</i>	41,6	1,7	56,7	25,6	2,2	72,2

TFN – taxa de falso negativo; TFP – taxa de falso positivo; TEF – taxa de eficiência

A Tabela 11 mostra a correlação entre amostras classificadas dentro do modelo de um país e a verdadeira origem daquelas amostras utilizando Duplex para os dados de ambos instrumentos portáteis. Os modelos do México apresentaram os maiores erros em classificar amostras mexicanas como sendo mexicanas, discriminando erroneamente 18 amostras, utilizando o MicroNIR™, e 13 amostras, com MicroPHAZIR™. Os modelos com o segundo maior erro de classificação foram Brasil e o Peru, ambos errando 15 amostras cada um, classificadas como de outros países para o MicroNIR™ e o modelo do Peru errando 10 amostras para o MicroPHAZIR™. Os modelos do Brasil tanto para MicroNIR™ quanto para o MicroPHAZIR™ apresentaram os maiores erros em classificar amostras de outros países como sendo do Brasil, tendo 22 e 9 amostras, respectivamente MicroNIR™ e MicroPHAZIR™.

Tabela 11. Correlação entre o número de espectros de amostras identificados na etapa de validação pelo modelo SIMCA de cada país de acordo com os dados de cada espectrômetro, utilizando-se o algoritmo Duplex.

Instrumento, <i>País predito</i>	País em questão				
	<i>Bolívia</i>	<i>Brasil</i>	<i>Guatemala</i>	<i>México</i>	<i>Peru</i>
MicroPHAZIR™					
<i>Bolívia</i>	30	0	0	0	0
<i>Brasil</i>	0	101	0	7	2
<i>Guatemala</i>	0	0	130	0	1
<i>México</i>	0	1	0	65	7
<i>Peru</i>	0	0	0	6	26
MicroNIR™					
<i>Bolívia</i>	26	5	0	7	1
<i>Brasil</i>	4	88	0	7	11
<i>Guatemala</i>	0	7	130	0	0
<i>México</i>	2	1	0	60	3
<i>Peru</i>	0	2	0	4	21

Analisando o resultado geral, os modelos MicroPHAZIR-Duplex obtiveram, resultados satisfatórios. Foi possível deduzir que as TEFs mais baixas para MicroNIR, devem-se ao fato de suas classes estarem muito próximas umas das outras, isto é, sobrepostas, dificultando a separação das nuvens espectrais das amostras. Portanto, a avaliação pelo PLS-DA possui maior capacidade discriminatória para essas classes sobrepostas.

4.1.2.3. Comparação entre os modelos para identificação de origem

Observando a Tabela 12, foi possível identificar qual condição (Instrumento + algoritmo de seleção de amostra + modelo quimiométrico) apresentou os melhores resultados para cada país. O modelo para Bolívia obteve melhor resultado em duas situações: MicroPHAZIR-Duplex-PLSDA (TEF de 100%) e MicroPHAZIR-Duplex-SIMCA (TEF de 100%). Já o modelo para o Brasil, o melhor resultado foi MicroPHAZIR-Duplex-PLSDA (TEF de 100%). Para a Guatemala, a melhor condição ocorreu em três situações, todas com TEFs de 100%: MicroNIR-KS-PLSDA, MicroNIR-Duplex-PLS-DA e MicroNIR-KS-SIMCA. Para o México a condição que obteve melhor resultado (TEF de 94,9%) foi MicroNIR-KS-PLS-DA e para o Peru a melhor condição (TEF de 99,4%) foi MicroNIR-KS-PLSDA.

Tabela 12. Conjunto de taxas de eficiências dos modelos PLS-DA e SIMCA para os dados dos dois espectrômetros portáteis (MicroNIR™ e MicroPHAZIR™) das amostras dos países a partir dos dois algoritmos de seleção de amostra (Kennard-Stone e Duplex).

Modelo Algoritmo	PLS-DA		SIMCA	
	Kennard-Stone (%)	Duplex (%)	Kennard-Stone (%)	Duplex (%)
MicroNIR				
<i>Bolívia</i>	99,3	94,7	46,7	80,3
<i>Brasil</i>	90,1	83,7	87,3	81,3
<i>Guatemala</i>	100,0	100,0	100,0	98,8
<i>México</i>	94,9	77,4	80,8	75,1
<i>Peru</i>	99,4	94,6	58,3	56,7
MicroPhazir				
<i>Bolívia</i>	99,6	100,0	96,7	100,0
<i>Brasil</i>	98,4	100,0	86,8	94,8
<i>Guatemala</i>	96,5	97,7	98,4	99,6
<i>México</i>	89,6	94,8	73,1	80,0
<i>Peru</i>	91,2	91,7	54,2	72,2

De forma global, o instrumento e algoritmo de seleção de amostras que apresentaram melhor desempenho foram o MicroPhazir e Duplex, destacados em negrito na Tabela 12. Com relação ao modelo quimiométrico, o PLS-DA apresentou

um melhor desempenho, sendo significativamente superior para os países Peru e México.

4.1.2.4. Conclusões parciais.

A caracterização de madeira é, em demasia complexa em função da heterogeneidade natural. Essa primeira parte do estudo, focou a determinação da procedência/origem da mesma espécie madeireira, pode ser considerada bem-sucedida e com boas perspectivas para aplicação na fiscalização do comércio de mogno. Contudo, ainda são necessários estudos de robustez e aplicação piloto em campo, para assegurar que os modelos desenvolvidos possam reproduzir os mesmos ótimos resultados apresentados em análises de apreensões e fiscalização de madeira.

4.2. Discriminação das espécies *S. macrophylla* e *S. humilis*

4.2.1. Espectros NIR das madeiras de *S. macrophylla* e *S. humilis*

A Figura 23 apresenta a média dos espectros NIR brutos, isto é, sem pré-processamento, das espécies *S. macrophylla* e *S. humilis*, designadas comumente de mogno. Foi perceptível que os formatos dos espectros foram muito semelhantes para ambas espécies, não havendo como determinar visivelmente qual espectro pertence a uma espécie específica, uma vez que a diferença de intensidade reflete apenas uma variação de linha de base, a qual não pode ser atribuída como característica da espécie.

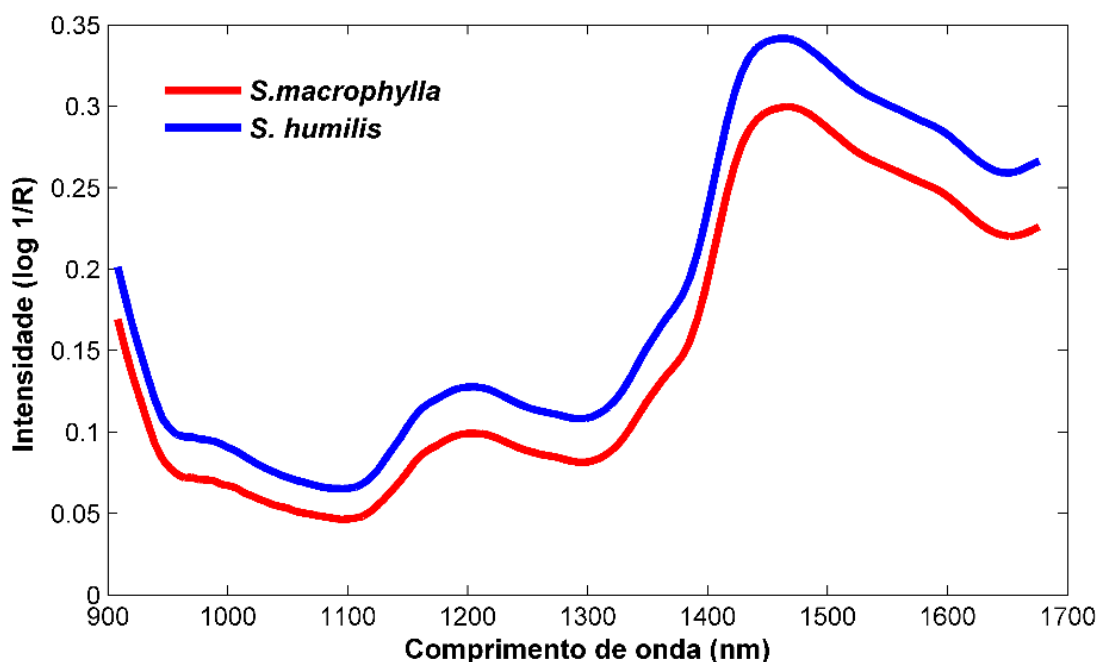


Figura 23. Média dos espectros NIR das madeiras das espécies *S. macrophylla* e *S. humilis* ou mogno antes do pré-processamento.

A Figura 24 mostra a média dos espectros pré-processados com primeira derivada das amostras de cada espécie estudada. Após a correção da variação de linha de base proporcionada pelo pré-processamento, ainda não foi possível fazer uma discriminação entre os espectros de ambas espécies. Entretanto, é perceptível algumas variações nas principais bandas: entre 1.110 e 1.200 nm, região espectral associada ao estiramento assimétrico do C-H e HC=CH da lignina, segundo sobretudo do estiramento do C-H dos grupos de CH₃ do conjunto de acetil éster nas hemiceluloses e segundos sobretudo dos estiramentos de C-H das moléculas de celulose e lignina. E entre 1.300 e 1.400 nm, intervalo relacionado aos primeiros sobretudo do estiramento do C-H e deformação angular em celulose e hemicelulose.

Novamente, não foi possível classificar as amostras entre uma espécie ou outra, tornando-se necessário utilizar uma análise multivariada dos dados espectrais (Schwanninger et al, 2011).

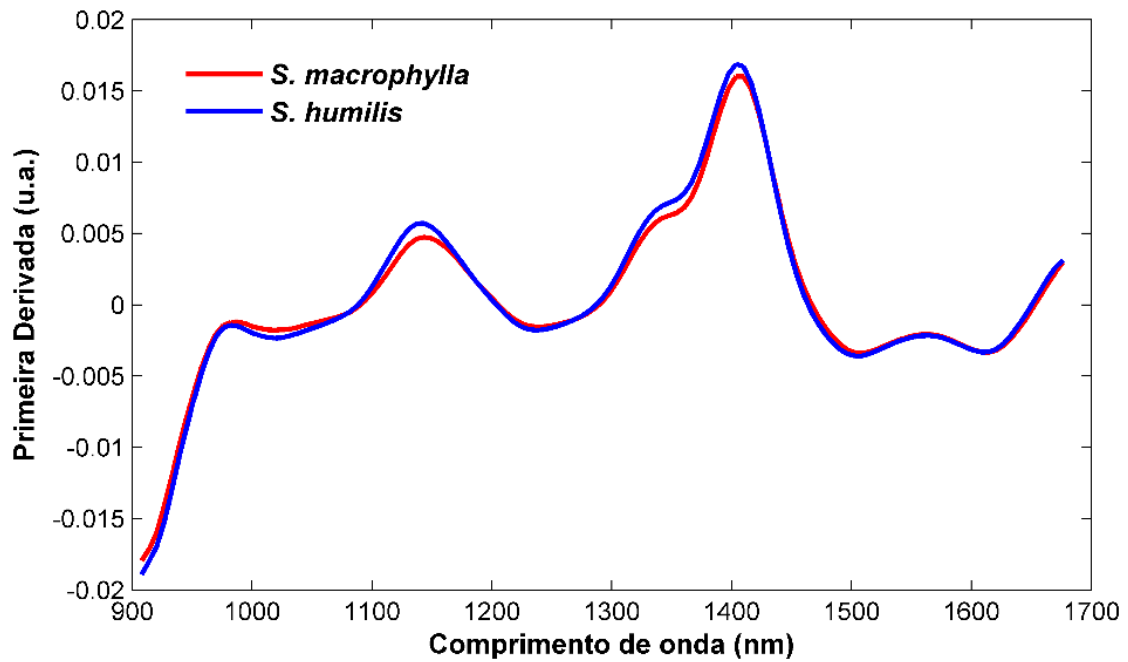


Figure 24. Média dos espectros NIR das madeiras das espécies *S. macrophylla* e *S. humilis* ou mogno após o pré-processamento com primeira derivada SG (Savitzky-Golay).

Uma comparação entre os dados das amostras de ambas espécies do espectrômetro MicroNIR™ e MicroPhazir™ não foi realizada em função de tempo hábil para análise, tornando-se um estudo futuro para observar se há variação entre os resultados dos equipamentos portáteis.

4.2.2. Análise Quimiométrica

4.2.2.1. PLS-DA

A partir da Tabela 13, é possível observar os parâmetros obtidos após a construção dos modelos PLS-DA das espécies *S. macrophylla* e *S. humilis*.

Tabela 13. Resultados dos parâmetros dos modelos PLS-DA e número de *outliers* identificados para a discriminação das espécies *S. macrophylla* e *S. humilis*.

Parâmetros	Espécies	
	<i>S. macrophylla</i>	<i>S. humilis</i>
VL	13	12
LD	0,526	0,519
OUT_Trein	3	1
OUT_Val	3	39

VL: número de variáveis latentes; LD: limite de discriminação; OUT_trein: número de *outliers* excluídos no conjunto de treinamento; OUT_val: número de *outliers* excluídos no conjunto de validação.

Os números de variáveis latentes são próximos e altos. O valor alto de VL se tornou necessário para conseguir atingir a especificidade da classificação. Os valores de limite de discriminação para os modelos de ambas espécies foram próximos um do outro e próximos de 0,5 explicando que as dispersões das distribuições dos valores de classes estimados são similares e separando bem as espécies, sendo modelos binários.

O modelo de *S. humilis* apresentou uma grande quantidade de *outliers* de validação (Tabela 13), isso pode ser explicado pelo alto teor de umidade das amostras de um dos locais de coleta, alterando de forma brusca os espectros dessas amostras e tornando-as *outliers*. Em consequência, os *outliers* de treinamento foram poucos e o modelo de *S. macrophylla* apresentou baixos valores de *outliers*, sendo o percentual de *outliers* 29% para *S. humilis* e 1,25% para *S. macrophylla*.

Analisando as informações da Tabela 14, ambos modelos apresentaram TEFs (Taxa de Eficiência) de 100%. Mostrando que as amostras de cada espécie foram discriminadas sem nenhum falso positivo ou negativo. Comprovou-se que a análise multivariada de dados espectrais consegue ser específica a ponto de conseguir distinguir madeira de espécies diferentes de um mesmo gênero.

Tabela 14. Figuras de mérito para os modelos PLS-DA das espécies *S. macrophylla* e *S. humilis* calculadas a partir do conjunto de validação.

Figuras de Mérito	Espécies	
	<i>S. macrophylla</i>	<i>S. humilis</i>
TFN (%)	0,0	0,0
TFP (%)	0,0	0,0
TEF (%)	100,0	100,0

Observando a Figura 25 que mostra os gráficos de Hotelling T^2 e resíduos Q para os modelos das amostras da espécie *S. macrophylla* (Fig. 25.a) e *S. humilis* (Fig. 25.b), algumas amostras foram identificadas como *outliers*, pois estão fora do limite de confiança de 99.9%, o que invalida o uso dos modelos para essas amostras. O modelo de *S. humilis* (Fig. 25.b) apresentou apenas uma amostra que não está apta para ser utilizada; a qual pode ter sido influenciada por diversos fatores tais como: umidade, erro instrumental e de medida, dentre outros, os quais podem ter contribuído para sua invalidação no modelo. Já o modelo para *S. macrophylla* apresentou uma quantidade maior de espectros identificados como *outliers*, invalidando também o uso do modelo para essas amostras. Entretanto, esses dois modelos não apresentaram um número grande de amostras identificadas como *outliers*, indicando que os modelos são adequados para classificar esse conjunto de validação.

Em seguida, foram analisados os gráficos de valores de classe estimados para avaliar o desempenho do modelo na identificação de amostras. Os gráficos estão representados na Figura 26, mostrando que houve uma classificação totalmente assertiva, uma vez que todas os espectros de validação de *S. macrophylla* (Fig. 26.a) e *S. humilis* (Fig. 26.b) apresentaram valores de classe estimados, acima do limite de discriminação para os seus respectivos modelos, concordando com as taxas de eficiência (TEFs) de 100% (ver Tabela 14) para ambos os modelos.

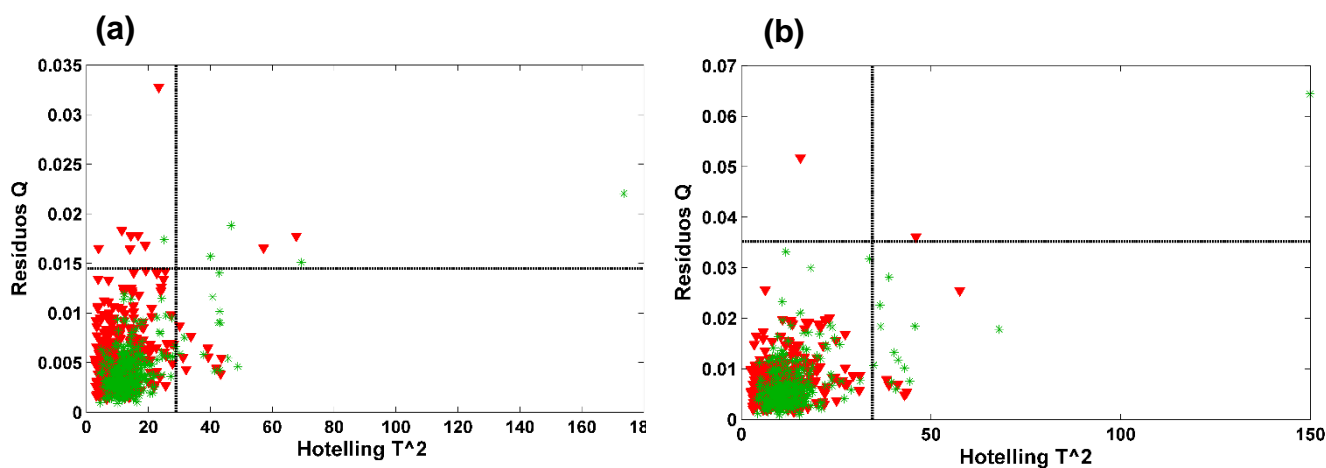


Figura 25. Gráfico dos valores de T² de Hotelling e resíduos Q das amostras de **(a)** *S. macrophylla* e **(b)** *S. humilis* obtidos com PLS-DA e algoritmo Duplex a partir dos dados do MicroNIRTM. Legenda: (▼) *S. macrophylla*, (*) *S. humilis*, (--) Limites de confiança (99,9%).

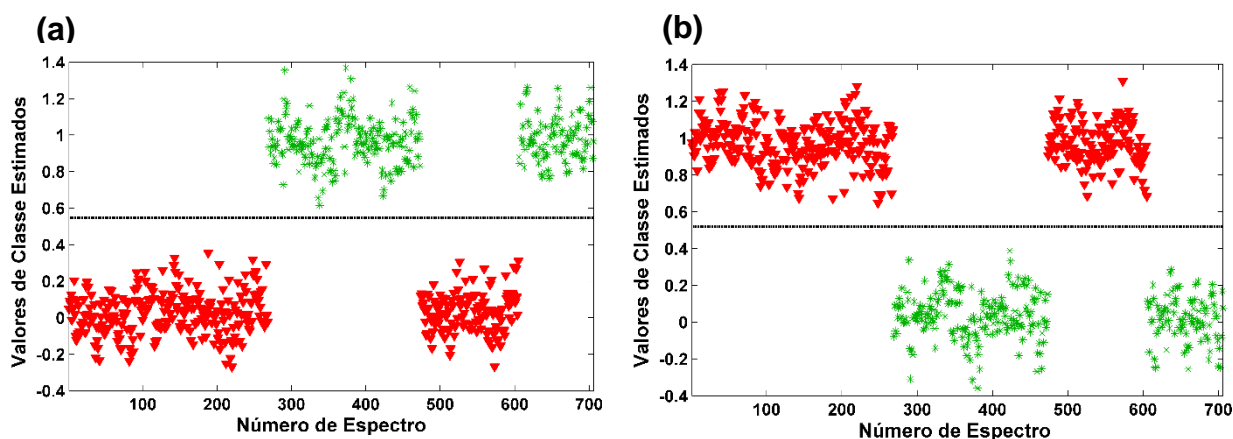


Figura 26. Distribuição dos valores de classe estimados para os conjuntos de treinamento e validação obtidos com PLS-DA a partir dos dados do MicroNIRTM. **(a)** *S. humilis* e **(b)** *S. macrophylla*. Legenda: (▼) *S. macrophylla* e (*) *S. humilis*, (--) Limite de discriminação.

4.2.2.2. SIMCA

Para fins de comparação, a análise com o modelo SIMCA também foi feita para as duas espécies de mogno (*S. macrophylla* e *S. humilis*). A Tabela 15 apresenta o resultado das figuras de mérito para esse modelo, mostrando que houve uma queda no valor das TEFs comparada com o modelo PLS-DA (Tabela 14), queda de 1,5%. Porém, o modelo está acima de 90% de acurácia, mostrando que a análise SIMCA também foi efetiva na modelagem dos dados espectrais de espécies diferentes de uma mesma madeira.

Tabela 15. Figuras de mérito para os modelos PLS-DA das espécies *S. macrophylla* e *S. humilis* calculadas a partir do conjunto de validação.

Figuras de Mérito	Espécies	
	<i>S. macrophylla</i>	<i>S. humilis</i>
TFN (%)	1,5	0,0
TFP (%)	0,0	1,5
TEF (%)	98,5	98,5

A Tabela 16 apresenta uma correlação entre as classes que as espécies pertencem e o modelo dessas amostras. Apenas dois espectros de amostras se apresentaram como sendo da espécie *S. humilis*, os quais, entretanto, pertenciam a classe das *S. macrophylla*, originando um falso negativo para o modelo de *S. macrophylla*.

Tabela 16. Correlação entre o número de espectros de amostras identificados na etapa de validação pelo modelo SIMCA de cada espécie.

Espécie predita	Espécie em questão	
	<i>S. macrophylla</i>	<i>S. humilis</i>
<i>S. macrophylla</i>	130	0
<i>S. humilis</i>	2	101

A análise de amostras de duas espécies consideradas mogno genuíno (*S. macrophylla* e *S. humilis*) utilizando os dados do MicroNIRTM apresentou valores de TEFs de 100,0% para o modelo PLS-DA e 98,5% para o SIMCA, mostrando a possibilidade da separação entre amostras de espécies do mesmo gênero com ambos os modelos.

5. CONCLUSÕES GERAIS

Os espectros NIR e modelos PLS-DA e SIMCA se mostraram eficientes na identificação da procedência das madeiras de mogno (*S. macrophylla*) dos cinco países estudados. Para a análise PLS-DA, 85% dos modelos desenvolvidos obtiveram taxas de eficiência superiores a 90%, enquanto que para a análise SIMCA apenas 45% obtiveram TEFs acima de 90%. Quanto aos algoritmos de seleção de amostras avaliados, não foram observadas diferenças marcantes em termos de eficiência, mas de forma global o algoritmo Duplex foi levemente superior.

Fazendo uma comparação dos espectrômetros utilizados nesse trabalho, pode-se concluir que ambos foram capazes de determinar a procedência das amostras. Contudo, devido a uma diferença de eficiência muito sutil e pela maior facilidade de operação e visualização do resultado, pode-se sugerir o MicroPhazir como equipamento mais indicado para uma continuação do projeto visando sua aplicação piloto em postos de fiscalização.

As espécies consideradas mogno genuíno que foram analisadas nesse estudo, em ambos modelos PLS-DA e SIMCA obtiveram TEFs de 100,0% e 98,5%, respectivamente. Estes resultados evidenciam a especificidade da tecnologia NIRS, através da qual foi possível discriminar-se espécies de madeiras que possuem um gênero taxonômico em comum, ambas consideradas mogno genuíno.

Com o objetivo de aplicar a tecnologia NIRS nos postos de fiscalização para identificação de espécies florestais transportadas, extraídas e comercializadas e ainda, com a finalidade de torna-la um método de certificação de madeira, é preciso ainda incluir mais amostras nos modelos para garantir a construção de um modelo robusto e representativo. Espera-se obter uma maior quantidade de amostras pertencentes aos cinco países estudados e novas amostras advindas de outros países para serem incluídas nos modelos de discriminação e modelagem. E ainda, estabelecer condições mínimas controladas de análise no ambiente de campo.

REFERÊNCIAS BIBLIOGRÁFICAS

- Amigo, J.M., Babamoradi, H., Elcoroaristizabal, S. (2015) Hyperspectral images analysis: A tutorial. *Anal. Chim. Acta.* 896: 34-52.
- Annual Book of ASTM Standards (2012) E1655-05. Standards Practices for Infrared Multivariate Quantitative Analysis.
- Aenugu, H.P.R.; Kumar, D.S; Srisudharson, Parthiban, N.; S.S. Ghosh, and D. Banji, Near Infra Red Spectroscopy – An Overview, *International Journal of ChemTech Research* 3 (2011) 825–836.
- Bächle, H., Zimmer, B., Winderisen, E., Wegener, G. (2010) Evaluation of thermally modified beech and spruce wood and their properties by FT-NIR spectroscopy. *Wood Sci. Technol.* 44: 421-433.
- Balladio, D.; Consonni, V. (2013), Classification tools in chemistry. Part 1: linear modeols. *PLS-DA, Anal. Methods*, 5, 3790-3798.
- Barreto, P. and C. Souza Jr. (2001). “Controle do desmatamento e da exploração de madeira da Amazônia: diagnóstico e sugestões”. Ministério do Meio Ambiente. Instituto Brasileiro de Recursos Naturais Renováveis. PPG7/Promanejo. Relatório técnico, componente III.
- Barker, M., Rayens, W. (2003) Partial least squares for discrimination. *J. Chemom.* 17: 166-173.
- Barton, F. E. (2016) Near infrared equipment through the ages and into the future. *NIR News* 27:41-44.
- Bergo, M.C.J., Pastore, T.C.M., Coradin, V.T.R., Wiedenhoeft, A.C., Braga, J.W.B. (2016) NIRS identification of *Swietenia macrophylla* is robust across specimens from 27 countries. *IAWA J.* 37: 420-430.
- Borin, A., Poppi, R. J. (2004) Multivariate quality control of lubricating oils using Fourier transform infrared spectroscopy. *J. Braz. Chem. Soc.* 15: 570-576.
- Botelho, B.G., Reis, N., Oliveira, L.S., Sena, M.M. (2015) Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA. *Food Chem.* 181: 31-37.
- Braga, J.W.B., Pastore T.C.M., Coradin, V.T.R., Camargos, J.A.A., Silva, A.R. da (2014) The use of near infrared spectroscopy to identify solid wood specimens of *Swietenia macrophylla* (CITES appendix II). *IAWA J.* 32: 285-296.
- Brasil (2003), Decreto nº 4.722, de 5 de junho, Ministério do Meio Ambiente.

- Brasil (2014), Decreto nº 6.101, portaria nº43, de 31 de janeiro, Ministério do Meio Ambiente.
- Brasil (2016), Decreto nº 34, de 24 de agosto, Ministério do Meio Ambiente.
- Brereton, R.G., Lloyd, G.R. (2014) Partial least squares discriminant analysis: taking the magic away. *J. Chemom.* 28: 213-225.
- Brereton, R. G.; *Chemometrics: Data Analysis for the Laboratory and Chemical Plant.* John Wiley & Sons: England, 2003.
- Brereton, R.G. (2000) Introduction to multivariate calibration in analytical chemistry. *Analyst.* 125: 2125-2154.
- Coradin, V.T.R., Camargos, J.A.A, Marques, L.F., Silva Jr, E.R. (2009) Madeiras similares ao mogno (*Swietenia macrophylla* King.): chave ilustrada para identificação anatômica em campo. Serviço Florestal Brasileiro, Brasil.
- Degen, B., Ward, S.E., Lemes, M.R.; Navarro, C.; Cavers, S., Sebbenn, A.M. (2013) Verifying the geographic origin of mahogany (*Swietenia macrophylla*) with DNA-fingerprints. *Forensic Sci. Int. Genet.* 7: 55-62.
- Dinwoodie, J. M.; *Timber: Its nature and behaviour*, 2nd ed.; E & FN Spon: New York, 2000.
- Dormontt, E.E., Boner, M., Braun, B., Breulmann, G., Degen, B., Espinoza, E., Gardner, S., Guillery, P., Hermanson, J.C., Koch, G., Lee, S.L., Kanashiro, M.; Rimbawanto, A., Thomas, D., Wiedenhoeft, A. C., Yin, Y., Zahnen, J., Lowe, A. J. (2015) Forensic timber identification: It's time to integrate disciplines to combat illegal logging. *Biol. Cons.* 191: 790-798.
- Dinwoodie, J. M.; *Timber: Its nature and behaviour*, 2nd ed.; E & FN Spon: New York, 2000.
- Esteves, B., Pereira, H. (2008) Quality assessment of heat-treated wood by NIR spectroscopy. *HolzRoh-Werkstoff.* 66: 323-332.
- Ferreira, M.M.C. (2015) *Quimiometria – Conceitos, Métodos e Aplicações*, 1ª ed., Editora Unicamp: Campinas.
- Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O.; *Quim. Nova.* 1999, 22, 724.
- Grogan, J., P. Barreto and A. Veríssimo (2002). "Mahogany in the Brazilian Amazon: Ecology and Perspectives on Management". Imazon, Belém.
- Hein, P.R.G., Lima, J. T., Chaix, G. (2010) Effects of sample preparation on NIR spectroscopic estimation of chemical properties of *Eucalyptus urophylla* S. T. Blake wood. *Holzforschung* 64:45-54.

- Hoöltken, A.M., Schröder, H., Wischmewki, N., Degen, B., Magel, E., Fladung, M. (2012) Development of DNA-based methods to identify CITES-protected timber species: a case study in the Meliaceae family. *Holzforschung* 66: 97-104.
- Hung, K.-H., Lin, C.-H. Ju, L.-P. (2017) Tracking the geographical origin of timber by DNA fingerprinting: a study of the endangered species *Cinnamomum kanehirae* in Taiwan. *Holzforschung*, 71(11):853-862.
- Jackson, J.E., Mudholkar, G.S. (1979) Control procedures for residuals associated with principal component analysis. *Technometrics* 21, 341-349.
- Jiang, Z.H., Huang A.M., Fei, B.H., Rei, H.Q. (2006) The analysis of wood microfibril angle by near infrared spectroscopy and X-ray diffractometry. *Spectroc. Spectr. Anal.* 26: 1230-1233.
- Jiao, L., Yin, Y., Cheng, Y., Jiang, X. (2014) DNA barcoding for identification of the endangered species *Aquilaria sinensis*: comparison of data from heated or aged wood samples. *Holzforschung*, 68(4):487-494.
- Jolliffe, I.T., (2002) *Principal Components Analysis*, 2nd Edition, Springer, New York, USA.
- Kelley, S.S., Rials, T.G., Snell, R., Groom, L.H., Sluiter, A. (2004) Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood. *Wood Sci. Technol.* 38: 257-276.
- Haider, K., Lim, S., Flaig, W.; *Holzforschung* 18, 18-88 (1964).
- Kennard, R.W.; Stone, L.A. (1969), *Computer aided design of experiments*, *Technometrics*, 11, 137-148.
- Lister, S.J., Dhanoa, M.S., Stewart, J.L., Gill, M. (2000) Classification and comparison of *Gliricidia* provenances using near infrared reflectance spectroscopy. *Anim. Feed Sci. Technol.* 86: 221-238.
- Lopes, R. E. C.; (2015) *Discriminação de madeiras similares por NIRS e PLS-DA considerando variações de temperatura e umidade*. Dissertação de Mestrado. Universidade de Brasília.
- Lowe, A.J., Dormontt, E.E., Bowie M., Degen, B., Gardner, S., Thomas, D., Clarke, C., Rimbawanto, A., Wiedenhoeft A., Yin. S., Sasaki, N. (2016) Opportunities for improved transparency in the timber trade through scientific verification. *BioScience*. 66: 990-998.
- Lu, W.Z., Yuan, H.F., Xu, G.T., Qiang, D.M. (2007) *Modern Near Infrared Spectroscopy Analytical Technology*, 2nd Edn. Chinese Petrochemical Press, Beijing, China, p.77.

- Martins, A.R., Talhavini M., Vieira, M.L., Zacca, J.J., Braga J.W.B. (2017) Discrimination of whisky brands and counterfeit identification by UV–Vis spectroscopy and multivariate data analysis. *Food Chem.* 229:142-151.
- Meder, R., Brawner, J. T., Downes, G. M., Ebson, N. (2011) Towards the in-forest assessment of Kraft pulp yield: comparing the performance of laboratory and hand-held instruments and their value in screening trials. *J. Near Infrared Spectrosc.* 19: 421-429.
- Michael H.A., Schröder, H., Wischnewski, N., Degen, B., Magel, E., Fladung, M. (2012) Development of DNA-based methods to identify CITES-protected timber species: a case study in the Meliaceae family. *Holzforschung*, 66(1):97-104.
- Nisgoski, S., Schardosin, F.Z., Batista, F.R.R., Muniz, G.I. de B., Carneiro, M.E. (2016) Potential use of NIR spectroscopy to identify *Cryptomeria japonica* varieties from southern Brazil. *Wood Sci. Technol.* 50:71-80.
- Norgaard, L., Saudland, A., Wagner, J., Nielsen, J.P., Munck, L., Engelsen, S.B. (2000) Interval partial least squares regression (iPLS): A comparative chemometric study with an example from near infrared spectroscopy. *Appl. Spectrosc.* 54: 413-419.
- Nuroniah, H.S., Gailing, O., Finkeldey, R. (2016) Development of a diagnostic DNA marker for the geographic origin of *Shorea leprosula*. *Holzforschung*. 71: 1-10.
- Oliveri, P., López, M.I., Casolino, M.C., Ruisánchez, I., Callao, M.P., Medini, L., Lanteri, S. (2014) Partial least square density modeling (PLS-DM) – A new class-modeling strategy applied to the authentication of olives in brine by near-infrared spectroscopy. *Anal. Chim. Acta.* 851: 30-36.
- Oliveri, P., Downey, G. (2012) Multivariate class modeling for the verification of food-authenticity claims. *Trends Anal. Chem.* 35: 74-86.
- Pasquini, C. (2003) Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. *J. Braz. Chem. Soc.* 14: 198-219.
- Pastore, T.C.M., Braga, J.W.B., Coradin, V.T.R., Magalhães, W.L.E, Okino, E.Y.A., Camargo, J.A.A., Muñoz, G.I.B., Bressan, O.A., Davrieux, F. (2011) Near infrared spectroscopy (NIRS) as a potential tool for monitoring trade of similar woods: Discrimination of true mahogany, cedar, andiroba, and curupixá. *Holzforschung*. 65: 73-80.
- Pastore, T. C. M.; Tese de Doutorado, Universidade de Brasília, Brasília, 2004.

- Rana, R., Müller, G., Naumann, A., Polle, A. (2008) FTIR spectroscopy in combination with principal component analysis or cluster analysis as a tool to distinguish beech (*Fagus sylvatica* L.) trees grown at different sites. *Holzforschung*, 62(5):530-538.
- Ravidran, P., Costa, A., Soares, R., Wiedenhoft, A. C. (2018) Classification of CITES-listed and other neotropical Meliaceae wood images using convolutional neural networks. *Plant Methods* 14:25.
- Reich, G. (2005) Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. *Advanced Drug Delivery Reviews*. 57: 1109 – 1143.
- Ribeiro, L.P.D., Silva, A.P.M. da, Lima, A.A. de, Silva, E. de O., Rinnan, A., Pasquini, C. (2016) Non-destructive determination of quality traits of cashew apples (*Anacardium occidentale* L.) using portable near infrared spectrophotometer. *J. Near Infrared Spectrosc.* 24: 77-82.
- Rodionova, O.Y., Titova, A.V., Pomerantsev, A.L. (2016) Discriminant analysis is an inappropriate method of authentication. *Trends Anal. Chem.* 78: 17-22.
- Rutiaga-Quiñones, J. G.; *Chemische und biologische Untersuchungen zum Verhalten dauerhafter Holzarten und ihrer Extrakte gegenüber holzabbauenden Pilzen*, Buchverlag Gräfelfing:München, 2001.
- Salazar, R., 1998: *Swietenia humilis* Zucc. Nota Técnica sobre manejo de semillas forestales No 35. CATIE
- Santos, C., Lopo, M., Páscoa, R.N.M., Lopes, J.A.A. (2013) Review on the applications of portable near-infrared spectrometers in agro-food industry. *Applied Spectroscopy*. 67: 1215-1233.
- Sandak, A., Sandak, J., Negri M. (2011) Relationship between near-infrared (NIR) spectra and the geographic provenance of timber. *Wood Sci. Technol.* 45:35-48.
- Sandak, A., Sandak, J., Prądzń W., Zborowska, M., Negri, M. (2009) Near infrared spectroscopy as a tool for characterization of wood surface. *Folia For. Pol.* 40: 31-40.
- Sandak, J., Sandak, A., Cantini C., Autino, A., Negri, M. (2015) Differences in wood properties of *Picea abies* L. Karst in relation to site of provenance and population genetics. *Holzforschung* 69: 385-397.
- Sandermann, W. (1962) Biosynthetische Untersuchungen an verschiedenen Kiefernarten. *Holzforschung* 16(3):65-74.

- Schwanninger, M., Rodrigues, J.C., Fackler K. (2011) A review of band assignments in near infrared spectra of wood and wood components. *J. Near Infrared Spectrosc.* 19: 287-308.
- Shou, G., Zhang, W., Gu, Y., Chao, D. (2014) Application of near infrared spectroscopy for discrimination of similar rare woods in the Chinese market. *J. Near Infrared Spectrosc.* 22: 423-432.
- Siau, J. F.; Transport processes in wood, Springer-Verlag: New York, 1984
- Silva, A.R. da; Pastore, T.M.C., Braga, J.W.B., Davrieux, F., Okino, E.Y.A., Camargos, J.A.A., Coradin, V.T.R., Prado, A.G.S. do (2013) Assessment of total phenols and extractives of mahogany wood by near infrared spectroscopy (NIRS). *Holzforschung* 67:1-8.
- Silva, D.C.; Pastore, T. M.; Soares, L.F.; Barros, F.A.S.; Bergo, M.C.J.; Coradin, V.T.H.; Gontijo, A.B.; Sosa, M.H.; Chacón, C.B.; Braga, J.W.B. (2018) Determination of the country of origin of true mahogany (*Swietenia macrophylla* King) woods in five South Latin American countries by handheld NIR devices and multivariate data analysis. *Holzforschung*.
- Silva, V.A.G. da, Talhavini, M., Zacca, J.J., Maldaner, A.O., Peixoto, I.C.F., Braga, J.W.B. (2014) Non-destructive identification of different types and brands of blue pen inks in cursive handwriting by visible spectroscopy and PLS-DA for forensic analysis. *Microchem. J.* 116: 235-243.
- Silva, V.A.G. da, Talhavini, M., Zacca, J.J., Trindade, B.R., Braga, J.W.B. (2014) Discrimination of black pen inks on writing documents using visible reflectance spectroscopy and PLS-DA. *J. Braz. Chem. Soc.* 25: 1552-1564
- Snee R.D (1977), Validation of regression models: methods and examples, *Technometrics*, 19, 415-428.
- Soares, L., Silva, D.C. da, Bergo, M.C.J., Coradin, V.T.R., Braga, J.W.B., Pastore, T.C.M. (2017) Avaliação de espectrômetro NIR portátil e PLS-DA para discriminação de seis espécies similares de madeira amazônica. *Química Nova.* 40: 418-426.
- Sousa, M.H., Megliano, M.M., Camargos, J.A.A., Sousa, M.R. (1997) Madeiras tropicais brasileiras. Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis. Laboratório de Produtos Florestais, Brasília.
- Sousa, L. C; Gomide, J. L.; Milagres, F. R.; Almeida, D. P. (2011) Desenvolvimento de modelos de calibração NIRS para minimização das análises de madeira de *Eucalyptus* spp. *Ciência Florestal*, v.21, n.3, p.591-599.

- Stern, W.L. (1988) Index Xylariorum 3. Institutional wood collections of the world. IAWA Bull. 9: 209–210.
- Teófilo, R.F., Martins, J.P.A., Ferreira, M.M.C. (2009) Support vector regression for functional data in multivariate calibration problems. J. Chemom. 23: 32-48.
- Tibola, C. S.; Medeiros, E. P.; Simeone, M. L. F.; Oliveira, M. A.; (2018) Espectroscopia no infravermelho próximo para avaliar indicadores de qualidade tecnológica e contaminantes em grãos. Embrapa.
- Tsuchikawa, S. (2007) A review of recent near infrared research for wood and paper. Appl. Spectrosc. Rev. 42: 43-71.
- Tsuchikawa, S., Inoue, K., Norma, J., Hhayashi, K. (2003) Application of near infrared spectroscopy to wood discrimination, J. Wood Sci. 49:29-35.
- Tsuchikawa, S., Kobori, H. (2015) A review of recent application of near infrared spectroscopy to wood science and technology. J. Wood Sci. 61:213-220.
- Tsuchikawa, S., Schwanninger, M. (2013) A review of recent near-infrared research for wood and paper (Part 2). Applied Spectroscopy Reviews. 48: 560-587.
- United Nations Office on Drugs and Crime, Global Programme for Combating Wildlife and Forest Crime, Best Practice Guide for Forensic Timber Identification, UNODC, New York, 2016.
- Wang, S.-N., Zhang, F.-D., Huang, A.-M., Zhou, Q. (2017) Distinction of four *Dalbergia* species by FTIR, 2nd derivative IR, and 2D-IR spectroscopy of their ethanol-benzene extractives. Holzforschung, 70(6):503-510.
- Westad F.; Marini F.(2015) Validation of chemometrics models – a tutorial. Anal Chim Acta. 893: 14-24.
- Yang, Z., Jiang, Z.H., Fei, B.H., Liu, J.L. (2005) Application of near infrared (NIR) spectroscopy to wood science. Scientia Silvae Sinicae. 41: 29-36.
- Yang, Z.; Zhang, M.; Li, K.; Chen, L. (2016), Rapid detection of knot defects on wood surface by Near Infrared Spectroscopy coupled with Partial Least Square Discriminant Analysis, BioResource, 11 (1), 2557-2567.
- Yu, M., Liu, K., Zhou, L., Zhao, L., Liu, S. (2016) Testing three proposed DNA barcodes for the wood identification of *Dalbergia odorifera*. Holzforschung 70: 127–136.