



MARIANA RAMOS DE ALMEIDA

**ESPECTROSCOPIA RAMAN E QUIMIOMETRIA COMO FERRAMENTAS
ANALÍTICAS PARA QUÍMICA FORENSE E PALEONTOLOGIA**

CAMPINAS

2015



**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE QUÍMICA**

MARIANA RAMOS DE ALMEIDA

**ESPECTROSCOPIA RAMAN E QUIMIOMETRIA COMO FERRAMENTAS
ANALÍTICAS PARA QUÍMICA FORENSE E PALEONTOLOGIA**

ORIENTADOR: PROF. DR. RONEI JESUS POPPI

**TESE DE DOUTORADO APRESENTADA AO
INSTITUTO DE QUÍMICA DA UNICAMP PARA
OBTENÇÃO DO TÍTULO DE DOUTORA EM CIÊNCIAS.**

**ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA
POR MARIANA RAMOS DE ALMEIDA, E ORIENTADA PELO PROF.DR. RONEI JESUS POPPI.**

Assinatura do Orientador

CAMPINAS

2015

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Química
Simone Lucas Gonçalves de Oliveira - CRB 8/8144

AL64e Almeida, Mariana Ramos de, 1985-
Espectroscopia Raman e quimiometria como ferramentas analíticas para química forense e paleontologia / Mariana Ramos de Almeida. – Campinas, SP : [s.n.], 2015.

Orientador: Ronei Jesus Poppi.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Química.

1. Espectroscopia Raman. 2. Espectroscopia Raman de imagem. 3. Classificação multivariada. 4. Resolução multivariada de curvas. I. Poppi, Ronei Jesus. II. Universidade Estadual de Campinas. Instituto de Química. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Raman spectroscopy and chemometrics as analytical tools for forensic chemistry and paleontology

Palavras-chave em inglês:

Raman spectroscopy
Raman spectroscopic imaging
Multivariate classification
Multivariate curve resolution

Área de concentração: Química Analítica

Titulação: Doutora em Ciências

Banca examinadora:

Ronei Jesus Poppi [Orientador]
Luiz Fernando Cappa de Oliveira
Maria Fernanda Pimentel Avelar
Márcia Cristina Breitzkreitz
Italo Odoni Mazali

Data de defesa: 28-01-2015

Programa de Pós-Graduação: Química

Dedico esta tese ao meu irmão Deives Ramos de Almeida, pela sua atenção e cuidado com nossa família e por todo incentivo e compreensão nas minhas escolhas.

“... Não queira apressar as coisas, tudo tem seu tempo, seu momento e hora... Vá sem pressa, andando, plantando e sentindo o cheiro da terra, o vento no rosto e apreciando a paisagem. O que planta com pressa perde o privilégio de coisas simples, e acaba perdendo a essência que teve ao dar o primeiro passo...”

Yla Fernandes

AGRADECIMENTOS

Ao prof. Dr. Ronei Jesus Poppi pela orientação, confiança, oportunidade de trabalho, paciência, pelo conhecimento compartilhado e pela convivência agradável durante esses anos;

À Universidade Estadual de Campinas e ao Instituto de Química pela disponibilização de toda infraestrutura para realização deste trabalho;

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, pela bolsa concedida;

Aos colegas do Laboratório de Quimiometria em Química Analítica, LAQQA (2011-2015): Márcia Breitz, Luciana Terra, Luciana Oliveira, Monica Lopez, Débora Forchetti, Marina De Gea, Thiago Ávila, Guilherme Alexandrino, Paulo Filgueiras, Guilherme Sabin, André Marcelo, Carlos Teixeira, Carlos Diego, José Augusto, pelo convívio e momentos agradáveis compartilhados;

Ao técnico do laboratório Humberto Machado pela sua disposição em ajudar nas mais diversas situações;

Ao perito criminal e amigo Deleon Nascimento Correa pelas amostras nos trabalhos de forense, pelas discussões e trocas de ideias;

Ao perito criminal Dr. Jorge J. Zacca pela colaboração no trabalho dos explosivos;

Ao prof. Dr. Lauro E. S. Barata e ao pesquisador Dr. Carlos Fidélis pela colaboração no trabalho dos óleos essenciais;

Ao pesquisador Dr. José Xavier Neto e a *Msc.* Lara Maldanis pela colaboração no trabalhos dos fósseis;

Aos professores Dr. Fabio Augusto e Dr. Dosil Pereira de Jesus pelo compartilhamento da experiência docente durante a realização do Programa de Estágio Docente (PED);

À minha família pela compreensão das minhas escolhas, da minha ausência e pelo apoio incondicional;

Agradeço aos meus padrinhos Cândida e Fausto e aos meus primos Fausto e Guilherme por todo o apoio até aqui, por me incentivarem sempre;

À amiga Sara Cristina de Souza pela convivência agradável, compreensão nos momentos de tensão, companheirismo e amizade.

Ao Sr. Sebastião Souza e Sra. Helena Souza pela amizade e pelo carinho;

Aos amigos Felipe Sampaio, Juliana Souza, Odirley Topan, Bruna Nardy, Cristiane Grossi e Fernando Ferrarezzi pelos bons momentos compartilhados;

A todos que de diferentes maneiras tornaram minha trajetória muito mais agradável.

CURRICULUM VITAE

Dados pessoais

Mariana Ramos de Almeida

E-mail: mariramosalmeida@gmail.com

Nome em citações bibliográficas: ALMEIDA, M. R.; Almeida, Mariana R.

Link para o Lattes: <http://lattes.cnpq.br/6690913086860156>

Formação Acadêmica

2011-2015 Doutorado em Ciências (Área de concentração: Química Analítica)

Universidade Estadual de Campinas, UNICAMP

2009-2011 Mestrado em Química (Área de concentração: Físico-Química)

Universidade Federal de Juiz de Fora, UFJF.

2004-2008 Graduação em Química – Bacharel e Licenciatura

Universidade Federal de Juiz de Fora, UFJF.

Produção científica no doutorado

Artigos científicos:

1. Mariana R. Almeida, Deleon N. Correa, Jorge J. Zacca, Lucio P. L. Logrado, Ronei J. Poppi. Detection of explosives on the surface of banknotes by Raman hyperspectral imaging and independent component analysis. *Analytica Chimica Acta*, 860, 15-22, 2015.

2. Mariana R. Almeida, Deleon N. Correa, Werickson F.C. Rocha, Francisco J.O. Scafi, Ronei J. Poppi. Discrimination between authentic and counterfeit banknotes using Raman spectroscopy and PLS-DA with uncertainty estimation. *Microchemical Journal*, 109, 170-177, 2013.

3. Mariana R. Almeida, Carlos H.V. Fidelis, Lauro E.S. Barata, Ronei J. Poppi. Classification of Amazonian rosewood essential oil by Raman spectroscopy and PLS-DA with reliability estimation. *Talanta*, 117, 305-311, 2013.

Trabalhos científicos apresentados em congressos nacionais e internacionais (2011-2014)

1. Mariana R. Almeida; Deleon N. Correa, Jorge J. Zacca, Ronei J. Poppi. Multivariate resolution methods in Raman imaging applied to explosives detection. XIV Chemometrics in Analytical Chemistry Conference, 2014, Richmond, Virginia, EUA.

2. Mariana R. Almeida, Carlos A. Teixeira, Deleon N. Correa, Ronei J. Poppi. Determination of intersecting lines in questioned documents by surface-enhanced Raman spectroscopy imaging and MCR-ALS. XIV Chemometrics in Analytical Chemistry Conference, 2014, Richmond, Virginia, EUA.

3. Mariana R. Almeida, Deleon N. Correa, Jorge J. Zacca. Ronei J. Poppi. Metodologia para identificação de explosivos em superfície de cédulas por espectroscopia Raman e Quimiometria. 17^º Encontro Nacional de Química Analítica, 2013, Belo Horizonte, MG.

4. Mariana R. Almeida, Carlos H. V. Fidelis, Lauro E. S. Barata, Ronei J. Poppi. Classificação do óleo essencial extraído de diferentes partes da árvore amazônica Aniba Rosaedora empregando espectroscopia Raman e PLS-DA. 5^º Congresso Iberoamericano de Química Analítica – 2^º Congresso Uruguayo de Química Analítica, 2012, Montevideo, Uruguai.

5. Mariana R. Almeida, Deleon N. Correa, Werickson F. C. Rocha, Francisco J. O. Scafi, Ronei J. Poppi. Emprego da espectroscopia Raman e quimiometria na discriminação entre cédulas autênticas e falsas. 16^º Encontro Nacional de Química Analítica, 2011, Campos do Jordão, SP.

Participação no Programa de estágio docente (PED)

1) QA 416 - Química Analítica IV (PED C, 2013)

2) QA 682 - Química Analítica Instrumental II (PED B, 2014)

3) QA 282 - Química Analítica Clássica (PED B, 2014)

RESUMO

ESPECTROSCOPIA RAMAN E QUIMIOMETRIA COMO FERRAMENTAS ANALÍTICAS PARA QUÍMICA FORENSE E PALEONTOLOGIA

A motivação para o desenvolvimento dessa tese foi a busca por métodos de análise não destrutivos, com nenhum ou mínimo preparo de amostra e que permitam a obtenção do máximo de informação com a realização de uma única análise na área de forense e paleontologia. Em forense, a espectroscopia Raman e o método de classificação supervisionado PLS-DA (Análise Discriminante por Mínimos Quadrados Parciais) foram empregados para construir modelos de classificação. O primeiro modelo foi construído para discriminar cédulas autênticas de cédulas falsas. A análise foi baseada na caracterização das tintas usadas na confecção das cédulas. O segundo modelo de classificação foi construído para diferenciar o óleo essencial extraído de diferentes partes (caule, folhas e galhos) da árvore amazônica *Aniba Rosaeodora*. A confiabilidade dos modelos foi avaliada pelo cálculo do intervalo de confiança, que foram calculados usando a técnica de reamostragem *bootstrap*. Os resultados obtidos mostraram que os modelos de classificação podem ser usados como método complementar à inspeção forense clássica e método de triagem. O desempenho dos modelos de classificação foi avaliado pelo cálculo de sensibilidade, especificidade, eficiência e coeficiente de Mathew. A espectroscopia Raman de imagem e o método de análise de componentes independentes (ICA) foram empregados para a identificação de explosivos em superfícies de cédulas. O método ICA foi avaliado como método de resolução de curvas para extrair os perfis espectrais e as imagens Raman dos constituintes presentes nas superfícies analisadas. Os resultados obtidos mostraram que o método ICA é adequado para resolução de curvas, uma vez que alcançou desempenho equivalente ao método clássico MCR-ALS (Resolução Multivariada de Curvas com Mínimos Quadrados Alternados). O limite de detecção da metodologia apresentada foi de $50 \mu\text{g}\cdot\text{cm}^{-2}$ para o explosivo TNT. Por fim, a espectroscopia Raman de imagem foi empregada no estudo da composição química de fósseis de peixes, com o objetivo de obter informações sobre características biológicas. Os resultados mostraram informações sobre a composição química do fóssil estudado.

ABSTRACT

RAMAN SPECTROSCOPY AND CHEMOMETRICS AS ANALYTICAL TOOLS FOR FORENSIC CHEMISTRY AND PALEONTOLOGY

The motivation for the development of this thesis was to search for non-destructive testing methods, with none or minimal sample preparation and allowing them to obtain maximum information with the completion of a single analysis for forensic and paleontology. In the forensics, the Raman spectroscopy and the PLS-DA (discriminant analysis by Partial Least Squares) classification method were explored to build classification models. The first model was built to discriminate authentic and counterfeit banknotes. The analysis was based on the characterization of inks used in the confection of the banknotes. The second classification model was built to differentiate the essential oil extracted from different parts (wood, leaves and branches) of the Brazilian tree *Aniba rosaeodora*. The reliability of the models was evaluated by calculating the confidence interval, which was calculated using the bootstrap resampling technique. The results show that the classification models can be used as a complementary method to classical forensic inspection and a screening method. The performance of classification models was evaluated by calculating sensitivity, specificity, efficiency and Matthew coefficient. In a third application, Raman hyperspectral imaging and the independent component analysis (ICA) method were used for identification of explosives on the surfaces of banknotes. The ICA method was evaluated as curve resolution method to extract the Raman spectral profiles and the images of the constituents present in the analyzed surfaces. The results showed that the ICA method is appropriate for curves resolution, once achieved equivalent performance to the classical MCR-ALS (Multivariate Curve Resolution with Alternating Least Squares) method. The methodology presented limit of detection of $50 \mu\text{g}\cdot\text{cm}^{-2}$ for the explosive TNT. Finally, the Raman hyperspectral imaging was applied in paleontology to study the chemical composition of fish fossil with the aim of obtaining information on biological characteristics. The results showed information about the chemical composition of fossil studied.

SUMÁRIO

LISTA DE ABREVIATURAS	xxi
LISTA DE TABELAS	xxv
LISTA DE FIGURAS.....	xxvii
CAPÍTULO 1: Introdução geral.....	1
1.1 Motivação	3
1.2. Química Forense.....	5
1.3. Paleontologia.....	7
1.4. Apresentação da Tese	9
CAPÍTULO 2: Espectroscopia Raman.....	11
2.1. Efeito Raman	13
2.2. Espectroscopia Raman de imagem	18
2.3. Instrumentação.....	22
CAPÍTULO 3: Quimiometria	29
3.1. Introdução	31
3.2. Análise por Componentes Principais (PCA).....	32
3.3. Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA)	34
3.4. Intervalos de confiança	38
3.5. Avaliação do desempenho do modelo de classificação	41
3.6 Métodos de Resolução de Curvas	42
3.6.1. Resolução Multivariada de Curvas com Mínimos Quadrados Alternados (MCR-ALS)	43
3.6.2. Análise de Componentes Independentes (ICA)	46
3.6.3. MCR-BANDS	49
3.6.4. Determinação do número de componentes	49
3.6.5. Figuras de mérito.....	51
CAPÍTULO 4: Objetivos.....	53
4.1 Objetivos Gerais.....	55
4.2 Objetivos Específicos	55
CAPÍTULO 5: Aplicação da Espectroscopia Raman e modelos de classificação multivariados em química forense	57
5.1. Discriminação entre cédulas autênticas e falsificadas usando espectroscopia Raman e PLS-DA com estimativa da incerteza.....	59

5.1.1 Introdução.....	59
5.1.2 Parte Experimental.....	61
5.1.3. Tratamento dos dados.....	63
5.1.2. Resultados e Discussão.....	64
5.1.5. Conclusões.....	78
5.2 Classificação do óleo essencial extraído da árvore amazônica <i>Aniba Rosaedora</i> por espectroscopia Raman e PLS-DA com estimativa da confiabilidade.....	79
5.2.1 Introdução.....	79
5.2.2 Parte Experimental.....	82
5.2.3 Tratamento dos dados.....	82
5.2.4 Resultados e Discussões.....	83
5.2.5. Conclusões.....	92
CAPÍTULO 6: Aplicação da Espectroscopia Raman de imagem e métodos de resolução multivariada de curvas em estudos forenses e paleontológicos.....	95
6.1 Metodologia para detecção de explosivos em superfícies de cédulas por espectroscopia Raman de imagem e métodos de resolução de curvas.....	97
6.1.1 Introdução.....	97
6.1.2. Parte experimental.....	100
6.1.2.1 Amostras.....	100
6.1.2.2. Imagens Raman.....	102
6.1.2.3. Tratamento dos dados.....	103
6.1.2.4. Limite de Detecção.....	104
6.1.3 Resultados e Discussões.....	104
6.1.5 Conclusões.....	120
6.2. Aplicação da espectroscopia Raman de imagem e quimiometria em Paleontologia.....	122
6.2.1 Introdução.....	122
6.2.2 Parte Experimental.....	124
6.2.3. Tratamento dos dados.....	125
6.2.4. Resultados e Discussões.....	125
6.2.5. Conclusões.....	129
CAPÍTULO 7: Conclusões Gerais.....	131
CAPÍTULO 8: Referências.....	137

LISTA DE ABREVIATURAS

ANFO – do inglês, *ammonium nitrate/ fuel oil* (nitrato de amônio em óleo combustível)

AOTF – do inglês, *acousto optic tunable filter* (filtros ópticos acústicos sintonizáveis)

BSS - do inglês, *blind signal separation* (separação cega de sinais)

CCD – do inglês, *charge-coupled device* (dispositivo de carga acoplada)

CV – do inglês, *cross-validation* (validação cruzada)

EFA – do inglês, *evolving factor analysis* (análise de fatores evolucionários)

FAST – do inglês, *fiber array spectral translation* (translação de matriz de fibras espectrais)

FT – do inglês, *Fourier Transform* (Transformada de Fourier)

HMX – do inglês, *high melting point explosive* (ciclo-1,3,5,7-tetrametileno-2,4,6,8-tetranitramina)

IBAMA – Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis

ICA – do inglês, *Independent component analysis* (análise de componentes independentes)

iMCR – do inglês, *interval multivariate curve resolution* (Resolução Multivariada de Curvas por intervalos)

ISO – do inglês, *International Organization for Standardization* (Organização Internacional para Padronização)

LCTF – do inglês, *liquid crystal tunable filter* (Filtros de cristal líquido sintonizáveis)

LOF – do inglês, *lack of fit* (falta de ajuste)

MCR-ALS – do inglês, *multivariate curve resolution with alternating least-squares* (resolução multivariada de curvas com mínimos quadrados alternados)

MF – do inglês, *mean field* (campo médio)

MSEC – do inglês, *mean square error of calibration* (erro médio quadrático de calibração)

MSECV - do inglês, *mean square error of cross-validation* (erro médio quadrático de validação cruzada)

NASA – do inglês, *National Aeronautics and Space Administration* (Administração Nacional da Aeronáutica e do Espaço)

OSC – do inglês, *orthogonal signal correction* (correção ortogonal do sinal)

PCA – do inglês, *principal component analysis* (análise por componentes principais)

PC – do inglês, *principal componente* (componente principal)

PDF – do inglês, *pseudo-degree of freedom* (pseudo-graus de liberdade)

PERC – perclorato de amônio

PETN – do inglês, *pentaerythritol tetranitrate* (tetranitrato de pentaeritritol)

PLS – do inglês, *partial least square* (mínimos quadrados parciais)

PLS-DA – do inglês, *partial least square – discriminant analysis* (análise discriminante por mínimos quadrados parciais)

RDX – do inglês, *research department explosive* (ciclo-1,3,5-trimetileno-2,4,6-trinitramina)

RMSECV – do inglês, *root mean square error of cross-validation* (raiz quadrada do erro médio quadrático de validação cruzada)

RMSEP – do inglês, *root mean square error of prediction* (raiz quadrada do erro médio quadrático de previsão)

SIMCA – do inglês, *soft independent modeling of class analogy* (modelagem independente e flexível por analogia de classes)

SIMPLISMA – do inglês, *SIMPL*e-to-use *InteractiveS Mixture Analysis*

SNV – do inglês, *standart normal variate* (variação normal padrão)

SVD – do inglês, *singular values decomposition* (decomposição em valores singulares)

TNP - trinitrofenol

TNT – 2,4,6-trinitrotolueno

VIP – do inglês, *variable importance in the projection* (projeção da importância da variável)

VL – variável latente

ν - estiramento

δ - deformação

LISTA DE TABELAS

Tabela 1: Resultados obtidos pelos métodos MCR-ALS e MF-ICA	108
Tabela 2: Valores máximos e mínimos de f obtidos pelo MCR-BANDS para as soluções do MCR-ALS e MF-ICA.....	116
Tabela 3: Resultados obtidos pelo MF-ICA para as amostras forenses	117

LISTA DE FIGURAS

Figura 1: Formas de espalhamentos da radiação que ocorrem quando uma radiação monocromática interage com uma amostra.....	13
Figura 2: Ilustração esquemática mostrando os diagramas de energia envolvidos nos processos de espalhamentos Raman e Rayleigh.....	15
Figura 3: Arranjo tridimensional mostrando a organização dos dados obtidos pelo mapeamento Raman (<i>point-mapping</i>) sobre a área de uma amostra.....	19
Figura 4: Comparação entre as técnicas de imagem Raman: (a) e (b) mostram a aquisição de dados no modo <i>point</i> e <i>line mapping</i> , respectivamente; (c) e (d) mostram a aquisição dos dados no modo <i>imaging</i> , sendo (c) usando filtros de cristal líquido para separação do comprimento de onda e (d) com o uso de matrizes de fibra de translação espectral.....	20
Figura 5: Desdobramento do arranjo tridimensional para duas dimensões e esquema do tratamento de dados que pode ser empregado para obtenção da informação espectral e espacial dos dados obtidos com imagem Raman.	21
Figura 6: Diagrama do espectrômetro Raman dispersivo RamanStation, modelo 400F da Perkin Elmer.	24
Figura 7: Diagrama óptico do espectrômetro dispersivo RamanStation.	25
Figura 8: Esquema do espectrômetro echelle com o arranjo de dispositivos de cargas acopladas (CCD).	26
Figura 9: Opções de amostragem para obtenção dos espectros Raman: <i>single-point</i> (1); <i>super-point</i> (2); <i>point-mapping</i> (3); <i>line-mapping</i> (4); <i>depth-profile</i> (5).	27
Figura 10: Organização dos dados de espectroscopia Raman em uma matriz de dados para aplicação de ferramentas quimiométricas.	32
Figura 11: Decomposição da matriz X pela PCA.....	33
Figura 12: Análise de componentes principais aplicada à uma matriz X, gerando uma matriz de escores T, uma matriz de pesos P e uma matriz de resíduos E.	33
Figura 13: Decomposição das matrizes X e Y para construção do modelo PLS.....	35
Figura 14: Esquema do particionamento das amostras para validação cruzada.....	36
Figura 15: Esquema para construção de um modelo de classificação usando PLS-DA.....	37
Figura 16: Distribuição dos valores previstos de y para determinação do valor limite entre as classes seguindo o teorema de Bayes.	38
Figura 17: Representação esquemática do <i>bootstrap</i>	39
Figura 18: Métodos de resolução de curvas aplicados a dados de espectroscopia Raman de imagem.....	43
Figura 19: MCR com matriz aumentada pelas linhas (a) e pelas colunas (b).....	45
Figura 20: Esquema da decomposição da matriz X para aplicação do iMCR-ALS.....	46
Figura 21: Esquema mostrando a análise de fatores evolucionária para a escolha do número de componentes nos métodos de resolução de curvas.....	50
Figura 22: Estatística dos artigos publicados na última década na área de documentos questionados mostrando a tendência da utilização de métodos de separação e métodos espectrométricos.	59

Figura 23: Estatística de cédulas da primeira família do real retidas nos anos de 2010, 2011 e 2012.	60
Figura 24: Cédula ilustrativa de R\$ 50 com as regiões analisadas por espectroscopia Raman.	63
Figura 25: Espectros Raman das áreas analisadas da cédula autêntica de R\$ 50: (A) calcografia; (B) área laranja; (C) área vermelha.	64
Figura 26: Estrutura molecular da (a) ftalocianina e de pigmentos dos grupos (b) diazo e (c) azo.	65
Figura 27: Espectro Raman da região calcográfica da cédula autêntica (A) e das cédulas falsas impressas em impressoras a laser (B) e jato de tinta (C).	66
Figura 28: Gráfico dos escores: (a) área calcográfica; (b) área laranja e (c) área vermelha. Notas autênticas (*), notas falsas (▼) e notas digitalizadas e impressas no laboratório em papel alcalino (■).	68
Figura 29: (a) Gráfico dos <i>loadings</i> da PC1 (A), PC2 (B) e PC4 (C) para a região de calcografia; (b) Gráfico dos <i>loadings</i> da PC1 (A), PC2 (B) e PC4 (C) para a região laranja; (c) Gráfico dos <i>loadings</i> da PC1 (A), PC2 (B) e PC3 (C) para a região vermelha da cédula de R\$ 50.	69
Figura 30: (a) Resultados do modelo de classificação para a tinta calcográfica e intervalo de confiança para o conjunto de calibração: notas autênticas e notas falsas fabricadas no laboratório; (b) Conjunto de validação externa e intervalo de confiança: notas autênticas e notas falsas apreendidas.	72
Figura 31: (a) Resultados do modelo de classificação para região de coloração laranja e intervalo de confiança para o conjunto de calibração: notas autênticas e notas falsas impressas no laboratório; (b) Conjunto de validação externa e intervalo de confiança: notas autênticas e notas falsas apreendidas.	74
Figura 32: (a) Resultados do modelo de classificação para a tinta vermelha e intervalo de confiança para o conjunto de calibração: notas autênticas e notas falsas impressas no laboratório em papel alcalino; (b) Conjunto de validação externa e intervalo de confiança: notas autênticas e notas falsas apreendidas.	75
Figura 33: (a) Modelo de classificação e intervalo de confiança para o conjunto de calibração : notas impressas em impressoras laser e jato de tinta; (b) Conjunto de validação externa e intervalo de confiança: notas falsas apreendidas.	77
Figura 34: Estrutura molecular do linalol.	80
Figura 35: Espectro Raman do óleo essencial extraído do caule (-); das folhas (-) e dos ramos (-).....	84
Figura 36: Gráfico dos escores da PC1 x PC2 x PC3 dos óleos extraídos do caule (▼), das folhas (•) e dos ramos (■).	86
Figura 37: Resultado do modelo do PLS-DA para as amostras de calibração (a) e validação (b). As amostras acima da linha vermelha são classificadas como óleo extraído do caule (●) e abaixo como óleo extraído dos ramos e folhas (•).	88
Figura 38: Gráfico dos <i>loadings</i> da primeira, segunda, terceira e quarta variável latente versus as variáveis (bandas Raman) para o modelo PLS-DA.	91
Figura 39: Escores do VIP para o modelo PLS-DA.	92
Figura 40: Estatística de ataques a bancos no Brasil entre 2011 e 2014.	99
Figura 41: Estrutura química dos explosivos estudados.	100

Figura 42: Cédulas suspeitas de serem de explosão de caixa eletrônico.....	101
Figura 43: Explosão simulada de caixa eletrônico. Explosivo ANFO (a); Cédula de R\$ 50 antes da explosão (b); explosivo juntamente com o dispositivo de segurança usado para manchar as cédulas (c); caixa metálica simulando caixa eletrônico (d); cédula de R\$ 50 pós explosão.	102
Figura 44: Espectros Raman dos explosivos (a) TNT; (b) PETN; (c) TNP; (d) RDX; (e) HMX; (f) ANFO; (g) perclorato de amônio; (h) pólvora negra.	106
Figura 45: Comparação entre os espectros Raman dos explosivos puro (linha preta) com os espectros recuperados pelo MF-ICA (linha vermelha) para os explosivos (a) TNT; (b) PETN; (c) TNP; (d) RDX; (e) HMX; (f) ANFO; (g) PERC (h) pólvora negra.	110
Figura 46: Mapas de concentração obtidos para os explosivos (a) TNP; (b) HMX; (c) PERC; (d) PETN; (e) RDX; (f) TNT; (g) ANFO e (h) pólvora negra pelo método MF-ICA nas amostras preparadas em laboratório.	112
Figura 47: Esquema da decomposição da matriz para aplicação do iMCR-ALS e o resultado obtido pelo explosivo TNT.....	115
Figura 48: Mapas de concentração obtidos no cálculo do limite de detecção do explosivo TNT para as concentrações (a) 0,25 mg.cm ⁻² ; (b) 0,20 mg.cm ⁻² ; (c) 0,15 mg.cm ⁻² ; (d) 0,10 mg.cm ⁻² ; (e) 0,05 mg.cm ⁻²	117
Figura 49: Amostra suspeitas de ter explosivo, perfil espectral e imagem Raman recuperados pelo método MF-ICA de um dos componentes, mostrando a presença de pólvora e sua distribuição na cédula.	118
Figura 50: Área da cédula onde foi encontrado o explosivo (a); imagem Raman e perfil espectral recuperados pelo método MF-ICA do pigmento da cédula (b); imagem Raman, perfil espectral (linha vermelha) recuperados pelo método MF-ICA do explosivo ANFO e espectro Raman do explosivo ANFO (linha preta) (c); imagem Raman e perfil espectral recuperados pelo método MF-ICA do dispositivo de segurança (d).	120
Figura 51: Localização da Bacia Sedimentar do Araripe.	124
Figura 52: Conjunto de espectros Raman antes do pré-processamento dos dados.	126
Figura 53: Conjunto de espectros Raman após o pré-processamento dos dados.....	127
Figura 54: Áreas mapeadas do fóssil e os perfis espectrais recuperados pelo método MF-ICA e seus respectivos mapas de concentração. A1 corresponde a matriz na qual o fóssil está inserido; A2 é a região do fóssil que supostamente localizava o coração; A3 corresponde uma região que provavelmente localizava-se o aparelho digestivo; A4 é a coluna vertebral; A5 é a região dorsal.....	128

CAPÍTULO 1: Introdução geral

1.1 Motivação

Métodos de análise não destrutivos, com nenhum ou mínimo preparo da amostra, que sejam rápidos, de baixo custo e que permitam a obtenção do máximo de informação com a realização de uma única análise despertam interesse em vários campos de pesquisa. Uma das técnicas que preenche praticamente todos os requisitos citados acima é a espectroscopia Raman.

A espectroscopia Raman fornece informações em nível molecular sobre os compostos presentes em uma amostra. A técnica é baseada no espalhamento inelástico da radiação eletromagnética monocromática que interage com as moléculas, onde a radiação é espalhada com uma frequência diferente da incidida. Essa variação de frequência corresponde à diferença de energia entre dois estados vibracionais da molécula¹. Dessa forma, os espectros Raman apresentam bandas que são características de uma determinada ligação química da molécula e fornecem uma impressão digital da amostra, permitindo assim sua identificação.

Grande parte do sucesso da espectroscopia Raman na química analítica se deve a utilização de ferramentas quimiométricas. A quimiometria é definida como uma área que emprega o uso de ferramentas matemáticas e estatísticas para interpretação de dados químicos multivariados². O desenvolvimento dessa área se deu devido ao aumento na capacidade da instrumentação analítica, de processamento dos computadores e no aumento da habilidade de aquisição de dados. Uma das primeiras e principais aplicações da quimiometria foi na área de espectroscopia molecular. Com o uso das ferramentas quimiométricas foi possível analisar dados de maior complexidade, dados com baixa resolução do sinal analítico e análise simultânea de vários compostos. A combinação da espectroscopia Raman com a quimiometria permite a criação de metodologias mais rápidas, redução do uso de reagentes, menor custo de análise e simplificação no preparo das amostras³.

A utilização de métodos quimiométricos nos dados extraídos da espectroscopia Raman permite a construção de modelos de classificação, modelos de regressão e separação matemática de sinais. Os modelos de classificação são de grande importância para a área de química forense, onde informações como falso/verdadeiro

são requisitadas. No entanto, existe uma lacuna nos parâmetros de validação de modelos de classificação multivariados, uma vez que poucos trabalhos investigam metodologias para avaliar a confiabilidade dos resultados^{4,5,6}. Apesar de muitas publicações descreverem aplicações de espectroscopia Raman e quimiometria em dados forenses, essas análises ainda não são usadas em laboratórios de rotina. Os motivos que dificultam a implementação dessas análises são falta de validação dos modelos desenvolvidos e falta de conhecimento dos analistas em espectroscopia Raman e quimiometria.

Toda vez que um método de análise é desenvolvido, os resultados obtidos precisam ser validados, comparados e suas incertezas calculadas, esses parâmetros são uma expressão quantitativa da qualidade dos dados. Devido à importância da estimativa da incerteza em dados analíticos, múltiplas propostas têm sido publicadas na literatura para o cálculo de incerteza de dados provenientes de análise multivariada, como por exemplo: método de propagação de erro, métodos de reamostragem e adição de ruído⁷. O número de aplicações de métodos quimiométricos de classificação é vasto na literatura, contudo, a maioria dos trabalhos somente atribui o objeto à classe; poucos trabalhos avaliam a incerteza da classificação.

O uso de ferramentas de separação matemática de sinais tem sido de grande importância em dados de espectroscopia Raman, principalmente na área de imagem^{8,9,10}, onde é gerada uma grande quantidade de dados. A extração das informações que são relevantes tem sido possível pela aplicação de métodos de resolução multivariada de curvas¹¹.

Nesse contexto, a presente tese almeja mostrar estratégias que podem ser adotadas para a validação de modelos de classificação multivariados e demonstrar como métodos de resolução de curvas podem ser aplicados à espectroscopia Raman de imagem para resolver problemas forenses e paleontológicos.

1.2. Química Forense

Em 1921 foi publicado o primeiro livro texto de química forense, onde se definiu a química forense ou química legal como a ciência que utiliza o conhecimento químico aplicado à resolução de problemas legais ou judiciais, ou seja, a química exercida a serviço da lei¹². É importante ressaltar que a química forense não está relacionada apenas a análise química na investigação criminal, mas está relacionada à análise de qualquer material que pode dar origem a um processo legal.

Dessa forma, o desenvolvimento de novas metodologias para análises de drogas, fármacos, agrotóxicos, alimentos, tintas, documentos, bebidas, combustíveis, resíduos de disparo de arma de fogo, explosivos, crimes ambientais, com diferentes formas de apresentação das amostras, são atribuições da química forense. As análises são feitas a fim de identificar as substâncias presentes, sua quantidade e verificar a conformidade das amostras¹³.

As pesquisas no campo da química forense são atualmente temas de grande interesse para a comunidade científica. É uma linha de pesquisa em amplo desenvolvimento devido à necessidade constante de atualização. As técnicas empregadas na área de química forense são a espectrometria de massas, cromatografia líquida e gasosa, eletroforese capilar, espectrofotometria UV-Vis, fluorescência, espectroscopia no Infravermelho e Raman, espectrometria atômica (absorção e emissão), raios X e microscopia eletrônica¹⁴. Na escolha da técnica é importante considerar: o tipo de informação pretendida e custo de operação e manutenção. Dentro desse contexto, duas técnicas vêm ganhando destaque, a espectroscopia Raman, que fornece informação sobre as ligações químicas presentes e a fluorescência de raios X que propicia informação sobre elementos químicos contidos na amostra.

Os problemas da química forense envolvem geralmente separações e identificação de misturas complexas e de materiais coletados em locais de ocorrência policial que quase sempre são materiais impuros e de composição desconhecidas. As matrizes das amostras variam do previsível ao imprevisível. Outro ponto importante em química forense é a necessidade de manter a integridade e a segurança das amostras.

Dessa forma, os métodos analíticos empregados devem ser minimamente destrutíveis ou não-destrutíveis¹⁵.

Com base no exposto acima, a espectroscopia Raman reúne características que são exigidas na área forense: permite a identificação de substâncias químicas, orgânicas ou inorgânicas; pode ser aplicada a qualquer estado físico da matéria, em soluções, materiais opacos, transparentes ou translúcidos; o sinal Raman não é afetado pela presença de água; as amostras não precisam passar por pré-tratamento e possui disponibilidade de equipamentos portáteis.

Nas últimas décadas, a espectroscopia Raman tem sido amplamente empregada na área de química forense¹⁶⁻²⁴. Num primeiro momento os espectros Raman eram usados para complementar as informações fornecidas pelos espectros no infravermelho. Atualmente os espectros Raman são empregados para proposta de identificação de substâncias ilícitas²⁵; de explosivos^{16,17}; de fibras²⁶; de corantes e pigmentos para a área de documentoscopia^{19,24} e de obra de arte²⁷; para caracterização de fluídos biológicos²⁸; e para identificar medicamentos falsos²⁹. Com o emprego das ferramentas quimiométricas, novas formas de utilização da espectroscopia Raman como ferramenta analítica foram surgindo, principalmente na área de reconhecimento de padrões, onde se encontram a maior parte dos trabalhos.

De forma geral, as pesquisas em química forense abordam modificações nas técnicas e métodos existentes. Uma das principais etapas desse processo é a validação dos métodos para demonstrar a confiabilidade e atender às necessidades das comunidades científica e jurídica¹⁴. As normas de validação para métodos univariados são bem estabelecidas³⁰⁻³². No caso da validação de métodos multivariados, existe a norma E1655-00 da ASTM³³ para desenvolvimento e validação de modelos de calibração multivariada empregando espectroscopia no infravermelho. Para modelos de classificação multivariados, poucos trabalhos científicos investigam metodologias para avaliar a confiabilidade dos resultados^{4,5}. Na literatura existem alguns guias específicos, como por exemplo, para determinações qualitativas na indústria farmacêutica a partir de dados de espectroscopia no infravermelho próximo³⁴. Uma das propostas dessa tese é a construção de intervalos de confiança para modelos multivariados de classificação

com o objetivo de melhor definir os critérios de validação de métodos qualitativos multivariados.

1.3. Paleontologia

A paleontologia é a ciência que se dedica ao estudo de fósseis. O termo fóssil é usado para descrever os restos ou vestígios de plantas e/ou animais; o termo deriva do latim *fossilis* e significa algo desenterrado. O estudo de fósseis tem como objetivo estudar as mudanças na geografia da Terra, no clima, nos ecossistemas e de que modo o conjunto dessas alterações influenciou a evolução das formas de vida pré-históricas. O conhecimento da composição química de plantas e animais extintos pode fornecer informações para entender a decomposição dos fósseis e compreender processos que levam à alterações significativas de compostos endógenos³⁵⁻³⁷.

A análise química do fósseis pode explicar ainda como os sistemas ou organismos adaptaram e evoluíram ao longo do tempo, e prever como o sistema atual poderá evoluir. Através da análise química podem-se resolver assinaturas bioquímicas e ambientais de seres extintos. A análise química de fósseis pode mostrar as variáveis que controlam e influenciam a decomposição, sendo as amostras paleontológicas verdadeiras experiências geoquímicas³⁸.

A microscopia eletrônica tem sido a técnica padrão para a análise de materiais geológicos. Os avanços nessa técnica possibilitaram o estudo da morfologia em estruturas delicadas preservadas de fósseis³⁸. A microscopia confocal de varredura a laser tem possibilitado a análise *in situ* com resolução em microescala da morfologia tridimensional de fósseis³⁹, porém, a técnica não permite a identificação da composição química de fósseis.

Para análise química, a maioria das técnicas analíticas disponíveis para detectar, resolver, quantificar e mapear os compostos químicos de fósseis pode danificar a amostra. Quando se trata de fósseis, onde muitas vezes as amostras são raras, de acesso limitado ou de valor inestimável (ou a combinação dos três), a técnica analítica a ser utilizada deve levar em consideração essas características da amostra.

A espectroscopia Raman é uma técnica analítica consagrada em geoquímica para a identificação, estudo molecular e estrutural de minerais⁴⁰. Recentemente, a técnica foi introduzida na paleontologia, e se mostrou útil não só nos estudos da mineralogia dos fósseis, mas também para documentar a estrutura molecular e maturidade geoquímica do querogênio de espécies orgânicas⁴¹. O querogênio é resultante das transformações da matéria orgânica disseminada em sedimentos primitivos. O reconhecimento da estrutura e composição dos tecidos moles em fósseis está sendo inovado pela aplicação da espectroscopia Raman, que tem sido utilizada principalmente na identificação de matéria orgânica, principalmente, em ossos^{35,42,43}. O termo tecido mole ou não resistente refere-se a presença de tecidos conjuntivo, epitelial, muscular, vascular e nervoso⁴⁴.

Em aplicações geoquímicas, os espectros Raman pontuais da região selecionada são suficientes para fornecer as informações analíticas desejadas. Para a análise química de fósseis, a espectroscopia Raman de imagem tem se revelado particularmente útil, pois permite a avaliação de uma área maior da amostra. Schopf e colaboradores apresentam vários trabalhos aplicando espectroscopia Raman de imagem confocal bidimensional e tridimensional em amostras de fósseis^{39,41,45}. Na imagem Raman bidimensional, um grande número de espectros é adquirido sobre uma área definida da amostra para fornecer um mapa de distribuição dos componentes químicos da área. Na imagem Raman tridimensional, uma nova dimensão é adicionada, sendo adquiridos espectros Raman em diferentes profundidades. O emprego da espectroscopia Raman de imagem auxiliou na identificação de filamentos microbianos em fósseis de 3,4 bilhões de anos, sendo esses fósseis os mais antigos já registrados⁴⁵.

A espectroscopia Raman também tem sido proposta como ferramenta para busca por evidências de vida em Marte. A busca por compostos orgânicos de origem biológica é um dos principais objetivos dos programas de exploração astrobiológico da NASA (*National Aeronautics and Space Administration*)⁴⁶.

Os trabalhos da literatura na área de paleontologia mostram o tratamento de dados para obtenção das imagens Raman de forma univariada, onde uma banda Raman é utilizada para obter a distribuição de determinado constituinte na amostra. No

entanto, o uso de abordagens multivariadas podem fornecer mais informações sobre o conjunto de dados estudados. Nessa tese a espectroscopia Raman de imagem e ferramentas quimiométricas de resolução multivariada de curvas foram aplicadas para estudar a composição de fósseis, ampliando o conhecimento sobre novas ferramentas para o estudo químico de fósseis.

1.4. Apresentação da Tese

A apresentação da tese segue da seguinte forma: exposição dos aspectos fundamentais da espectroscopia Raman normal e de imagem, um breve histórico do instrumentos usados para obtenção do sinal Raman, assim como a descrição do espectrômetro Raman empregado na obtenção dos dados dessa tese. O capítulo seguinte introduz as ferramentas quimiométricas empregadas no desenvolvimento dessa tese, onde são discutidos os aspectos teóricos da análise discriminante por mínimos quadrados parciais (PLS-DA) para a construção de modelos de classificação e dos métodos de resolução multivariada de curvas para a construção das imagens Raman.

As aplicações foram divididas em duas partes. A primeira parte mostra o desenvolvimento de modelos de classificação multivariados empregando dados da espectroscopia Raman. O primeiro modelo de classificação foi construído para discriminar cédulas autênticas de cédulas falsas; O segundo modelo foi desenvolvido com o objetivo de identificar o óleo essencial extraído do caule da árvore amazônica *Aniba Rosaeodora Ducke*. A incerteza na classificação de cada amostra foi estimada empregando o método de reamostragem *bootstrap*.

A segunda parte mostra o uso dos métodos de resolução multivariada de curvas, ICA (análise de componentes independentes) e MCR-ALS (resolução multivariada de curvas por mínimos quadrados alternados) aplicados aos dados de espectroscopia Raman de imagem. Os estudos forenses foram realizados para detecção de explosivos em superfícies de cédulas. Os dois métodos de resolução de curvas foram comparados nesse estudo. Na área de paleontologia, a espectroscopia Raman de imagem e o

método ICA foram avaliadas como novas ferramentas para estudos de fósseis, a fim de obter informação sobre características biológicas.

Em cada aplicação é apresentada uma breve introdução, os objetivos específicos, a descrição do procedimento experimental, os resultados obtidos e as conclusões. Por fim, são apresentadas as conclusões gerais do trabalho.

CAPÍTULO 2: Espectroscopia Raman

2.1. Efeito Raman

Quando a radiação eletromagnética interage com os elétrons e núcleos dos átomos, a radiação pode ser absorvida, transmitida ou espalhada. Essas interações dão origem a uma série de técnicas espectroscópicas, que possibilitam obter informações sobre estrutura molecular, níveis de energia, ligações químicas, identificação e quantificação de elementos químicos e moléculas.

Quando a radiação é espalhada, a maior parte dos fótons são espalhados elasticamente, ou seja, a radiação espalhada tem a mesma energia que a radiação incidida. Esse tipo de espalhamento não traz nenhuma informação a respeito da estrutura e composição molecular e é chamado de espalhamento Rayleigh. No entanto, uma pequena fração da radiação pode ser espalhada inelasticamente, ou seja, com energia diferente da incidida. Quando a energia espalhada é menor que a incidida, o espalhamento é chamado de Stokes, quando a energia é maior, denomina-se espalhamento anti-Stokes. A Figura 1 ilustra o procedimento de incidência de fótons com uma radiação monocromática sobre a amostra e os diferentes tipos de espalhamentos que podem ocorrer.

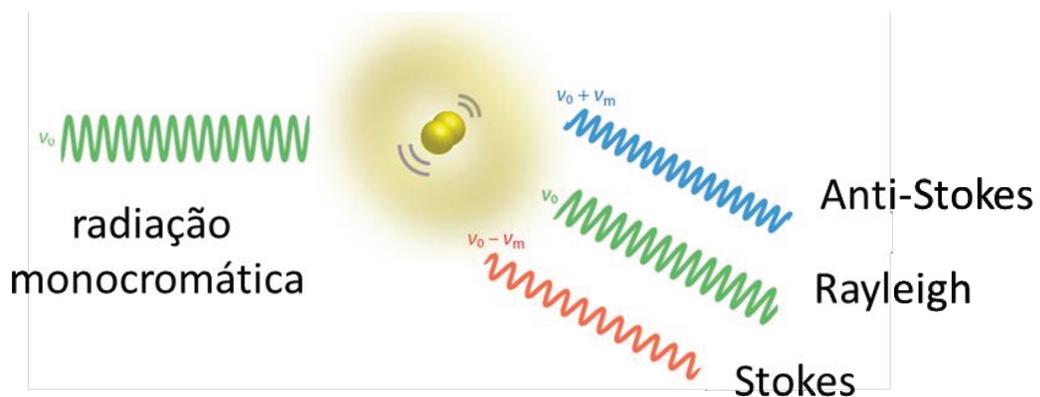


Figura 1: Formas de espalhamentos da radiação que ocorrem quando uma radiação monocromática interage com uma amostra, imagem adaptada da referência ⁴⁸.

O espalhamento inelástico é conhecido como efeito Raman e foi demonstrado experimentalmente pela primeira vez em 1928, pelo cientista indiano Sir Chandrasekhara Venkata Raman, após ter sido previsto por Smekal em 1923⁴⁷. A descoberta do efeito Compton em 1923, levou Raman a considerar a possibilidade de um fenômeno similar, o espalhamento inelástico da radiação para a região visível do

espectro eletromagnético. Raman imaginou que seria possível a radiação visível interagir com a matéria de modo que houvesse variação na energia do fóton incidente. Raman utilizou a luz solar, um espectrômetro de bolso e o olho humano como detector. A luz solar foi focalizada através de uma série de lentes em um recipiente contendo um líquido transparente purificado e um filtro azul foi colocado antes da amostra para deixar passar apenas a radiação de maior energia do espectro. Observando a amostra em uma direção perpendicular à direção de iluminação, verificou-se um traço luminoso devido ao espalhamento da radiação, demonstrando assim um novo tipo de fenômeno, o efeito Raman ⁴⁹. A importância da descoberta de Raman foi reconhecida em 1930 quando ele foi agraciado com o prêmio Nobel de Física.

O sinal Raman está relacionado a variação do momento de dipolo induzido na molécula pelo campo elétrico da radiação:

$$E = E_0 \cos 2\pi \nu_0 t \quad (1)$$

onde E_0 é a amplitude vibracional e ν_0 é a frequência da radiação incidente (frequência do laser). O campo elétrico oscilante induz na molécula um momento de dipolo oscilante P (equação 2), cuja frequência será a mesma do campo elétrico externo:

$$P = \alpha E = \alpha E_0 \cos 2\pi \nu_0 t \quad (2)$$

onde α é a constante de proporcionalidade chamada de polarizabilidade. A polarizabilidade pode ser entendida como uma medida da facilidade de deformação da nuvem eletrônica na presença de um campo elétrico.

Levando em conta a vibração molecular, os movimentos nucleares irão induzir flutuações na polarizabilidade com frequências que correspondem aos vários modos normais de vibração, ou seja, se a molécula vibra com frequência ν_m , o deslocamento do núcleo para uma molécula diatômica (q) pode ser descrito de acordo com a equação 3:

$$q = q_0 \cos 2\pi \nu_m t \quad (3)$$

Na equação 3 ν_m é a frequência vibracional da coordenada interna da ligação, q . A polarizabilidade pode ser considerada como uma função linear de q e expandida numa série:

$$\alpha = \alpha_0 + \left(\frac{\delta\alpha}{\delta q}\right) q + \dots \quad (4)$$

O primeiro termo da equação 4 é a polarizabilidade na posição de equilíbrio e o segundo termo é a variação da polarizabilidade com a coordenada q em relação à posição de equilíbrio.

Combinando a equação 2, 3 e 4 obtemos:

$$P = \alpha_0 E_0 \cos(2\pi \nu_0 t) + \frac{1}{2} \left(\frac{\delta \alpha}{\delta q} \right) q_0 E_0 \{ \cos[2\pi(\nu_0 + \nu_v)t] + \cos[2\pi(\nu_0 - \nu_v)t] \} \quad (5)$$

O primeiro termo da equação 5 contém somente a frequência da radiação incidente e corresponde ao espalhamento elástico Rayleigh, após a interação do fóton com a molécula, esta volta ao mesmo nível de energia inicial e o fóton é espalhado sem modificação de frequência. Somente no segundo termo aparece o espalhamento Raman com frequências $\nu_0 + \nu_v$ (espalhamento anti-Stokes) nesse caso o fóton encontra a molécula já num estado excitado e após a interação a molécula decai para o estado fundamental, sendo essa diferença de energia cedida ao fóton que é espalhado com energia $h\nu_0 + e_v$. No espalhamento Stokes de frequência $\nu_0 - \nu_v$, a molécula no estado fundamental sofre colisão com o fóton de energia $h\nu_0$, passa para um estado intermediário (virtual), pois não corresponde a nenhum estado estacionário da molécula. Em seguida a molécula decai para um estado vibracional excitado de energia e_v e assim o fóton espalhado $h\nu_0 - e_v$, tem energia menor que o fóton incidente. Os espalhamentos elástico e inelástico (Stokes e anti-Stokes) são mostrados esquematicamente na Figura 2.

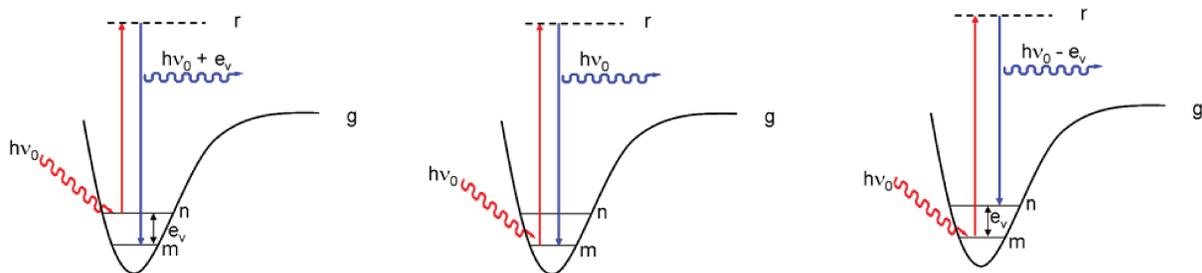


Figura 2: Ilustração esquemática mostrando os diagramas de energia envolvidos nos processos de espalhamentos Raman e Rayleigh. Onde $h\nu_0$ é a energia incidente, m e n são o primeiro e o segundo estado vibracional da ligação química, respectivamente; e_v é a diferença de energia entre os dois estados vibracionais m e n ; r representa um estado virtual.

A relação de intensidades entre as bandas Stokes e anti-Stokes pode ser mostrada pela função de distribuição de Boltzmann, a relação entre as intensidades anti-Stokes (I_{AS}) e Stokes (I_S) é dada pela equação 6⁴⁷. Uma vez que, a população do estado excitado que origina as bandas anti-Stokes é muito menor que a população do estado fundamental, as bandas Stokes são mais intensas que as bandas anti-Stokes. No entanto, as intensidades são comparáveis para frequências vibracionais de energia baixa, mas para frequências de maior energia é difícil observar as bandas anti-Stokes.

$$\frac{I_{AS}}{I_S} = \left(\frac{\nu_0 + \nu_v}{\nu_0 - \nu_v} \right) \exp\left(-\frac{e_v}{kT} \right). \quad (6)$$

Na transição entre os estados vibracionais m e n (Figura 2) devem ser consideradas as integrais das componentes $(\alpha_{ij})_{mn}$ em relação ao modo vibracional. Para que a transição vibracional Raman seja permitida, é necessário que uma das componentes do momento de transição (seis componentes no total) seja diferente de zero:

$$\left(\alpha_{ij} \right)_{mn} = \int \psi_m \alpha_{ij} \psi_n d\tau \quad (7)$$

Combinando a equação anterior com a equação 4, obtém-se:

$$\left(\alpha_{ij} \right)_{mn} = \left(\alpha_{ij} \right)_0 \int \psi_m \psi_n d\tau + \left(\frac{d\alpha_{ij}}{dq} \right)_0 \int \psi_m q \psi_n d\tau \quad (8)$$

Para ser observado o efeito de espalhamento Raman é necessário que pelo menos uma das integrais da equação 8 seja diferente de zero. O primeiro termo da equação é igual a zero, devido a ortogonalidade entre Ψ_m e Ψ_n . Para o segundo termo ser diferente de zero é necessário atender a duas condições. A primeira é:

$$\left(\frac{\delta\alpha}{\delta q} \right)_0 \neq 0 \quad (9)$$

ou seja, é necessário que ocorra uma variação da polarizabilidade com o pequeno deslocamento da coordenada q em relação a sua posição de equilíbrio. A polarizabilidade é uma propriedade tensorial (estabelece uma relação linear entre vetores P e E) que determina a atividade dos modos vibracionais na espectroscopia Raman.

A segunda condição é:

$$\int \psi_m q \psi_n d\tau \neq 0 \quad (10)$$

Nestas condições, o modo vibracional é ativo no espectro Raman⁴⁷.

O espectro Raman mostra um conjunto de transições vibracionais permitidas de uma molécula. O número de bandas, as frequências e a intensidade relativa dessas bandas estão associados com as vibrações das ligações químicas.

As principais restrições da espectroscopia Raman são a pequena secção de choque e a interferência causada pela fluorescência da amostra. A secção de choque é a probabilidade de interação entre duas partículas, nesse caso a molécula e o fóton, quanto maior a interação, maior a intensidade. O aumento da eficiência do espalhamento Raman e a supressão da fluorescência têm sido alcançados com o uso de outros mecanismos como o efeito Raman ressonante e o efeito de intensificação de espalhamento por meio de superfície (SERS, do inglês, *Surface Enhanced Raman Scattering*)⁵⁰.

No efeito Raman normal ou ordinário, a intensidade Raman depende apenas da polarizabilidade do estado eletrônico fundamental, nesse caso a energia do fóton de excitação é muito menor que a energia necessária para a transição eletrônica da molécula (Figura 2). Quando a energia do fóton de excitação é próxima da energia de uma transição eletrônica permitida, produz-se uma intensificação do sinal Raman da espécie química em questão. Esse efeito é conhecido como Raman ressonante.

Outro efeito de intensificação do sinal Raman é o efeito SERS, que apresenta grande interesse no estudo de espécies adsorvidas em superfícies metálicas, pois apresenta intensificação do sinal da ordem de 10^6 vezes. O efeito de intensificação causado pela interação da molécula com a superfície foi descrito pela primeira vez por Fleischmann em 1974⁵⁰. A observação de uma única molécula através do efeito SERS em 1997 foi um forte estímulo para as pesquisas nesta área, uma vez que demonstrou que a intensidade do sinal Raman pode competir com o sinal de fluorescência⁵¹. Nas últimas décadas tem-se observado um grande desenvolvimento e aplicações de SERS na área química, de materiais e principalmente nas ciências biológicas. As aplicações analíticas do efeito SERS dependem de superfícies que possam produzir espectros reprodutíveis e os estudos se concentram na busca de métodos de síntese de

nanopartículas com formas e tamanhos otimizados⁵². Um trabalho de revisão recente sobre o assunto pode ser encontrado na literatura⁵³.

A combinação do efeito SERS e do efeito Raman ressonante leva a uma intensificação ainda maior do sinal Raman, resultando no efeito SERRS (SERRS, do inglês, *Surface Enhanced Resonance Raman Scattering*). A intensificação ocorre quando o espectro SERS é excitado com uma radiação que coincide com a radiação da banda de absorção da molécula. A soma dos efeitos de intensificação das bandas Raman permite a obtenção de espectros em concentrações ainda menores.

2.2. Espectroscopia Raman de imagem

O acoplamento de um microscópio ótico ao espectrômetro Raman levou ao desenvolvimento da microscopia Raman na década de 90, que permitiu a exploração de duas características interessantes e úteis: a capacidade de produzir imagem e a confocalidade¹⁵. Com isso foi possível combinar a especificidade química dos espectros Raman com a possibilidade de investigação espacial, o que permitiu a análise e discriminação de componentes em amostras heterogêneas.

O uso de espectrômetros Raman confocais para aquisição de espectros permite que seja feita a rejeição da fluorescência da amostra, uma vez que as informações obtidas pelo detector são em grande parte da região onde o laser está focalizado na amostra. Isso é conseguido através de dois pequenos orifícios chamados de *pinholes* que rejeitam os sinais que estão fora de foco fazendo uma filtragem espacial. A confocalidade permite obter perfil de distribuição lateral e axial⁵⁴.

A espectroscopia Raman de imagem é uma combinação entre a espectroscopia Raman pontual com a tecnologia de imagem digital. Dessa forma é possível visualizar a composição química da amostra e a distribuição espacial dos constituintes simultaneamente. As imagens Raman podem ser obtidas nos modos, *mapping* e *imaging*⁵⁵.

No caso do modo *mapping*, uma plataforma móvel é usada para movimentar a amostra na direção x e y, o mapeamento pode ser *point-mapping*, onde uma área da

amostra é selecionada e um espectro é obtido em cada ponto xy da superfície amostrada, sendo esses pontos denominados pixels. No final, os espectros obtidos são agrupados na forma de um arranjo tridimensional, como mostrado na Figura 3, onde xy são as coordenadas do pixel e a terceira dimensão z corresponde aos números de onda do espectro. Outra forma de mapeamento é o *line-mapping*, nesse caso, o laser ilumina uma linha da superfície da amostra e os espectros são obtidos nesta linha. Apesar deste método ser mais rápido, problemas nas intensidades podem surgir devido à iluminação desigual da amostra⁵⁵. As Figuras 4a e 4b ilustram os procedimentos de mapeamentos citados acima.

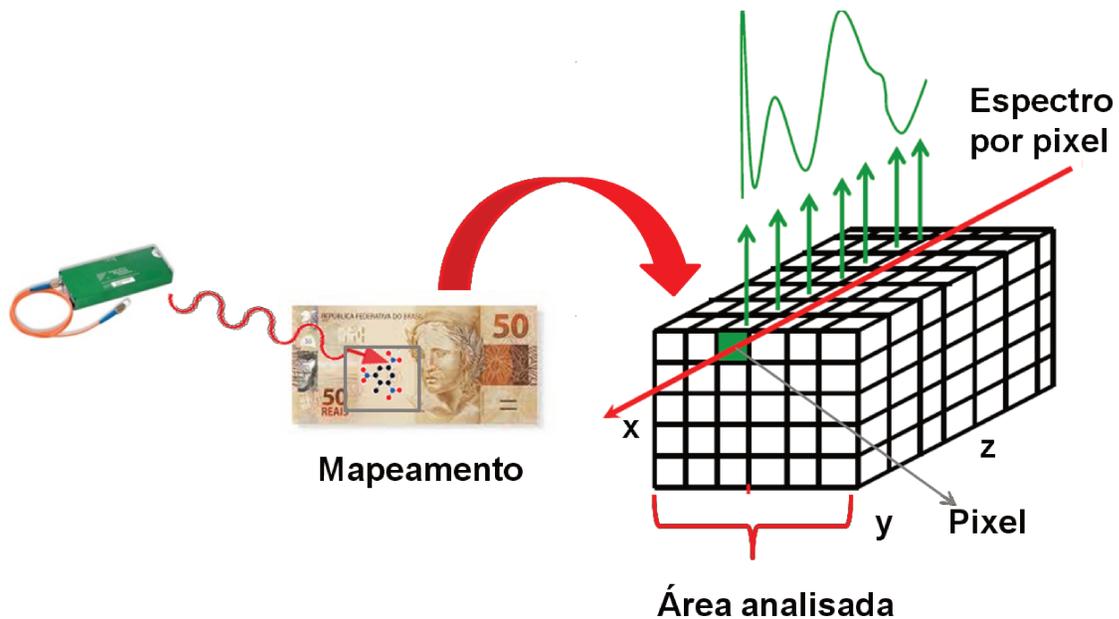


Figura 3: Arranjo tridimensional mostrando a organização dos dados obtidos pelo mapeamento Raman (*point-mapping*) sobre a área de uma amostra

No modo *imaging*, o laser ilumina toda a superfície da amostra e uma imagem é obtida por comprimento de onda, como mostrado na Figura 4c. A principal vantagem desse método é a rapidez para obtenção dos espectros. O modo *imaging* é adequado para aplicações nas quais componentes majoritários são analisados⁵⁴. A resolução das imagens depende das características do instrumento, como os filtros empregados para separação dos comprimentos de onda que podem ser filtros ópticos acústicos (AOTF) ou de cristal líquido (LCTF) e do número de pixels no detector. A resolução espacial das imagens é restringida pelo limite de difração de *Bragg* e depende do comprimento de

onda da radiação empregada, do índice de refração do meio no qual a luz se propaga e da lente usada na focalização da radiação na amostra. Uma abordagem mais recente do modo *imaging* são as matrizes de fibra ótica. Dentro dessa tecnologia destacam-se as matrizes de fibra de translação espectral (FAST)⁵⁶. Nessa estratégia de obtenção de imagem, duas dimensões espaciais e uma dimensão espectral de dados são obtidas ao mesmo tempo com um conjunto de fibras. A radiação espalhada é captada a partir de um campo de visão iluminado globalmente próximo a uma matriz bidimensional de fibras ópticas. A Figura 4d ilustra o procedimento de coleção de dados dessa técnica. A extremidade da matriz de fibra está em contato com uma matriz paralela linear que está inserida dentro de uma fenda na entrada de um espectrômetro dispersivo equipado com um detector de imagem CCD. Na aquisição de uma única imagem são obtidos informações espacial e espectral simultaneamente. A grande vantagem do uso do conjunto de fibras é a velocidade quando se compara com as outras estratégias de imagem Raman, a desvantagem é que a informação espacial é limitada. A Figura 4 mostra uma comparação entre as técnicas de imagem Raman discutidas.

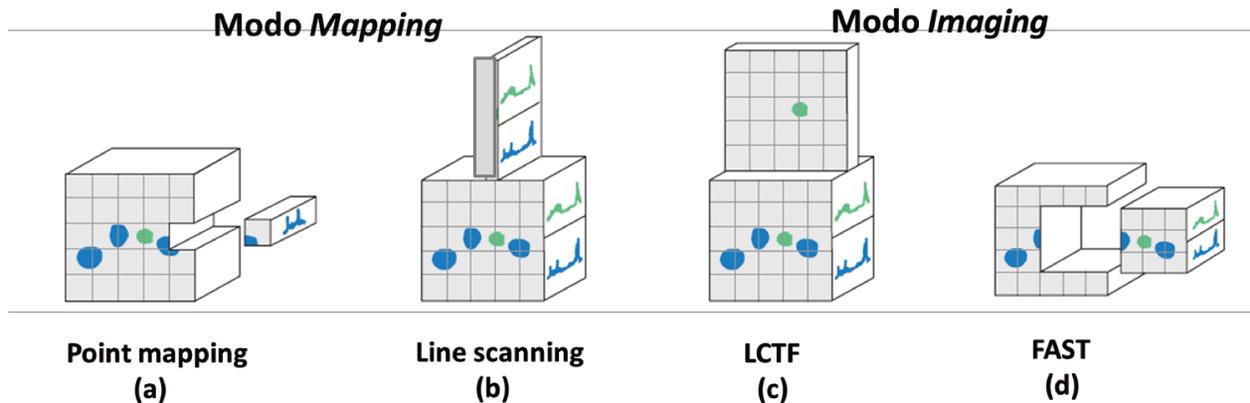


Figura 4: Comparação entre as técnicas de imagem Raman: (a) e (b) mostram a aquisição de dados no modo *point* e *line mapping*, respectivamente; (c) e (d) mostram a aquisição dos dados no modo *imaging*, sendo (c) usando filtros de cristal líquido para separação do comprimento de onda e (d) com o uso de matrizes de fibra de translação espectral. Figura retirada da literatura⁵⁶.

Tanto o procedimento *mapping* quanto o procedimento *imaging* geram imagens químicas, sendo os dados obtidos organizados em um arranjo tridimensional. Para extração das informações de interesse como, por exemplo, a concentração de um ou mais constituintes da amostra ou o perfil espectral do analito puro, o arranjo de dados

tridimensional é desdobrado para um arranjo bidimensional na forma de uma matriz de dados, de forma a não perder informação. A Figura 5 mostra o procedimento para processamento de dados de imagem Raman.

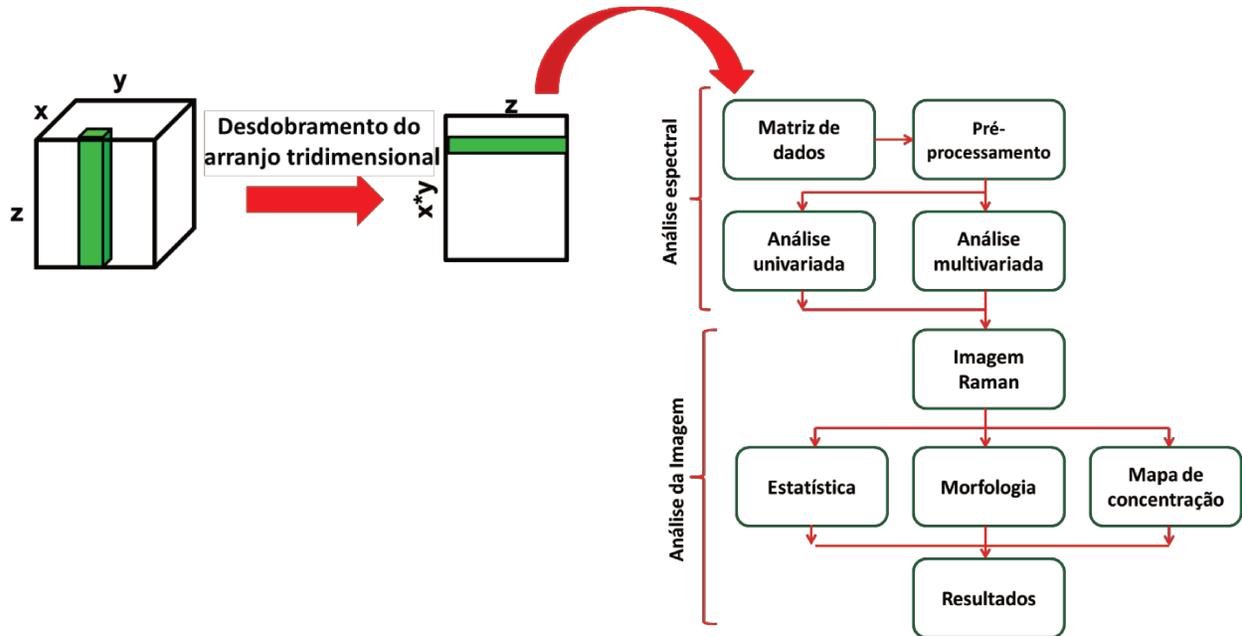


Figura 5: Desdobramento do arranjo tridimensional para duas dimensões e esquema do tratamento de dados que pode ser empregado para obtenção da informação espectral e espacial dos dados obtidos com imagem Raman. Figura adaptada da literatura⁵⁵

O primeiro passo para tratamento das imagens é o pré-processamento dos dados, que é realizado com o objetivo de remover artefatos que não fornecem informações sobre a composição da amostra. Existem vários métodos para pré-processamento^{57,58}, sendo que a escolha do mais adequado não é trivial e depende do objetivo da análise.

Após o pré-processamento dos dados, as informações espectrais e de concentração dos compostos que constituem uma amostra podem ser extraídas usando métodos uni ou multivariados. O método univariado é adequado quando existe uma banda vibracional característica para cada composto presente, do contrário, os métodos multivariados mostram-se mais apropriados, uma vez que permitem o uso de toda a faixa espectral para extrair informações. Os perfis espectrais recuperados fornecem informações sobre a composição química da amostra e a matriz com os perfis de concentração pode ser redobrada para retornar ao espaço bidimensional e fornecer a distribuição de cada componente na amostra analisada. O uso de ferramentas

quimiométricas em dados de imagem Raman é importante para obter o máximo da informação disponível.

2.3. Instrumentação

Para obtenção dos espectros Raman até 1950 eram utilizadas lâmpadas de vapor de mercúrio (435 nm) como fonte de radiação excitante. No entanto, essas lâmpadas emitiam uma série de linhas indesejadas, fazendo-se necessário o uso de filtros para isolar as linhas de interesse, entretanto, vários compostos absorviam essas linhas ou eram decompostos fotoquimicamente, impedindo a obtenção dos espectros Raman de alguns compostos. Os espectrógrafos empregavam prismas como elemento dispersor e os espectros eram registrados em chapas fotográficas⁴⁷.

A radiação laser como fonte de excitação na espectroscopia Raman foi introduzida na década de 60. D.L. Wood e o brasileiro Sérgio Porto utilizaram pela primeira vez a radiação de um laser pulsado de rubi para obter o efeito Raman⁵⁹. Mais tarde Porto e Kogelnik registraram um espectro Raman utilizando laser contínuo de He-Ne com excitação em 632 nm. O uso de laser como fonte de excitação para espectroscopia Raman trouxe novas perspectivas para a técnica, dando início a uma nova fase.

No início da década de 70, os lasers de Ar e Kr estavam disponíveis em espectrômetros Raman comerciais que possuíam duplo ou triplo monocromador e detecção por fotomultiplicadoras⁴⁷. Nos anos 80, o uso da radiação no infravermelho próximo como fonte de excitação e técnicas de transformada de Fourier possibilitaram a eliminação da fluorescência. O emprego de detectores multicanal de alto desempenho conduziu a um espetacular aumento na sensibilidade da técnica. Essas melhoras na instrumentação levaram a universalização da espectroscopia Raman que passou a ser aplicada a sistemas biológicos, biomédicos, arqueológicos, obras de arte e outros.

Um espectrômetro Raman registra a radiação que é inelasticamente espalhada em função da sua energia e mostra cada componente espectral e sua respectiva intensidade em um gráfico chamado de espectro Raman. A radiação espalhada na amostra contém vários comprimentos de onda e necessita ser discriminada para ser

apresentada como bandas. As componentes espectrais podem ser separadas através de redes de difração e prismas que compõe os equipamentos Raman dispersivos. Esse tipo de espectrômetro opera com laser na região do visível e início do infravermelho próximo (785 nm) e utilizam detectores multicanal, como o CCD, esse tipo de detector registra uma janela do espectro por vez e é constituído por uma matriz de fotodiodo, onde cada fotodiodo é denominado de pixel. Outro modo de discriminação da radiação espalhada é através dos equipamentos interferométricos, onde se cria um padrão de interferência gerando um interferograma durante o movimento dos espelhos. Nesse caso, a radiação é refletida em um espelho fixo e em outro móvel. O espectro é obtido através da aplicação da transformada de Fourier no interferograma. Os equipamentos com excitação no infravermelho próximo são interferométricos. O uso de laser na região do infravermelho próximo evita o problema da fluorescência, no entanto, como a intensidade da radiação espalhada é proporcional a quarta potência da frequência da radiação incidida, o efeito Raman é fraco nessa região por isso é necessário o uso de um interferômetro, onde é possível somar os espectros e melhorar a sua qualidade. Os detectores usados em equipamentos FT-Raman são geralmente fotodiodos semicondutores como Germânio resfriado com nitrogênio líquido.

Como visto, os avanços na instrumentação facilitaram a tarefa de obter um espectro Raman. Os lasers como fonte de radiação monocromática apresentam atributos que os tornam uma excelente fonte de excitação, entre os quais podemos destacar monocromaticidade, direcionalidade e coerência. Os novos tipos de detectores, como os detectores multicanais (CCD), o uso de microscópio óptico acoplado ao espectrômetro, a utilização de fibra óptica para medidas remotas, o desenvolvimento da espectroscopia FT-Raman e a minituarização da instrumentação têm tornado a espectroscopia Raman uma técnica cada vez mais popular para a caracterização rápida e fácil de amostras, abrindo assim novas possibilidades de campos de pesquisa¹.

O instrumento empregado no desenvolvimento desta tese foi um espectrômetro dispersivo RamanStation, modelo 400F da Perkin Elmer. A Figura 6 mostra o esquema do espectrômetro RamanStation, composto por um laser de diodo que fornece radiação

em 785 nm, espelhos, filtros, lentes, espectrômetro echelle e um detector multicanal baseado em dispositivos de carga acoplada (CCD).

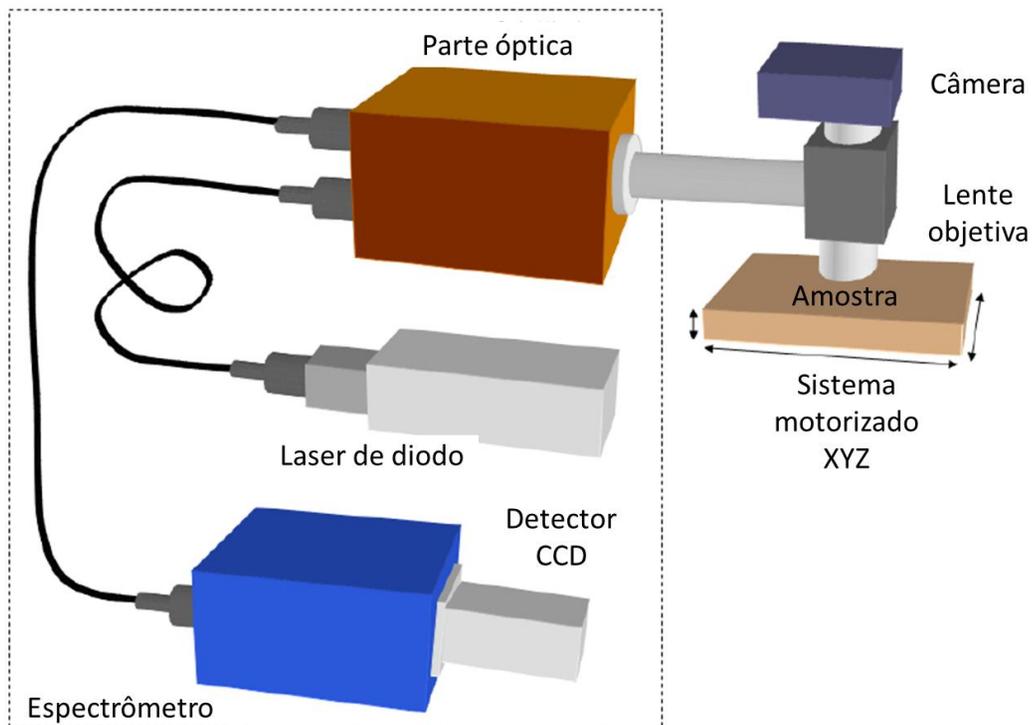


Figura 6: Diagrama do espectrômetro Raman dispersivo RamanStation, modelo 400F da Perkin Elmer.

Para obtenção dos espectros Raman a radiação do laser passa através de uma lente objetiva que focaliza a radiação na amostra. A radiação espalhada pela amostra é coletada pela mesma lente objetiva, passa por um filtro que elimina a radiação do espalhamento Rayleigh e é conduzida até o espectrômetro. A Figura 7 mostra o diagrama óptico do espectrômetro.

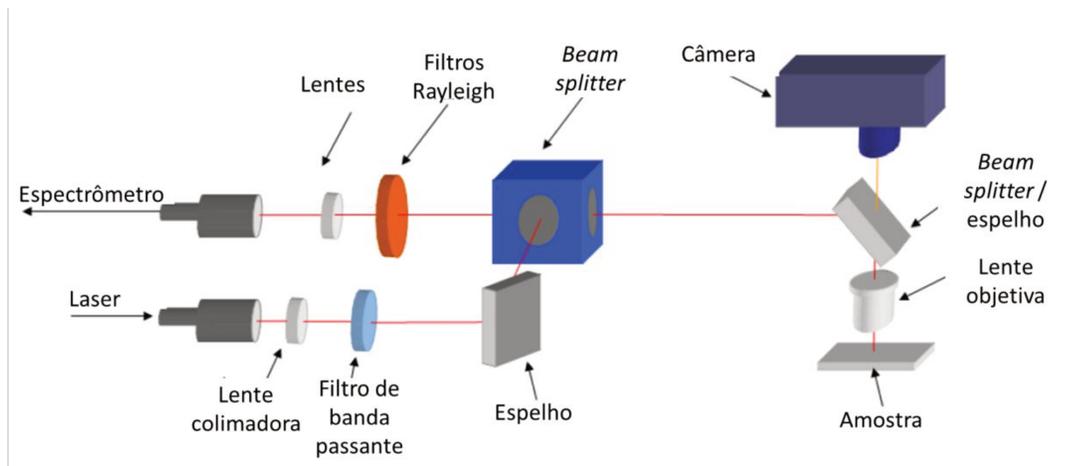


Figura 7: Diagrama óptico do espectrômetro dispersivo RamanStation.

Quando a radiação atinge o espectrômetro echelle, ela é dispersada por duas grades. A primeira grade dispersa a radiação horizontalmente em direção a segunda grade, onde a radiação é dispersada verticalmente para o detector CCD. A radiação é espalhada sobre um número de faixas no CCD que são lidas simultaneamente, dessa forma, a aquisição do espectro na faixa espectral de 200 a 3200 cm^{-1} é feita simultaneamente. A Figura 8 mostra a representação esquemática do espectrômetro echelle com o arranjo de dispositivos de cargas acopladas.

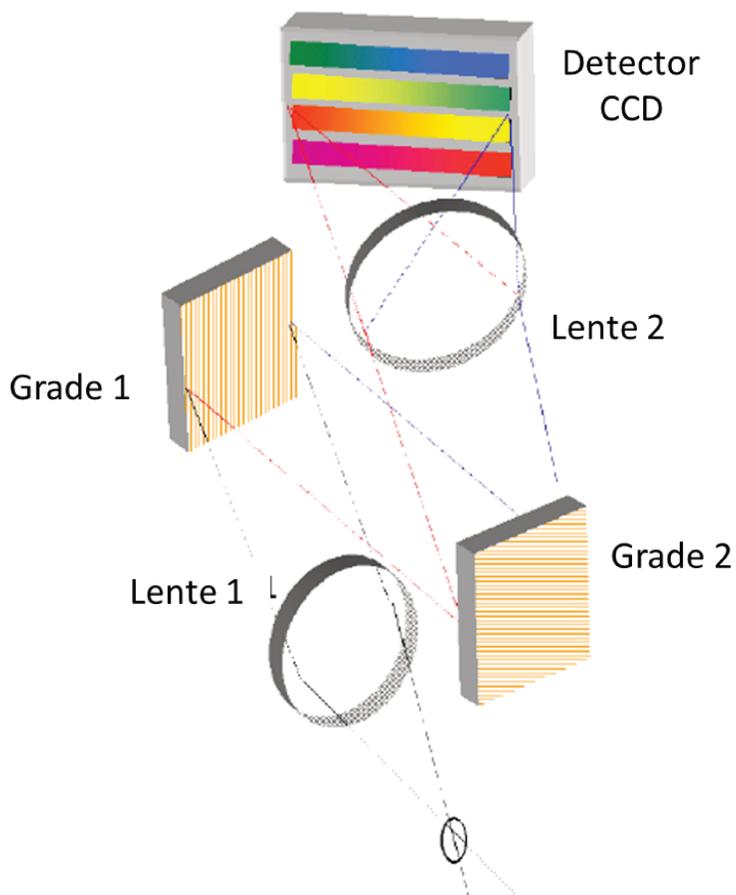


Figura 8: Esquema do espectrômetro echelle com o arranjo de dispositivos de cargas acopladas (CCD).

Nesta tese a amostragem das amostras para obtenção dos espectros Raman foi feita de duas formas: para obtenção dos espectros Raman pontuais empregou-se o modo *single-point* (Figura 9), onde uma lente objetiva focaliza a radiação laser na amostra, o tamanho do *spot* do laser na amostra é de $100\ \mu\text{m}$. Para a técnica de imagem Raman foi empregado o modo *point-mapping* (Figura 9), onde uma área xy da amostra é limitada e os espectros são obtidos pelo movimento do compartimento de amostra em passos micrométricos, de modo a cobrir a região da amostra selecionada. A Figura 9 ilustra os processos de amostragem que podem ser empregados com o equipamento RamanStation. Outros modos possíveis são o *super-point*, onde a área de iluminação da amostra pelo laser é de $600\ \mu\text{m}$, o *line-mapping* que foi discutido anteriormente e o modo *depth-profile*, onde os espectros são obtidos em diferentes distâncias entre o laser e a amostra.

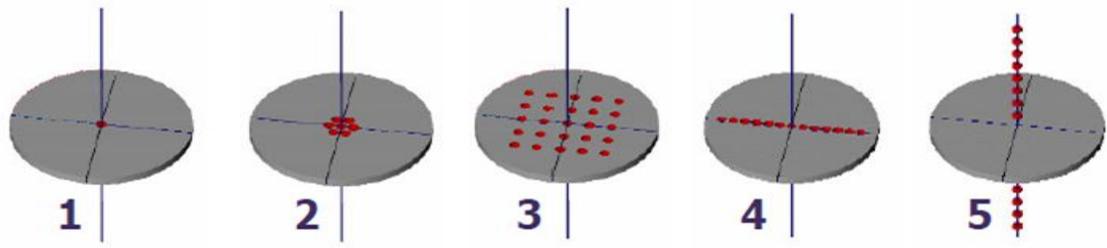


Figura 9: Opções de amostragem para obtenção dos espectros Raman: *single-point* (1); *super-point* (2); *point-mapping* (3); *line-mapping* (4); *depth-profile* (5).

CAPÍTULO 3: Quimiometria

3.1. Introdução

O uso de estatística multivariada permite que muitas variáveis sejam analisadas simultaneamente, o que possibilita extrair muito mais informações de um conjunto de dados. Em Química esse tipo de análise é conhecido como quimiometria e é definida como o uso de ferramentas estatísticas e matemáticas em dados químicos². Os métodos quimiométricos podem ser aplicados para planejamento e otimização de experimentos; para a extração e interpretação de dados químicos multivariados com o objetivo de reconhecimento de padrões, classificação, modelagem, processamento de imagens e outros.

Os métodos de reconhecimento de padrões permitem identificar similaridades e diferenças nas propriedades das amostras e/ou variáveis e classificá-las de acordo com tais características. Esses métodos são divididos em dois segmentos: os de treinamento não-supervisionado e os de treinamento supervisionado.

Os métodos de treinamento não-supervisionados não requerem nenhuma suposição inicial sobre a estrutura dos dados, o objetivo é encontrar agrupamentos naturais. A Análise por Componentes Principais (PCA, do inglês, *Principal Component Analysis*) é um dos métodos mais utilizados e difundidos para esse tipo de estudo⁶⁰.

Os métodos de treinamento supervisionados são aqueles que necessitam de alguma suposição inicial sobre o sistema em estudo e são empregados para prever se uma amostra pertence a uma determinada classe previamente estabelecida, como por exemplo: SIMCA (do inglês, *Soft Independent Modeling of Class Analogy*) e a Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA, do inglês, *Partial Least Squares Discriminant Analysis*)⁶¹.

Para utilização das ferramentas quimiométricas os dados químicos são arranjados na forma de tabelas ou matrizes de dados, como mostrado na Figura 10, onde as linhas representam as amostras e as colunas referem-se as variáveis (exemplo, intensidade Raman em diferentes números de onda). Durante uma análise química, uma grande quantidade de dados pode ser obtida, um espectro Raman coletado para 50 amostras na faixa de 3500 a 200 cm^{-1} com resolução de 4 cm^{-1} produz uma matriz de dados de dimensões $50 \times 825 = 41250$ números. Desta forma, se faz

necessária uma maneira de extrair informações importantes de matrizes de dados com tais dimensões.

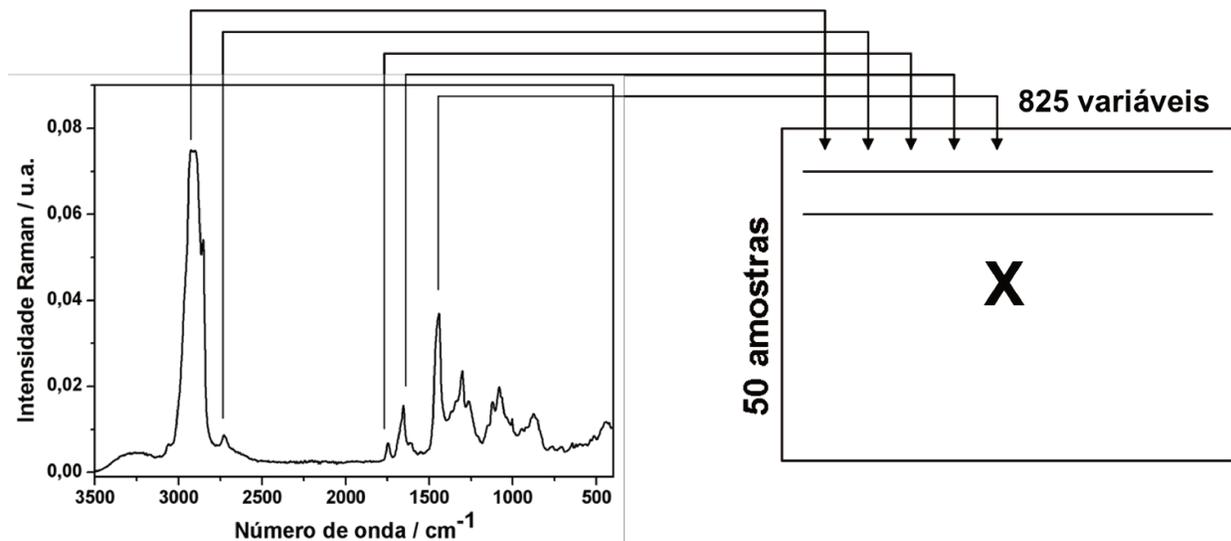


Figura 10: Organização dos dados de espectroscopia Raman em uma matriz de dados para aplicação de ferramentas quimiométricas.

Nessa tese são apresentados somente os métodos empregados para tratamento dos dados obtidos.

3.2. Análise por Componentes Principais (PCA)

A PCA é um dos métodos mais importantes na quimiometria, sendo a base para vários outros métodos. O objetivo da PCA é transformar grandes matrizes de dados em matrizes menores que podem ser facilmente interpretadas, permitindo extrair o máximo de informação. A PCA converte as informações obtidas em gráficos que mostram a relação entre as amostras e entre as variáveis do sistema em estudo. Geometricamente, a PCA pode ser definida como uma técnica de projeção, na qual uma matriz **X** é projetada num subespaço de dimensões reduzidas. A PCA faz uma decomposição da matriz de dados em termos da soma de várias matrizes de posto 1, na qual posto significa um número que expressa a dimensionalidade da matriz. A matriz

pode ser escrita como produto de dois vetores, escores \mathbf{t} e *loadings* \mathbf{p} para “ \mathbf{a} ” componentes principais.

$$\mathbf{X} = t_1\mathbf{p}'_1 + t_2\mathbf{p}'_2 + t_a\mathbf{p}'_a + \mathbf{E} \quad (11)$$

A Figura 11 mostra graficamente a decomposição da matriz de dados \mathbf{X} .

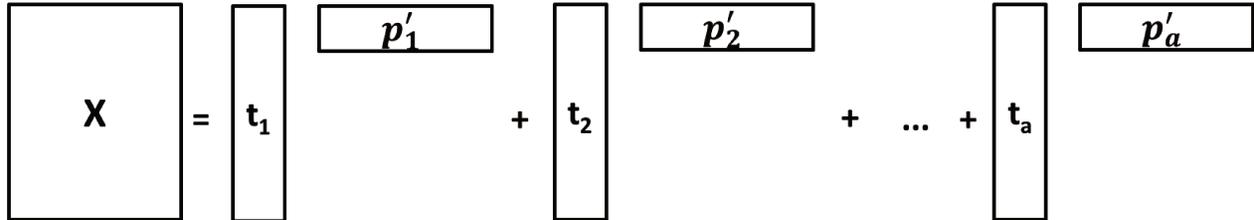


Figura 11: Decomposição da matriz \mathbf{X} pela PCA

A forma matricial de representar a equação anterior é mostrada na equação 12 e descrita esquematicamente na Figura 12:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} \quad (12)$$

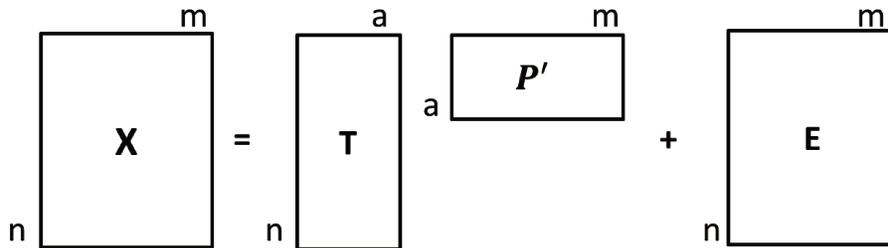


Figura 12: Análise de componentes principais aplicada à uma matriz \mathbf{X} , gerando uma matriz de escores \mathbf{T} , uma matriz de pesos \mathbf{P} e uma matriz de resíduos \mathbf{E} .

onde \mathbf{X} é a matriz original dos dados com n linhas (amostras) e m colunas (variáveis); \mathbf{T} é a matriz dos escores com n linhas e a colunas (número de componentes principais, novas coordenadas no novo sistema de eixos), \mathbf{P} é a matriz dos *loadings* (informação da contribuição de cada variável original na formação dos novos eixos) com a linhas e m colunas. E a matriz \mathbf{E} é a matriz de resíduos, contendo o que não foi modelado pela PCA.

A PCA tenta agrupar aquelas variáveis que estão altamente correlacionadas em uma nova variável chamada de componente principal. As componentes principais são uma combinação linear das variáveis originais. Usualmente a grande fração da

variância é descrita nos primeiros componentes principais, sendo possível visualizar os dados pelo gráfico dos escores de um componente contra o outro. O gráfico de escores mostra as similaridades e as diferenças entre as amostras e o gráfico dos *loadings* mostra a contribuição de cada variável original para o modelo das componentes principais.

A PCA pode ser usada como um método de pré-processamento, a fim de reduzir a dimensionalidade do conjunto de dados. Ao invés de usar todas as variáveis originais, o conjunto de dados pode ser representado pelas componentes principais, as quais são empregadas para muitas finalidades diferentes, como regressão linear, análise discriminante, redes neurais artificiais, análise de componentes independentes, resolução multivariada de curvas, e muitas outras⁶⁰.

A PCA é um método versátil que possibilita uma visão geral dos dados multivariados, revelando relações entre variáveis e as amostras, permitindo detecção de amostras anômalas, encontrando padrões e mostrando variáveis importantes de um processo.

3.3. Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA)

O PLS-DA é considerado um método supervisionado, no qual se deve ter um conhecimento inicial sobre as classes do conjunto de amostras⁶¹. As classes são definidas baseadas no conhecimento prévio do sistema ou através de uma análise exploratória, por exemplo, empregando a Análise de Componentes Principais.

O modelo de classificação PLS-DA pode ser construído a partir de algoritmos da regressão multivariada por Mínimos Quadrados Parciais (PLS). O PLS é um tipo de calibração inversa, onde se busca uma relação direta entre a resposta instrumental e a propriedade de interesse (qualitativa ou quantitativa). A PCA é aplicada nas duas matrizes de dados, matriz $\mathbf{X}_{(I \times J)}$ (resposta instrumental) e na matriz ou vetor $\mathbf{Y}_{(I \times K)}$ (propriedade de interesse) com o objetivo de reduzir a dimensão dos dados:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} \quad (13)$$

$$\mathbf{Y} = \mathbf{UQ}' + \mathbf{F} \quad (14)$$

onde **T** e **U** são as matrizes dos escores, **P** e **Q** são as matrizes dos *loadings*; **E** e **F** são os resíduos de **X** e **Y**, respectivamente. A Figura 13 mostra graficamente a decomposição das matrizes **X** e **Y**.

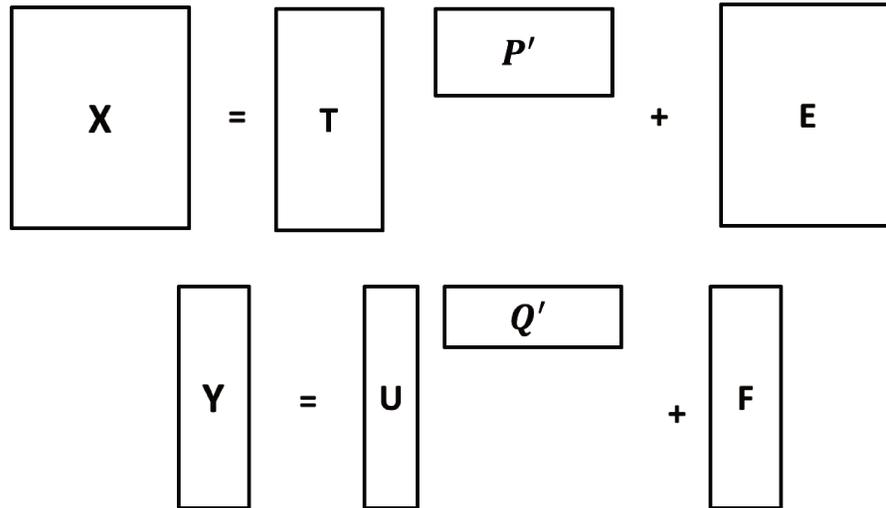


Figura 13: Decomposição das matrizes **X** e **Y** para construção do modelo PLS.

No PLS existe um compromisso entre a explicação da variância em **X** e sua correlação com **Y**, dessa forma ocorre uma leve rotação no eixo das componentes principais para aumentar a correlação, após a rotação, as componentes principais passam a ser nomeadas de variáveis latentes. Os escores **T** são ortogonais e são estimados como uma combinação linear entre as variáveis originais **X** com os coeficientes pesos **W*** (equação 15).

$$T = XW^* \quad (15)$$

Os escores **T** são bons previsores de **Y** e assume-se que **Y** e **X** são modelados pelas mesmas variáveis latentes.

$$Y = TQ' + F \quad (16)$$

Os resíduos de **Y**, **F**, expressam os desvios entre a resposta observada e a modelada. As equações 15 e 16 podem ser reescritas como:

$$Y = XW^*Q' + F = X\beta + F \quad (17)$$

Onde:

$$\beta = W^* Q' \quad (18)$$

$$W^* = W(P'W)^{-1} \quad (19)$$

$$W = \frac{X'U}{U'U} \quad (20)$$

W é definido em termos de um conjunto de peso dos *loadings* que maximiza a covariância descrita entre X e Y ⁶². Maiores detalhes da regressão PLS são dados por Wold e *et. al*⁶³.

Um passo fundamental para a construção de um modelo PLS-DA é a determinação do número correto de variáveis latentes. A escolha do número de variáveis latentes é comumente feita através de validação cruzada, onde uma parte das amostras (ou apenas uma) do conjunto de dados é utilizada para validação. A Figura 14 ilustra o processo da escolha do número de variáveis latentes: separa as amostras em conjunto de calibração e validação e constrói-se o modelo PLS com as amostras de calibração. Testa as amostras que foram separadas (conjunto de validação) e estima-se os erros de previsão empregando-se diferentes números de variáveis latentes. Repete-se o processo até que todas as amostras tenham sido previstas, ou seja, tenham participado do conjunto de validação. O número de variáveis latentes escolhido é o que fornece um menor erro de previsão, por exemplo, na Figura 14 seriam escolhidas 3 variáveis latentes. Se for utilizado menos que 3 o modelo pode apresentar falta de ajuste e mais que 3 variáveis latentes o modelo será sobre ajustado.

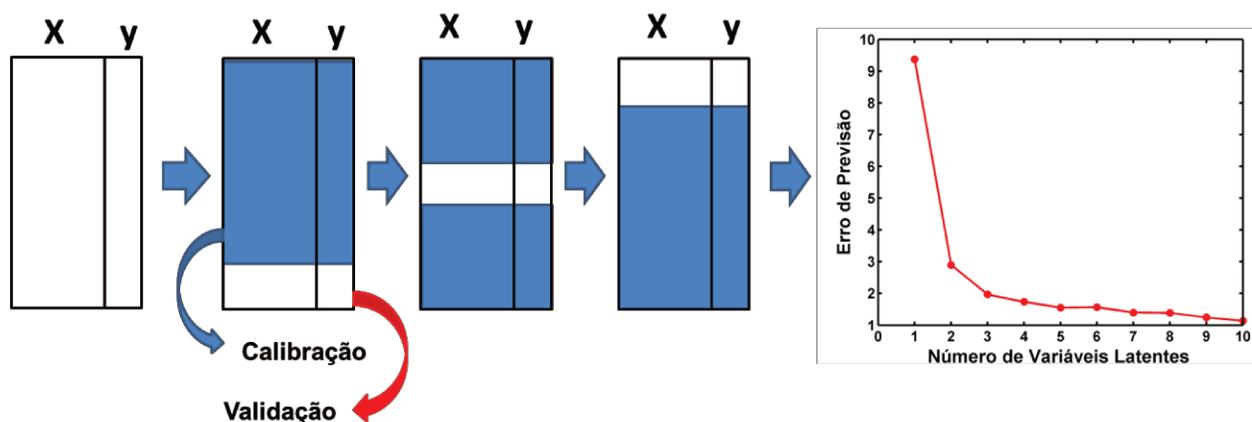


Figura 14: Esquema do particionamento das amostras para validação cruzada.

Como o PLS-DA é um método de classificação, a matriz ou vetor Y da propriedade de interesse será codificada por 0 e 1 quando se tem duas classes ($Y=2$).

Para $Y > 2$, pode-se construir vários modelos com a codificação 0 e 1, ou utilizar o algoritmo PLS2⁶⁴, construindo uma matriz ($I \times C$), onde I representa o número de amostras e C a classe de cada amostra representada por 0 ou 1. A Figura 15 ilustra o procedimento para PLS-DA para dados de espectroscopia Raman. A matriz X é composta pelos espectros Raman, onde as linhas representam as amostras e as colunas as intensidades Raman; o vetor y é construído com valores de 1 ou 0, sendo 1 para as amostras que pertencem a uma classe que deseja classificar e 0 para as demais amostras.

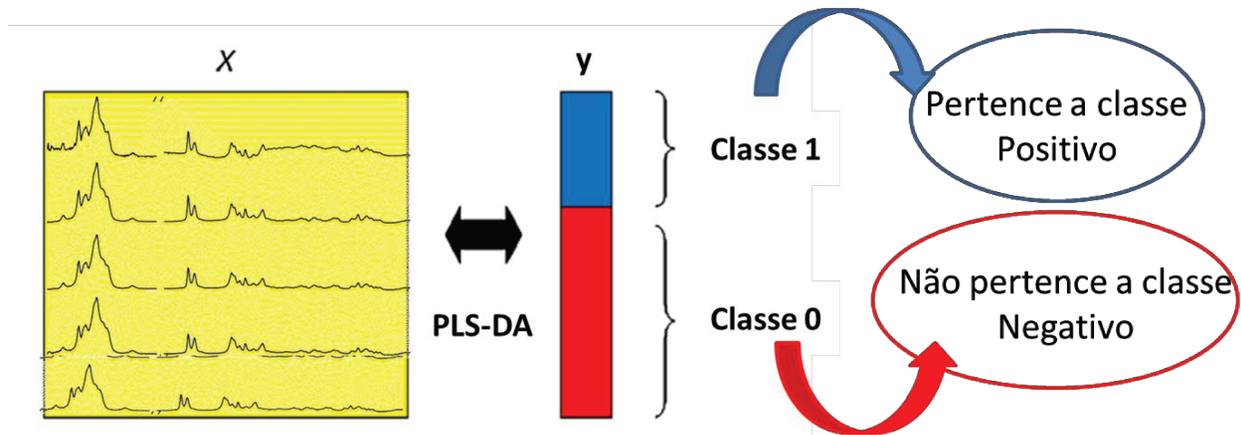


Figura 15: Esquema para construção de um modelo de classificação usando PLS-DA. A matriz X contém as informações químicas e o vetor y é construído de acordo com a classe que pertence cada amostra. Para um modelo de classificação binário podem ser atribuídos valores de 1 para as amostras que pertencem a determinada classe e 0 para as amostras que não pertencem a classe avaliada.

O valor obtido, y previsto, pelo modelo PLS-DA será um número dado pela equação 17, não exatamente 0 ou 1. Dessa forma é necessário estabelecer valores limites entre as classes, esses valores são estimados empregando o Teorema de Bayes⁶⁵. Nesse caso, considera-se que os valores de y previstos para cada classe seguem uma distribuição normal. O valor limite entre as classes é selecionado no ponto em que as duas distribuições se cruzam, como mostrado na Figura 16, esse valor é estabelecido de forma a minimizar o número de resultados falsos positivos e falsos negativos.

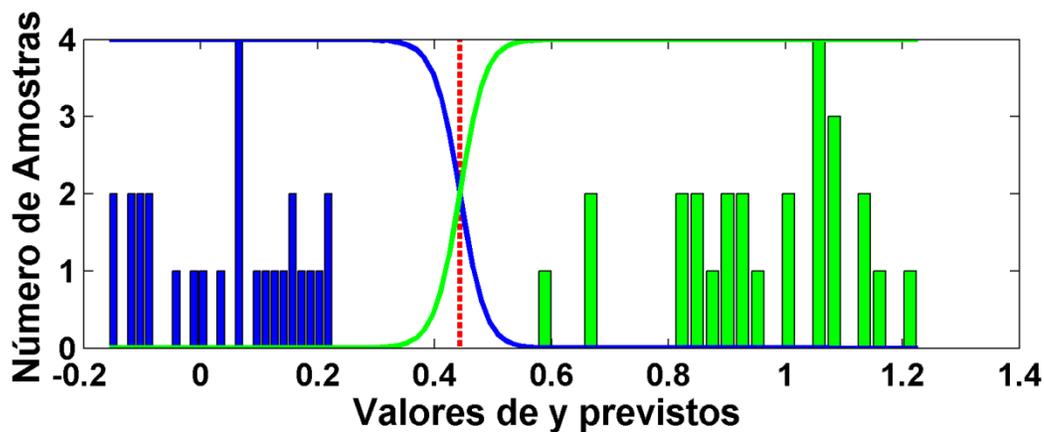


Figura 16: Distribuição dos valores previstos de y para determinação do valor limite entre as classes seguindo o teorema de Bayes.

O cálculo da confiabilidade na classificação é um parâmetro importante, não somente deve-se atribuir uma amostra a uma classe, mas também é necessário saber a incerteza da atribuição. Nessa tese, a técnica de reamostragem *bootstrap* é proposta para o cálculo dos intervalos de confiança de cada amostra do modelo de classificação PLS-DA.

3.4. Intervalos de confiança

O método de reamostragem *bootstrap* foi introduzido por Efron⁶⁶ para estimar intervalos de confiança de determinados parâmetros onde não era possível por outras técnicas, principalmente quando se dispunha de um número pequeno de amostras. Nestes casos, a técnica do *bootstrap* trata um conjunto de amostras como uma população. As novas amostras são geradas a partir de uma população finita E (e_1, e_2, \dots, e_n) com B substituições aleatórias com reposição a partir de um conjunto E .

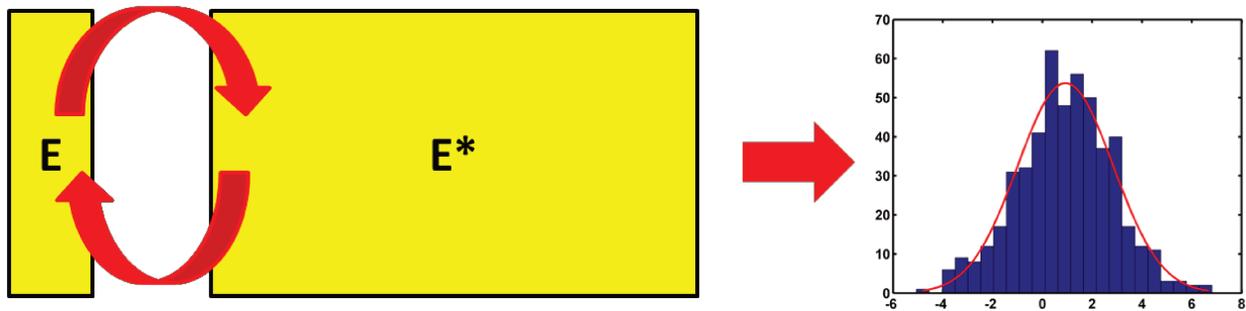


Figura 17: Representação esquemática do *bootstrap*.

Ao realizar este esquema de reamostragem várias vezes, uma boa estimativa dos parâmetros estatísticos pode ser obtida, como viés, desvio padrão e intervalo de confiança.

Existem dois tipos de *bootstrap* para estimar a incerteza de modelos de regressão, *bootstrap* por objetos e *bootstrap* residual. Aqui estamos interessados apenas no *bootstrap* residual, onde trabalhos da literatura mostram melhores resultados^{64,67}. A estratégia seguida foi desenvolvida de acordo com o procedimento apresentado por Pereira *et. al.*⁶⁸ com modificações no cálculo dos resíduos.

O primeiro passo para estimar as incertezas das previsões dos valores de y estimados pelo modelo PLS-DA é calcular os resíduos, de acordo com a equação 21:

$$F = Y_{ref} - Y_{PLS} \quad (21)$$

Sendo Y_{ref} o valor de referência e Y_{PLS} os valores previstos pelo modelo.

De acordo com Faber⁶⁷, os resíduos devem ser corrigidos pelos graus de liberdade devido à diferença entre os valores de referência e o modelado serem consideravelmente menores que os desvios dos valores esperados.

$$F = \frac{Y_{ref} - Y_{PLS}}{\sqrt{\left(1 - \frac{gl}{N}\right)}} \quad (22)$$

Onde gl é o número de graus de liberdade, N o número de amostras do conjunto de calibração. O número de graus de liberdade envolvidos consiste em mais uma dificuldade para os cálculos dos intervalos de confiança e várias abordagens tem sido propostas para estimar os graus de liberdade em modelos de regressão multivariados. No presente trabalho foi empregado os pseudo-graus de liberdade (PDF, do inglês *Pseudo-Degree of Freedom*) definido por van der Voet⁶⁹ que leva em consideração a

diferença entre o erro médio quadrático de calibração estimado por validação cruzada (MSECV) e pela previsão das amostras de calibração (MSEC), de modo que quanto maior essa diferença, menor será o número de graus de liberdade, conforme equação abaixo:

$$gl = PDF = N \left(1 - \sqrt{\frac{MSEC}{MSECV}} \right) \quad (23)$$

O MSECV e o MSEC são calculados de acordo com as equações 24 e 25, respectivamente.

$$MSECV = \frac{\sum_{i=1}^I (y_{ref} - y_{pred,cv})^2}{N-1} \quad (24)$$

$$MSEC = \frac{\sum_{i=1}^I (y_{ref} - y_{cal})^2}{N-1} \quad (25)$$

Onde y_{ref} são os valores de referência; o $y_{pred,cv}$ são os valores previstos pelo modelo na validação cruzada, y_{cal} os valores previstos na calibração e N o número de amostra do conjunto de calibração.

As amostras *bootstraps* são geradas a partir de substituições aleatórias com reposição dos valores dos resíduos corrigidos (Equação 22). Os resíduos *bootstraps* são adicionados ao Y_{PLS} gerando uma matriz Y^* ($N \times B$):

$$Y^* = Y_{PLS} + F^* \quad (26)$$

Como resultado, o desvio padrão, o viés, os intervalos de confiança, os testes de hipóteses e outras análises podem ser obtidos pelo método do *bootstrap*.

Um novo modelo PLS pode ser calculado a partir do Y^* . Um novo coeficiente de regressão *bootstrap* (β^*) é estimado e permite o cálculo de novos valores de \hat{Y}^* . Novos resíduos são calculados:

$$\hat{F}^* = Y^* - \hat{Y}^* \quad (27)$$

Para obter o intervalo de confiança para cada amostra, o método do percentil foi utilizado. O intervalo percentil é definido pelos percentis α e $1-\alpha$, de acordo com a equação 28.

$$\hat{F}_{B(\frac{\alpha}{2})}^* \leq y_a \leq \hat{F}_{B(1-\frac{\alpha}{2})}^* \quad (28)$$

Onde B é número de replicações, α a confiança do intervalo (0,05), sendo $B(\frac{\alpha}{2})$ -ésimo valor da ordenação das B replicações de \hat{F}^* .

3.5. Avaliação do desempenho do modelo de classificação

O desempenho do modelo de classificação foi avaliado através de tabelas de contingência, onde é possível extrair parâmetros como a taxa de falso positivo e falso negativo, sensibilidade, especificidade, eficiência e coeficiente de correlação de Matthew^{70,71,72}. Nos estudos descritos nesta tese foi possível utilizar a tabela de contingência porque a procedência das amostras era conhecida.

Apesar de alguns parâmetros terem o mesmo nome, eles diferem da definição empregada em química analítica quantitativa e metrologia. A razão de falso positivo é a probabilidade de uma amostra negativa ser classificada como positiva. A razão de falso positivo é calculada da seguinte forma:

$$\text{razão de falso positivo} = \frac{fp}{vn + fp} \times 100 \quad (29)$$

Onde fp é o número de amostras falso positivas e vn é o número de amostras verdadeira negativas.

De forma similar, a taxa de falso negativo é a probabilidade de uma amostra positiva ser classificada como negativa.

$$\text{razão de falso negativo} = \frac{fn}{vp + fn} \times 100 \quad (30)$$

Onde fn é o número de amostras falso negativas e vp é o número de amostras verdadeira positivas.

Equações similares podem ser usadas para calcular sensibilidade e especificidade. A sensibilidade é a habilidade de o modelo classificar corretamente as amostras positivas e pode ser calculada de acordo com a equação abaixo:

$$\text{Sensibilidade} = \frac{vp}{vp + fn} \quad (31)$$

A especificidade é a capacidade do modelo em identificar corretamente as amostras negativas, e é calculada com a equação abaixo:

$$\text{Especificidade} = \frac{vn}{vn + fp} \quad (32)$$

A eficiência e o coeficiente de correlação de Matthew são parâmetros estatísticos que fornecem um único valor para medir o desempenho do modelo de classificação. A eficiência combina todas as informações obtidas na sensibilidade e especificidade:

$$Eficiência = \frac{vp + vn}{vp + vn + fp + fn} \quad (33)$$

O coeficiente de correlação de Matthew resume a qualidade da tabela de contingência em um único valor numérico, este valor pode ser comparado, mesmo que as classes tenham tamanhos diferentes.

$$Coeficiente\ de\ Matthew = \frac{(vp \times vn) - (fp \times fn)}{\sqrt{(vp + fp) \times (vp + fn) \times (vn + fp) \times (vn + fn)}} \quad (34)$$

A equação 34 retorna valores entre -1 e 1, onde o valor +1 representa uma classificação perfeita, zero representa uma classificação aleatória e -1 uma classificação inversa.

3.6 Métodos de Resolução de Curvas

Os métodos de resolução de curvas são classificados como métodos de processamento de sinais que são aplicados a um conjunto de dados com o objetivo de resolver misturas de sinais¹⁰. De forma geral, os métodos de resolução multivariada de curvas decompõem um conjunto de dados químicos espectrais de acordo com a equação:

$$\mathbf{X} = \mathbf{CS}' + \mathbf{E} \quad (35)$$

onde \mathbf{X} representa a matriz de dados, \mathbf{C} a matriz referente ao perfil de concentração, \mathbf{S} a matriz relacionada aos perfis espectrais e \mathbf{E} o erro. O método mais empregado para essa proposta em dados químicos é a resolução multivariada de curvas com mínimos quadrados alternados (MCR-ALS). Nesta tese é feita uma comparação entre o MCR-ALS e o método de análise de componentes independentes (ICA). A Figura 18 ilustra a decomposição de uma matriz de dados usando métodos de resolução de curvas. A matriz \mathbf{S} mostra os espectros puros obtidos e é usada para proposta de identificação,

enquanto a matriz de concentração \mathbf{C} é redobrada para recuperar o espaço do pixel fornecendo o mapa de distribuição de cada componente na amostra analisada.

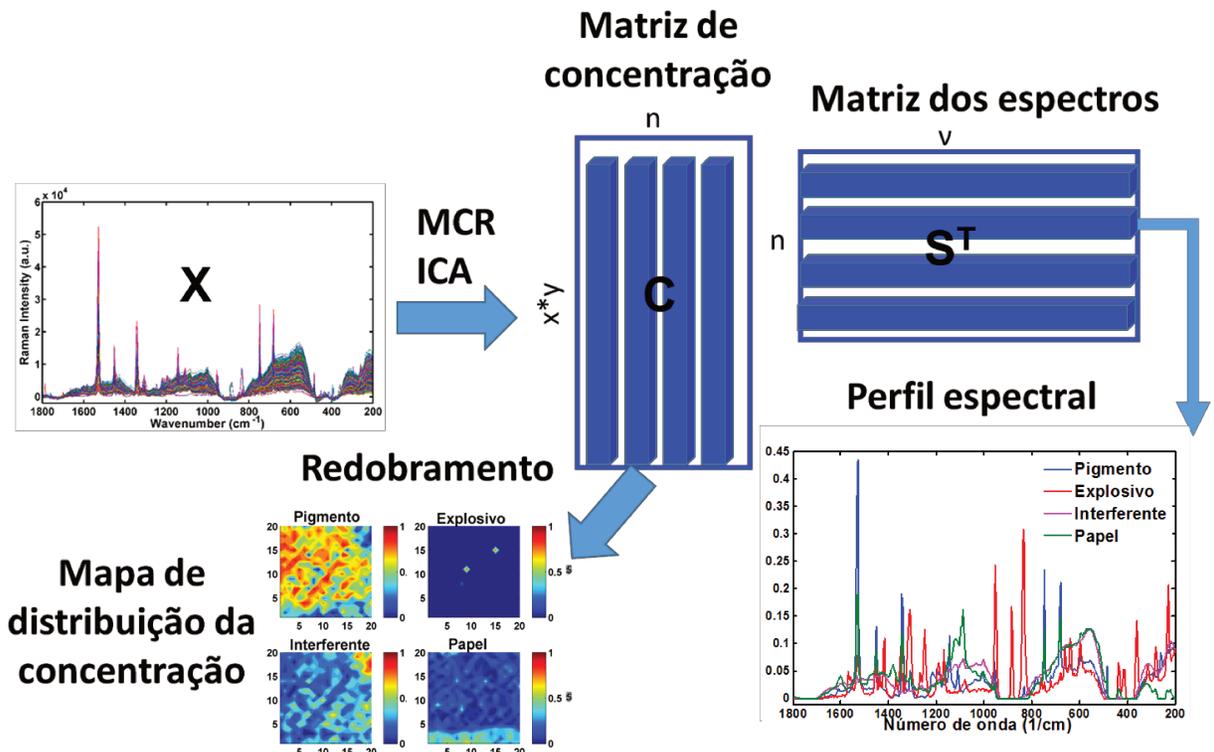


Figura 18: Métodos de resolução de curvas aplicados a dados de espectroscopia Raman de imagem.

3.6.1. Resolução Multivariada de Curvas com Mínimos Quadrados Alternados (MCR-ALS)

O MCR-ALS é um método de processamento de sinais que tem o objetivo de encontrar os sinais e as concentrações dos componentes puros presentes em uma mistura. As matrizes \mathbf{C} e \mathbf{S} são encontradas de forma iterativa, de forma que o produto entre \mathbf{C} e \mathbf{S} retorne a matriz \mathbf{X} com o menor erro possível⁷³.

O cálculo é realizado para um número pré-determinado de componentes, o qual pode ser determinado por decomposição em valores singulares (SVD), análise de componentes principais (PCA) ou por conhecimento prévio do conjunto de dados. No

MCR-ALS é necessário fornecer as estimativas iniciais de **C** ou **S**, que podem ser obtidas pelo conhecimento do sistema, ou quando não se conhece, por métodos de seleção de variáveis puras, como o SIMPLISMA⁷³ ou pelos *loadings* da PCA.

Após a escolha do número de componentes e da estimativa inicial de **C** ou **S**, o processo de otimização é realizado resolvendo de forma iterativa as equações abaixo:

$$\mathbf{C} = \mathbf{X}^* \mathbf{S}^+ \quad (36)$$

$$\mathbf{S} = \mathbf{X}^* \mathbf{C}^+ \quad (37)$$

As matrizes **C**⁺ e **S**⁺ são as pseudoinversas das matrizes **C** e **S**, respectivamente.

A cada iteração uma matriz **C** e uma matriz **S** são estimadas. Se a estimativa inicial foi para a matriz **S**, a equação 35 é resolvida para a matriz **C**:

$$\mathbf{X}^* (\mathbf{S}')' = \mathbf{C} \mathbf{S}' \mathbf{S}'^+ = \mathbf{C} \quad (38)$$

Quando a estimativa inicial é realizada para matriz **C**, a equação 35 é resolvida para a matriz **S**⁺:

$$\mathbf{C}^+ \mathbf{X}^* = \mathbf{C}^+ \mathbf{C} \mathbf{S}' = \mathbf{S}' \quad (39)$$

As soluções do MCR não são únicas, assim, diferentes combinações entre **C** e **S** podem gerar a mesma solução matemática para a equação 35, porém, muitas dessas soluções podem não apresentar sentido químico. Essa falta de convergência é chamada de ambiguidade rotacional⁷⁴.

Para corrigir ou diminuir o número de soluções possíveis é necessário adicionar restrições como não negatividade, balanço de massa à concentração, seletividade e posto local para convergência das soluções⁷⁴. A restrição de não negatividade estabelece valores positivos para as concentrações (matriz **C**) e para os sinais analíticos (matriz **S**), uma vez que somente valores positivos tem sentido físico. O balanço de massa faz com que o somatório dos perfis de concentração seja constante durante o processo de otimização. No caso de amostras muito homogêneas, o algoritmo MCR-ALS não consegue convergir para perfis dos constituintes puros devido à falta de uma área seletiva da amostra onde cada constituinte poderia estar presente. Nesses casos, o número de componentes químicos (posto químico) presentes na amostra é diferente do número de componentes distinguível pelo algoritmo e o MCR-ALS não converge para as soluções adequadas.

O uso de restrições nem sempre é possível ou suficiente e as soluções podem não convergir para o resultado correto. Para corrigir esse problema, outras estratégias têm sido adotadas, por exemplo, o uso de matriz aumentada, na qual a decomposição dos dados é feita em diferentes matrizes simultaneamente, como mostrada na Figura 19. Essa estratégia aumenta a variabilidade na matriz \mathbf{X} e minimiza o problema de deficiência de posto⁷⁴.

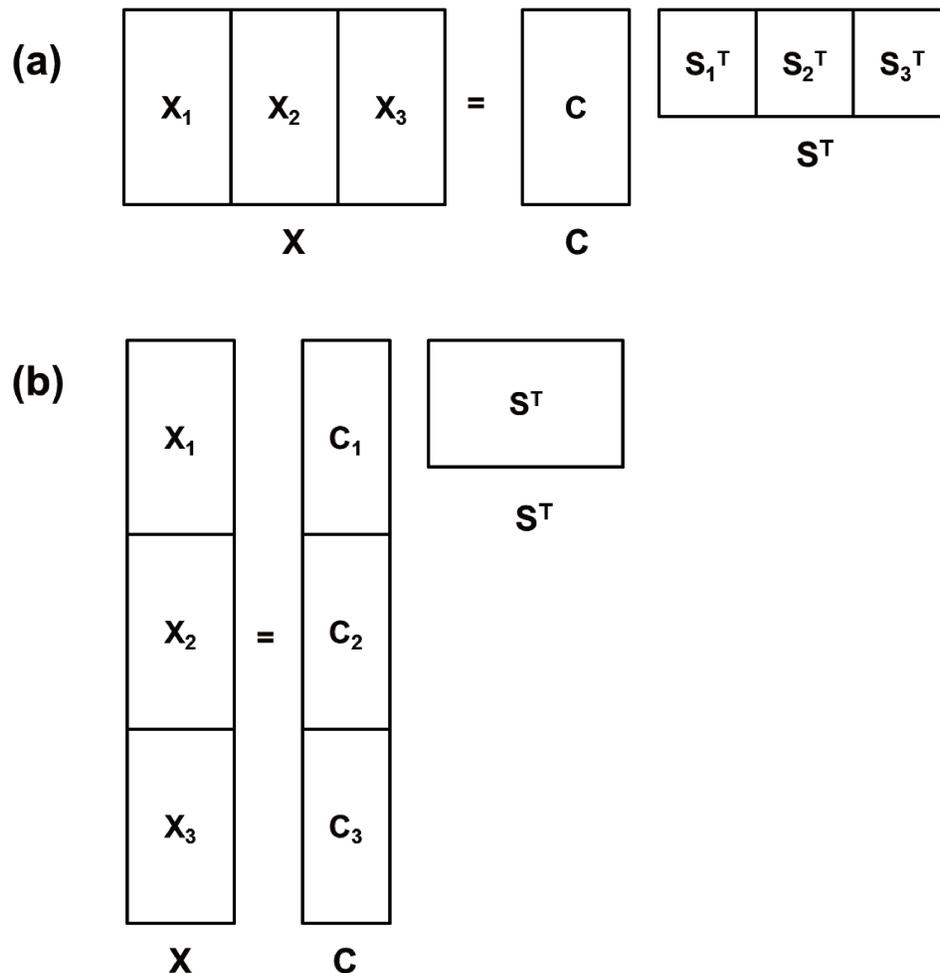


Figura 19: MCR com matriz aumentada pelas linhas (a) e pelas colunas (b), figura adaptada da referência⁷⁴.

Outra alternativa que pode ser empregada para corrigir a falta de convergência do MCR-ALS é o iMCR, ou seja, o MCR em intervalos⁷⁵. Nesse caso, a matriz de dados é dividida em submatrizes, e em cada submatriz o MCR-ALS é empregado, como mostrado na Figura 20.

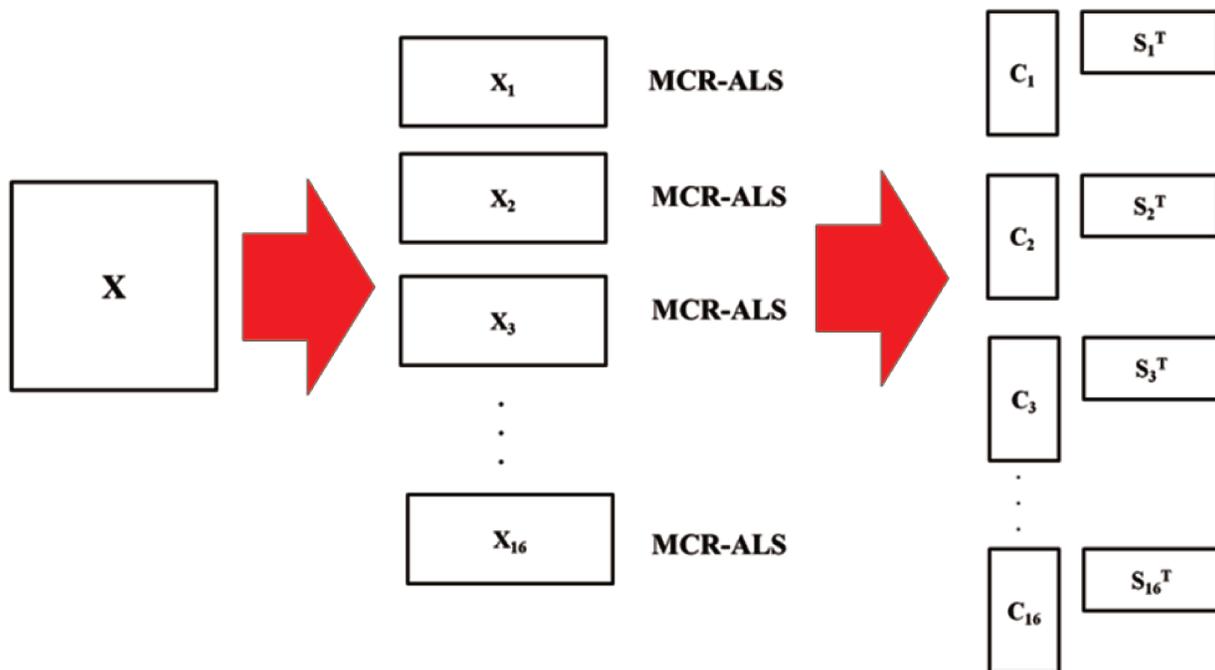


Figura 20: Esquema da decomposição da matriz \mathbf{X} para aplicação do iMCR-ALS.

3.6.2. Análise de Componentes Independentes (ICA)

A ICA é um método de separação cega das fontes, que foi primeiramente empregado na separação de sinais de áudio em telecomunicações⁷⁶ e nos últimos anos tem sido empregado em química analítica⁷⁷. A ICA assume que um conjunto de sinais observados \mathbf{X} é formado por uma combinação linear de fontes desconhecidas \mathbf{S} :

$$\mathbf{X} = \mathbf{A}\mathbf{S}' \quad (40)$$

O objetivo da ICA é recuperar as fontes puras \mathbf{S} , utilizando apenas as informações de \mathbf{X} , por esse motivo o método é chamado de separação cega de sinais (BSS, do inglês, *blind signal separation*).

A ICA parte do princípio que as fontes originais são estatisticamente independentes e estima as componentes de forma que estas sejam as mais independentes possíveis. O conceito de independência estatística diz que duas variáveis aleatórias são independentes quando o conhecimento de uma delas não traz

nenhuma informação a respeito da outra. Uma vez que não se tem conhecimento dos coeficientes da mistura e dos componentes independentes, é necessário realizar uma transformação linear \mathbf{W} (matriz de separação) que quando aplicada a \mathbf{X} produz os componentes \mathbf{S} independentes:

$$\hat{\mathbf{S}} = \mathbf{W}\mathbf{X} \quad (41)$$

O problema da ICA se resume em encontrar a matriz de separação \mathbf{W} . A estimativa de \mathbf{W} é realizada principalmente através de algoritmos iterativos de otimização, que vão maximizar ou minimizar uma determinada função custo.

Após a estimativa de \mathbf{W} , é feita a ortogonalização e obtém-se a matriz de mistura \mathbf{A} e estima-se os componentes independentes através de \mathbf{X} .

$$\mathbf{A} = \mathbf{W}^{-1} \quad (42)$$

$$\hat{\mathbf{S}} = \mathbf{W}\mathbf{X} = \mathbf{W}\mathbf{A}\mathbf{S}' = \mathbf{S}' \quad (43)$$

A otimização é feita com as restrições de independência e geralmente é utilizado o princípio de não gaussianidade para estimar a independência estatística. O princípio de não gaussianidade vem do teorema do limite central, onde a função de densidade de probabilidade da soma de duas variáveis aleatórias é sempre mais próxima de uma distribuição Gaussiana do que as funções de densidade de probabilidade das variáveis originais⁷⁷.

Considerando que os sinais são uma combinação linear das fontes, as fontes são as componentes com as funções de densidade de probabilidade distantes da Gaussiana. Um dos parâmetros que é utilizado para estimar se uma distribuição é ou não Gaussiana é a curtose, que é um parâmetro estatístico de quarta-ordem que assume valores iguais à zero para distribuições Gaussianas⁷⁷.

Inúmeros algoritmos foram propostos para encontrar as componentes independentes de um conjunto de dados⁷⁸. Os algoritmos apresentam características que os tornam mais indicados para determinadas aplicações. Nessa tese, o algoritmo MF-ICA (*Mean field ICA*) proposto por Hojen-Sorensen *et al* foi usado⁷⁹. O que levou a utilização desse algoritmo entre as várias opções disponíveis foi a possibilidade de adicionar restrições de não-negatividade. No MF-ICA o conjunto de sinais \mathbf{X} assume ser uma mistura linear das fontes com adição de ruído Gaussiano τ , tem-se então que:

$$\mathbf{X} = \mathbf{A}\mathbf{S}' + \tau \quad (44)$$

No MF-ICA, a razão de verossimilhança $P(X|A, \Sigma, S)$ para a matriz de mistura \mathbf{A} , matriz de covariância Σ e as fontes dos sinais puros \mathbf{S} são calculadas a partir da matriz \mathbf{X} . Os valores obtidos para as fontes \mathbf{S} e para os parâmetros \mathbf{A} e Σ fornecem a máxima razão de verossimilhança. A principal dificuldade é calcular a estimativa da matriz de sinais puros \mathbf{S} . Para melhorar essa estimativa, algumas restrições podem ser adicionadas, como o uso da não negatividade em $p(\mathbf{S})$ e $p(\mathbf{A})$. O uso de probabilidades condicionais e de probabilidades a priori permite que a matriz dos sinais puros seja estimada através de probabilidade a posteriori usando o teorema de Bayes e a matriz das misturas e o nível de ruído sejam estimadas pela maximização da função de verossimilhança. As equações para estimar os componentes independentes usando esta abordagem seriam:

$$\hat{\mathbf{S}} = \langle \mathbf{S} \rangle \quad (45)$$

$$\hat{\mathbf{A}} = \mathbf{X} \langle \mathbf{S}' \rangle \langle \mathbf{S} \mathbf{S}' \rangle^{-1} \quad (46)$$

$$\Sigma = \frac{1}{n} \langle (\mathbf{X} - \hat{\mathbf{A}} \hat{\mathbf{S}}) (\mathbf{X} - \hat{\mathbf{A}} \hat{\mathbf{S}})' \rangle \quad (47)$$

Onde $\langle . \rangle$ indica a média a posteriori da matriz \mathbf{S} dadas as matrizes \mathbf{A} e de covariância. O MF-ICA é um método de resolução iterativo, alternativamente a estimativa de $\hat{\mathbf{S}}$ pode ser feita iniciando tanto por $\hat{\mathbf{A}}$ ou por Σ . No processo de otimização, o objetivo é estimar $\hat{\mathbf{S}}$ e $\hat{\mathbf{A}}$ de forma a ter um mínimo Σ .

No método ICA não é possível determinar as variâncias das componentes independentes, como consequência não se pode determinar sua ordem. A escolha do número de componentes também é uma etapa crítica para o modelo. Vários métodos têm sido propostos na literatura, no entanto, muitos deles necessitam do conhecimento prévio da matriz \mathbf{A} ou \mathbf{S} . Para a determinação do número de componentes independentes em dados químicos têm sido propostos os métodos de ICA por bloco, o critério de Durbin-Watson⁸⁰ e a raiz quadrada das somas dos resíduos⁷⁸.

Para dados de espectroscopia vibracional, a matriz \mathbf{S} recupera os espectros puros de cada constituinte da amostra e a matriz \mathbf{A} (equivalente a matriz \mathbf{C} no MCR-ALS) o perfil de concentração dos constituintes na amostra.

3.6.3. MCR-BANDS

A presença de ambiguidade rotacional é uma das principais causas de soluções incorretas para os métodos de resolução de curvas. A ambiguidade rotacional, como definida anteriormente, é a indeterminação nas soluções \mathbf{C} e \mathbf{S}' do modelo. Isto é, um número infinito de soluções pode ser obtido da combinação linear entre \mathbf{C} e \mathbf{S} , que quando são multiplicadas, produzem o mesmo resultado, ou seja, a matriz \mathbf{X}^* (sem ruído). A ambiguidade rotacional pode ser representada matematicamente pela equação 48:

$$\mathbf{X}^* = \mathbf{C}_{anterior} \times \mathbf{S}'_{anterior} = (\mathbf{C}_{anterior} \times \mathbf{T}^{-1}) \times (\mathbf{T} \times \mathbf{S}'_{anterior}) = \mathbf{C}_{novo} \times \mathbf{S}'_{novo} \quad (48)$$

De acordo com a equação acima, qualquer matriz $\mathbf{T}(N,N)$ fornece um novo conjunto de soluções equivalentes.

O método MCR-BANDS⁸¹ foi empregado para avaliar a quantidade de ambiguidade rotacional associado a soluções do MCR-ALS e do MF-ICA. O MCR-BANDS é baseado na maximização e minimização do sinal de cada componente na mistura. A contribuição do sinal de cada componente em relação ao conjunto de sinal é calculada usando a expressão:

$$f_n = \frac{\|\mathbf{C}_n \mathbf{S}'_n\|}{\|\mathbf{C} \mathbf{S}'\|} \quad (49)$$

onde \mathbf{C} e \mathbf{S}' são os perfis de concentração e espectral recuperados pelo método; \mathbf{C}_n e \mathbf{S}'_n são os perfis de concentração e espectral do componente n . Quando a diferença entre os valores máximos e mínimos de f_n são próximos de zero, não há presença de ambiguidade rotacional nas soluções.

3.6.4. Determinação do número de componentes

A escolha do número de componentes é uma etapa importante nos métodos de resolução de curvas. O número de componentes deve ser igual ao número de espécies que produzem sinal analítico no conjunto de dados. A determinação do número de componentes pode ser feita através de métodos de análise exploratória.

Nesta tese, a escolha do número de componentes para os métodos de resolução multivariada de curvas foi feita por Análise de Fatores Evolucionários⁸² (EFA, do inglês, *Evolving Factor Analysis*). A ideia central da EFA é acompanhar a evolução de posto (dimensionalidade) da matriz, ou seja, o número de componentes do sistema. A matriz \mathbf{X} com os espectros Raman, por exemplo, é iniciada com três amostras, o que corresponde a três pixels. É calculado um modelo PCA e o número de componentes é selecionado. Novas amostras são adicionadas a matriz \mathbf{X} e para cada amostra adicionada, a PCA é realizada e o número de componentes selecionado. A Figura 21 ilustra o procedimento da EFA. Uma mudança no número de componentes principais indica a presença de um espectro de outro constituinte na matriz de dados, ou seja, o posto da matriz aumenta uma unidade. Procede-se dessa maneira até que todas as amostras do conjunto de dados tenham sido adicionadas e analisadas.

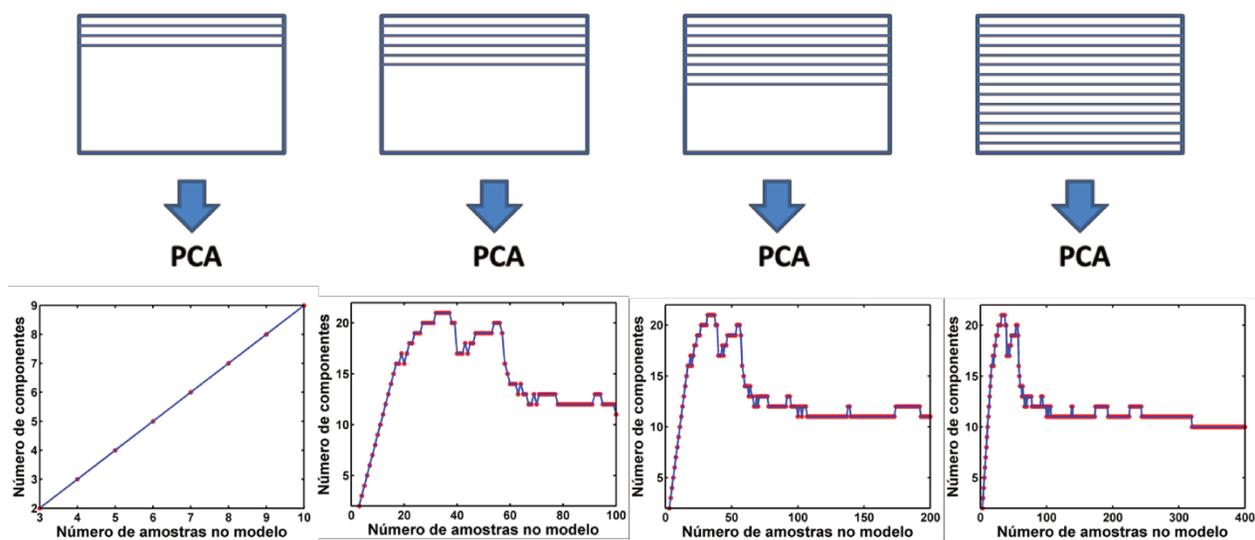


Figura 21: Esquema mostrando a análise de fatores evolucionária para a escolha do número de componentes nos métodos de resolução de curvas.

3.6.5. Figuras de mérito

Os métodos de resolução multivariada de curvas foram avaliados e comparados usando o parâmetro da porcentagem da variância explicada (R^2), calculada pela equação 50, a qual mostra quanto de \mathbf{X} foi explicado pelos modelos MF-ICA e MCR-ALS:

$$R^2 = 100 \times \left(1 - \frac{\sum_{ij} e_{ij}^2}{\sum_{ij} x_{ij}^2} \right) \quad (50)$$

onde e_{ij} são os elementos da matriz \mathbf{E} e x_{ij} são os elementos da matriz e dados \mathbf{X} .

A diferença entre a matriz de dados experimentais e os dados reproduzidos pelo produto $\mathbf{C} \cdot \mathbf{S}^T$ dos métodos de resolução de curvas é chamada de falta de ajuste (LOF), e foi calculada de acordo com a expressão abaixo:

$$LOF(\%) = 100 \sqrt{\frac{\sum_{ij} e_{ij}^2}{\sum_{ij} x_{ij}^2}} \quad (51)$$

Os coeficientes de correlação foram calculados para os perfis espectrais dos componentes recuperados pelos métodos de resolução de curvas e os espectros de referência e, segundo a equação:

$$r = \frac{\mathbf{s}^T \mathbf{y}}{\|\mathbf{s}\| \|\mathbf{y}\|} \quad (52)$$

onde \mathbf{s} é o espectro recuperado pelo método de resolução de curvas e \mathbf{y} o espectro do constituinte puro.

CAPÍTULO 4: Objetivos

4.1 Objetivos Gerais

Os objetivos gerais do presente trabalho foram aplicar a espectroscopia Raman e espectroscopia Raman de imagem juntamente com técnicas quimiométricas como ferramentas analíticas a serem empregadas em estudos forenses e paleontológicos. Na área de forense foram desenvolvidos modelos de classificação empregando PLS-DA com cálculo da incerteza de classificação de cada amostra usando reamostragem por *bootstrap* e métodos de resolução de curvas aplicados a dados de imagem Raman com a proposta de identificação de explosivos em cédulas. Na área de paleontologia, imagens Raman obtidas de amostras de fósseis de peixes foram tratadas com métodos de resolução multivariada de curvas para recuperar informações químicas e espaciais com a proposta de correlacionar a composição dos fósseis com características biológicas.

4.2 Objetivos Específicos

- 1) Empregar a espectroscopia Raman para caracterização de tintas usadas na elaboração de cédulas de R\$ 50 autênticas e falsificadas. Com base no perfil espectral das tintas desenvolver modelos de classificação empregando PLS-DA para discriminar cédulas autênticas de cédulas falsas e identificar o tipo de impressora empregada na falsificação;
- 2) Identificar a fonte do óleo essencial extraído da árvore amazônica *Aniba Rosaedora Ducke*. Obter os espectros Raman dos óleos extraídos da madeira, das folhas e ramos. Construir modelo de classificação empregando o PLS-DA para distinguir óleo da madeira do óleos das folhas e galhos;
- 3) Calcular a incerteza na classificação de cada amostra dos modelos PLS-DA usando o método de reamostragem *bootstrap*;

4) Desenvolver uma metodologia utilizando espectroscopia Raman de imagem e métodos quimiométricos de resolução de curvas para detecção de explosivos em superfície de cédulas;

5) Comparar os métodos de resolução de curvas MF-ICA e MCR-ALS para tratamento de imagens Raman;

6) Aplicar a espectroscopia Raman de imagem e métodos de resolução multivariada de curvas em análise de fósseis de peixes para obter informação sobre a composição química e distribuição espacial dos componentes.

**CAPÍTULO 5: Aplicação da Espectroscopia Raman e modelos de
classificação multivariados em química forense**

5.1. Discriminação entre cédulas autênticas e falsificadas usando espectroscopia Raman e PLS-DA com estimativa da incerteza

5.1.1 Introdução

A documentoscopia é a área da ciência forense responsável pela análise e identificação de diversos tipos de falsificações, contrafações e adulterações em documentos, moedas, selos, cartões de crédito, cheques, contratos, entre outros⁸³. As técnicas utilizadas em documentoscopia são: inspeção ocular, através de lupas e microscópios, raios ultravioleta, luz infravermelha, cromatografia gasosa acoplada ao espectrômetro de massas, cromatografia líquida de alta eficiência, espectroscopia no infravermelho e fluorescência^{83,84}. Nos últimos anos observou-se um aumento no uso de técnicas espectroscópicas para análise de documentos questionados em relação a técnicas de separações⁸⁵ como mostra a Figura 22.

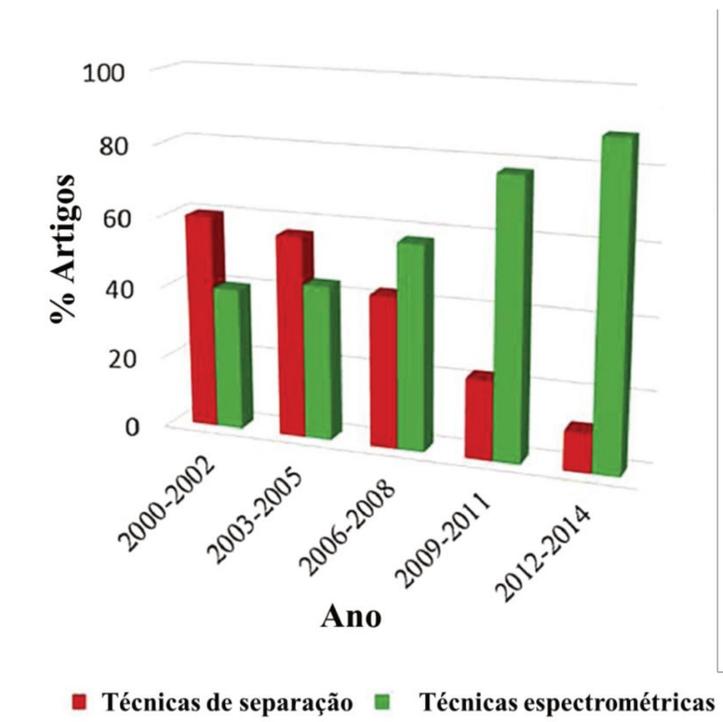


Figura 22: Estatística dos artigos publicados na última década na área de documentos questionados mostrando a tendência da utilização de métodos de separação e métodos espectrométricos. Figura adaptada da literatura⁸⁵.

A circulação de cédulas falsas é uma preocupação constante, uma vez que pode afetar a economia de um país. As cédulas possuem vários recursos de segurança que podem ser reconhecidos pela pessoa comum e são empregados na identificação de sua autenticidade. No entanto, com a popularização das tecnologias digitais, os falsificadores chegaram a um nível de qualidade que tornou praticamente impossível distinguir notas falsas das originais a olho nu. No Brasil, a cédula de R\$ 50 é a favorita dos falsificadores, representando 70% de todas as notas contrafeitas apreendidas⁸⁶, como pode ser observado na Figura 23. Toda vez que uma cédula falsa é identificada, esta é retirada de circulação. Atualmente, o volume de cédulas falsas não tem impactado a economia do país, mas traz prejuízos consideráveis aos cidadãos comuns.

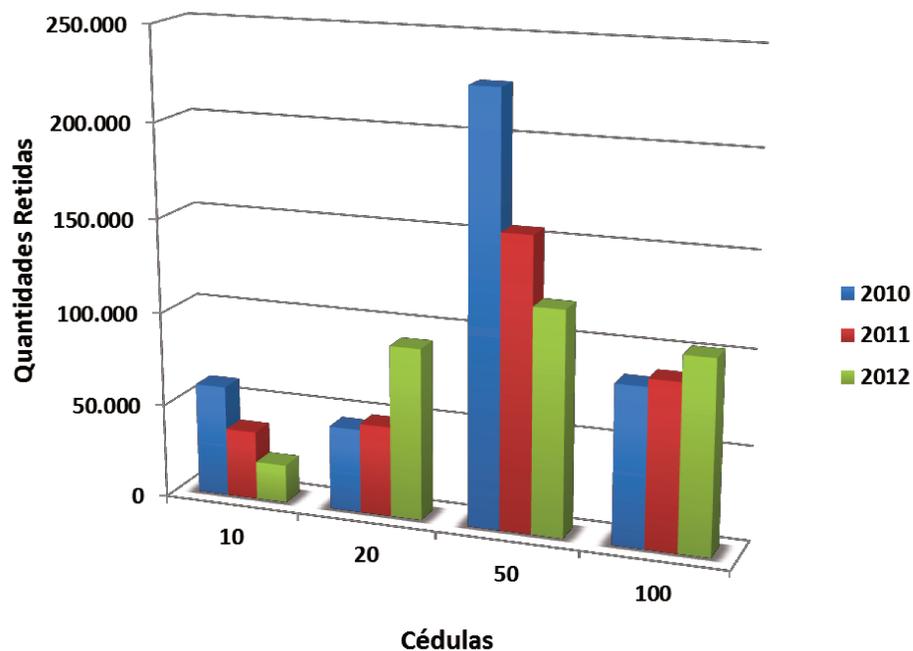


Figura 23: Estatística de cédulas da primeira família do real retidas nos anos de 2010, 2011 e 2012. Fonte: Banco Central do Brasil⁸⁶.

A proposta de métodos para o estudo de cédulas questionadas é pouco discutido na literatura. Trabalhos empregando espectroscopia no infravermelho e espectrometria de massas para identificação de cédulas falsificadas, incluindo Dólar, Euro e Real podem ser encontrados^{87,88}. Na última década tem-se observado um crescente interesse em espectroscopia Raman para amostras forenses, principalmente na identificação de pigmentos^{89,85}. Os trabalhos publicados na área de documentos questionados mostram

aplicações da técnica na área de análise de tintas de canetas, tintas de impressoras e cruzamento de traços⁹⁰. No estudo das cédulas falsificadas, o entendimento da composição química das tintas utilizadas para elaboração das notas é uma informação importante e a espectroscopia Raman apresenta-se como técnica apropriada para tal proposta.

Apesar de existir um número considerável de publicações da espectroscopia Raman em análises forenses, a espectroscopia Raman não é utilizada ainda em laboratórios forenses de rotina. Os motivos vão do uso restrito das ferramentas quimiométricas até à falta de validação de métodos com propostas qualitativas.

Devido à importância da estimativa da incerteza em dados analíticos, múltiplas propostas têm sido publicadas na literatura para o cálculo de incerteza de dados provenientes de análise multivariada, como por exemplo: método de propagação de erro, métodos de reamostragem e adição de ruído⁷.

Nesta aplicação foi empregada a espectroscopia Raman para caracterização das tintas usadas na elaboração de cédulas de R\$ 50 autênticas e falsificadas. Após a caracterização das tintas foram construídos modelos de classificação para discriminar entre cédulas autênticas e falsas baseados nos perfis espectrais das tintas. Os modelos desenvolvidos foram validados com um conjunto de novas amostras. A incerteza na classificação de cada amostra foi calculada usando o método de reamostragem *bootstrap*.

5.1.2 Parte Experimental

5.1.2.1 Amostras

As cédulas autênticas e falsificadas de R\$ 50 da primeira família foram fornecidas pela Superintendência da Polícia Técnico-Científica do Estado de São Paulo. Para conhecer o perfil dos pigmentos de impressoras comerciais, cédulas foram digitalizadas no laboratório e impressas em dois tipos de impressoras: jato de tinta e a laser de várias marcas encontradas no mercado. As impressões foram realizadas em

papel branco alcalino. O conjunto de amostras foi composto por 44 cédulas autênticas, 59 cédulas falsas apreendidas e 28 cédulas digitalizadas e impressas em diferentes impressoras. As amostras foram analisadas diretamente no espectrômetro Raman sem nenhuma preparação da amostra.

5.1.2.2 Espectros Raman

Os espectros das cédulas autênticas e falsas foram coletados com um espectrômetro dispersivo RamanStation 400F da Perkin Elmer equipado com detector CCD refrigerado a -50°C por resfriamento termoelétrico (sistema Peltier). Foi empregado um laser diodo com comprimento de onda de excitação em 785 nm e 250 mW de potência máxima na fonte. As amostras foram expostas ao laser por 3s e foram feitas 12 acumulações na região de 3200 a 400 cm^{-1} com uma resolução espectral de 4 cm^{-1} . A potência do laser foi ajustada de forma a não degradar a amostra, sendo usadas 75 e 50% da potência máxima nominal. Três diferentes áreas da cédula de R\$ 50 foram selecionadas para esse estudo: área vermelha, área laranja e área calcográfica, como mostrado na Figura 24. As áreas estudadas mostradas na Figura 24 são imagens ampliadas e mostram a região da cédula onde a radiação laser foi incidida para obtenção dos espectros.

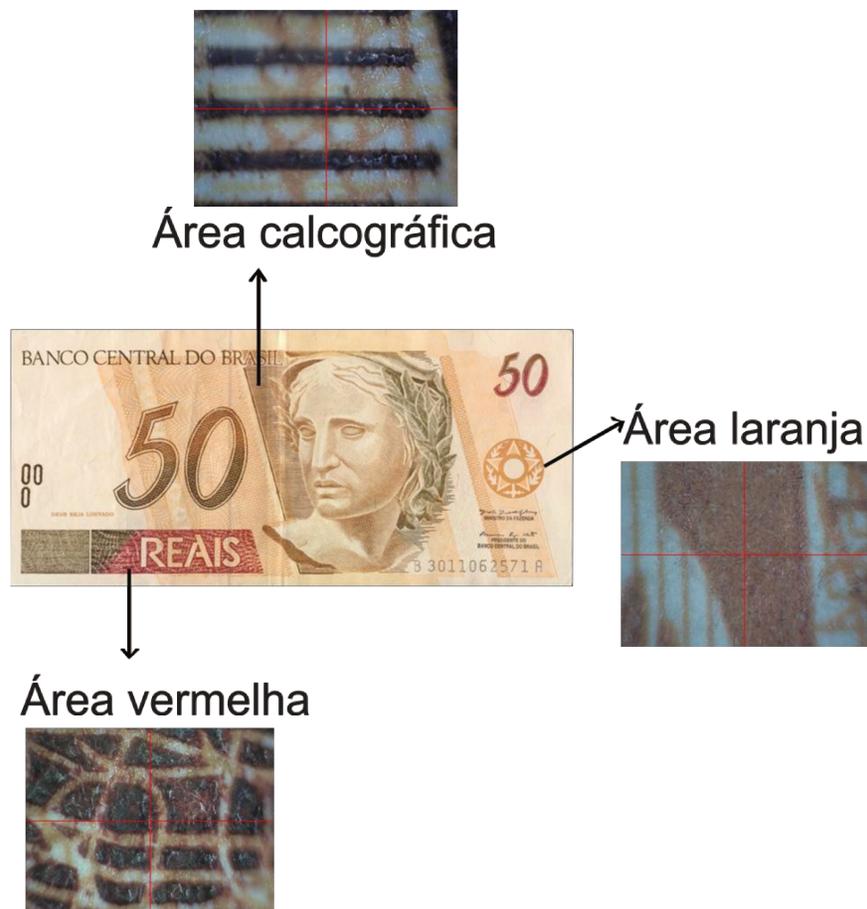


Figura 24: Cédula ilustrativa de R\$ 50 com as regiões analisadas por espectroscopia Raman.

5.1.3. Tratamento dos dados

Cada espectro Raman foi obtido em duplicata e a correção da linha base foi aplicada antes dos cálculos quimiométricos. Para a construção dos modelos de classificação os dados foram primeiramente pré-processados. A matriz \mathbf{X} com as intensidades Raman foi centrada na média e normalizada empregando o método de normalização SNV (do inglês, *Standart Normal Variate*). O vetor \mathbf{Y} que representa a classe que pertence cada amostra foi centrado na média. A Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA) foi aplicada para construir os modelos de classificação. Os dados foram aleatoriamente divididos em conjunto de calibração e validação. Foram construídos dois modelos: no primeiro modelo, o vetor \mathbf{Y} foi construída com 0 para as cédulas falsificadas e 1 para as cédulas autênticas; um

segundo modelo foi desenvolvido com as amostras fabricadas no laboratório, os valores na matriz **Y** foram 1 para as amostras impressas em impressoras a laser e 0 para as amostras confeccionadas em impressora jato de tinta. O número de variáveis latentes para os modelos PLS-DA foi selecionado por validação cruzada. O limite entre as classes foi calculado por Teorema de Bayes através da função *plsdthres* do software PLS toolbox⁹¹. O cálculo do intervalo de confiança para cada amostra foi feito com o *bootstrap* residual.

5.1.2. Resultados e Discussão

Os espectros Raman das três diferentes áreas estudadas da cédula autêntica são mostrados na Figura 25. O espectro Raman das tintas calcográficas e laranja apresentam bandas características dos pigmentos da classe das ftalocianinas e diarilida⁹². As estruturas químicas são mostradas na Figura 26a e 26b, respectivamente. O perfil do espectro Raman indica que a cor da tinta calcográfica e da cor laranja é resultado da mistura desses dois pigmentos.

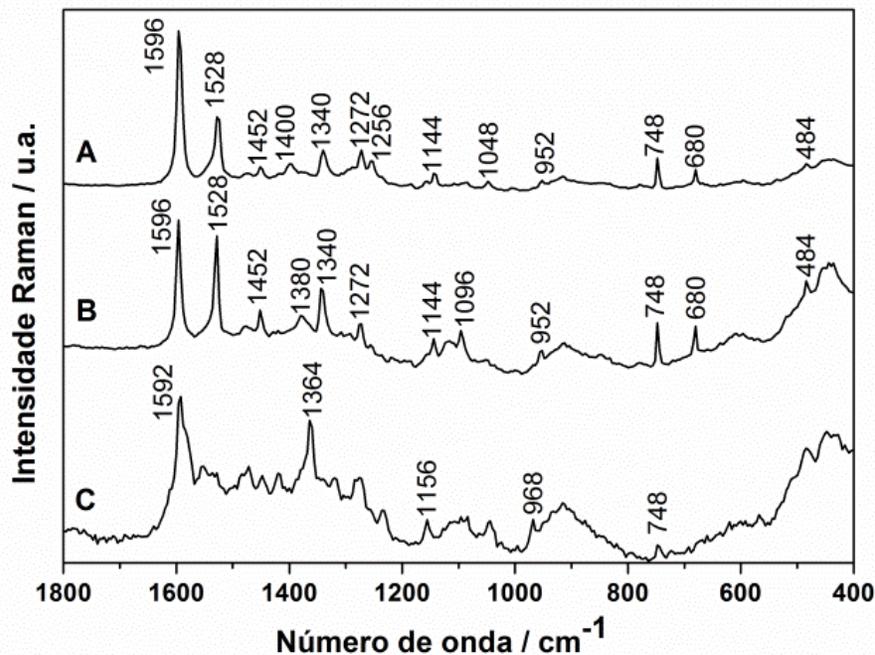
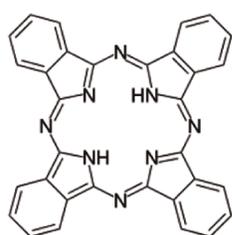
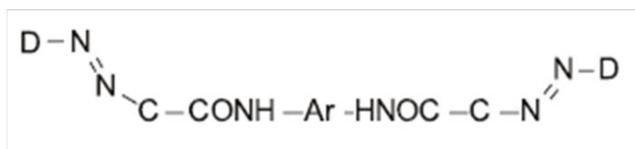


Figura 25: Espectros Raman das áreas analisadas da cédula autêntica de R\$ 50: (A) calcografia; (B) área laranja; (C) área vermelha.

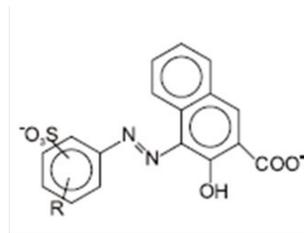
Os pigmentos da classe das ftalocianinas são amplamente utilizadas como materiais de geração de carga em dispositivos para impressão, tais como copiadoras e impressoras eletrofotográficas⁹³. As principais bandas Raman usadas para identificar o pigmento ftalocianina estão localizadas em 1528, 1340, 748, 680 e 484 cm^{-1} . A banda em 1528 cm^{-1} corresponde ao estiramento C=C e C=N do macrociclo; respiração e deformação do macrociclo aparecem em 680 e 750 cm^{-1} , respectivamente. O espectro Raman do pigmento diarilida apresenta quatro bandas características em 1596, 1398, 1256 e 953 cm^{-1} . A banda em 1596 cm^{-1} é atribuída a vibrações do anel aromático. A banda Raman em 1398 cm^{-1} é atribuída a deformação da ligação C-H; o sinal em 1256 cm^{-1} refere-se ao modo conhecido como Amida III, com contribuição dos modos de deformação N-H e estiramento C-N⁹⁴. Por último a banda que aparece em 953 cm^{-1} é devida ao estiramento simétrico da benzamida⁹⁵.



(a)



(b)



(c)

Figura 26: Estrutura molecular da (a) ftalocianina e de pigmentos dos grupos (b) diazo e (c) azo.

A análise do espectro Raman coletado da tinta laranja indica que o pigmento ftalocianina está presente em maior quantidade em relação ao pigmento diarilida, como sugere as bandas mais intensas que distinguem a ftalocianina em comparação com a tinta calcográfica. A proporção de cada pigmento é modificada para obter a cor desejada.

O espectro Raman da tinta vermelha revela bandas características de pigmentos do grupo azo (Figura 26c). O sinal em 1592 e 1364 cm^{-1} são atribuídos ao estiramento C-C e estiramento do grupo azo (N=N), respectivamente. As bandas Raman referentes à vibração de modos do naftaleno aparecem em 968 e 748 cm^{-1} . A banda de baixa intensidade em 1156 cm^{-1} pode ser atribuída ao estiramento simétrico SO_2 e indica pigmento do tipo laka⁹⁵.

A Figura 27 mostra os espectros Raman da tinta calcográfica da cédula autêntica e das cédulas digitalizadas e impressas em impressoras do tipo laser e jato de tinta no laboratório. O espectro Raman da cédula autêntica foi discutido anteriormente. Analisando o espectro Raman da cédula produzida por impressora do tipo laser, observam-se bandas do pigmento ftalocianina. A tinta calcográfica apresenta coloração marrom, provavelmente devido a mistura de ftalocianina e outro pigmento não identificado. O pigmento de ftalocianina é azul e seu espectro Raman pode apresentar intensificação de bandas com o uso de laser na região do visível (vermelho), devido ao efeito Raman ressonante⁹³. Como o outro pigmento não apresenta banda de absorção próxima de 785 nm não apresenta efeito ressonante e seu espectro pode ter sido ocultado pelas bandas da ftalocianina.

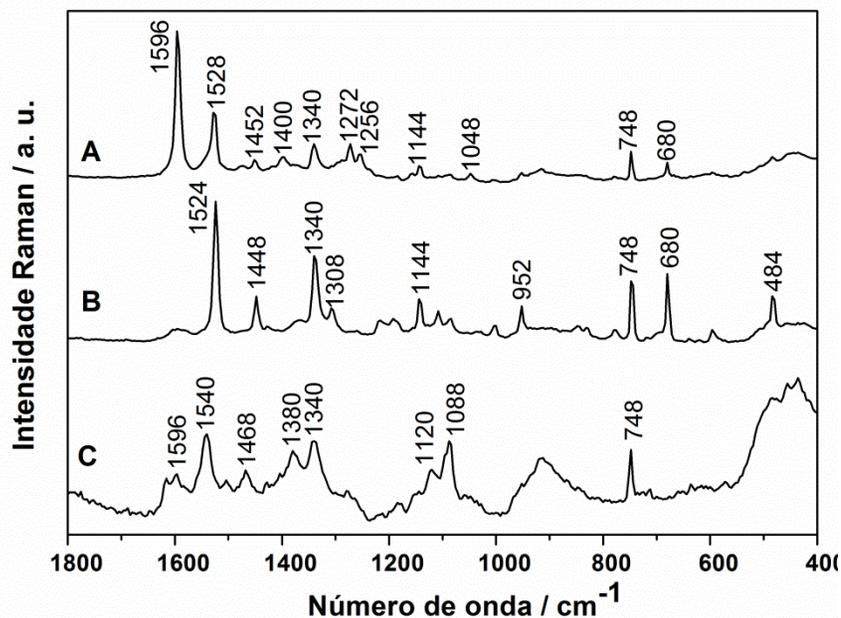


Figura 27: Espectro Raman da região calcográfica da cédula autêntica (A) e das cédulas falsas impressas em impressoras a laser (B) e jato de tinta (C).

O espectro Raman da cédula de R\$ 50 impressa em impressora do tipo jato de tinta mostrou bandas características de pigmento do tipo azo, indicado pela banda em 1596 cm^{-1} e outro composto do grupo azo como sugere a banda em 1380 cm^{-1} (estiramento N=N).

A análise dos espectros Raman da região de calcografia permite a diferenciação dos pigmentos utilizados na impressão de cédulas autênticas e falsas, no entanto, é

interessante o uso de outras ferramentas para classificar corretamente e discriminar as amostras. A análise quimiométrica pode ser empregada para criar modelos de classificação automáticos, validar os dados medidos e fornecer informações mais confiáveis. Os espectros Raman obtidos permitem ainda a construção de banco de dados que podem ser usados nos modelos de classificação.

Com este propósito, a análise exploratória baseada na PCA, foi aplicada aos dados. O modelo PCA foi construído com as cédulas autênticas, as cédulas falsas fornecidas pela Superintendência de Polícia e com as amostras confeccionadas no laboratório usando impressoras do tipo laser e jato de tinta.

Para construção do modelo PCA para a área calcográfica quatro componentes principais foram selecionados, explicando 90,2% da variância total das amostras. Na Figura 28a é possível avaliar os escores do modelo da PCA, com a projeção dos escores da PC1 x PC2 x PC4 observa-se a formação de três grupos diferentes entre as cédulas. Nota-se que as amostras autênticas formam um único grupo (*), para as amostras falsificadas (▼) é observado um único agrupamento também, o que indica que as amostras presentes no modelo apresentam espectros Raman com mesmas características. As amostras preparadas no laboratório (■) foram separadas de acordo com o tipo de impressora. As amostras impressas em impressoras do tipo laser ficaram próximas das cédulas falsas apreendidas. Esse resultado sugere que as cédulas de R\$ 50 apreendidas pela Polícia foram fabricadas nesse tipo de impressora. Para as notas impressas em impressora jato de tinta verificou-se a formação de um terceiro grupo, no qual duas amostras ficaram afastadas das demais, indicando que essas amostras possuem composição de pigmentos diferentes. A presença da mistura de diferentes pigmentos no conjunto das amostras estudadas justifica a análise multivariada na discriminação entre as cédulas autênticas e falsas.

Nas Figuras 28b e 28c estão representados os escores para as áreas de coloração laranja e vermelha das cédulas. Para essas áreas também foi observada a formação de grupos de acordo com o tipo de impressão empregado na confecção das cédulas.

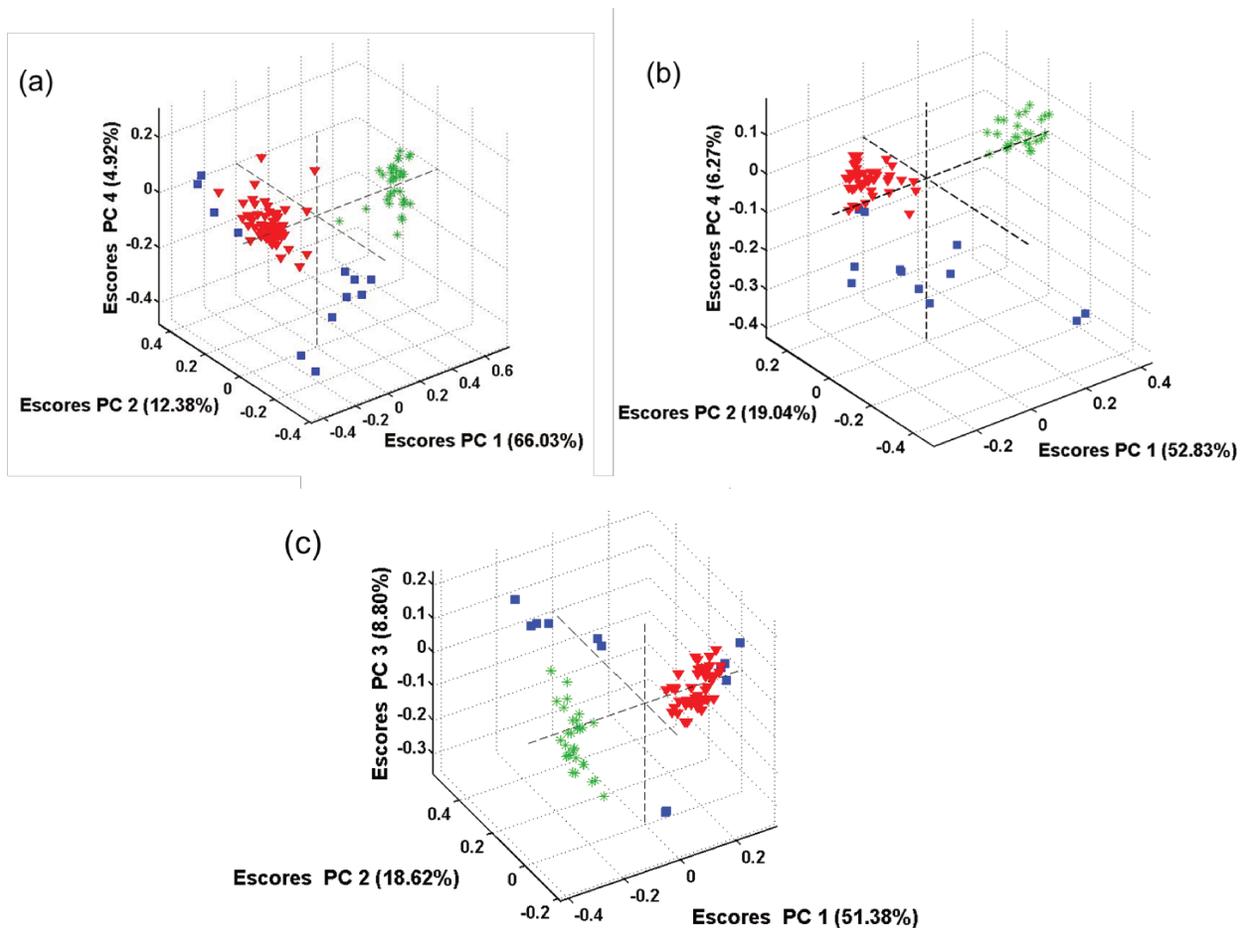


Figura 28: Gráfico dos escores: (a) área calcográfica; (b) área laranja e (c) área vermelha. Notas autênticas (*), notas falsas (▼) e notas digitalizadas e impressas no laboratório em papel alcalino (■).

As variáveis responsáveis pela formação dos agrupamentos das amostras podem ser observadas pelo gráfico dos *loadings* das componentes principais, apresentados na Figura 29. Os *loadings* da PC1 para a área calcográfica mostraram que a variável com maior contribuição é a banda Raman em 1596 cm^{-1} , atribuída ao pigmento diarilida e que está presente na cédula genuína, como discutido na seção anterior. As variáveis que contribuem para a segunda componente correspondem as bandas Raman em $1521, 1340, 748$ e 640 cm^{-1} , que são correspondente ao pigmento ftalocianina, presente na cédula original de R\$ 50 e principalmente nas cédulas falsas fabricadas em impressora do tipo laser. E finalmente os *loadings* da PC4 são representados pelas bandas Raman $1542, 1520, 1084, 910$ e 748 cm^{-1} . O perfil do gráfico dos *loadings* da quarta componente indica a contribuição das amostras

produzidas na impressora jato de tinta. Os *loadings* da região laranja mostraram que as variáveis que possuem maior peso na PC1 são as bandas Raman do pigmento diarilida, o pigmento ftalocianina apresenta contribuição nos *loadings* da PC2 e os *loadings* da PC4 têm contribuição desses dois pigmentos. Os *loadings* da tinta vermelha possuem contribuição de pigmentos do tipo azo, como discutido anteriormente.

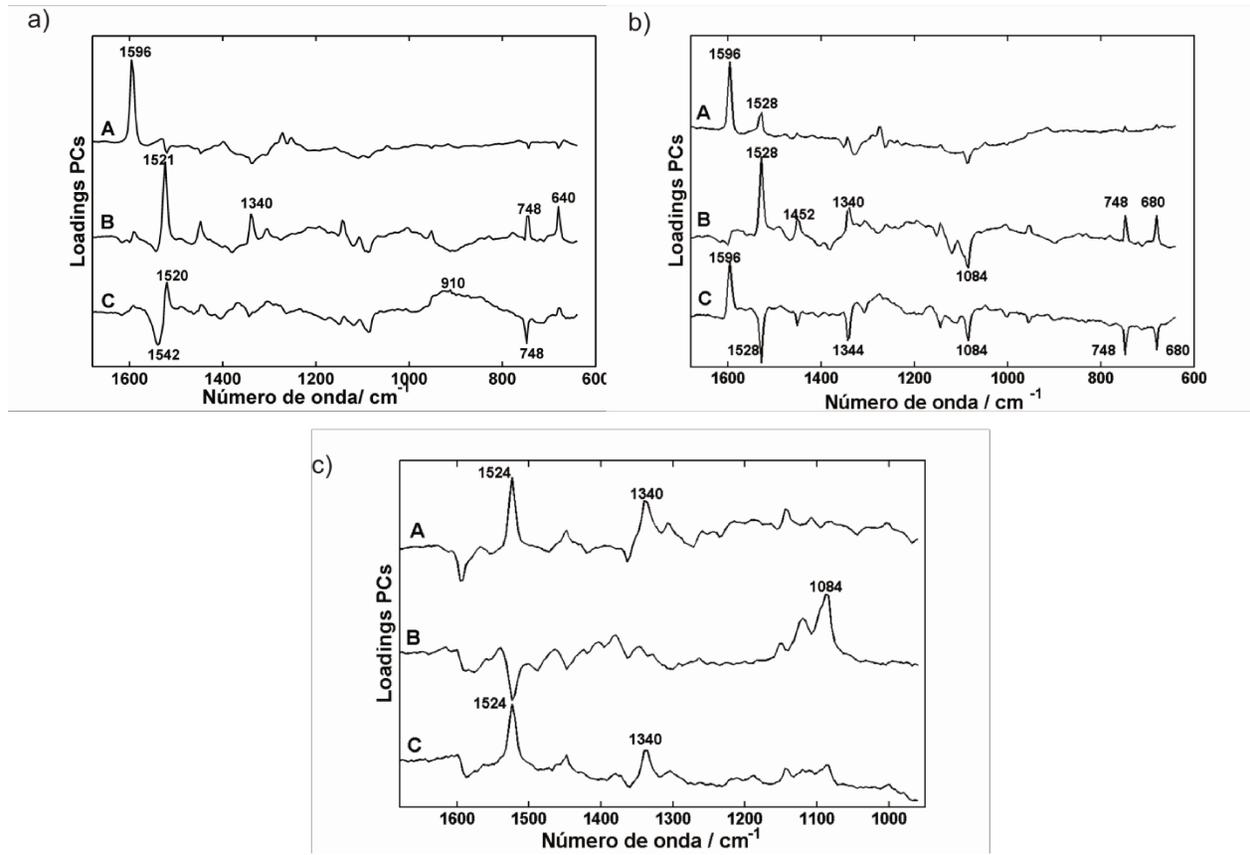


Figura 29: (a) Gráfico dos *loadings* da PC1 (A), PC2 (B) e PC4 (C) para a região de calcografia; (b) Gráfico dos *loadings* da PC1 (A), PC2 (B) e PC4 (C) para a região laranja; (c) Gráfico dos *loadings* da PC1 (A), PC2 (B) e PC3 (C) para a região vermelha da cédula de R\$ 50.

A análise exploratória empregando a PCA e os espectros Raman das cédulas genuínas e falsas permitiu a separação das amostras de acordo com o tipo de impressora empregada na fabricação das cédulas.

Após o estágio preliminar de extração de características para obter as informações relevantes usando a PCA, modelos PLS-DA foram desenvolvidos para classificar novas observações. Dois modelos foram construídos, um para discriminar entre cédula autêntica e falsa e um segundo modelo para identificar o tipo de impressora empregada na falsificação.

O primeiro modelo foi construído para classificar as cédulas em autênticas ou falsas. O conjunto de calibração foi composto por 30 amostras de cédulas de R\$ 50 autênticas e 28 amostras falsas produzidas no laboratório. A validação cruzada foi empregada para definir a dimensionalidade do modelo, sendo o número de variáveis latentes escolhido pela análise do gráfico de RMSECV (do inglês, *Root Mean Squared Error of Cross-Calibration*) versus o número de variáveis latentes.

Quatro variáveis latentes foram selecionadas, representando 88,8% da variância explicada em **X** e 98,9% da variância em **Y**. O valor limite que separa as classes foi calculado baseado na distribuição das amostras de calibração, esse valor é calculado de forma que as classes são divididas com uma menor probabilidade de classificação errada para as previsões futuras. Para isso assume-se uma distribuição normal dos valores previstos para cada classe. O valor limite foi de 0,46, acima desse valor as amostras são classificadas como autênticas e abaixo desse valor como falsas.

Em geral, o cálculo dos parâmetros RMSEC e RMESEP (erros de previsão) que são atribuídos como uma estimativa da exatidão para modelos PLS, não podem ser considerados para prever a incerteza das amostras futuras⁷. Ou seja, esses parâmetros não são suficientes para avaliar o desempenho do modelo desenvolvido, é importante também considerar sua confiabilidade. Então, após a otimização dos parâmetros do modelo de calibração, intervalos de confiança foram calculados para cada amostra.

Para este propósito, a técnica do *bootstrap* residual foi empregada. Devido à sua natureza estocástica, o método *bootstrap* apresenta resultados ligeiramente diferentes quando aplicado repetidamente para o mesmo problema (a não ser se os mesmos números aleatórios são reutilizados). Esta propriedade é certamente um problema para os metrologistas. Uma das maneiras de compensar essa variabilidade está relacionada com o ajuste do número de tentativas, que deve ser grande o suficiente para garantir a confiabilidade dos resultados do *bootstrap*. Levando essa característica do método em consideração, cada simulação foi realizada com 1000 testes aleatórios para cada amostra de calibração. Esse número foi otimizado de forma a obter um menor desvio padrão. O procedimento de reamostragem foi repetido 10 vezes. Todas as amostras foram consideradas como uma nova observação e o intervalo de confiança com o *bootstrap* foi computado.

Os resultados das amostras de calibração do PLS-DA do primeiro modelo para a área de tinta calcográfica são mostrados na Figura 30a. Os resultados do conjunto de validação externa com estimativa da incerteza são mostrados na Figura 30b. As amostras de validação classificadas como falsificadas pelo modelo foram fornecidas pela Superidentência da Polícia Técnico-Científica do estado de São Paulo e classificadas como falsas por análises forenses clássicas, baseado em testes de inspeção óptica por um profissional da área.

A estimativa da incerteza para cada amostra, mostrada pelas barras de erro na Figura 30, permite uma melhor confiança na classificação das amostras de validação. Os limites de confiança superior e inferior para esse conjunto de dados ficou na faixa de 0,030 e 0,031, e todas as amostras foram corretamente classificadas. Os resultados mostram a confiabilidade do modelo para discriminar cédulas de R\$ 50 autênticas das falsificadas.

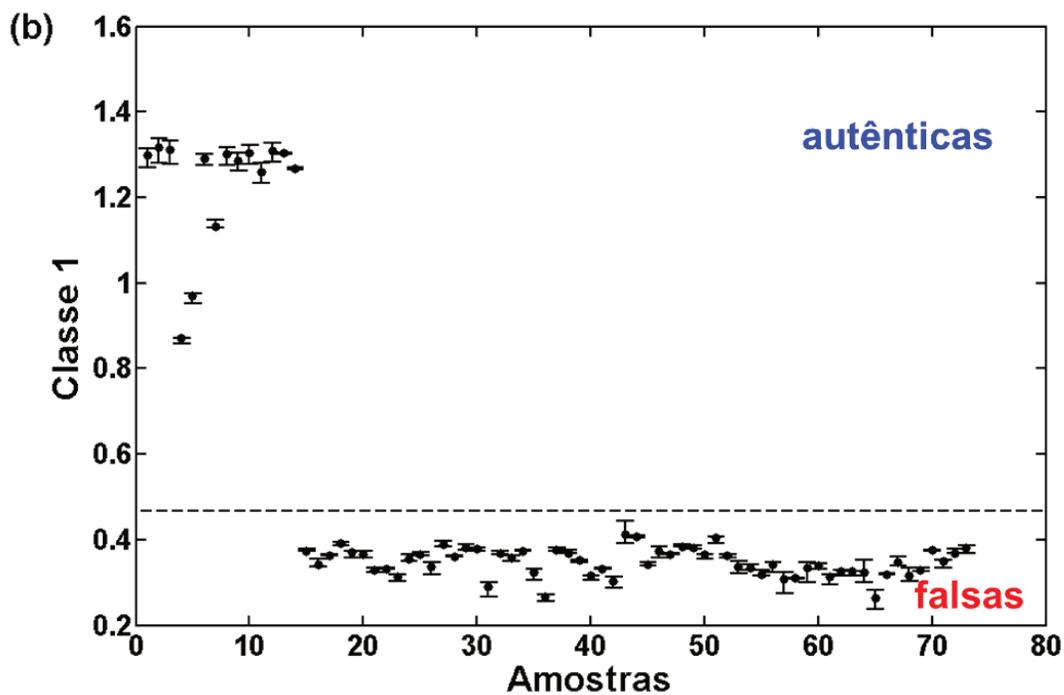
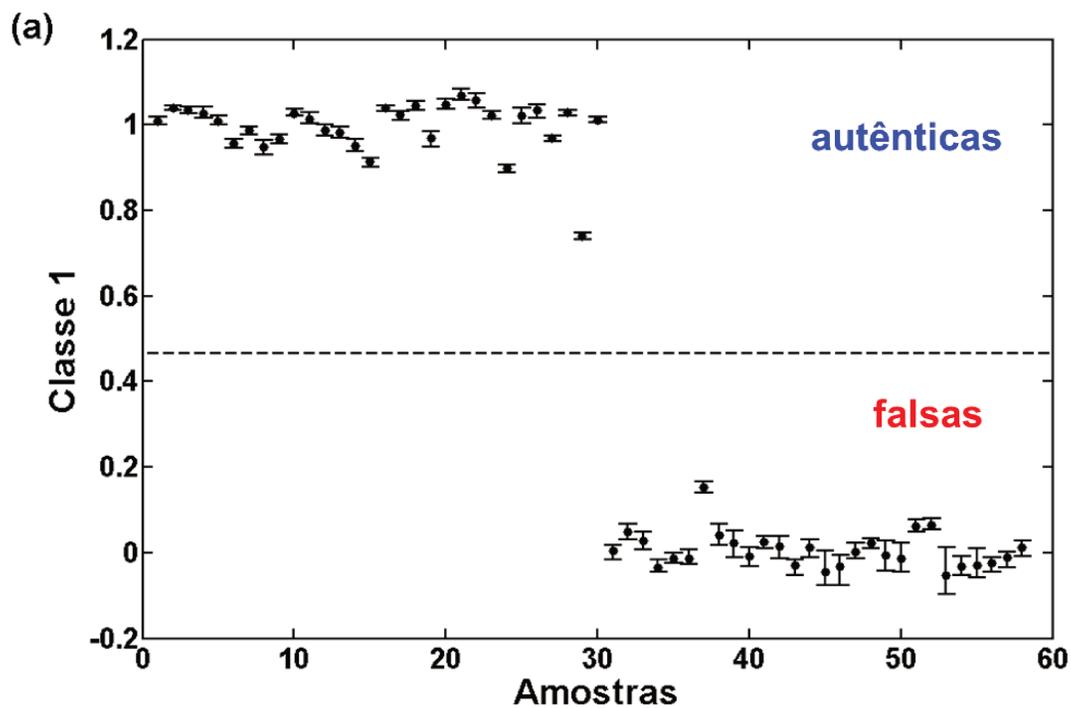


Figura 30: (a) Resultados do modelo de classificação para a tinta calcográfica e intervalo de confiança para o conjunto de calibração: notas autênticas e notas falsas fabricadas no laboratório; (b) Conjunto de validação externa e intervalo de confiança: notas autênticas e notas falsas apreendidas.

Os resultados para as áreas de coloração laranja e vermelha são apresentados nas Figuras 31 e 32, respectivamente. Todas as amostras foram corretamente classificadas para os conjuntos de calibração das duas áreas estudadas. Para a região de coloração vermelha observou-se que três amostras do conjunto de validação externa apresentaram intervalos de confiança que ultrapassaram o limite estabelecido entre as classes, essas amostras foram consideradas como amostras anômalas devido ao espectro Raman com alto ruído e não podem ser classificadas com confiança.

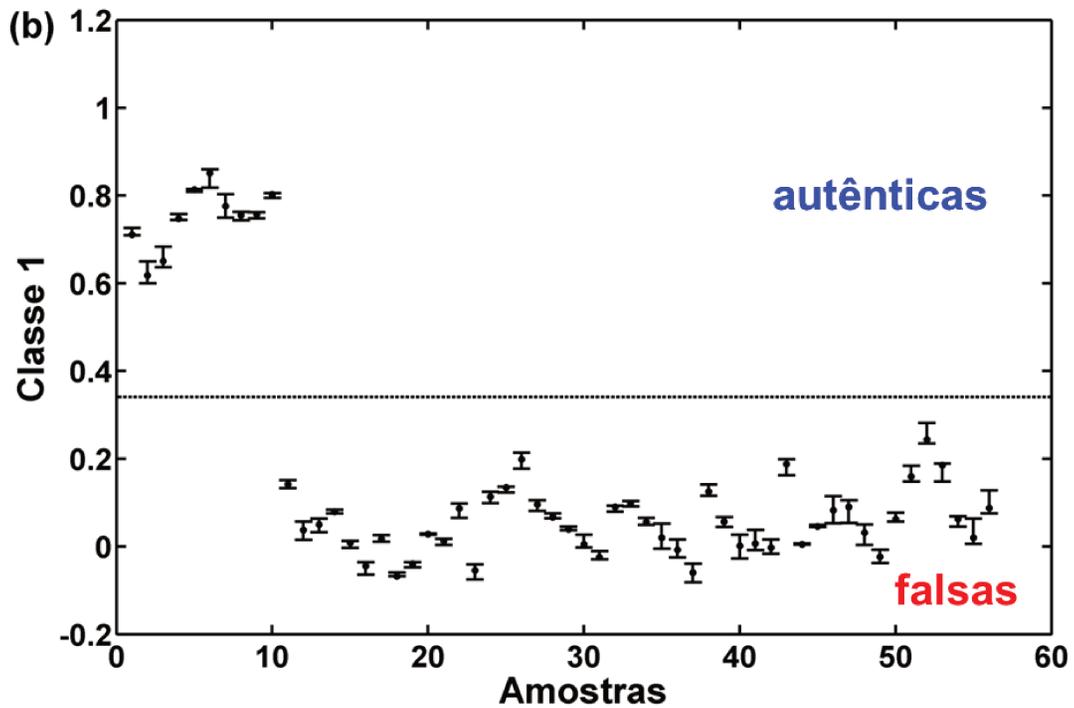
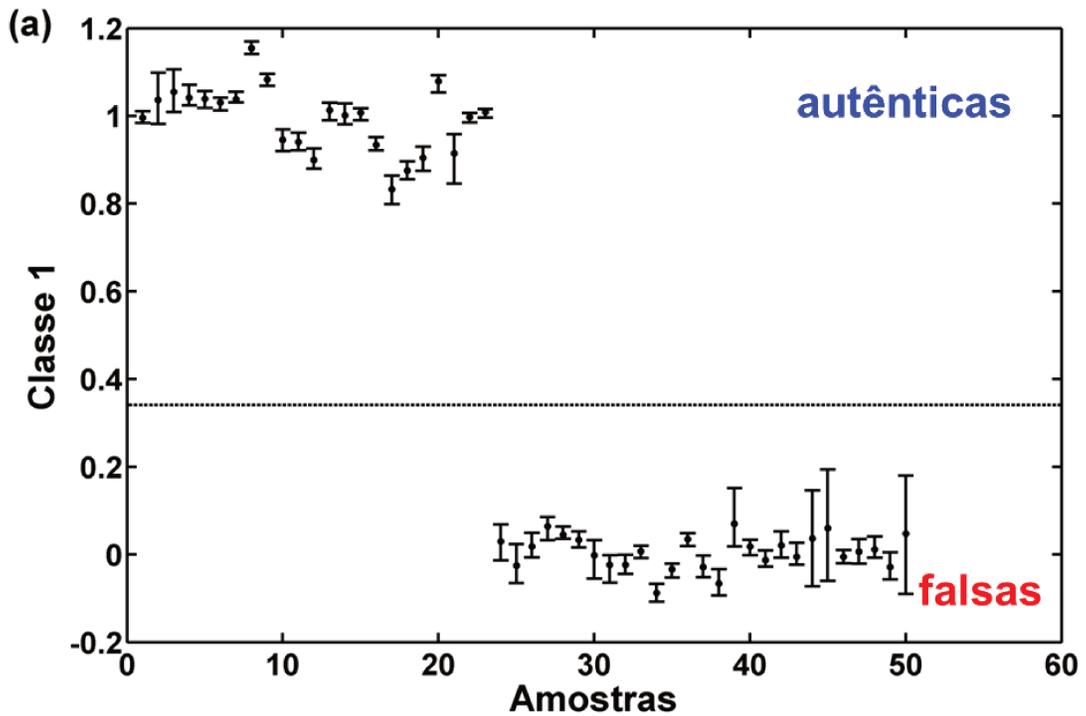


Figura 31: (a) Resultados do modelo de classificação para região de coloração laranja e intervalo de confiança para o conjunto de calibração: notas autênticas e notas falsas impressas no laboratório; (b) Conjunto de validação externa e intervalo de confiança: notas autênticas e notas falsas apreendidas.

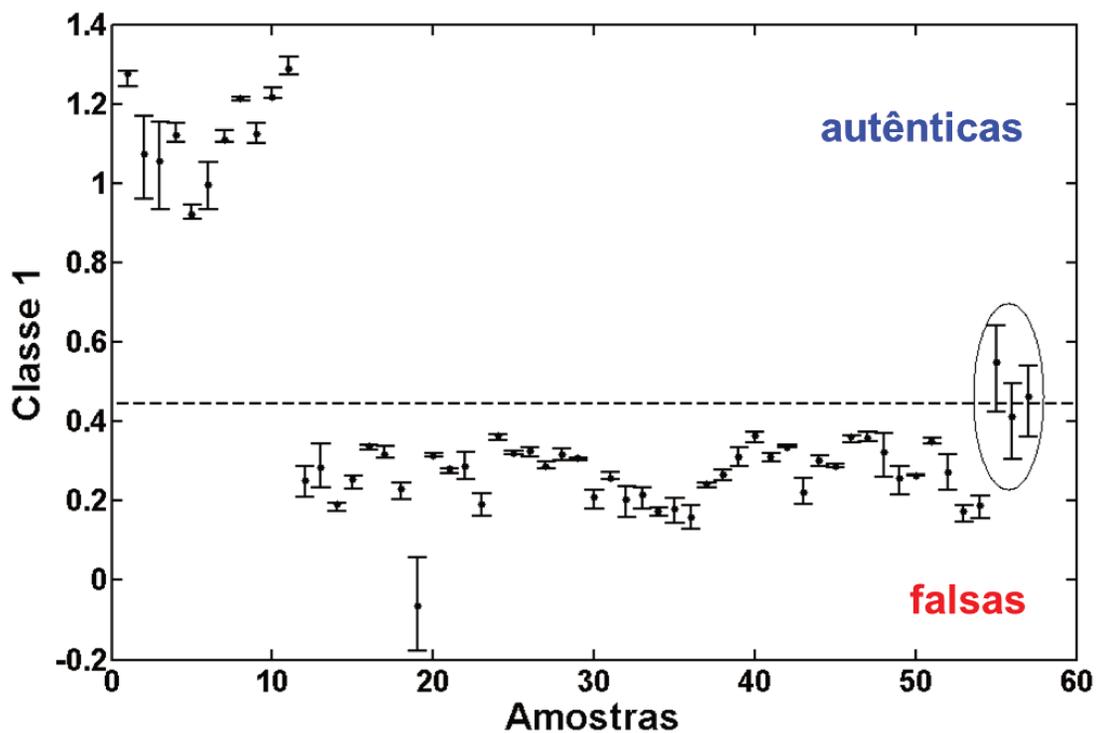
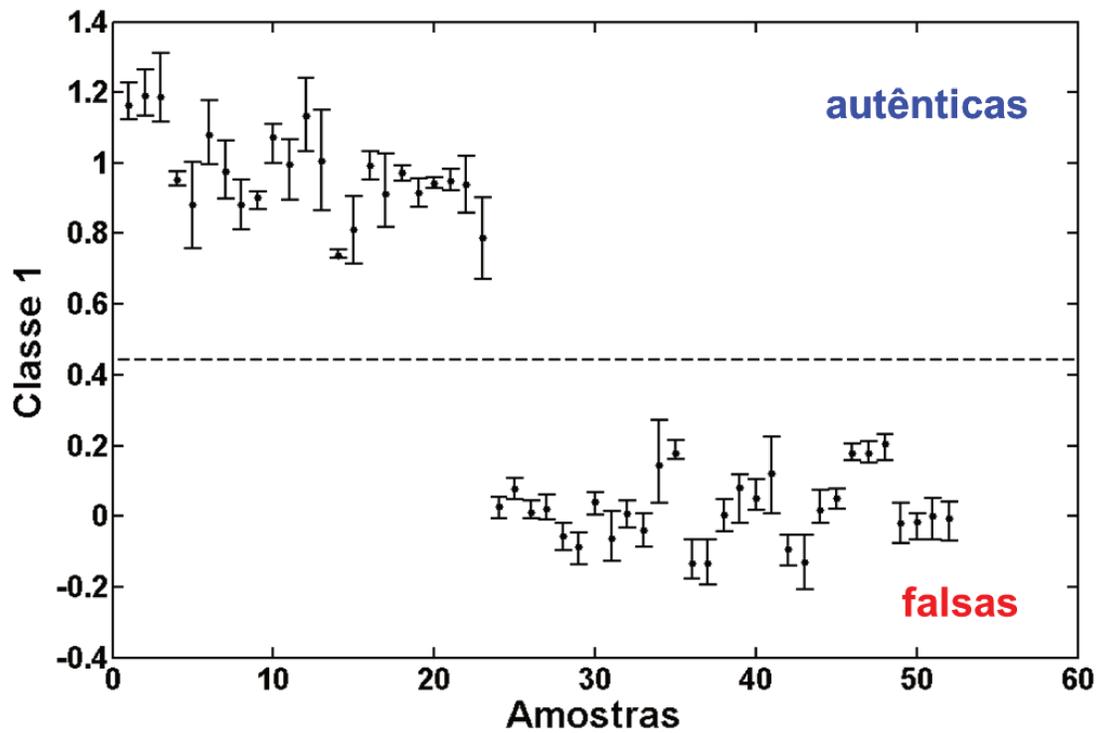


Figura 32: (a) Resultados do modelo de classificação para a tinta vermelha e intervalo de confiança para o conjunto de calibração: notas autênticas e notas falsas impressas no laboratório em papel alcalino; (b) Conjunto de validação externa e intervalo de confiança: notas autênticas e notas falsas apreendidas.

Comparando os modelos de classificação desenvolvidos para as três regiões estudadas, observa-se que os melhores resultados são obtidos para a região de tinta calcográfica e laranja. Uma justificativa para tal resultado é que a composição química dessas áreas e/ou a concentração usada nas cédulas autênticas é diferente das tintas empregadas em impressoras do tipo laser e jato de tinta usadas nesse estudo.

Após identificar se as notas eram autênticas ou falsas, um segundo modelo de classificação foi desenvolvido para classificar o tipo de impressora utilizado na fabricação das cédulas falsas. Para esse modelo, foi analisada a região de calcografia das cédulas. Nesse caso, o conjunto de calibração foi composto por 28 amostras de cédulas impressas em impressoras do tipo laser e jato de tinta. Foram selecionadas quatro variáveis latentes devido ao menor erro obtido na classificação. A Figura 33a mostra os resultados para as amostras de calibração. O valor limite entre as classe foi de 0,32, representado pela linha pontilhada no gráfico da Figura 33. Acima dessa linha as amostras são classificadas como impressas em impressoras do tipo laser e abaixo como impressoras jato de tinta.

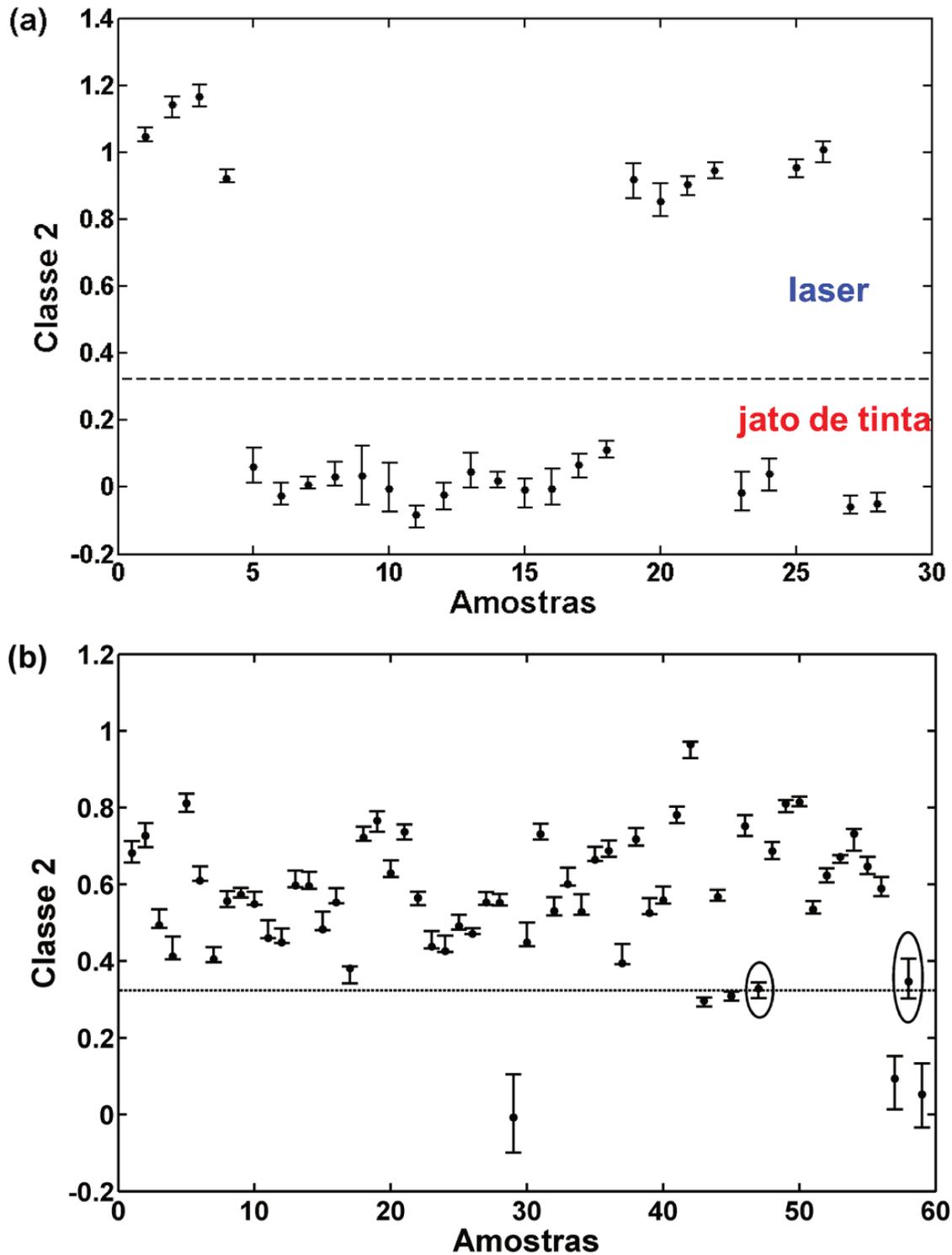


Figura 33: (a) Modelo de classificação e intervalo de confiança para o conjunto de calibração : notas impressas em impressoras laser e jato de tinta; (b) Conjunto de validação externa e intervalo de confiança: notas falsas apreendidas.

A estimativa da incerteza foi calculada pela técnica de reamostragem usando o *bootstrap* residual. Após a construção do modelo de classificação, as cédulas falsas apreendidas foram empregadas a fim de identificar o tipo de impressora usada na fraude. A maioria das cédulas foi classificada como fabricadas por impressora do tipo laser, esse tipo de impressora fornece uma impressão de melhor qualidade. Cinco amostras foram classificadas como originárias de impressora jato de tinta. Baseado no cálculo do intervalo de confiança para cada amostra, os limites de confiança superior e inferior foram na faixa de 0,029 e 0,0729, respectivamente. Os valores dos limites de confiança de duas amostras excederam o limite entre as classes, levando a incerteza na classificação das amostras. Esse resultado sugere que o tipo de pigmento ou o tipo de impressora dessas amostras não foi incluído no modelo de calibração ou devido a ruídos no espectro Raman dessas amostras levaram a uma incerteza na classificação.

5.1.5. Conclusões

Os modelos de classificação multivariados desenvolvidos permitiram distinguir as cédulas entre autênticas e falsas, e o tipo de impressora empregado na falsificação com base na informação química da tinta. As regiões analisadas da cédula que apresentaram melhores resultados foram a da tinta calcográfica e da tinta laranja, por apresentar perfil do espectro Raman da cédula autêntica diferente do perfil espectral das cédulas falsificadas nas amostras estudadas. Os modelos de classificação consideraram a incerteza nas previsões de cada amostra utilizando técnica de reamostragem *bootstrap*. Dessa forma foi possível classificar as amostras com avaliação da confiança e ainda identificar amostras anômalas.

O modelo de classificação desenvolvido com base no PLS-DA e nos espectros Raman pode ser empregado como um método complementar à inspeção judicial clássica. Os dados obtidos geram informação rápida, de forma não destrutiva e permitem a construção de um banco de dados.

5.2 Classificação do óleo essencial extraído da árvore amazônica *Aniba Rosaeodora* por espectroscopia Raman e PLS-DA com estimativa da confiabilidade

5.2.1 Introdução

A árvore amazônica *Aniba Rosaeodora Ducke*, popularmente conhecida como pau-rosa, fornece um óleo essencial valioso para a indústria de perfumes finos, sendo o perfume mais conhecido o Chanel nº 5. Desde o início da sua exploração, por volta de 1920, a extração do óleo é feita com a derrubada das árvores e a extração do óleo da madeira. Após décadas de exploração predatória, a espécie *Aniba Rosaeodora* encontra-se atualmente na lista de espécie de comércio controlado da Convenção Internacional sobre Espécies da Flora e Fauna ameaçadas de extinção⁹⁶. Isso significa um maior cuidado e fiscalização na comercialização do óleo essencial.

Em 2011, foi publicada uma instrução normativa que estabelece procedimentos e exigências para a exploração da espécie *Aniba Rosaeodora*⁹⁷. Dados coletados pelo Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA) mostraram que a diferença entre o volume de árvores autorizado para a exploração e a quantidade relativa de óleo essencial exportado era da ordem de 500%, mostrando irregularidades no processo de extração e comercialização do óleo essencial⁹⁶.

Segundo a legislação de crimes ambientais (Lei 9.605/98), nos artigos de 38 a 53 que contemplam crimes contra a flora, a extração, corte, aquisição, venda, exposição para fins comerciais de madeira, lenha, carvão e outros produtos de origem vegetal sem a devida autorização ou em desacordo com esta caracteriza-se crime ambiental. Dessa forma, a proposta de um método de identificação rápida da fonte do óleo essencial da espécie *Aniba Rosaeodora* encontra-se dentro do contexto forense. Nesse caso, a diferenciação entre a fonte dos óleos é importante para mostrar que o óleo possui uma origem legal.

Em 2004, Barata e colaboradores⁹⁸ publicaram um trabalho com alternativas sustentáveis para a produção do óleo essencial de pau-rosa, sendo que uma dessas

alternativas é a produção do óleo essencial usando as folhas e galhos, onde o óleo pode ser extraído sem a derrubada das árvores. Essas investigações relativas às fontes sustentáveis alternativas representam uma questão extremamente importante considerando a situação ambiental atual e a ameaça de extinção da espécie.

O óleo essencial de pau-rosa é economicamente relevante por ser uma valiosa fonte de linalol (3,7 dimetil-1,6-octadieno-3-ol), Figura 34. O linalol e seus ésteres, como o acetato de linalila, são compostos de cheiro intenso e agradável. Outros constituintes ocorrem em menor quantidade, incluindo os terpenos β -pineno, limoneno, terpinen-4-ol e α -terpineol⁹⁹. O linalol é um álcool terciário no qual é possível a existência de dois enantiômeros devido ao átomo de carbono assimétrico: (-)-linalol e (+)-linalol. No óleo de pau-rosa brasileiro podem ser encontradas ambas as formas em diferentes proporções.

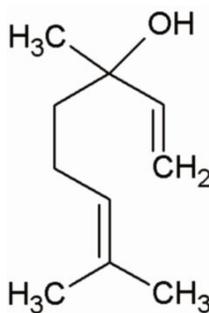


Figura 34: Estrutura molecular do linalol.

A maioria dos estudos encontrados na literatura é baseada na extração do óleo da madeira. Barata e colaboradores têm estudado o óleo essencial extraído das folhas e galhos^{98,99,100}. Em 2006, um trabalho empregando cromatografia gasosa com espectrometria de massas e cromatografia gasosa enantioseletiva acoplada à olfatométrica mostrou que o óleo extraído das folhas do pau-rosa pode substituir o óleo extraído da madeira, uma vez que apresenta perfil cromatográfico e qualidade olfativa similar ao linalol da madeira⁹⁹. Foi mostrado que o linalol apresenta-se em alta concentração no óleo essencial extraído da madeira (85%) e das folhas (81%). Em outro trabalho, a caracterização do óleo extraído das folhas de plantas jovens foi feito por cromatografia gasosa bidimensional, o estudo mostrou que as plantas jovens podem produzir óleo essencial com a qualidade requerida pela indústria com

rendimento razoável^{100,101}. Esses resultados mostram que é viável a extração do óleo essencial das folhas e galhos da árvore *Aniba Rosaeodora* para a indústria de perfumes.

A técnica de cromatografia gasosa é o principal método para estudo da composição de óleo essencial. No entanto, o uso da espectroscopia Raman tem ganhado espaço nesse campo, apresentando vantagens analíticas como alta velocidade, baixo custo, pouco ou nenhum resíduo gerado. Daferera e *et al*¹⁰² mostraram que os principais componentes dos óleos essenciais de algumas plantas da família *Lamiaceae* podem ser reconhecidos e diferenciados por espectroscopia FT-Raman. Em um trabalho de revisão¹⁰³, a espectroscopia Raman foi empregada para identificar e quantificar substâncias de interesse econômico das plantas. O trabalho mostrou que os principais constituintes podem ser identificados através das características espectrais exibidas sem a necessidade de uma separação física. O óleo essencial de diferentes espécies de eucalipto também foi analisado por espectroscopia Raman¹⁰⁴ sendo possível discriminar os óleos de diferentes espécies baseado nos seus componentes majoritários. Os trabalhos descrevem a espectroscopia Raman como uma ferramenta valiosa e rápida para a classificação química de amostras desconhecidas de óleo essenciais e para o controle de qualidade de óleo essencial de diferentes espécies e aromas.

Como já discutido, a maioria dos estudos que utilizam dados espectroscópicos e métodos quimiométricos com a proposta de classificação somente atribui o objeto a classe, poucos trabalhos na literatura avaliam a confiabilidade dos métodos de reconhecimento de padrões. Nesse trabalho, o método de reamostragem *bootstrap* foi empregado para estimar a incerteza na classificação da origem do óleo essencial de pau-rosa.

A espectroscopia Raman e o método supervisionado de classificação PLS-DA foram empregados para a caracterização e identificação da fonte do óleo essencial da árvore *Aniba Rosaeodora*. Um modelo de classificação foi construído para distinguir o óleo procedente do caule, das folhas e galhos. Os intervalos de confiança foram estimados para cada amostra classificada. O desempenho do modelo de classificação foi avaliado através de tabelas de contingência, parâmetros como razão de falso

positivo, falso negativo, sensibilidade, especificidade e coeficiente de correlação de Matthew foram calculados.

5.2.2 Parte Experimental

5.2.2.1 Amostras

As amostras de óleo essencial utilizadas neste trabalho foram provenientes de dois estados brasileiros: Amazonas e Pará. O conjunto de amostras analisado foi composto por vinte amostras de óleo extraído do caule, sessenta e cinco amostras de óleo de folhas, doze amostras de óleo de ramos e seis amostras de óleo extraído das folhas e ramos. As medidas foram realizadas diretamente nos recipientes de vidro que continham as amostras, sem retirar a amostra do seu interior.

5.2.2.2 Espectros Raman

Os espectros Raman do óleo essencial foram coletados em um espectrômetro dispersivo Raman Station 400F Perkin Elmer equipado com um detector CCD e um laser diodo com comprimento de onda de excitação em 785 nm com potência máxima de 250 mW na fonte. Os espectros foram obtidos com 3s de exposição ao laser e 20 acumulações na região de 3200 a 200 cm^{-1} com resolução espectral de 4 cm^{-1} . Cada espectro foi obtido em duplicata, foram feitas correções automáticas da linha base e subtração do sinal Raman do vidro.

5.2.3 Tratamento dos dados

Para remover variações sistemáticas indesejáveis nos dados, vários pré-processamentos foram testados, tais como métodos de normalização, derivadas, filtros

de alisamento e correção ortogonal do sinal (OSC, do inglês *orthogonal signal correction*). O método PLS-DA foi aplicado para a classificação. O modelo PLS-DA foi calculado a partir dos algoritmos da regressão por mínimos quadrados parciais (PLS). Para a construção do modelo os espectros Raman (matriz **X**) foram divididos em dois conjuntos: conjunto de calibração e conjunto de validação. Um vetor **Y** foi criado com zero para os óleos extraídos das folhas e ramos e um para os espectros dos óleos extraídos do caule. O limite entre as classes foi calculado por Teorema de Bayes através da função *plsdthres* do software PLS toolbox⁹¹. O cálculo do intervalo de confiança para cada amostra foi feito com o *bootstrap* residual.

5.2.4 Resultados e Discussões

O linalol é o principal componente do óleo essencial da espécie *Aniba Rosaeodora* encontrada no Brasil⁹⁹. Devido à porcentagem de linalol contida no óleo essencial (70-90%), seu espectro Raman apresenta o perfil característico desse composto, a Figura 35 mostra os espectros Raman do óleo essencial extraído do caule, das folhas e dos ramos.

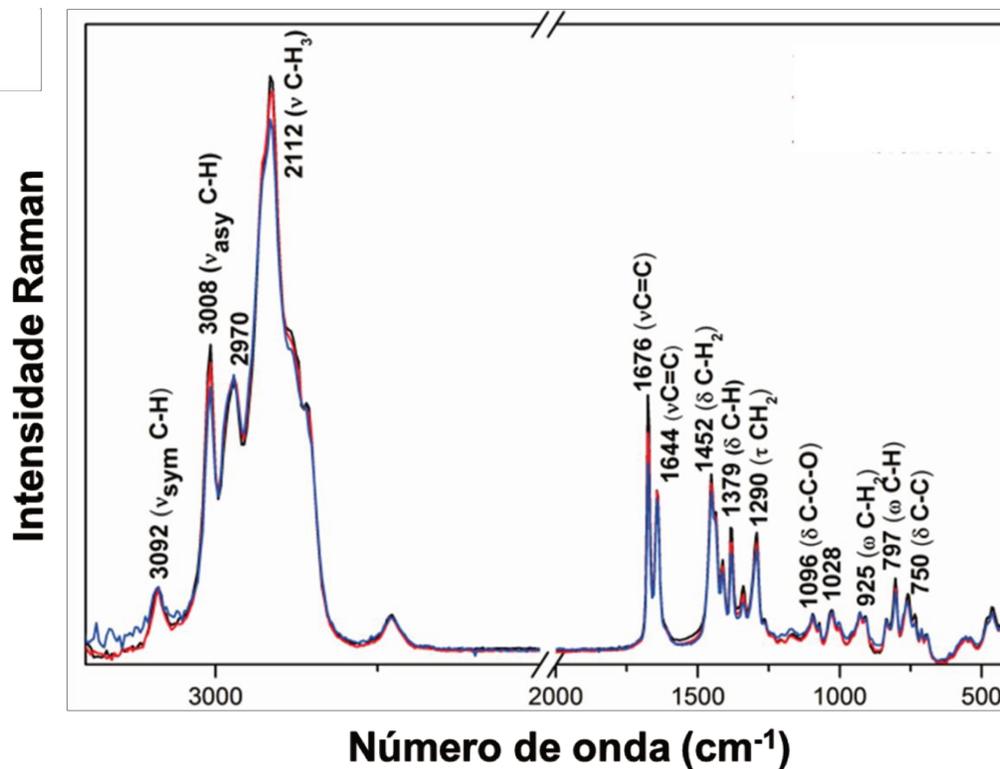


Figura 35: Espectro Raman do óleo essencial extraído do caule (-); das folhas (-) e dos ramos (-).

A tentativa de atribuição das principais bandas vibracionais foi baseada em dados da literatura^{102,103,104,105}. A molécula de linalol contribui com um intenso sinal na região de 3000 cm^{-1} devido ao estiramento das ligações C-H. A presença de uma banda de baixa intensidade em 3092 cm^{-1} é atribuída ao estiramento simétrico =C-H do grupo vinil. O estiramento assimétrico C-H é observado em 3008 cm^{-1} . Essa banda vibracional encontra-se deslocada para maior frequência devido a ligação dupla C=C. A banda presente em 2912 cm^{-1} está associada com a vibração C-H do grupo metileno. No entanto, não é interessante usar essa região para identificação da origem do óleo, uma vez que a maioria dos compostos terpenos possuem bandas intensas nessa região.

A região entre $1800\text{ a }400\text{ cm}^{-1}$ é rica em informação estrutural e vários trabalhos utilizam esta região para caracterização de diferentes óleos essenciais^{102,103,104,105}. Os espectros Raman obtidos para as amostras de óleo essencial mostram uma banda intensa em 1676 cm^{-1} devido ao estiramento da ligação C=C dissubstituída, e uma banda de menor intensidade em 1644 cm^{-1} atribuída ao estiramento da dupla C=C

monossubstituída. A quantidade de linalol presente no óleo e a presença de outros terpenos podem causar mudanças nas intensidades dessas duas bandas as quais podem ser usadas para diferenciação entre óleos essenciais. As outras bandas presentes na região de impressão digital são descritas como vibrações de deformação de ângulo dos grupos CH₂ e CH₃.

Comparando os espectros Raman do óleo essencial extraído do caule, das folhas e dos ramos, mostrado na Figura 35, não é possível observar diferenças visuais nos perfis dos espectros. Dessa forma, o uso de ferramentas quimiométricas foi necessário para a construção do modelo de classificação para o óleo essencial extraído de diferentes partes da árvore *Aniba Rosaeodora*.

Primeiramente foi realizada uma análise exploratória dos dados empregando a PCA, que é um dos primeiros tratamentos realizados em um conjunto de dados para encontrar agrupamentos naturais, reduzir a dimensão dos dados e obter somente informações relevantes. Nesse trabalho a PCA foi realizada com 103 espectros Raman do óleo essencial do caule, das folhas e ramos na faixa espectral de 3200 a 200 cm⁻¹, obtendo uma matriz de dados de 103x770 (amostras x variáveis). Seis componentes principais explicando 86,9% da variância total explicada foram selecionados. Na Figura 36 é apresentado o gráfico de escores do modelo PCA com a projeção de PC1 x PC2 x PC3.

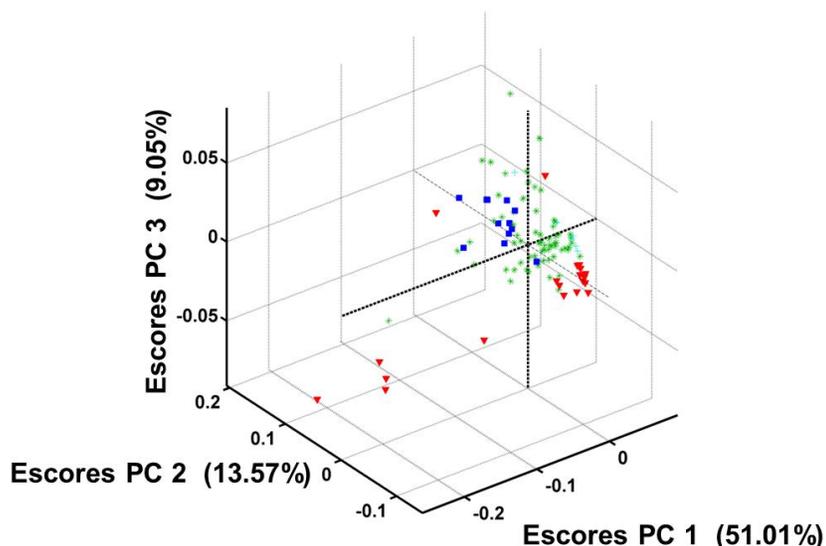


Figura 36: Gráfico dos escores da PC1 x PC2 x PC3 dos óleos extraídos do caule (▼), das folhas (*) e dos ramos (■).

A análise dos escores da PCA não mostrou a formação clara de agrupamentos, nos quais poderiam ser usados para distinguir a fonte do óleo. Observou-se que seis amostras ficaram afastadas. Os espectros Raman dessas amostras foram avaliados separadamente e comparando-se com os espectros de outras amostras notou-se uma mudança nas intensidades de algumas bandas. Houve um aumento de intensidade das bandas em 2978 e 1644 cm^{-1} e uma diminuição na banda em 3008 e 1676 cm^{-1} . Essas mudanças de intensidade podem indicar um processo de oxidação dos óleos essenciais, principalmente do composto linalol ou a diminuição desse composto associado com a região de cultivo, idade da árvore e época de extração do óleo essencial. Apesar de não apresentar uma tendência para separação das amostras, a análise da PCA mostrou a presença de amostras anômalas que foram retiradas do conjunto de dados antes da construção do modelo de classificação.

Analisando os gráficos de *loadings* da primeira componente principal, as regiões de 3000 e 1650 cm^{-1} apresentaram maior peso para a distribuição das amostras no gráfico de escores corroborando com a análise anterior. Apesar de não ser possível fazer a distinção entre as fontes de óleo, a PCA possibilita uma análise exploratória das

amostras e pode ser empregada como metodologia de controle de qualidade para identificação de amostras anômalas.

A grande variabilidade das amostras explica a necessidade de muitos componentes principais para explicar uma variância significativa, mesmo que, a princípio, os dados sejam compostos por três tipos de amostras. A PCA não forneceu um resultado satisfatório em relação à fonte do óleo, mesmo com a retirada das amostras anômalas. Dessa forma, é necessário o emprego de um método supervisionado de classificação, em que se deve ter conhecimento inicial sobre as classes que irão ser modeladas.

Para a análise discriminante com o PLS o conjunto de dados, agora composto por 97 amostras, foi dividido em dois subconjuntos: de calibração (3/4 das amostras) e de validação (1/4 das amostras). A partir do conjunto de calibração foram criadas as regras de classificação e a validação do modelo foi feita com um conjunto externo de amostras. O conjunto de calibração foi composto de 14 amostras de óleo extraído do caule ($Y=1$), 44 amostras de óleo das folhas e 8 amostras de óleo dos ramos ($Y=0$). A dimensionalidade do modelo foi determinada por validação cruzada empregando as amostras de calibração. O número de variáveis latentes foi selecionado usando o gráfico de RMSECV versus o número de variáveis latentes.

Devido à variabilidade inerente das amostras, que contém variações de muitas fontes e de diferentes tipos, seria interessante o uso de métodos de pré-processamento para separar a informação relevante e obter modelos mais simples e fáceis para interpretação. Nesse caso, a correção ortogonal do sinal (OSC, do inglês, *orthogonal signal correction*) aparece como uma opção para remover a variação sistemática na matriz \mathbf{X} que não está correlacionada com o vetor \mathbf{Y} . A OSC foi empregada ao conjunto de dados, no entanto, foi observado um sobreajuste do modelo de classificação. Quando o conjunto de validação externa (conjunto teste) foi aplicado ao modelo, este não foi capaz de classificar corretamente as amostras.

A opção de pré-processamento que forneceu um menor erro de classificação foi a normalização com vetor unitário, segunda derivada com alisamento *Savitzky-Golay* e dados centrados na média para a matriz \mathbf{X} . Os resultados do modelo PLS-DA para as amostras de calibração e validação são mostrados na Figura 37.

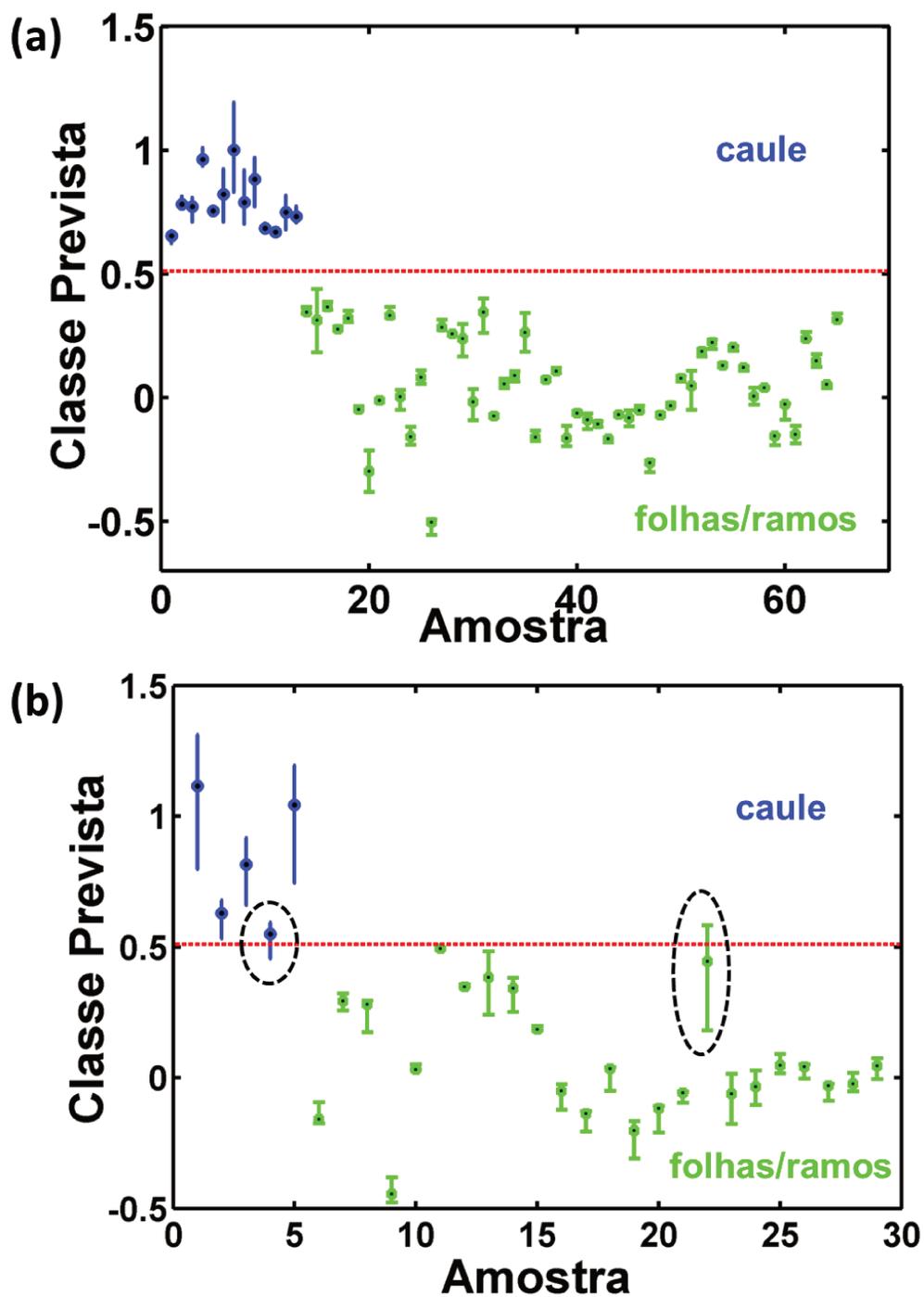


Figura 37: Resultado do modelo do PLS-DA para as amostras de calibração (a) e validação (b). As amostras acima da linha vermelha são classificadas como óleo extraído do caule (•) e abaixo como óleo extraído dos ramos e folhas (*). As barras representam o intervalo de confiança de 95% de cada amostra calculado por *bootstrap*.

O valor limite calculado que separa as classes é representado pela linha horizontal na Figura 37, com valor de 0,53 sendo que acima desse valor as amostras são classificadas como óleo extraído do caule e abaixo como óleo oriundo das folhas e dos ramos.

O cálculo da confiabilidade para modelos de classificação é de bastante interesse para conhecer a incerteza da atribuição de uma amostra a determinada classe. Com base nessa proposta, a técnica de *bootstrap* residual foi empregada. O número de *bootstrap* foi otimizado para obter o menor desvio padrão e resultados reprodutíveis; dessa forma foram testados 500, 1000 e 2000 reamostragens e foi observado que após 1000 reamostragens os valores dos desvios padrões variam muito pouco. Assim, os cálculos de intervalo de confiança para cada amostra foram feitos com 1000 reamostragens. Todas as amostras foram consideradas como uma nova observação e os intervalos de 95% de confiança foram computados. As barras de erros mostradas na Figura 37 são os intervalos de confiança para a classificação de cada amostra e permitem uma melhor avaliação do modelo. As amostras destacadas no conjunto de validação apresentam resultados inconclusivos, uma vez que a incerteza na classificação ultrapassa o limite entre as classes. Nesse caso, o modelo de classificação não pode classifica-las corretamente.

O desempenho do modelo de classificação foi avaliado empregando tabelas de contingência para determinação das taxas de falsos resultados, sensibilidade, especificidade, eficiência e cálculo do coeficiente de correlação de Matthew. Os resultados mostraram similaridades entre os parâmetros do conjunto de calibração e o conjunto de validação, indicando que não ocorreu sobre ajuste do modelo desenvolvido. Para o conjunto de calibração nenhuma previsão errada foi observada. Considerando as incertezas na classificação das amostras do conjunto de validação, a taxa de falso resultados foi de 6,5%.

Alguns parâmetros podem contribuir para classificações erradas, como por exemplo, a região que limita as classes. A escolha errada do valor limite entre as classes pode levar a erro do tipo 1 (resultados falso positivos) e erro do tipo 2 (resultados falso negativos). Ao aumentar o valor limite entre as classes a probabilidade de obter resultados falso positivos, ou seja, de classificar amostras de óleo extraído das

folhas e galhos como óleo extraído do caule é reduzida, mas aumenta-se a chance de resultados falso negativos, como classificar amostras de óleo extraído do caule como óleo extraídos das folhas. Em um teste de *screening*, uma resposta positiva é confirmada por um método de referência, por exemplo, cromatografia gasosa, mas respostas negativas são consideradas diretas para diminuir custo e tempo de análise. Dessa forma, é preferível que a probabilidade dos resultados falso negativos seja menor. Neste trabalho é importante evitar os resultados falso negativos, uma vez que classificaria óleo extraído do caule como óleo extraído de folhas e ramos.

Para o conjunto de calibração nenhuma resposta errada foi observada, no entanto, a taxa de falso positivo para o conjunto de validação foi 4% e de falso negativo de 16%. Outro parâmetro que podemos avaliar é a estatística das respostas verdadeiras, nesse caso, a sensibilidade e especificidade. A especificidade do conjunto de validação foi de 96%, essa foi a probabilidade de uma amostra de óleo das folhas e galhos serem classificadas corretamente. A sensibilidade do método foi 83,3%, que indica a porcentagem das amostras de óleos de caule classificados corretamente.

A eficiência de um modelo de classificação combina todas as informações fornecidas pela sensibilidade e especificidade. Quando o método é muito sensível para positivo gera muitos falso positivos, e vice versa, dessa forma é interessante avaliar a eficiência do modelo, que nesse caso foi de 89,7%. O coeficiente de correlação de Matthew também fornece um valor único na avaliação do desempenho do modelo de classificação. O valor desse coeficiente para o conjunto de validação foi de 0,79, no qual pode ser considerado como um bom desempenho.

Para construir os modelos de classificação foram usadas as informações fornecidas pelos espectros Raman. A contribuição das bandas Raman para a classificação das amostras foi avaliada pelo gráfico de *loadings* das variáveis latentes mostrado na Figura 38. As principais regiões do espectro Raman que contribuíram para a diferenciação entre as fontes de óleo foram 1650, 1450, 1000 e 700 cm^{-1} .

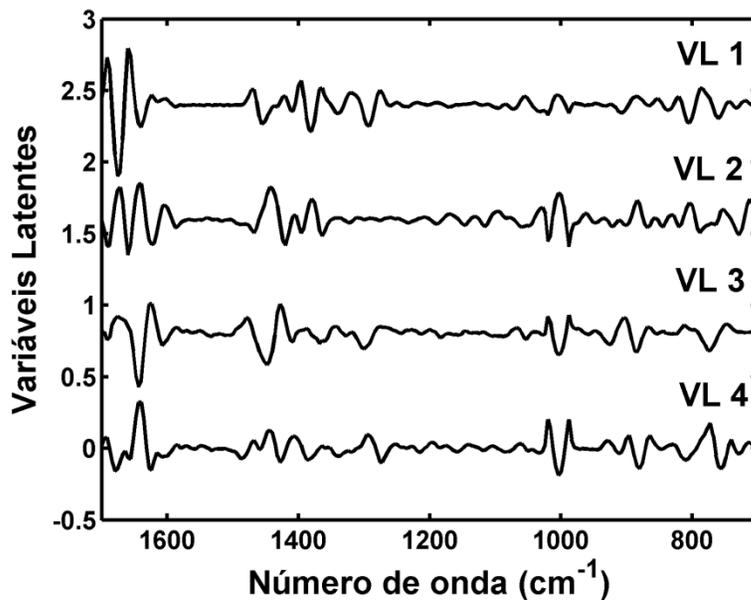


Figura 38: Gráfico dos *loadings* da primeira, segunda, terceira e quarta variável latente versus as variáveis (bandas Raman) para o modelo PLS-DA.

Outro método usado para avaliar a importância das variáveis é a análise do gráfico VIP (do inglês, *Variable Importance in the Projection*). A Figura 39 mostra os escores VIP para o modelo PLS-DA, os valores $VIP > 1$ mostram a região espectral que pode ser considerada importante para o melhor desempenho do modelo de classificação. As regiões espectrais que apresentaram valores de VIP maiores que 1 foram as regiões de 1650, 1450, 1000 e 750 cm^{-1} .

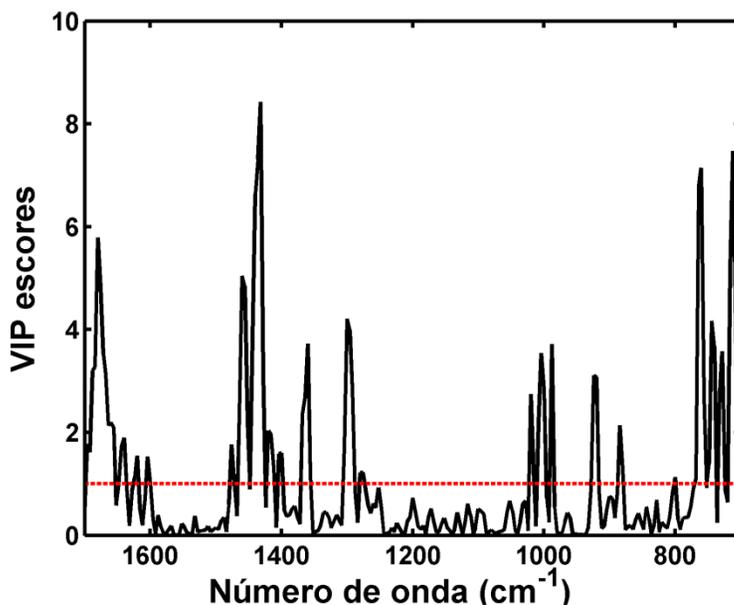


Figura 39: Escores do VIP para o modelo PLS-DA.

Esses resultados sugerem que a diferença entre as fontes do óleo é devido a variação das quantidades dos terpenos presentes, especialmente, o linalol. Segundo Zellner *et al*⁹⁹, o óleo extraído das folhas é caracterizado por uma maior concentração de sesquiterpenos em relação ao óleo do caule, que contém uma maior quantidade de monoterpenos. Os sesquiterpenos apresentam bandas Raman na região de 1650, 1450 e 750 cm^{-1} , e essas regiões mostraram maior contribuição para a discriminação da fonte dos óleos na análise dos *loadings* e do gráfico VIP.

5.2.5. Conclusões

A metodologia proposta baseada na caracterização do óleo essencial de pau-rosa extraído de diferentes partes da árvore *Aniba Rosaeodora*, empregando espectroscopia Raman e PLS-DA permite distinguir entre os óleos oriundos do caule do óleo das folhas e galhos. O modelo foi validado com um conjunto de amostras testes e a avaliação do seu desempenho foi feita através da tabela de contingência, parâmetros como falso positivo, falso negativo, sensibilidade, especificidade, acurácia, eficiência e coeficiente de correlação de Matthew foram calculados. Os resultados mostraram que o

modelo de classificação é aplicável para a finalidade que se destina, com 96,5% das amostras classificadas corretamente. Sendo uma metodologia adequada para triagem das amostras. A incerteza da classificação de cada amostra foi estimada usando métodos de reamostragem *bootstrap*, dando maior confiabilidade aos resultados.

Os resultados apresentados mostram que a espectroscopia Raman com o auxílio da ferramenta quimiométrica PLS-DA tem potencial para complementar os procedimentos de referência usados para análise de óleos essenciais, como cromatografia gasosa, oferecendo vantagens como alta velocidade de análise, menores custos e pouca geração de resíduos, uma vez que não é necessário um tratamento de amostras antes das análises.

A classificação pode ser empregada como um método rápido e automatizado para distinguir os óleos extraídos de diferentes partes da árvore amazônica *Aniba Rosaeodora*. A metodologia é rápida, simples e ainda oferece a possibilidade de análise *in situ* usando instrumentos portáteis.

**CAPÍTULO 6: Aplicação da Espectroscopia Raman de
imagem e métodos de resolução multivariada de curvas em
estudos forenses e paleontológicos**

6.1 Metodologia para detecção de explosivos em superfícies de cédulas por espectroscopia Raman de imagem e métodos de resolução de curvas

6.1.1 Introdução

O estudo de explosivos e seus resíduos em diferentes superfícies é de grande interesse para a área de química forense, de segurança nacional e ambiental¹⁰⁶. O interesse em análises forenses é devido à utilização desses artefatos em ações criminosas. Uma substância é considerada um explosivo se produz por reação grande volume de gás e calor em um curto espaço de tempo. Os explosivos podem ser uma combinação de componentes baratos e de fácil obtenção, que podem ser facilmente transportados, escondidos e possuem um alto poder de destruição.

A análise dos explosivos para fins forenses é feita para identificar o material que constitui um dispositivo de explosão apreendido, sua origem e vestígios em peças de perícia. A identificação é feita com o propósito de estabelecer ligações entre suspeitos e cenas do crime. Geralmente, a análise para detecção de explosivos é feita pela identificação de nitrito, um indicativo da presença desses compostos. O método mais utilizado é o teste por via úmida de Gries¹⁰⁷. No entanto, o teste identifica nitrito de qualquer procedência e portanto, os resultados nem sempre são conclusivos.

A identificação de explosivos utilizando técnicas espectroscópicas vibracionais, como infravermelho e Raman têm sido discutida na literatura¹⁰⁶. Ali e colaboradores fizeram análise *in-situ* de explosivos presos em fibras de tecidos sem tingir e tingidas utilizando Raman confocal e observaram diferenças espectrais em ambas as fibras¹⁸. Embora os resultados tenham permitido a identificação com respeito à aplicação espectroscópica, este tipo de amostra caracteriza-se por estar em quantidades microscópicas presas nas fibras e segundo os autores, a dificuldade de encontrá-las dentro dos tecidos foi significativa.

O uso de espectrômetros Raman portáteis utilizando laser na região de 785 e 1064 nm tem sido empregado em estudos forenses, sendo de grande relevância, uma vez que alguns tipos de aplicações requerem esse tipo de tecnologia devido à presença

de fluorescência. Sharma e colaboradores trabalharam com um equipamento portátil para monitorar hidrocarbonetos e explosivos no ambiente¹⁰⁸. Métodos para detectar e analisar artefatos a base de peróxido também tem sido discutidos na literatura¹⁰⁹.

Nos últimos anos, a detecção *stand-off* por espectroscopia Raman em explosivos tem sido o foco das pesquisas nesse campo¹¹⁰. Para detectar o sinal Raman na presença de interferentes de fundo a grandes distâncias, os sistemas Raman *stand-off* utilizam lasers pulsados para intensificar o sinal nos detectores CCD. Nesse caso, o uso de metodologias de processamento e resolução de sinais é requerido para reduzir a interferência de fundo e separar o sinal do analito do sinal dos interferentes.

O uso da espectroscopia Raman de imagem para identificação de explosivos também tem sido mostrado em alguns trabalhos da literatura. Emmons e colaboradores¹¹¹ aplicaram imagem Raman para detectar quantidades traço de explosivos em impressões digitais. Em outro trabalho do mesmo grupo, espectros Raman de explosivos foram registrados em materiais fortemente espalhadores, como poliestireno e policarbonato e foi empregado um algoritmo para subtração de fundo. Os trabalhos mostraram que a imagem Raman é uma poderosa ferramenta na detecção de explosivos em quantidades traço, mesmo na presença de interferentes e sem destruição da impressão, deixando-a disponível para outras análises.

O objetivo desse estudo foi desenvolver uma metodologia utilizando espectroscopia Raman de imagem e métodos quimiométricos de resolução de curvas para identificação de explosivos em superfície de cédulas. O uso da quimiometria se justifica devido a grande quantidade de dados gerados, baixa resolução espectral e baixa intensidade do sinal dos compostos de interesse. Nesse trabalho, métodos de resolução de curvas como a resolução multivariada de curvas (MCR) e a análise de componentes independentes (ICA) foram empregados para extrair informação das imagens Raman.

A motivação desse trabalho foi o aumento de ataques a caixas eletrônicos utilizando substâncias explosivas para o roubo de cédulas. De acordo com relatórios da Pesquisa Nacional de ataques a bancos, de 2011 para 2014 houve um aumento de 102% nos arrombamentos a caixa eletrônicos¹¹². No primeiro semestre de 2011 foram registrados 838 casos contra 1693 no ano de 2014, como pode ser observado na

Figura 40. O desenvolvimento de métodos analíticos inequívocos que possam identificar explosivos em diferentes superfícies podem ajudar a rastrear a origem do explosivo e contribuir para a investigação criminal.

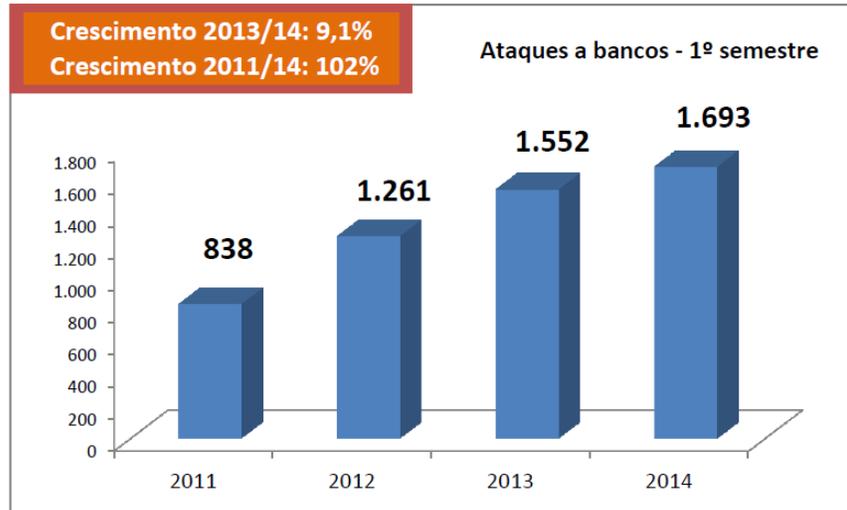


Figura 40: Estatística de ataques a bancos no Brasil entre 2011 e 2014. Fonte: Notícias da imprensa, secretaria de segurança pública e sindicatos¹¹².

O estudo foi conduzido nas seguintes etapas: primeiramente foram obtidos os espectros Raman dos explosivos 2,4,6-trinitrotolueno (TNT), tetranitrato de pentaeritritol (PETN), trinitrofenol ou ácido pícrico (TNP), ciclo-1,3,5-trimetileno-2,4,6-trinitramina (RDX), ciclo-1,3,5,7-tetrametileno-2,4,6,8-tetranitramina (HMX), nitrato de amônio em óleo combustível (ANFO), perclorato de amônio (PERC) e pólvora negra composta por carbono, enxofre e nitrato. As estruturas químicas dos explosivos estudados estão representadas na Figura 41.

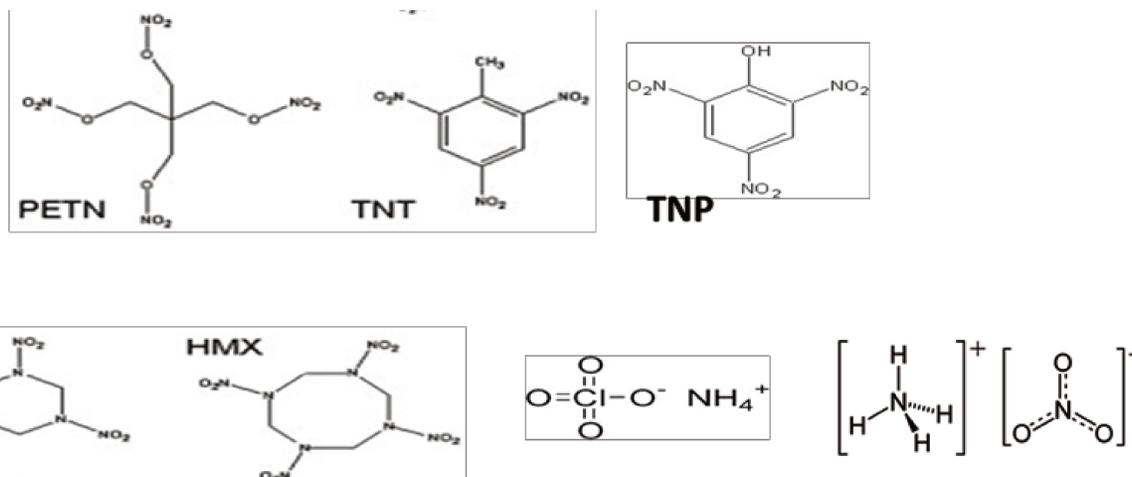


Figura 41: Estrutura química dos explosivos estudados

Em seguida, as imagens Raman das cédulas contaminadas com explosivos no laboratório foram obtidas e ferramentas quimiométricas de resolução de curvas foram aplicadas com o objetivo de extrair o espectro puro do explosivo, sua distribuição na superfície analisada e o limite de detecção da metodologia. Os métodos empregados foram a resolução multivariada de curvas (MCR) e a análise de componentes independentes (ICA). O desempenho dos métodos foi avaliado em termos da variância explicada pelo modelo, da quantidade de falta de ajuste e dos coeficientes de correlação entre os perfis espectrais recuperados e os espectros Raman dos explosivos puros. A quantidade de ambiguidade rotacional das soluções do MCR e do ICA foi avaliada pelo método MCR-BANDS. Por último, a metodologia foi aplicada à amostras forenses, na tentativa de encontrar explosivos e seus resíduos em cédulas suspeita de serem de explosões de caixa eletrônico.

6.1.2. Parte experimental

6.1.2.1 Amostras

As amostras dos explosivos TNT, PETN, TNP, RDX, HMX, ANFO, perclorato de amônio (PERC) e pólvora negra (BP) foram fornecidas pela Polícia Federal. Cédulas

foram contaminadas em laboratório com uma pequena quantidade de explosivo que foi espalhada em uma determinada área da cédula e a região foi mapeada usando espectroscopia Raman. Dois tipos de amostras forenses foram analisadas por espectroscopia Raman de imagem. Cédulas suspeitas de serem provenientes de explosão de caixa eletrônico, mostradas na Figura 42, foram fornecidas pela Superintendência da Polícia Técnico-Científica do Estado de São Paulo.



Figura 42: Cédulas suspeitas de serem de explosão de caixa eletrônico

A segunda amostra forense foi obtida de uma explosão simulada realizada por uma equipe forense da Polícia Federal. Para a simulação, a cédula foi inserida dentro de uma caixa metálica com 7,3g do explosivo ANFO e 1mL do corante utilizado como dispositivo de segurança. A Figura 43 mostra a cédula antes e após a explosão.

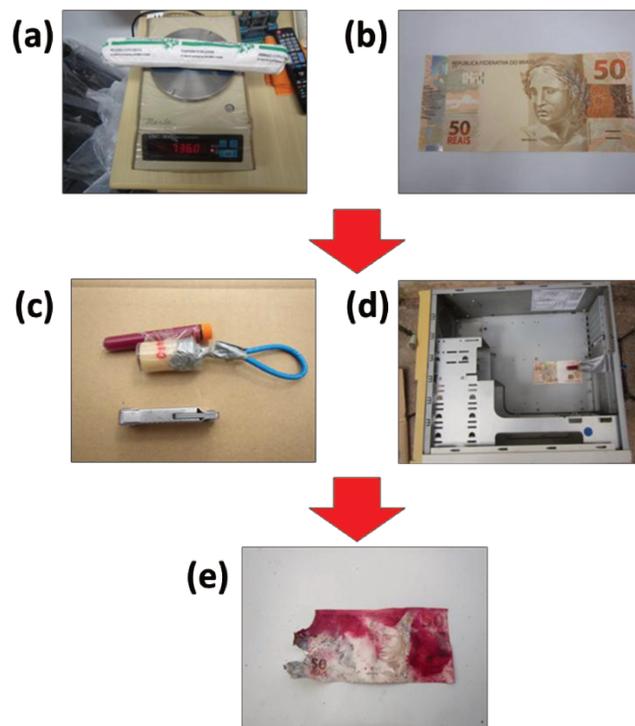


Figura 43: Explosão simulada de caixa eletrônico. Explosivo ANFO (a); Cédula de R\$ 50 antes da explosão (b); explosivo juntamente com o dispositivo de segurança usado para manchar as cédulas (c); caixa metálica simulando caixa eletrônico (d); cédula de R\$ 50 pós explosão.

6.1.2.2. Imagens Raman

Os espectros e as imagens Raman foram obtidos em um espectrômetro dispersivo RamanStation da PerkinElmer, equipado com detector CCD e com laser de excitação em 785 nm. Os parâmetros do equipamento foram otimizados de forma a obter a melhor relação sinal-ruído para os espectros. A potência do laser foi ajustada para 30%, o que corresponde uma potência nominal de aproximadamente 30 mW chegando na amostra, a resolução espectral foi de 4 cm^{-1} . Para o mapeamento Raman foi usado o modo *point-mapping*. Uma área de 4 cm^2 foi selecionada e os espectros foram adquiridos a cada 1 mm para as cédulas contaminadas no laboratório. O espectro de cada pixel foi coletado com 10s de exposição ao laser e 2 acumulações na região de 1800 a 200 cm^{-1} . Para as amostras forenses foram mapeadas 6 áreas de 4 cm^2 cada, esse procedimento foi necessário uma vez que não se conhecia a localização do

explosivo. O tempo gasto para cada área de 4 cm² mapeada foi de aproximadamente 3 horas.

6.1.2.3. Tratamento dos dados

Os dados obtidos por espectroscopia Raman de imagem apresentam-se na forma de um arranjo tridimensional, onde x e y são as dimensões espaciais e a terceira dimensão z corresponde ao número de onda do espectro. Para aplicação das ferramentas quimiométricas nas imagens hiperespectrais é necessário o desdobramento para duas dimensões. Dessa forma, o arranjo tridimensional foi desdobrado para duas dimensões, sendo que essa operação não acarreta perda da informação espacial. Antes da aplicação dos métodos multivariados, os dados foram pré-processados usando um filtro para retirada de *spikes*, que são gerados por raios cósmicos que atingem a Terra e detectados pelos detectores CCD. O algoritmo utilizado foi desenvolvido pelo nosso grupo de pesquisa¹¹³. Após a remoção dos *spikes*, os espectros foram normalizados para comprimento unitário para corrigir as diferenças de intensidades, as quais são geradas pela presença de diferentes compostos na amostra.

O método usado para resolução de curvas foi a Análise de Componentes Independentes (ICA)⁷⁶ e o algoritmo escolhido para obter os componentes independentes foi o ICA por campo médio (MF-ICA). O motivo da escolha deste algoritmo foi a possibilidade de adição de restrições, como a não negatividade.

O número de componentes, ou seja, o posto da matriz foi escolhido usando o método EFA. A variância explicada de até 1% de cada modelo PCA foi considerada como significativa. Esse valor foi escolhido levando em consideração a relação sinal ruído dos espectros Raman e a possível presença de resíduo de explosivos em amostras reais.

No conjunto de dados onde observou-se problemas de ambiguidade nas soluções, a estratégia do MCR por intervalos (iMCR)⁷⁵ foi adotada para diminuir e/ou eliminar a presença de ambiguidade nas soluções. Para utilização do iMCR, a matriz principal \mathbf{X} (400x402) foi dividida em 16 novas submatrizes de tamanho 25x402 e o

MCR-ALS foi aplicado para cada submatriz. Após análise com o MCR, as 16 submatrizes de \mathbf{C} e \mathbf{S}^T foram somadas e obteve-se a matriz com as dimensões iniciais.

A quantidade de ambiguidade rotacional nas soluções dos métodos de resolução de curvas foi calculada pelo método MCR-BANDS.

As rotinas para MCR-ALS e MCR-BANDS foram obtidas no site http://www.cid.csic.es/homes/rtaqam/tmp/WEB_MCR/download.html. O algoritmo MF-ICA foi obtido no site <http://cogsys.imm.dtu.dk/toolbox/ica/>.

As rotinas obtidas e as desenvolvidas no laboratório foram tratadas em ambiente Matlab, versão 7.8 (R2009a).

6.1.2.4. Limite de Detecção

O limite de detecção foi estimado para a metodologia proposta. Diferentes concentrações do explosivo TNT solubilizado em acetonitrila foram preparadas. Um volume de 2 mL de cada concentração foi adicionado na superfície da cédula e após a evaporação do solvente o mapeamento Raman foi realizado. As concentrações do explosivo TNT nas cédulas foram de 0,25, 0,20, 0,15, 0,10, 0,05 e 0,01 mg.cm⁻².

6.1.3 Resultados e Discussões

O espectro Raman de um composto tem seu perfil determinado pelo tipo de átomos presentes, pela forma como estão ligados e suas interações com os átomos vizinhos. Como consequência, os espectros Raman são ferramentas valiosas na identificação de materiais, uma vez que, as vibrações dos átomos acontecem em valores de frequências que são característicos para os tipos de grupos funcionais presentes. As substâncias classificadas como explosivos geralmente contém átomos de oxigênio, nitrogênio, carbono e hidrogênio. O oxigênio está ligado, na maioria dos compostos, ao nitrogênio, formando ligações NO, NO₂ e NO₃, como pode ser verificado na Figura 41. A frequência vibracional desses grupos depende das espécies químicas

vizinhas, dessa forma, cada molécula de explosivo possui um espectro Raman único e característico que pode ser empregado para sua identificação irrefutável.

Para uma correta interpretação dos espectros Raman, algumas considerações no procedimento de análise devem ser adotadas. Primeiro, para comparação dos espectros obtidos com dados da literatura ou banco de dados é importante que estas sejam feitas entre espectros obtidos nas mesmas condições experimentais, principalmente de radiação de excitação, uma vez que podem existir diferenças de intensidade das bandas Raman de um mesmo composto adquirido com diferentes energias de excitação. Outro ponto importante para a correta interpretação dos espectros Raman é a presença de cristais. Os microcristais podem apresentar alterações nas intensidades das bandas dependendo da sua orientação. Amostras que se apresentam na forma de pequenos cristais devem ter seu espectro Raman registrado em vários pontos, e por esse motivo, os espectros Raman dos padrões dos explosivos foram registrados em triplicata em diferentes pontos.

A Figura 44 mostra os espectros Raman dos padrões dos explosivos TNT, PETN, TNP, RDX, HMX, ANFO, perclorato de amônio e pólvora negra. Observa-se que o conjunto de bandas dos espectros Raman de cada explosivo apresenta perfis particulares, constituindo uma impressão digital para cada explosivo. A atribuição dos modos vibracionais foi feita com comparações de dados da literatura^{114,115}.

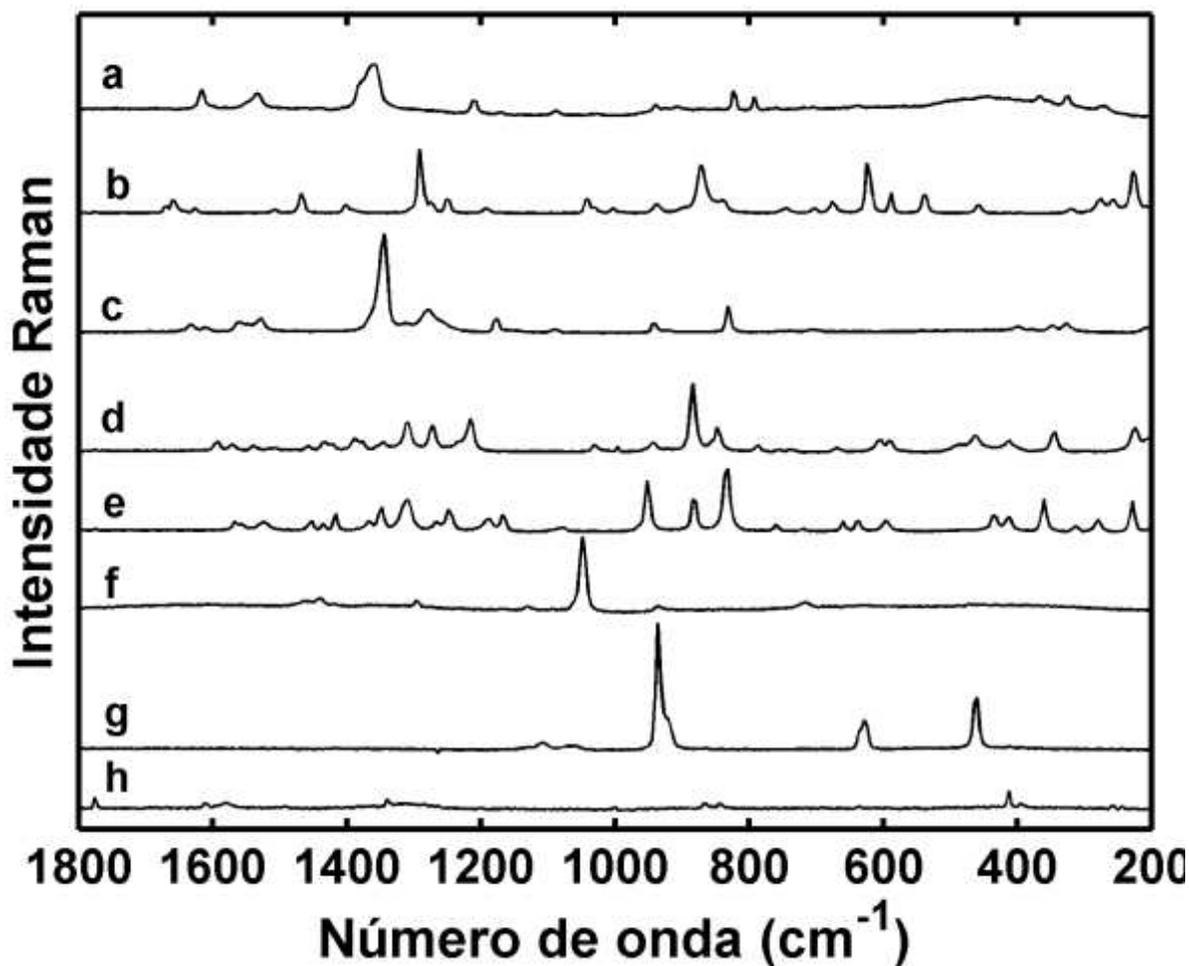


Figura 44: Espectros Raman dos explosivos (a) TNT; (b) PETN; (c) TNP; (d) RDX; (e) HMX; (f) ANFO; (g) perclorato de amônio; (h) pólvora negra.

O espectro Raman do explosivo TNT (Figura 36a) apresenta uma banda intensa em 1356 cm^{-1} referente ao $\nu_s(\text{NO}_2)$ que pode ser empregada para identificação inequívoca desse composto. A Figura 36b mostra o espectro Raman do PETN que apresenta várias bandas, sendo as mais intensas e características as bandas em 1290 e 872 cm^{-1} referentes aos modos de vibração da ligação N-O e em 642 cm^{-1} relacionado com modo de deformação C-C-C. As bandas em 1344 e 1276 cm^{-1} na Figura 36c são características do TNP e são atribuídas ao $\nu_s(\text{NO}_2)$ e $\nu(\text{C-C})_{\text{anel}}$, respectivamente. O explosivo RDX apresenta uma banda Raman com intensidade alta em 884 cm^{-1} , atribuída ao $\nu(\text{C-N-C})_{\text{anel}}$ e um conjunto de bandas em 1212 , 1276 e 1312 cm^{-1} que podem ser empregadas para identificação desse composto, como mostra a Figura 36d.

As bandas Raman em 836, 884 e 952 cm^{-1} da Figura 30e podem ser usadas para distinguir o explosivo HMX. O espectro Raman do explosivo ANFO (Figura 36f) apresenta uma banda com intensidade alta em 1044 cm^{-1} e é atribuída ao modo de vibração $\nu(\text{NO}_3)$. O modo de vibração $\nu_s(\text{ClO}_4)$ do perclorato de amônio aparece em 940 cm^{-1} , com intensidade maior que as demais bandas presentes, como pode ser observado na Figura 36g. A pólvora negra pode ser identificada pela presença das bandas no espectro Raman em 1600, 1350, 870 e 475 cm^{-1} , como mostra a Figura 36h que corresponde aos três constituintes principais, carvão (carbono), nitrato de potássio e enxofre, respectivamente.

Após conhecer os perfis espectrais dos explosivos, os arranjos de dados tridimensionais obtidos no mapeamento das cédulas contaminadas com os explosivos foram analisados usando os métodos de resolução de curvas. Para aplicar os métodos multivariados foi necessário desdobrar os dados em uma matriz. Em seguida, o conjunto de dados foi pré-processado usando normalização de vetor unitário. Finalmente, o método ICA foi aplicado aos dados para extrair a matriz espectral e a matriz de concentrações dos constituintes puros. Os espectros recuperados dos explosivos foram usados para fins de identificação e a matriz de concentração foi redobrada para recuperar a localização do pixel e fornecer o mapa de distribuição para cada um dos constituintes.

A escolha do número de componentes foi feita usando a estratégia EFA. Na imagem Raman das cédulas contaminadas com explosivos é esperado pelo menos três componentes, o papel, o pigmento da impressão e o explosivo acrescentado. Levando em conta essas informações, o conhecimento prévio do sistema pode ser adotado para a escolha do número de componentes. No entanto, o uso de uma estratégia, como o método EFA, para a escolha do número de componentes é justificado para amostras forenses reais, onde há presença de contaminantes e o analito pode estar presente em baixa concentração. Usando o número de componentes selecionados pelo EFA, os métodos MF-ICA e MCR-ALS foram aplicados nos conjuntos de dados. A Tabela 1 mostra o número de componentes utilizados para recuperar o espectro de cada explosivo, observa-se que este número foi diferente para os diferentes explosivos mostrando a importância de um critério de escolha. A diferença entre o número de

componentes para os explosivos pode ser relacionada a diferentes áreas mapeadas da cédula, presença de outros compostos e variações sistemáticas do mapeamento.

Tabela 1: Resultados obtidos pelos métodos MCR-ALS e MF-ICA

Dados	N ^o		LOF(%)		R ²		r ^a	
	componentes		ICA	MCR	ICA	MCR	ICA	MCR
	ICA	MCR						
TNP	6		4,52	4,56	99,79	99,79	0,99	0,99
HMX	7		2,27	2,48	99,95	99,94	0,87	0,93
PERC	12		2,12	2,34	99,95	99,94	0,94	0,94
PETN	13		4,78	5,36	99,71	99,77	0,95	0,92
RDX	12		4,86	5,14	99,76	99,74	0,93	0,93
TNT	8		3,18	2,66 ^b	99,93	99,89 ^b	0,83	0,80 ^b
ANFO	6		7,40	6,45	99,45	99,58	0,79	0,80
Pólvora negra	6		4,52	4,58	99,79	99,79	0,78	0,74

^a Coeficiente de correlação entre o espectro recuperado e o espectro real

^b resultados obtidos aplicando i-MCR (MCR por intervalo)

No método ICA não é necessário fornecer uma estimativa inicial do perfil espectral ou de concentração. Essa característica do algoritmo é importante no caso de amostras forenses, onde o analista não sabe quais compostos estão presentes na amostra. Para restringir as soluções do algoritmo MF-ICA e obter um erro mínimo, restrições de não negatividade foram aplicados aos perfis espectrais (matriz **S**) e de concentração (matriz **C**). A Tabela 1 mostra os parâmetros de otimização alcançados pelo algoritmo MF-ICA. Os valores de falta de ajuste foram inferiores a 7% e o percentual de variância explicada ficou próximo de 100% para todos os explosivos. Esses resultados mostram, do ponto de vista matemático, que o MF-ICA recuperou adequadamente os dados.

A Figura 45 compara os perfis espectrais recuperados para o componente que representa o explosivo com o espectro do explosivo puro. A Tabela 1 mostra o coeficiente de correlação entre o explosivo puro e os perfis recuperados. Para a maioria dos explosivos estudados, o coeficiente de correlação entre o espectro do explosivo

puro e o espectro recuperado pelo MF-ICA está próximo de 1. Esses resultados indicam uma recuperação adequada dos espectros dos explosivos, mostrando que os perfis espectrais recuperados possuem sentido químico.

A comparação dos espectros recuperados com os espectros de referência mostra que os explosivos podem ser identificados pelos seus espectros Raman recuperados. Mesmo quando o coeficiente de correlação é próximo de 0,8, como observado para TNT, ANFO e pólvora, as bandas Raman características desses explosivos permitem a identificação inequívoca dos mesmos.

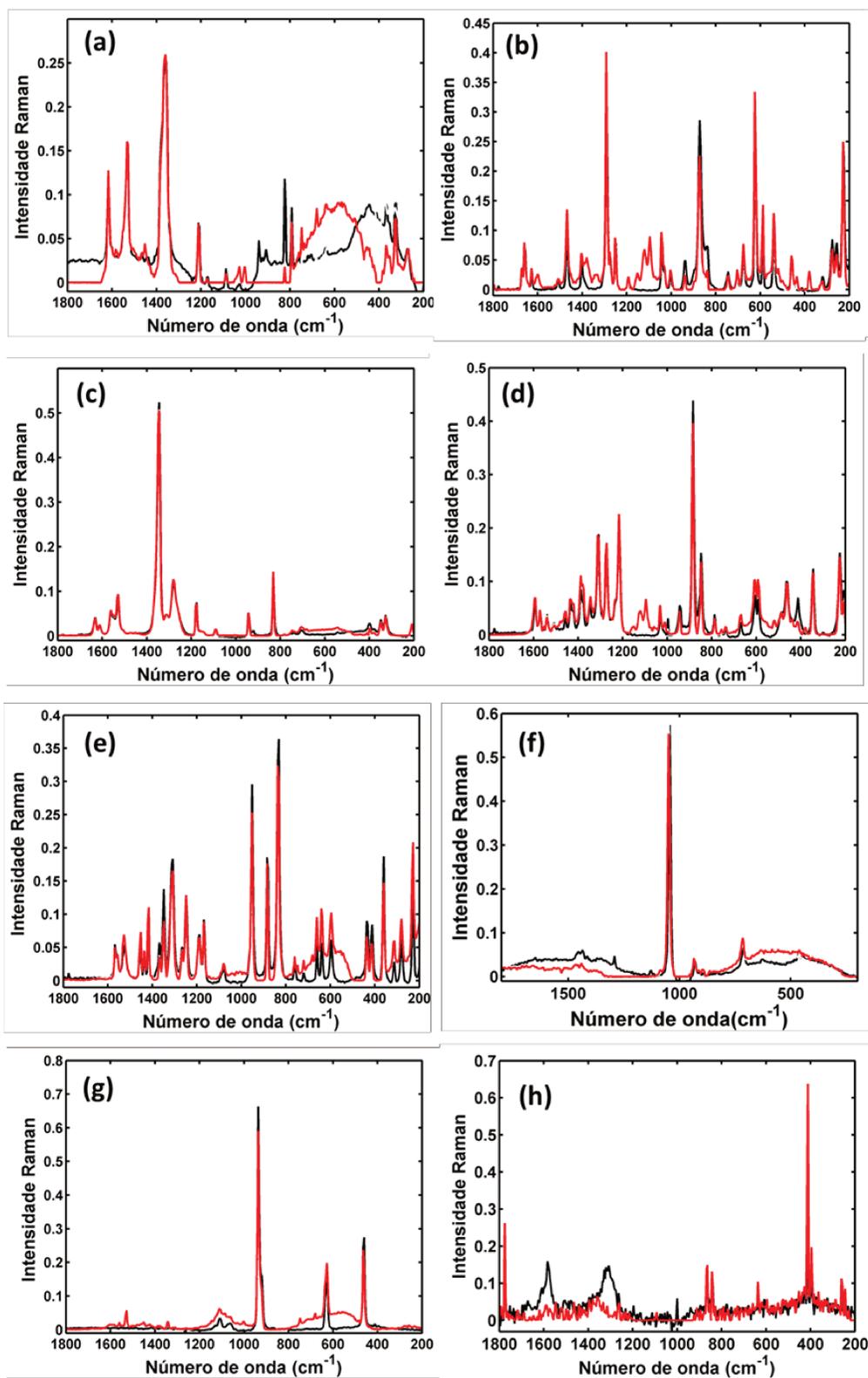


Figura 45: Comparação entre os espectros Raman dos explosivos puro (linha preta) com os espectros recuperados pelo MF-ICA (linha vermelha) para os explosivos (a) TNT; (b) PETN; (c) TNP; (d) RDX; (e) HMX; (f) ANFO; (g) PERC (h) pólvora negra.

Nos espectros recuperados pelo MF-ICA é possível observar regiões espectrais em que o espectro do explosivo puro é diferente do espectro recuperado pelo MF-ICA. Essas diferenças espectrais são devidas à ambiguidade rotacional, em que diferentes combinações de **C** e **S** retornam a matriz **X** sem representar a solução real. Se fossem utilizados os espectros com estas diferenças, os valores de concentração obtidos não estariam corretos. Dessa forma, para construir os mapas de concentração, foram usadas as regiões dos espectros recuperados que não apresentam ambiguidade rotacional. As matrizes **C** dos perfis de concentração foram redobradas e dessa forma, obtidos os mapas de concentração relativa dos componentes. A Figura 46 mostra os mapas de concentração estimados pelo MF-ICA para os explosivos. As cores denotam a concentração do explosivo em cada pixel.

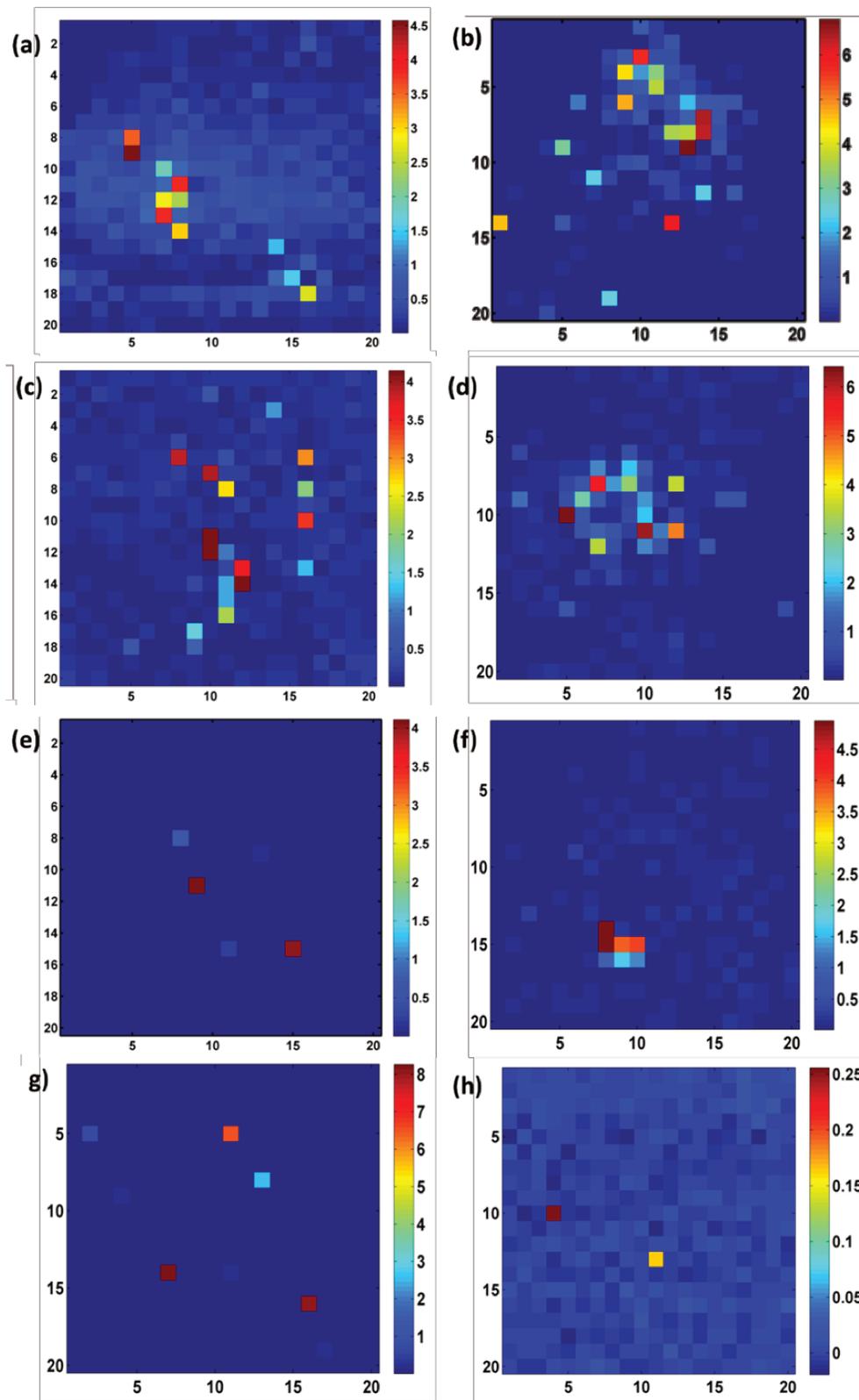


Figura 46: Mapas de concentração obtidos para os explosivos (a) TNP; (b) HMX; (c) PERC; (d) PETN; (e) RDX; (f) TNT; (g) ANFO e (h) pólvora negra pelo método MF-ICA nas amostras preparadas em laboratório.

Para fins de comparação, o mesmo conjunto de dados foi analisado pelo algoritmo MCR-ALS. O número de componentes utilizados foi o mesmo que o do MF-ICA. Os parâmetros de otimização são mostrados na Tabela 1. Comparando os valores de falta de ajuste das imagens resolvidas pelos métodos MCR-ALS e MF-ICA para os explosivos TNP, HMX, PERC e pólvora observa-se que os valores obtidos são semelhantes para os dois métodos, isso mostra que as soluções obtidas pelos dois métodos são equivalentes do ponto de vista matemático. Para os explosivos PETN e RDX, os valores de falta de ajuste obtidos pelo MF-ICA são sutilmente menores do que as soluções obtidas pelo MCR-ALS. Já para o explosivo ANFO a falta de ajuste foi levemente mais alta na solução obtida pelo método MF-ICA. Para o conjunto de dados que contém o explosivo TNT, a falta de ajuste foi maior na solução do MCR-ALS (7,5%) do que na do ICA (3%).

Em termos de variância explicada, os valores obtidos foram altos (99%) para todos os conjuntos de dados e semelhantes para os dois métodos. Os valores dos coeficientes de correlação foram próximos ou superiores a 0,8 para todos os explosivos estudados. Os dois métodos de resolução de curvas forneceram resultados semelhantes, exceto para o explosivo TNT, para o qual o MCR-ALS não recuperou o perfil espectral do explosivo puro. A justificativa para esse resultado é baixa contribuição do sinal do explosivo TNT para o modelo global. O coeficiente de correlação do perfil recuperado pelo MCR-ALS e espectro do TNT foi de 0,5. No caso do algoritmo MF-ICA, o coeficiente de correlação foi de 0,83, e foi possível identificar o explosivo através das bandas Raman em 1532 e 1357 cm^{-1} .

Para resolver o problema da ambiguidade rotacional, um conjunto de restrições está disponível para o algoritmo MCR-ALS, tais como não negatividade, unimodalidade, *closure* (balanço de massa), seletividade, posto local e a estratégia de matriz aumentada. Restrição de não negatividade e normalização foram empregadas para as soluções do MCR-ALS, e com exceção do conjunto de dados do explosivo TNT, as soluções obtidas pelo método MCR-ALS foram satisfatórias. Outras estratégias de restrições, como seletividade e posto local que forneceriam baixa ambiguidade e estimativas mais confiáveis, não são simples de implementar em imagens

hiperespectrais de amostras forenses, uma vez que é difícil obter informação de posto local, onde é necessário conhecer se um pixel tem um ou mais componentes.

Para corrigir a falta de convergência do MCR-ALS para o conjunto de dados do TNT, foi empregado o MCR-ALS em intervalos (iMCR)⁷⁵. A matriz de dados de dimensão 400 x 402 foi dividida em 16 submatrizes de 25 x 402 e para cada submatriz o MCR-ALS foi empregado, como esquematizado na Figura 47. A escolha de 16 submatrizes foi empírica, blocos de dimensões maiores foram testados, mas não foi observado melhora na solução comparada com a matriz total. As soluções de três blocos mostraram perfis espectrais do explosivo TNT com coeficiente de correlação de 0,8. As principais bandas do TNT foram recuperadas, permitindo a identificação do explosivo. Dessa forma, o problema da ambiguidade foi corrigido pelo método iMCR. Os resultados obtidos pelo iMCR podem ser vistos na Figura 47.

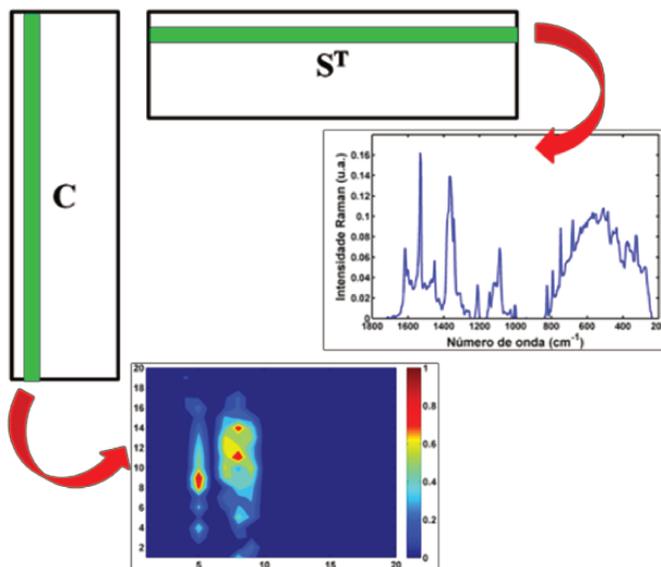
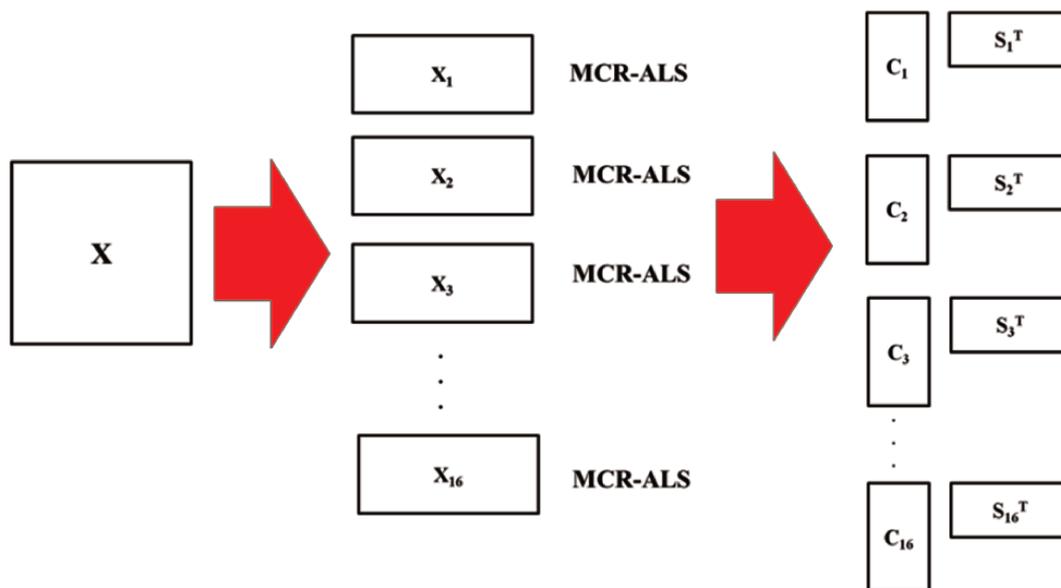


Figura 47: Esquema da decomposição da matriz para aplicação do iMCR-ALS e o resultado obtido pelo explosivo TNT.

Para verificar se as soluções do MF-ICA e do MCR-ALS são únicas, ou seja, se há presença de ambiguidade rotacional nas soluções fornecidas foi empregado o método MCR-BANDS. Os resultados da função f (função de contribuição de cada componente) mostraram que os valores máximos e mínimos dessa função são próximos, indicando baixa ambiguidade rotacional para as soluções do MF-ICA e MCR-ALS. Isto é, os espectros resolvidos pelos métodos de resolução de curvas são

próximos do real. Esses resultados podem ser verificados na Figura 45. Na proposta qualitativa desse trabalho, a baixa ambiguidade rotacional nos perfis espectrais não foi um problema, uma vez que foi possível a identificação do explosivo. Para fins quantitativos, a eliminação da ambiguidade rotacional é importante para a correta quantificação da concentração no pixel.

Tabela 2: Valores máximos e mínimos de f obtidos pelo MCR-BANDS para as soluções do MCR-ALS e MF-ICA.

Dados	f_{max} MF-ICA	f_{min} MF-ICA	f_{max} MCR-ALS	f_{min} MCR-ALS
TNP	0,143	0,143	0,144	0,143
HMX	0,068	0,068	0,065	0,063
PERC	0,082	0,082	0,077	0,073
PETN	0,151	0,151	0,152	0,150
RDX	0,114	0,114	0,111	0,108
TNT	0,360	0,277	0,111	0,093
ANFO	0,094	0,092	0,093	0,093
BP	0,083	0,079	0,046	0,046

O limite de detecção para a metodologia proposta foi estimado como sendo a menor quantidade de explosivo que pode ser detectada. Amostras de cédulas foram contaminadas com concentrações de 0,25, 0,20, 0,15, 0,10, 0,05 e 0,01 mg.cm⁻² do explosivo TNT e mapeadas. Os dados foram pré-processados e o algoritmo MF-ICA foi aplicado aos dados. Os resultados são mostrados na Figura 48, sendo que o menor valor detectado no pixel foi de 50 µg.cm⁻². Até essa concentração é possível identificar esse explosivo pela metodologia proposta.

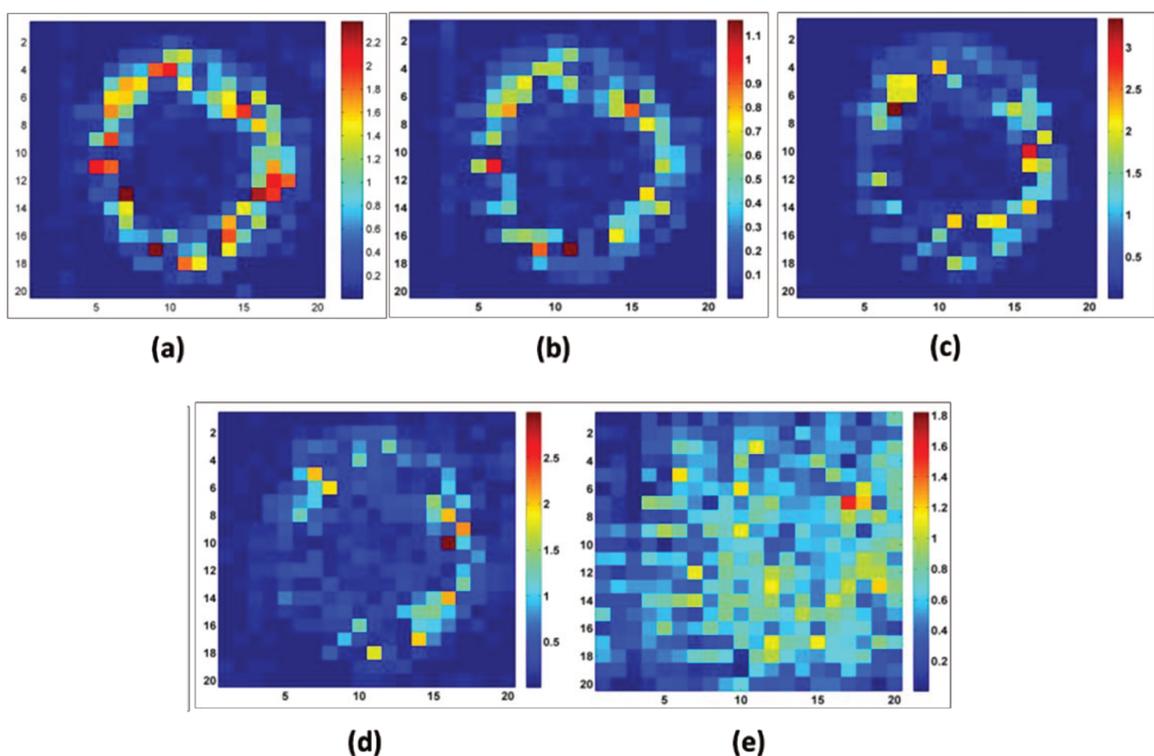


Figura 48: Mapas de concentração obtidos no cálculo do limite de detecção do explosivo TNT para as concentrações (a) $0,25 \text{ mg.cm}^{-2}$; (b) $0,20 \text{ mg.cm}^{-2}$; (c) $0,15 \text{ mg.cm}^{-2}$; (d) $0,10 \text{ mg.cm}^{-2}$; (e) $0,05 \text{ mg.cm}^{-2}$.

A segunda parte do trabalho consistiu na análise de amostras forenses. O primeiro conjunto de amostras analisadas foi fornecido pela Polícia Científica do Estado de São Paulo consistindo de amostras suspeitas de terem explosivos. Nesse caso não conhecia-se a identidade do explosivo. Após o mapeamento Raman, os dados foram pré-processados e analisados com o método de resolução de curvas MF-ICA. O número de componentes foi escolhido usando o método EFA, a Tabela 3 mostra o número de componentes usados e os parâmetros de otimização para as soluções do método MF-ICA.

Tabela 3: Resultados obtidos pelo MF-ICA para as amostras forenses

Dados	Nº Componentes	LOF	R ²	r
Cédulas com pólvora	8	4,18	99,82	0,64
Cédulas com ANFO	8	15,80	92,09	0,48

O número de componentes escolhido pelo método EFA foi igual a 8. A falta de ajuste do modelo foi de 4,18% e a variância explicada pelo modelo foi próxima de 100%. A Figura 49 mostra as amostras e áreas mapeadas e o espectro recuperado do explosivo com seu respectivo mapa de concentração.

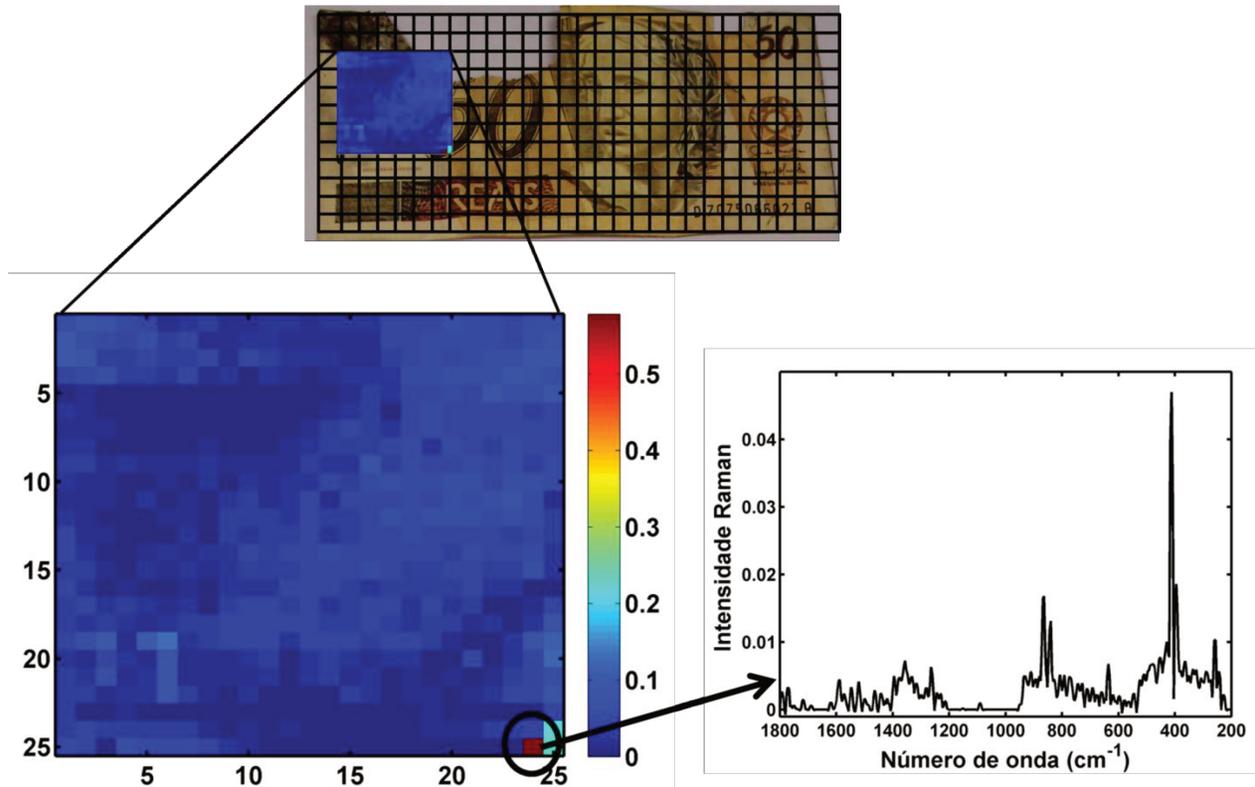


Figura 49: Amostra suspeitas de ter explosivo, perfil espectral e imagem Raman recuperados pelo método MF-ICA de um dos componentes, mostrando a presença de pólvora e sua distribuição na cédula.

O espectro recuperado mostra a presença de pólvora negra nas cédulas analisadas. A pólvora é um composto formado por grafite, nitrato de potássio e enxofre, sendo 75% de sua composição de nitrato de potássio. Seus resíduos pós-exploração são nitritos, nitratos, carbono na forma de fuligem negra, potássio e enxofre. A pólvora pode ser confeccionada artesanalmente e é empregada para disparar a dinamite, um explosivo bastante utilizado em desmonte de rochas.

O coeficiente de correlação foi de 0,64, algumas regiões de ambiguidade rotacional foram observadas no espectro recuperado pelo MF-ICA, no entanto, a região

entre 1000 e 400 cm^{-1} apresentou baixa ambiguidade rotacional e as bandas em 870 e 475 cm^{-1} foram usadas para identificação do nitrato de potássio e enxofre, dois constituintes da pólvora. Com base nos resultados apresentados, pode-se afirmar que existe evidência da presença de explosivos nas cédulas apreendidas.

A segunda amostra forense estudada foi obtida de uma explosão simulada de cédulas de R\$ 50 com o explosivo ANFO. Na simulação da explosão, o pigmento usado como dispositivo de segurança foi empregado. Esse dispositivo é utilizado em caixas eletrônicos e em caso de arrombamento é acionado causando manchas nas cédulas para inutilizá-las.

Após o mapeamento da cédula pós-explosão, os dados foram tratados usando o algoritmo MF-ICA. As soluções do MF-ICA mostraram a presença de uma banda com intensidade alta na região de 1040 cm^{-1} , o que caracteriza a presença de nitrato. A Tabela 3 mostra os parâmetros obtidos pelo método MF-ICA. O número de componentes para a área onde foi encontrado a presença de nitrato foi de 8 componentes, este número foi selecionado pelo método EFA. A falta de ajuste na modelagem dos dados foi de 15,80% e a variância explicada foi de 92,09%. Devido à presença de ambiguidade rotacional nas soluções do MF-ICA, o coeficiente de correlação obtido foi de 0,48. No entanto, como a proposta do trabalho é qualitativa, a identificação da banda de nitrato indica a presença de explosivo.

O explosivo empregado nessa explosão foi o ANFO, que é composto por nitrato de amônio com óleo combustível. A quantidade de resíduo gerada por esse explosivo é pequena, dado o alto rendimento da explosão. Ainda assim, o uso da espectroscopia Raman juntamente com métodos de resolução de curvas permitiu identificar a presença do explosivo. Os resultados do MF-ICA são mostrados na Figura 50. O mapa de concentração de cada componente recuperado permite identificar a localização desses componentes na superfície analisada.

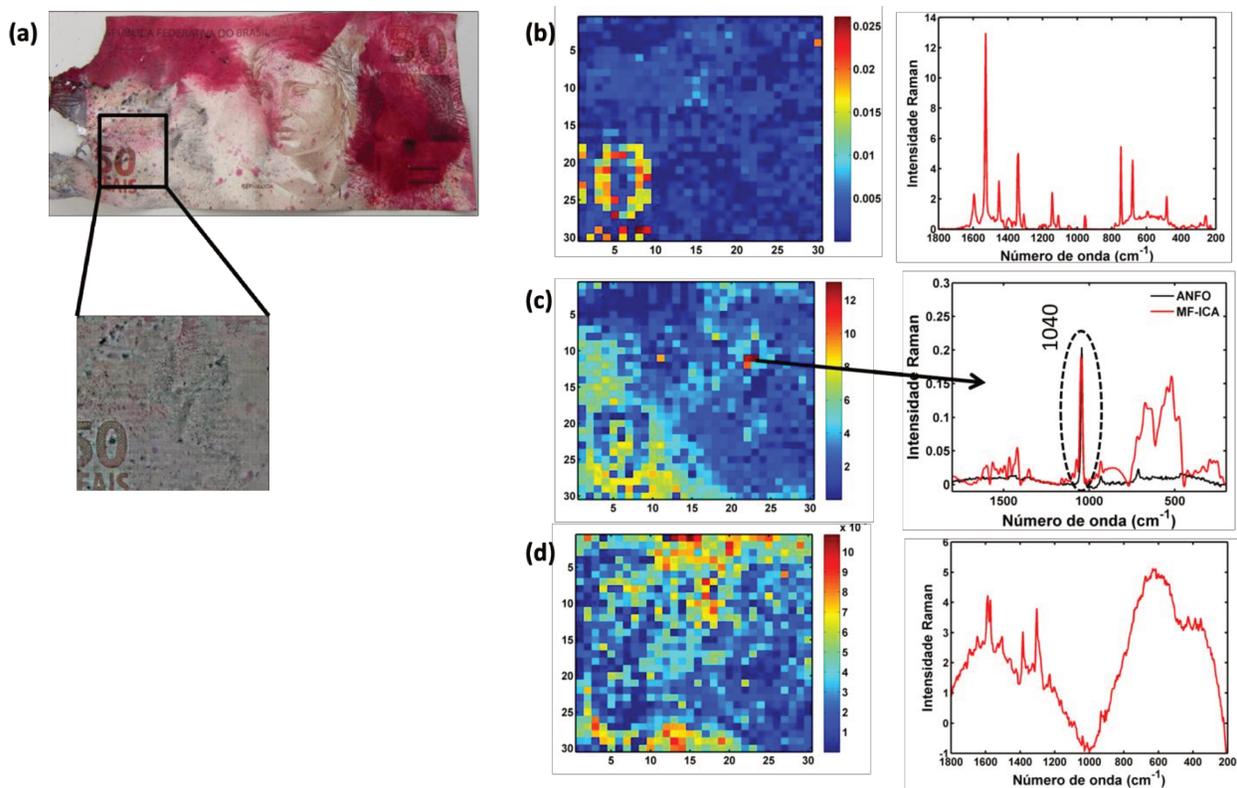


Figura 50: Área da cédula onde foi encontrado o explosivo (a); imagem Raman e perfil espectral recuperados pelo método MF-ICA do pigmento da cédula (b); imagem Raman, perfil espectral (linha vermelha) recuperados pelo método MF-ICA do explosivo ANFO e espectro Raman do explosivo ANFO (linha preta) (c); imagem Raman e perfil espectral recuperados pelo método MF-ICA do dispositivo de segurança (d).

6.1.5 Conclusões

O uso da espectroscopia Raman de imagem apresenta grande potencial para o estudo de amostras forenses. O uso da metodologia proposta permite obter informações espectrais e espaciais dos explosivos em superfícies de cédulas e essa conclusão pode abranger outras superfícies. As análises permitem obter informação química e física sem preparo e sem destruição da amostra. A metodologia apresentada nesse trabalho pode ser estendida para dados obtidos por espectroscopia Raman com detecção *stand-off*.

Os métodos de resolução de curvas empregados, MCR e o ICA, se mostraram adequados e os resultados convergiram para soluções equivalentes na maioria das

amostras estudadas. No entanto, foi observado que em baixas concentrações de um constituinte, o MCR-ALS apresentou limitações não convergindo para a solução correta. Dessa forma, foi necessário a adoção de outra estratégia, como o emprego do método iMCR. A aplicação do MCR-ALS nas submatrizes mostrou-se eficiente e foi possível recuperar o espectro do explosivo. A desvantagem apresentada foi o tempo gasto na recuperação dos perfis espectrais. Já o MF-ICA conseguiu recuperar adequadamente todos os perfis espectrais dos conjuntos de dados estudados com baixa ambiguidade rotacional.

Para as amostras forenses, os perfis espectrais recuperados dos explosivos apresentaram ambiguidade rotacional, entretanto, foi possível identificar a presença de explosivo pela frequência vibracional do nitrato. Dessa forma, a espectroscopia Raman se mostrou uma técnica sensível na identificação de explosivos.

6.2. Aplicação da espectroscopia Raman de imagem e quimiometria em Paleontologia

6.2.1 Introdução

Os fósseis, objetos de estudo da paleontologia, são restos ou vestígios de animais, plantas ou micro-organismos. Muitas vezes são as únicas evidências de vida de espécies que existiram e agora não existem mais. O entendimento do processo de fossilização pode fornecer informações de processos evolutivos e geológicos¹¹⁶.

O processo de fossilização pode ocorrer de várias maneiras e envolve várias etapas. De forma geral, inicia-se com a morte do organismo, passando por processos enzimáticos e microbiológicos que acarretam na decomposição dos tecidos moles. O organismo pode ainda passar por processo de desarticulação, transporte e ser soterrado ou ser soterrado logo após sua morte. O soterramento acontece quando a água ou outro agente transporta sedimentos que recobrem os organismos, passando a ocorrer processos físicos e químicos, esses processos são chamados de diagênese e dão início à formação do fóssil. Em determinadas condições ocorre a substituição da matéria orgânica por minerais presentes no ambiente, onde o fóssil mantém a forma do organismo, sendo esse processo de fossilização chamado de mineralização¹¹⁶.

Nas últimas décadas, técnicas analíticas têm sido empregadas no estudo de fósseis. As análises químicas são realizadas a fim de identificar sua composição, compreender o mecanismo de fossilização, estudar a diagênese e seus efeitos na preservação e formação dos fósseis, estudar processos metamórficos e investigar características biológicas que mudaram ao longo do tempo^{25,36,27}.

A espectroscopia vibracional, infravermelho e Raman, tem sido abordagens analíticas usadas para gerar informações moleculares de fósseis^{43,117,118}. Os trabalhos, geralmente combinam informações morfológicas usando microscopia eletrônica de varredura e informação molecular com os espectros vibracionais.

As pesquisas paleontológicas empregando espectroscopia Raman são importantes para identificar a origem biológica de fósseis, que é feita pela presença de

material carbonáceo encontrado nestes materiais, como o querogênio. Este material pode ser definido como a matéria orgânica residual após o processo de diagênese. No entanto, para identificar a origem biológica de fósseis é importante diferenciar entre material carbonáceo de origem biótica e abiótica. Porém, ainda existe uma controvérsia quanto a aplicação da espectroscopia Raman na identificação do tipo de carbono, sua estrutura e interpretação⁴⁵. Uma vez que processos geoquímicos e metamórficos levam a matéria orgânica de origem biológica e não biológica a produtos semelhantes⁴⁶. O material carbonáceo (independente da origem) pode sofrer, por exemplo, mudanças em sua composição química e estrutural devido ao aumento do grau metamórfico, sendo convertido em grafite¹¹⁹.

Na última década, a literatura científica têm divulgado trabalhos que mostram a preservação de tecidos moles em fósseis^{118,120,121,122}. Apesar de esses resultados serem vistos com um pouco de ceticismo, essas descobertas só foram possíveis com o emprego de técnicas analíticas na paleontologia, sendo possível acessar informações das espécies biológicas para obter informações de processos físicos até a nível molecular. Estas pesquisas mostram a importância da combinação de estudos morfológicos e químicos para uma maior compreensão dos organismos extintos e dos processos geológicos que atuaram sobre eles, assim como para entender a evolução da vida no planeta.

O objetivo desse estudo foi aplicar espectroscopia Raman de imagem e ferramentas quimiométricas de resolução multivariada de curvas para estudar a composição de fósseis de peixes.

Os fósseis estudados são do período Cretáceo, que corresponde ao período de 144 a 65 milhões de anos. O estudo sobre a composição de peixes fossilizados do período Cretáceo por técnicas vibracionais é pouco abordado na literatura, dessa forma, esse trabalho pretende ampliar o conhecimento sobre novas ferramentas para o estudo químico desse tipo de fósseis, assim como obter informação sobre suas características biológicas.

As amostras são do peixe *Rhacolepis buccalis*, espécie extinta há 125 milhões de anos e foram encontrados na Bacia Sedimentar do Araripe, localizada no extremo sul do estado do Ceará na região nordeste do Brasil, como mostra a Figura 51.



Figura 51: Localização da Bacia Sedimentar do Araripe.

A bacia do Araripe é conhecida como o principal depósito fossilífero do país, sendo os fósseis encontrados na região classificados como grau de preservação excepcional, esta característica faz com que os fósseis dessa região constituam material paleontológico reconhecido internacionalmente.

6.2.2 Parte Experimental

6.2.2.1 Amostras

Três amostras de fósseis tridimensionais do peixe *Rhacolepsis buccalis* foram fornecidas pelo pesquisador Dr. José Xavier Neto do Laboratório Nacional de Biociências do Centro Nacional de Pesquisas em Energia e Materiais.

6.2.2.2 Imagem Raman

As imagens Raman foram obtidas no espectrômetro Raman dispersivo RamanStation da PerkinElmer, equipado com detector CCD de excitação em 785 nm.

Várias regiões dos fósseis foram mapeadas, sendo essas regiões denominadas de área A1 (4mm²) que é a matriz na qual o fóssil está inserido; área A2 (2 mm²) é a região supostamente onde ficava localizado o coração; área A3 (2 mm²) é a região que seria o aparelho digestivo (estômago e intestino); área A4 (2 mm²) é a coluna vertebral e a área A5 (2 mm²) compreende a região dorsal. Para cada área foi realizado um mapeamento Raman com espectros adquiridos a cada 0,2 mm. O espectro de cada pixel foi coletados com 3s de exposição ao laser (20 mW) e 6 acumulações na região de 1800 a 200 cm⁻¹ e resolução espectral de 4 cm⁻¹.

6.2.3. Tratamento dos dados

Após o procedimento de desdobramento do arranjo tridimensional e remoção dos *spikes*, como descrito anteriormente, foi realizada a correção da linha de base usando o algoritmo de alisamento *Savitzky-Golay*¹²³. Uma linha de base foi estimada usando o filtro de *Savitzky-Golay* com largura da janela maior que 47 pontos e polinômio de 2^a ordem, a linha de base estimada foi subtraída do conjunto de dados original. Para estimar a linha de base empregando métodos de suavização é importante observar que o tamanho da janela precisa ser grande para remoção do ruído e das bandas Raman, mas não tão grande para não distorcer os espectros quando for realizada a subtração.

Os espectros foram normalizados para comprimento unitário para corrigir diferenças de intensidades entre os espectros devido a mudanças no foco. A análise de componentes independentes (MF-ICA) foi empregada para obter os perfis dos compostos puros e sua distribuição nos fósseis.

6.2.4. Resultados e Discussões

Após o desdobramento do arranjo tridimensional para uma matriz de dados, foi realizada a etapa de pré-processamento. Os espectros Raman obtidos apresentaram efeitos secundários como ruído de fundo e *spikes* de raios cósmicos. Estes efeitos

dificultam a interpretação e modelagem dos dados e precisam ser corrigidos antes da aplicação das ferramentas quimiométricas. A Figura 52 ilustra o conjunto de dados obtidos no mapeamento dos fósseis. Para eliminar esses efeitos foram realizadas etapas de pré-processamento. A primeira etapa foi a retirada de *spikes* usando o algoritmo *anti-spike*¹¹³ desenvolvido pelo nosso grupo. Após a retirada dos *spikes*, foi realizada a correção da linha de base usando o filtro de suavização de *Savitzky-Golay*.

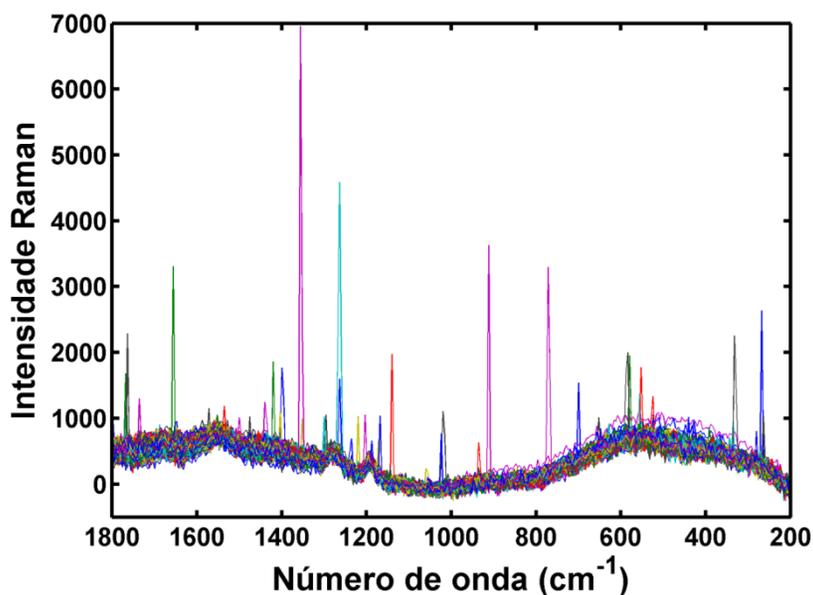


Figura 52: Conjunto de espectros Raman antes do pré-processamento dos dados.

A Figura 53 mostra o conjunto de dados após a remoção dos *spikes* e correção da linha de base. Após o pré-processamento dos dados, o método de resolução de curvas ICA foi empregado e os perfis espectrais puros e de concentração foram obtidos. O número de componentes foi escolhido usando a estratégia EFA (Sessão 6.1.3). Nesse caso, 4 componentes foram selecionados para as áreas mapeadas, no entanto, observou-se que somente um ou dois componentes possuíam significado químico.

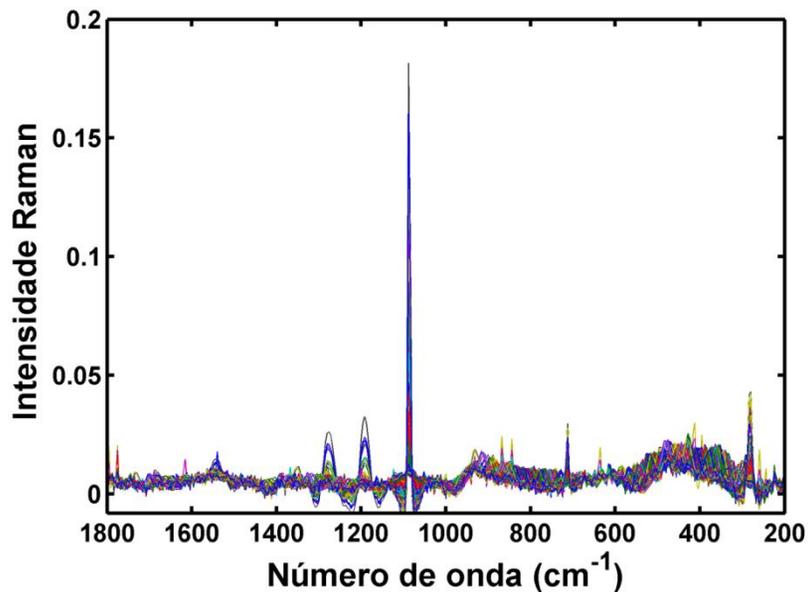


Figura 53: Conjunto de espectros Raman após o pré-processamento dos dados.

A Figura 54 mostra os principais resultados obtidos para as áreas mapeadas de uma amostra do fóssil. O espectro Raman recuperado do mapeamento da região (A1) apresenta bandas características de carbonato de cálcio, CaCO_3 (calcita). A intensa banda em 1087 cm^{-1} é devido ao estiramento simétrico da ligação CO_3^{2-} , a banda em 715 cm^{-1} é atribuída a deformação CO_3^{2-} e o sinal Raman observado em 284 cm^{-1} é devido a vibrações externas do CO_3^{2-} ¹²⁴. O perfil espectral da calcita também foi recuperado no mapeamento realizado na região da coluna vertebral (A4) e da região dorsal (A5). O carbonato de cálcio não é um componente biogênico, sendo conseqüentemente um indicador de alteração diagenética. A presença de calcita no fóssil indica que o processo de fossilização deu-se por mineralização, onde a matéria orgânica foi substituída por matéria mineral. Esse resultado era esperado, uma vez que o fóssil se originou a partir de rochas sedimentares que contêm carbonatos, como evidenciado pelo espectro Raman recuperado da região da matriz (região A1).

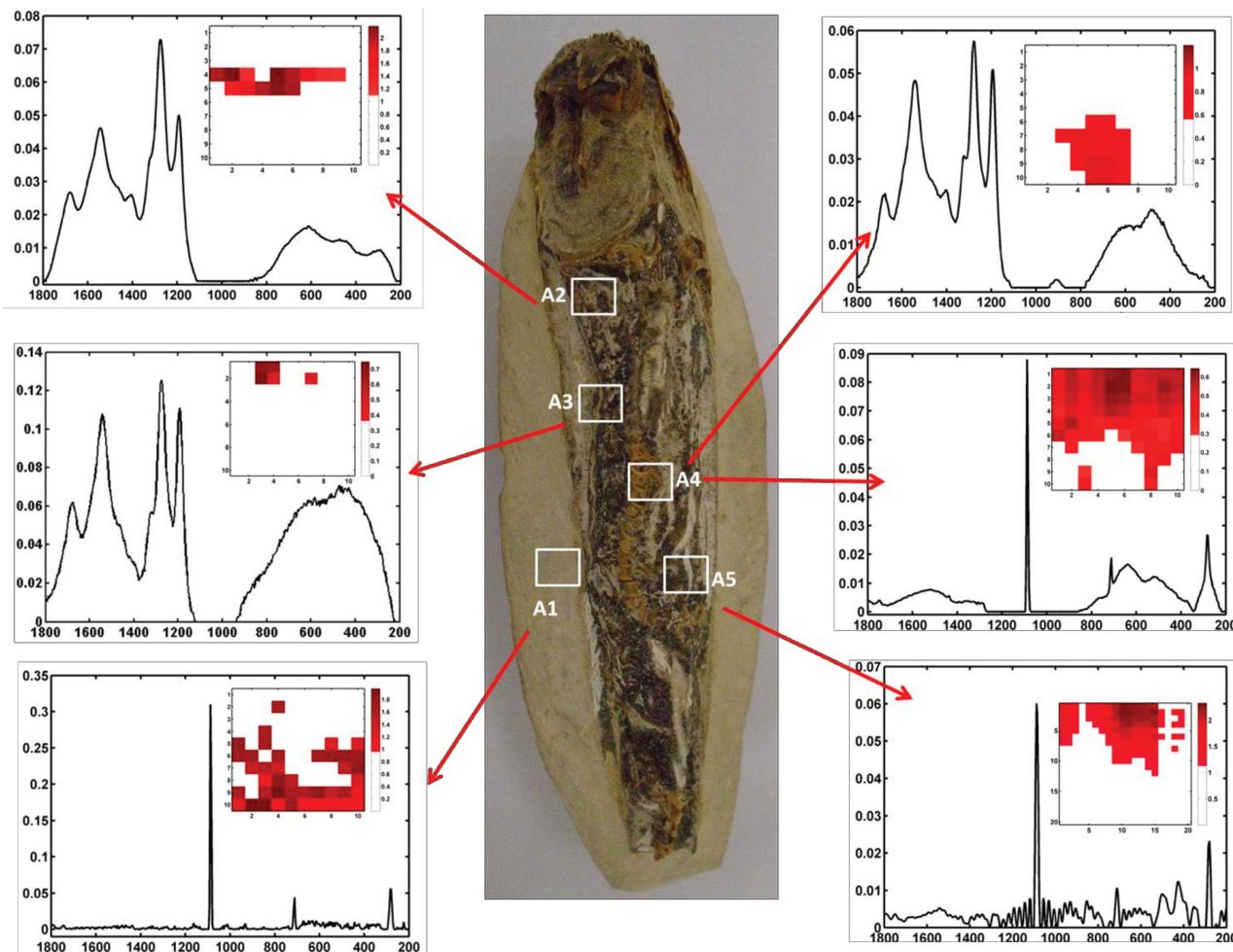


Figura 54: Áreas mapeadas do fóssil e os perfis espectrais recuperados pelo método MF-ICA e seus respectivos mapas de concentração. A1 corresponde a matriz na qual o fóssil está inserido; A2 é a região do fóssil que supostamente localizava o coração; A3 corresponde uma região que provavelmente localizava-se o aparelho digestivo; A4 é a coluna vertebral; A5 é a região dorsal.

Outro perfil espectral que chamou atenção foi obtido nas regiões que supostamente localizava-se o coração (A2), o estômago (A3) e na região da coluna vertebral (A4). Os perfis espectrais dessas regiões mostraram bandas interessantes na região de 1650 e 1200 cm^{-1} . Geralmente essas bandas são atribuídas a presença de materiais carbonáceos em fósseis^{39,45,46,125}.

Trabalhos da literatura mostram que a presença de bandas nessa região em fósseis, principalmente de ossos, está relacionada com a presença de resíduos de colágeno^{42,43}. A princípio espera-se que o processo de fossilização seja iniciado pela decomposição dos compostos orgânicos, no entanto, a literatura tem mostrado que

sobre condições ambientais específicas é possível a preservação destes compostos mesmo após milhares de anos^{126,121,120}. Os fatores que influenciam na possível preservação dos tecidos moles são a decomposição, soterramento rápido e mineralização rápida. No caso, a mineralização precoce por carbonato, caso da amostra estudada, é o processo mais frequente de preservação de tecidos moles.

A importância dos resultados obtidos com a análise do fóssil do peixe *Rhacolepsis buccalis* é a possível presença de matéria orgânica em regiões que supostamente ficariam localizados o coração e o sistema digestivo. Estes resultados abrem campo para investigação em evolução empregando a técnica de espectroscopia Raman de imagem e pode fornecer informação sobre características biológicas do peixe *Rhacolepsis buccalis*.

As imagens Raman apresentadas na Figura 54 mostram a localização dos espectros recuperados na área do fóssil analisada. A região de coloração vermelha indica a presença do perfil espectral em questão. Os resultados apresentados aqui só foram possíveis de se obter com a aplicação de métodos de pré-processamento dos espectros Raman e métodos quimiométricos de resolução de curvas.

6.2.5. Conclusões

Nesse trabalho foi explorada a potencialidade da espectroscopia Raman de imagem em conjunto com a análise de componentes independentes para a geração de imagens química de fósseis. Os resultados apresentados podem ser usados para obter informação sobre a morfologia e composição química dos fósseis.

Os resultados mostraram que a matriz dos fósseis encontrados na bacia do Araripe é constituída principalmente de calcita, composto também encontrado no fóssil. Indicativos da material carbonáceo foram encontrados principalmente na região do coração, do estomago e da coluna vertebral.

O método de resolução multivariada de curvas, ICA, se mostrou útil na recuperação dos perfis espectrais e mapas de concentração. No entanto, foi necessária uma etapa de pré-processamento dos dados para retirada dos *spikes*, correção da linha

de base e normalização das intensidades de espalhamento Raman. A etapa de pré-processamento permitiu a eliminação de efeitos físicos e conseqüentemente, a interpretação dos espectros.

A metodologia apresentada é interessante para a área de paleontologia, uma vez que as amostras são raras. Um procedimento experimental que permite obter informação química e espacial sem a destruição da amostra é importante para o desenvolvimento científico na área.

CAPÍTULO 7: Conclusões Gerais

Nessa tese foi proposto o desenvolvimento de metodologias analíticas para a área de química forense e paleontologia empregando espectroscopia Raman. Tendo como motivação a importância do desenvolvimento de métodos de análise não-destrutivos, com pouco ou nenhum preparo de amostra e que permitam a obtenção do máximo de informação com a realização de uma única análise.

Na área forense, modelos de classificação foram construídos com o auxílio da ferramenta quimiométrica PLS-DA para discriminar cédulas autênticas de cédulas falsas e identificar a origem do óleo extraído da árvore amazônica *Aniba Rosaedora*. O modelo de classificação desenvolvido para discriminação de cédulas foi baseado nos espectros Raman dos pigmentos de três regiões da cédula de R\$ 50, denominadas de região laranja, vermelha e calcográfica. Essa última forneceu os melhores resultados, mostrando que a composição química da tinta calcográfica ou a concentração usada nas cédulas autênticas é diferente das tintas empregadas em impressoras do tipo laser e jato de tinta usadas na fabricação de cédulas falsas. Dois pigmentos foram identificados na composição da tinta calcográfica, um pigmento da classe das ftalocianinas e o pigmento diazo diarilida.

Combinando a espectroscopia Raman e a análise discriminante usando PLS foi possível determinar se os óleos essenciais extraídos da árvore *Aniba Rosaedora* eram provenientes do caule ou das folhas e galhos. A análise dos *loadings* das variáveis latentes e o gráfico da importância das variáveis (VIP) permitiu identificar que a diferença entre as fontes do óleo está na variação das quantidades dos terpenos presentes, especialmente o linalol. O óleo extraído das folhas e galhos é caracterizado por uma maior concentração de sesquiterpenos em relação ao óleo do caule, que contém uma maior quantidade de monoterpenos, esses resultados estão de acordo com dados obtidos na literatura por cromatografia gasosa.

O cálculo da incerteza na classificação de cada amostra classificada foi realizado com método de reamostragem *bootstrap* e permitiu a construção de modelos de classificação mais confiáveis, diminuindo as chances de classificação errada. A proposta do cálculo de incerteza para modelos de classificação multivariados preenche uma lacuna na validação desses modelos e abre novas perspectivas para a aplicação da espectroscopia Raman e quimiometria em química forense.

Os modelos de classificação desenvolvidos podem ser empregados como teste de triagem para a análise em documentoscopia e de óleos essenciais, oferecendo vantagens como alta velocidade de análise, menor custo e sem geração de resíduos, uma vez que não é necessário um tratamento de amostras antes das análises.

O uso da espectroscopia Raman de imagem como ferramenta analítica pode auxiliar na tomada de decisões para questões forenses. O emprego de métodos de resolução de curvas, como o ICA, mostrou-se adequado para a recuperação do perfil espectral e de concentração dos explosivos nas imagens Raman. A detecção de explosivos em cédulas com a metodologia proposta pode ser feita com exatidão, uma vez que o coeficiente de correlação entre o espectro Raman do explosivo puro e do perfil espectral recuperado pelo ICA foi acima de 0,8. O limite de detecção foi de 50 $\mu\text{g}\cdot\text{cm}^{-2}$ no pixel para o explosivo TNT.

Na comparação entre os métodos de resolução de curvas empregados no tratamento das imagens Raman, os resultados convergiram para soluções equivalentes na maioria das amostras estudadas. No entanto, foi observado que em baixas concentrações de um constituinte, o método MCR-ALS não convergiu para a solução correta. Dessa forma, foi necessária a adoção de outra estratégia, como o MCR por intervalos, onde foi possível a recuperação do constituinte em questão mesmo em baixa concentração.

Por fim, a espectroscopia Raman de imagem foi empregada no estudo da composição química de fósseis, com o objetivo de obter informações sobre os processos de diagênese e auxiliar na compreensão da evolução dos seres vivos dentro de uma abordagem interdisciplinar. Os resultados mostraram a presença de calcita no fóssil indicando que o processo de fossilização deu-se por mineralização, onde a matéria orgânica é substituída por matéria mineral. Perfis espectrais que estão relacionados com a presença de material carbonáceo também foram recuperados das imagens Raman.

As metodologias propostas empregando espectroscopia Raman de imagem e métodos multivariados de resolução de curvas permitem obter informação espectral e espacial de substâncias explosivas em superfícies de cédulas e informações da

composição de fósseis de peixes extintos há mais de 125 milhões de anos, sendo uma abertura para um novo campo de investigação em paleontologia.

CAPÍTULO 8: Referências

1. McCreery, R. L. *Raman spectroscopy for chemical analysis*. 437 (John Wiley & Sons, 2000).
2. Massart, D.L.; Vandeginste, B.G.M.; Buydens, L.M.C.; De Jong, S.; Lewi, P.J.; Smeyers-Verbeke, J. *Handbook of Chemometrics and Qualimetrics: Part A*. 886 (Elsevier, 1997).
3. Moros, J., Garrigues, S., Guardia, M. D. La. Vibrational spectroscopy provides a green tool for multi-component analysis. *TrAC Trends Anal. Chem.* **29**, 578–591 (2010).
4. Pérez, N. F., Ferré, J., Boqué, R. Calculation of the reliability of classification in discriminant partial least-squares binary classification. *Chemom. Intell. Lab. Syst.* **95**, 122–128 (2009).
5. Preisner, O., Lopes, J. A., Menezes, J. C. Uncertainty assessment in FT-IR spectroscopy based bacteria classification models. *Chemom. Intell. Lab. Syst.* **94**, 33–42 (2008).
6. Villa, J. L., Boqué, R., Ferré, J. Calculation of the probability of correct classification in probabilistic bagged k-Nearest Neighbours. *Chemom. Intell. Lab. Syst.* **94**, 51–59 (2008).
7. Zhang, L., Garcia-Munoz, S. A comparison of different methods to estimate prediction uncertainty using Partial Least Squares (PLS): A practitioner's perspective. *Chemom. Intell. Lab. Syst.* **97**, 152–158 (2009).
8. Piqueras, S., Duponchel, L., Tauler, R., de Juan, a. Monitoring polymorphic transformations by using in situ Raman hyperspectral imaging and image multiset analysis. *Anal. Chim. Acta* **819**, 15–25 (2014).
9. Vajna, B. *et al.* Testing the performance of pure spectrum resolution from Raman hyperspectral images of differently manufactured pharmaceutical tablets. *Anal. Chim. Acta* **712**, 45–55 (2012).
10. Duponchel, L., Elmi-Rayaleh, W., Ruckebusch, C., Huvenne, J. P. Multivariate curve resolution methods in imaging spectroscopy: influence of extraction methods and instrumental perturbations. *J. Chem. Inf. Comput. Sci.* **43**, 2057–67 (2003).
11. Vidal, M., Amigo, J. M. Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemom. Intell. Lab. Syst.* **117**, 138–148 (2012).
12. Bell, S. Forensic chemistry. *Annu. Rev. Anal. Chem.* **2**, 297–319 (2009).
13. Bell, S. *Forensic Chemistry*. 671 (Pearson Education, 2006).

14. Brettell, T. a, Butler, J. M., Almirall, J. R. Forensic science. *Anal. Chem.* **83**, 4539–56 (2011).
15. Branco, R. P. O. *Química Forense: ampliando o horizonte da perícia*. 232, Volume 2 (Millenium, 2012).
16. Botti, S. *et al.* Trace level detection and identification of nitro-based explosives by surface-enhanced Raman spectroscopy. *J. Raman Spectrosc.* **44**, 463–468 (2013).
17. Ali, E. M. A., Edwards, H. G. M., Scowen, I. J. Raman spectroscopy and security applications: the detection of explosives and precursors on clothing. *J. Raman Spectrosc.* **40**, 2009–2014 (2009).
18. Ali, E. M. A., Edwards, H. G. M., Scowen, I. J. In-situ detection of single particles of explosive on clothing with confocal Raman microscopy. *Talanta* **78**, 1201–3 (2009).
19. Braz, A., López-López, M., García-Ruiz, C. Raman spectroscopy for forensic analysis of inks in questioned documents. *Forensic Sci. Int.* **232**, 206–12 (2013).
20. Guedes, A. *et al.* Raman Microspectroscopy of Genuine and Fake Euro Banknotes. *Spectrosc. Lett.* **46**, 569–576 (2013).
21. López-López, M., Ferrando, J. L., García-Ruiz, C. Comparative analysis of smokeless gunpowders by Fourier transform infrared and Raman spectroscopy. *Anal. Chim. Acta* **717**, 92–9 (2012).
22. López-López, M., Ferrando, J. L., García-Ruiz, C. Dynamite analysis by Raman spectroscopy as a unique analytical tool. *Anal. Chem.* **85**, 2595–600 (2013).
23. López-López, M., García-Ruiz, C. Infrared and Raman spectroscopy techniques applied to identification of explosives. *TrAC Trends Anal. Chem.* **54**, 36–44 (2014).
24. Skenderović Božičević, M., Gajović, A., Zjakić, I. Identifying a common origin of toner printed counterfeit banknotes by micro-Raman spectroscopy. *Forensic Sci. Int.* **223**, 314–20 (2012).
25. West, M. J., Went, M. J. Detection of drugs of abuse by Raman spectroscopy. *Drug Test. Anal.* **3**, 532–8 (2011).
26. Yu, M. M. L., Sandercock, P. M. L. Principal component analysis and analysis of variance on the effects of Entellan New on the Raman spectra of fibers. *J. Forensic Sci.* **57**, 70–4 (2012).

27. Ali, E. M. A., Edwards, H. G. M. Analytical Raman spectroscopy in a forensic art context: the non-destructive discrimination of genuine and fake lapis lazuli. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **121**, 415–9 (2014).
28. McLaughlin, G., Doty, K. C., Lednev, I. K. Discrimination of human and animal blood traces via Raman spectroscopy. *Forensic Sci. Int.* **238**, 91–5 (2014).
29. Been, F., Roggo, Y., Degardin, K., Esseiva, P., Margot, P. Profiling of counterfeit medicines by vibrational spectroscopy. *Forensic Sci. Int.* **211**, 83–100 (2011).
30. Instituto Nacional de Metrologia (INMETRO). Orientações sobre validação de métodos de ensaios químicos, DQO-CGRE-008, 2003.
31. International Organization of Standardization (ISO). General requirements for the competence of testing and calibration laboratories (ISO/IEC 17025), 1995.
32. CITAC/EURACHEM Guide. Guide of quality in analytical chemistry – An aid to accreditation, 2nd ed., 2002.
33. Annual Book of ASTM Standards. Standards practices for infrared, multivariate, quantitative analysis, E1655, vol 03.06, ASTM International: West Conshohocken, 2000.
34. European Medicines Agency. Guideline on the use of near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submissions and variations, 2014.
35. Thomas, D. B., Fordyce, R. E., Frew, R. D., Gordon, K. C. A rapid , non-destructive method of detecting diagenetic alteration in fossil bone using Raman spectroscopy. *J. Raman Spectrosc.* **38**, 1533–1537 (2007).
36. Allwood, A. C., Walter, M. R., Marshall, C. P. Raman spectroscopy reveals thermal palaeoenvironments of c.3.5 billion-year-old organic matter. *Vib. Spectrosc.* **41**, 190–197 (2006).
37. Rouchon, V. *et al.* Raman and FTIR spectroscopy applied to the conservation report of paleontological collections: identification of Raman and FTIR signatures of several iron sulfate species such as ferrinatrite and sideronatrite. *J. Raman Spectrosc.* **43**, 1265-1274 (2012).
38. Bergmann, U., Manning, P. L., Wogelius, R. a. Chemical mapping of paleontological and archeological artifacts with synchrotron X-rays. *Annu. Rev. Anal. Chem.* **5**, 361–89 (2012).
39. Schopf, J. W., Kudryavtsev, A. B., Sergeev, V. N. Confocal Laser Scanning Microscopy and Raman Imagery. *J. Paleontol.* **84**, 402–416 (2010).

40. McMillan, P. F. Raman spectroscopy in mineralogy and geochemistry. *Ann. Rev. Earth Planet Sci.* **17**, 255–283 (1989).
41. Schopf, J. W., Kudryavtsev, A. B. Confocal laser scanning microscopy and Raman imagery of ancient microscopic fossils. *Precambrian Res.* **173**, 39–49 (2009).
42. Kuczumow, a. *et al.* Investigation of chemical changes in bone material from South African fossil hominid deposits. *J. Archaeol. Sci.* **37**, 107–115 (2010).
43. France, C. a. M., Thomas, D. B., Doney, C. R., Madden, O. FT-Raman spectroscopy as a method for screening collagen diagenesis in bone. *J. Archaeol. Sci.* **42**, 346–355 (2014).
44. Gobbo, S. R. Tecidos moles (não resistentes): como se fossilizam? *Terrae Didática* 2–13 (2013).
45. Schopf, J. W., Kudryavtsev, A. B. Biogenicity of Earth's earliest fossils: A resolution of the controversy. *Gondwana Res.* **22**, 761–771 (2012).
46. Marshall, C. P., Edwards, H. G. M., Jehlicka, J. Understanding the application of Raman spectroscopy to the detection of traces of life. *Astrobiology* **10**, 229–43 (2010).
47. Sala, O. *Fundamentos da espectroscopia Raman e infravermelho*. 276 (UNESP, 2008).
48. Opilik, L., Schmid, T., Zenobi, R. Modern Raman imaging: vibrational spectroscopy on the micrometer and nanometer scales. *Annu. Rev. Anal. Chem.* **6**, 379–98 (2013).
49. Faria, D. L. A., Santos, L. G. C., Gonçalves, N. S. Uma demonstração sobre o espalhamento inelástico de luz: repetindo o experimento de Raman. *Quim.Nova* **20**, 319–323 (1997).
50. Fleischmann, M. Hendra, P.J., McQuillan, A. J. Raman Spectra of Pyridine adsorbed at a Silver Electrode. *Chem. Phys. Lett.* **26**, 2–5 (1974).
51. Kneipp, K. *et al.* Single Molecule Detection Using Surface-Enhanced Raman Scattering (SERS). *Phys. Rev. Lett.* **78**, 1667–1670 (1997).
52. Fan, M., Andrade, G. F. S., Brolo, A. G. A review on the fabrication of substrates for surface enhanced Raman spectroscopy and their applications in analytical chemistry. *Anal. Chim. Acta* **693**, 7–25 (2011).
53. Schlücker, S. Surface-enhanced Raman spectroscopy: concepts and chemical applications. *Angew. Chem. Int. Ed. Engl.* **53**, 4756–95 (2014).

54. Salzer, R., Siesler, H. W. *Infrared and Raman Spectroscopic Imaging*. (Wiley-VCH, 2009).
55. Lee, E. *Raman Imaging*. **168**, (Springer Berlin Heidelberg, 2012).
56. Stewart, S., Priore, R. J., Nelson, M. P., Treado, P. J. Raman imaging. *Annu. Rev. Anal. Chem.* **5**, 337–60 (2012).
57. Lasch, P. Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging. *Chemom. Intell. Lab. Syst.* **117**, 100–114 (2012).
58. Vidal, M., Amigo, J. M. Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemom. Intell. Lab. Syst.* **117**, 138–148 (2012).
59. Lins, W. A., Junior, O. F. Contribuição do físico brasileiro Sergio Porto para as aplicações do laser e sua introdução no Brasil. *Ver. Bras. Ensino Fis.* **32**, 3601-01-3601-10 (2010).
60. Bro, R., Smilde, A. K. Principal component analysis. *Anal. Methods* **6**, 2812 (2014).
61. Barker, M., Rayens, W. Partial least squares for discrimination. *J. Chemom.* **17**, 166–173 (2003).
62. Martens, H., Martens, M. Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food Qual. Prefer.* **11**, 5–16 (2000).
63. Wold, S., Sjostrom, M. PLS-regression : a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **58**, 109–130 (2001).
64. Brereton, R. G., Lloyd, G. R. Partial least squares discriminant analysis: taking the magic away. *J. Chemom.* **28**, 213–225 (2014).
65. Hibbert, D. B., Armstrong, N. An introduction to Bayesian methods for analyzing chemistry data. *Chemom. Intell. Lab. Syst.* **97**, 211–220 (2009).
66. Efron, B., Tibshirani, R. J. *An Introduction to the Bootstrap*. 456 (Chapman and Hall/CRC, 1994).
67. Fernández Pierna, J., Jin, L., Wahl, F., Faber, N., Massart, D. . Estimation of partial least squares regression prediction uncertainty when the reference values carry a sizeable measurement error. *Chemom. Intell. Lab. Syst.* **65**, 281–291 (2003).

68. Pereira, A. C., Reis, M. S., Saraiva, P. M., Marques, J. C. Madeira wine ageing prediction based on different analytical techniques: UV-vis, GC-MS, HPLC-DAD. *Chemom. Intell. Lab. Syst.* **105**, 43–55 (2011).
69. Voet, H. V. D. Pseudo-Degrees of Freedom for Complex Predictive Models: The Example of Partial Least Squares. *J. Chemom.* **208**, 195–208 (1999).
70. Mil'man, B. L., Konopel'ko, L. A. Uncertainty of Qualitative Chemical Analysis: General Methodology and Binary Test Methods. *J. Anal. Chem.* **59**, 1128–1141 (2004).
71. Pulido, a., Ruisánchez, I., Boqué, R., Rius, F. X. Uncertainty of results in routine qualitative analysis. *TrAC Trends Anal. Chem.* **22**, 647–654 (2003).
72. EURACHEM / CITAC Guide : The Expression of Uncertainty in Qualitative Testing Expression of Uncertainty in Qualitative Testing Contents. (2003).
73. Jaumot, J., Gargallo, R., de Juan, A., Tauler, R. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemom. Intell. Lab. Syst.* **76**, 101–110 (2005).
74. De Juan, A., Tauler, R. Multivariate Curve Resolution (MCR) from 2000: Progress in Concepts and Applications. *Crit. Rev. Anal. Chem.* **36**, 163–176 (2006).
75. Pilon, A. C., Carneiro, R. L., Carnevale Neto, F., da S Bolzani, V., Castro-Gamboa, I. Interval multivariate curve resolution in the dereplication of HPLC-DAD data from *Jatropha gossypifolia*. *Phytochem. Anal.* **24**, 401–6 (2013).
76. Hyvärinen, A., Oja, E. Independent component analysis: algorithms and applications. *Neural Networks* **13**, 411–430 (2000).
77. Rutledge, D. N., Jouan-Rimbaud Bouveresse, D. Independent Components Analysis with the JADE algorithm. *TrAC Trends Anal. Chem.* **50**, 22–32 (2013).
78. Wang, G., Ding, Q., Hou, Z. Independent component analysis and its applications in signal processing for analytical chemistry. *TrAC Trends Anal. Chem.* **27**, 368–376 (2008).
79. Højen-Sørensen, P. A. D. F. R., Winther, O., Hansen, L. K. Mean-field approaches to independent component analysis. *Neural Comput.* **14**, 889–918 (2002).
80. Jouan-Rimbaud Bouveresse, D., Moya-González, a., Ammari, F., Rutledge, D. N. Two novel methods for the determination of the number of components in independent components analysis models. *Chemom. Intell. Lab. Syst.* **112**, 24–32 (2012).

81. Jaumot, J., Tauler, R. MCR-BANDS: A user friendly MATLAB program for the evaluation of rotation ambiguities in Multivariate Curve Resolution. *Chemom. Intell. Lab. Syst.* **103**, 96–107 (2010).
82. Maeder, M., Zilian, A. Evolving factor analysis, a new multivariate technique in chromatography. *Chemom. Intell. Lab. Syst.* **3**, 205–213 (1988).
83. Tocchetto, D. *Documentoscopia*. 736 (Millenium, 2014).
84. Romão, W. *et al.* Química forense: perspectiva sobre novos métodos analíticos aplicados à documentoscopia, balística e drogas de abuso. *Quim.Nova* **34**, 1717–1728 (2011).
85. Calcerrada, M., García-Ruiz, C. Analysis of questioned documents: A review. *Anal. Chim. Acta* **853C**, 143–166 (2015).
86. Falsificação de Cédulas, Banco Central do Brasil, Disponível em <http://www.bcb.gov.br/?MECIRESTFALSAS>, acessado em 15/12/2014.
87. Vila, a., Ferrer, N., Mantecón, J., Bretón, D., García, J. F. Development of a fast and non-destructive procedure for characterizing and distinguishing original and fake euro notes. *Anal. Chim. Acta* **559**, 257–263 (2006).
88. Eberlin, L. S. *et al.* Instantaneous chemical profiles of banknotes by ambient mass spectrometry. *Analyst* **135**, 2533–9 (2010).
89. Das, R. S, Agrawal, Y. K. Raman spectroscopy: Recent advancements, techniques and applications. *Vib. Spectrosc.* **57**, 163–176 (2011).
90. Kudelski, A. Analytical applications of Raman spectroscopy. *Talanta* **76**, 1–8 (2008).
91. Wise, B. M. *et al.* PLS Tollbox for use with MATLAB™, (Eigenvector Research, Inc, 2006)
92. Poon, K. W. C., Dadour, I. R., Mckinley, A. J. In situ chemical analysis of modern organic tattooing inks and pigments by micro-Raman spectroscopy. *J. Raman Spectrosc.* **39**, 1227–1237 (2008).
93. Tackley, D. R., Dent, G., Ewen Smith, W. Phthalocyanines: structure and vibrations. *Phys. Chem. Chem. Phys.* **3**, 1419–1426 (2001).
94. Schulte, F., Brzezinka, K.-W., Lutzenberger, K., Stege, H., Panne, U. Raman spectroscopy of synthetic organic pigments used in 20th century works of art. *J. Raman Spectrosc.* **39**, 1455–1463 (2008).

95. Vandenberghe, P., Moens, L., Edwards, H. G. M., Dams, R. Raman spectroscopic database of azo pigments and application to modern art studies. *J. Raman Spectrosc.* **31**, 509–517 (2000).
96. Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis, Notícias: Pau-rosa terá controle internacional mais rigoroso. Disponível em www.ibama.gov.br/publicadas/pau-rosa-tera-controle-internacional-mais-rigoroso, acessado 17/12/2014.
97. IBAMA. INSTRUÇÃO NORMATIVA N 9, DE 25 DE AGOSTO DE 2011. Disponível em www.jusbrasil.com.br/diarios/29993619/pg-73-secao-1-diario-oficial-da-uniao-dou-de-26-08-2011.
98. May, P. H., Barata, L. E. S. Rosewood Exploitation in the Brazilian Amazon : Options for Sustainable Production. *Economic Botany* **58**, 257–265 (2004).
99. Zellner, A. *et al.* Evaluation of Leaf-Derived Extracts as an Environmentally Sustainable Source of Essential Oils by Using Gas Chromatography - Mass Spectrometry and Enantioselective Gas Chromatography - Olfactometry. *Anal. Chem.* **78**, 883–890 (2006).
100. Fidelis, C. H. V., Sampaio, P. T. B., Krainovic, P. M., Augusto, F., Barata, L. E. S. Correlation between maturity of tree and GC×GC–qMS chemical profiles of essential oil from leaves of *Aniba rosaeodora* Ducke. *Microchem. J.* **109**, 73–77 (2013).
101. Fidelis, C. H. V., Augusto, F., Sampaio, P. T. B., Krainovic, P. M., Barata, L. E. S. Chemical characterization of rosewood (*Aniba rosaeodora* Ducke) leaf essential oil by comprehensive two-dimensional gas chromatography coupled with quadrupole mass spectrometry. *J. Essent. Oil Res.* **24**, 245–251 (2012).
102. Daferera, D. J., Tarantilis, P. A., Polissiou, M. G. P. Characterization of Essential Oils from Lamiaceae Species by Fourier Transform Raman Spectroscopy. *J. Agric. Food Chem.* **50**, 5503–5507 (2002).
103. Schulz, H., Baranska, M. Identification and quantification of valuable plant substances by IR and Raman spectroscopy. *Vib. Spectrosc.* **43**, 13–25 (2007).
104. Baranska, M. *et al.* Vibrational spectroscopic studies to acquire a quality control method of Eucalyptus essential oils. *Biopolymers* **78**, 237–48 (2005).
105. Seidler-Lozykowska, K., Baranska, M., Baranski, R., Krol, D. Raman analysis of caraway (*Carum carvi* L.) single fruits. Evaluation of essential oil content and its composition. *J. Agric. Food Chem.* **58**, 5271–5 (2010).

106. López-López, M., García-Ruiz, C. Infrared and Raman spectroscopy techniques applied to identification of explosives. *TrAC Trends Anal. Chem.* **54**, 36–44 (2014).
107. Glasttstein, B. Kit for detecting traces of explosives - contg. DMSO soln. of tetra:alkyl:ammonium hydroxide and azo dye forming system, EP264252-A2 ; EP264252-A ; JP63109369-A ; US4788039-A ; AU8780477-A ; BR8706484-A ; IL80311-A ; EP264252-B ; DE3769430-G ; ES2021722-B ; CA1309933-C ; JP2580200-B2.
108. Sharma, S. K., Misra, A. K., Sharma, B. Portable remote Raman system for monitoring hydrocarbon, gas hydrates and explosives in the environment. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **61**, 2404–12 (2005).
109. Stewart, S. P. *et al.* Determination of hydrogen peroxide concentration using a handheld Raman spectrometer: Detection of an explosives precursor. *Forensic Sci. Int.* **216**, e5–8 (2012).
110. Moros, J., Lorenzo, J. A., Novotný, K., Laserna, J. J. Fundamentals of stand-off Raman scattering spectroscopy for explosive fingerprinting. *J. Raman Spectrosc.* **44**, 121–130 (2013).
111. Tripathi, A. *et al.* Semi-automated detection of trace explosives in fingerprints on strongly interfering surfaces with Raman chemical imaging. *Appl. Spectrosc.* **65**, 611–9 (2011).
112. Estatística de ataques a banco. Contraf-CUT CNVT. 7th National Survey on Attacks on Banks in Brazil. 2013. Disponível em www.contrafcut.org.br.
113. Sabin, G. P., Souza, A. M., Breitreitz, M. C., Poppi, R. J. Desenvolvimento de um algoritmo para identificação e correção de spikes em espectroscopia Raman. *Quim. Nova* **35**, 612–615 (2012).
114. Lewis, I. R., Daniel, N. W., Griffiths, P. R. Interpretation of Raman Spectra of Nitro-Containing Explosive Materials. Part I: Group Frequency and Structural Class Membership. *Appl. Spectrosc.* **51**, 1854–1867 (1997).
115. Gruzdkov, Y. A., Gupta, Y. M. Vibrational Properties and Structure of Pentaerythritol Tetranitrate. *J. Phys. Chem. A* **105**, 6197–6202 (2001).
116. Schweitzer, M. H., Avci, R., Collier, T., Goodwin, M. B. Microscopic, chemical and molecular methods for examining fossil preservation. *Comptes Rendus Palevol*, **7**, 159–184 (2008).
117. Jehlička, J. , Edwards, H. G. M. Raman spectroscopy as a tool for the non-destructive identification of organic minerals in the geological record. *Org. Geochem.* **39**, 371–386 (2008).

118. Lindgren, J. *et al.* Microspectroscopic evidence of cretaceous bone proteins. *PLoS One* **6**, e19445 (2011).
119. Bernard, S. *et al.* Exceptional preservation of fossil plant spores in high-pressure metamorphic rocks. *Earth Planet. Sci. Lett.* **262**, 257–272 (2007).
120. Alvarado-Ortega, J. *et al.* Exceptional Preservation of Soft Tissues in Cretaceous Fishes From the Tlayua Quarry, Central Mexico. *Palaios* **22**, 682–685 (2007).
121. Reisz, R. R. *et al.* Embryology of Early Jurassic dinosaur from China with evidence of preserved organic remains. *Nature* **496**, 210–4 (2013).
122. Edwards, N. P. *et al.* Infrared mapping resolves soft tissue preservation in 50 million year-old reptile skin. *Proc. Biol. Sci.* **278**, 3209–18 (2011).
123. Schulze, H. G., Foist, R. B., Okuda, K., Ivanov, A., Turner, R. F. B. A small-window moving average-based fully automated baseline estimation method for Raman spectra. *Appl. Spectrosc.* **66**, 757–64 (2012).
124. Sun, J., Wu, Z., Cheng, H., Zhang, Z., Frost, R. L. A Raman spectroscopic comparison of calcite and dolomite. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **117**, 158–62 (2014).
125. Sforza, M. C., van Zuilen, M. A., Philippot, P. Structural characterization by Raman hyperspectral mapping of organic carbon in the 3.46 billion-year-old Apex chert, Western Australia. *Geochim. Cosmochim. Acta* **124**, 18–33 (2014).
126. Janko, M., Stark, R. W., Zink, A. Preservation of 5300 year old red blood cells in the Iceman. *J. R. Soc. Interface* **9**, 2581–90 (2012).