

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Programa de Pós-Graduação em Física

Thiago de Oliveira Mendes

**Métodos multivariados para a elucidação de informações analíticas em
amostras complexas.**

Juiz de Fora
2015

Thiago de Oliveira Mendes

Métodos multivariados para a elucidação de informações analíticas em amostras complexas.

Tese apresentada ao Programa de Pós-Graduação em Física da Universidade Federal de Juiz de Fora, na área de concentração em Física da matéria condensada, como requisito parcial para obtenção do título de Doutor em Física.

Orientadora: Dra. Maria José Valenzuela Bell

Coorientador: Dr. Marcone Augusto Leal de Oliveira

Juiz de Fora

2015

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Mendes, Thiago de Oliveira.

Métodos multivariados para a elucidação de informações analíticas em amostras complexas. / Thiago de Oliveira Mendes. – 2015.

119 f. : il.

Orientadora: Dra. Maria José Valenzuela Bell

Coorientador: Dr. Marccone Augusto Leal de Oliveira

Tese (Doutorado) – Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós-Graduação em Física, 2015.

1. Análise multivariada. 2. Infravermelho. 3. Raman. 4. Técnicas de separação. 5. Amostras complexas. I. Bell, Maria José Valenzuela, orient. II. Oliveira, Marccone Augusto Leal de, coorient. III. Título.

Thiago de Oliveira Mendes

Métodos multivariados para a elucidação de informações analíticas em amostras complexas.

Tese apresentada ao Programa de Pós-Graduação em Física da Universidade Federal de Juiz de Fora, na área de concentração em Física da matéria condensada, como requisito parcial para obtenção do título de Doutor em Física.

Aprovada em: 4 de agosto de 2015

BANCA EXAMINADORA

Professora Dra. Dra. Maria José Valenzuela Bell -
Orientadora
Universidade Federal de Juiz de Fora

Professor Dr. Dr. Marcone Augusto Leal de Oliveira -
Coorientador
Universidade Federal de Juiz de Fora

Professor Dr. Alvaro Vianna Novaes de Carvalho
Teixeira
Universidade Federal de Viçosa

Professor Dr. Italo Tuler Perrone
Universidade Federal de Viçosa

Professor Dr. Virgílio de Carvalho dos Anjos
Universidade Federal de Juiz de Fora

Professora Dra. Georgia Maria Amaral Junqueira
Universidade Federal de Juiz de Fora

AGRADECIMENTOS

Agradeço a minha esposa Brenda, pessoa com a qual eu mais discuti, em todos os âmbitos, todas as aplicações aqui apresentadas. Companheira na vida e também nas atividades acadêmicas. Sou privilegiado de ter a sua companhia. Obrigado.

Agradeço a oportunidade que me foi dada pela minha orientadora, a professora Dra. Maria José Valenzuela Bell, pela convivência nestes seis anos de trabalho, pelos ensinamentos e pelo apoio as minhas decisões. Ao chefe de grupo, professor Dr. Virgílio de Carvalho dos Anjos, pelas conversas e por me acolher na equipe do Laboratório de Espectroscopia de Materiais. Ao professor Dr. Marcone Augusto Leal de Oliveira pela onipresente coorientação. Muito obrigado a vocês! Agradeço ainda às agências que financiaram projetos, viagens e bolsa de estudos: UFJF, CNPQ, Capes e Fapemig.

Quanto a mobilidade acadêmica no ano 2012, agradeço ao meu coorientador que propiciou o contato com a professora Dra. Marina Franco Maggi Tavares (USP) e a Dra. Coral Barbas Arribas (CEMBIO, Universidad San Pablo - CEU). Agradeço também a todo o aprendizado compartilhado pelo professor Santiago Angulo Díaz-Parreño (San Pablo - CEU) e a Fundação EADS-CASA pela bolsa de estudos. O aprendizado foi enorme.

Agradeço a contribuição dos professores Dr. Virgílio e Dr. Marco Antônio Moreira Furtado por compor a banca examinadora de pré-defesa de Tese, e aos colegas de grupo Dr. Geraldo Henriques Silva e Dra. Mariana Brandão pelas correções no texto.

Por fim, mas não menos importante, agradeço as minhas famílias. A que me educou: minha mãe Maria Aparecida, meu pai José Calixto, meu irmão mais velho Alexandre e a caçula Ana Carolina, são os pilares de minha formação. Sogra e cunhada também são família, Inês e Cindy, um beijo, e também ao meu sogro Ely, *in memoriam*. A família que eu encontrei ao longo destes anos de faculdade, em especial aos três grupos de pesquisa que eu participei durante a pós-graduação, família LEM: Wesley, Geraldo, Rodrigo, Leonarde, Alysson, Mariana, Thamiris, Luiz, Leandro e Roney; família GQAQ: Rafael, Sandro, Fernando, Paula, Carina, Isaura e Lucas; família CEMBIO: Marcel, Thiago, Dani, Juliana, Claudia, Gisele, Kenia, Davi e Alicia.

E aos mentores que orientaram em diferentes estágios de minha formação acadêmica, seguramente me inspiram e moldaram meu perfil de pesquisa: José Luiz M. Valle, José Roberto Tagliati e José Maurício de Andrade (monitoria, tutoria e estágio - licenciatura em Física), José Paulo R.F. Mendonça (iniciação científica - bacharelado em Física), Ana Carolina de M. Zeri (bolsa de verão LNLS), Frederico José A. Lopes (especialização em Educação Matemática), Coral Barbas e Santiago Angulo (sanduíche - Espanha) e Maria José V. Bell, Virgílio de C. dos Anjos e Marcone A.L. de Oliveira (mestrado e doutorado em Física).

Se o vento soprar de uma única direção,
a árvore crescerá inclinada.

RESUMO

O estudo de amostras complexas, como alimentos, fármacos e fluidos biológicos necessita de métodos robustos de análise de dados, uma vez que estas amostras podem conter informações de dezenas a milhares de constituintes. A pesquisa desenvolvida neste trabalho tem como temática principal a aplicação de métodos multivariados de análise para extrair informações a partir dos espectros vibracionais (infravermelho e Raman) ou de massas, de amostras complexas. Técnicas de separação tais como CE, GC e LC são utilizadas como abordagem auxiliar para fornecer informações aos modelos multivariados propostos. Cinco aplicações são discutidas no texto. A primeira faz o uso da PCA sobre cromatogramas de GC de diferentes amostras de óleo de soja e azeite de oliva, buscando realizar uma análise exploratória das duas classes de óleos vegetais. Na sequência são analisadas misturas com adições controladas de óleo de soja em azeite de oliva pelas técnicas de absorção infravermelha e espalhamento Raman, com o objetivo de propor uma regressão PLS que determine o percentual de adição. A segunda aplicação aborda a busca de marcadores vibracionais, infravermelho e Raman, que possam ser utilizados para a quantificação do percentual de gordura em amostras de leite fluido. Neste estudo foi realizada uma análise do perfil de ácidos graxos presentes no leite pela técnica de GC, com o objetivo de selecionar uma molécula representativa de triacilglicerol que foi utilizada para simular espectros vibracionais por modelos teóricos. A partir dos espectros obtidos por simulação computacional foi proposta a atribuição de modos vibracionais presentes na gordura do leite. Foi também proposto um modelo MLR preditor do percentual de gordura em amostras de leite fluido, cujas variáveis vibracionais foram previamente selecionadas a partir de um modelo PLS. Outra aplicação foi a utilização de modelo PLS sobre espectros Raman para a determinação simultânea das concentrações de rifampicina, isoniazida, etambutol e pirazinamida em amostras de comprimidos utilizados no tratamento anti-tuberculose. A quarta aplicação está focada na quantificação do percentual de soro de leite adicionado fraudulentamente em amostras de leite por análise do perfil de ácidos graxos por GC e CE. Uma análise discriminante foi utilizada para selecionar marcadores de ácidos graxos que foram monitorados por regressão MLR. Por fim, é apresentada a elaboração de um tutorial completo de análise univariada e multivariada de dados, desenvolvido em *R-software*, para a determinação de biomarcadores em experimentos de metabolômica baseados em análises de espectros de massas. Todas estas aplicações têm como interseção o uso de métodos multivariados de análise de dados como ferramenta principal para propor novos marcadores vibracionais/metabólitos, assim como, métodos alternativos para a quantificação de diferentes analitos.

Palavras-chave: Análise multivariada. Infravermelho. Raman. Técnicas de separação. Amostras complexas.

ABSTRACT

The study of complex samples such as food, drugs, and biological fluids require a robust data analysis methods, since these samples may contain information tens of thousands of constituents. The research developed in this work has as main theme the application of multivariate analysis methods to extract information from the vibrational spectra (infrared and Raman) or mass spectra of complex samples. Separation techniques such as CE, GC and LC are used as an auxiliary approach to providing information to the proposed multivariate models. Five applications are discussed in the text. The first makes use of PCA on GC chromatograms of samples of soybean oil and olive oil, searching for an exploratory analysis of the two classes of vegetable oils. Following are analyzed with mixtures of controlled additions of soybean oil in olive oil, by infrared absorption and Raman scattering techniques, with the aim of proposing a PLS regression to determine the percentage of addition. A second application addresses the search vibrational markers, infrared and Raman, that may be used to quantify the percentage of fat in milk fluid samples. In this study we carried out a profile analysis of fatty acids present in milk by CG technique, for the purpose of selecting a representative molecule of triacylglycerol was used to simulate vibrational spectra by theoretical models. From the spectra obtained by computer simulation was proposed assignment of vibrational modes present in milk fat. It was also proposed an MLR predictor model of the percentage of fat in fluid samples, whose vibrational variables were previously selected from a PLS model. Another application has been the use of PLS model of Raman spectra for the simultaneous determination of concentrations of rifampicin, isoniazid, ethambutol and pyrazinamide in tablet samples used in anti-tuberculosis treatment. The fourth application is focused on quantifying the percentage of whey added fraudulently in milk samples for analysis of the fatty acid profile by GC and CE. A discriminant analysis was used to select fatty acids markers that were monitored by MLR regression. Finally, it shows the preparation of a complete tutorial of univariate and multivariate data analysis, developed in R-software, for the determination of biomarkers in metabolomics experiments based on mass spectra analysis. All these applications have the intersection as the use of multivariate methods of analyzing data as the main tool to propose new vibrational / metabolite markers as well as alternative methods for quantification of various analytes.

Key-words: Multivariate analysis. Infrared. Raman. Separation Techniques. Complex samples.

LISTA DE ILUSTRAÇÕES

Figura 1 – Análise de Componentes Principais.	19
Figura 2 – Interferência entre dois feixes de luz que percorrem caminhos diferentes.	25
Figura 3 – Representação das funções $f(t)$, $g(\omega)$ e $G(\omega)$	30
Figura 4 – Medida da transmitância - MIR e NIR.	31
Figura 5 – Medida do espalhamento Raman.	31
Figura 6 – Modelo para absorção da radiação eletromagnética.	32
Figura 7 – Modelo de molécula diatômica.	34
Figura 8 – Diagrama de níveis de energia de bandas no infravermelho.	35
Figura 9 – Regras de seleção para Infravermelho e Raman.	36
Figura 10 – Efeito Raman.	37
Figura 11 – Esquema de separação em cromatografia a gás.	39
Figura 12 – Esquema de separação em cromatografia a líquido.	40
Figura 13 – Esquema de separação em eletroforese capilar.	40
Figura 14 – Perfil de ácidos graxos obtidos pela técnica de GC-FID.	44
Figura 15 – PCA das amostras de óleos vegetais a partir dos dados de GC-FID.	46
Figura 16 – Espectros das soluções estoque de OS e AOEV.	48
Figura 17 – PCA dos espectros NIR de óleos vegetais.	50
Figura 18 – Regressão PLS a partir dos espectros NIR.	51
Figura 19 – PCA dos espectros MIR de óleos vegetais.	52
Figura 20 – Regressão PLS a partir dos espectros MIR.	53
Figura 21 – PCA dos espectros Raman de óleos vegetais.	54
Figura 22 – Regressão PLS a partir dos espectros Raman.	55
Figura 23 – Perfil de AG pela técnica de GC para uma amostra de leite de vaca.	60
Figura 24 – Molécula TAG otimizada por cálculo computacional.	62
Figura 25 – Espectros Raman e infravermelho da gordura do leite de vaca.	63
Figura 26 – Espectros Raman de amostras de leite fluido com diferentes concentrações de gordura e gráfico de variáveis importantes da regressão PLS.	65
Figura 27 – Resultados para quantificação a partir dos espectros Raman com regressão MLR.	66
Figura 28 – Espectros de infravermelho de amostras de leite fluido com diferentes concentrações de gordura e gráfico de variáveis importantes da regressão PLS.	67
Figura 29 – Resultados para quantificação a partir dos espectros de infravermelho com regressão MLR.	68
Figura 30 – Análise de fármacos anti-tuberculose pela técnica de LC.	72
Figura 31 – Estrutura molecular dos quatro fármacos analisados.	72
Figura 32 – Espectros Raman do conjunto de fármacos analisados.	74

Figura 33 – Correlação entre os valores previstos pelo modelo PLS-Raman <i>versus</i> e valores verdadeiros obtidos pelo método LC.	76
Figura 34 – Perfil de AG de leite de vaca e soro de leite obtidos por GC.	80
Figura 35 – Análise multivariada por OPLS-DA a partir dos dados de GC das amostras de leite e de soro de leite.	81
Figura 36 – Análise das amostras de leite e soro de leite por CE.	82
Figura 37 – Quantificação de AG por CE para leite e soro de leite.	83
Figura 38 – Resultados para a quantificação do percentual de soro de leite adicionado ao leite fluido, por CE e regressão MLR.	84
Figura 39 – Análise exploratória de publicações pelo diagrama de Venn.	85
Figura 40 – Forma da entrada da tabela de dados no <i>software R</i>	87
Figura 41 – Fluxograma para obtenção da tabela que será utilizada na análise estatística dos dados.	88
Figura 42 – Mapa de valores perdidos.	89
Figura 43 – Diagrama de tomada de decisão para análise estatística univariada.	91
Figura 44 – Diagrama para análise de dados por estatística multivariada.	98
Figura 45 – Proporção de variância explicada pelo modelo PCA em função do número de componentes principais.	100
Figura 46 – Análise do diagrama de <i>scores</i> para o modelo de PCA.	101
Figura 47 – Análise do diagrama de <i>scores</i> para o modelo PLS-DA.	102
Figura 48 – Resumo dos resultados gráficos para análise multivariada obtidos pelo <i>software SIMCA</i>	103
Figura 49 – Artigo - Óleos vegetais - NIR, MIR e Raman.	115
Figura 50 – Artigo - Resíduos de fármacos em leite - NIR.	115
Figura 51 – Artigo - Gordura Trans - CE e GC.	116
Figura 52 – Artigo - Instrumentação.	116
Figura 53 – Artigo - Ensino de física.	116

LISTA DE TABELAS

Tabela 1 – Graus de liberdade de moléculas poliatômicas.	34
Tabela 2 – Quantificação de ácidos graxos por GC.	45
Tabela 3 – p -valores do teste T^2 de Hotelling para a comparação entre os vetores de composição química quantificados por GC-FID.	46
Tabela 4 – Valores previstos pelo modelo de regressão PLS.	56
Tabela 5 – Quantificação dos AG da amostra de leite cru analisados por GC-FID.	61
Tabela 6 – Atribuição dos modos vibracionais Raman para a gordura de leite de vaca.	64
Tabela 7 – Atribuição dos modos vibracionais de absorção infravermelha para a gordura de leite de vaca.	64
Tabela 8 – Limites de detecção e quantificação Raman e infravermelho para os modos vibracionais selecionados.	68
Tabela 9 – Coeficientes de determinação e RMSE do ajuste PLS para os fármacos RIF, INH, PYZ e EMB.	75
Tabela 10 – Predição das concentrações de RIF, INH, PYZ e EMB em comprimidos comerciais.	77
Tabela 11 – Resumo de resultados da análise univariada. <i>R-software</i>	95
Tabela 12 – Resumo dos resultados da análise univariada no <i>software</i> Matlab.	97
Tabela 13 – Resultados para a análise estatística multivariada. <i>R-Software</i>	102
Tabela 14 – Resultados para a análise estatística multivariada. <i>Software</i> SIMCA.	103

LISTA DE ABREVIATURAS E SIGLAS

AG	Ácidos graxos
AOEV	Azeite de oliva extra virgem
CE	Eletroforese capilar
DA	Análise discriminante
DAD	Arranjo de diodos
EA	Ácido esteárico
EG	Extrato de gordura
EMB	Etambutol
FAME	Ésteres metílicos de ácidos graxos
FID	Ionização por chama
FT	Transformada de Fourier
GC	Cromatografia a gás
INH	Isoniazida
LA	Ácido linoleico
LC	Cromatografia a líquido
LD	Limite de detecção
LNA	Ácido linolênico
LQ	Limite de quantificação
LV	Variável latente
MIR	Infravermelho médio
MLR	Regressão linear múltipla
MS	Espectrometria de massas
MTB	Mycobacterium tuberculosis
NIPALS	Iterativo mínimos quadrados parciais não-linear
NIR	Infravermelho próximo

OA	Ácido oleico
OPLS-DA	Análise discriminante por projeção em estruturas latentes ortogonal
OS	Óleo de soja
PA	Ácido palmítico
PC	Componente Principal
PCA	Análise de componentes principais
PI	Padrão interno
PLS	Mínimos quadrados parciais
PLS-DA	Análise discriminante por projeção em estruturas latentes
POA	Ácido palmitoléico
PYZ	Pirazinamida
Q^2	Fração de variância que pode ser prevista
QC	Amostras de controle de qualidade
R^2X	Percentual de variância explicada da matriz X
R^2Y	Percentual de variância explicada da matriz Y
RIF	Rifampicina
$RMSE_{cv}$	Raiz do erro quadrático médio obtido por validação cruzada
RMSEC	Raiz do erro quadrático médio para o conjunto de calibração
RMSEP	Raiz do erro quadrático médio para o conjunto de predição
RSD	Desvio padrão relativo
TAG	Triacilglicerol
TR	Tempo de retenção
VIP	Variáveis mais importantes na projeção

SUMÁRIO

1	INTRODUÇÃO	15
2	MÉTODOS MULTIVARIADOS	17
2.1	REGRESSÃO LINEAR MÚLTIPLA	17
2.2	ANÁLISE DE COMPONENTES PRINCIPAIS	18
2.3	REGRESSÃO POR MÍNIMOS QUADRADOS PARCIAIS	19
2.4	ANÁLISE DISCRIMINANTE	20
2.5	PARÂMETROS QUE INDICAM A QUALIDADE DE MODELOS MULTIVARIADOS	21
3	TÉCNICAS ESPECTROSCÓPICAS	24
3.1	ESPECTROSCOPIA COM TRANSFORMADA DE FOURIER	24
3.2	ABSORÇÃO INFRAVERMELHA	31
3.3	ESPALHAMENTO RAMAN	35
4	TÉCNICAS DE SEPARAÇÃO	39
4.1	CROMATOGRAFIA A GÁS	39
4.2	CROMATOGRAFIA A LÍQUIDO	39
4.3	ELETROFORESE CAPILAR	40
5	APLICAÇÕES	41
5.1	QUANTIFICAÇÃO DA ADULTERAÇÃO DE AZEITE DE OLIVA COM ÓLEO DE SOJA POR ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO, MÉDIO E RAMAN	41
5.1.1	Objetivos	41
5.1.2	Materiais e métodos	42
5.1.3	Resultados e discussões	44
5.1.4	Conclusões	56
5.2	ANÁLISE DO PERFIL LIPÍDICO DE AMOSTRAS DE LEITE DE VACA EM ESPECTROS DE INFRAVERMELHO MÉDIO E RAMAN	57
5.2.1	Objetivos	58
5.2.2	Materiais e métodos	58
5.2.3	Resultados e discussões	60
5.2.4	Conclusões	69

5.3	QUANTIFICAÇÃO SIMULTÂNEA DOS TEORES DE RIFAMPICINA, ISONIAZIDA, PIRAZINAMIDA E ETAMBUTOL EM COMPRIMIDOS PARA TRATAMENTO ANTI-TUBERCULOSE POR ESPECTROSCÓPIA RAMAN	69
5.3.1	Objetivos	70
5.3.2	Materiais e métodos	70
5.3.3	Resultados e discussões	71
5.3.4	Conclusões	78
5.4	ANÁLISE DO PERFIL DE LIPÍDEOS EM AMOSTRAS DE LEITE DE VACA E SORO DE LEITE POR GC E CE COMO METODOLOGIA ALTERNATIVA PARA DETERMINAÇÃO DE FRAUDE	78
5.4.1	Objetivos	78
5.4.2	Materiais e métodos	79
5.4.3	Resultados	79
5.4.4	Conclusões	84
5.5	TUTORIAL PARA ANÁLISE DE DADOS DE ESTUDOS DE METABOLÔMICA	85
5.5.1	Objetivos	86
5.5.2	Materiais e métodos	86
5.5.3	Resultados e discussões	87
5.5.4	Conclusões	104
6	CONSIDERAÇÕES GERAIS E PERSPECTIVAS	105
	REFERÊNCIAS	107
	APÊNDICE A – Trabalhos publicados	115
	APÊNDICE B – Participação em Eventos	117
	ANEXO A – Colaborações em andamento	119

1 INTRODUÇÃO

A compreensão de sistemas simplificados, ou idealizados, é o ponto de partida para o entendimento de casos mais realísticos, uma vez que o nível de dificuldade do tratamento algébrico e das interpretações gráficas quando se tem um número reduzido de variáveis é mais simples que problemas que envolvem muitas variáveis. Por outro lado um número representativo de publicações compreendidas na área de ciências naturais, como física, química e biologia, busca uma compreensão mais global dos sistemas em estudo, ou seja, no lugar de avaliar parâmetros isoladamente, tentam avaliar um conjunto de variáveis simultaneamente [1–5].

Obter informações que representam todo o conjunto amostral analisando um número tão expressivo de variáveis exige um tratamento estatístico dos dados diferenciado. Avaliar espectros de amostras complexas como fármacos, alimentos e amostras biológicas, em geral, pode exigir a análise de um número muito elevado de variáveis espectrais de uma só vez. Neste âmbito métodos multivariados de análise [6], como a Análise de Componentes Principais (PCA) [7–10] e a regressão por mínimos quadrados parciais (PLS) [11, 12] tornam-se cada vez mais utilizados, pois permitem explorar as regiões espectrais mais significativas a partir de métodos de redução de dimensão ou ainda ajustar modelos de previsão baseados em toda a informação espectral.

Esta tese tem como objetivo aplicar diferentes métodos multivariados de análise sobre dados de técnicas espectroscópicas, infravermelho médio (MIR), próximo (NIR) e Raman e também de dados oriundos de técnicas de separação, como cromatografia a gás (GC), cromatografia a líquido (LC) e eletroforese capilar (CE), aplicados na análise de óleos vegetais, leite de vaca, fármacos e fluidos biológicos.

Inicialmente são apresentados os princípios da análise multivariada e a física envolvida nas técnicas espectroscópicas utilizadas, a espectroscopia com Transformada de Fourier (FT), a absorção infravermelha e o espalhamento Raman e também, brevemente, as técnicas químicas de separação, GC, LC e CE que foram utilizadas, alternadamente, em todos os estudos como técnicas de referência para ajustar os modelos estatísticos e para comparar resultados obtidos por diferentes técnicas.

Na sequência cinco aplicações são discutidas. Um estudo que aborda o tema da adulteração de amostras de azeite de oliva extra-virgem pela adição de óleo de soja, utilizando espectroscopia NIR, MIR e Raman juntamente com métodos multivariados de análises e GC para caracterização das amostras. Um estudo da fração lipídica do leite de vaca, com caracterização inicial por GC, simulação computacional dos espectros Raman e infravermelho da molécula de triacilglicerol mais abundante no leite além de medições experimentais MIR e Raman para propor um modelo de previsão da concentração de gordura no leite fluido.

A terceira aplicação propõe a quantificação simultânea de quatro fármacos utilizados no tratamento anti-tuberculose através de uma regressão PLS a partir dos espectros Raman de comprimidos. A quarta aplicação apresenta uma nova abordagem da quantificação da fraude no leite por adição de soro de leite, baseada na análise do perfil de ácidos graxos e, como última aplicação, é proposto um tutorial para análise de dados que auxilia a busca por novos biomarcadores em experimentos de metabolômica com dados de espectrometria de massas (MS).

Portanto, ao final, busca-se mostrar como métodos multivariados podem ser utilizados em estudos experimentais para extrair informações, principalmente quando não se conhece muito bem a amostra em estudo, sugerindo novos marcadores vibracionais/metabolitos, assim como métodos alternativos para a quantificação de diferentes analitos.

2 MÉTODOS MULTIVARIADOS

A determinação da concentração de uma espécie química, por exemplo, pela aplicação direta da Lei de Beer-Lambert [13], em um único comprimento de onda, é uma abordagem com bons resultados para análise de diluições sucessivas de padrões, amostras altamente derivatizadas e purificadas, ou analitos que absorvam exclusivamente em um único comprimento de onda.

Tais condições permitem concluir que a abordagem univariada se aplica a casos isolados, uma vez que a natureza de propriedades químicas são inerentemente multivariadas, por exemplo, em uma determinação espectroscópica muitas interações ocorrem simultaneamente (momento de dipolo, absorção, espalhamento, conformações moleculares, interação com a matriz, temperatura, etc) de modo que restringir uma quantificação a um só comprimento de onda pode acarretar em perda de informação. Será que é satisfatório prever a atividade biológica de um composto em uma amostra de plasma medindo exclusivamente uma propriedade? E a interação deste composto com os demais fica negligenciada? É viável desenvolver uma nova droga farmacêutica e caracterizá-la por apenas um único modo vibracional?

Análises por um só marcador estão mais propícias a falsos positivos ou negativos. Neste sentido a abordagem multicomponente torna-se mais fidedigna ao analito estudado e permite caracterizar, e quantificar, com maior representatividade e exatidão. Esta tese tem como determinações experimentais a análise de amostras complexas reais. Por isso, na sequência, serão descritos os principais métodos de análise multivariada.

2.1 REGRESSÃO LINEAR MÚLTIPLA

A regressão, ou curva analítica, é uma relação matemática que estabelece uma função entre um conjunto de variáveis (por exemplo concentração, \mathbf{Y}) e um conjunto de medidas (por exemplo absorvâncias, \mathbf{X}). A Regressão Linear Múltipla (MLR) [14, 15], equação 2.1, é uma ferramenta utilizada quando um analito absorve em mais de um comprimento de onda ou quando mais de uma quantificação é feita em uma única amostra, além de reduzir problemas de ruídos na medição e/ou interferentes desconhecidos.

$$Y = XB \tag{2.1}$$

Em espectroscopia um exemplo seria analisar uma amostra com mais de um analito de interesse, e cada um destes analitos tendo uma ou mais absorções características em diferentes comprimentos de onda. No lugar de uma regressão para cada analito a abordagem MLR permite a quantificação de todos, de uma só vez, por meio de um cálculo matricial, onde, para tal, é necessário conhecer de antemão as regiões de absorção de interesse e a quantidade de cada analito para a construção da curva de calibração. A

matriz dos coeficientes da regressão MLR, matriz \mathbf{B} , é obtida pela equação 2.2.

$$B = (X^T X)^{-1} X^T Y \quad (2.2)$$

Uma vez obtida a matriz dos coeficientes da regressão MLR (eq. 2.2) um novo conjunto de amostras pode ser analisado, e as concentrações dos analitos que se deseja determinar podem ser obtidas diretamente pela eq. 2.1. Para inverter a matriz $X^T X$ é necessário um conjunto representativo de amostras, com diferentes níveis de concentração e que estes não estejam correlacionados, para garantir que os vetores de concentrações sejam independentes uns dos outros, (matriz singular). Uma vantagem da simplicidade dos cálculos de MLR é que estes podem ser implementados em planilhas de cálculos como Microsoft Excel ou LibreOffice Calc, sem a necessidade de *softwares* específicos.

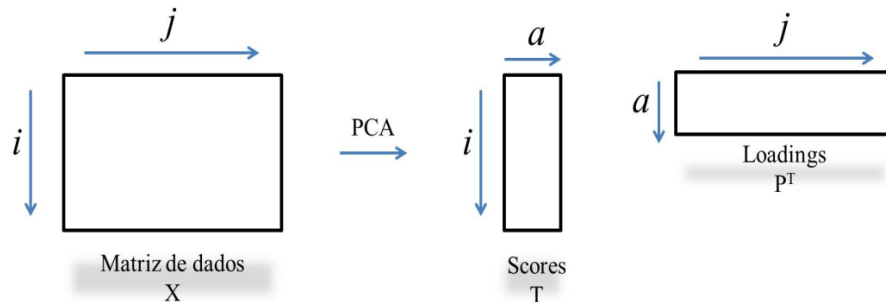
2.2 ANÁLISE DE COMPONENTES PRINCIPAIS

A PCA é uma técnica que define um novo conjunto de variáveis, que consiste numa combinação linear das variáveis originais em que, o primeiro eixo deste novo subespaço tem a característica de ser a direção que contem a maior variação dos dados originais. As demais variáveis deste subespaço são ortogonais as anteriores e descrevem a variação restante dos dados originais. Desta forma, PCA é um método de redução de dimensão. Os dados de entrada numa análise PCA podem ser uma matriz de espectros completos \mathbf{X} , de um grupo de amostras. A ideia da PCA é propor um novo espaço de variáveis, de menor dimensão, que contenha a informação mais relevante dos dados originais. Neste novo espaço, as coordenadas das amostras são chamadas de *scores* \mathbf{T} e os coeficientes das combinações lineares das variáveis originais são chamadas de *loadings* \mathbf{P} . O termo Componente Principal (PC) pode se referir tanto para *scores* como para *loadings*. A multiplicação da matriz de *scores* e *loadings* se aproxima da matriz de dados originais \mathbf{X} .

$$\tilde{X} = T_a P_a^T \quad (2.3)$$

sendo o subscrito a , na eq. 2.3, indica o número de componentes utilizado no modelo. O maior número de PCs é o menor número de linhas e colunas da matriz $a_{max} = \min(n, p)$, se $a = a_{max}$ a aproximação $\tilde{X} = X$ se torna uma igualdade. A Figura 1, a seguir, ilustra a ideia da decomposição da matriz de dados originais em *scores* e *loadings*.

Figura 1 – Análise de Componentes Principais.



i representa o número de amostras e j o número de variáveis.

Fonte: O AUTOR.

As PCs são combinações ortogonais da variáveis definidas de acordo com: a variância dos *scores* seja máxima, a soma euclidiana das distâncias entre os *scores* seja máxima e a reconstrução da matriz \mathbf{X} seja o mais próxima da matriz original ($\|X - \tilde{X}\|$ seja mínimo). Para satisfazer este critério são utilizados métodos computacionais, estas rotinas já estão implementadas em softwares como Matlab, R e SIMCA.

Para um usuário de análise PCA não é necessário conhecer todos os passos matemáticos para obtenção do modelo multivariado. Este processo é obtido por uma sequência de cálculos de álgebra de matrizes, onde se busca que a diferença $\|X - \tilde{X}\|$ seja mínima com o uso de um número pequeno de PCs. Neste caso a obtenção da matriz de *loadings* e de *scores* é feita por cálculo computacional por algoritmos distintos. Os algoritmos mais utilizados para cálculo de PCA é a Rotação de Jacobi, a Decomposição em Valores Singulares e o algoritmo Iterativo de Mínimos Quadrados Parciais Não-linear (NIPALS) [16, 17].

A análise por PCA é dita análise não-supervisada, ou seja, é baseada unicamente na matriz de espectros \mathbf{X} , e, se for considerado o espectro completo, a análise torna-se uma caracterização *fingerprint* [18]. De posse da matriz de espectros o algoritmo PCA separa a informação espectral mais relevante do ruído, em um número pequeno de novas variáveis. Esta abordagem é amplamente utilizada para análise gráfica, de visualização e reconhecimento de padrões de comportamento de amostras (gráfico de *scores*), além da contribuição de cada variável original no modelo PCA (gráfico de *loadings*).

2.3 REGRESSÃO POR MÍNIMOS QUADRADOS PARCIAIS

A regressão por PLS é uma técnica de obtenção de curva analítica que relaciona uma matriz de espectros \mathbf{X} com um vetor ou uma matriz de concentrações \mathbf{Y} . É dito um método supervisionado pois necessita de uma informação além dos dados espectrais, por exemplo, as concentrações de cada amostra. Diferentemente da regressão MLR, a PLS não necessita do conhecimento prévio de quais bandas espectrais devem ser consideradas

no modelo, desta forma, todo o espectro pode ser utilizado na regressão, caracterizando como abordagem *fingerprint*.

Em síntese a análise PLS parte da seguinte motivação: Decompor a matriz de dados \mathbf{X} num novo espaço, chamado de espaço de variáveis latentes (LV), contendo a matriz de *scores* \mathbf{T} e *loadings* \mathbf{P} , eq. 2.4, que expliquem a máxima variação dos espectros originais, de maneira análoga a redução de dimensão feita pela abordagem PCA. E o mesmo para a matriz de concentrações, decompondo \mathbf{Y} em *scores* \mathbf{U} e *loadings* \mathbf{Q} , eq. 2.5.

$$X = TP^T \quad (2.4)$$

$$Y = UQ^T \quad (2.5)$$

A regressão PLS busca então uma função que correlacione os dados espectrais \mathbf{X} com as concentrações de \mathbf{Y} , ou seja, uma relação matemática idêntica a regressão MLR ($Y = XB$), na qual a matriz dos coeficientes da regressão, matriz \mathbf{B} , é uma função das matrizes \mathbf{T} , \mathbf{U} , \mathbf{P} e \mathbf{Q} .

Este procedimento é feito por métodos computacionais e não serão descritos aqui pois não são necessários para a interpretação dos resultados obtidos pela regressão PLS. Estes detalhes podem ser obtidos pela análise de diferentes algoritmos de cálculo PLS, como o algoritmo Kernel, NIPALS para PLS, SIMPLS e outros [16, 17].

Após a construção de um modelo PLS com os dados de calibração, um novo conjunto de amostras pode ser aplicado ao modelo para se então prever as concentrações de interesse. Além da quantificação, a análise do gráfico de *loadings* do modelo PLS permite inferir quais regiões do espectro são as mais importantes para se quantificar um determinado analito, propondo assim marcadores vibracionais para uma certa propriedade. Da mesma forma que na análise PCA, os cálculos para obtenção do modelo de regressão PLS necessitam de softwares dedicados, os mais utilizados para este fim são Matlab, R e SIMCA

2.4 ANÁLISE DISCRIMINANTE

Na análise de padrões de comportamento, a variável dependente \mathbf{Y} de uma regressão PLS não é mais contínua, como exemplo o problema da classificação de amostras em grupos de amostras (como saudável/enfermo, falso/positivo, grupo A/grupo B), a análise é dita análise discriminante (DA) [19–21]. Neste caso, a matriz \mathbf{Y} é então codificada para valores como zero e um, menos um e um, menos um, zero e um (0/1, -1/1, -1, 0 e 1) ou ainda pode ser descrita por funções de probabilidades, como nas abordagens de support vector machine (SNV) e redes neurais artificiais (ANN).

A DA é extremamente útil quando se deseja estudar características de grupo de amostras, busca de biomarcadores de doenças, efeitos de fármacos no organismo, determinação de origem de amostras e outros.

Numa análise discriminante por projeção em estruturas latentes (PLS-DA), ou PLS-DA ortogonal (OPLS-DA) [21, 22], a matriz de classes previstas \mathbf{Y} pode assumir valores contínuos, que estão entre os valores codificados das variáveis categóricas originais. Por isso DA é mais utilizada para visualizar o comportamento global de amostras, no diagrama de *scores*, do que para prever se a amostra é de uma classe ou outra, uma vez que a predição pode acarretar na amostra não estar em nenhum dos dois grupos.

A diferença entre o diagrama de *scores* de uma análise PCA e de uma análise PLS-DA é que na análise PCA, não supervisionada, as amostras podem se agrupar por qualquer parâmetro não controlado que está contido na matriz de espectros \mathbf{X} . Já na análise PLS-DA além da matriz \mathbf{X} é fornecida uma matriz com a classe de cada amostra, matriz \mathbf{Y} , sendo então uma análise supervisionada, neste caso o algoritmo PLS-DA busca correlacionar os dados espectrais de \mathbf{X} com as classes de \mathbf{Y} , e a separação das amostras num diagrama de *scores* é maximizada pela informação da classe, um parâmetro controlado.

Se as amostras se separam em grupos num diagrama de *scores* de uma análise PLS-DA, pode-se então analisar o gráfico de *loadings* para então justificar, com as variáveis originais, o motivo da separação dos grupos devido a variável controlada.

Para esclarecer isso vamos a um exemplo: imagine um conjunto de amostras de plasma humano, de homens e mulheres, com IMC alto e baixo, diagnosticados com câncer e outros sem câncer. Um gráfico de *scores* de PCA de espectros Raman destas amostras pode conter agrupamentos por sexo, IMC, enfermidade ou formação de grupos combinados ou nenhuma formação de grupos uma vez que a variabilidade das amostras é muito grande. Por outro lado, se for ajustado, com os mesmos dados, um modelo PLS-DA com a informação de qual indivíduo tem câncer e qual não tem, o modelo multivariado vai buscar quais regiões espectrais estão correlacionadas com esta informação e então é mais provável que num diagrama de *scores* as amostras se separem pela enfermidade. Se isso ocorrer pode-se extrair do modelo quais regiões espectrais são as mais significativas para a discriminação de classes com câncer e sem câncer, propondo então marcadores vibracionais para a enfermidade.

2.5 PARÂMETROS QUE INDICAM A QUALIDADE DE MODELOS MULTIVARIADOS

Após o ajuste dos modelos de multivariados, alguns parâmetros podem ser analisados para indicar a qualidade ou não dos respectivos ajustes [15–17]. Os coeficientes descritos a seguir estão relacionados com desvios entre os valores verdadeiros, obtidos

por um método de referência, e os valores previstos pelas curvas analíticas. Em modelos multivariados é comum serem apresentados alguns destes para reforçar a qualidade das regressões. São eles:

Erro padrão da predição (SEP):

$$SEP = \sqrt{\frac{1}{z-1} \sum_{i=1}^z (y_i - \hat{y}_i - bias)^2} \quad (2.6)$$

com

$$bias = \frac{1}{z-1} \sum_{i=1}^z (y_i - \hat{y}_i) \quad (2.7)$$

onde y_i representa um valor experimental verdadeiro e \hat{y}_i é um valor previsto pelo modelo e z é o número de medições. O SEP , ou SEP_{teste} , chamado de erro da padrão da predição é, normalmente maior que o SEP_{CV} , que é calculado a partir de rodadas de validação cruzada. Desta forma o SEP_{teste} é dito mais otimista que o SEP . Se a eq. 2.6 for utilizada sobre o conjunto de dados da calibração então o resultado é chamado de erro padrão da calibração SEC , todos podem ser utilizados para estimar erros de novos casos.

O erro quadrático médio (MSE) é a média aritmética dos quadrados dos erros.

$$MSE = \frac{1}{z} \sum_{i=1}^z (y_i - \hat{y}_i)^2 \quad (2.8)$$

$MSEC$ (ou $MSEC_{CAL}$) refere-se a resultados estimados sobre o conjunto de calibração, $MSECV$ (ou $MSEC_{CV}$) para validação cruzada, e $MSEP$ ou MSE_{teste} para os resultados de predição e conjunto de teste. MSE menos o quadrado do $bias$ é o quadrado do SEP .

$$SEP^2 = MSE - bias^2 \quad (2.9)$$

A raiz do erro quadrático médio (RMSE) é obtido tomando a raiz quadrada do MSE. E de maneira análoga, pode ser calculado para o conjunto de calibração (RMSEC ou RMSECAL), validação cruzada (RMSECV ou RMSECV) e para o grupo de predição ou teste (RMSEP, ou $RMSE_{teste}$).

$$RMSE = \sqrt{\frac{1}{z} \sum_{i=1}^z (y_i - \hat{y}_i)^2} \quad (2.10)$$

MSE é mais utilizado durante o desenvolvimento e otimização dos modelos porque ele não tem as mesmas unidades de medida da propriedade analisada. Um outro parâmetro similar ao MSE é a medida da soma dos quadrados dos erro residual previsto (PRESS).

$$PRESS = \sum_{i=1}^z (y_i - \hat{y}_i)^2 = z.MSE \quad (2.11)$$

Além destes coeficientes, todos relacionados a desvios entre valores verdadeiros e valores previstos, existem também os coeficientes R^2X e R^2Y que representam a porção de

variância explicada pelo modelo multivariado para a matriz de espectros e de concentrações, respectivamente. Os valores de R^2X e R^2Y são maiores para a primeira PC, ou LV, pois representam a direção de maior variação do conjunto de dados, as PCs e LVs subsequentes tem contribuições menor para explicar a variância da matriz original individualmente, mas a soma das contribuições tanto de R^2X como de R^2Y para todas as componentes do modelo ajustado deve ser o mais próximo de um. o Mesmo para o coeficiente Q^2 que representa a capacidade de previsão dos modelos PLS.

Após a análise destes coeficientes comuns à análise multivariada, em estudos quantitativos é de praxe apresentar um gráfico que contenha no eixo vertical os valores obtidos pelo modelo de previsão contra os valores verdadeiros no eixo horizontal. Os valores previstos são satisfatórios quando existe uma correlação linear entre os pontos deste gráfico. Tal ajuste linear pode ser também avaliado pela análise do coeficiente de determinação R^2 .

$$R^2 = \frac{\sum_{i=1}^z (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^z (y_i - \bar{y})^2} \quad (2.12)$$

Quanto mais próximo de um for o valor de R^2 melhor terá sido o ajuste do modelo. Entretanto ainda assim deve-se avaliar se a curva analita não apresenta tendências nas previsões. Isso pode ser feito calculando-se o resíduo (*res*) de cada estimativa e após isso analisar o gráfico de resíduos. Não havendo tendências na distribuição dos resíduos, graficamente pode-se dizer que não há indícios de falta de ajuste no modelo de previsão.

$$res_i = y_i - \hat{y}_i \quad (2.13)$$

Existe também um teste de hipóteses que pode ser utilizado sobre o conjunto de dados de resíduos para inferir se o modelo não apresenta falta de ajuste, isso complementa a análise do gráfico de resíduos [23]. Para este fim, pode ser utilizado o teste de Durbin-Watson [24], no qual para p -valores acima de 0,05 conclui-se que, ao nível de 95% de confiança não há evidencias de falta de ajuste no modelo. Este teste deve ser realizado após a verificação do pré suposto da normalidade dos dados (teste de Shapiro-Wilk [25], por exemplo).

Já para a regressão MLR, preferencialmente no caso de poucos parâmetros, pode-se também mensurar os limites de detecção (LD) e quantificação (LQ) através da relação:

$$LD = 3 \times \frac{Sd_{ruído} \times C_a}{Altura_{sinal}} \quad (2.14)$$

$$LQ = 10 \times \frac{Sd_{ruído} \times C_a}{Altura_{sinal}} \quad (2.15)$$

onde, $Sd_{ruído}$ é o desvio padrão da linha base calculado próximo ao pico de interesse, C_a é a concentração do analito e $Altura_{sinal}$ é a altura do pico em questão.

3 TÉCNICAS ESPECTROSCÓPICAS

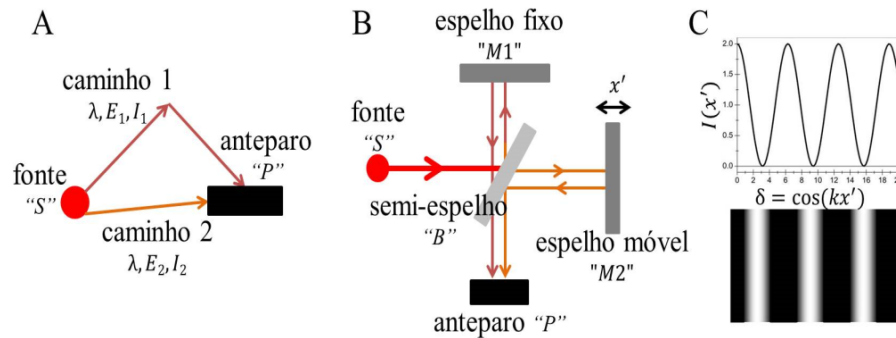
A espectroscopia no infravermelho e Raman são técnicas complementares. O princípio físico envolvido em cada uma é diferente, isso se deve ao fato de que as regras de seleção envolvidas em cada fenômeno se diferem. Embora ambas revelem informações acerca dos modos vibracionais moleculares, existem alguns modos vibracionais que aparecem em somente uma das técnicas, por isso a sua utilização em conjunto é recomendada para melhor caracterizar um estudo. MIR e NIR são baseadas na absorção de radiação, de modo que uma vibração é ativa no infravermelho apenas se o momento de dipolo molecular da amostra em questão muda com a radiação incidente, sendo assim particularmente sensível às funções polares. MIR apresenta diferentes bandas fundamentais dos modos de vibração molecular, enquanto NIR revela bandas de combinação e sobretons. De forma diferente, a espectroscopia de Raman baseia-se na dispersão de luz, e a observação de uma vibração Raman ativa requer uma alteração na polarizabilidade molecular, que é um efeito mais sensível à espécies homonucleares [26].

Neste trabalho todas as medidas espectroscópicas foram realizadas em equipamentos com FT. Estes equipamentos têm a característica de obtenção de medições rápidas e número elevado de acumulações por amostra reduzindo, assim, o ruído nas medições. O mecanismo de obtenção dos espectros nesses tipos de espectrofotômetros é diferente de equipamentos dispersivos, que utilizam uma rede de difração (com a função de um prisma), para separar as linhas espectrais, como ocorre nos tradicionais espectrofotômetros que operam na região do ultravioleta e visível (UV-Vis). Por isso será apresentada, na seção 3.1, a Teoria envolvida num espectrofotômetro do tipo FT e na sequência as interações envolvidas na espectroscopia vibracional.

3.1 ESPECTROSCOPIA COM TRANSFORMADA DE FOURIER

A espectroscopia com FT se baseia no fenômeno da interferência entre dois feixes de luz. A interferência consiste na interação de movimentos ondulatórios com mesmas frequências de oscilação e amplitude, mas que mantém, entre si, uma diferença de fase δ . Esta diferença de fase pode ser originada, por exemplo, por dois feixes de luz originados de uma fonte em comum, mas que percorrem caminhos diferentes. Quando estes feixes se encontram eles podem estar defasados e então se observa o fenômeno da interferência. Para visualizar esta situação basta analisar a Figura 2.

Figura 2 – Interferência entre dois feixes de luz que percorrem caminhos diferentes.



Em "A" têm-se uma representação didática de dois feixes de luz que emergem de uma fonte e depois se encontram, representando a formação da interferência. "B" é um esquema do Interferômetro de Michelson, neste caso a diferença de caminho entre os dois feixes se dá pelo movimento de um espelho móvel, e em "C" está representado o padrão de franjas claras e escuras formado num anteparo, este padrão pode ser representado por uma função cosseno que depende a diferença de caminho entre os feixes de luz que se interferem.

Fonte: O AUTOR.

A intensidade de luz I que chega no ponto P , é dada pela média temporal do vetor de Poynting \vec{S} , cuja magnitude fornece a densidade de fluxo de energia associada a uma onda eletromagnética. Esta grandeza pode ser medida, e tem dimensão de *Watt/metro*, no sistema internacional de unidades de medida, ou seja:

$$I = \langle \vec{S} \rangle \quad (3.1)$$

O vetor de Poynting é definido por [27]:

$$\vec{S} = \vec{E} \times \vec{H} \quad (3.2)$$

Nesta equação, \vec{E} representa o vetor campo elétrico e \vec{H} o vetor campo magnético. Como a magnitude de \vec{H} se relaciona com \vec{E} por

$$H = \epsilon c E \quad (3.3)$$

tem-se então:

$$\vec{S} = \epsilon c \vec{E} \times (\vec{E} \times \hat{k}) \quad (3.4)$$

onde ϵ é a permissividade elétrica do meio no qual a onda eletromagnética se propaga, c é a velocidade de propagação da onda, \hat{k} é o vetor de onda unitário, que fornece a direção de propagação da onda eletromagnética. Pelo princípio da superposição, o campo eletromagnético resultante \vec{E} que chega no anteparo P , da Figura 2, é dado por:

$$\vec{E} = \vec{E}_1 + \vec{E}_2 \quad (3.5)$$

onde

$$E_1 = A e^{(kx - \omega t)} \quad (3.6)$$

e

$$E_2 = B e^{(kx - \omega t + \delta)} \quad (3.7)$$

onde a letra grega δ presente na exponencial do campo \vec{E}_2 representa a diferença de fase entre as duas ondas \vec{E}_1 e \vec{E}_2 provocada pela diferença de caminho no percurso da fonte S até a chegada no anteparo P , da Figura 2.

Desta forma, um detector colocado no anteparo P faria a seguinte medida de intensidade de luz que ali chega:

$$I = \langle \vec{S} \rangle = \epsilon c \langle \vec{E} \cdot \vec{E} \rangle \quad (3.8)$$

como o vetor campo elétrico \vec{E} é resultante das contribuições provenientes de \vec{E}_1 e \vec{E}_2 , pode-se escrever:

$$I = \epsilon c (\langle E_1^2 \rangle + \langle E_2^2 \rangle + \langle E_1 \cdot E_2^* \rangle + \langle E_1^* \cdot E_2 \rangle) \quad (3.9)$$

ou

$$I = \epsilon c (I_1 + I_2 + 2\sqrt{I_1 I_2} \cos \delta) \quad (3.10)$$

Nesta equação os termos I_1 e I_2 são, respectivamente, a intensidade de luz proveniente do feixe de luz que viajou pelos caminhos 1 e 2 da Figura 2. O termo $2\sqrt{I_1 I_2} \cos \delta$ é chamado termo de interferência.

Esta equação mostra que a intensidade de luz que chega no detector depende do comportamento da função cosseno onde, para que se tenha uma irradiância máxima em P é necessário que a diferença de caminho δ entre os dois feixes seja igual a $0, \pm 2\pi, \pm 4\pi, \dots$ Com a mesma análise conclui-se que tem-se mínimos de intensidade de luz em P quando a diferença de caminho δ é igual a $\pm\pi, \pm 3\pi, \dots$ Uma visualização deste comportamento é apresentada na Figura 2-C.

A diferença de fase δ presente na equação que introduz a diferença de caminho percorrida pelos dois feixes pode ser escrita como:

$$|\delta| = kx' \quad (3.11)$$

Para o caso de irradiâncias iguais a equação (3.10) se reduz a:

$$I(x') = \epsilon c (I_1 + I_2 + 2\sqrt{I_1 I_1} \cos kx') = I_0 (1 + \cos kx') \quad (3.12)$$

A este comportamento da intensidade $I(x')$, que pode ser máxima ou mínima, dependendo da diferença de fase δ , Figura 2-C, entre os dois movimentos ondulatórios \vec{E}_1 e \vec{E}_2 , dá-se o nome de interferência construtiva e interferência destrutiva, respectivamente, e aos padrões de regiões claras e regiões escuras alternadamente, dá-se o nome de franjas ou figuras de interferência.

Interferômetros são configurações que usam este fenômeno da interferência para a determinação de algumas grandezas físicas.

Existem dois tipos de interferômetros: interferômetros de divisão de frente de onda, quando o feixe é dividido ao passar por diferentes aberturas, colocadas lado a lado (semelhante ao experimento de Young de dupla fenda) e interferômetros de divisão de amplitude, quando o feixe é dividido em superfícies parcialmente refletoras (como no interferômetro de Michelson, Figura 2-B).

O interferômetro de Michelson, esquematizado na Figura 2-B, ficou amplamente conhecido por sua contribuição para a formulação da teoria da relatividade especial de Einstein quando Michelson e Morley provaram a inexistência do éter. Suas aplicações, como a medida de pequenos deslocamentos e do índice de refração da luz, ficaram limitadas por muitos anos até o advento da computação, que tornou possível os cálculos matemáticos necessários nas integrações da espectroscopia de Fourier. Num interferômetro de Michelson, um feixe luminoso proveniente da fonte S incide em um semi-espelho divisor de feixe B (*beamsplitter*) e é dividido em dois feixes. Um viaja até o espelho fixo $M1$ e o outro viaja até o espelho móvel $M2$. Os feixes refletem nestes espelhos $M1$ e $M2$ e retornam até o divisor de feixe B , estes dois feixes então se recombinam e viajam até um ponto de observação em P .

Analisando a expressão (3.12), experimentalmente, pode-se introduzir um detector na saída do interferômetro. Este detector serviria para registrar a intensidade de luz em função do deslocamento x' do espelho $M2$ do interferômetro, I_0 independe do movimento deste espelho e o termo $\cos kx'$ é o termo de interferência.

Entretanto, para aplicações espectroscópicas, no lugar de uma fonte de luz monocromática deve-se utilizar uma fonte de luz não-monocromática, que tenha emissão em uma determinada região de interesse do espectro eletromagnético, para, então, obter um padrão de interferência proveniente de diferentes comprimentos de onda λ , ou seja, uma análise espectral.

Neste caso, com a utilização de uma fonte de luz não monocromática, a expressão da irradiância (eq. 3.12) que chega ao detector precisa ser modificada, sendo necessária a adição de um termo que pondere a intensidade de luz emitida pela fonte em cada comprimento de onda λ_n do espectro no lugar da contribuição devido a um único comprimento de onda λ , este termo é chamado de distribuição espectral $G(k)$. Fazendo tal modificação obtêm-se:

$$I(x') = \int_0^{\infty} G(k)dk[1 + \cos kx'] \quad (3.13)$$

A função cosseno, presente no integrando, pode ser escrita em termos de exponenciais, desta forma:

$$I(x') = \int_0^{\infty} G(k)dk + \frac{1}{2} \left(\int_0^{\infty} G(k)e^{ikx'} dk + \int_0^{\infty} G(k)e^{-ikx'} dk \right) \quad (3.14)$$

observando que a primeira integral independe da diferença de caminho x' e, fazendo uma troca de variáveis k por $-k$, apenas na última integral, pode-se escrever:

$$I(x') = I_0 + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sqrt{\frac{\pi}{2}} G(k) e^{ikx'} dk \quad (3.15)$$

que é o mesmo que

$$W(x') = I(x') - I_0 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sqrt{\frac{\pi}{2}} G(k) e^{ikx'} dk \quad (3.16)$$

a função $W(x')$ recebe o nome de interferograma. Seu comportamento descreve a intensidade de luz que chega ao detector em função da diferença de caminho x' provocada pelo movimento de um dos braços ($M2$) do interferômetro.

É interessante a descrição matemática da função $W(x')$, pois a forma de $W(x')$ na expressão (3.16) é chamada de Transformada de Fourier da função $G(k)$ e, pelo teorema da inversão de Fourier obtêm-se a relação inversa:

$$G(k) = \frac{1}{\pi} \int_{-\infty}^{\infty} W(x') e^{-ikx'} dx' \quad (3.17)$$

A função $G(k)$, que recebe o nome de distribuição espectral, fornece a intensidade de luz que chega ao detector em cada número de onda k . Desta maneira a função $G(k)$ descreve o espectro eletromagnético que é irradiado pela fonte S .

Este passo matemático que muda o espaço das funções de $W(x')$ para $G(k)$, ou seja, um interferograma para um espectro, dá origem ao nome da técnica "espectroscopia com transformada de Fourier". A espectroscopia com FT consiste em gravar $W(x')$ como uma função de x' e transformar matematicamente esta relação para obter $G(k)$ em função de k .

Na prática, $W(x')$ é substituída por uma função $f(t)$, pois como o espelho do interferômetro se move com velocidade constante, $v = x'/t$, a aquisição de dados para o interferograma é feita em determinados intervalos de tempo t , por isso a expressão para $W(x')$ é conhecida como domínio de tempo. De forma análoga, a função $G(k)$ é substituída por uma $g(\omega)$, pois como $\omega = ck$, é usual registrar a dependência em frequência ω , e a expressão fica conhecida como domínio das frequências. Como $f(t)$ e $g(\omega)$ são pares de FT, uma função opera em um espaço e a outra no espaço recíproco. Desta forma tem-se então:

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(\omega) e^{-i\omega t} d\omega \quad (3.18)$$

e

$$g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{i\omega t} dt \quad (3.19)$$

No entanto, as equações (3.18) e (3.19) não podem ser empregadas como estão escritas porque não é possível mover o espelho a um comprimento infinito e, porque a

radiação não assume valores negativos. Neste ponto então é necessário um tratamento matemático chamado de apodização [28, 29], a apodização tem por objetivo avaliar o interferograma que será convertido em um espectro por meio do cálculo da FT apenas em uma região apreciável, ou seja, cortar o interferograma em uma certa região com a finalidade de reduzir picos laterais que não contribuem para a análise espectral. Esta truncagem pode ser feita por um corte quadrado, como na função de apodização *Boxcar*, mas também pode ser feita a apodização *Triangular* ou ainda utilizar funções mais complexas que suavizam mais ainda este corte, como a função *NB Strong*, *Happ Genzel* e outras [30].

Uma boa aproximação para resolver o problema dos limites de integração, é assumir que em um espectrofotômetro real a irradiância que emerge do interferômetro se comporta como um pulso finito de ondas senoidais. Consideremos então um trem de ondas se repetindo n vezes tantas quantas forem os pontos onde ocorrem interferências construtivas devido ao movimento do espelho do interferômetro. Estas variações podem ser escritas em termos da função:

$$f(t) = \begin{cases} e^{-i\omega_0 t} & \text{se } -\frac{\tau_0}{2} < t < \frac{\tau_0}{2} \\ 0 & \text{senão} \end{cases} \quad (3.20)$$

Tal aproximação para a função $f(t)$, na eq.(3.20), de assumir um determinado comportamento em uma região e ser nula em todo o restante, é análogo ao procedimento utilizado na apodização *Boxcar*. Substituindo a eq. (3.20) em (3.19), tem-se:

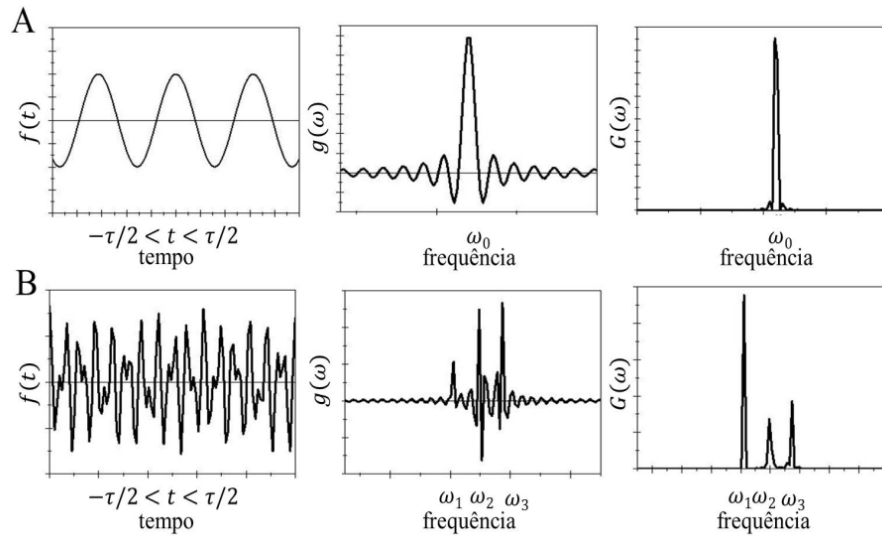
$$g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\frac{\tau_0}{2}}^{\frac{\tau_0}{2}} e^{i(\omega - \omega_0)t} dt \quad (3.21)$$

desprezando a parte imaginária e o fato da paridade ímpar da função seno, obtêm-se:

$$g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\frac{\tau_0}{2}}^{\frac{\tau_0}{2}} \cos [(\omega - \omega_0)t] dt = \sqrt{\frac{2}{\pi}} \frac{\sin [(\omega - \omega_0)\frac{\tau_0}{2}]}{\omega - \omega_0} \quad (3.22)$$

Graficamente, Figura 3, tem-se: para uma fonte monocromática, a parte real de $f(t)$, eq.(3.20) que representa o "trem de ondas finito", $g(\omega)$ eq.(3.21) que ilustra a transformada de Fourier de $f(t)$ e finalmente, $G(\omega) = |g(\omega)|^2$ que representa a densidade de energia ou distribuição espectral.

Figura 3 – Representação das funções $f(t)$, $g(\omega)$ e $G(\omega)$.



Em "A" esta representado o padrão formado por uma fonte de luz monocromática e em "B" por uma fonte policromática. Na sequência é ilustrado um trem de ondas finito (sequência de franjas), a Transformada de Fourier deste trem de ondas, e ao final, a distribuição de energia espectral.

Fonte: O AUTOR.

Para o caso de uma fonte de luz não monocromática (caso real, em espectroscopia por FT) pode ser feita a mesma aproximação, de um "trem de ondas finito", onde deve-se corrigir que a contribuição em $f(t)$ não se deve apenas a uma única frequência ω_0 como na eq.(3.20) mas sim a uma distribuição contínua de frequências ω_n que caracterizam a emissão de uma lâmpada em uma determinada região do espectro eletromagnético. Supondo uma lâmpada hipotética com emissões bem definidas apenas em torno das frequências $\omega_0, \omega_1, \omega_2$, tem-se os esboços da forma de $f(t)$, $g(\omega)$ e $G(\omega)$, mostrados na Figura 3-B.

As medições no MIR e NIR foram medidas de transmitância, eq. 3.23. Neste modo de medida, a intensidade de luz que emerge da fonte do interferômetro, I_0 , é medida e registrada sem a presença da amostra. Após isso a amostra é então colocada no caminho ótico logo após o interferômetro e então é registrado a porção de luz que atravessa a amostra, denotada por I_T .

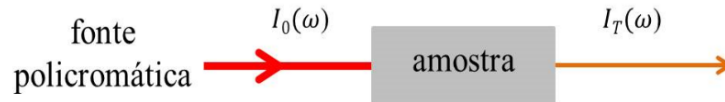
$$T = \frac{I_T}{I_0} \quad (3.23)$$

Esta razão, a transmitância, esta relacionada com a porção de energia que foi absorvida pela amostra. A análise espectroscópica consiste em avaliar quais regiões do espectro foram absorvidas ou não, e suas intensidades.

À medida que a luz atravessa um meio que contém uma espécie absorvedora, um decréscimo de sua intensidade ocorre na proporção que a espécie é excitada. Para uma solução de determinada concentração c , quanto mais longo for o caminho óptico b mais unidades absorvedoras estarão no caminho, e maior será a atenuação. Da mesma forma,

para um caminho óptico b fixo, quanto maior for a concentração c de absorventes, maior será a atenuação [31]. Abaixo está representado uma medida de transmitância.

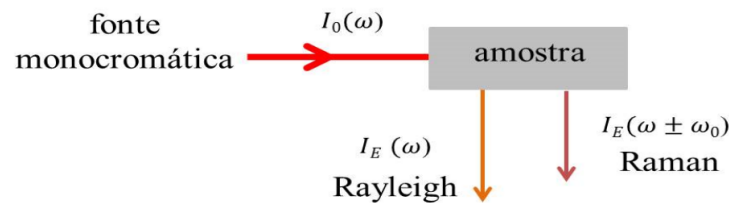
Figura 4 – Medida da transmitância - MIR e NIR.



Fonte: O AUTOR.

Já nas medições por espectroscopia Raman, a análise gráfica dos espectros é a mesma que nos espectros MIR e NIR. O diferencial é que não é mais uma medida de transmitância, mas sim uma medida de luz espalhada. Na espectroscopia Raman a porção de luz que emerge do interferômetro não atravessa a amostra. O processo físico envolvido na espectroscopia Raman será discutido mais adiante, mas o mecanismo utilizado no processo de medição consiste em incidir uma fonte de luz monocromática na amostra, luz laser, ocorre uma interação e é, então, colhido num interferômetro a porção de luz que é espalhada pela amostra num certo ângulo. Na figura abaixo representado um esquema de medida Raman.

Figura 5 – Medida do espalhamento Raman.



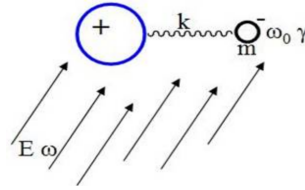
Fonte: O AUTOR.

Na espectroscopia Raman a intensidade de luz medida é proporcional a quarta potência da diferença entre as frequência da luz laser incidente e da luz espalhada, $I \propto (\omega \pm \omega_0)^4$. Por isso frequentemente a medida por espectroscopia Raman também é chamada deslocamento Raman (*Raman shift*). Mais detalhes sobre o espalhamento Raman serão mostrados nas seções que seguem.

3.2 ABSORÇÃO INFRAVERMELHA

O objetivo desta seção é apresentar a formulação teórica do fenômeno físico envolvido na interação entre uma onda eletromagnética e o meio material, centrando a discussão no que diz respeito à absorção molecular. Um modelo clássico para a abordagem desta seção é apresentado pela figura abaixo:

Figura 6 – Modelo para absorção da radiação eletromagnética.



Elétron de massa m oscilando em torno do núcleo, com frequência natural de oscilação ω_0 quando um campo elétrico E de frequência ω é aplicado.

Fonte: O AUTOR.

A equação de movimento para o elétron com carga $-e$, massa m preso ao núcleo atômico por uma força elástica que o faz oscilar com frequência ω_0 em torno da posição de equilíbrio, sujeito a um amortecimento forçado, pela atuação de um campo elétrico \vec{E} , descrito por [32]:

$$m \frac{d^2 \vec{x}}{dt^2} + \gamma m \frac{d\vec{x}}{dt} + m\omega_0^2 \vec{x} = -e\vec{E} \quad (3.24)$$

Usando a aproximação de que o elétron não tem inércia, ou seja, ele irá oscilar com a mesma frequência ω do campo \vec{E} aplicado, tem-se:

$$x = x_0 e^{-i\omega t} \quad (3.25)$$

onde \vec{x} representa o vetor posição do elétron e γ , presente na força de arraste, é a constante de amortecimento proporcional a velocidade de oscilação do elétron. Substituindo a solução de (3.25) em (3.24), chega-se à:

$$\vec{x} = -\frac{e}{m} (\omega_0^2 - \omega^2 - i\gamma\omega)^{-1} \vec{E} \quad (3.26)$$

Como a contribuição para o momento de dipolo, por elétron é:

$$\vec{P} = -e\vec{x} \quad (3.27)$$

usando a eq. (3.26) pode-se escrever o momento de dipolo \vec{P} como:

$$\vec{P} = \frac{e^2}{m} (\omega_0^2 - \omega^2 - i\gamma\omega)^{-1} \vec{E} \quad (3.28)$$

Generalizando este resultado para N moléculas por unidade de volume e cada átomo com Z elétrons, tem-se a eq. (3.28) modificada para

$$\vec{P} = \sum \vec{P}_j = \frac{Ne^2}{m} \sum f_j (\omega_j^2 - \omega^2 - i\gamma_j\omega)^{-1} \vec{E} \quad (3.29)$$

onde tem-se f_j elétrons por molécula e a soma deste é Z elétrons por molécula ($Z = \sum f_j$).

Nesta equação, (3.29), pode-se ver claramente que o comportamento da polarização da molécula muda abruptamente quando a frequência ω do campo \vec{E} aplicado se aproxima

da frequência ω_j natural de oscilação do elétron. Quando isso ocorre, o termo entre parênteses, no denominador da eq. (3.29), tem um mínimo. Logo, o valor de \vec{P} é um máximo. Para que ocorra uma absorção no infravermelho, MIR ou NIR, é necessário uma mudança no momento de dipolo molecular, isto é, $\left(\frac{d\vec{P}}{dq}\right) \neq 0$, onde q é a coordenada generalizada.

Fazendo o uso das equações de Maxwell no meio material [27]:

$$\nabla \times \vec{E} = -\mu \frac{\partial \vec{H}}{\partial t} \quad (3.30)$$

$$\nabla \times \vec{H} = \varepsilon \frac{\partial \vec{E}}{\partial t} + \frac{\partial \vec{P}}{\partial t} + \vec{J} \quad (3.31)$$

$$\nabla \cdot \vec{E} = -\frac{1}{\varepsilon} \nabla \cdot \vec{P} \quad (3.32)$$

$$\nabla \cdot \vec{H} = 0 \quad (3.33)$$

tomando o rotacional de (3.30) e com o uso de (3.31) e (3.29), chega-se a:

$$\nabla^2 \vec{E} = \mu \varepsilon \left(1 + \frac{Ne^2}{\varepsilon m} \cdot \frac{1}{\omega_j^2 - \omega^2 - i\gamma_j \omega} \right) \frac{\partial^2 \vec{E}}{\partial t^2} \quad (3.34)$$

Com a eq. (3.34) e a forma do vetor campo elétrico, $E = E_0 e^{i(kx - \omega t)}$, chega-se facilmente a:

$$\kappa^2 = \frac{\omega^2}{c^2} \left(1 + \frac{Ne^2}{\varepsilon m} \cdot \frac{1}{\omega_j^2 - \omega^2 - i\gamma_j \omega} \right) \quad (3.35)$$

ou seja, conclui-se que o número de onda κ é uma quantidade complexa. Desta forma, tem-se:

$$\kappa = k_{\text{real}} + ik_{\text{imag}} = k + i\alpha \quad (3.36)$$

Este resultado implica na atenuação do campo elétrico dado por:

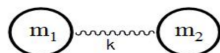
$$E = E_0 e^{-\alpha x} e^{i(kx - \omega t)} \quad (3.37)$$

Ou seja, a natureza complexa do número de onda κ , eq. (3.24), obtido pelo sistema baseado na Figura (6) implica em uma atenuação do campo elétrico dado pela eq. (3.37). Esta atenuação ocorre quando existe uma ressonância entre a frequência ω do campo aplicado no material e a frequência natural de oscilação ω_0 do meio material. Esta atenuação, ou perda na energia do campo \vec{E} que atravessa um determinado material, é a quantidade a ser medida em algumas técnicas espectroscópicas, como a que será discutida adiante, e que caracteriza inequivocadamente a amostra em estudo, pois a absorção de energia é uma característica única de cada tipo de grupamento químico, como uma assinatura da estrutura molecular do objeto em estudo.

A determinação algébrica das frequências naturais de oscilação de uma molécula não é uma tarefa fácil, a determinação mais simples que pode ser feita é baseada em um

sistema de massa-mola, como o apresentado na Figura 7 abaixo, este sistema é chamado de oscilador harmônico para molécula diatômica:

Figura 7 – Modelo de molécula diatômica.



Fonte: O AUTOR.

A frequência de oscilação para este sistema é dada por [26]:

$$f = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}}, \quad (3.38)$$

onde μ é a massa reduzida do sistema composto por m_1 e m_2 , dois átomos, e k é a constante de ligação entre esses átomos.

Para moléculas maiores, as energias de absorção, ou o espectro característico, são obtidos por meio de cálculos computacionais complexos ou a partir de uma medida espectroscópica feita com a substância pura. Cada molécula contém N átomos e um total de $3N$ graus de liberdade, que correspondem às coordenadas cartesianas de cada átomo que compõe a molécula. Desta forma, moléculas com muitos átomos, tem um número elevado de modos vibracionais, vide Tabela 1.

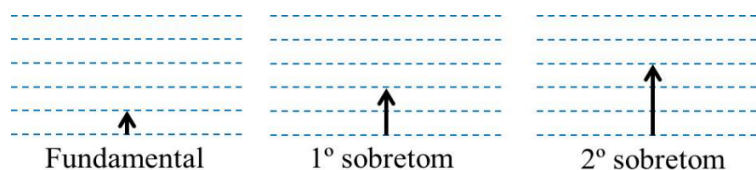
Tabela 1 – Graus de liberdade de moléculas poliatômicas.

Tipos de graus de liberdade	Linear	não-Linear
Translacional	3	3
Rotacional	2	3
Vibracional	$3N - 5$	$3N - 6$
Total	$3N$	$3N$

Fonte: O AUTOR.

Tais modos de vibracionais se referem aos modos fundamentais de vibração molecular. Estes modos fundamentais de vibração são característicos na região do infravermelho médio do espectro eletromagnético, e podem ser observados tanto em experimentos de espectroscopia MIR como Raman. Já região do infravermelho próximo é a região de bandas de combinação e de sobretons dos modos fundamentais, como representado no diagrama de energia abaixo.

Figura 8 – Diagrama de níveis de energia de bandas no infravermelho.



Fonte: O AUTOR.

As bandas características de algumas funções químicas são facilmente encontradas na literatura, [13,26] entretanto as atribuições em espectros experimentais de amostras reais não ocorrem exatamente nos números de onda tabelados. Isso se deve principalmente ao radical no qual está ligada a função química e a interações com a vizinhança, chamados efeitos de matriz, por isso torna-se difícil a atribuição de modos vibracionais de moléculas complexas.

Agora que se sabe um pouco sobre a interação física que ocorre num processo de absorção molecular passa-se para a próxima seção afim de discutir sobre a física do espalhamento Raman. Nesta seção será mostrada também uma comparação entre a absorção infravermelha e o espalhamento Raman (Figura 9).

3.3 ESPALHAMENTO RAMAN

A espectroscopia Raman baseia-se na dispersão da luz, na qual, uma molécula tem vibração Raman ativa se existe uma mudança na polarizabilidade molecular, que é um efeito mais sensível para espécies homonucleares. Os espectros Raman mostram diferentes bandas fundamentais dos modos de vibração molecular e podem ser observados alguns modos que não aparecem em medidas de absorção infravermelha. Isso se deve a diferença nas regras de seleção de cada efeito. Absorção infravermelha requer uma mudança no momento de dipolo molecular e o espalhamento Raman uma mudança na polarizabilidade [26,33]. A Figura 9 faz um comparativo entre as duas técnicas.

Figura 9 – Regras de seleção para Infravermelho e Raman.

Vibração molecular					
$\frac{d\alpha}{dq}$	$\neq 0$	$\neq 0$	$\neq 0$	$= 0$	$= 0$
Raman ativo	sim	sim	sim	não	não
$\frac{d\vec{P}}{dq}$	$= 0$	$\neq 0$	$= 0$	$\neq 0$	$\neq 0$
Infravermelho ativo	não	sim	não	sim	sim

Fonte: Adaptado de [33,34].

Um modelo clássico para explicar o fenômeno envolvido no espalhamento Raman, que é um espalhamento inelástico da radiação incidente, origina-se quando uma fonte de luz monocromática de frequência ω , e campo elétrico dado por $E = E_0 \cos(\omega t)$ interage com a matéria. Levando em conta o momento de dipolo induzido $\mathbf{P} = \alpha \mathbf{E}$, a polarizabilidade α pode ser expandida em série de Taylor em torno da coordenada generalizada q da seguinte forma [26,33]:

$$\alpha = \alpha_0 + \left(\frac{d\alpha}{dq} \right)_0 q + \dots \quad (3.39)$$

Sendo ω_V e ω , respectivamente, um nível de frequência vibracional e a frequência da radiação laser incidente e a coordenada q escrita na forma $q = q_0 \cos(\omega_V t)$, a expressão para o momento de dipolo induzido fica:

$$P = (\alpha_0)(E_0 \cos(\omega t)) + \left(\frac{d\alpha}{dq} \right)_0 (q_0 \cos(\omega_V t))(E_0 \cos(\omega t)) \quad (3.40)$$

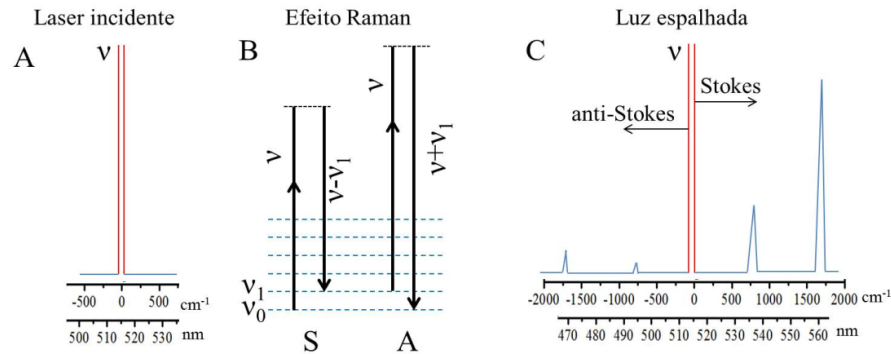
utilizando a identidade trigonométrica $\cos(a)\cos(b) = \frac{1}{2}(\cos(a+b) + \cos(a-b))$, pela eq. 3.40, temos

$$P = \alpha_0 E_0 \cos(\omega t) + \frac{1}{2} \left(\frac{d\alpha}{dq} \right)_0 q_0 E_0 [\cos(\omega + \omega_V)t + \cos(\omega - \omega_V)t] \quad (3.41)$$

O primeiro termo da eq.(3.41) aparece somente a frequência ω da radiação incidente, este é o termo de espalhamento elástico (Rayleigh). O termo entre colchetes representa o espalhamento inelástico Raman, neste termo tem-se a contribuição $(\omega + \omega_V)$ chamado de espalhamento Raman Stokes e o termo $(\omega - \omega_V)$, chamado de espalhamento Raman anti-Stokes.

A Figura 10 ilustra o efeito Raman. O efeito Raman ocorre quando a molécula é excitada pelo fóton do laser incidente e, como o comprimento de onda do feixe de excitação é, em geral, afastado de uma banda de absorção, a molécula pode atingir inúmeros valores de energia que estão abaixo do nível energético correspondente ao primeiro estado excitado eletrônico. Este é um processo não quantizado e estes níveis intermediários de energia são chamados de estados virtuais.

Figura 10 – Efeito Raman.



"A" representa o laser que incide na amostra, "B" a interação da radiação com a matéria e os níveis de energia para Raman Stokes e Raman anti-Stokes, e "C" uma representação de um espectro Raman. A escala inferior é absoluta (em nanômetros) e a superior é relativa ao comprimento de onda do laser incidente (deslocamento Raman).

Fonte: O AUTOR.

O efeito Raman é observado quando a molécula retorna deste estado virtual para um nível de menor energia, espalhando um fóton que pode ser detectado. Raman Stokes ocorre quando a molécula sai do nível fundamental de energia, sobe para um estado virtual e decai até um nível vibracional. Ao retornar a este nível vibracional, a molécula emite um fóton com energia dada pela diferença entre a energia do fóton incidente e a do nível vibracional. A porção devido ao espalhamento anti-Stokes ocorre quando a molécula já se encontra em um nível vibracional excitado, então sobe a um nível virtual e na sequência volta ao nível fundamental de energia, espalhando um fóton com energia dada pela energia do fóton incidente acrescida da energia do nível vibracional que se encontrava a molécula inicialmente.

Classicamente as contribuições para o espalhamento Stokes e anti-Stokes deveriam ter a mesma intensidade, mas observa-se que a Stokes é mais intensa do que a anti-Stokes, isso se deve ao fato que é mais provável a molécula estar no seu nível fundamental de energia [35], vide distribuição de Maxwell-Boltzmann ou eq. (3.42). Entretanto pode-se aumentar a intensidade da contribuição anti-Stokes com um aumento da temperatura, uma vez que, mais moléculas, nestas condições, vão estar no primeiro estado excitado vibracional.

$$\frac{I_{(anti-Stokes)}}{I_{(Stokes)}} = \left(\frac{\omega + \omega_V}{\omega - \omega_V} \right)^4 e^{-hc\omega_V/K_B T} \quad (3.42)$$

Como analisado anteriormente, eq.(3.40), o efeito Raman é observado se existe mudança na polarizabilidade, ou seja, se $\left(\frac{d\alpha}{dq}\right) \neq 0$. Ao final a intensidade Raman observada é proporcional a intensidade da radiação incidente e a quarta potência do termo $(\omega \pm \omega_V)$. Estas intensidades Raman são da ordem de 0,001% da intensidade da radiação incidente [35], todavia a radiação espalhada encontra-se na região do visível ou NIR do espectro eletromagnético, que possui detectores muito sensíveis, de modo que é possível ter espectros Raman de igual qualidade a dos espectros de infravermelho.

4 TÉCNICAS DE SEPARAÇÃO

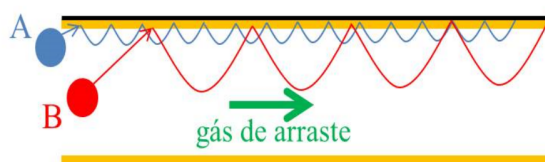
Nesta seção serão apresentadas brevemente as técnicas de separação que foram utilizadas como métodos de referência para as determinações espectroscópicas, sendo a cromatografia a gás utilizada nos dois trabalhos apresentados na seção de aplicações e a cromatografia a líquido e a eletroforese capilar utilizados em trabalhos apresentados nos apêndices.

Para estas três técnicas analíticas são feitas considerações gerais sobre o mecanismo de separação. Não é propósito deste trabalho discutir completamente estas técnicas, uma vez que são utilizadas apenas para a obtenção de medições de referência e não como objetos de pesquisa. E ainda, foram todas acompanhadas, ou realizadas, por especialistas em cada uma destas técnicas, por meio de colaboração de pesquisa. Uma vez que não é pretensão do autor tornar-se um especialista, mas sim um usuário de métodos bem estabelecidos.

4.1 CROMATOGRAFIA A GÁS

A GC é uma técnica de separação na qual uma mistura volatilizada de compostos (fase gasosa) elui por um tubo capilar pela passagem forçada de um gás de arraste, sem interação química com este gás (geralmente N_2 , H_2 , He ou Ar). Estes diferentes compostos presentes na mistura interagem com um filme com polaridade variável (fase estacionária) que reveste a parede interna do tubo capilar, de modo que moléculas que possuem maior afinidade com o filme demoram mais a percorrer o comprimento total do capilar, dessa forma dá-se a migração dos compostos em diferentes tempos. Por isso o eixo horizontal de uma análise por GC é chamado de tempo de retenção (TR). No final deste tubo capilar podem existir diferentes tipos de detectores, como ionização por chama (FID), detector de massas (MS) e outros [13]. Este esquema de separação é representado na Figura 11.

Figura 11 – Esquema de separação em cromatografia a gás.



"A" representa um analito com maior afinidade com a fase estacionária e "B" um analito com menor afinidade. "B" percorre a coluna em um tempo menor.

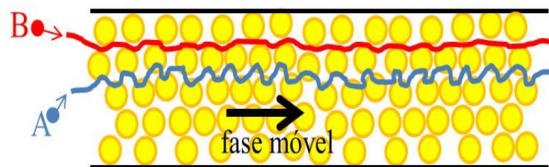
Fonte: O AUTOR.

4.2 CROMATOGRAFIA A LÍQUIDO

A LC baseia-se na separação de uma mistura de compostos em meio líquido quando estes atravessam uma coluna preenchida por uma fase sólida finamente dividida (como

uma esponja). Nesta técnica a separação ocorre devido a duas interações: do analito presente na mistura com o material do interior da coluna (fase estacionária) ou com a interação com a solução que gera o fluxo de análises (fase móvel). Um analito que tem maior interação com a fase estacionária fica mais tempo retido na coluna enquanto que um analito que tem maior interação com a fase móvel pode, por exemplo, se solubilizar mais e atravessar a coluna num tempo menor. A detecção é feita no final da coluna onde é registrado o tempo que o analito passou pelo detector (tempo de retenção - TR). A detecção pode ser por arranjo de diodos (DAD), MS e outros [13]. Este esquema de separação é representado na Figura 12.

Figura 12 – Esquema de separação em cromatografia a líquido.



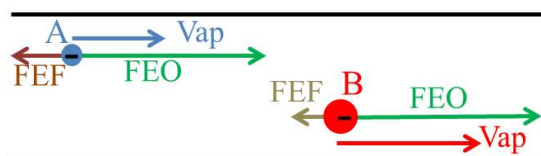
"A" representa um analito com maior afinidade com a fase estacionária e "B" com a fase móvel. "B" percorre a coluna em um tempo menor.

Fonte: O AUTOR.

4.3 ELETROFORESE CAPILAR

Na CE a separação ocorre em meio líquido, pela migração diferenciada de compostos carregados que compõe uma mistura, dentro de um tubo capilar submetido a aplicação de um campo elétrico. Nestas condições ocorre a formação do fluxo da solução de eletrólito (tampão) e a separação dos analitos se dá em função da sua relação carga/tamanho. A detecção ocorre próximo à extremidade oposta da injeção do tubo capilar, através de uma janela óptica de detecção no caso de detector DAD, ou por detecção de condutividade elétrica sem contato (C4D), por MS, e outros [13]. Este esquema de separação é representado na Figura 13.

Figura 13 – Esquema de separação em eletroforese capilar.



FEO é o fluxo eletrosmótico, independe da natureza do analito, FEF é o fluxo eletroforético, intrínseco de cada analito e velocidade aparente V_{ap} é dada por $\vec{V}_{ap} = \vec{FEO} + \vec{FEF}$. "B" percorre o capilar em um tempo menor.

Fonte: O AUTOR.

5 APLICAÇÕES

Nesta seção serão discutidos cinco estudos que utilizam métodos multivariados como ferramentas para extrair informações de espectros NIR, MIR e Raman e de dados de técnicas de separação.

5.1 QUANTIFICAÇÃO DA ADULTERAÇÃO DE AZEITE DE OLIVA COM ÓLEO DE SOJA POR ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO, MÉDIO E RAMAN

O azeite de oliva, produzido a partir da prensagem da oliva (*Olea Europaea*), é uma fonte de alimento nutritiva e saborosa, saudável e amplamente utilizada em todo o mundo. O consumo regular do azeite está associado com a redução de fatores de risco e ocorrência de doenças degenerativas e crônicas como diabetes, obesidade e enfermidades cardiovasculares como hipertensão arterial, síndrome metabólica e aterosclerose [36,37].

Além do consumo fresco, o azeite é utilizado como um ingrediente em formulações industriais de temperos, molhos, recheios, coberturas, massas e vários produtos de panificação. Seu uso tem se tornado comum no desenvolvimento de filmes comestíveis [38] e embalagens biodegradáveis, sendo usado como uma barreira física de proteção antimicrobiana para frutas, legumes e produtos minimamente processados [39–41]. A indústria farmacêutica tem constantemente avaliado a sua utilização para a produção de cremes hidratantes para a pele [42, 43], proteção contra a radiação UV [44], tratamento de estrias na gravidez [45] e de produtos para auxiliar na manutenção e conservação das membranas celulares [36].

Este conjunto de propriedades funcionais e características tecnológicas, associados com um preço relativamente elevado de venda, torna o azeite um alvo potencial para a ação de agentes fraudadores. A fraude comercial de azeite de oliva, bem como de outros alimentos, tornou-se cada vez mais uma prática comum, na qual os interesses do agente fraudador estão principalmente na oportunidade de fazer lucros rápidos, mesmo se for necessária a utilização de produtos químicos e materiais inadequados para consumo humano ou contaminados com microorganismos patogênicos, pondo em risco a saúde do consumidor. Neste sentido torna-se necessária a utilização de métodos analíticos de identificação e quantificação de suspeita de fraude no produto, além da verificação dos padrões de qualidade exigidos pela legislação.

5.1.1 Objetivos

O objetivo deste estudo foi a avaliação da resposta de técnicas de espectroscopia vibracional, NIR, MIR e Raman, associadas à análises multivariada de dados por PLS

para prever e quantificar a fraude por adição de óleo de soja em amostras de azeite de oliva.

5.1.2 Materiais e métodos

Oito amostras de óleos vegetais, todos de diferentes marcas e regiões produtoras, foram adquiridas no mercado local (Juiz de Fora, Brasil), sendo quatro amostras de óleo de soja (OS) e quatro de azeite de oliva extra-virgem (AOEV). A partir deste conjunto de amostras, duas soluções estoque foram preparadas, uma de AOEV, preparada a partir de uma mistura em volumes iguais das quatro amostras diferentes e a outra, uma solução estoque de OS preparada de forma análoga.

Estas dez amostras (quatro amostras de AOEV, quatro amostras de OS, e as duas soluções estoque) foram analisadas individualmente, em duplicata, pela técnica de GC. Tal análise foi necessária para identificar o perfil de ácidos graxos (AG) de cada amostra e verificar se as soluções estoque representam bem o perfil dos AG encontrados em cada marca diferente de óleo comercial.

As análises pela técnica de GC foram realizadas em um cromatógrafo fabricado pela Shimadzu modelo GC-2010 Plus (Shimadzu, Kyoto, Japão) equipado com injetor tipo *split-splitless*, auto injetor AOC 20-I e detecção por FID. Este equipamento está instalado na Central Analítica, localizado no departamento de química da Universidade Federal de Juiz de Fora. As análises por esta técnica foram realizadas em colaboração com a doutoranda Brenda Lee Simas Porto, discente do programa de pós-graduação em química da Universidade Federal de Juiz de Fora.

As condições de análise pela técnica de GC foram: coluna capilar com fase estacionária de líquido iônico SLB-IL111 (Supelco, Bellefonte, PA, EUA) com 14,0 m de comprimento, 0,10 mm de diâmetro interno e 0,08 μm de espessura do filme, FID e injetor a 250 °C, temperatura da coluna contante em 168 °C, pressão de 143 kPa, injeção em modo *split* em uma proporção de 1 : 500, e volume de injeção de 1 μL [46].

Para as análises feitas por GC-FID, cerca de 15,0 mg de cada amostra foram pesados em tubo de vidro (em duplicata) e as amostras foram esterificadas por catálise ácida [47]. A composição qualitativa de ácidos graxos (AG) presente em cada amostra foi determinada medindo o TR dos seus picos e comparando com os TR dos respectivos padrões de ésteres metílicos de ácidos graxos (FAME). Os padrões de FAME considerados foram o ácido palmítico (PA), ácido palmitoleico (POA), ácido esteárico (EA), ácido oleico (OA), ácido linoleico (LA), e ácido linolênico (LNA) (Sigma Aldrich, St. Louis, MO, EUA). A composição quantitativa foi realizada por normalização de área e expressa como uma percentagem em massa [48].

Após a caracterização pela técnica de GC-FID das amostras adquiridas, misturas

foram preparadas a partir das duas soluções estoque a fim de simular uma possível fraude no produto. Tais misturas foram então analisadas pelas técnicas espectroscópicas de NIR, MIR e Raman. Para obter um modelo de previsão do percentual de OS adicionado ao AOEV, dois conjuntos de amostras foram preparados, um utilizado para-se obter o modelo de calibração PLS, e o outro para a validação do modelo. Estas amostras foram preparadas em balões volumétricos de 10,0 ml, fazendo-se misturas de OS e AOEV, num range de 0,0 – 30,0 e 70,0 – 100,0% de adulteração, com incrementos de 1,5% nestas faixas. Outras seis amostras, com percentuais de adulteração de 40, 45, 50, 55, 60, e 65% de adição de OS foram preparadas a fim de cobrir toda a faixa intermediária das percentagens de mistura. Desta forma foram preparadas um total de 48 que foram utilizadas para o modelo de calibração PLS. Outras 12 amostras foram preparadas e utilizadas para validar o modelo. Estas amostras de validação foram preparadas com incrementos que compreendem toda a faixa experimentada, mas com percentuais diferentes daqueles utilizados para o modelo de calibração.

Para as medições por técnicas espectroscópicas não foi necessário nenhum procedimento de preparo de amostra. Espectros NIR foram obtidos com um espectrômetro com transformada de Fourier modelo FT-NIR MPA fabricado pela Bruker, equipado com detector de Te-InGaAs e ótica de quartzo, configurado para o modo de transmitância. Este equipamento está instalado no Laboratório de Espectroscopia de Materiais, localizado no departamento de física da Universidade Federal de Juiz de Fora. As condições para as análises por NIR foram: região espectral de $4000 - 12000 \text{ cm}^{-1}$, resolução de 4 cm^{-1} , 32 acumulações por amostra no modo transmitância. Cada amostra foi colocada em um *vial* de vidro com 8,0 mm de diâmetro interno e este colocado diretamente no compartimento do equipamento. Como fundo foi utilizado o porta amostras vazio.

Para os espectros MIR foi utilizado um espectrômetro com transformada de Fourier modelo FT-MIR Vertex70 fabricado pela Bruker, equipado com detector de RT-DLaTGS e ótica de KBr, configurado para o modo de transmitância. Este equipamento está instalado no Laboratório de Espectroscopia de Materiais, localizado no departamento de física da Universidade Federal de Juiz de Fora. As condições de aquisição no MIR foram: região espectral de $350 - 4000 \text{ cm}^{-1}$, resolução de 4 cm^{-1} , 32 acumulações por amostra, utilizando o porta amostras vazio como fundo. Os espectros foram obtidos colocando uma gota de cada amostra sobre a superfície de uma janela de KBr. Após cada análise a janela de KBr foi limpa com diclorometano (Vetec, Xerém, RJ, Brazil) para então ser reutilizada.

Espectros Raman foram obtidos com um espectrômetro com transformada de Fourier modelo FT-Raman RFS-100 fabricado pela Bruker, equipado com detector de germânio resfriado com nitrogênio líquido e radiação de excitação no comprimento de onda de 1064 nm a partir de um laser Nd:YAG, configurado para o modo retroespalhamento. Este equipamento está instalado no Núcleo de Espectroscopia e Estrutura Molecular,

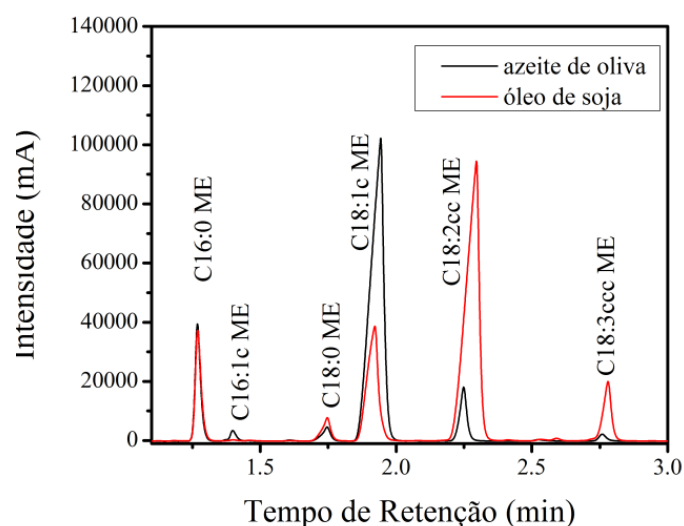
localizado no departamento de química da Universidade Federal de Juiz de Fora. Os espectros Raman foram obtidos na região espectral de $50 - 3500 \text{ cm}^{-1}$, com resolução de 4 cm^{-1} , 50 acumulações por amostra e energia do laser de 200 mW .

5.1.3 Resultados e discussões

Inicialmente foi realizada uma análise exploratória do perfil de AG de todas as amostras comerciais adquiridas e das respectivas soluções estoque por CG. O objetivo era de avaliar as similaridades e diferenças entre todo o conjunto de amostras, bem como o grau de representatividade das soluções estoque que foram utilizadas na etapa de adulteração intencional de AOEV por adição de OS.

A Figura 14 mostra o perfil de AG da solução estoque de AOEV e de OS. A análise das amostras dos óleos pela técnica de GC mostrou que os FAMES de PA (C16:0ME), POA (C16:1cME), EA (C18:0ME), OA (C18:1cME), LA (C18:2ccME), e LNA (C18:3cccME) foram encontrados em todas as amostras de AOEV. Os mesmos FAMES foram encontrados nas amostras de OS, exceto o POA, que não foi observado em qualquer amostra de OS. Estes FAMES foram identificados por comparação dos TR das amostras e seus respectivos padrões.

Figura 14 – Perfil de ácidos graxos obtidos pela técnica de GC-FID.



Fonte: O AUTOR.

Cerca de 98% da composição de óleos vegetais é devida a AG, e destes, aproximadamente 95%, para as amostras analisadas, referem-se a seis diferentes AG. Os ésteres destes seis ácidos graxos detectados foram quantificados, por normalização de área em todas as dez amostras. A Tabela 2 apresenta os resultados para a quantificação destes

FAMEs. Pode-se perceber que existem semelhanças entre as quatro amostras de OS e as quatro amostras de AOEV que compõem as soluções estoque.

Tabela 2 – Quantificação de ácidos graxos por GC.

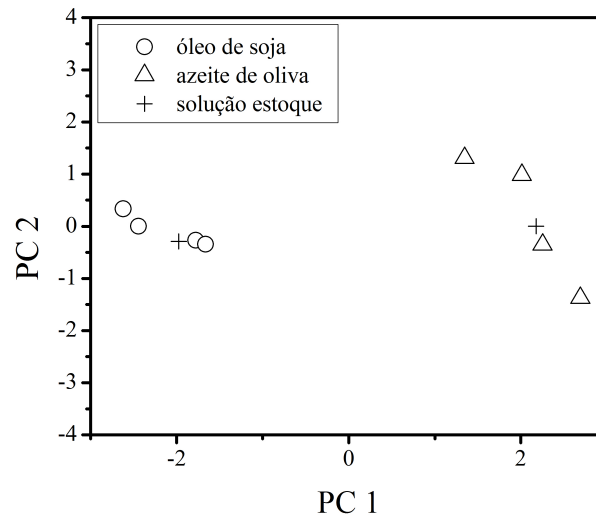
Amostras	FA massa/massa de óleo (g/100g)					
	C16:0	C16:1c	C18:0	C18:1c	C18:2cc	C18:3ccc
óleo de soja 1	10,6	-	3,6	20,9	53,3	6,6
óleo de soja 2	10,8	-	2,7	20,9	54,9	5,7
óleo de soja 3	10,0	-	3,0	19,5	55,1	7,4
óleo de soja 4	10,7	-	2,5	22,2	53,3	6,2
estoque óleo de soja	10,6	-	2,7	20,9	54,7	6,7
azeite de oliva 1	15,9	2,1	1,9	52,0	21,7	1,4
azeite de oliva 2	12,0	1,2	2,2	71,7	7,1	0,9
azeite de oliva 3	14,9	1,6	2,3	59,7	15,4	1,1
azeite de oliva 4	11,3	0,8	2,4	69,1	10,0	1,4
estoque azeite de oliva	14,3	1,6	2,4	62,9	12,7	1,1
RSD	0,4	0,3	0,1	1,1	0,8	0,3

Fonte: O AUTOR.

As duas soluções estoque analisadas por GC mostraram resultados semelhantes para a concentração de AG quando comparadas com as respectivas amostras de óleos vegetais originais. Os valores de desvio padrão relativo (RSD), de uma medição em duplicata, estão todos abaixo de 1,1%, evidenciando a precisão e confiabilidade das medições.

Pode-se perceber na Tabela 2 que os FAMEs de LA (C18:2cc) e LNA (C18:3ccc) tem intensidades relativas mais elevadas no OS, enquanto o OA está presente em maiores quantidades no AOEV, e ainda que o POA (C16:1c) está presente apenas no AOEV. Entretanto, analisar uma tabela de concentrações de seis analitos em dez diferentes amostras nem sempre é uma tarefa de fácil interpretação. Neste sentido foi realizada uma análise de PCA baseada nas quantificações obtidas pela técnica de GC. Este modelo de PCA foi elaborado a partir de uma matriz de dados com dez observações e seis variáveis, sendo ajustada com três PCs e coeficiente $R^2X = 0,996$. A Figura 15 é o diagrama de *scores* deste modelo PCA que pode ser utilizado para avaliar a semelhança química entre diferentes marcas de OS, AOEV e as soluções estoque.

Figura 15 – PCA das amostras de óleos vegetais a partir dos dados de GC-FID.



Fonte: O AUTOR.

O diagrama de *scores* (Figura 15) mostra uma clara separação em dois grupos de amostras. O lado esquerdo do diagrama corresponde ao grupo de amostras de OS, enquanto que os pontos presentes no lado direito do diagrama são todos referentes às amostras de AOEV. A PCA fornece uma boa indicação visual das posições relativas das amostras adquiridas e das soluções estoque. A inspeção visual revela que cada solução estoque é de fato bastante representativa do seu grupo, uma vez que está posicionada aproximadamente no centro de cada grupo, podendo assim ser usada nas diluições subsequentes.

Além disso, o teste T^2 de Hotelling [49, 50] foi usado para comparar os óleos individuais adquiridos. O teste é um equivalente multivariado do teste- t de Student, e foi utilizado para avaliar se as amostras vêm do mesmo grupo ou não. Como resultado, as comparações entre os vetores de composição de AG não foram estatisticamente diferentes para qualquer par, o que indica a homogeneidade das amostras dentro de cada grupo. Os resultados para este teste estatístico está mostrado na Tabela 3. Todas as comparações foram feitas depois de verificar a hipótese de normalidade dos dados pelo teste de Shapiro-Wilk.

Tabela 3 – p -valores do teste T^2 de Hotelling para a comparação entre os vetores de composição química quantificados por GC-FID.

Amostras	Interação entre as amostras									
	1:2	1:3	1:4	1:SE	2:3	2:4	2:SE	3:4	3:SE	4:SE
óleo de soja	0,765	0,972	0,842	0,879	0,819	0,929	0,901	0,885	0,918	0,969
azeite de oliva	0,661	0,940	0,724	0,999	0,698	0,460	0,661	0,658	0,940	0,727

Fonte: O AUTOR.

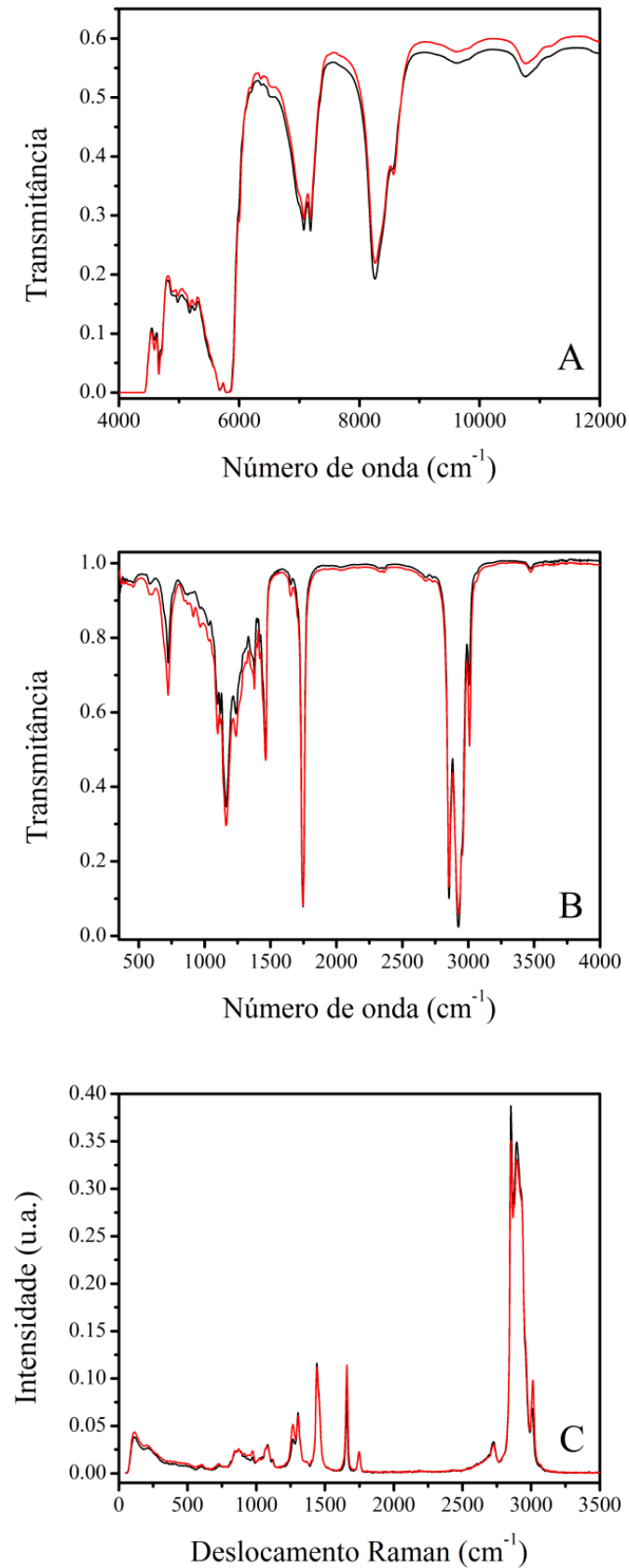
Por outro lado, para a comparação OS versus AOEV, o teste T^2 de Hotelling foi significativo ($p < 0,001$), isto é, o conjunto formado pelas amostras de OS difere significativamente do grupo constituído por amostras de AOEV.

Neste ponto do trabalho as considerações que podem ser feitas a partir dos resultados obtidos pela técnica de GC são: (1) a composição de AG, em perfil e concentração, é diferente para amostras de OS e AOEV; (2) apesar das amostras serem de diferentes fabricantes e regiões, elas possuem perfil muito próximo entre si. Logo, pode-se inferir que a utilização de um conjunto maior de amostras não influenciaria o resultado, bem como a utilização da solução estoque preparada em laboratório pode ser considerada representativa para o universo amostral, tanto para OS quanto para AOEV. Desta forma justifica-se a utilização das soluções estoque como boa escolha para a elaboração das adições de OS em AOEV, simulando uma possível adulteração no produto.

Ao contrário da análise por GC, que está limitada a considerações apenas sobre o perfil de AG, as técnicas espectroscópicas oferecem a possibilidade de análise sem a etapa preparo da amostra, ou seja, o perfil espectral leva em conta não só a contribuição de AG, mas também de todos os outros componentes presentes na amostra, uma vez que a análise é feita com a amostra bruta. O tema da análise de adulteração de AOEV pela adição fraudulenta de OS via técnicas espectroscópicas foi realizado concomitantemente pelas técnicas de espectroscopia NIR, MIR e Raman a partir das duas soluções estoque caracterizadas anteriormente.

A Figura 16 mostra o perfil espectral obtido para as amostras de solução estoque de OS e AOEV, utilizando as técnicas NIR (Fig. 16-A), MIR (Fig. 16-B), e Raman (Fig. 16-C), respectivamente. Pode se notar que os espectros vibracionais são também muito semelhantes. Este fato era esperado, devido à semelhança da composição química das duas amostras. Neste caso, os espectros obtidos refletem características gerais dos óleos vegetais, sem um marcador molecular para distinguir claramente as duas classes de óleos.

Figura 16 – Espectros das soluções estoque de OS e AOEV.



Em "A" estão apresentados os espectros NIR, em "B" espectros MIR e em "C" espectros Raman. Linha preta para azeite de oliva e linha vermelha para óleo de soja.

Fonte: O AUTOR.

A atribuição dos grupos funcionais observados para as absorções IR foram 723 cm^{-1} (*cis*-CH=CH tesoura fora do plano), 1163 cm^{-1} (-C-O estiramento; -CH₂ tesoura), 1463 cm^{-1} (-CH₂ tesoura), 1747 cm^{-1} (-C=O estiramento), 2854 e 2925 cm^{-1} (estiramento simétrico e assimétrico em -CH₂) e 3008 cm^{-1} (*cis*=C-H estiramento) [51, 52]. Para os espectros NIR, bandas em torno de 5677 e 5806 cm^{-1} são devidos às bandas de combinação e primeiro sobretom de C-H de metileno de grupos alifáticos de óleos, e seu segundo sobretom é observado em 8258 cm^{-1} . As bandas em torno de 4661 cm^{-1} podem ser atribuídas à bandas de combinação de estiramentos C=C e C-H da insaturação *cis* de AG e em 7077 e 7189 cm^{-1} são atribuídos a bandas de combinação de C-H de metileno [53, 54], já para o espalhamento Raman foram observadas bandas em 1266 cm^{-1} (*cis*-CH=CH rotação), 1303 cm^{-1} (=CH₂ torção), 1440 cm^{-1} (=CH₂ tesoura), 1658 cm^{-1} (C=C estiramento), 1747 cm^{-1} (éster C=O estiramento), 2854 e 2896 cm^{-1} (CH₂ e CH₃ estiramento simétrico), e 3012 cm^{-1} (=CH estiramento assimétrico) [55, 56].

Entretanto o objetivo não é prolongar uma discussão em torno da atribuição de modos vibracionais encontrados em cada espectro. Um exemplo para mostrar por que não fazê-lo neste estudo é o caso do POA (C16:1cME). Como foi demonstrado na análise por GC (Tabela 2), este AG foi detectado apenas nas amostras de AOEV. Com base nesta informação, este seria um bom candidato como marcador vibracional para quantificar a adulteração. No entanto, o sinal presente no espectro Raman em 1658 cm^{-1} , característico do estiramento C=C, presente no POA, aparece tanto no espectro do OS quanto no de AOEV. Isto pode ser explicado pelo aumento de LA (C18:2ccME) nas amostras de OS, quando comparado com o AOEV, o que acaba equilibrando este pico Raman, de modo que apenas com uma inspeção visual não é possível encontrar um marcador vibracional para quantificar esta adulteração.

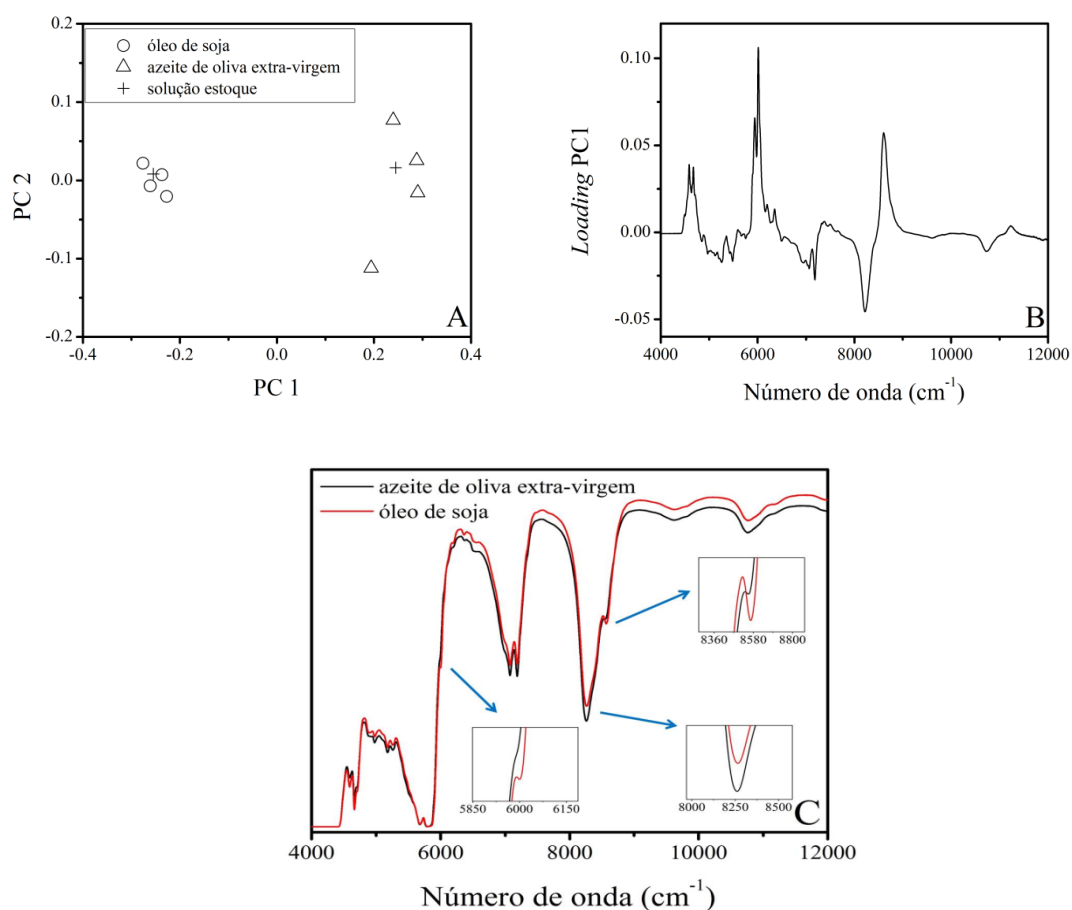
Devido a sobreposição dos espectros vibracionais de OS e AOEV, não foi possível encontrar um marcador que possa ser utilizado para o monitoramento da adulteração de AOEV por adição de OS. Neste caso métodos de análise multivariada podem ser utilizados como ferramentas para auxiliar a encontrar quais regiões do espectro estão correlacionadas com a adição de OS e também para elaborar modelos matemáticos de previsão do percentual de OS adicionado no AOEV.

Os resultados obtidos pelas técnicas de análise multivariada aplicadas às medidas espectroscópicas serão apresentados na sequência NIR, MIR e Raman. Para ambas as técnicas os espectros foram primeiramente centrados na média, as análises de PCA foram feitas com um conjunto de quatro amostras de OS, quatro de AOEV e suas respectivas solução estoque. As regressões PLS foram feitas com as amostras do conjunto de calibração, num total de 48 amostras compreendendo níveis de adulteração de zero a 100%, além de outras 12 diferentes amostras utilizadas para a validação de cada modelo.

Abaixo é mostrado a PCA para os dados de NIR. Este modelo PCA foi ajustado

com duas PCs e coeficiente $R^2X = 0,987$. A análise da Figura 17 permite concluir que a técnica de NIR juntamente com a análise multivariada de dados por PCA é apropriada para diagnosticar diferenças espectrais nos dois conjuntos de amostras, OS e AOEV. Apesar dos espectros serem visualmente muito semelhantes o diagrama de *scores* separa completamente as amostras de AOEV das amostras de OS. Como a separação dos dois grupos de amostras ocorre ao longo do eixo horizontal (direção da primeira PC), o gráfico de *loadings* para esta PC apresenta quais regiões do espectro são as mais importantes para explicar tal separação. Ou seja, a análise de PCA sugere marcadores vibracionais responsáveis por discriminar os dois grupos de amostras.

Figura 17 – PCA dos espectros NIR de óleos vegetais.



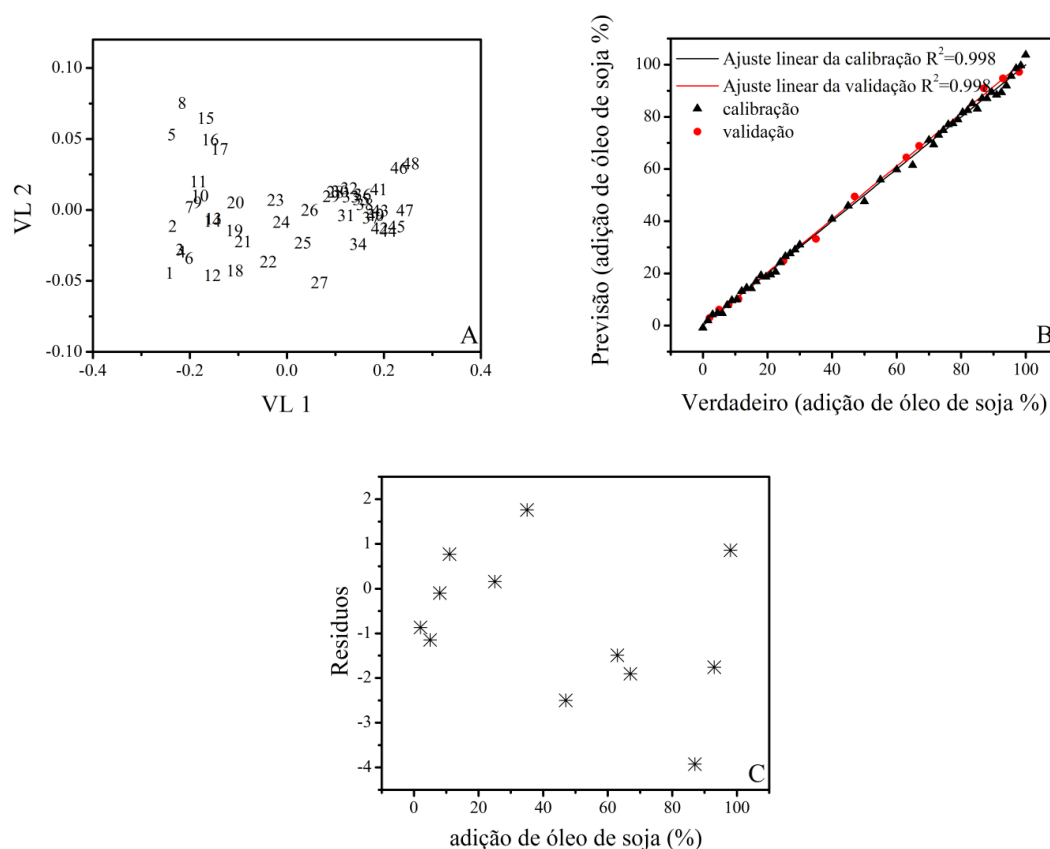
Em "A" têm-se o diagrama *descores* do modelo PCA, "B" o gráfico de *loadings* para a PC-1 e em "C" as regiões de maior variação no espectro.

Fonte: O AUTOR.

A PCA é uma análise não supervisionada, ou seja, é baseada somente nas informações espectrais. Como o modelo multivariado encontrou uma distinção natural entre as amostras, então é esperado que frações de misturas entre os dois grupos de amostras possam ser previstos, pois estarão distribuídos em regiões no espaço de PCA entre os dois grupos previamente classificados. Na sequência foi elaborado um modelo de regressão PLS

ajustado com duas LV, e coeficientes de correlação da matriz de espectros $R^2X = 0,913$, da matriz de percentual de adição de óleo de soja $R^2Y = 0,998$ e poder de predição $Q^2 = 0,998$. Os resultados para este modelo estão apresentados na Figura 18, abaixo.

Figura 18 – Regressão PLS a partir dos espectros NIR.



Em "A" têm-se o diagrama de *scores* do modelo PLS, "B" o gráfico de correlação entre os valores previstos *versus* valores verdadeiros e em "C" o gráfico de resíduos para o conjunto de validação.

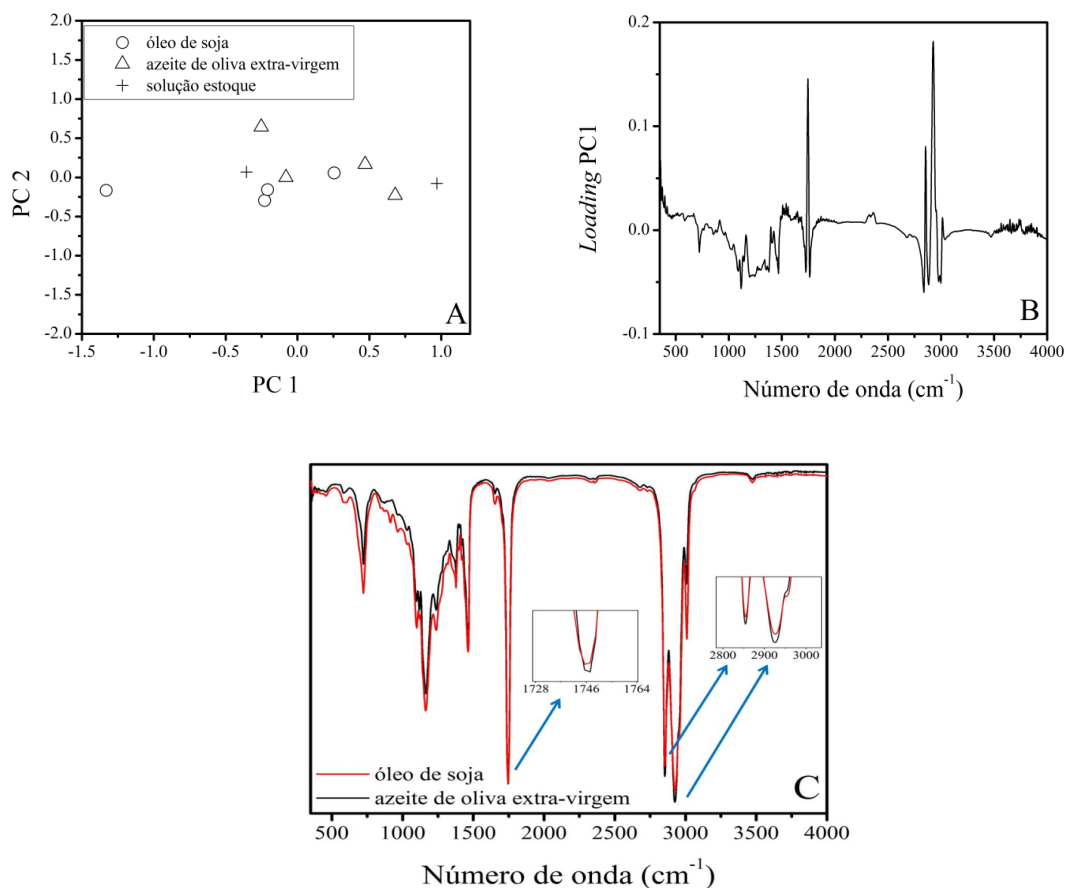
Fonte: O AUTOR.

Cada amostra do conjunto de calibração é apresentada no gráfico de *scores* do modelo PLS, Figura 18-A, o aumento do rótulo da amostra (de 1 a 48) indica um aumento no percentual de adição de OS. Os extremos 1 e 48 indicam, respectivamente, 0% e 100% OS. Como esta numeração cresce ao longo da horizontal é possível esperar que o modelo PLS forneça um bom ajuste. Este ajuste é mostrado na Figura 18-B que obteve coeficiente de determinação $R^2 = 0,998$ para os conjunto de calibração e $R^2 = 0,998$ para o de validação. Para verificar se não existe falta de ajuste no modelo, foram calculados os resíduos para o conjunto de validação, cujo gráfico é mostrado na Figura 18-C. Como os resíduos estão distribuídos aleatoriamente, pode-se dizer que não há evidências de falta de ajuste no modelo previsão PLS-NIR. Outros coeficientes que indicam a qualidade global do ajuste feito foram $RMSEP = 1,76$ e $RMSE_{cv} = 1,71$.

A mesma sequência de estudo foi empregada com os dados de espectroscopia MIR.

A Figura 19 apresenta os resultados da análise PCA com os dados de infravermelho médio. O modelo PCA foi ajustado com duas PCs e $R^2X = 0,902$.

Figura 19 – PCA dos espectros MIR de óleos vegetais.



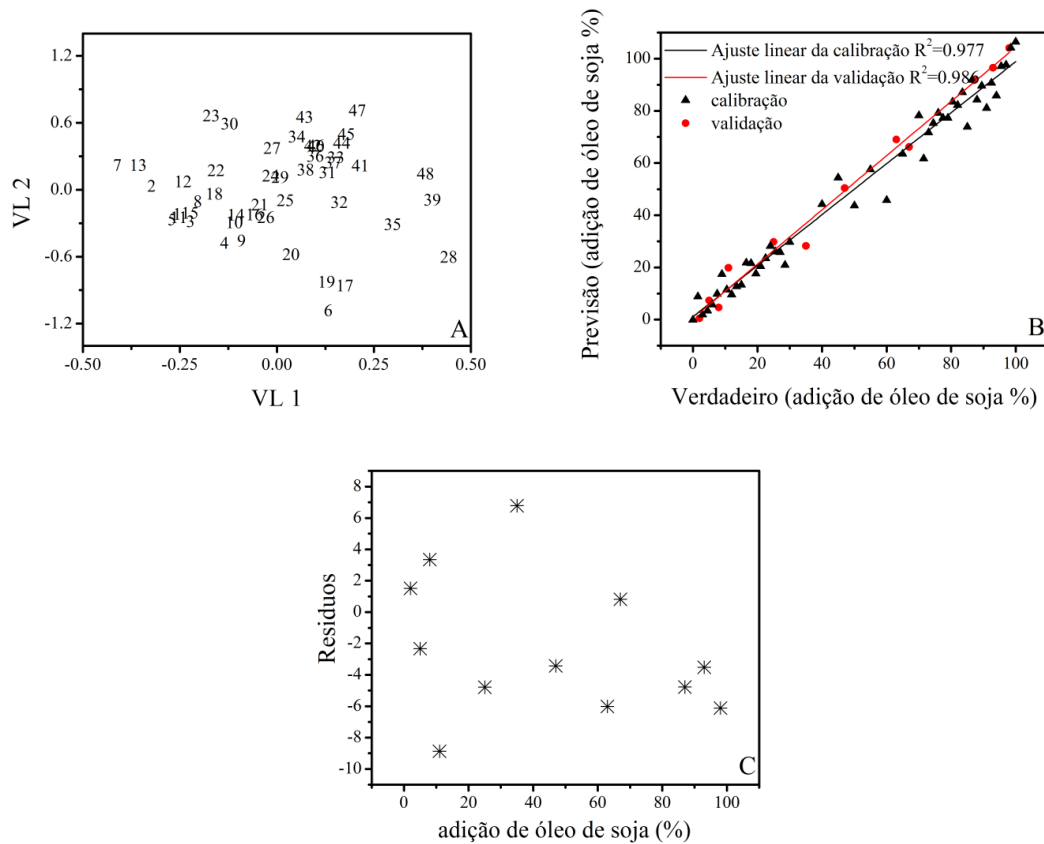
Em "A" têm-se o diagrama de *scores* do modelo PCA, "B" o gráfico de *loadings* para a PC-1 e em "C" as regiões de maior variação do espectro.

Fonte: O AUTOR.

Neste caso, diferentemente do resultado da PCA-NIR, a PCA com os espectros MIR não separa claramente as amostras em dois grupos. O gráfico de *scores* da Figura 19-A apresenta as amostras mais dispersas, mas ainda assim, com uma tendência de alocar as amostras de OS no lado esquerdo do diagrama e as de AOEV no lado direito. Como esta separação não é evidenciada, o respectivo gráfico de *loadings*, 19-B, não deve ser utilizado como candidatos a marcadores vibracionais para discriminar OS de AOEV, esta visualização apenas indica quais regiões apresentaram maior variação nos espectros de transmitância.

A regressão PLS com os dados MIR foi ajustada com três PCs e coeficientes $R^2X = 0,956$, $R^2Y = 0,978$ e $Q^2 = 0,964$. A análise PLS é apresentada na Figura 20.

Figura 20 – Regressão PLS a partir dos espectros MIR.



Em "A" têm-se o diagrama de *scores* do modelo PLS, "B" o gráfico de valores previstos *versus* valores verdadeiros e em "C" o gráfico de resíduos para o conjunto de validação.

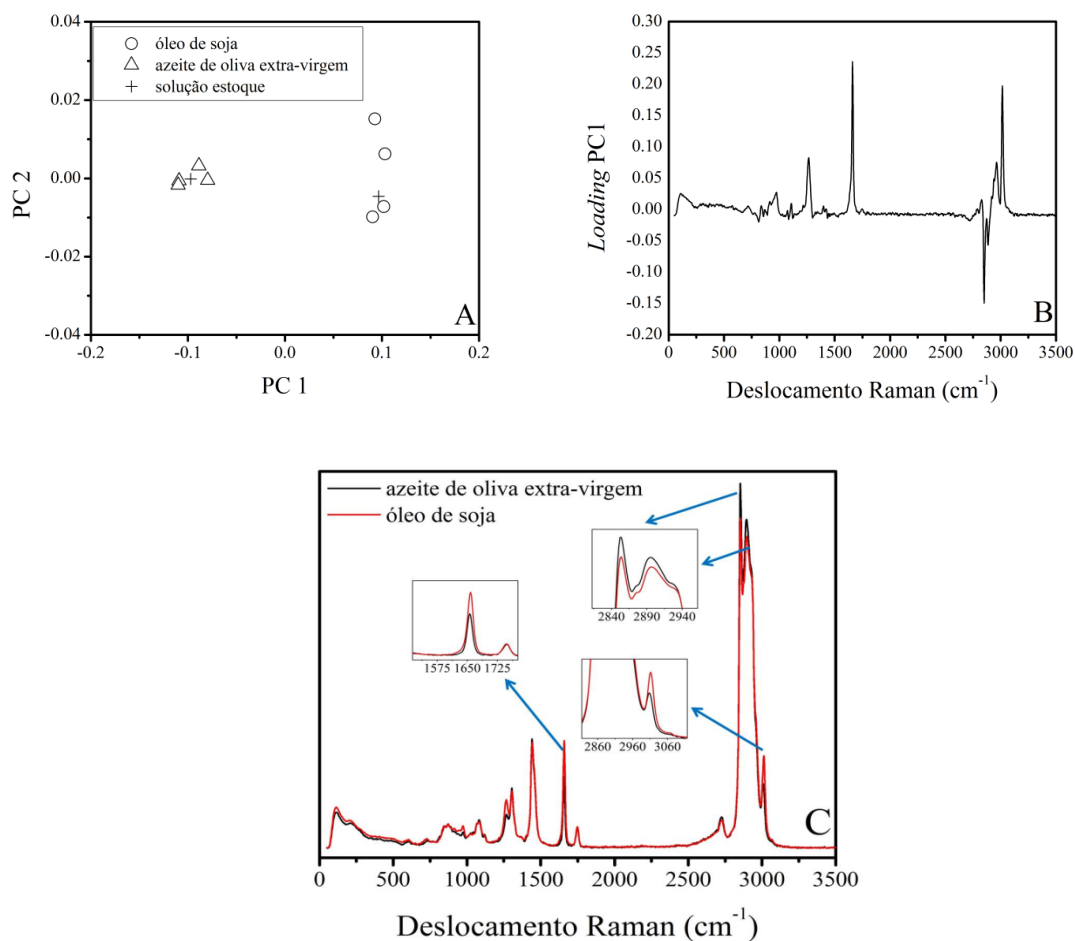
Fonte: O AUTOR.

Como foi indicado na PCA, as observações obtidas no ajuste PLS não seguem uma tendência linear de aumento no número dos rótulos das amostras, como mostra o gráfico de *scores* 20-A. Isto repercute em uma maior dispersão no ajuste apresentado na Figura 20-B e em uma maior amplitude nos resíduos, mostrado no eixo-y da Figura 20-C.

Apesar da maior dispersão nos resultados obtidos para o ajuste PLS-MIR, foram obtidos coeficientes de determinação $R^2 = 0,977$ para o conjunto de calibração e $R^2 = 0,986$ para o conjunto de validação. O gráfico de resíduos para o conjunto de validação também não apresentou tendência, indicando não haver falta de ajuste no modelo. Os coeficientes de erro global corroboram com as observações pois, apresentaram valores de $RMSEP = 4,89$ e $RMSE_{cv} = 14,71$.

Novamente a mesma sequência de estudo foi utilizada, desta vez aplicada às análises por espectroscopia Raman. Assim a análise de PCA a partir dos espectros Raman é mostrada na Figura 21. Esta PCA foi ajustada com duas PCs e $R^2X = 0,984$.

Figura 21 – PCA dos espectros Raman de óleos vegetais.



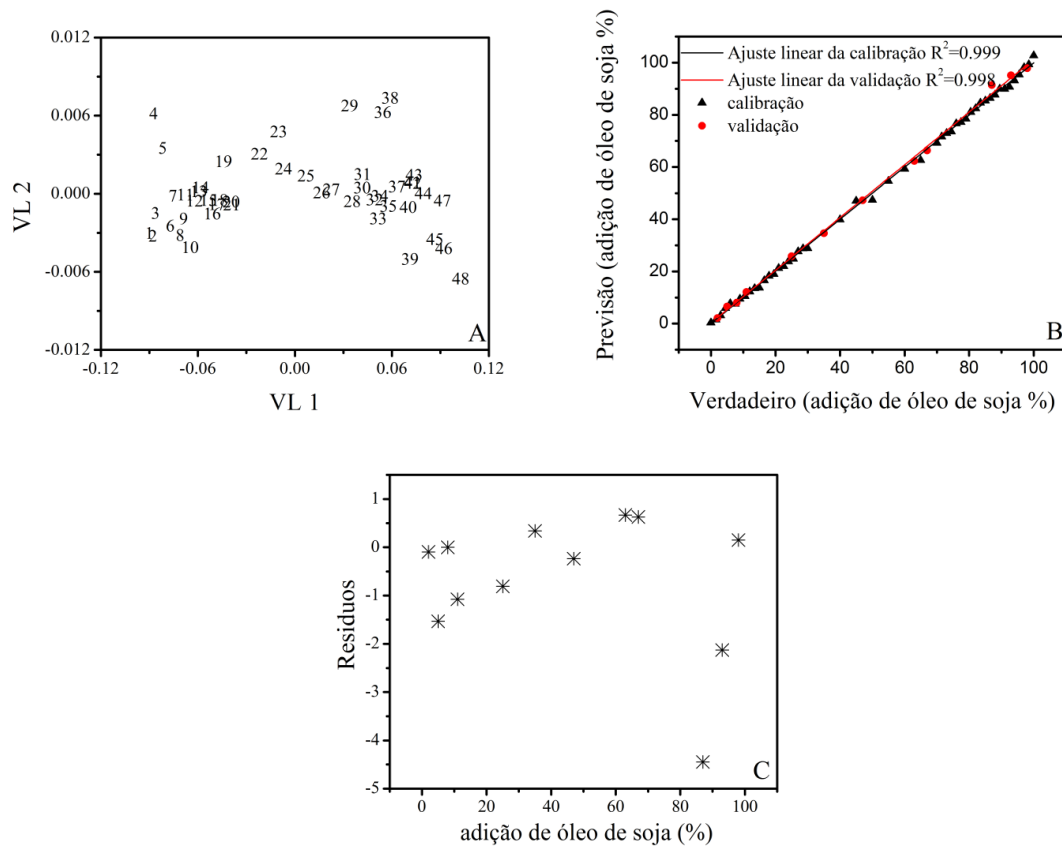
Em "A" têm-se o diagrama de *scores* do modelo PCA, "B" o gráfico de *loadings* para a PC-1 e em "C" as regiões de maior variação do espectro.

Fonte: O AUTOR.

O diagrama de *scores* para os dados de Raman mostra uma clara separação dos grupos de amostras de óleo de soja e azeite de oliva, indicando que a metodologia pode ser utilizada para aferir misturas destas duas classes de óleos vegetais, assim como pôde-se observar para a PCA-NIR. O gráfico de *loadings* obtido para a primeira componente principal indica que, de todo o espectro Raman, três picos são os mais importantes para explicar a separação destas classes de óleos. Todavia, nosso objetivo é propor um modelo preditor do percentual de adição de OS em AOEV, para isso é necessário um modelo supervisionado como o PLS.

O ajuste da regressão PLS foi obtido a partir de duas LV, coeficientes de variância $R^2X = 0,959$, $R^2Y = 0,999$ e capacidade de previsão $Q^2 = 0,998$. A Figura 22 mostra os resultados mais importantes para a análise PLS.

Figura 22 – Regressão PLS a partir dos espectros Raman.



Em "A" têm-se o diagrama de *scores* do modelo PLS, "B" o gráfico de valores previstos *versus* valores verdadeiros e em "C" o gráfico de resíduos para o conjunto de validação.

Fonte: O AUTOR.

Semelhantemente aos resultados obtidos por NIR, o gráfico de *scores* para o modelo PLS-Raman, Figura 22-A, mostra os rótulos das amostras crescendo ordenadamente ao longo do eixo horizontal, indicando que o modelo PLS está bem ajustado e compreende toda a gama de concentrações experimentadas. A Figura 22-B mostra os coeficientes de determinação $R^2 = 0,999$ para o conjunto de calibração e $R^2 = 0,998$ para o de validação. O gráfico de resíduos não apresenta tendência, o que indicanda que não há falta de ajuste no modelo PLS-Raman e os coeficientes relacionados ao erro global de predição do modelo foram $RMSEP = 1,57$ e $RMSE_{cv} = 1,61$.

Finalmente, a tabela 4 apresenta de forma resumida, os valores previstos para o conjunto de validação, obtidos pelos três modelos criados a partir dos espectros NIR, MIR e Raman, comparando-os com os valores verdadeiros.

Tabela 4 – Valores previstos pelo modelo de regressão PLS.

Grupo de validação	NIR	MIR	Raman
2,0	2,9	0,5	2,1
5,0	6,1	7,3	6,5
8,0	8,1	4,7	8,0
11,0	10,2	19,9	12,1
25,0	24,8	29,8	25,8
35,0	33,2	28,2	34,7
47,0	49,5	50,4	47,2
63,0	64,5	69,0	62,3
67,0	68,9	66,2	66,4
87,0	90,9	91,8	91,4
93,0	94,8	96,5	95,1
98,0	97,1	104,1	97,9

Fonte: O AUTOR.

Os valores obtidos a partir da previsão PLS para cada técnica espectroscópica foram comparados através de um teste-*t* pareado juntamente com os valores verdadeiros de adição de OS preparados em laboratório. O *p*-valor obtido foi de 0,103 para a comparação (NIR vs. verdadeiro), 0,108 (MIR vs. verdadeiro) e 0,117 (Raman vs. verdadeiro). Como resultado, pode-se dizer que não há diferenças significativas, ao nível de 95% de confiança, entre as previsões feitas pelos modelos de regressão PLS e os valores verdadeiros de adição de OS preparados em laboratório. O que indica que as abordagens podem ser utilizada para a predição da adulteração em misturas de OS em AOEV, até mesmo a metodologia MIR, que apresentou uma maior dispersão dos dados, como mencionado anteriormente.

A técnica de MIR mostrou uma maior dispersão, provavelmente devido a dificuldade de homogeneização da gota de óleo colocada sobre a superfície da janela de KBr e também ao processo de reutilização da janela após a limpeza, para cada análise subsequente. As análises por NIR e Raman não apresentaram esta dificuldade experimental, pois todas as análises foram feitas diretamente em frascos de vidro, colocados no compartimento do equipamento, sem reutilização imediata.

5.1.4 Conclusões

Os modelos de previsão PLS apresentaram resultados satisfatórios para todas as técnicas espectroscópicas experimentadas. Todos os coeficientes de determinação foram maiores que 0,9 e somente a metodologia MIR apresentou uma dispersão maior nos resultados. As técnicas espectroscópicas tem a vantagem de não necessitar de uma preparação prévia das amostras para análise, as amostras podem ser reutilizadas (não destrutiva) e as análises apresentam grande frequência analítica, uma vez que as medições,

todas realizadas em equipamentos FT, têm tempo de análise de aproximadamente um minuto. Este trabalho mostrou que a espectroscopia NIR e Raman combinadas com algumas ferramentas multivariadas podem ser úteis na indústria e por órgãos de fiscalização como técnicas alternativas para determinar possíveis fraudes no produto, uma vez que são simples de aplicar e não requerem conhecimento químico específico para realização das medições.

5.2 ANÁLISE DO PERFIL LIPÍDICO DE AMOSTRAS DE LEITE DE VACA EM ESPECTROS DE INFRAVERMELHO MÉDIO E RAMAN

O leite é uma mistura de nutrientes extremamente complexa, especula-se que o leite contenha cerca de cem mil constituintes distintos [57], contudo, sua composição principal pode ser descrita por cerca de 12 a 13% de gordura, proteína e lactose, em solução ou suspensão em aproximadamente 87% de água [58,59]. A quantificação dos componentes do leite é crucial para o conhecimento de sua qualidade nutricional, para determinar a sua aptidão para o consumo humano, processamento e ainda para determinar possíveis adulterações. Nesse sentido, novas técnicas para avaliar a qualidade, autenticidade e quantificação dos constituintes presentes no leite são um objetivo contínuo dos setores de controle de qualidade, da indústria de laticínios e agências governamentais.

A gordura presente no leite é um constituinte que recebe destaque, pois é o principal elemento que agrega valor ao produto. Leite rico em gordura tem melhor valor de mercado. A gordura também é o constituinte que sofre a maior variação percentual, isso está associado à alimentação e raça do rebanho, bem como à estação do ano. Por isso, são necessários métodos eficientes e rápidos de quantificação do percentual de gordura no leite fluido.

Os métodos tradicionais utilizados para quantificar o teor de gordura em amostras de leite são o butirômetro de Guerber [60] e o método de Roesse-Gottlieb [61,62], além da metodologia oficial de análise de AG por GC-FID [48]. No entanto, técnicas espectroscópicas tem um grande potencial para tais análises, pois apresentam como principais vantagens a alta frequência analítica, a ausência de etapas de preparo de amostra e a análise não-destrutiva, características importantes para se obter resultados rápidos em produtos perecíveis como o leite.

A fração lipídica do leite ocorre principalmente (cerca de 98%) [63,64] sob a forma de triacilglicerol (TAG), formado pela junção de uma molécula de glicerol ($C_3O_3H_8$) com três AG. A análise desta molécula pode fornecer informações para quantificação do percentual de gordura no leite, além de possíveis adulterações no produto, caso seja adicionada uma porção lipídica de outra origem.

Neste sentido, a proposta deste trabalho é um estudo da fração lipídica que constitui

o leite de vaca fluido por métodos espectroscópicos, com o intuito de propor novos métodos de análise, contribuindo assim para o controle de qualidade do produto.

5.2.1 Objetivos

Determinar qual a molécula TAG mais abundante em amostras de leite de vaca por GC-FID. Utilizar tal molécula como *input* para simulação dos espectros Raman e infravermelho, comparar os espectros simulados com espectros experimentais da gordura extraída do leite para então realizar as atribuições dos modos vibracionais encontrados. Na sequência, utilizar uma abordagem de análise multivariada em um conjunto de espectros de leite fluido com diferentes níveis de concentração de gordura para selecionar quais regiões espectrais estão correlacionadas com a variação do percentual de gordura presente no leite. Finalmente propor regiões espectrais específicas para esta quantificação de gordura e ainda um modelo de regressão múltipla para verificar a potencialidade destes marcadores vibracionais.

5.2.2 Materiais e métodos

Neste estudo uma amostra representativa de leite cru de vaca foi obtida a partir de uma mistura em partes iguais de vinte amostras diferentes, proveniente de tanques de armazenamento de leite cru, adquiridos na área rural da cidade de Juiz de Fora, MG, Brasil. Esta mistura foi utilizada para obter o extrato de gordura (EG) que foi analisado pelas técnicas de GC-FID, infravermelho e Raman.

Para a análise por GC, foram coletados em duplicata 1,0 mL de leite cru, proveniente da mistura descrita anteriormente. A gordura foi extraída utilizando o método descrito por Hara e Radin [65], com solução de extração preparada por adição de três partes de hexano para duas partes de isopropanol, num balão volumétrico. Neste ponto do processo de preparo de amostras têm-se o EG do leite. Após isso, as amostras foram esterificadas de acordo com Christie [47], usando catálise básica.

As análises pela técnica de GC foram realizadas no cromatógrafo fabricado pela Shimadzu modelo GC-2010 Plus (Shimadzu, Kyoto, Japão) equipado com injetor tipo *split-splitless*, auto injetor AOC 20-I e detecção por FID. Este equipamento está instalado na Central Analítica, localizado no departamento de química da Universidade Federal de Juiz de Fora. Medidas realizadas em colaboração com a aluna do programa de pós-graduação em Química da UFJF, Brenda Lee Simas Porto.

O método utilizado nas análises de GC consiste em uma coluna capilar de sílica fundida (CP-SIL 88 para o FAME; 100 m x 0,25 mm x 0,2 mm, Agilent Technologies, Palo Alto, USA) com as seguintes condições cromatográficas: volume de injeção de 1,0 μL no modo split com fluxo de 20 mL/min à 250 $^{\circ}\text{C}$, a temperatura do detector FID foi fixada em 270 $^{\circ}\text{C}$, as temperaturas do forno foram inicialmente programadas para 80 $^{\circ}\text{C}$,

em seguida, um aumento de $4\text{ }^{\circ}\text{C}/\text{min}$ até $220\text{ }^{\circ}\text{C}$, que foi mantida durante 5 min. Depois disso, a temperatura aumentou $4\text{ }^{\circ}\text{C}/\text{min}$ até $240\text{ }^{\circ}\text{C}$ e foi mantida durante 10 min. O gás de arraste utilizado foi o hidrogênio com fluxo de $1,0\text{ mL}/\text{min}$ e a pressão foi de $140,3\text{ kPa}$. Os compostos foram identificados por comparação do TR de uma análise do padrão FAME-37 (Sigma (St. Louis, MO, EUA) sob as mesmas condições. Os ácidos graxos foram quantificados por normalização de área e os valores expressos em g por 100 g de gordura [48].

Para as análises de infravermelho e Raman do EG não foi necessário o passo de esterificação. Desta forma o processo de extração foi realizado, como já descrito, partindo de um volume inicial de $5,0\text{ mL}$ de leite cru. Posteriormente foi analisado também um conjunto de seis amostras de leite cru fluido com diferentes percentuais de gordura, em conformidade com o procedimento de calibração descrito pela *International Dairy Federation* [66] usando amostras modificadas.

Os espectros MIR foram obtidos utilizando um espectrômetro com transformada de Fourier modelo FT-MIR Vertex70 fabricado pela Bruker, equipado com detector de RT-DLaTGS e ótica de KBr, configurado para o modo de transmitância. Este equipamento está instalado no Laboratório de Espectroscopia de Materiais, localizado no departamento de física da Universidade Federal de Juiz de Fora. As análises foram realizadas numa faixa de aquisição de $3500\text{-}900\text{ cm}^{-1}$, resolução de 4 cm^{-1} e 32 acumulações por amostra. Para espectros de EG, foi utilizado uma gota do extrato de gordura colocado diretamente em uma janela de KBr, o fundo foi feito com o compartimento de amostra vazio. Para as análises de leite fluido foi utilizado um kit de células desmontável com janelas de CaF_2 e espaçador de Teflon de $0,05\text{ mm}$ de espessura, neste caso foi utilizada água deionizada como fundo.

Espectros Raman foram obtidos com um espectrômetro com transformada de Fourier modelo FT-Raman RFS-100 fabricado pela Bruker, equipado com detector de germanium resfriado com nitrogênio líquido e radiação de excitação no comprimento de onda de 1064 nm a partir de um laser Nd:YAG, configurado para o modo retroespalhamento. Este equipamento está instalado no Núcleo de Espectroscopia e Estrutura Molecular, localizado no departamento de química da Universidade Federal de Juiz de Fora. As condições de análise foram: 100 acumulações por amostra, 4 cm^{-1} da resolução compreendendo a faixa espectral de $3500\text{-}100\text{ cm}^{-1}$ e potência do laser de 500 mW . Os espectros Raman do extrato de gordura e das análises de leite fluido foram obtidos acondicionando as amostras dentro de frascos de vidro com $8,0\text{ milímetros}$ de diâmetro interno. Este frasco foi colocado diretamente no compartimento de amostras do equipamento.

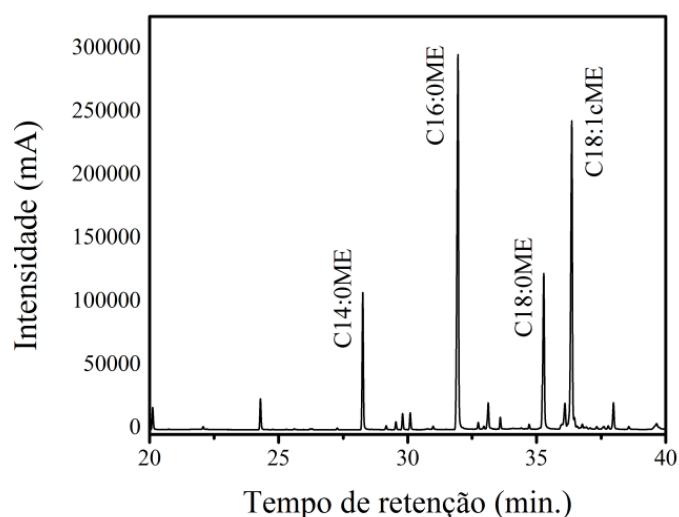
Os cálculos teóricos dos espectros Raman e infravermelho da molécula TAG majoritária no leite de vaca foram feitos em colaboração com a Dra. Georgia Maria Amaral Junqueira, do Laboratório de Simulação Computacional do Departamento de Física da

Universidade Federal de Juiz de Fora. Inicialmente foi feita uma análise conformacional da molécula usando Dinâmica Molecular com campo de força AMBER. A geometria de menor energia foi otimizada por Teoria de Funcional Densidade (DFT) com o funcional B3LYP [67, 68] com valência [6-31G] [69]. Todos os cálculos foram feitos utilizando o conjunto de *software* GAMES [70].

5.2.3 Resultados e discussões

O cromatograma da gordura do leite analisada pela técnica clássica de GC é mostrado na Figura 23. As espécies mais abundantes de AG estão destacados na figura.

Figura 23 – Perfil de AG pela técnica de GC para uma amostra de leite de vaca.



Fonte: O AUTOR.

A análise por GC mostrou que três AG são responsáveis por cerca de 70% da composição total de AG provenientes da gordura do leite. Estes três AG majoritários e suas proporções de abundância na amostra de leite são C16:0 ácido palmítico (31,09%), C18:1c ácido oleico (26,77%), C18:0 ácido esteárico (12,03%), e numa sequência foi identificado C14:0 ácido mirístico (8,43%) e outros dezesseis AG sem um percentual representativo em comparação com os já citados. Estes resultados corroboram com uma ampla discussão apresentada por Contarini *et al* [71]. A quantificação completa dos AG identificados é apresentada na Tabela 5.

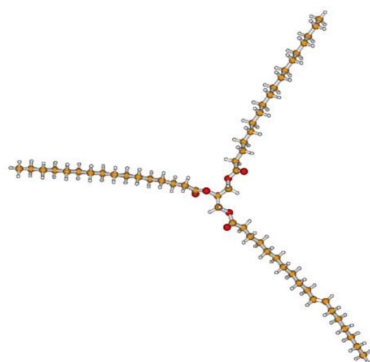
Tabela 5 – Quantificação dos AG da amostra de leite cru analisados por GC-FID.

AG	TR (min)	Fórmula molecular	Nome comum	massa (g/100 g)
C4:0	10,54	C ₄ H ₈ O ₂	ácido butírico	0,86
C6:0	12,73	C ₆ H ₁₂ O ₂	ácido capríco	0,82
C8:0	16,10	C ₈ H ₁₆ O ₂	ácido caprílico	0,56
C10:0	20,12	C ₁₀ H ₂₀ O ₂	ácido cáprico	1,30
C11:0	22,07	C ₁₁ H ₂₂ O ₂	ácido undecílico	0,16
C12:0	24,29	C ₁₂ H ₂₄ O ₂	ácido láurico	1,75
C14:0	28,26	C ₁₄ H ₂₈ O ₂	ácido mirístico	8,43
C14:1c	29,80	C ₁₄ H ₂₆ O ₂	ácido miristoleico	0,92
C15:0	30,09	C ₁₅ H ₃₀ O ₂	ácido pentadecílico	0,99
C16:0	31,95	C₁₆H₃₂O₂	ácido palmítico	31,09
C16:1c	33,12	C ₁₆ H ₃₀ O ₂	ácido palmitoleico	1,89
C17:0	33,58	C ₁₇ H ₃₄ O ₂	ácido margárico	0,65
C18:0	35,27	C₁₈H₃₆O₂	ácido esteárico	12,03
C18:1t	36,09	C ₁₈ H ₃₄ O ₂	ácido vacênico	2,06
C18:1c	36,36	C₁₈H₃₄O₂	ácido oleico	26,77
C18:2cc	37,97	C ₁₈ H ₃₂ O ₂	ácido linoleico	2,00
C20:0	38,57	C ₂₀ H ₄₀ O ₂	ácido araquídico	0,16
γC18:3ccc	39,64	C ₁₈ H ₃₀ O ₂	ácido γ-linolênico	0,15
C20:1c	40,04	C ₂₀ H ₃₈ O ₂	ácido eicosenóico	0,40
αC18:3ccc	4,30	C ₁₈ H ₃₀ O ₂	ácido α-linolênico	1,00

Fonte: O AUTOR.

Desta maneira a análise por GC destacou os ácidos graxos palmítico, oleico e esteárico como os mais abundantes. Foi então proposta uma molécula de triacilglicerol a partir destes três AG, resultando no TAG de fórmula molecular C₅₅H₁₀₄O₆. Esta molécula TAG já foi reportada na literatura como presente no leite de vaca [72], sendo assim, foi selecionada como *input* para os cálculos dos espectros Raman e infravermelho, por simulação computacional. A Figura 24 mostra a conformação otimizada por simulação computacional da molécula TAG, formada por estes três AG majoritários.

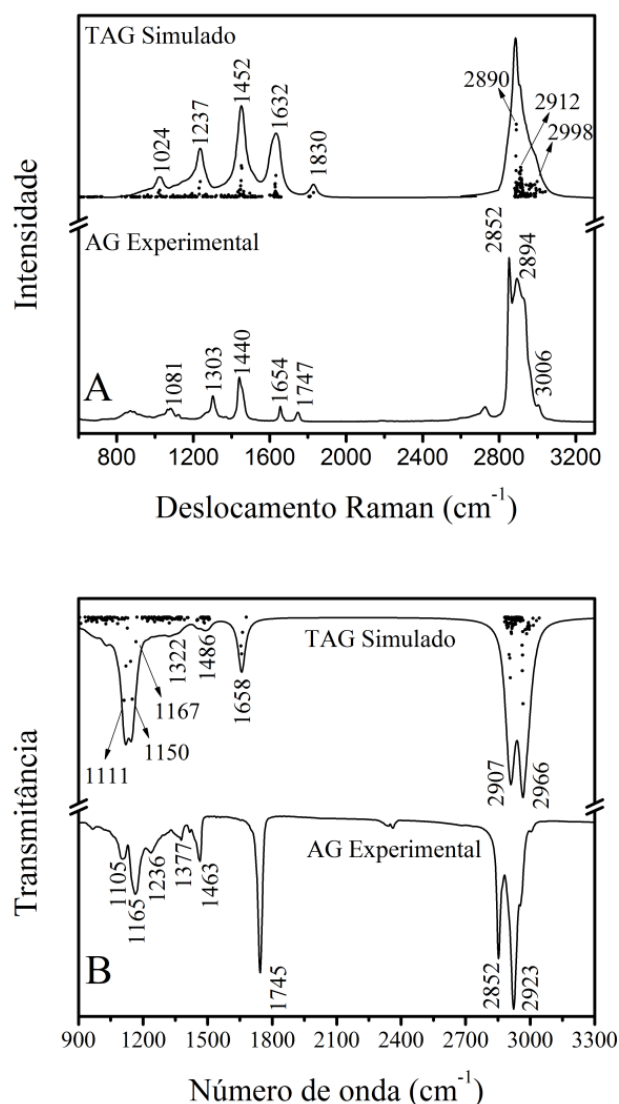
Figura 24 – Molécula TAG otimizada por cálculo computacional.



Fonte: LABORATÓRIO DE SIMULAÇÃO COMPUTACIONAL DO DEP. DE FÍSICA DA UFJF - Dra. GEORGIA JUNQUEIRA.

Na sequência, a simulação computacional apresentou os espectros teóricos Raman e infravermelho desta molécula TAG simulada no vácuo e suas respectivas atribuições de modos vibracionais. A Figura 25-A mostra um comparativo do espectro Raman simulado a partir da molécula TAG selecionada pela técnica de GC e o espectro Raman obtido experimentalmente do EG do leite de vaca. O mesmo comparativo é feito na Figura 25-B, para as análises de infravermelho médio.

Figura 25 – Espectros Raman e infravermelho da gordura do leite de vaca.



Em "A" estão representados espectros Raman e em "B" espectros infravermelho. A parte superior representa o perfil obtido por cálculo computacional da molécula TAG e abaixo são os espectros experimentais do EG do leite.

Fonte: O AUTOR.

Pode se ver claramente que a forma global da linha do espectro Raman e infravermelho da molécula TAG simulada assemelha-se muito ao perfil obtido experimentalmente para o EG do leite. Há um desvio ao se comparar os picos do perfil experimental e teórico devido ao fato de que as simulações foram feitas para uma molécula isolada, no vácuo. Desta forma não foram consideradas a contribuição de outros AG que estão presentes no perfil experimental, os efeitos de matriz devido a interações com outras moléculas e até mesmo o processo utilizado para extração da porção lipídica. Ainda assim, devido a forma de linha ser reproduzida, é possível utilizar as atribuições obtidas teoricamente para inferir atribuições sobre os espectros experimentais. A Tabela 6 apresenta a atribuição de modos

vibracionais, obtidos por simulação, para o espectro Raman da molécula TAG. É proposta uma correspondência destes modos vibracionais com os espectros Raman experimentais do EG e do leite fluido (o espectro Raman de leite fluido será apresentado ainda nesta seção, Figura 27).

Tabela 6 – Atribuição dos modos vibracionais Raman para a gordura de leite de vaca.

TAG*	EG**	Leite fluido**	Modo vibracional*
2998	3006	-	H-CH ₂ estiramento assimétrico
2912	2894	2894	H-C estiramento assimétrico
2890	2852	2852	H-C estiramento simétrico
1830	1747	1747	C=O estiramento
1632	1654	1654	C=C estiramento
1452	1440	1440	CH ₂ tesoura
1237	1303	1301	CH ₂ torção
1024	1081	1083	CH ₃ torção

*obtidos por simulação computacional. **resultados experimentais. valores em cm^{-1}

Fonte: O AUTOR.

Moléculas TAG são conhecidas por apresentar sinais Raman fortes, uma vez que têm grandes volumes de polarizabilidade originado a partir das porções de radicais acila presentes na composição do TAG [73, 74]. Efeitos de matriz são esperados, e discutidos sob diferentes aspectos [75–77]. Na sequência é apresentada a Tabela 7, na qual é destacada as regiões mais importantes do espectro infravermelho para a molécula de TAG obtido teoricamente, bem como as medições experimentais do EG do leite e do leite cru fluido (o espectro infravermelho do leite fluido será apresentado ainda nesta seção, Figura 29).

Tabela 7 – Atribuição dos modos vibracionais de absorção infravermelha para a gordura de leite de vaca.

TAG*	EG**	Leite fluido**	Modo vibracional*
2966	2923	2922	C-H estiramento assimétrico
2907	2852	2852	C-H estiramento simétrico
1658	1745	1739	C=O estiramento
1486	1463	1456	CH ₂ flexão simétrica
1322	1377	1379	CH ₂ flexão assimétrica (fora do plano)
1167	1236	1245	CH ₂ (C=O)OCH ₂ estiramento assimétrico + CH ₂ flexão simétrica (fora do plano)
1150	1165	1155	C-C + -C-O estiramento + -CH ₂ torção
1111	1105	-	CH ₂ flexão assimétrica (fora do plano)

*obtidos por simulação computacional. **resultados experimentais. valores em cm^{-1}

Fonte: O AUTOR.

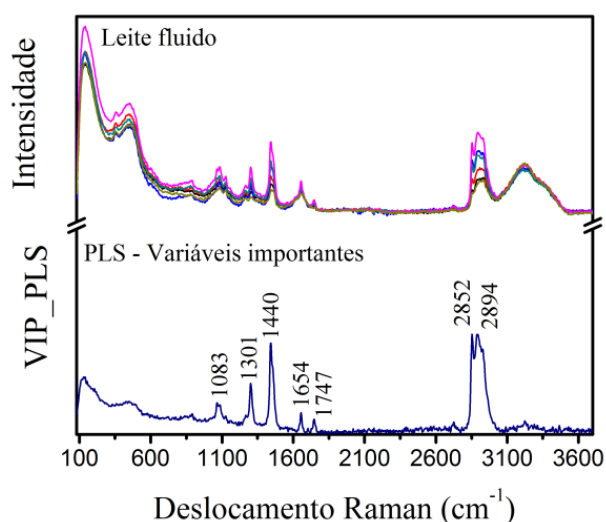
A complexidade química da molécula analisada e o fato da composição do leite ser uma mistura de diferentes tipos de gorduras, minerais, vitaminas, proteínas, açúcares

e outros [58, 59], torna difícil a comparação exata dos números de onda obtidos por simulação e experimentação. No entanto, como já mencionado, o perfil encontrado é muito semelhante, o que torna a atribuição viável. Uma comparação das atribuições previstas na Tabela 7 com os relatados na literatura, mostram que a maioria das atribuições estão de acordo com os resultados declarados na literatura tanto para Raman [78, 79] como infravermelho [80, 81].

Neste ponto o perfil do espectro Raman e infravermelho, tanto do extrato de gordura, como do leite cru fluido, de amostras de leite, é conhecido. As bandas relacionadas com a porção de gordura já foram identificadas. Foi então analisado um conjunto de seis amostras de leite cru fluido com diferentes teores de gordura, a saber: 1,75%, 2,49%, 3,20%, 5,75%, 6,01% e 7,36%, com o objetivo de verificar se existem bandas de gordura que variam, umas mais que as outras, com o aumento da concentração de gordura nas amostras.

A Figura 26, mostrada a seguir, está dividida em duas partes para facilitar a comparação. Acima é mostrado o espectro Raman das amostras de leite fluido com diferentes concentrações de gordura e abaixo esta representado o gráfico das variáveis mais importantes na projeção (VIP) [82–84] de uma regressão PLS ajustada para predição do teor de gordura a partir do espectro completo. A análise do gráfico VIP permite extrair quais variáveis são as mais eficientes para predizer a concentração de gordura a partir de espectros Raman de amostras de leite cru.

Figura 26 – Espectros Raman de amostras de leite fluido com diferentes concentrações de gordura e gráfico de variáveis importantes da regressão PLS.



Fonte: O AUTOR.

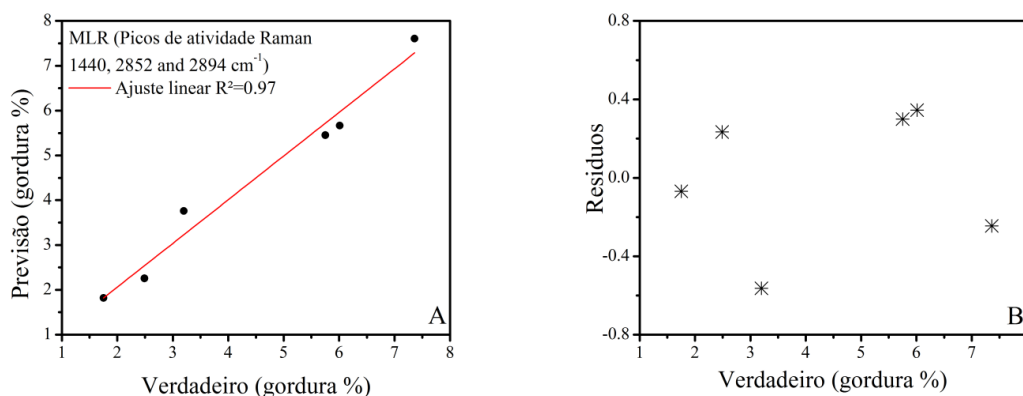
Devido à sobreposição de algumas faixas espectrais e variações devido à mudança na concentração de outros componentes não controlados, uma inspeção visual não é muito

eficiente para selecionar quais são os marcadores vibracionais mais significativos para explicar o percentual de gordura no leite. Por isso a estratégia foi utilizar uma regressão PLS calibrada para prever o teor de gordura. Na verdade, a regressão PLS correlaciona as regiões espectrais que variam à medida que muda uma dada propriedade ou substância investigada, neste caso, com o percentual de gordura do leite.

O modelo PLS foi ajustado a partir de uma matriz contendo todos os seis espectros Raman centrados na média, com duas LV, e coeficientes $R^2X = 0,96$, $R^2Y = 0,98$ e $RMSEC = 0,44$. As magnitudes dos modos vibracionais Raman em 1440 cm^{-1} , 2852 cm^{-1} e 2894 cm^{-1} são mais elevadas do que as outras variáveis na análise VIP (Figura 26). Por isso, estes podem ser os modos mais importantes para quantificar o percentual de gordura no leite de vaca.

A fim de verificar a resposta destes marcadores vibracionais Raman, foi realizado uma regressão MLR a partir destes deslocamentos Raman em 1440 , 2852 e 2894 cm^{-1} . Como resultado, a curva de correlação entre os valores reais da percentagem de gordura no leite fluido e os valores previstos pelo modelo MLR, Figura 27-A, apresentou coeficiente $R^2 = 0,97$ e ausência de falta de ajuste pela análise do gráfico de resíduos Figura 27-B. Tais parâmetros apontam como positiva a utilização destes três modos Raman ativos para a quantificação de gordura.

Figura 27 – Resultados para quantificação a partir dos espectros Raman com regressão MLR.

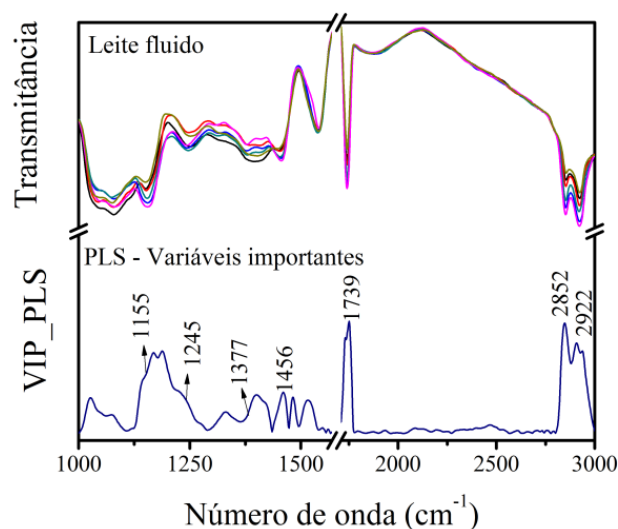


Em "A" está representado o gráfico de correlação entre os valores previstos pelo modelo MLR *versus* valores verdadeiros obtidos pelo analisador eletrônico LactoScope; e "B" é o gráfico de resíduos para o mesmo conjunto de pontos.

Fonte: O AUTOR.

As mesmas seis amostras de leite fluido, com diferentes concentrações de gordura, foram analisadas por espectroscopia no infravermelho, como ilustrado na Figura 28. A região abaixo de 1000 cm^{-1} não é mostrada na mesma porque a janela do CaF_2 não é transparente nesta região do espectro. O intervalo entre 1570 e 1710 cm^{-1} , bem como a região acima de 3000 cm^{-1} , também não são mostrados devido à saturação causada pela forte absorção da água presente no leite fluido.

Figura 28 – Espectros de infravermelho de amostras de leite fluido com diferentes concentrações de gordura e gráfico de variáveis importantes da regressão PLS.

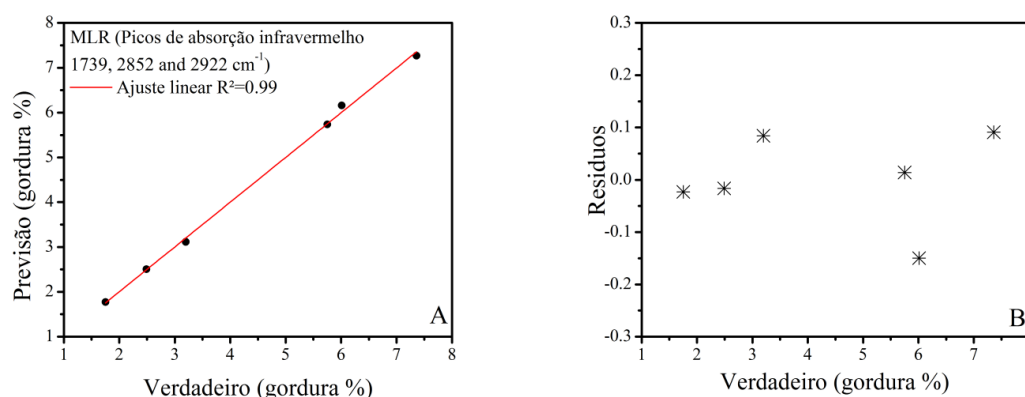


Fonte: O AUTOR.

De forma análoga ao realizado com os espectros Raman de leite fluido, foi proposto um modelo de regressão PLS a partir de uma matriz com todos os seis espectros de infravermelho, centrados na média, com duas LV e $R^2X = 0,89$, $R^2Y = 0,98$ e $RMSEC = 0,36$. A análise do gráfico VIP, Figura 28, mostra claramente que os modos vibracionais no infravermelho em 1739 cm^{-1} , 2852 e 2922 cm^{-1} podem ser utilizados como marcadores para quantificar o percentual de gordura no leite de vaca fluido. Os modos vibracionais abaixo de 1658 cm^{-1} observados na molécula TAG, 25-B, se sobrepõe com outros componentes do leite, tais como a água em torno de 1660 cm^{-1} , proteínas em 1550 cm^{-1} , e a lactose em 1040 cm^{-1} [80,81]. Desta forma, esta região abaixo de 1658 cm^{-1} apresenta dificuldades para prever a concentração de gordura no leite fluido.

Avaliou-se um ajuste de regressão MLR, baseado nos três modos de infravermelho selecionados pela análise VIP, 1739 , 2852 e 2922 cm^{-1} . Os resultados, apresentados abaixo na Figura 29-A-B, confirmam que os modos escolhidos são bons marcadores vibracionais, com coeficiente $R^2 = 0,99$ e resíduos distribuídos aleatoriamente.

Figura 29 – Resultados para quantificação a partir dos espectros de infravermelho com regressão MLR.



Em "A" está representado o gráfico de correlação entre os valores previstos pelo modelo MLR *versus* valores verdadeiros obtidos pelo analisador eletrônico LactoScope; e "B" é o gráfico de resíduos para o mesmo conjunto de pontos.

Fonte: O AUTOR.

A Tabela 8 apresenta os valores de LD e LQ para as bandas vibracionais utilizadas na regressão MLR. Pode-se perceber que ambos, LD e LQ, possuem valores abaixo dos medidos em todas as amostras, visto que a faixa de concentração experimentada foi de 1,75 à 7,36%. Tais índices auxiliam na verificação da qualidade das quantificações espectroscópicas realizadas, por serem estes parâmetros utilizados na validação tradicional de modelos de regressão.

Tabela 8 – Limites de detecção e quantificação Raman e infravermelho para os modos vibracionais selecionados.

Técnica	$\nu(cm^{-1})$	LD*(%)	LQ*(%)
Raman	1440	0,23	0,77
	2852	0,27	0,88
	2894	0,19	0,62
Infravermelho	1739	0,02	0,06
	2852	0,1	0,34
	2922	0,06	0,19

*índices calculados utilizando uma amostra de concentração intermediária às medidas no estudo, com 3,2% de gordura.

Fonte: O AUTOR.

Além disso, o teste de Durbin-Watson foi realizado sobre o conjunto de determinação do percentual de gordura em amostras de leite fluido, tanto na abordagem Raman como infravermelho, e obteve *p*-valores de 0,41 (infravermelho) e 0,48 (Raman). Como ambos foram maiores que 0,05, pode-se dizer que, ao nível de 95% de confiança, não há evidências de correlação na distribuição dos resíduos, indicando que ambas as regressões MLR não apresentam falta de ajuste.

5.2.4 Conclusões

A análise por cromatografia a gás permitiu revelar qual a porção majoritária de ácidos graxos presentes em amostras de leite cru fluido, ácido palmítico, oleico e esteárico. Com base neste resultado foi proposta uma molécula TAG, $C_{55}H_{104}O_6$, para serem simulados os espectros Raman e infravermelho. Estes cálculos teóricos permitiram realizar atribuições vibracionais nas medidas experimentais de extrato de gordura e de leite fluido. A análise multivariada por PLS e a lista de variáveis mais importantes deste modelo multivariado permitiu propor três marcadores vibracionais para a análise Raman (1440 , 2852 e 2894 cm^{-1}) e outros três marcadores para a análise no infravermelho (1739 , 2852 e 2922 cm^{-1}) que descrevem com exatidão as concentrações de gordura em amostras de leite fluido.

As análises com duas técnicas espectroscópicas diferentes apresentaram marcadores moleculares diferentes, o que é esperado pela característica de complementaridade entre a técnica Raman e infravermelho. Isso pode ajudar a uma melhor quantificação da gordura do leite ao usar ambas as técnicas concomitantemente, pode também ser muito útil para o monitoramento da qualidade do leite, além de tornar possível a detecção de fraude por adição de gordura de outra origem.

5.3 QUANTIFICAÇÃO SIMULTÂNEA DOS TEORES DE RIFAMPICINA, ISONIAZIDA, PIRAZINAMIDA E ETAMBUTOL EM COMPRIMIDOS PARA TRATAMENTO ANTI-TUBERCULOSE POR ESPECTROSCOPIA RAMAN

A tuberculose é uma enfermidade espalhada em todo o mundo, é a segunda principal causa de morte por doenças infecciosas, atrás apenas do vírus HIV, portanto, é considerada pela Organização Mundial de Saúde (OMS) um dos maiores problemas de saúde. Em 2013 foram 9 milhões de novos casos desta enfermidade, sendo 11% coinfectados com HIV e 1,5 milhões de mortes foram estimadas [85].

Em 1882, Robert Koch isolou a *Mycobacterium tuberculosis* (MTB), agente responsável pela transmissão da tuberculose. A infecção afeta principalmente os pulmões e a transmissão é dada pela MTB em gotículas expelidas pelos pacientes com tuberculose pulmonar bacilífera. Sessenta e três anos após o isolamento da MTB, em 1945, com o desenvolvimento da estreptomicina, o tratamento da tuberculose tornou-se possível. A primeira terapia de combinação para o tratamento da tuberculose apareceu em 1970, que consiste em isoniazida (INH), estreptomicina e ácido p-amino salicílico. Depois disso, a inclusão de rifampicina (RIF) e pirazinamida (PZY) aumentou as taxas de sucesso e reduziu o tempo do tratamento [86].

O tratamento combinado com vários fármacos é essencial [87] para alcançar a cura do paciente sem que haja recrudescência e também para evitar o aparecimento de mutantes

resistentes aos medicamentos. Desde 2010 o tratamento de novos casos de tuberculose consiste em um regime de dois meses de uso de quatro medicamentos de primeira linha, RIF, INH, PYZ, e etambutol (EMB) e mais quatro meses de RIF e INH [86].

A combinação de dose fixa simplifica a prescrição de medicamentos e a gestão da oferta, podendo até mesmo reduzir o risco de resistência microbacteriana devido ao tratamento inadequado. Portanto, para garantir um tratamento adequado, a OMS e a União Internacional Contra Tuberculose e Doenças Pulmonares recomendam o uso de combinação de dose fixa [75].

Estudos anteriores de espectroscopia Raman incluem avaliações individuais das moléculas de EMB [88, 89], RIF [90–92], INH [93, 94] e PYZ [95], porém não há relatos prévios na literatura sobre a quantificação simultânea dessas quatro drogas.

5.3.1 Objetivos

Desenvolvimento de um método baseado em espectroscopia Raman, rápido e eficiente para a determinação simultânea da RIF, INH, PYZ e EMB em comprimidos de combinação de dose fixa. Após isso, comparar os resultados com o método tradicional de análises por cromatografia a líquido para verificar a qualidade do método proposto.

5.3.2 Materiais e métodos

Padrões de referência RIF, INH, EMB PYZ e foram obtidos a partir da *US Pharmacopeia* (Rockville, EUA). Foi preparado o excipiente contendo todos os componentes da formulação na ausência dos princípios ativos. Diferentes quantidades de mistura excipiente e comprimidos em pó e princípios ativos foram misturados para preparar um conjunto de calibração (trinta amostras) e de validação (sete amostras) para serem utilizados no ajuste do modelo de quantificação baseado em espectroscopia Raman. Ambos os conjuntos foram também analisados através do método de LC-DAD.

Cinco amostras comerciais de comprimidos de dose fixa combinada, de três lotes diferentes (SL2076, SL2094 e SL2145) fabricados pela Svizera Labs (Mumbai, Índia) e distribuídos pela Svizera Europe BV (Almere, Holanda) foram analisadas por ambas as técnicas para serem comparados.

As análises de LC foram realizadas num cromatógrafo fabricado pela Agilent modelo 1200 (Santa Clara, EUA), equipado com uma bomba quaternária, um forno de coluna, um amostrador automático e um detector de arranjo de diodos. Este equipamento está instalado no Laboratório de Controle de Qualidade, localizado no departamento de produtos farmacêuticos da Universidade Federal de Minas Gerais. Medidas realizadas em colaboração com a doutoranda Paula Rocha Chellini.

As condições de análise por LC foram: separação realizada numa coluna Purospher

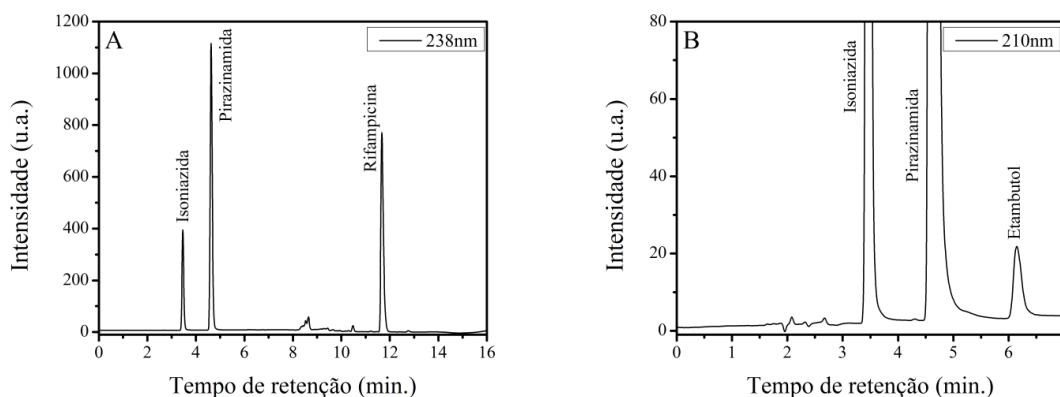
STAR RP18e (250 x 4,6 mm d.i., 5 μm de tamanho de partícula) fabricada pela Merck (Darmstadt, Alemanha), a 40 °C. Detecção UV foi feita a 238 nm para a RMP e INH e à 210 nm para PYZ e EMB. A fase móvel A foi uma mistura de tampão de fosfato de sódio monobásico 20 mM com 0,2% de trietilamina (o pH foi ajustado para 7,0 com ácido fosfórico) e acetonitrila (96:4, v/v) e a fase móvel B foi acetonitrila. O gradiente de eluição foi realizado da seguinte forma: primeiramente 100% da fase móvel A durante 5 min, em seguida a fase móvel B foi elevada até 52% em 7 min., a fase móvel B foi mantida a este nível por 12 min, em 12,01 min a fase móvel A foi transferido de volta para 100% até 17 min. (reequilíbrio). A taxa de fluxo foi de 1,5 mL/min e o volume de injeção foi de 25 μL .

As análises por espectroscopia Raman foram feitas com um espectrômetro com transformada de Fourier modelo FT-Raman RFS-100 fabricado pela Bruker, equipado com detector de germanium resfriado com nitrogênio líquido e radiação de excitação no comprimento de onda de 1064 nm a partir de um laser Nd:YAG, configurado para o modo retroespalhamento. Este equipamento está instalado no Núcleo de Espectroscopia e Estrutura Molecular, localizado no departamento de química da Universidade Federal de Juiz de Fora. Os parâmetros otimizados de aquisição de Raman utilizados neste estudo foram 100 acumulações, resolução de 4 cm^{-1} , potência de laser de 300 mW e espectros obtidos na faixa de 3500-50 cm^{-1} . Para tais análises cada comprimido foi pulverizado, acondicionado em frascos de vidro com 8,0 mm de diâmetro interno e colocados diretamente no compartimento do equipamento para a obtenção dos espectros.

5.3.3 Resultados e discussões

A Figura 30 ilustra uma análise realizada pela técnica LC com abordagem de detecção em duplo comprimento de onda. Os tempos de retenção médio foram 3,5 min para INH, 4,6 min para PYZ, 6,1 min para EMB e 11,6 min para RIF. Não será feita uma discussão maior sobre as análises pela técnica de LC pois, neste trabalho, ela foi utilizada apenas para fornecer os valores de referência das concentrações dos analitos supracitados, tanto para o modelo de calibração por espectroscopia Raman quanto para comparação dos resultados de quantificação.

Figura 30 – Análise de fármacos anti-tuberculose pela técnica de LC.



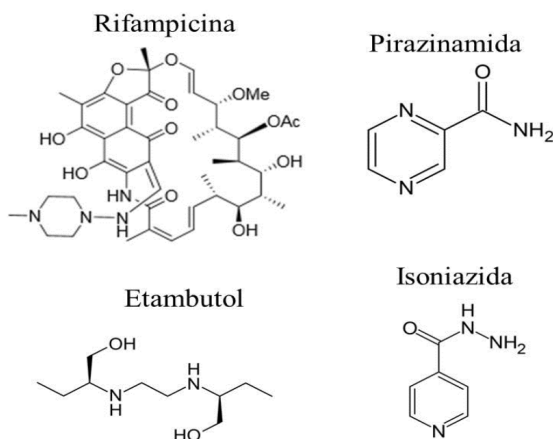
Em "A" detecção de RIF e INH e em "B" detecção de EMB e PYZ.

Fonte: LABORATÓRIO DE CONTROLE DE QUALIDADE DO DEP. DE FARMÁCIA DA UFMG - Dra. PAULA CHELLINI.

Ao contrário do método de LC, a metodologia Raman aqui descrita não necessita de dissolução da amostra, uma vez que os espectros Raman são obtidas apenas colocando a amostra, em forma de pó, diretamente no equipamento. Vale a pena ressaltar que a princípio a ideia era analisar o comprimido inteiro, contudo, após testes iniciais, constatou-se que o involucro do comprimido se degradava com a incidência do laser, por tanto, optou-se por pulverizá-lo antes das análises.

Como já mencionado, o número de modos vibracionais para cada molécula é dado por $3N-6$, desta forma análise de moléculas complexas pode revelar um número muito elevado de modos Raman ativos. A Figura 31 mostra a estrutura molecular dos quatro fármacos alvo deste estudo.

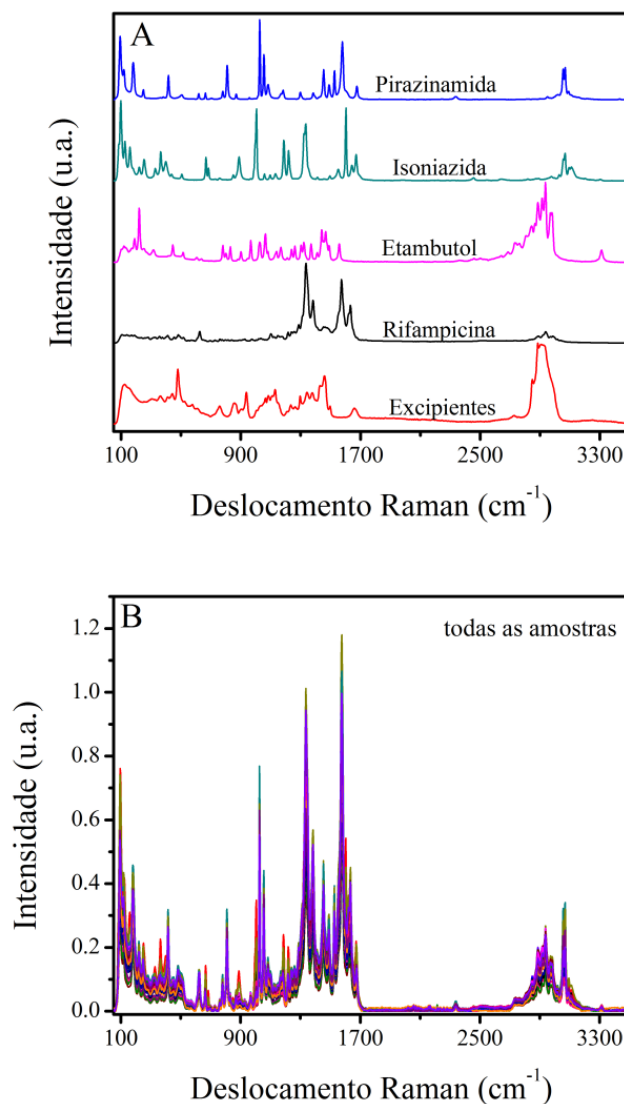
Figura 31 – Estrutura molecular dos quatro fármacos analisados.



Fonte: <http://en.wikipedia.org/>. Acessado em junho de 2015.

Em outras palavras, devido à complexidade da estrutura química dos compostos analisados os respectivos espectros Raman, mesmo dos padrões isolados (Figura 32-A), apresentam várias bandas vibracionais ativas no Raman. Além disso, a análise de comprimidos em dose fixa contendo os quatro princípios ativos acrescidos da mistura de excipientes, Figura 32-B, sofre efeitos de matriz e sobreposição de bandas, acarretando mais uma vez em muitos modos vibracionais e conseqüentemente, uma difícil atribuição de bandas. Foram testadas diferentes condições de aquisição espectral. Observou-se que resolução maior que 6 cm^{-1} omite alguns modos vibracionais e resolução abaixo de 2 cm^{-1} incorpora um ruído muito elevado na medição sem novos modos de vibração. O número de varreduras também foi testado e observou-se que 50 varreduras por amostra é muito baixo de acumulações, o que resulta numa variação na linha base dos espectros. Nenhum ganho significativo na qualidade dos espectros foi observada usando 150 acumulações ou mais (200, 300 e 400 acumulações foram testadas) e como o número de acumulações é diretamente proporcional ao tempo de análise, valores excessivamente elevados não promovem uma boa relação com a frequência analítica. Potência do laser acima 500 mW resultou em um maior fundo de luminescência e uma definição da linha de base ruim. Sendo assim os espectros foram obtidos com resolução de 4 cm^{-1} , 100 acumulações por amostra, potência de laser de 300 mW e na faixa de $3500\text{-}50\text{ cm}^{-1}$.

Figura 32 – Espectros Raman do conjunto de fármacos analisados.



Em "A" espectros de amostras padrão e mistura de excipientes e em "B" coleção de espectros das amostras de dose fixa.

Fonte: O AUTOR.

Alguns modos vibracionais presentes nos espectros acima foram previamente descritos. Kalkar *et al* [95] descreve modos de vibração de PYZ em 3055 cm^{-1} (estiramento CH), 809 cm^{-1} (C=O no plano), 1027 cm^{-1} 1579 cm^{-1} (vibrações de anel). Espectro Raman de INH foi relatado [93, 94] com vibrações ativas em 1334 cm^{-1} , 1604 cm^{-1} e próximo de 670 cm^{-1} . EMB também foi mencionado [89] com bandas em 3311 cm^{-1} (estiramento OH), 1263 cm^{-1} (CH no plano), 800 cm^{-1} (CH fora do plano) e 220 cm^{-1} (OH flexão fora do plano). Finalmente, análises RIF relataram [90–92] intensas bandas ao redor $1300\text{--}1400\text{ cm}^{-1}$ e em 1570 cm^{-1} . No entanto, a utilização de modos vibracionais singulares num espectro com inúmeras contribuições para quantificar simultaneamente cada medicamento pode não ser eficaz. Neste caso, uma análise multivariada utilizando o espectro completo

é recomendável para caracterizar a amostra e quantificar simultaneamente os quatro fármacos alvo. Esta é justamente a proposta deste estudo. Não se pretende realizar aqui uma atribuição completa dos modos de vibração, pois estes não serão utilizados como marcadores vibracionais individuais.

A Figura 32-B mostra a coleção completa de espectros Raman das 42 amostras analisadas, com concentrações previamente determinadas pelo método de LC. Cada um desses espectros foi utilizado para a abordagem estatística multivariada. A curva de calibração PLS foi ajustada a partir de uma matriz de dados de ordem 30x2052, onde cada linha da matriz representa uma amostra, neste caso foram 30, e cada coluna um número de onda do espectro Raman, neste caso 2052. A matriz \mathbf{Y} de concentrações foi então uma matriz 30x4. Todos os espectros foram centrados na média, o número de LV foi cinco e os coeficientes obtidos para o modelo PLS foram $R^2X = 0.98$, $R^2Y = 0.95$ e $Q^2 = 0.90$.

Os coeficientes relacionados a predição de cada fármaco é apresentado na Tabela 9.

Tabela 9 – Coeficientes de determinação e RMSE do ajuste PLS para os fármacos RIF, INH, PYZ e EMB.

Coeficientes	RIF	INH	PYZ	EMB
R_c^2	0,95	0,92	0,98	0,98
R_v^2	0,97	0,91	0,96	0,96
RMSEC	7,51	11,04	6,40	7,25
RMSEP	4,31	4,64	5,85	5,72
Durbin-Watson	0,11	0,63	0,32	0,19

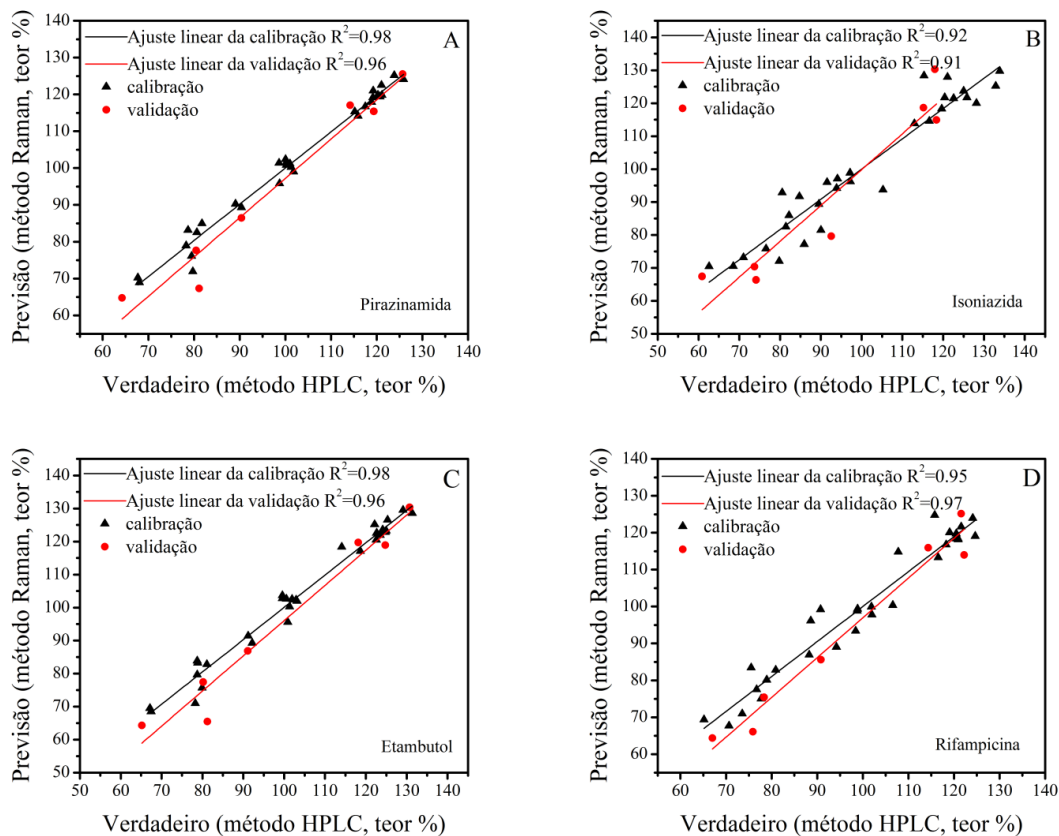
Fonte: O AUTOR.

Os valores de R^2 foram superiores a 0,9 para todas as comparações, indicando uma boa correlação entre as quantificações feitas pelo método tradicional por LC e os valores obtidos por regressão PLS a partir dos espectros de Raman. Os coeficientes RMSEC e RMSEP representam um erro global no modelo PLS de modo que os valores obtidos estão adequados para o intervalo investigado, que compreende uma faixa de variação de cerca de 60 a 135% das especificações definidas. Finalmente o teste de Durbin-Watson foi realizado após a verificação da normalidade dos dados (teste de Shapiro-Wilk: 0,82 RIF, 0,94 INH, 0,10 PYZ e 0,07 EMB) e apresentou todos os p -valores maiores do que 0,05, indicando que, ao nível de 95% de confiança, não existe evidência de correlação na distribuição dos resíduos entre os valores previstos pelo método de Raman, mostrando que o modelo está bem ajustado.

A Figura 33 mostra os gráficos de correlação entre os valores obtidos pela regressão PLS a partir dos espectros Raman contra os valores obtidos pelo método tradicional por HPLC, para o conjunto de calibração e de validação, para os quatro fármacos separadamente. Tais resultados de quantificação simultânea de pirazinamida, isoniazida, etambutol e

rifampicina foram obtidos em comprimidos de dose fixa para o tratamento anti-tuberculose, por espectroscopia Raman e regressão PLS.

Figura 33 – Correlação entre os valores previstos pelo modelo PLS-Raman *versus* e valores verdadeiros obtidos pelo método LC.



Fonte: O AUTOR.

Após isso também foi analisado um conjunto de amostras comerciais para tratamento anti-tuberculose. Sua normalidade foi verificada pelo teste de Shapiro-Wilk (0,16 RIF, 0,91 INH, 0,80 PYZ e 0,84 EMB). A partir disso foi realizado um teste-*t* pareado que indicou não haver diferenças estatísticas entre as determinações pela metodologia proposta e pela metodologia 10 por HPLC. Estes resultados são mostrados na tabela a seguir.

Tabela 10 – Predição das concentrações de RIF, INH, PYZ e EMB em comprimidos comerciais.

Amostra	RIF			INH			PYZ			EMH		
	LC	Raman	dif(%)	LC	Raman	dif(%)	LC	Raman	dif(%)	LC	Raman	dif(%)
1	96,14	96,78	-0,66	93,44	90,99	2,62	96,06	98,50	-2,54	93,58	100,13	-7,01
2	100,34	100,60	-0,26	108,05	100,95	6,57	109,07	102,56	5,96	105,30	103,54	1,67
3	100,57	100,68	-0,11	97,30	103,20	-6,07	103,16	102,30	0,84	100,87	103,12	-2,23
4	101,73	106,14	-4,34	107,04	108,11	-1,00	99,98	107,07	-7,09	99,89	109,56	-9,69
5	104,16	95,61	8,21	98,57	94,68	3,94	104,18	95,69	8,15	100,77	96,40	4,33
teste-t		0,78			0,59			0,68			0,39	

Fonte: O AUTOR.

5.3.4 Conclusões

Foi proposto um método baseado para quantificação simultânea de rifampicina, isoniazida, pirazinamida e etambutol por espectroscopia Raman e regressão PLS. Os quatro compostos foram analisados em um curto período de tempo, de aproximadamente 2 minutos. A análise estatística dos dados mostrou que não há diferenças entre as quantificações obtidas pelo modelo PLS e pelo método tradicional por cromatografia a líquido. O método mostrou ser simples, confiável, não destrutivo por não necessitar da dissolução da amostra e ainda é amigável segundo a química verde, por não exigir consumo de reagente, e pode ser utilizado para controle de formulações farmacêuticas que possuem estes quatro fármacos de primeira linha de medicamentos anti-tuberculose.

5.4 ANÁLISE DO PERFIL DE LIPÍDEOS EM AMOSTRAS DE LEITE DE VACA E SORO DE LEITE POR GC E CE COMO METODOLOGIA ALTERNATIVA PARA DETERMINAÇÃO DE FRAUDE

O monitoramento de produtos lácteos que utilizam soro de leite proveniente da fabricação de queijo é um problema muito discutido no Brasil, porque são produtos que podem estar relacionado ao tema de fraude. A fraude por adição de soro de leite em leite fluido é de difícil detecção porque leite e soro de leite tem características físicas e químicas muito semelhantes. No Brasil, os métodos oficiais utilizados para determinar este tipo de fraude considera técnicas analíticas, tais como LC e CE para quantificação da caseinomacropéptido [96], todavia estas análises não podem ser questionáveis visto que outros fatores também podem alterar a concentração de caseinomacropéptido que não estão relacionados à adulteração fraudulenta.

Uma questão que se pode levantar é: existe uma forma de realizar uma correlação da adulteração do leite fluido por adição fraudulenta de soro de leite que não seja pelos métodos já descritos na literatura? Ao tentar responder à esta questão este estudo apresenta um novo método para a quantificação do percentual de soro adicionado ao leite. Tal método se baseia na análise de ácidos graxos para que esta seja utilizada como metodologia alternativa no controle da qualidade do leite.

5.4.1 Objetivos

Avaliar o perfil de AG em amostras de leite e de soro de leite pela técnica de cromatografia a gás e através de ferramentas de análise multivariada buscar um grupo de AG que possam ser utilizados como marcadores para a quantificação do percentual de adição de soro de leite em amostras de leite fluido.

5.4.2 Materiais e métodos

Para este estudo, um conjunto de sete amostras de leite cru foi adquirido a partir de diferentes produtores na região da cidade de Juiz de Fora - MG. Foram realizados testes preliminares para verificação da qualidade do leite. Com base nos resultados de acidez Dornic, pH, crioscopia, níveis de gordura, proteínas e lactose, todas as amostras estavam dentro dos padrões exigidos pela legislação local. Assim, foram utilizados sete amostras de leite cru de vaca para a análise dos AG por GC, além de uma mistura, formada por partes iguais de cada uma das sete amostras iniciais, que será utilizada na segunda etapa do trabalho, para a curva analítica pelo método de CE.

As amostras de soro de leite, de cada amostra de leite adquirida, e da mistura de leite, foram obtidas por coagulação enzimática usando quimosina produzida por *Aspergillus niger var. awamori* (coagulante HA-LA, Chr. Hansen, Valinhos, Brasil). 1 mL de coagulante foi adicionado a 1 L de leite, de acordo com as recomendações do rótulo. O tempo de incubação foi de 40 min a 36°C. Depois a massa foi removida e o soro de leite foi filtrado através de um filtro de papel e recolhido. Por fim, foi realizado um tratamento térmico no soro por 5 min a 80°C antes da utilização. Em suma, as sete amostras de leite cru e a fração de soro de leite de cada uma destas amostras foram analisadas por GC.

As análises pela técnica de cromatografia à gás foram realizadas no cromatógrafo fabricado pela Shimadzu modelo CG-2010 Plus (Shimadzu, Kyoto, Japão) equipado com injetor tipo split-splitless, auto injetor AOC 20-I e detecção por ionização em chama. Este equipamento está instalado na Central Analítica, localizado no departamento de química da Universidade Federal de Juiz de Fora. Medidas realizadas em colaboração com a doutoranda do programa de pós-graduação em Química da UFJF, Brenda Lee Simas Porto.

Após isso a amostra formada pela mistura de leite cru e o seu respectivo soro de leite, foi utilizado para preparar adulterações intencionais, adição de soro de leite em leite, em diferentes percentuais (4,0, 8,0, 12,0, 16,0 e 20,0%) simulando cinco níveis de adulteração analisados em duplicatas autênticas por CE.

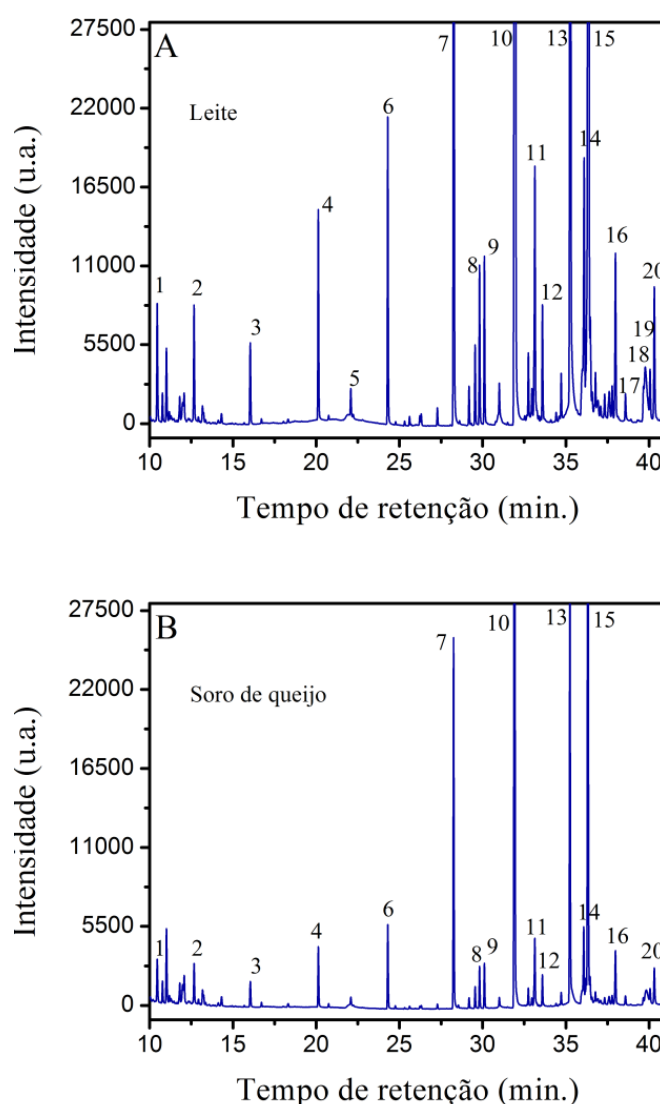
As análises pela técnica de eletroforese capilar foram realizadas num equipamento fabricado pela Agilent Technologies modelo HP3D-CE (Agilent Technologies, Palo Alto, USA) equipado com detector de arranjo de diodos. Este equipamento está instalado na Central Analítica, localizado no departamento de química da Universidade Federal de Juiz de Fora. Medidas realizadas em colaboração com a doutoranda Brenda Lee Simas Porto.

5.4.3 Resultados

A revisão da literatura mostra que a composição lipídica do leite de vaca é de aproximadamente 150 diferentes AG. Destes, cerca de vinte AG majoritários são mostrados

no cromatograma na Figura 34-A e de forma equivalente, também foram encontrados nas amostras de soro de leite 34-B. O soro de leite apresenta uma magnitude inferior para alguns AG porque parte da fração lipídica é retida na massa de queijo, a qual é formada no processo de obtenção do soro de leite. Devido à complexidade dos cromatogramas das amostras de leite e do soro do leite se torna difícil conceber um modelo de previsão de adulteração, baseado na integração de cada AG em cada amostra, por isso serão utilizados métodos de análise multivariada para auxiliar na elucidação de quais AG são mais importantes para serem analisados com maior rigor.

Figura 34 – Perfil de AG de leite de vaca e soro de leite obtidos por GC.



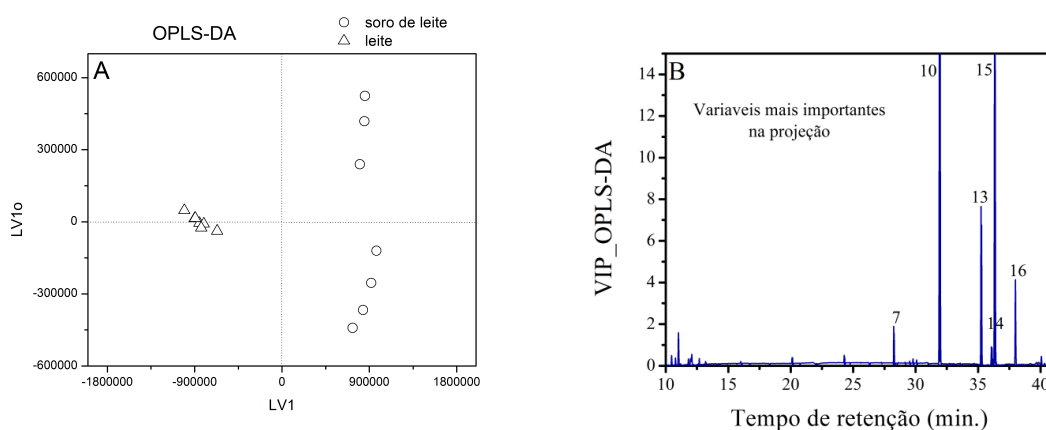
Fonte: O AUTOR.

Um modelo de estatística de OPLS-DA foi utilizado a partir de uma tabela contendo quatorze cromatogramas (sete amostras de leite e sete amostras de soro de leite), as classes

discriminantes foram leite e soro de leite. Não foi utilizada nenhuma transformação dos dados, estes foram apenas centrados na média. O ajuste OPLS-DA obteve coeficientes $R^2X = 0,972$, $R^2Y = 0,990$ e $Q^2 = 0,986$.

O diagrama de *scores*, mostrado na Figura 35-A, apresenta uma clara separação entre os dois grupos de amostras, deixando no lado esquerdo do diagrama apenas amostras de leite e, no lado direito, apenas amostras de soro de leite. A diferença entre cada amostra específica é muito menor que as diferenças entre o grupo de leite e soro de leite como um todo. Esta separação era esperada porque a concentração de AG no soro é muito menor do que no leite, no entanto, a análise OPLS-DA não foi utilizada para mostrar um resultado previsível, mas sim para apontar dentro do universo de AG encontrados no leite, quais são os potenciais marcadores responsáveis para separar estes dois grupos de amostras. Neste ponto, a Figura 35-B, mostra o gráfico das variáveis mais importantes na projeção OPLS-DA. A análise desta figura mostra que entre todos os AG presentes na amostra, apenas seis (C14:0, C16:0, C18:0, C18:1, C18:2 e C18:3) são mais relevantes para separar o grupo de amostras de leite do grupo de soro de leite. Os outros AG têm maior variação entre amostras dentro de cada grupo e/ou estão no nível do ruído, não sendo destacados como bons marcadores para discriminar os grupos analisados.

Figura 35 – Análise multivariada por OPLS-DA a partir dos dados de GC das amostras de leite e de soro de leite.



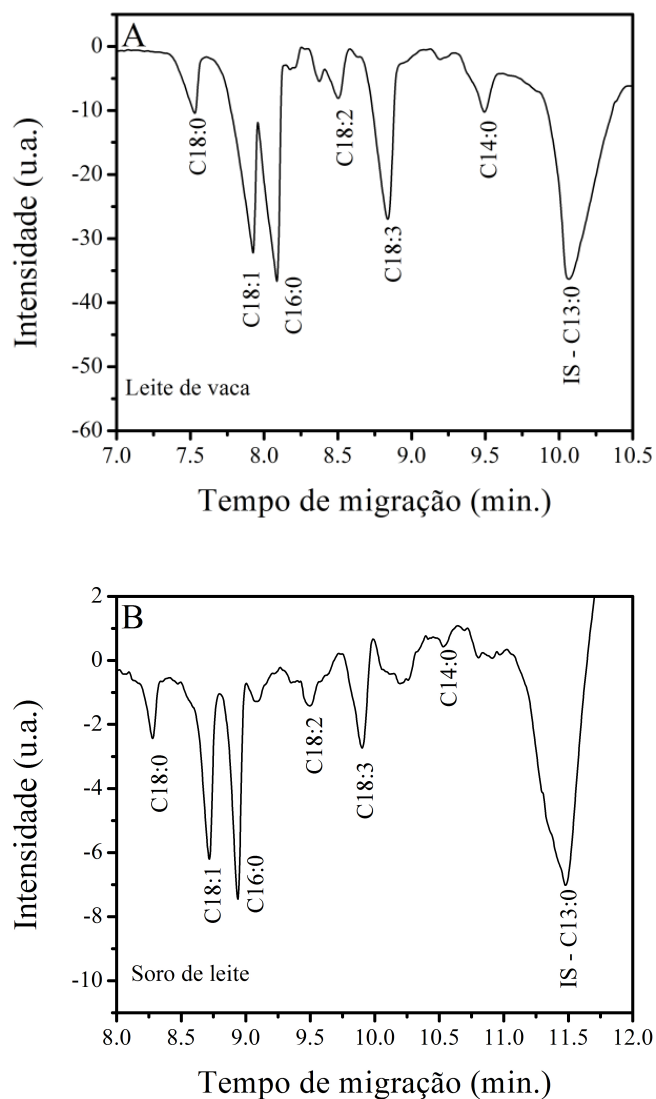
Em "A" está apresentado o diagrama *descotes* para o modelo OPLS-DA e em "B" o gráfico de variáveis importantes na projeção.

Fonte: O AUTOR.

Estes seis AG selecionados pela análise OPLS-DA em GC foram examinados por um método de análise mais rápido por CE, com um novo conjunto de amostras, a fim de verificar se tais AG seriam realmente bons marcadores para compor um modelo de predição do percentual de soro adicionado em leite fluido. A Figura 36 mostra o perfil de AG obtido por CE para amostras de soro e de leite. O método utilizado por CE foi capaz de analisar estes seis AG, previamente selecionados pela técnica de GC, com a diferença

que a detecção ocorre num curto período de tempo, menor de 15 min de corrida, enquanto que para a análise por GC são necessários cerca de 60 min por amostra. Além disso, as análises por CE tem a vantagem de não necessitar da etapa de derivatização da amostra, que é primordial no procedimento de preparo de amostra para análise de AG por GC.

Figura 36 – Análise das amostras de leite e soro de leite por CE.



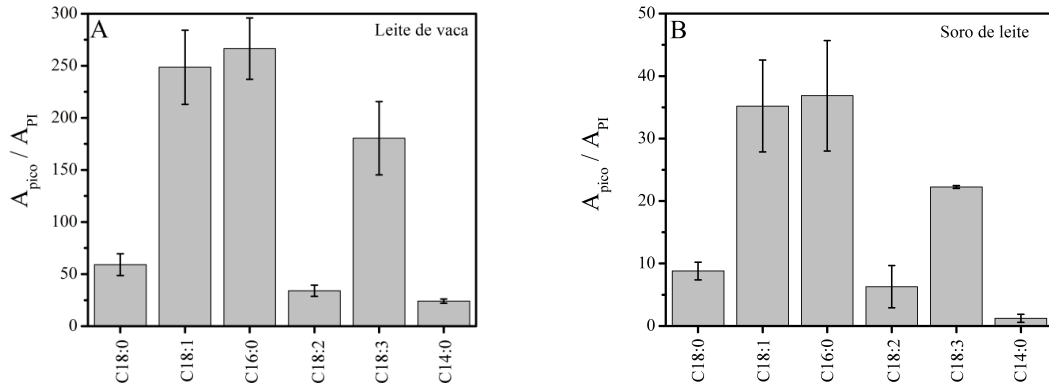
Em destaque estão rotulados os ácidos graxos selecionados como mais significativos para discriminar o soro e leite, obtidos pela análise multivariada dos resultados de GC. Em "A" é apresentado um eletroferograma de uma amostra de leite e em "B" de soro de leite.

Fonte: O AUTOR.

Na análise por CE, a área de cada pico de AG (C14:0, C16:0, C18:0, C18:1, C18:2 e C18:3) foi normalizada pela área do padrão interno (PI-C13:0). A Figura 37 mostra um exemplo de uma análise em duplicata para uma amostra de leite e uma de soro de leite. Este gráfico de barras fornece uma boa visão da distribuição dos seis AG monitorados. Leite e soro tem um perfil de AG muito semelhante, no entanto, a magnitude destas

concentrações é muito mais elevada em amostras de leite, ver eixo-y, e este fato pode ser usado para ajustar uma regressão capaz de prever a mistura destas amostras.

Figura 37 – Quantificação de AG por CE para leite e soro de leite.

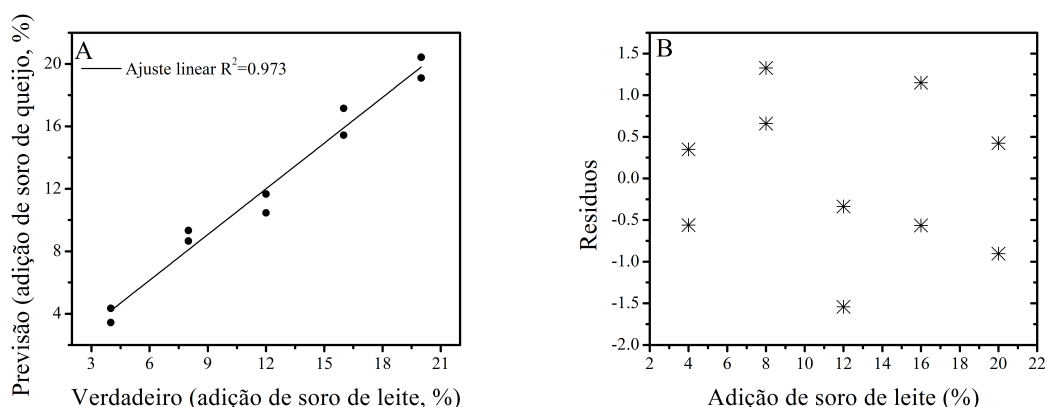


Os gráficos de barras representam a razão entre a área do pico de cada AG e da área do pico do PI de um eletroferograma. A barra de erro é o desvio padrão de análises em duplicatas.

Fonte: O AUTOR.

Uma tabela de dados foi preparada a partir da normalização da área de cada um dos seis AG monitorados. Esta tabela contém cinco níveis de adulteração intencional de leite por adição de soro de leite, todos analisados em duplicata, num total de dez pontos experimentais. Este conjunto de medidas foi utilizado para ajustar uma regressão MLR. A Figura 38-A mostra, no eixo horizontal, os valores reais da adulteração preparados em laboratório e, o eixo vertical representa os valores previstos pelo modelo MLR. Os valores obtidos pelo modelo MLR apresentaram excelente correlação com os valores reais, $R^2 = 0,973$, e o gráfico dos resíduos, mostrado na Figura 38-B, não apresenta tendência, isto é, indica que o modelo está bem ajustado.

Figura 38 – Resultados para a quantificação do percentual de soro de leite adicionado ao leite fluido, por CE e regressão MLR.



Em "A" está o gráfico de correlação entre os valores previstos pelo modelo MLR *versus* os valores verdadeiros de percentual de fraude preparados em laboratório. "B" é o gráfico dos resíduos para o mesmo conjunto de dados.

Fonte: O AUTOR.

A fim de complementar as informações do gráfico de resíduos, foi realizado o teste de Durbin-Watson (p -valor igual 0,896) após ser verificado o pressuposto da normalidade dos dados pelo teste de Shapiro-Wilk (p -valor igual 0,808). Como resultado têm-se que, não existem evidências, ao nível de 95% de confiança, de correlação na distribuição dos resíduos entre os valores previstos pelo modelo MLR e os valores verdadeiros, que reitera que o modelo está bem ajustado para tal predição.

5.4.4 Conclusões

O monitoramento da adulteração no leite fluido por adição de soro de leite foi investigado pela análise do perfil de AG das amostras. A análise OPLS-DA sobre o conjunto de cromatogramas de amostras de leite e soro revelou que os AG C14:0, C16:0, C18:0, C18:1, C18:2 e C18:3 são os mais significativos para serem monitorados sob a suspeita de tal adulteração.

Foi elaborado um ajuste de regressão múltipla com estes seis AG e o percentual de adição de soro foi quantificado com sucesso numa faixa de 4 a 20% de fraude, sugerindo assim um novo método que poderá ser utilizado no controle da qualidade do leite. É importante mencionar que a metodologia tradicional de detecção deste tipo de fraude, pela quantificação do CMP com a técnica de LC, apenas a porção de soro doce é monitorada. Diferentemente disso, a metodologia aqui descrita independe de qual dos três tipos de soro foi utilizado, soro doce, soro com acidez média e soro ácido. Entretanto, em níveis industriais, a porção lipídica presente no soro é mínima, de modo que, para o monitoramento deste tipo de fraude, pela avaliação do perfil de AG, são necessários métodos com reduzidos níveis de quantificação.

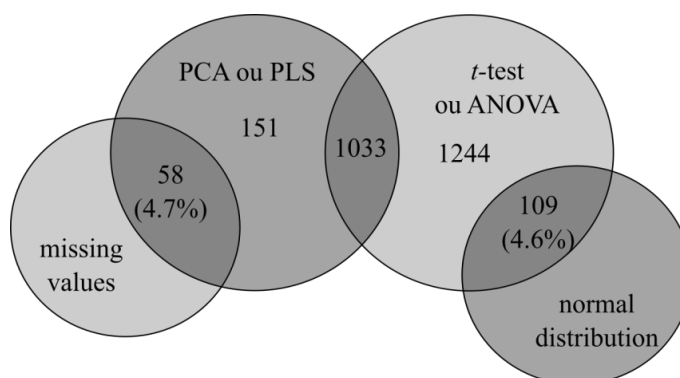
5.5 TUTORIAL PARA ANÁLISE DE DADOS DE ESTUDOS DE METABOLÔMICA

Experimentos de metabolômica visam identificar um conjunto de metabólitos produzidos ou modificados por um organismo, e seu objetivo é encontrar novos biomarcadores que possam ser usados para diagnosticar doenças, efeitos de dieta ou ação de drogas, entre outros. Devido à complexidade das amostras biológicas e a alta sensibilidade de técnicas de separação modernas utilizadas nestes estudos, tal como a CE, GC e LC, todos acoplados a detectores de espectrometria de massas (MS), fazem com que experimentos de metabolômica possam gerar tabelas de dados tipicamente da ordem de centenas a milhares de variáveis em uma única amostra [97–99].

Neste sentido a abordagem por análise multivariada de dados tem sido amplamente utilizada nesta área de pesquisa, permitindo facilitar a visualização dos dados e amostras e a obtenção de novos biomarcadores [100]. A fase de análise de dados é muito importante neste campo, e o procedimento utilizado neste passo tem um impacto direto na seleção dos candidatos a biomarcadores [101–105].

Uma pesquisa simples na literatura mostra que nem todas as publicações mencionam completamente as etapas utilizadas para a obtenção de novos resultados, de modo que pode ser questionado se os pressupostos são realmente verificados durante as análises de dados e se todos têm conhecimento de limitações originadas nas técnicas experimentais. A Figura 39 mostra este panorama. Ela apresenta o resultado de uma pesquisa em periódicos, na última década, na qual foram buscados alguns pressupostos necessários para análise de dados univariados, bem como a questão do método de substituição de valores em falta, típicos de espectros de massas desta área de pesquisa. As interseções indicam a fração, do percentual total de publicações, que mencionam na redação dos artigos os pressupostos para as respectivas abordagens.

Figura 39 – Análise exploratória de publicações pelo diagrama de Venn.



Pesquisa por palavras-chave "metabolomic" e "mass spectrometry", e palavras adicionais indicadas no diagrama.

Fonte: <http://www.sciencedirect.com/>. Acessado em janeiro de 2015.

Esta introdução mostra como uma discussão em torno da temática análise de dados

na área de metabolômica pode ser muito útil para contribuir para novas pesquisas além de informações mínimas que devem ser reportadas, de acordo com a MIAMET [106].

5.5.1 Objetivos

Apresentar um tutorial completo de análise de dados univariados e multivariados em um *software* livre (*R-software*) [107] para ser utilizado desde o momento posterior as medições experimentais até a revelação de candidatos a biomarcadores. Além disso, fomentar uma discussão em torno da etapa de análise de dados em experimentos de metabolômica.

5.5.2 Materiais e métodos

O conjunto de dados utilizado para a discussão estatística foi obtido a partir de uma experiência controlada que consiste em trinta amostras diferentes, formadas a partir de uma mistura de amostras de urina fortificada com vinte soluções individuais de padrões de metabolitos (L-serina, Creatinina, Betaína, L-valina, L-treonina, Creatina, L-Leucina, L-asparagina, Ornitina, L-histidina, L-glutamina, L-lisina, Ácido L-glutâmico, L-fenilalanina, 3-metil-histidina, L-tirosina, Dimetil-L-arginina, L-triptofano, L-carnosina e Cistina). Tais metabolitos foram adicionados em diferentes níveis de concentrações, simulando a variabilidade encontrada numa amostra real. Este conjunto de amostras foi analisado por CE-MS.

As análises por eletroforese capilar foram realizadas num equipamento fabricado pela Agilent modelo CE-7100 acoplado a uma fonte de ionização por *electrospray* e detector de massas por tempo de voo modelo Agilent 6224 (Agilent Technologies, Wilmington, EUA). Este equipamento está instalado no *Centro de Excelencia en Metabolómica y Bioanálisis*, localizado na *Facultad de Farmacia da Universidad CEU San Pablo, Madrid España*. As medidas foram realizadas durante o programa de mobilidade acadêmica promovido pela fundação EADS-CASA em colaboração com grupo de pesquisa liderado pela professora doutora Coral Barbas Arribas, ao longo do ano de 2012.

Entre os vinte metabolitos adicionados, oito foram selecionados para apresentar maior variabilidade entre as amostras (L-valina, L-treonina, L-asparagina, L-lisina, L-fenilalanina, 3-metil-histidina, Dimetil-L-arginina e L-Cistina), dividindo desta forma o conjunto de amostras em dois grupos de quinze indivíduos cada. Estes dois grupos, aqui chamados de grupo "A" e "B", simulam um experimento típico de metabolômica, como por exemplo um grupo "caso" e outro "controle", em que a questão a ser respondida é: qual(is) metabolitos podem ser usados para classificar os dois grupos, ou seja, os biomarcadores. Neste estudo, é conhecido com exatidão os metabolitos que fazem a diferença entre os dois grupos. Então, espera-se que estes metabolitos apareçam como candidatos a biomarcadores no final da análise estatística dos dados.

Além das trinta amostras foram incluídas cinco amostras de controle de qualidade (QC). Estes são usados para verificar a qualidade e a estabilidade do sistema analítico do experimento. Estes QC foram preparados a partir de uma mistura em partes iguais [108] de todas as trinta amostras do experimento. Por conseguinte, no total, a experiência incluiu trinta e cinco amostras: quinze amostras pertencentes ao grupo A, quinze amostras pertencentes ao grupo B e cinco QC.

Não é objetivo deste trabalho descrever o método analítico completo, a preparação da amostra e condições experimentais, uma vez que a abordagem aqui apresentada é puramente estatística. No entanto, o método de análise química pode ser encontrado na literatura [109–112]. Assim, esta experiência controlada gerou uma tabela de dados que vai ser utilizada como *input* para as discussões sobre análise de dados. No entanto, é crucial usar um experimento controlado quando quer testar a confiabilidade da análise estatística utilizada.

5.5.3 Resultados e discussões

A análise estatística é iniciada a partir de uma tabela de dados única, que combina as abundâncias de todos os íons detectados e alinhados com o conjunto de amostras. Para iniciar o processamento dos dados, sua entrada correta é crucial. Os dados devem ser inseridos no *software-R* de acordo com a Figura 40. As colunas devem conter as variáveis, íons ou massas detectadas, e as linhas devem conter as amostras.

Figura 40 – Forma da entrada da tabela de dados no *software R*.

		Variáveis			
		variavel 1 ↓	variavel 2 ↓	variavel 3 ... ↓	
Amostras	Grupo A	amostra A1 →	373.8027@7.2393	289.8401@7.2444506	205.879@7.2514496
		amostra A2 →	79743	361199	317461
		amostra A3 →	74980	362215	1
		
	Grupo B	amostra B1 →	83330	357576	308682
		amostra B2 →	70788	346862	314067
		amostra B3 →	1	345502	315541
		
	Amostras de Controle de Qualidade	qc1 →	71677	355254	320845
		qc2 →	68325	334659	318603
		qc3 →	1	312608	303652
		
			64551	316918	300870
		

Fonte: O AUTOR.

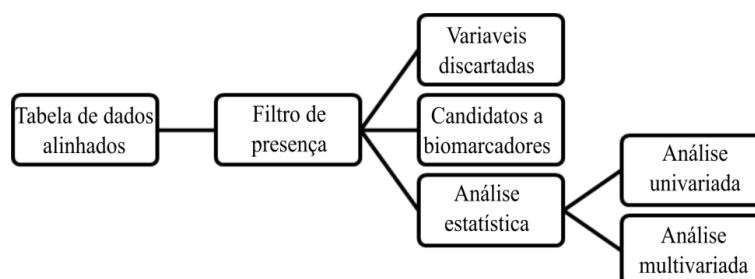
Após a tabela de dados ter sido organizada corretamente, deve ser informado ao programa o número de amostras em cada grupo de estudo, bem como o caractere utilizado para representar os valores em falta (*missing values*). Por exemplo, o *software Mass Profiler Professional* utilizado em equipamentos fabricados pela Agilent (como o utilizado neste estudo) exporta a matriz de dados alinhados com os valores em falta identificados por "1". Estes valores em falta devem ser substituídos por "NA" (não disponível). NA é

normalmente declarado com antecedência para operações aritméticas no *R-Software*. Não fazer essa substituição leva a uma análise equivocada dos dados. O código usado para inserir a tabela de dados, e fazer esta substituição é:

```
#abrindo a tabela de dados
alignment=read.table(file="alignment.txt", sep="\ t",header=T, dec=".")
dim(alignment)
groupA=15
groupB=15
qc=5
#identificando os valores em falta por NA
for (j in 1:(dim(alignment)[2])) {
for (i in 1:(dim(alignment)[1])) {
if(alignment[i,j]==1)
alignment[i,j]=NA }}
```

A Figura 41 mostra os passos necessários para se obter os dados para a análise estatística após o alinhamento dos mesmos. Estas etapas serão discutidas na sequência.

Figura 41 – Fluxograma para obtenção da tabela que será utilizada na análise estatística dos dados.



Fonte: O AUTOR.

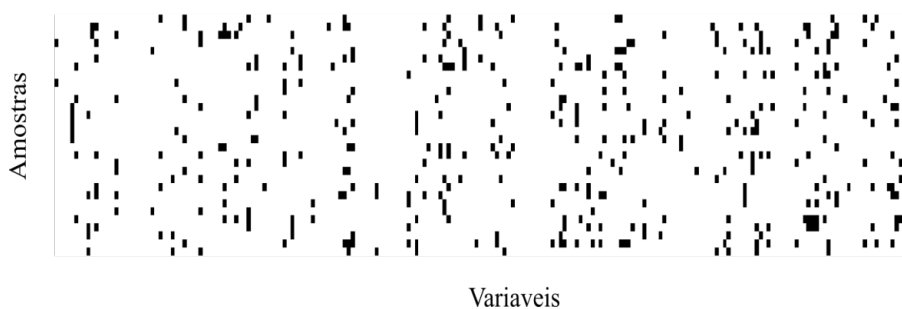
O efeito de sinais aleatórios pode ser minimizado através da utilização de um filtro que inclui apenas variáveis presentes em uma quantidade mínima de amostras em cada um dos grupos. Filtros com um percentual muito baixo inclui variáveis que não possuem características de grupo, gerando uma matriz com um número excessivo de valores em falta. Por outro lado, filtros com alto percentual pode gerar a perda de variáveis importantes, que poderiam contribuir para o estudo e para a interpretação da rota metabólica. Após a aplicação do filtro de presença muitas variáveis são descartadas. Na sequência é mostrado o código para utilização do filtro de presença de 75%, para reduzir o número excessivo de variáveis com valores em falta. Após este filtro a matriz de dados ficou, neste caso, com um total de 332 variáveis e 16,98% de valores em falta.

```
#filtro de presença de 75% em 1 de 2 grupos
filterreddata=alignment
for (j in 1:(dim(alignment)[2])) {
for (k in 1:(dim(filterreddata)[2])) {
ok1=sum(is.na(filterreddata[1:(groupA),k]))
ok2=sum(is.na(filterreddata[((groupA)+1):(( groupA)+( groupB)),k]))
if (ok1>round(0.25*( groupA)) & ok2>round(0.25*( groupB)) )
filterreddata=filterreddata[,-k]}
dim(filterreddata)
```

Um número excessivo de valores em falta também pode indicar a necessidade de rever os métodos utilizados para extrair os íons ou o método de alinhamento de dados. A inspeção da distribuição de valores em falta também pode ser útil para verificar a qualidade das medidas experimentais assim como o nível do filtro de presença utilizado. O código usado para gerar um mapa de valores perdidos é descrito abaixo e a Figura 42 representa a saída do programa R.

```
#dividindo a tabela de dados em amostras e QCs
data=filterreddata[1:( groupA + groupB),]
dataqc=filterreddata[(groupA + groupB+1):( groupA + groupB+qc),]
#para gerar o mapa de valores perdidos
library(Amelia)
missmap(data, legend = FALSE, col = c("black","white"),
main = "Missing Value Map",
y.cex = 0.8, x.cex = 0, x.labels = NULL, csvar = NULL, tsvar =
NULL, rank.order = FALSE)
```

Figura 42 – Mapa de valores perdidos.



Pontos em preto representam os valores perdidos. Esta representação contém 30 amostras (linhas horizontais) por 332 variáveis (colunas verticais), com um total de 16,98% de valores perdidos.

Fonte: O AUTOR.

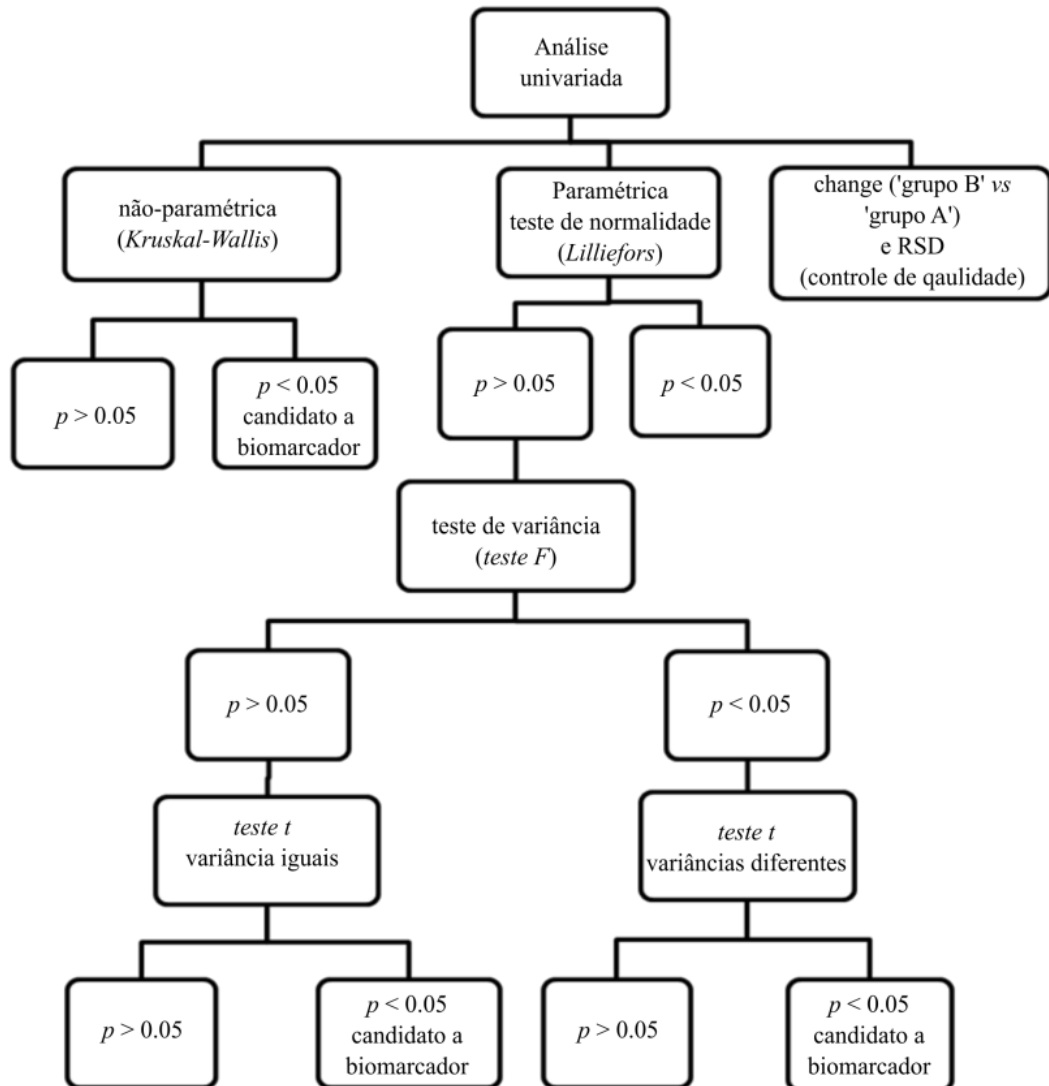
Neste mapa cada ponto preto representa um valor em falta. Assim, uma linha preta (vertical ou horizontal) é uma sequência de valores em falta, e por inspeção deste mapa é possível identificar *outliers* e candidatos a biomarcadores. A fim de não desvirtuar as conclusões do estudo, uma distribuição aleatória de valores em falta é desejável. A linha preta vertical é uma variável com valores em falta em excesso. Se estes valores em falta, variável por variável, concentram-se em apenas um dos grupos analisados, isso significa que esta é uma variável (metabolito), presente em apenas um grupo, o qual pode então ser caracterizado como um candidato a biomarcador.

Por outro lado, uma linha preta horizontal sobre o mapa de valores em falta pode ser interpretada como uma amostra com valores em falta em excesso. Essa amostra deve ser inspecionada e analisada novamente, se possível. Se não for possível, deve ser removida da tabela de dados para evitar que afete indevidamente as conclusões sobre o grupo de amostras.

Dando sequência ao diagrama mostrado na Figura 41, a análise univariada é realizada através da aplicação de um conjunto de testes de hipóteses, somente depois de obtida a tabela de dados filtrados. Os testes de hipóteses consistem em um procedimento estatístico com base na teoria de probabilidade, no qual um parâmetro é testado em um conjunto de valores. Duas hipóteses são consideradas: a hipótese nula H_0 (a ser testada); e a hipótese alternativa H_a que será aceita se o grupo não se insere nesta hipótese testada. A decisão se o conjunto teste pertence a H_0 ou H_a pode ser feita através da análise dos p -valores.

É importante observar que, para a correta aplicação de um teste de hipóteses, a verificação dos pressupostos deve ser realizada. Se os pressupostos não são verificados isso pode resultar em uma inferência incorreta sobre o conjunto de dados analisados. Para a análise univariada existe uma grande diferença entre a sequência de análise a ser feita, se os dados seguem uma distribuição normal (estatística paramétrica), ou não (estatística não-paramétrica). A cascata de processos de tomada de decisão em análise univariada é apresentada na Figura 43.

Figura 43 – Diagrama de tomada de decisão para análise estatística univariada.



Fonte: O AUTOR.

Para uma aplicação de rotina como a sugerida acima, o número de observações deve ser superior a quatro por grupo. Neste esquema, foi utilizado o teste de Lilliefors para uma verificação de normalidade [113]. Neste caso, a hipótese nula padrão, é que a amostra vem de uma distribuição normal, contra a hipótese alternativa de que a amostra não provém de uma distribuição normal.

O teste não-paramétrico utilizado foi o teste de Kruskal-Wallis [114]. Para este teste a função compara as medianas das amostras e retorna o p -valor para a hipótese nula de que todas as amostras provém da mesma população. A alternativa é que eles diferem por pelo menos um. O teste de Kruskal-Wallis, é uma versão não-paramétrica da clássica ANOVA, e pode ser utilizado para dois ou mais grupos.

O teste de verificação da homogeneidade de variâncias foi o teste-F [23]. Tal teste

verifica a hipótese de que duas amostras independentes, dos grupos A e B, são provenientes de distribuições normais com a mesma variância, contra a alternativa, que eles vêm de distribuições normais com variâncias diferentes. Depois disso, o teste-*t* [23] foi usado para verificar se a média das abundâncias de uma amostra do grupo A é diferente do grupo B. A hipótese nula do teste-*t* verifica se os dados dos grupos A e B são amostras aleatórias independentes, provenientes de distribuições normais, com médias iguais, contra a alternativa, de que as médias são diferentes.

Finalmente, o desvio padrão relativo (RSD) [115] foi calculado para cada variável de cada amostra do grupo de QC, para verificar a estabilidade na medição de uma variável individualmente. Também é calculado o *change*, que representa o quanto a média de abundância de uma variável de um grupo é maior ou menor que do outro grupo.

Esta sequência completa de verificação de testes de hipótese, cálculos de RSD e *change* apresentada na Figura 43 pode ser implementada em uma rotina R da seguinte forma:

```
#análise univariada
class=c(rep(1, groupA),rep(2, groupB))
class=factor(class)
variables=dim(data)[2]
compounds=c(data.frame(colnames(filterreddata)))
#calculo dos p-valores para o teste de normalidade de Lilliefors
pvalLillieforsgroup1=rep(0,variables)
pvalLillieforsgroup2=rep(0,variables)
group1=data[class==1,]
group2=data[class==2,]
library(nortest)
for (i in 1:variables)
{lillie1=lillie.test(group1[,i])
lillie2=lillie.test(group2[,i])
pvalLillieforsgroup1[i]=lillie1$p.value
pvalLillieforsgroup2[i]=lillie2$p.value}
#calculo dos p-valores para o teste de homogeneidade de variancias
#pelo teste-F
pvalF=rep(0,variables)
library(stats)
for (i in 1:variables)
{testf=var.test(data[,i]~class)
pvalF[i]=testf$p.value}
#calculo dos p-valores para o teste-t assumindo variancias diferentes
```

```

pvalTstudvardif=rep(0,variables)
library(stats)
for (i in 1:variables)
{tstudentvardif=t.test(data[,i]~class)
pvalTstudvardif[i]=tstudentvardif$p.value}
#calculo dos p-valores para o teste-t assumindo variancias iguais
pvalTstudvarequal=rep(0,variables)
library(stats)
for (i in 1:variables)
{tstudentvarequal=t.test(data[,i]~class,var.equal=TRUE)
pvalTstudvarequal[i]=tstudentvarequal$p.value}
#calculo dos p-valores para o teste-t não-paramétrico de Kruskal-Wallis
pvalkruskal=rep(0,variables)
library(stats)
for (i in 1:variables)
{kruskalwallis=kruskal.test(data[,i]~class)
pvalkruskal[i]=kruskalwallis$p.value}
#calculo do RSD no grupo QC e do change entre os grupos A e B
rsd=rep(0,variables)
change=rep(0,variables)
library(stats)
for (i in 1:variables)
{rsd[i]=100*sd(dataqc[,i], na.rm = TRUE)/mean(dataqc[,i],
na.rm = TRUE)
change[i]=100*(mean(group2[,i], na.rm = TRUE)/mean(group1[,i],
na.rm = TRUE)-1)}
#para organizar os resultados em uma tabela unica
pvalues=data.frame(compounds,pvalLillieforsgroup1,pvalLillieforsgroup2,
pvalF,pvalTstudvardif,pvalTstudvarequal,pvalkruskal,rsd,change)
#para salvar esta tabela
write.table(pvalues, "univariatestatistics.csv", sep="," , dec=".")

```

Ao final têm-se os resultados completos da análise univariada resumidos em uma tabela, como a Tabela 11. Esta tabela contem os p -valores de todos os testes de hipóteses, para cada uma das variáveis da matriz de dados filtrada. Outros testes de hipóteses podem ser implementados facilmente, com base no modelo sugerido aqui. A tomada de decisão pode ser efetuada analisando o p -valor de cada teste de hipótese. Em cada um destes ensaios, a H_0 é aceita quando tomamos a porção de p -valores $> 0,05$, e, por conseguinte, assumi-se a H_a quando se toma a porção de p -valores $< 0,05$. Todos os testes foram

realizados ao nível de 95% de confiança. No entanto, tanto o p -valor crítico, como o intervalo de confiança, podem ser modificados, dependendo do estudo. Por exemplo, intervalos de confiança de 90% ou 99% podem ser usados, ou a correção de Bonferroni pode ser usada para os p -valores [116,117]. Não é nosso objetivo discutir completamente cada um desses testes, mas sim apresentar a sua essência no processo de premissas que devem ser consideradas, e enfatizar que eles devem ser mencionados, quando utilizados.

Tabela 11 – Resumo de resultados da análise univariada. *R-software*.

Metabolito	<i>p</i> -valor normalidade		<i>p</i> -valor grupo B (Lilliefors)	<i>p</i> -valor variância (teste- <i>F</i>)	<i>p</i> -valor (teste- <i>t</i>) variância diferente	<i>p</i> -valor (teste- <i>t</i>) variância igual	<i>p</i> -valor não-paramétrico (Kruskal-Wallis)	RSD (%)	change (%)
	grupo A (Lilliefors)	grupo B (Lilliefors)							
Cistina	8.08 10 ⁻²	1.72 10 ⁻¹	1.90 10 ⁻²²	6.65 10 ⁻¹²	1.63 10 ⁻¹⁸	3.07 10 ⁻⁶	4.40	-97.74	
Fenilalanina	5.12 10 ⁻¹	5.23 10 ⁻²	3.73 10 ⁻¹¹	7.37 10 ⁻¹²	3.90 10 ⁻¹⁸	3.07 10 ⁻⁶	6.61	274.50	
Treonina	4.85 10 ⁻¹	1.23 10 ⁻²	4.79 10 ⁻⁷	2.81 10 ⁻⁷	1.94 10 ⁻⁹	3.07 10 ⁻⁶	2.23	-55.06	
Arginina	7.76 10 ⁻²	1.18 10 ⁻²	2.11 10 ⁻⁷	8.83 10 ⁻⁶	2.51 10 ⁻⁸	1.56 10 ⁻⁵	3.72	-57.71	
Valina	5.19 10 ⁻¹	2.72 10 ⁻¹	3.08 10 ⁻¹	2.96 10 ⁻⁷	2.03 10 ⁻⁷	1.76 10 ⁻⁵	5.05	47.35	
Lisina	2.81 10 ⁻¹	3.82 10 ⁻²	8.48 10 ⁻¹	1.24 10 ⁻⁴	5.72 10 ⁻⁵	2.13 10 ⁻⁴	11.57	-19.85	
Histidina	4.50 10 ⁻¹	4.68 10 ⁻¹	6.89 10 ⁻¹	3.39 10 ⁻²	3.38 10 ⁻²	4.01 10 ⁻²	8.85	10.01	
Asparagina	3.49 10 ⁻¹	3.96 10 ⁻⁵	4.20 10 ⁻⁴	2.68 10 ⁻¹	2.17 10 ⁻¹	4.54 10 ⁻²	4.02	25.65	

Fonte: O AUTOR.

É importante deixar claro que todos os testes são realizados em todas as variáveis, e para que uma variável possa ser considerada como candidata a biomarcador, a tomada de decisão é a função do analista, com base nos resultados dos testes apresentados. O processo de seleção de biomarcadores não é automatizado para evitar a perda de variáveis caracterizadas como um "falso negativo". A tabela completa com os resultados dos testes univariados, RSD e *change* é constantemente consultada durante o processo de interpretação biológica, por isso, deve conter todas as variáveis, para ajudar ao analista a considerar a contribuição de cada variável como mais ou menos importante.

A mesma sequência de análise de dados univariados foi realizada no *software* comercial Matlab. A Tabela 12 apresenta os resultados para efeitos de comparação com os obtidos pelo *software R*.

Tabela 12 – Resumo dos resultados da análise univariada no *software* Matlab.

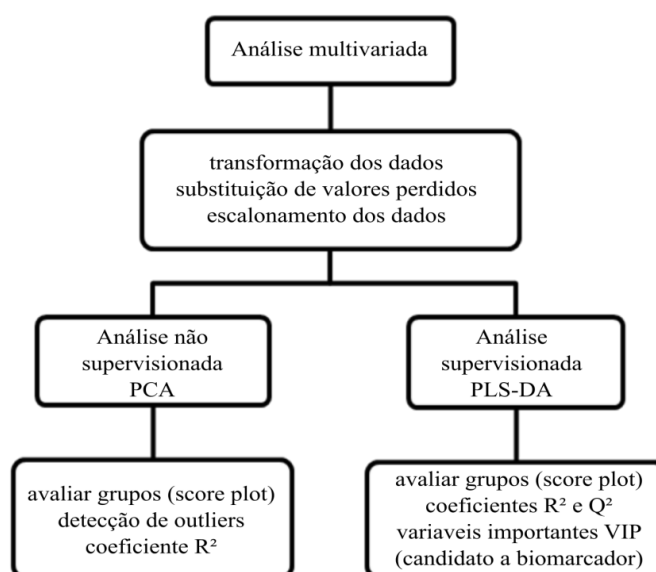
Metabolito	<i>p</i> -valor normalidade		<i>p</i> -valor normalidade	<i>p</i> -valor		<i>p</i> -valor	<i>p</i> -valor	<i>p</i> -valor
	grupo A (Lilliefors)	grupo B (Lilliefors)	grupo B (Lilliefors)	variância (teste- <i>F</i>)	(teste- <i>t</i>) variância diferente	(teste- <i>t</i>) variância igual	(teste- <i>t</i>) variância igual	não-paramétrico (Kruskal-Wallis)
Cistina	7.80 10 ⁻²	1.67 10 ⁻¹	1.67 10 ⁻¹	1.90 10 ⁻²²	6.65 10 ⁻¹²	1.63 10 ⁻¹⁸	1.63 10 ⁻¹⁸	3.07 10 ⁻⁶
Fenilalanina	5.00 10 ⁻¹	5.08 10 ⁻²	5.08 10 ⁻²	3.73 10 ⁻¹¹	7.37 10 ⁻¹²	3.90 10 ⁻¹⁸	3.90 10 ⁻¹⁸	3.07 10 ⁻⁶
Treonina	4.73 10 ⁻¹	1.20 10 ⁻²	1.20 10 ⁻²	4.79 10 ⁻⁷	2.81 10 ⁻⁷	1.94 10 ⁻⁹	1.94 10 ⁻⁹	3.07 10 ⁻⁶
Arginina	7.54 10 ⁻²	1.18 10 ⁻²	1.18 10 ⁻²	2.11 10 ⁻⁷	8.83 10 ⁻⁶	2.51 10 ⁻⁸	2.51 10 ⁻⁸	1.56 10 ⁻⁵
Valina	5.00 10 ⁻¹	2.64 10 ⁻¹	2.64 10 ⁻¹	3.08 10 ⁻¹	2.96 10 ⁻⁷	2.03 10 ⁻⁷	2.03 10 ⁻⁷	1.76 10 ⁻⁵
Lisina	2.82 10 ⁻¹	3.77 10 ⁻²	3.77 10 ⁻²	8.48 10 ⁻¹	1.24 10 ⁻⁴	5.72 10 ⁻⁵	5.72 10 ⁻⁵	2.13 10 ⁻⁴
Histidina	4.37 10 ⁻¹	4.56 10 ⁻¹	4.56 10 ⁻¹	6.89 10 ⁻¹	3.39 10 ⁻²	3.38 10 ⁻²	3.38 10 ⁻²	4.01 10 ⁻²
Asparagina	3.38 10 ⁻¹	1.00 10 ⁻³	1.00 10 ⁻³	4.20 10 ⁻⁴	2.68 10 ⁻¹	2.17 10 ⁻¹	2.17 10 ⁻¹	4.54 10 ⁻²

Fonte: O AUTOR.

As análises multivariadas aqui descritas são baseadas em métodos de redução de dimensão, como PCA e PLS-DA, que são amplamente utilizados em estudos metabolômica [100]. Estes métodos interpretam a tabela de dados como uma matriz de dados. Neste

caso, é necessário usar uma forma de substituição de valores em falta, porque, por métodos de redução de dimensões, é necessária uma matriz com observações completas. Este é um requisito de álgebra de matrizes, que deve ser realizado antes de se iniciar as análises. O diagrama da Figura 44 mostra, de uma forma simplificada, o processo de análise de dados multivariados. Os estágios de transformação, escalonamento e substituição de valores em falta podem ser feitos com o código a seguir, aplicado sobre a tabela de dados filtrados obtida anteriormente.

Figura 44 – Diagrama para análise de dados por estatística multivariada.



Fonte: O AUTOR.

```

#análise multivariada
#transformação logarítmica dos dados
mult.data=log10(filtereddata)
#algoritmo NIPALS para substituição de valores em falta
library(mixOmics)
nipals.datos = nipals(mult.data, reconst = TRUE)$rec
id.na = is.na(mult.data)
nipals.datos[!id.na] = mult.data[!id.na]
nipals.datos=data.frame(nipals.datos)
#escalonamento dos dados do tipo Pareto-scaling
library(BioMark)
X=nipals.datos
FUN=scalefun(sc.p = "pareto")
matrix=FUN(X)
  
```

Neste código foi utilizada a transformação logarítmica dos dados. A transformação

log é um tratamento comum, uma vez que está relacionada com a instrumentação utilizada em detectores de massa e com a natureza biológica dos dados. O escalonamento do tipo Pareto [2] está entre sem escalonamento e escalonamento do tipo variância única (UV), e dá à variável uma variação igual ao seu desvio-padrão. NIPALS foi utilizado para substituir os valores em falta, este método substitui os valores em falta através de um método chamado de projeção de componente único para variáveis com menos de 20% de valores em falta, e pela média condicional se o percentual de valores em falta é maior [118]. Todo este pré-tratamento aqui descrito tem um forte impacto sobre os modelos multivariados que são subsequentemente gerados, ou seja, influenciam no conjunto de metabolitos classificados como candidatos a biomarcadores.

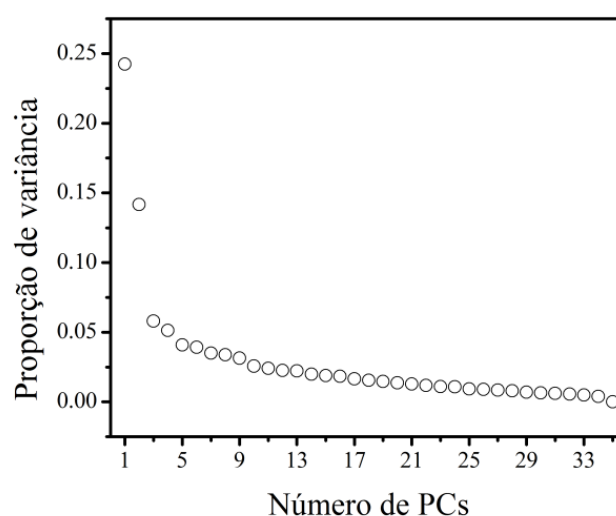
Agora, a matriz de dados está pronta para análise multivariada. Uma análise de PCA pode ser feita a partir do seguinte código R:

```
#modelo PCA
pca.clase=c(rep(1, groupA),rep(2, groupB),rep(3,qc))
library(stats)
pcaanalysis=prcomp(matrix, scale. = FALSE, center = FALSE)
```

Para gerar um gráfico da proporção de variância explicada pelo modelo PCA, como a Figura 45. Têm-se o seguinte código:

```
#gráfico da proporção de variância por PC
bmp(file = "PCAvariancePCs.bmp")
plot(summary(pcaanalysis)$importance[2,], ylab = "Proportion of variance",
xlab = "Number of PCs")
title(main = list("Proporcion of variance - PCA model", cex=1, col="black",
font=1))
dev.off()
```

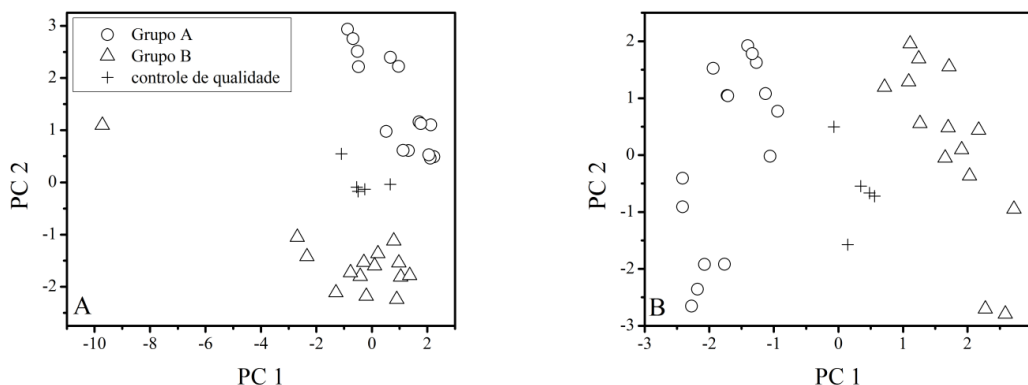
Figura 45 – Proporção de variância explicada pelo modelo PCA em função do número de componentes principais.



Fonte: O AUTOR.

Para gerar um gráfico de *scores* do modelo PCA, como o da Figura 46, têm-se o seguinte código:

```
#Gáficos de scores da PCA
bmp(file = "PCA.bmp")
plot(pcaanalysis$x, xlab = paste("PC 1 (Proportion of Variance R2 ",
round(100*summary(pcaanalysis)$importance[2,1], dig = 2), "%)", sep = "")
, ylab = paste("PC 2 (Proportion of Variance R2 ",
round(100*summary(pcaanalysis)$importance[2,2], dig = 2), "%)", sep = "")
, pch = pca.clase)
title(main = list("PCA", cex=1, col="black", font=1))
dev.off()
#resumo de coeficientes PCA
summary(pcaanalysis)
```

Figura 46 – Análise do diagrama de *scores* para o modelo de PCA.

Em "A" estão representadas todas as 30 amostras do desenho experimental e em "B" a amostra número 29 foi removida.

Fonte: O AUTOR.

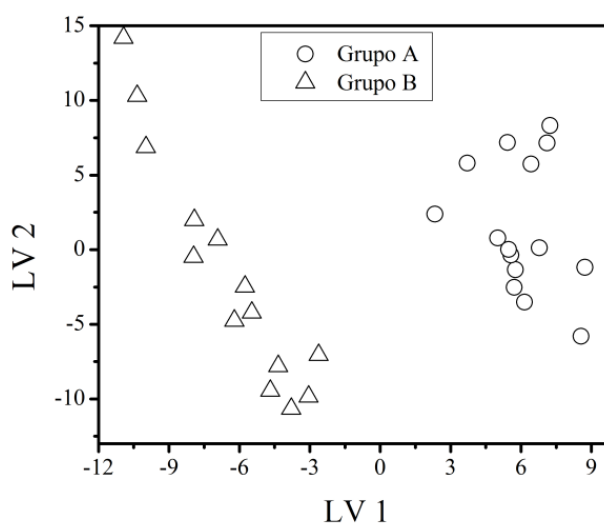
Na Figura 46-A pode-se perceber que uma amostra apresentou um comportamento discrepante das demais (triângulo do lado esquerdo do diagrama). Esta amostra é uma candidata a *outlier*. Seu eletroferograma foi inspecionado e percebeu-se uma flutuação muito elevada na linha base, desta forma tal amostra (número 29) foi removida do estudo para não influenciar nas análises subsequentes.

E na sequência, a ramificação da direita do diagrama representado na Figura 44 pode ser realizada pelo seguinte código, para a análise do gráfico de *scores* do modelo PLS-DA, como o representado na Figura 47.

```
#modelo PLS-DA
plsda.clase=c(rep(1, groupA),rep(2, groupB))
library(Discriminer)
pls.data=matrix[1:( groupA + groupB),]
pls.analysis=plsDA(pls.data, plsda.clase, autosel = FALSE)
#gráfico de scores
bmp(file = "PLSDA.bmp")
bmp(file = "PLSDA.bmp")
plot(pls.analysis$components, xlab = paste("PC 1; R2X global ( ",
round(100*pls.analysis$R2[2,2], dig = 2), "%); R2Y global ( ",
round(100*pls.analysis$R2[2,4], dig = 2), "%); Q2 global ( ",
round(100*pls.analysis$Q2[1,3], dig = 2), "%)", sep = "" ), ylab
= paste("PC 2"), pch = plsda.clase)
title(main = list("PLSDA", cex=1, col="black", font=1))
dev.off()
#resumo de coeficientes PLS-DA
```

```
summary(pls.analysis)
#slvar a lista de variáveis mais importantes VIP
write.table(pls.analysis$VIP, "VIP-plsda.csv", sep="\t", dec=".")
```

Figura 47 – Análise do diagrama de *scores* para o modelo PLS-DA.



Fonte: O AUTOR.

No final deste último código foi incluído um comando para salvar uma lista ordenada das VIP variáveis, Tabela 13, responsáveis por explicar a separação dos dois grupos de amostras observado na Figura 47. Esta lista obtida pela análise multivariada pode ser analisada concomitantemente com os resultados obtidos pela análise univariada, numa forma de complementariedade, com o intuito de uma maior possibilidade de encontrar novos biomarcadores, além de justificar novas revelações.

Tabela 13 – Resultados para a análise estatística multivariada. *R-Software*.

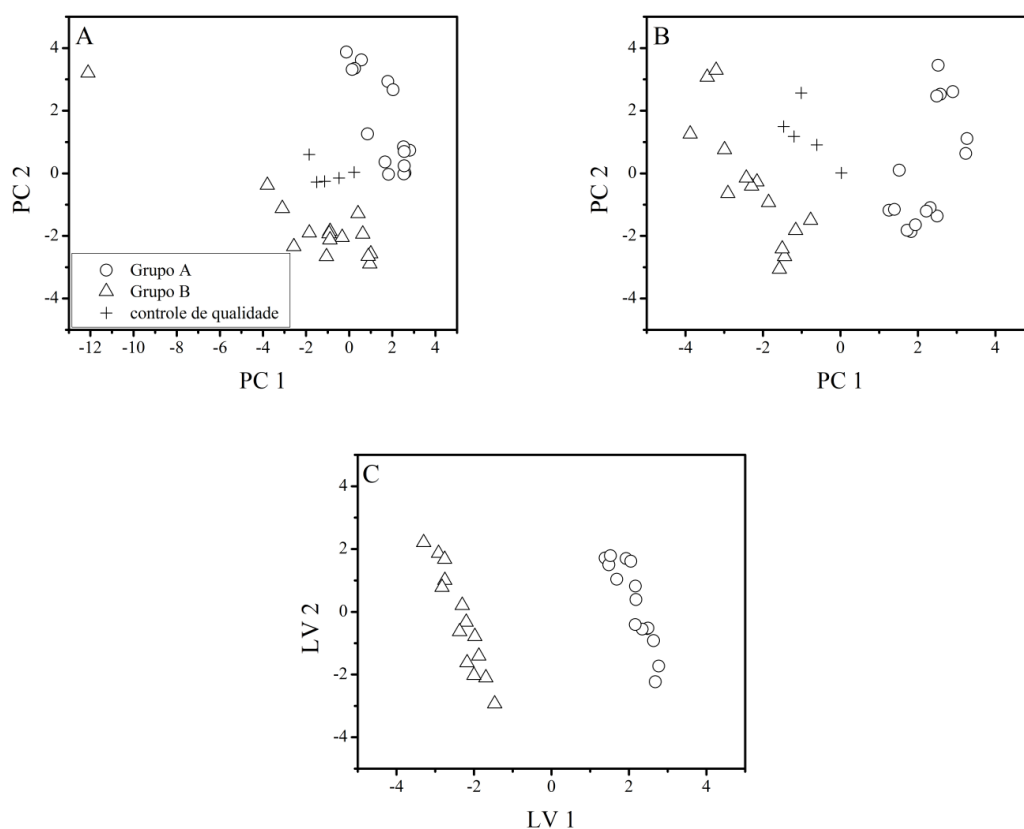
Metabolito	VIP (LV 1)	Posição
Cistina	2.8773	1°
Fenilalanina	2.8500	2°
Treonina	2.6163	3°
Arginina	2.5057	4°
Valina	2.2913	5°
Lisina	2.2809	6°
Histidina	1.4294	58°
Asparagina	0.3845	230°

Somente foram apresentados aqui os oito metabolitos previamente selecionados pelo desenho experimental.

Fonte: O AUTOR.

Para comparação com a rotina apresentada para a análise multivariada pelo *software* livre R, foi realizada uma análise em paralelo, com o mesmo conjunto de dados, com o *software* comercial SIMCA. A Figura 48 resume os principais resultados gráficos e em seguida a Tabela 14 mostra as variáveis mais importantes.

Figura 48 – Resumo dos resultados gráficos para análise multivariada obtidos pelo *software* SIMCA.



Em "A" análise do diagrama de *scores* do modelo PCA a partir das 30 amostras, em "B" PCA após a retirada da amostra número 29 e em "C" análise PLS-DA.

Fonte: O AUTOR.

Tabela 14 – Resultados para a análise estatística multivariada. *Software* SIMCA.

Metabolito	Posição
Cistina	1º
Fenilalanina	2º
Treonina	4º
Arginina	3º
Valina	12º
Lisina	38º
Histidina	84º
Asparagina	83º

Fonte: O AUTOR.

5.5.4 Conclusões

O procedimento utilizado na sequência de análise de dados tem um impacto direto sobre os resultados obtidos por um estudo. Estas discussões muitas vezes não são apresentadas na íntegra em publicações. O guia aqui mostrado, como um tutorial, permite esclarecer algumas etapas durante a análise de dados para a obtenção de novos biomarcadores e também serve como ferramenta didática para novos pesquisadores e alunos. Foi utilizado uma plataforma livre pelo fato de ser de fácil acesso. É importante encorajar tal discussão, de modo que trabalhos futuros possam apresentar os seus resultados de forma mais clara, e, conseqüentemente possam ser reproduzidos.

6 CONSIDERAÇÕES GERAIS E PERSPECTIVAS

Os casos de estudos apresentados neste trabalho mostram que métodos multivariados podem ser úteis para reconhecimento de padrões. Também para buscar variáveis responsáveis por discriminar classes de amostras, propondo assim, marcadores vibracionais e a quantificação de um analito que tenha atividade em diferentes regiões espectrais. No que tange a análise de amostras complexas tais como alimentos, fármacos e amostras biológicas, tais métodos são indispensáveis. Assim, são propostas determinações alternativas as já conhecidas, e novas, em alguns casos, com a característica de análises rápidas, mais representativas e menos propícia a desvios, como no caso de quantificações baseadas em um único comprimento de onda, e também podendo ser utilizadas para quantificações de mais de um analito simultaneamente.

No estudo com óleos vegetais, a abordagem PCA permitiu mostrar que as técnicas espectroscópicas podem ser utilizadas para quantificar misturas de óleo de soja em azeite de oliva extra-virgem, uma vez que estas amostras ficam completamente separadas no diagrama de *scores*, mesmo com seus espectros se sobrepondo em uma inspeção visual. Neste caso a análise multivariada permite tanto elucidar quais regiões do espectro são as mais importantes para serem monitoradas (isso foi analisado e justificado com modos vibracionais já descritos na literatura), como também todo o espectro pode ser utilizado para uma quantificação. Com uma regressão PLS foi possível prever o percentual de óleo de soja adicionado em azeite de oliva numa faixa de concentrações de zero a cem por cento.

No estudo do perfil de lipídeos de amostras de leite de vaca, foram propostos marcadores vibracionais, MIR e Raman, para quantificar a concentração de gordura total presente no leite. Tais marcadores foram selecionados pela análise de variáveis importantes de um modelo PLS ajustado para determinar o percentual de gordura em amostras de leite fluido. Como resultado, três modos vibracionais foram selecionados para cada técnica espectroscópica e, por uma regressão múltipla, foram utilizados para determinar o percentual de gordura em amostras de leite fluido, apresentado excelentes coeficientes de determinação quando comparado com as determinações feitas por um analisador eletrônico comercial. A abordagem PLS também foi utilizada para determinar com boa exatidão, comparado com determinações pela técnica de cromatografia à líquido, a concentração de rifampicina, isoniazida, etambutol e pirazinamida pela análise de um único espectro Raman de comprimidos comerciais utilizados para o tratamento anti-tuberculose.

Ao final é proposta uma metodologia para a determinação, e quantificação, da adição fraudulenta de soro de leite em amostras de leite fluido baseada no monitoramento do perfil de ácidos graxos. Neste estudo foi utilizada a abordagem de análise discriminante para selecionar um grupo de ácidos graxos que descreva bem a adulteração. Estes ácidos graxos selecionados através de análise multivariada foram testados e por fim resultaram

em bons marcadores para a determinação da fraude. Desta forma, esta tese se utiliza de diferentes técnicas experimentais e diferentes classes de amostras, tendo como interseção a utilização de métodos multivariados como facilitador na compreensão e quantificação de analitos.

Como perspectivas futuras, pretende-se finalizar a redação dos manuscritos dos trabalhos aqui descritos para a forma de artigos científicos. Além disso, esta tese abre uma linha de pesquisa tanto no Grupo de Espectroscopia e Engenharia de Materiais do dep. de Física como no Grupo de Química Analítica e Quimiometria do dep. de Química desta universidade, baseada em trabalhos de física e química aplicados, com abordagem de análise multivariada de amostras diversas, por técnicas espectroscópicas (NIR, MIR e Raman) em associação a técnicas de separação (CE, GC e LC).

Uma outra linha de pesquisa, após o conhecimento das matrizes de trabalho e da sugestão de variáveis importantes em determinações por métodos multivariados, é a de desenvolvimento de protótipos para quantificações específicas. Dois protótipos surgiram após o estudo de propriedades de amostras por métodos multivariados. O primeiro aplicado à quantificação dos teores de gordura, proteínas e lactose em amostras de leite fluido (este protótipo está em fase de desenvolvimento e escolha dos elementos ópticos a serem utilizados, como filtros e LEDs de infravermelho). Já o segundo baseia-se nas técnicas de separação para quantificação dos mesmos constituintes do leite supracitados, com o diferencial da detecção condutométrica sem contato, para contornar algumas dificuldades encontradas no protótipo baseado em medições de propriedades ópticas.

REFERÊNCIAS

- 1 MOROS, J.; GARRIGUES, S.; GUARDIA, M. d. l. Vibrational spectroscopy provides a green tool for multi-component analysis. *TrAC Trends in Analytical Chemistry*, v. 29, n. 7, p. 578–591, 2010.
- 2 SHAFFER, R. E. Multi- and megavariate data analysis. principles and applications, i. eriksson, e. johansson, n. kettaneh-wold and s. wold, umetrics academy, umeå, 2001, isbn 91-973730-1-x, 533pp. *Journal of Chemometrics*, v. 16, n. 5, p. 261–262, 2002.
- 3 WOLD, S.; SJÖSTRÖM, M. Chemometrics, present and future success. *Chemometrics and Intelligent Laboratory Systems*, v. 44, n. 1–2, p. 3–14, 1998.
- 4 LILAND, K. H. Multivariate methods in metabolomics – from pre-processing to dimension reduction and statistical analysis. *TrAC Trends in Analytical Chemistry*, v. 30, n. 6, p. 827–841, 2011.
- 5 KAROUI, R.; DOWNEY, G.; BLECKER, C. Mid-infrared spectroscopy coupled with chemometrics: A tool for the analysis of intact food systems and the exploration of their molecular structure-quality relationships - a review. *Chemical Reviews*, v. 110, n. 10, p. 6144–6168, 2010.
- 6 FORINA, M.; LANTERI, S.; CASALE, M. Multivariate calibration - review. *Journal of Chromatography A*, n. 1158, p. 61–93, 2007.
- 7 PEARSON, K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, v. 2, n. 6, p. 559–572, 1901.
- 8 HOTELLING, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psych.*, v. 24, 1933.
- 9 WOLD, S.; ESBENSEN, K.; GELADI, P. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, v. 2, n. 1–3, p. 37–52, 1987.
- 10 JOLLIFFE, I. *Principal Component Analysis*. [S.l.]: Springer, 2002. (Springer Series in Statistics).
- 11 GELADI, P.; KOWALSKI, B. R. Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, v. 185, n. 0, p. 1–17, 1986.
- 12 WOLD, S. et al. Multi-way principal components-and pls-analysis. *Journal of Chemometrics*, v. 1, n. 1, p. 41–56, 1987.
- 13 SKOOG, D. D. et al. *Fundamentos de química analítica*. [S.l.]: Cengage Learning, 2006.
- 14 PIMENTEL, M. F.; GALVAO, R. K. H.; ARAÚJO, M. C. U. Recomendações para calibração em química analítica, parte 2. calibração multianalito. *Química Nova*, v. 31, n. 2, p. 462–467, 2008.
- 15 BRERETON, R. G. *Chemometrics - Data Analysis for the Laboratory and Chemical Plant*. [S.l.]: John Wiley and Sons, 2003.

- 16 VARMUZA, K.; FILZMOSE, P. *Introduction to multivariate statistical analysis in chemometrics*. [S.l.]: CRC Press, 2008.
- 17 WEHRENS, R. *Chemometrics with R - Multivariate Data Analysis in the Natural Sciences and Life Sciences*. [S.l.]: Springer, 2011.
- 18 FIEHN, O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comparative and Functional Genomics*, v. 2, p. 155–168, 2001.
- 19 BARKER, M.; RAYENS, W. Partial least squares for discrimination. *Journal of Chemometrics*, v. 17, n. 3, p. 166–173, 2003.
- 20 FONVILLE, J. M. et al. The evolution of partial least squares models and related chemometric approaches in metabonomics and metabolic phenotyping. *Journal of Chemometrics*, v. 24, n. 11-12, p. 636–649, 2010.
- 21 BYLESJÖ, M. et al. Opls discriminant analysis: combining the strengths of pls-da and simca classification. *Journal of Chemometrics*, v. 20, n. 8-10, p. 341–351, 2006.
- 22 TRYGG, J.; WOLD, S. Orthogonal projections to latent structures (o-pls). *Journal of Chemometrics*, v. 16, n. 3, p. 119–128, 2002.
- 23 SNEDECOR, G. W.; COCHRAN, W. G. *Statistical Methods*. [S.l.]: Iowa State University Press, 1989.
- 24 MONTGOMERY, D.; PECK, E.; VINING, G. *Introduction to Linear Regression Analysis (4th ed.)*. [S.l.]: Wiley Sons, 2006.
- 25 SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, v. 52, n. 3-4, p. 591–611, 1965.
- 26 SALA, O. *Fundamentos da Espectroscopia Raman e no Infravermelho*. [S.l.]: Editora UNESP, 2008.
- 27 FOWLES, G. R. *Introduction to modern optics*. [S.l.]: Dover publications, 1989.
- 28 ZILIO, S. C. *Óptica Moderna - Fundamentos e aplicações*. e-book IFSC-USP, 2000. Disponível em: <<http://www.fotonica.ifsc.usp.br/ebook/e-book1.php>>.
- 29 HOMES, C. *Fourier Transform Infrared Spectroscopy*. These notes, 2007.
- 30 BRETZLAFF, R. S.; BAHDER, T. B. Apodization effects in fourier transform infrared difference spectra. *Revue de Physique Appliquée*, v. 21, n. 12, p. 833–844, 1986.
- 31 CUSTODIO, R.; KUBOTA, L. T.; ANDRADE, J. a. C. Lei dos processos de absorção da radiação. *Chemkeys*, p. 1–5, 2000.
- 32 JACKSON, J. D. *Classical electrodynamics*. [S.l.]: John Wiley and Sons, 1999.
- 33 LONG, D. A. *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules*. [S.l.]: John Wiley and Sons, 2002.
- 34 KUZMANY, H. *Solid-State Spectroscopy - An Introduction*. [S.l.]: Springer, 2009.

- 35 SKOOG, D. A.; HOLLER, F. J.; NIEMAN, T. A. *Princípios de Análise Instrumental*. [S.l.]: Bookmann, 2006.
- 36 LOPEZ, S. et al. Membrane composition and dynamics: A target of bioactive virgin olive oil constituents. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, v. 1838, n. 6, p. 1638–1656, 2014.
- 37 PRIORE, P.; SICULELLA, L.; GNONI, G. V. Extra virgin olive oil phenols down-regulate lipid synthesis in primary-cultured rat-hepatocytes. *The Journal of Nutritional Biochemistry*, v. 25, n. 7, p. 683–691, 2014.
- 38 PEREDA, M. et al. Polyelectrolyte films based on chitosan/olive oil and reinforced with cellulose nanocrystals. *Carbohydrate Polymers*, v. 101, n. 0, p. 1018–1026, 2014.
- 39 MA, W. et al. Characterization of gelatin-based edible films incorporated with olive oil. *Food Research International*, v. 49, n. 1, p. 572–579, 2012.
- 40 MARCOS, B. et al. Development of biodegradable films with antioxidant properties based on polyesters containing tocopherol and olive leaf extract for food packaging applications. *Food Packaging and Shelf Life*, v. 1, n. 2, p. 140–150, 2014.
- 41 VALENZUELA, C.; ABUGOCH, L.; TAPIA, C. Quinoa protein–chitosan–sunflower oil edible film: Mechanical, barrier and structural properties. *LWT - Food Science and Technology*, v. 50, n. 2, p. 531–537, 2013.
- 42 FRANCO, M. N. et al. Phenolic compounds and antioxidant capacity of virgin olive oil. *Food Chemistry*, v. 163, n. 0, p. 289–298, 2014.
- 43 LORENCINI, M. et al. Active ingredients against human epidermal aging. *Ageing Research Reviews*, v. 15, n. 0, p. 100–115, 2014.
- 44 D'ANGELO, S. et al. Hydroxytyrosol, a natural antioxidant from olive oil, prevents protein damage induced by long-wave ultraviolet radiation in melanoma cells. *Free Radical Biology and Medicine*, v. 38, n. 7, p. 908–919, 2005.
- 45 SOLTANIPOOR, F. et al. The effect of olive oil on prevention of striae gravidarum: A randomized controlled clinical trial. *Complementary Therapies in Medicine*, v. 20, n. 5, p. 263–266, 2012.
- 46 DELMONTE, P. et al. Separation characteristics of fatty acid methyl esters using slb-il111, a new ionic liquid coated capillary gas chromatographic column. *Journal of Chromatography A*, v. 1218, n. 3, p. 545–554, 2011.
- 47 CHRISTIE, W. W.; HAN, X. Chapter 7 - preparation of derivatives of fatty acids. In: _____. *Lipid Analysis (Fourth edition)*. [S.l.]: Woodhead Publishing, 2012. p. 145–158.
- 48 AOCS. *Determination of cis- and trans- Fatty Acids in Hydrogenated and Refined Oils and Fats by Capillary GLC*. Reapproved 1997 “Revised 2001”.
- 49 HOTELLING, H. The generalization of student's ratio. *The Annals of Mathematical Statistics*, v. 2, n. 3, p. 360–378, 1931.
- 50 CAMPBELL, G. P.; CURRAN, J. M. The interpretation of elemental composition measurements from forensic glass evidence iii. *Science and Justice*, v. 49, n. 1, p. 2–7, 2009.

- 51 GUILLÉN, M. D.; CABO, N. Characterization of edible oils and lard by fourier transform infrared spectroscopy. relationships between composition and frequency of concrete bands in the fingerprint region. *Journal of the American Oil Chemists' Society*, v. 74, n. 10, p. 1281–1286, 1997.
- 52 ROHMAN, A.; MAN, Y. B. C. The use of fourier transform mid infrared (ft-mir) spectroscopy for detection and quantification of adulteration in virgin coconut oil. *Food Chemistry*, v. 129, n. 2, p. 583–588, 2011.
- 53 AFSETH, N. K. et al. Raman and near-infrared spectroscopy for quantification of fat composition in a complex food model system. *Applied Spectroscopy*, v. 59, n. 11, p. 1324–1332, 2005.
- 54 MBA, O. et al. Application of near-infrared spectroscopy to characterize binary blends of palm and canola oils. *Industrial Crops and Products*, v. 61, n. 0, p. 472–478, 2014.
- 55 AFSETH, N. K.; WOLD, J. P.; SEGTAN, V. H. The potential of raman spectroscopy for characterisation of the fatty acid unsaturation of salmon. *Analytica Chimica Acta*, v. 572, n. 1, p. 85–92, 2006.
- 56 OLSEN, E. F. et al. Quantitative determination of saturated-, monounsaturated- and polyunsaturated fatty acids in pork adipose tissue with non-destructive raman spectroscopy. *Meat Science*, v. 76, n. 4, p. 628–634, 2007.
- 57 SILVA, P. H. F. Leite: Aspectos de composição e propriedades. *Química Nova na escola*, n. 6, p. 3–5, 1997.
- 58 JENNESS, R. Chapter one - the composition of milk. In: _____. *Nutrition and Biochemistry of Milk/maintenance*. [S.l.]: Academic Press, 1974. p. 3–107.
- 59 JENSEN, R. G. Chapter 1 - introduction. In: _____. *Handbook of Milk Composition*. San Diego: Academic Press, 1995. p. 1–3.
- 60 KLEYN, D. H. et al. Determination of fat in raw and processed milks by the gerber method: collaborative study. *J AOAC Int*, v. 84, n. 5, p. 1499–508, 2001. Kleyn, D H Lynch, J M Barbano, D M Bloom, M J Mitchell, M W Validation Studies United States Journal of AOAC International J AOAC Int. 2001 Sep-Oct;84(5):1499-508.
- 61 AOCS. *Roese-Gottlieb method*. 1990.
- 62 MANIRAKIZA, P.; COVACI, A.; SCHEPENS, P. Comparative study on total lipid determination using soxhlet, roese-gottlieb, bligh dyer, and modified bligh dyer extraction methods. *Journal of Food Composition and Analysis*, v. 14, n. 1, p. 93–100, 2001.
- 63 JENSEN, R. G.; FERRIS, A. M.; LAMMI-KEEFE, C. J. The composition of milk fat1. *Journal of Dairy Science*, v. 74, n. 9, 1991.
- 64 WALSTRA, P.; WOUTERS, J. T.; GEURTS, T. J. Lipids. In: _____. *Dairy Science and Technology*. Second edition. [S.l.]: CRC Press, 2006. p. 42.
- 65 HARA, A.; RADIN, N. Lipid extraction of tissues with a low-toxicity solvent. *Analytical Biochemistry*, v. 90, n. 1, p. 420–426, 1978.

- 66 IDF. *IDF141-C: Whole milk - Determination of milkfat, protein and lactose content - guidance on the operation of mid-infrared instruments*. 2000.
- 67 BECKE, A. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A*, v. 38, n. 6, p. 3098–3100, 1988. PRA.
- 68 LEE, C.; YANG, W.; PARR, R. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, v. 37, n. 2, p. 785–789, 1988. PRB.
- 69 RASSOLOV, V. A. et al. 6-31g* basis set for third-row atoms. *Journal of Computational Chemistry*, v. 22, n. 9, p. 976–984, 2001.
- 70 SCHMIDT, M. W. et al. General atomic and molecular electronic structure system. *Journal of Computational Chemistry*, v. 14, n. 11, p. 1347–1363, 1993.
- 71 CONTARINI, G. et al. Interlaboratory evaluation of milk fatty acid composition by using different GC operating conditions. *Journal of Food Composition and Analysis*, v. 32, n. 2, p. 131–140, 2013.
- 72 ZHOU, Q. et al. Chemical profiling of triacylglycerols and diacylglycerols in cow milk fat by ultra-performance convergence chromatography combined with a quadrupole time-of-flight mass spectrometry. *Food Chemistry*, v. 143, n. 0, p. 199–204, 2014. (Lucy).
- 73 MOTOYAMA, M. Structure and phase characterization of triacylglycerols by Raman spectroscopy. *National Institute of Livestock and Grassland Science Research*, n. 12, p. 19–68, 2012.
- 74 MOTOYAMA, M. et al. In situ Raman spectrometric analysis of crystallinity and crystal polymorphism of fat in porcine adipose tissue. *Journal of Agricultural and Food Chemistry*, v. 61, n. 1, p. 69–75, 2012.
- 75 SINGH, D. et al. Raman and DFT study of static, dynamic interactions and isotope effect in pyridazine h₂O/d₂O systems. *Vibrational Spectroscopy*, v. 49, n. 2, p. 242–250, 2009.
- 76 MISHRA, S. et al. Effect of methylation on wavenumber shift of ring breathing mode of pyrimidine in the solution of h₂O and d₂O by employing Raman difference spectroscopic technique and DFT approach. *Vibrational Spectroscopy*, v. 65, n. 0, p. 116–123, 2013.
- 77 ZHOU, S. et al. DFT studies of structure and vibrational spectra of 4-benzylidene-1-phenyl-2-selenomorpholino-1H-imidazol-5(4H)-one and its derivatives. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, v. 110, n. 0, p. 333–342, 2013.
- 78 MCGOVERIN, C. M. et al. Raman spectroscopic quantification of milk powder constituents. *Analytica Chimica Acta*, v. 673, n. 1, p. 26–32, 2010.
- 79 EL-ABASSY, R. M. et al. Fast determination of milk fat content using Raman spectroscopy. *Vibrational Spectroscopy*, v. 56, n. 1, p. 3–8, 2011.
- 80 NÓN, F. A. I.; GARRIGUES, S.; GUARDIA, M. de la. Nutritional parameters of commercially available milk samples by FTIR and chemometric techniques. *Analytica Chimica Acta*, v. 513, n. 2, p. 401–412, 2004.

- 81 LYNCH, J. M. et al. Precalibration evaluation procedures for mid-infrared milk analyzers¹. *Journal of Dairy Science*, v. 89, n. 7, p. 2761–2774.
- 82 CHONG, I.-G.; JUN, C.-H. Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, v. 78, n. 1–2, p. 103–112, 2005.
- 83 MEHMOOD, T. et al. A review of variable selection methods in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, v. 118, n. 0, p. 62–69, 2012.
- 84 AFANADOR, N. L.; TRAN, T. N.; BUYDENS, L. M. C. Use of the bootstrap and permutation methods for a more robust variable importance in the projection metric for partial least squares regression. *Analytica Chimica Acta*, v. 768, n. 0, p. 49–56, 2013.
- 85 WHO. *World Health Organization - Global tuberculosis report 2014*. Geneva: WHO Press, 2014.
- 86 WHO. *World Health Organization - Treatment of tuberculosis: guidelines*. [S.l.]: WHO Press, 2010.
- 87 CAMINERO, J. *Guidelines for clinical and operational management of drug-resistant Tuberculosis*. 2013.
- 88 CHERUKUVADA, S.; NANGIA, A. Polymorphism in an api ionic liquid: ethambutol dibenzoate trimorphs. *CrystEngComm*, v. 14, n. 23, p. 7840–7843, 2012.
- 89 RAJALAKSHMI, K.; GUNASEKARAN, S.; KUMARESAN, S. Vibrational assignment, homo–lumo and nbo analysis of (2s)-2-[(2-[(2S)-1-hydroxybutan-2-yl]aminoethyl)amino]butan-1-ol by density functional theory. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, v. 130, n. 0, p. 466–479, 2014.
- 90 ANGELONI, L.; MARZOCCHI, M. P.; SMULEVICH, G. Resonance raman spectra and excitation profiles of some rifamycins. *Journal of Raman Spectroscopy*, v. 15, n. 2, p. 90–96, 1984.
- 91 HOWES, B. D. et al. The influence of ph and anions on the adsorption mechanism of rifampicin on silver colloids. *Journal of Raman Spectroscopy*, v. 38, n. 7, p. 859–864, 2007.
- 92 JOHNSON, R. S. A resonance raman study on the interaction of rifampicin with escherichia coli rna polymerase. *Biochimica et Biophysica Acta (BBA) - General Subjects*, v. 839, n. 1, p. 16–25, 1985.
- 93 AKALIN, E.; AKYUZ, S. Vibrational structure of free and hydrogen bonded complexes of isoniazid: Ft-ir, ft-raman and dft study. *Journal of Molecular Structure*, v. 834–836, n. 0, p. 492–497, 2007.
- 94 AKYUZ, S.; AKYUZ, T. Ft-ir and ft-raman spectroscopic studies of adsorption of isoniazid by montmorillonite and saponite. *Vibrational Spectroscopy*, v. 48, n. 2, p. 229–232, 2008.
- 95 KALKAR, A. K.; BHOSEKAR, N. M.; KSHIRSAGAR, S. T. Polarized raman and infrared spectra of pyrazinamide. *Spectrochimica Acta Part A: Molecular Spectroscopy*, v. 45, n. 6, p. 635–641, 1989.

- 96 BRASIL. *Instrução Normativa MAPA N7*. 2010.
- 97 DUNN, W. B.; ELLIS, D. I. Metabolomics: Current analytical platforms and methodologies. *TrAC Trends in Analytical Chemistry*, v. 24, n. 4, p. 285–294, 2005.
- 98 IDLE, J. R.; GONZALEZ, F. J. Metabolomics. *Cell Metabolism*, v. 6, n. 5, p. 348–351, 2007.
- 99 ROESSNER, U. et al. *Metabolomics – The Combination of Analytical Biochemistry, Biology, and Informatics*. Burlington: Academic Press, 2011. (Comprehensive Biotechnology (Second Edition)).
- 100 ERIKSSON, L. et al. Using chemometrics for navigating in the large data sets of genomics, proteomics, and metabonomics (gpm). *Analytical and Bioanalytical Chemistry*, v. 380, n. 3, p. 419–429, 2004.
- 101 HENDRIKS, M. M. W. B. et al. Data-processing strategies for metabolomics studies. *TrAC Trends in Analytical Chemistry*, v. 30, n. 10, p. 1685–1698, 2011.
- 102 LILAND, K. H. Multivariate methods in metabolomics – from pre-processing to dimension reduction and statistical analysis. *TrAC Trends in Analytical Chemistry*, v. 30, n. 6, p. 827–841, 2011.
- 103 VEROUDEN, M. P. H. et al. Exploring the analysis of structured metabolomics data. *Chemometrics and Intelligent Laboratory Systems*, v. 98, n. 1, p. 88–96, 2009.
- 104 GOODACRE, R. et al. Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics*, v. 3, n. 3, p. 231–241, 2007.
- 105 KATAJAMAA, M.; OREŠIČ, M. Data processing for mass spectrometry-based metabolomics. *Journal of Chromatography A*, v. 1158, n. 1–2, p. 318–328, 2007.
- 106 BINO, R. J. et al. Potential of metabolomics as a functional genomics tool. *Trends in Plant Science*, v. 9, n. 9, p. 418–425, 2004.
- 107 R-SOFTWARE. *R Core Team - R: A Language and Environment for Statistical*. 2013. Disponível em: <<http://www.R-project.org/>>.
- 108 KAMLEH, M. A. et al. Optimizing the use of quality control samples for signal drift correction in large-scale urine metabolic profiling studies. *Analytical Chemistry*, v. 84, n. 6, p. 2670–2677, 2012.
- 109 ALBERICE, J. V. et al. Searching for urine biomarkers of bladder cancer recurrence using a liquid chromatography–mass spectrometry and capillary electrophoresis–mass spectrometry metabolomics approach. *Journal of Chromatography A*, v. 1318, n. 0, p. 163–170, 2013.
- 110 BALDERAS, C. et al. Plasma and urine metabolic fingerprinting of type 1 diabetic children. *Electrophoresis*, v. 34, n. 19, p. 2882–2890, 2013.
- 111 CANUTO, G. A. B. et al. Ce-esi-ms metabolic fingerprinting of leishmania resistance to antimony treatment. *Electrophoresis*, v. 33, n. 12, p. 1901–1910, 2012.

- 112 MORAES, E. P. et al. Metabolomic assessment with ce-ms of the nutraceutical effect of cystoseira spp extracts in an animal model. *Electrophoresis*, v. 32, n. 15, p. 2055–2062, 2011.
- 113 LILLIEFORS, H. W. On the kolmogorov-smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, v. 62, p. 399–402, 1967.
- 114 M., H.; HOBOKEN, W. D. A. *Nonparametric Statistical Methods*. NJ: John Wiley and Sons, 1999.
- 115 FASSEL, V. A. International union of pure and applied chemistry. analytical chemistry division. commission on spectrochemical and other optical procedures for analysis. nomenclature, symbols, units and their usage in spectrochemical analysis. ii. data interpretation. *Analytical Chemistry*, v. 48, n. 14, p. 2294–2294, 1976.
- 116 DETTMER, K.; ARONOV, P. A.; HAMMOCK, B. D. Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews*, v. 26, n. 1, p. 51–78, 2007.
- 117 URY, H. K. A comparison of four procedures for multiple comparisons among means (pairwise contrasts) for arbitrary sample sizes. *Technometrics*, v. 18, n. 1, p. 89–97, 1976.
- 118 NELSON, P. R. C.; TAYLOR, P. A.; MACGREGOR, J. F. Missing data methods in pca and pls: Score calculations with incomplete observations. *Chemometrics and Intelligent Laboratory Systems*, v. 35, n. 1, p. 45–65, 1996.

APÊNDICE A – Trabalhos publicados

Figura 49 – Artigo - Óleos vegetais - NIR, MIR e Raman.

Food Anal. Methods
DOI 10.1007/s12161-015-0121-y

Quantification of Extra-virgin Olive Oil Adulteration with Soybean Oil: a Comparative Study of NIR, MIR, and Raman Spectroscopy Associated with Chemometric Approaches

Thiago O. Mendes • Roney A. da Rocha •
Brenda L. S. Porto • Marcone A. L. de Oliveira •
Virgílio de C. dos Anjos • Maria J. V. Bell

Received: 10 September 2014 / Accepted: 1 February 2015
© Springer Science+Business Media New York 2015

Fonte: FOOD ANALITICAL METHODS, doi:10.1007/s12161-015-0121-y.

Figura 50 – Artigo - Resíduos de fármacos em leite - NIR.

PESQUISA	<i>Revista Brasileira de Ciências da Saúde</i>	
Research	DOI:10.4034/RBCS.2014.18.03.05	Volume 18 Número 3 Páginas 219-224 2014 ISSN 1415-2177
<h2>Análise de Resíduos de Diclofenaco Sódico Veterinário em Leite por Espectroscopia no Infravermelho Próximo</h2> <p>Analysys Of Residues Of Veterinary Sodium Diclofenac in Milk through Near Infrared Spectroscopy</p> <p>LEANDRO DA CONCEIÇÃO LUIZ^{1*} MARIA JOSÉ VALENZUELA BELL² RONEY ALVES DA ROCHA³ THIAGO DE OLIVEIRA MENDES⁴ VIRGÍLIO DE CARVALHO DOS ANJOS⁵</p>		

Fonte: REVISTA BRASILEIRA DE CIÊNCIAS DA SAÚDE, doi:10.4034/RBCS.2014.18.03.05.

Figura 51 – Artigo - Gordura Trans - CE e GC.

Food Control 55 (2015) 230–235



Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont



Fast screening method for the analysis of trans fatty acids in processed food by CZE-UV with direct detection 

Brenda Lee Simas Porto, Isaura Daniele Leite Faria, Thiago de Oliveira Mendes, Marcone Augusto Leal de Oliveira*

QQAQ (Grupo de Química Analítica e Quimiometria) Departamento de Química, Instituto de Ciências Exatas, Universidade Federal de Juiz de Fora, Cidade Universitária, CEP 36036-330, Juiz de Fora, MG, Brazil

ARTICLE INFO	ABSTRACT
<p><small>Article history:</small> Received 15 November 2014 Received in revised form 18 February 2015 Accepted 19 February 2015 Available online 27 February 2015</p>	<p>In the present work, an alternative method for trans fatty acids (TFA) analysis with direct UV detection using capillary zone electrophoresis (CZE-UV) is proposed. The background electrolyte used consisted of 12.0 mmol L⁻¹ of tetraborate buffer, 12.0 mmol L⁻¹ of Brij 35, 33% methanol and 17% acetonitrile. Trans fatty acids in different samples such as butter toffee, cake mix, stuffed wafers, chocolate and a mix for Brazilian cheese bread were successfully quantified within an analysis time of 13 min, taking into account the statistical approach based on response factor calculation using C19:1 (nonadecanoic acid) as</p>

Fonte: FOOD CONTROL, doi:10.1016/j.foodcont.2015.02.027.

Figura 52 – Artigo - Instrumentação.

<http://dx.doi.org/10.5935/0100-4042.20140263> *Quim. Nova*, Vol. 38, No. 1, 132-135, 2015

CONSTRUÇÃO DE UM DISPOSITIVO DE PRESSURIZAÇÃO A VÁCUO PARA A PREPARAÇÃO DE FASES ESTACIONÁRIAS MONOLÍTICAS VIA PROCESSO SOL-GEL

Sandro Pereira Ribeiro^a, Fernando Antonio Simas Vaz^a, Marcone Augusto Leal de Oliveira^{a*}, Thiago de Oliveira Mendes^b e José Paulo Rodrigues Furtado de Mendonça^b


^aDepartamento de Química, Instituto de Ciências Exatas, Universidade Federal de Juiz de Fora, 36036-330 Juiz de Fora – MG, Brasil
^bDepartamento de Física, Instituto de Ciências Exatas, Universidade Federal de Juiz de Fora, 36036-330 Juiz de Fora – MG, Brasil

Recibido em 26/06/2014; aceito em 08/08/2014; publicado na web em 22/09/2014

Fonte: QUÍMICA NOVA, doi:10.5935/0100-4042.20140263.

Figura 53 – Artigo - Ensino de física.

e-Mosaicos - Revista Multidisciplinar de Ensino, Pesquisa, Extensão e Cultura do Instituto de Aplicação Fernando Rodrigues da Silveira (CAp-UERJ)
 ANO 1 - V. 1 - N. 1 - JUNHO 2012 - ISSN: 2316-9303



e-Mosaicos
 Revista Multidisciplinar de Ensino, Pesquisa, Extensão e Cultura do Instituto de Aplicação Fernando Rodrigues da Silveira (CAp-UERJ)

OFICINA DE ELETRICIDADE: UMA PROPOSTA DE APLICAÇÃO E USO CONSCIENTE DA ENERGIA ELÉTRICA

Thiago de Oliveira Mendes¹
Cláudio Pires de Mendonça²
José Mauricio de Andrade³

UFJF

Fonte: E-MOSAICOS, ISSN:2316-9303.

APÊNDICE B – Participação em Eventos

Internacional Conference on Near Infrared Spectroscopy - Foz do Iguaçu - outubro 2015.

Use of FT-NIR to detect residues of veterinary drugs in milk. (Pôster), co-autor.

Congresso Nacional de Laticínios - Juiz de Fora - julho 2015.

Estudo de antimicrobianos em leite. Por que analisá-los? (Pôster), co-autor.

Internacional symposium on electro and liquid phase separation techniques (ITP) and Latin American symposium on biomedical and biopharmaceutical, and industrial applications of capillary electrophoresis (LACE) - Natal - outubro 2014.

Evaluation of milk adulteration by cheese whey using capillary electrophoresis-UV detection a chemometric approach: analysis of changes in fatty acids content. (Pôster), autor principal.

Construction of vacuum pressurization device for preparation of sol-gel monolithic stationary phases. (Pôster), co-autor.

Congresso brasileiro de ciências e tecnologia de alimentos (SBCTA) - Aracaju - outubro 2014.

Análise quantitativa da adulteração do azeite de oliva extra-virgem comercial com óleo de soja utilizando espectrometria de infravermelho próximo e análise multivariada. (Pôster - Trabalho vencedor do Prêmio Leopold Hartman), co-autor.

Workshop da pós-graduação em Física UFJF - Juiz de Fora - setembro 2014.

Busca conformacional das moléculas de palmítico, esteárico, oleico e triacilglicerol por métodos semi-empíricos e ab-iníto. (Pôster), co-autor.

Encontro nacional de física da matéria condensada (ENFMC) - Costa do Saúpe - maio 2014.

Evaluation of multivariate statistical models for data analysis in infrared and raman using R-software. (Pôster), autor principal.

Workshop da pós-graduação em Física UFJF - Juiz de Fora - outubro 2013.

R-Software para análise multivariada de dados de infravermelho e Raman. (Seção oral e pôster), autor principal.

Workshop on collaborative research between CEU San Pabro University & brazilian universities and research institutions - São Paulo - agosto 2013.

Controlled experiment for optimizing data analysis in CE-MS metabolomics. (Seção oral), autor principal.

Scientific meeting of the spanish society of chromatography and related techniques (SECyTA) - Tarragona - novembro 2012.

Controlled experiment for evaluation of statistical models for data analysis in metabolomics. (Pôster), autor principal.

Metabolomic approach to evaluate liver growth factor (LGF) effect in mice with emphysema caused by tobacco smoke exposure. (Pôster), co-autor.

Encontro nacional de química analítica (ENQA) - Campos do Jordão - outubro 2011.

Espectroscopia de infravermelho para quantificação dos teores de gordura, proteínas e lactose em amostras de leite líquido. (Pôster), autor principal. (Pôster), autor principal.

ANEXO A – Colaborações em andamento

Brenda Lee Simas Porto, doutoranda em Química pela Universidade Federal de Juiz de Fora. **Estudo relacionado a termo oxidação de óleos vegetais.** Contribuição: Medidas de infravermelho próximo e análise multivariada de dados.

Filomena Karla de Castro Monteiro, mestranda em Ciência e tecnologia do leite e derivados pela Universidade Federal de Juiz de Fora. **Pesquisa relacionada a análise de soro de leite em amostras de leite.** Contribuição: Medidas de infravermelho próximo e análise multivariada de dados.

Igor Moura de Paiva, doutorando em Pharmaceutical Science pela University of Alberta. **Pesquisa relacionada a análise de Exopolissacarídeo.** Contribuição: Medidas de infravermelho próximo e análise de dados com métodos multivariados.

Leandro da Conceição Luiz, doutorando em Física pela Universidade Federal de Juiz de Fora. **Pesquisa relacionada a análise de fármacos em amostras de leite.** Contribuição: Auxílio com seção experimental e interpretação de dados por métodos multivariados.

Liliane Pereira Pinto, doutoranda em Engenharia Biomédica pela Universidade do Vale do Paraíba. **Estudo relacionado ao envelhecimento da pele humana por espectroscopia Raman.** Contribuição: Análise de dados com métodos multivariados.

Wesley William Gonçalves Nascimento, professor do departamento de Farmácia da Universidade Federal de Juiz de Fora campus avançado de Governador Valadares. **Pesquisa relacionada a análise de ureia em amostras de leite.** Contribuição: Interpretação espectros de infravermelho médio e análise de dados por métodos multivariados.