



**UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

Tese de Doutorado

**TEOR DE ÓLEO DE ALHO EM CÁPSULAS GELATINOSAS POR
ESPECTROSCOPIA NO INFRAVERMELHO MÉDIO E TÉCNICAS
QUIMIOMÉTRICAS**

Valeria Visani

João Pessoa – PB – Brasil

Agosto/2017



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

Tese de Doutorado

**TEOR DE ÓLEO DE ALHO EM CÁPSULAS GELATINOSAS POR
ESPECTROSCOPIA NO INFRAVERMELHO MÉDIO E TÉCNICAS
QUIMIOMÉTRICAS**

Valeria Visani*

Tese apresentada ao Programa de Pós-Graduação em Química da Universidade Federal da Paraíba como parte dos requisitos para obtenção do título de Doutor em Ciências, área de concentração Química Analítica.

Orientador: Prof. Dr. Mário César Ugulino de Araújo

2º Orientador: Profª Drª Fernanda Araújo Honorato

Coorientador: Dr. Ricardo Saldanha Honorato

João Pessoa – PB – Brasil

Agosto/2017

Catálogo na Publicação
Seção de Catalogação e Classificação

V822t Visani, Valeria.
Teor de óleo de alho em cápsulas gelatinosas por espectroscopia no infravermelho médio e técnicas quimiométricas / Valeria Visani. - João Pessoa, 2017.
142 f. : il. -

Orientador: Dr. Mário César Ugulino de Araújo.
Tese (Doutorado) - UFPB/CCEN/PPGQ


1. Química. 2. Óleo de alho. 3. Análise screening.
4. Calibração multivariada. 5. Classificação multivariada.
I. Título.

UFPB/BC


CDU - 54(043)

Teor de óleo de alho em cápsulas gelatinosas por espectroscopia no infravermelho médio e técnicas quimiométricas.

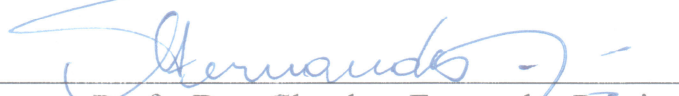
Tese de Doutorado apresentada pela aluna Valeria Visani e aprovada pela banca examinadora em 24 de agosto de 2017.



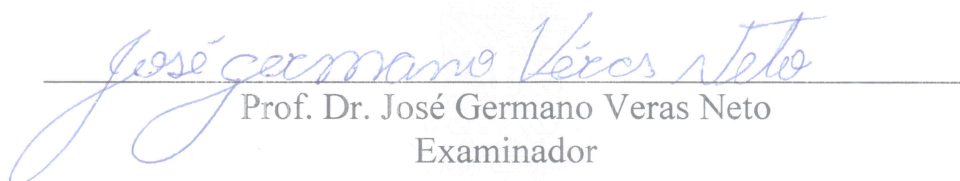
Prof. Dr. Mário César Ugulino de Araújo
Orientador/Presidente




Profa. Dra. Fernanda Araújo Honorato
2ª. Orientadora



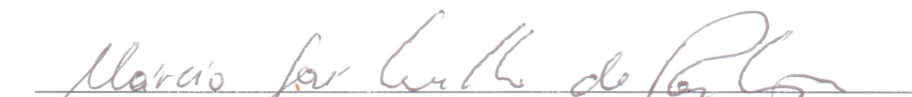
Profa. Dra. Claudete Fernandes Pereira
Examinadora



Prof. Dr. José Germano Veras Neto
Examinador



Profa. Dra. Lilians de Fátima Bezerra Lira de Pontes
Examinadora



Prof. Dr. Márcio José Coelho de Pontes
Examinador

“Tenho a impressão de ter sido uma criança brincando à beira-mar, divertindo-me em descobrir uma pedrinha mais lisa ou uma concha mais bonita que as outras, enquanto o imenso oceano da verdade continua misterioso diante de meus olhos” (Isaac Newton)

À minha inesquecível e amada mãe, responsável por quem sou hoje, pelo exemplo de força e de virtudes, por todo o amor e cuidados dedicados a mim, por sempre acreditar e me apoiar.

Com amor, saudade e gratidão,

Dedico.

AGRADECIMENTOS

- A Deus, pelo dom da vida.
- Ao meu amado marido, por todo o amor, dedicação, apoio e paciência. Por acreditar em mim e me ajudar a sempre seguir em frente diante das dificuldades. Você desperta o que há de melhor em mim e ilumina a minha vida.
- À minha família, pelo amor e incentivo, por estarem sempre ao meu lado, torcendo por minha vitória.
- Ao Prof. Dr. Mário César Ugulino de Araújo, meu orientador desde a iniciação científica, que me acolheu de braços abertos após eu ter trilhado outros caminhos. Obrigada por todos os ensinamentos e por sempre acreditar em mim.
- À minha querida amiga e orientadora Prof.^a Dr^a Fernanda Araújo Honorato, por me confiar esse desafio, pela amizade, por toda a orientação, paciência e dedicação. Você é um exemplo de pessoa e profissional. Tenho um orgulho enorme de ser sua orientanda.
- Ao amigo e coorientador Ricardo Saldanha Honorato, pelos ensinamentos e discussões relevantes para o andamento da tese.
- À minha grande amiga Graça, irmã que escolhi, por toda a amizade, incentivo e apoio incondicional. Por em determinados momentos confiar mais em mim que eu mesma. Você é um grande exemplo de determinação e coragem.
- À Anninha, doce amiga que o doutorado me presenteou. Por ser minha companheira e cúmplice nesse árduo caminho que é o doutorado. Pelas conversas sérias e nem tão sérias assim, pelo constante encorajamento, por toda a ajuda que me deu nesse percurso e pela amizade inestimável.
- A Marcelino, pela grande ajuda na parte experimental.
- Aos amigos do LAQA e do LAC, pelo companheirismo e auxílio.
- À professora Fernanda Pimentel e aos demais professores e funcionários do LAC, pela acolhida durante o desenvolvimento do experimental.
- Aos professores e funcionários do DQ/UFPB.
- E à CAPES pela bolsa concedida.

SUMÁRIO

Lista de figuras.....	IX
Lista de tabelas.....	XV
Lista de abreviaturas.....	XVII
Resumo.....	XX
Abstract.....	XXI
Capítulo 1. INTRODUÇÃO	2
1.1. Objetivo.....	4
1.1.1. Objetivo geral.....	4
1.1.2. Objetivos específicos.....	4
Capítulo 2. FUNDAMENTAÇÃO TEÓRICA	7
2.1. Alho.....	7
2.2. Suplementos de alho.....	11
2.3. Análise da composição de alho e de seus produtos por GC e HPLC....	14
2.4. Espectroscopia no infravermelho.....	18
2.4.1 Espectroscopia no Infravermelho Médio (MIR).....	19
2.4.1.1. Aplicações do MIR.....	20
2.5. Análise multivariada dos dados.....	23
2.5.1. Pré-processamento dos dados.....	23
2.5.2. Técnicas de reconhecimento de padrões.....	27
2.5.2.1. Técnicas de reconhecimento de padrões não supervisionadas	27
2.5.2.2. Técnicas de reconhecimento de padrões supervisionadas.....	29
2.5.2.2.1. Parâmetros de desempenho.....	30
2.5.2.2.2. Análise discriminante por mínimos quadrados parciais (PLS-DA)	32
2.5.2.2.3. Análise discriminante linear (LDA).....	33
2.5.3. Calibração multivariada.....	36
2.5.3.1. Regressão linear múltipla (MLR).....	38
2.5.3.1.1. SPA-MLR.....	40
2.5.3.2. Regressão por componentes principais (PCR).....	42
2.5.3.3. Regressão por mínimos quadrados parciais (PLS).....	45
2.5.3.4. Validação de modelos de calibração multivariada.....	47

2.5.3.4.1. Ensaio de Recuperação	49
2.5.4. Resolução multivariada de curvas com mínimos quadrados alternantes (MCR-ALS).....	50
2.5.4.1. Ambiguidade rotacional e restrições.....	52
2.5.4.2. Matrizes aumentadas	56
2.6. Espectroscopia MIR aplicada à análise de óleos comestíveis	57
Capítulo 3. EXPERIMENTAL	68
3.1. Amostras e reagentes	68
3.1.1. Preparação das misturas binárias de treinamento ou calibração	68
3.1.2. Preparação dos ensaios de recuperação	69
3.1.3. Cápsulas de óleo de alho	71
3.2. Aquisição dos espectros FT-IR	71
3.3. Procedimento quimiométrico.....	73
3.3.1. Classificação.....	73
3.3.2. Calibração Multivariada	75
3.3.3. Tratamento dos dados	77
Capítulo 4. RESULTADOS E DISCUSSÃO	79
4.1. Análise espectral.....	79
4.2. Análise de componentes principais (PCA)	83
4.3. Classificação de cápsulas gelatinosas de óleo de alho quanto ao seu teor.....	91
4.3.1. PLS-DA.....	91
4.3.2. SPA-LDA	96
4.4. Quantificação do teor de óleo de alho em cápsulas gelatinosas	100
4.4.1. Modelos PLS1.....	100
4.4.2. Modelos SPA-MLR	107
4.4.3. Modelos MCR-ALS	115
Capítulo 5. CONCLUSÃO	122
5.2. Propostas futuras	123
REFERÊNCIAS BIBLIOGRÁFICAS	124
ANEXO: Artigo publicado no periódico <i>Microchemical Journal</i> referente à classificação das cápsulas de óleo de alho quanto ao seu teor.	

ÍNDICE DE FIGURAS

Figura 2.1. Formação da alicina a partir da aliina e alinase. <i>Fonte:</i> adaptado de MENDES, 2008.	10
Figura 2.2. Compostos sulfurados formados a partir da degradação da (a) alicina: (b) dialil sulfeto (DAS), (c) dialil dissulfeto (DADS), (d) dialil trissulfeto (DATS), (e) 2-vinil-[4H]-1,3-ditiina (2-VDT), (f) 3-vinil-[4H]-1,2-ditiina (3-VDT), (g) (E)-ajoeno, (h) (Z)-ajoeno.	11
Figura 2.3. Exemplo de curva ROC. Pontos na curva correspondem a valores de sensibilidade e especificidade obtidas variando-se o <i>threshold</i> . A linha pontilhada indica classificação randômica.	33
Figura 2.4. Organização dos dados em calibração multivariada.	38
Figura 2.5. Ilustração de calibração PCR e previsão de novas amostras. (1) Decomposição de matriz de dados X em matrizes de escores e pesos. (2) Cálculo do vetor de regressão. (3) Projeção dos espectros de predição no espaço da calibração. (4) Previsão de propriedade de interesse das novas amostras. <i>Fonte:</i> adaptado de SUN, 2009.	44
Figura 2.6. Esquema de decomposição de dados por MCR-ALS para um sistema de dois componentes. <i>Fonte:</i> JUAN; JAUMOT; TAULER, 2014.	51
Figura 2.7. Perfis antes e após a aplicação das restrições de não negatividade, unimodalidade e sistema fechado (<i>closure</i>).	54
Figura 2.8. Modelo bilinear de MCR-ALS aumentado (a) pelas linhas, (b) pelas colunas e (c) pelas linhas e pelas colunas ao mesmo tempo. <i>Fonte:</i> JUAN; TAULER, 2006.	56
Figura 3.1. Ilustração da preparação dos ensaios de recuperação.	70
Figura 3.2. Espectrômetro FT-IR Spectrum 400 da Perkin Elmer com o acessório UATR. <i>Fonte:</i> Própria autora.	72
Figura 3.3. Ilustração da obtenção dos dados para a determinação da exatidão do método por MCR-ALS, desenvolvido para quantificação do teor de óleo de alho em cápsulas gelatinosas.	76
Figura 4.1. Espectros obtidos do padrão de óleo de alho e da média das 4 marcas de óleo de soja na região de 4000 cm^{-1} a 650 cm^{-1}	79

Figura 4.2. Espectros obtidos das 123 misturas em diferentes concentrações e das 136 cápsulas na região de 4000 cm^{-1} a 650 cm^{-1} 80

Figura 4.3. Espectros das 123 misturas de treinamento e 136 amostras de cápsulas gelatinosas de óleo de alho nas regiões compreendidas entre 3100 cm^{-1} e 2750 cm^{-1} e entre 1850 cm^{-1} a 650 cm^{-1} 81

Figura 4.4. Espectros das misturas de treinamento e cápsulas gelatinosas de óleo de alho pré-processadas por (a) SNV, (b) MSC e (c) 1ª derivada Savitzky Golay, com polinômio de 2ª ordem e janela de 11 pontos. 82

Figura 4.5. Gráfico dos escores de PC1 x PC2 para as 123 misturas de treinamento e 136 amostras de cápsulas de óleo de alho pré-processadas por SNV, utilizando toda a faixa espectral. (●) Misturas de treinamento da classe 1; (■) Misturas de treinamento da classe 2; (▲) Misturas de treinamento da classe 3; (○) Amostras da classe 1; (□) Amostras da classe 2; (△) Amostras da classe 3. 83

Figura 4.6. Gráfico dos escores de PC1 x PC2 para as 123 misturas de treinamento e 136 amostras de cápsulas de óleo de alho pré-processadas por MSC utilizando toda a faixa espectral. (●) Misturas de treinamento da classe 1; (■) Misturas de treinamento da classe 2; (▲) Misturas de treinamento da classe 3; (○) Amostras da classe 1; (□) Amostras da classe 2; (△) Amostras da classe 3. 84

Figura 4.7. Gráfico dos escores de PC1 x PC2 para as 123 misturas de treinamento e 136 amostras de cápsulas de óleo de alho pré-processadas por 1ª derivada Savitzky Golay com polinômio de 2ª ordem e janela de 11 pontos, utilizando toda a faixa espectral. Misturas de treinamento da classe 1 - círculos; Misturas de treinamento da classe 2 - quadrados; Misturas de treinamento da classe 3 - triângulos; (☆) Amostras da classe 1; (★) Amostras da classe 2; (☆) Amostras da classe 3. Óleo de soja da marca 1 – verde, Óleo de soja da marca 2 – cinza, Óleo de soja da marca 3 – magenta, Óleo de soja da marca 4 – ciano. 84

Figura 4.8. Gráfico dos pesos para PC1 e PC2 para as 123 misturas de treinamento e 136 amostras de cápsulas de óleo de alho pré-processadas por SNV (a) e (b), por MSC (c) e (d), e por 1ª derivada Savitzky Golay com polinômio de 2ª ordem e janela de 11 pontos (e) e (f), utilizando toda a faixa espectral. . 86

Figura 4.9. Gráfico dos escores de PC1 x PC2 para as 123 misturas de treinamento e 136 amostras de cápsulas de óleo de alho pré-processadas por (a) MSC e (b) 1ª derivada Savitzky Golay com polinômio de 2ª ordem e janela de 11 pontos, na região de 1000 cm^{-1} a 900 cm^{-1} . (●) Misturas de treinamento da classe 1; (■) Misturas de treinamento da classe 2; (▲) Misturas de treinamento

da classe 3; (○) Amostras da classe 1; (□) Amostras da classe 2; (▲) Amostras da classe 3. Amostras circuladas em verde apresentam comportamento diferente das demais. 87

Figura 4.10. Gráfico dos pesos para PC1 e PC2 para as 123 misturas de treinamento e 136 amostras de cápsulas de óleo de alho pré-processadas por 1ª derivada Savitzky Golay com polinômio de 2ª ordem e janela de 11 pontos (a) e (b) e por MSC (c) e (d), na região espectral de 1000 cm⁻¹ a 900 cm⁻¹. 88

Figura 4.11. Gráfico de Influência (a) e médias dos espectros dos lotes de cápsulas (b) para os dados pré-processados por MSC, na região de 1000 cm⁻¹ a 900 cm⁻¹. (●) Misturas de treinamento da classe 1; (■) Misturas de treinamento da classe 2; (▲) Misturas de treinamento da classe 3; (○) Amostras da classe 1; (□) Amostras da classe 2; (▲) Amostras da classe 3. Amostras circuladas em verde em (a) correspondem ao grupo de cápsulas separadas das demais cápsulas na **Figura 4.9**. Em (b) a curva em vermelho refere-se à média dos espectros das cápsulas circuladas em verde em (a). 89

Figura 4.12. Espectros médios das amostras cujas análises foram realizadas no mês de seu vencimento (em azul), das amostras do mesmo lote 1 ano após sua data vencimento (em vermelho) e das amostras de mesma marca e lote diferente, dentro do prazo de validade (em verde). Círculos e setas em vermelho estão indicando regiões afetadas pela oxidação das amostras. 90

Figura 4.13. Gráfico (a) da taxa de classificações erradas em função do número de variáveis latentes utilizado na construção do modelo para os dados pré-processados por MSC, em toda a região espectral, e de sensibilidade e especificidade em função da variação do *threshold* selecionado para (b) classe 1, (c) classe 2 e (d) classe 3. Sensibilidade em azul e especificidade em vermelho. 92

Figura 4.14. Respostas calculadas para as classes (a) 1, (b) 2 e (c) 3, versus misturas de treinamento e amostras de teste, para os dados pré-processados por MSC em toda a região espectral. (●) Misturas de treinamento da classe 1; (■) Misturas de treinamento da classe 2; (▲) Misturas de treinamento da classe 3; (○) Amostras da classe 1; (□) Amostras da classe 2; (▲) Amostras da classe 3. A linha verde pontilhada indica o valor do *threshold* para cada classe. 93

Figura 4.15. Gráfico dos escores das misturas de treinamento e amostras de teste nas duas primeiras variáveis latentes, obtidos pelo modelo PLS-DA, para os dados pré-processados por MSC em toda a região espectral. (●) Misturas de treinamento da classe 1; (■) Misturas de treinamento da classe 2; (▲) Misturas de treinamento da classe 3; (○) Amostras da classe 1; (□) Amostras da classe 2; (▲) Amostras da classe 3. 94

Figura 4.16. Variáveis selecionadas pelo algoritmo SPA-MLR, para os dados pré-processados por (a) SNV, (b) MSC e (c) 1ª derivada. 96

Figura 4.17. Espectros médios das misturas de treinamento a cada concentração, e das cápsulas classificadas em classes não condizentes com seus rótulos. 99

Figura 4.18. Gráfico das misturas de treinamento e amostras de teste e suas intensidades na variável selecionada pelo SPA-LDA, para os dados pré-processados por MSC. Os *thresholds* utilizados para separar as classes são indicados por linhas pontilhadas verdes. (●) Misturas de treinamento da classe 1; (■) Misturas de treinamento da classe 2; (▲) Misturas de treinamento da classe 3; (○) Amostras da classe 1; (□) Amostras da classe 2; (△) Amostras da classe 3. 99

Figura 4.19. Gráficos de RMSECV em função do número de variáveis latentes utilizado nos modelos, para os dados pré-processados por 1ª derivada (a) e por MSC (b), em toda a região espectral. 101

Figura 4.20. Gráficos de escores (a) e (b) e de coeficientes de regressão (c) e (d) obtidos pelos modelos PLS1, aplicados aos dados pré-processados por 1ª derivada (a) e (c) e por MSC (b) e (d), em toda a região espectral. 101

Figura 4.21. Gráficos de escores obtidos pelos modelos PLS1, aplicados aos dados pré-processados por (a) 1ª derivada e (b) MSC, em toda a região espectral, com VL 2 explicando a variância dos dados em relação às marcas de soja utilizadas nas misturas. 102

Figura 4.22. Gráficos de escores (a) e (b); e dos RMSECV em função do número de variáveis latentes (c) e (d); obtidos pelos modelos PLS1, aplicados aos dados pré-processados por 1ª derivada (a) e (b) e por MSC (b) e (d), na região de 940 cm^{-1} a 900 cm^{-1} 103

Figura 4.23. Gráficos de escores obtidos pelos modelos PLS1, aplicados aos dados pré-processados por (a) 1ª derivada e (b) MSC, na região de 940 cm^{-1} a 900 cm^{-1} , com as marcas dos óleos de soja utilizadas nas misturas diferenciados por cor. 103

Figura 4.24. Gráficos das concentrações medidas versus as concentrações previstas na construção dos modelos de calibração PLS1, em toda a região espectral (a) e (b); e na região de 940 cm^{-1} a 900 cm^{-1} (c) e (d), para os pré-processados por 1ª derivada (a) e (c) e por MSC (b) e (d). Círculos azuis são os valores previstos na calibração e os vermelhos na validação. 104

Figura 4.25. Gráficos das concentrações medidas versus as concentrações previstas na etapa de previsão, em toda a região espectral (a) e (c); e na região

de 940 cm^{-1} a 900 cm^{-1} (b) e (d). Dados pré-processados por 1ª derivada (a) e (b) e por MSC (c) e (d). 105

Figura 4.26. Número de variáveis selecionadas pelo SPA-MLR com os dados pré-processados por 1ª derivada (polinômio de 2ª ordem com janela de 11 pontos) (a) utilizando toda a região espectral e (b) a região de 940 cm^{-1} a 900 cm^{-1} 107

Figura 4.27. Variáveis selecionadas pelo SPA-MLR, utilizando toda a região espectral, com os dados pré-processados por 1ª derivada (polinômio de 2ª ordem com janela de 11 pontos). 108

Figura 4.28. Variáveis selecionadas pelo SPA-MLR, utilizando a região espectral de 940 cm^{-1} a 900 cm^{-1} , com os dados pré-processados por 1ª derivada (polinômio de 2ª ordem com janela de 11 pontos). 108

Figura 4.29. Gráfico das concentrações de referência versus as concentrações previstas na construção do modelo de calibração SPA-MLR, usando validação cruzada completa, para os dados pré-processados por 1ª derivada em (a) toda a região espectral e (b) na região de 940 cm^{-1} a 900 cm^{-1} 109

Figura 4.30. Variáveis selecionadas pelo SPA-MLR, utilizando toda a região espectral, com os dados pré-processados por MSC. 110

Figura 4.31. Variáveis selecionadas pelo SPA-MLR, utilizando a região espectral de 940 cm^{-1} a 900 cm^{-1} , com os dados pré-processados por MSC. 110

Figura 4.32. Gráfico das concentrações de referência versus as concentrações previstas na construção do modelo de calibração SPA-MLR, para os dados pré-processados por MSC em (a) toda a região espectral e (b) na região de 940 cm^{-1} a 900 cm^{-1} 111

Figura 4.33. Gráficos dos resíduos das misturas de calibração, para os dados pré-processados por 1ª derivada Savitzky Golay com polinômio de 2ª ordem e janela de 11 pontos (a) e (b); e por MSC (c) e (d), em toda a região espectral (a) e (c) e na região 940 cm^{-1} a 900 cm^{-1} (b) e (d). 111

Figura 4.34. Gráfico das concentrações de referência versus as concentrações previstas das 136 amostras de cápsulas comerciais de óleo de alho obtidas pelo modelo SPA-MLR, para os dados pré-processados por 1ª derivada em (a) toda a região espectral e (b) na região de 940 cm^{-1} a 900 cm^{-1} 112

Figura 4.35. Gráfico das concentrações de referência versus as concentrações previstas das 136 amostras de cápsulas comerciais de óleo de alho, obtidas para os modelos SPA-MLR, para os dados pré-processados por MSC em (a) toda a região espectral e (b) na região de 940 cm^{-1} a 900 cm^{-1} 112

- Figura 4.36.** Gráficos dos resíduos das 136 cápsulas de óleo de alho, para os dados pré-processados por 1ª derivada Savitzky Golay com polinômio de 2ª ordem e janela de 11 pontos (a) e (b); e por MSC (c) e (d), em toda a região espectral (a) e (c) e na região de 940 cm^{-1} a 900 cm^{-1} (b) e (d). 113
- Figura 4.37.** Configuração de matriz aumentada utilizada na análise MCR-ALS. 115
- Figura 4.38.** Ilustração das etapas da análise MCR-ALS. 117
- Figura 4.39.** Espectros recuperados e experimentais do padrão de óleo de alho e do óleo de soja. 118
- Figura 4.40.** Gráfico da regressão linear entre as concentrações previstas pelo MCR-ALS e as concentrações de referência para as misturas de calibração. 118
- Figura 4.41.** Gráfico da regressão linear entre as concentrações previstas pelo MCR-ALS e as concentrações de referência para as cápsulas comerciais de óleo de alho (em azul) e para as recuperações (em verde). 119

ÍNDICE DE TABELAS

Tabela 2.1. Produção em 2014 dos dez maiores produtores de alho.....	7
Tabela 2.2. Representação geral de uma matriz de confusão.....	30
Tabela 3.1. Quantidades de misturas para cada fração mássica (m/m) com cada marca de óleo de soja.	69
Tabela 3.2. Quantidade de ensaios de recuperação para cada marca de cápsula e cada nível de fortificação.....	70
Tabela 3.3. Quantidades de cápsulas, marcas e lotes de acordo com suas concentrações (m/m).....	71
Tabela 3.4. Quantidades de misturas de treinamento por classe.	74
Tabela 3.5. Número de marcas, lotes e quantidades de amostras comerciais para cada classe.	74
Tabela 4.1. Matriz de confusão para as amostras de cápsulas de óleo de alho, obtida a partir do modelo PLS-DA construído para os dados pré-processados por MSC, em toda a região espectral.	94
Tabela 4.2. Valores de exatidão, especificidade e sensibilidade obtidos para as amostras de cápsulas de óleo de alho, a partir do modelo PLS-DA construído para os dados pré-processados por MSC, em toda a região espectral.....	95
Tabela 4.3. Matriz de confusão para o conjunto de teste, obtida a partir dos modelos SPA-LDA, para os três pré-processamentos utilizados.	97
Tabela 4.4. Parâmetros de desempenho da classificação do conjunto de teste, obtidos a partir dos modelos SPA-LDA.	97
Tabela 4.5. Resultados da validação dos modelos PLS para os dados pré-processados por MSC e 1ª derivada, em toda a região espectral e na região de 940 cm ⁻¹ a 900 cm ⁻¹ , para a faixa de 0,00% a 1,75% m/m.....	104
Tabela 4.6. Resultados da previsão das amostras de cápsulas de óleo de alho, a partir dos modelos PLS construídos para os dados pré-processados por MSC e 1ª derivada, em toda a região espectral e na região de 940 cm ⁻¹ a 900 cm ⁻¹ , na faixa de 0,00% a 1,75% m/m.....	105
Tabela 4.7. Resultados da previsão das amostras comerciais de cápsulas de óleo de alho, obtidas pelos modelos SPA-MLR, para os dados pré-processados	

por 1ª derivada e por MSC, utilizando toda a região espectral e a região de 940
cm⁻¹ a 900 cm⁻¹, na faixa de 0,00% a 1,75% m/m.....113

LISTA DE ABREVIATURAS

- AGE – Extrato de alho envelhecido (do inglês, *Aged Garlic Extract*)
- ASTM - *American Society for Testing and Materials*
- ATR – Reflectância total atenuada (do inglês, *Attenuated Total Reflectance*)
- AUC – Área sob a curva (do inglês, *Area Under the Curve*)
- CA – Análise de clusters (do inglês, *Clusters Analysis*)
- DA – Análise discriminante (do inglês, *Discriminant Analysis*)
- DADS – Dialil dissulfeto
- DAS – Dialil sulfeto
- DATS – Dialil trissulfeto
- EFA – Análise de fatores evolucionários (do inglês, *Evolving Factor Analysis*)
- FD - Função discriminante
- FID – Detector por ionização em chama (do inglês, *Flame Ionization Detector*)
- FIR – Infravermelho distante (do inglês, *Far Infrared*)
- FPD – Detector fotométrico de chama (do inglês, *Flame Photometric Detector*)
- FT - Transformada de Fourier (do inglês, *Fourier Transformed*)
- FTIR – Infravermelho com transformada de Fourier (do inglês, *Fourier Transformed Infrared*)
- GA – Algoritmo genético (do inglês, *Genetic Algorithm*)
- GC – Cromatografia gasosa (do inglês, *Gas Chromatography*)
- HCA – Análise de clusters hierárquico (do inglês, *Hierarchical Cluster Analysis*)
- HPLC – Cromatografia Líquida de Alta Eficiência (do inglês, *High Performance Liquid Chromatography*)
- HS-SPME - Microextração em fase sólida por *headspace* (do inglês, *Headspace - Solid Phase Microextraction*)
- IR – Infravermelho (do inglês, *Infrared*)
- iPLS - Regressão por mínimos quadrados parciais por intervalo (do inglês, *Interval Partial Least-Squares Regression*)

LDA - Análise discriminante (do inglês, *Linear Discriminant Analysis*)

MCR-ALS – Resolução multivariada de curvas por mínimos quadrados alternantes (do inglês, *Multivariate Curve Resolution with Alternating Least-Squares*)

MLR - Regressão linear múltipla (do inglês, *Multiple Linear Regression*)

MIR – Infravermelho médio (do inglês, *Mid Infrared*)

MS – Espectroscopia de massa (do inglês, *Mass Spectroscopy*)

MSC – Correção de espalhamento multiplicativo (do inglês, *Multiplicative Scattering Correction*)

MWHD - Hidrodestilação assistida por micro-ondas (do inglês, *Microwave-assisted hydrodistillation extraction*)

NIR – Infravermelho próximo (do inglês, *Near Infrared*)

PCR – Regressão em componentes principais (do inglês, *Principal Component Regression*)

PCA - Análise de Componentes Principais (do inglês, *Principal Component Analysis*)

PCs - Componentes principais (do inglês, *Principal Components*)

PLS – Regressão por mínimos quadrados parciais (do inglês, *Partial Least-Squares Regression*)

PLSCM – Modelagem de classes por regressão por mínimos quadrados parciais (do inglês, *Partial Least Squares Class Model*)

PLS-DA - Análise Discriminante por Mínimos Quadrados Parciais (do inglês, *Partial Least Squares – Discriminant Analysis*)

RMSECV – Raiz quadrada do erro médio quadrático de validação cruzada (do inglês, *Root Mean Squares Error of Cross Validation*)

RMSEP – Raiz quadrada do erro médio quadrático de previsão (do inglês, *Root Mean Squares Error of Prediction*)

RMSEV – Raiz quadrada do erro médio quadrático de validação (do inglês, *Root Mean Squares Error of Validation*)

ROC – Curvas de características de operação do receptor (do inglês, *Receiver Operating Characteristics*)

RPD - Relação de Desempenho do Desvio (do inglês, *Residual Prediction Deviation* ou *Relative Predictive Determinant*)

SAC - S-alilcisteína

SAMC - S-alil mercaptocisteína

SD – Destilação a vapor (do inglês, *Steam Distillation*)

SDE - Destilação seguida de extração por solvente (do inglês, *Simultaneous Distillation and Solvent Extraction*)

SDV - Desvio padrão dos erros de validação (do inglês, *Standard Deviation of Validation*)

SG – Derivada de Savitsky Golay (do inglês, *Savitsky Golay Derivative*)

SIMCA - Modelagem independente e flexível por analogia de classe (do inglês, *Soft Independent Modeling of Class Analogy*)

SNV – Variação normal padrão (do inglês, *Standard Normal Variate*)

SPA – Algoritmo das Projeções Sucessivas (do inglês, *Successive Projection Algorithm*)

SPME – Micro extração em fase sólida (do inglês, *Solid Phase Microextraction*)

STW – *Stepwise*

SVD – Decomposição em valores singulares (do inglês, *Singular Values Decomposition*)

UATR – Acessório de reflectância total atenuada (do inglês, *Universal Attenuated Total Reflectance*)

USE - Extração por ultrassom (do inglês, *Ultrasound-Assisted Extraction*)

UV – Ultravioleta (do inglês, *Ultraviolet*)

2 – VDT – 2-vinil-[4H]-1,3-ditiina

3 – VDT - 3-vinil-[4H]-1,2-ditiina

RESUMO

O óleo de alho, comercializado na forma de cápsulas gelatinosas, contém diversas substâncias consideradas benéficas à saúde, que atuam na redução do colesterol e da pressão arterial, no tratamento de gripe e diabetes e na prevenção do desenvolvimento de tumores e doenças cardiovasculares. Entretanto, divergências a respeito de sua funcionalidade tem sido relatadas e, uma possível causa, é a falta de padronização desses produtos. Considerando que não há legislação estabelecida quanto ao controle de qualidade do óleo de alho, comercializado em cápsulas, reveste-se de importância o desenvolvimento de métodos simples e rápidos para a análise desses produtos. Neste trabalho, propõe-se métodos de análise *screening* e de quantificação, utilizando espectroscopia na região do infravermelho médio por reflectância total atenuada (ATR) e métodos quimiométricos, para a classificação e quantificação de cápsulas de óleo de alho em função de suas concentrações. O Algoritmo das Projeções Sucessivas (SPA) aliado à Análise Discriminante Linear (LDA), obteve os melhores resultados de classificação das cápsulas em três classes, por meio de modelos construídos utilizando apenas a variável selecionada 924 cm^{-1} , onde a banda existente é atribuída a vibrações de estiramento de C-S-C e deformação de $=\text{CH}_2$, presentes nos constituintes majoritários do óleo de alho. Para as 136 cápsulas testadas, obteve-se uma taxa de concordância com as informações contidas em seus rótulos. Na quantificação do teor de óleo de alho das cápsulas, o método MCR-ALS (Resolução de Curvas Multivariadas com Mínimos Quadrados Alternantes), empregando configuração aumentada na direção das colunas (*column wise*), mostrou-se capaz de resolver as misturas, recuperando os espectros de óleo de alho e de soja com correlações de 0,8914 e 0,9997, respectivamente. Como resultado, MCR-ALS obteve falta de ajuste de 3,63% e variância explicada de 99,87%, limites de detecção e de quantificação iguais a 0,09% e 0,27%, respectivamente, RPD (Relação de Desempenho do Desvio) igual a 8,89 e os ensaios de recuperação forneceram índices de 98% a 99%, indicando boa exatidão do método. Assim como na classificação, as concentrações previstas para algumas marcas de cápsulas, revelaram-se em discordância com os valores declarados em seus rótulos, inclusive entre diferentes lotes da mesma marca. Deste modo, justifica-se a necessidade de controle dos órgãos de fiscalização junto aos fabricantes, para garantir a qualidade desses produtos. Os resultados apresentados sugerem que os métodos propostos são eficazes para investigar cápsulas gelatinosas de óleo de alho quanto ao seu teor.

Palavras-chave: óleo de alho, infravermelho médio, técnicas quimiométricas de classificação e calibração multivariada, SPA-LDA, MCR-ALS.

ABSTRACT

Garlic oil, in the form of gelatin capsules, contains various substances considered beneficial to health, which act to lower cholesterol and blood pressure, in the treatment of flu and diabetes, and to prevent the development of tumors and cardiovascular diseases. However, disagreements over its functionality have been reported and, one possible reason is the lack of standardization of these products. Considering that there is no established legislation on the quality control of garlic oil, commercialized in capsules, it is important to develop simple and fast methods for the analysis of these products. In this work, we propose screening and quantification analysis methods, using attenuated total reflectance (ATR) spectroscopy and chemometric methods for the classification and quantification of garlic oil capsules as a function of their concentrations. The Successive Projections Algorithm (SPA), together with Linear Discriminant Analysis (LDA), obtained the best results of classifying capsules in three classes, using models constructed using only the selected variable 924 cm^{-1} , where the existing band is assigned to stretching vibrations of C-S-C and deformation of $=\text{CH}_2$, present in the major constituents of garlic oil. For the 136 capsules tested, we obtained a rate of 88.2% of agreement with the information contained in their labels. In the quantification of the garlic oil content of the capsules, the MCR-ALS (Multivariate Curve Resolution with Alternating Least Squares) method, using a column wise configuration, was able to solve the mixtures, recovering the spectra of garlic and soybean oil with correlations of 0.8914 and 0.9997, respectively. As a result, MCR-ALS obtained a lack of fit of 3.63% and explained variance of 99.87%, limits of detection and quantification equals to 0,09% and 0,27%, respectively, RPD (Residual Prediction Deviation) equal to 8,89 and the recovery trials provided indexes of 98% to 99%, indicating good accuracy of the method. As in the classification, the concentrations predicted for some brands of capsules, proved to be in disagreement with the values declared on their labels, including between different batches of the same brand. In this way, it is justified the need of control of the inspection organs with the manufacturers, to guarantee the quality of these products. The results presented suggest that the proposed methods are effective in investigating gelatinous capsules of garlic oil as to their content.

Key words: garlic oil, medium infrared, chemometric techniques of classification and multivariate calibration, SPA-LDA, MCR-ALS.

CAPÍTULO 1

INTRODUÇÃO



Capítulo 1. INTRODUÇÃO

A Pesquisa Nacional de Saúde realizada em 2013 pelo Ministério da Saúde, em parceria com o IBGE - Instituto Brasileiro de Geografia e Estatística – apontou que, apesar dos brasileiros estarem buscando um estilo de vida mais saudável, sua dieta é deficiente de frutas e verduras e baseada em alimentos gordurosos (PORTAL BRASIL, 2014; FUNDAÇÃO OSWALDO CRUZ, 2017). Aliado aos maus hábitos alimentares, o estresse da vida moderna, originado pela sobrecarga de trabalho, além da falta de atividades físicas, têm propiciado o surgimento de inúmeras doenças como hipertensão, diabetes, problemas cardíacos etc. Com isso, variadas formas de melhorar a qualidade de vida, a saúde e o bem-estar, aliviar sintomas associados às doenças crônicas e aos efeitos colaterais de medicamentos, têm sido utilizadas pela população. Em um estudo realizado nos Estados Unidos nos anos de 2002, 2007 e 2012, em média 33,7% das 88.962 pessoas investigadas utilizavam alguma técnica para melhorar a saúde (CLARKE *et al.*, 2015). Entre essas técnicas, a mais utilizada era o consumo de suplementos alimentares na forma de comprimidos, cápsulas ou líquidos, demonstrando que os consumidores se sentem cada vez mais atraídos por produtos que proporcionem benefícios adicionais à saúde.

Os suplementos à base de alho eram os mais populares e utilizados pelas famílias norte-americanas em 1997 (AMAGASE *et al.*, 2001), mantendo-se ainda entre os 10 mais consumidos em pesquisas realizadas em 2012 (CLARKE *et al.*, 2015). No Brasil, uma pesquisa inédita sobre o consumo de suplementos alimentares, realizado entre março e abril de 2015, divulgado pela Associação Brasileira da Indústria de Alimentos Para Fins Especiais e Congêneres (ABIAD), indicou que 54% dos lares brasileiros utilizam suplementos alimentares e o óleo de alho está entre os mais consumidos (REVISTA EXAME, 2016).

Atualmente, o óleo de alho é classificado pela ANVISA como “Novo Alimento”, denominação dada aos alimentos ou substâncias sem histórico de consumo no país, ou alimentos com substâncias já consumidas e que venham a ser adicionadas ou utilizadas em níveis muito superiores aos atualmente observados nos alimentos (BRASIL. ANVISA - AGÊNCIA NACIONAL DE

VIGILÂNCIA SANITÁRIA, 1999a). Os novos alimentos não possuem alegação de propriedade funcional e deve constar nos rótulos de suas embalagens a advertência do Ministério da Saúde de que não existem evidências científicas comprovadas de que o alimento previna, trate ou cure doenças.

Apesar de existirem diversas pesquisas a respeito da funcionalidade do óleo de alho e de outros produtos derivados do alho, divergências são encontradas quanto às suas propriedades benéficas (QIDWAI; ASHFAQ, 2013). Essas incongruências podem estar relacionadas à falta de regulamentação e à padronização desses produtos quanto ao teor de compostos bioativos presentes.

Em geral, duas técnicas são utilizadas para identificação e quantificação dos componentes do alho: a cromatografia gasosa (GC) e cromatografia líquida de alta eficiência (HPLC). Apesar da GC ser utilizada em diversos trabalhos na área (YU; WU; LIOU, 1989; LEE; KIM; LEE, 2003; ABU-LAFI *et al.*, 2004; CALVO-GÓMEZ; MORALES-LÓPEZ; LÓPEZ, 2004; KIMBARIS *et al.*, 2006), sabe-se que a alta temperatura empregada na realização das análises causa a degradação de alguns compostos sulfurados existentes, acarretando a detecção de compostos diferentes dos originalmente presentes na amostra em análise (LAWSON; WANG; HUGHES, 1991). Esse inconveniente é contornado quando se utiliza HPLC, porém, assim como na GC, para realização das análises é necessária a aquisição dos padrões de cada componente a ser investigado, o operador deve ser altamente qualificado e tanto o equipamento quanto sua manutenção são dispendiosos. Além disso, etapas de preparo das amostras são necessárias, tornando a análise ainda mais demorada e complexa.

Nesse contexto, devido à importância em se determinar a concentração de óleo de alho contida em cápsulas gelatinosas, tanto para estabelecer uma relação apropriada entre as pesquisas a respeito de sua funcionalidade, quanto para assegurar a qualidade desses produtos aos seus consumidores, faz-se necessário o desenvolvimento de métodos mais simples e eficientes para essa finalidade.

A espectroscopia no infravermelho médio (MIR), é uma técnica robusta e versátil que, associada a técnicas quimiométricas de reconhecimento de padrões e de calibração multivariada, vem sendo amplamente utilizada no

desenvolvimento de métodos analíticos que possam ser utilizados no controle de qualidade de óleos comestíveis. A substituição dos métodos clássicos de análise de seus parâmetros de qualidade, como índice de iodo, acidez, anisidina, saponificação, conteúdo de ácidos graxos e *trans*, por métodos baseados na espectroscopia MIR e na análise multivariada dos dados, apresenta diversos benefícios como a análise simultânea de mais de um parâmetro, rapidez nas análises, evita ou reduz a necessidade de preparo de amostras e a utilização de reagentes perigosos, diminui a quantidade de descartes, etc.

Face ao exposto, a espectroscopia MIR pode ser de grande valia para a avaliação do teor de óleo de alho em cápsulas gelatinosas, quer seja por meio de análise *screening*, empregando técnicas de reconhecimento de padrões supervisionadas, para identificar em que faixas de teor de óleo de alho as cápsulas estão contidas, ou por meio de técnicas de calibração multivariada para sua determinação quantitativa.

1.1. Objetivo

1.1.1. Objetivo geral

Desenvolver métodos baseados em espectroscopia no infravermelho médio (*mid infrared* - MIR), para classificar e quantificar cápsulas gelatinosas de óleo de alho de diferentes marcas quanto ao seu teor, por meio de técnicas de reconhecimento de padrões e de calibração multivariada.

1.1.2. Objetivos específicos

- Realizar análise exploratória dos dados, utilizando a técnica de reconhecimento de padrões não supervisionada PCA, com o intuito de avaliar a tendência natural de agrupamento das amostras e identificar regiões espectrais mais informativas para classificar e quantificar as cápsulas quanto ao teor de óleo de alho.

- Classificar diferentes marcas de cápsulas gelatinosas disponíveis comercialmente quanto ao teor de óleo de alho, utilizando espectros obtidos no infravermelho médio e as técnicas de reconhecimento de padrões supervisionadas PLS-DA e SPA-LDA.
- Validar os modelos de classificação utilizando sensibilidade, especificidade e exatidão como parâmetros de desempenho.
- Quantificar o teor de óleo de alho em diferentes marcas de cápsulas gelatinosas, utilizando espectroscopia no infravermelho médio e as técnicas de calibração multivariada PLS, SPA-MLR e MCR-ALS com restrição de correlação.
- Validar os modelos de calibração em função da exatidão (obtida por meio de ensaios de recuperação), dos limites de detecção e de quantificação e em função da presença de erros sistemáticos (*bias*), por meio do teste t.

CAPÍTULO 2

FUNDAMENTAÇÃO TEÓRICA



Capítulo 2. FUNDAMENTAÇÃO TEÓRICA

2.1. Alho

Pertencente à família Alliaceae, que compreende aproximadamente 700 espécies (ROSE *et al.*, 2005; AVIELLO *et al.*, 2009), o alho ordinário, cujo nome científico é *Allium sativum* L., é nativo da Ásia Central (PETROVSKA; CEKOVSKA, 2010). Apesar da China ser o maior produtor na atualidade, seu cultivo difundiu-se entre os continentes, sendo produzido em vários países como Estados Unidos, Espanha, Egito, Índia, República da Coreia, Argentina, Brasil, etc. A produção em toneladas no ano de 2014 dos dez maiores produtores de alho é apresentada na **Tabela 2.1**. O Brasil está em 13º lugar no ranking com uma produção de 93.679 toneladas (MAPS OF WORLD, 2017).

Tabela 2.1. Produção em 2014 dos dez maiores produtores de alho.

Ranking	País	Produção em toneladas, 2014
1º	China	19.984.724
2º	Índia	1.252.000
3º	República da Coreia	353.761
4º	Bangladesh	312.000
5º	Egito	263.167
6º	Rússia	256.406
7º	Birmânia	208.900
8º	Ucrânia	191.140
9º	Espanha	177.420
10º	Estados Unidos	175.450
11º	Uzbequistão	154.130
12º	Argentina	148.953
13º	Brasil	93.679

Fonte: Adaptado de MAPS OF WORLD (2017).

O alho é utilizado e reconhecido como medicamento natural contra inúmeras doenças há mais de 5.000 anos (QIDWAI; ASHFAQ, 2013). As inscrições mais antigas a respeito da existência do alho foram encontradas em tumbas egípcias. Antes de sua evolução mercantil, o alho era a planta mais utilizada pelos egípcios como medicamento e, após seu enriquecimento, continuaram utilizando-o para alimentar seus escravos, como forma de suprir a necessidade de vitaminas e fortalecê-los para o trabalho.

Na Roma Antiga, utilizavam alho para curar tuberculose, febre, para regular o funcionamento do intestino, contra cólicas e para evitar mordidas de cobra, espalhando seu suco pelo corpo (AMAGASE *et al.*, 2001). Os gregos tratavam gangrena e doenças do trato respiratório com alho, devido a seu efeito bactericida, usavam-no como laxante, diurético, contra parasitas intestinais e seus atletas utilizavam-no como parte essencial de sua dieta, pois acreditavam que esse alimento conferia força e resistência física (AVIELLO *et al.*, 2009; PETROVSKA; CEKOVSKA, 2010).

Na África, pescadores esfregavam seus corpos com extrato de alho para protegê-los contra crocodilos (MAJEWSKI, 2014). Os antigos Israelitas usavam alho para estimular a fome, equilibrar a pressão sanguínea e contra parasitas (PETROVSKA; CEKOVSKA, 2010). Em antigos textos chineses, o consumo regular de alho foi descrito como a razão da longevidade de seu povo (AVIELLO *et al.*, 2009). Na Rússia foi usado durante a primeira guerra mundial para tratar gangrenas e durante a segunda guerra mundial foi chamado de “penicilina russa” devido a seu efeito bactericida (AVIELLO *et al.*, 2009).

Embora a medicina natural muitas vezes seja utilizada baseada na cultura e crendice popular, a potencialidade do alho quanto às propriedades antissépticas foi confirmada em diferentes momentos da história. Em 1858, Louis Pasteur reconheceu que o alho era capaz de matar bactérias tais como *Helicobacter pylori*, responsável por doenças no estômago como úlcera, gastrite e câncer (PETROVSKA; CEKOVSKA, 2010). Entre 1913 e 1918 o alho mostrou-se eficaz no tratamento da cólera, febre tifoide, difteria e na prevenção da gripe espanhola, (PETROVSKA; CEKOVSKA, 2010). Mais recentemente, inúmeras pesquisas científicas têm demonstrado que o alho é um poderoso remédio

natural e pode ser utilizado como agente hipocolesteremiante, hipolipemiante, antioxidante, hipotensor, hipoglicemiante, antitrombótico, anticarcinogênico, bactericida e fungicida (AGARWAL, 1996; KREST; KEUSGEN, 1999; AMAGASE *et al.*, 2001; BOREK, 2001a, 2006; HASSAN, 2004; REINHART *et al.*, 2008; TOULOUPAKIS; GHANOTAKIS, 2010; ILIC *et al.*, 2011; REHMAN; SAMYA, 2012; SINGH; KUMAR; DHAKAL, 2015). Logo, ele atua na prevenção de alguns tipos de câncer, resfriados, diabetes, acidente vascular cerebral, doenças cardíacas, arteriosclerose, no tratamento de infecções por diversas bactérias gram positivas e negativas etc.

Atualmente, o alho é considerado alimento funcional pela ANVISA (AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA), definido como aquele alimento que, além de oferecer funções nutricionais básicas, quando se tratar de nutriente, produz efeitos metabólicos e ou fisiológicos e ou benéficos à saúde, sem necessidade de supervisão médica (BRASIL. ANVISA - AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA, 1999b). Contém teores elevados de alguns minerais como fósforo, potássio, enxofre e zinco, teores moderados de selênio e baixos teores de cálcio, magnésio, sódio, ferro e manganês. Possui ainda em sua composição vitaminas A, B e C, além de uma variedade de aminoácidos (AGARWAL, 1996; RAHMAN, 2003). Porém, as propriedades benéficas do alho estão associadas, principalmente, aos compostos organosulfurados presentes em sua composição e aos produtos formados decorrentes de sua manipulação, bem como da sinergia entre eles (BLOCK, 1985; RAHMAN, 2007; RATTI *et al.*, 2007; AVIELLO *et al.*, 2009; BUTT *et al.*, 2009; PETROVSKA; CEKOVSKA, 2010; GUERCIO *et al.*, 2014).

A aliina (sulfóxido de (+)-S-alil-L-cisteína) é o principal composto organosulfurado presente no alho intacto, entre os sulfóxidos que o compõem. Em média, um bulbo de alho possui aproximadamente 1% de aliina (AMAGASE *et al.*, 2001) e, quando submetido a processos físicos, ou seja, cortado, esmagado ou triturado, a membrana que separa a aliina da enzima alinase se rompe e ocorre uma reação entre esses dois compostos dando origem à alicina (dialil tiosulfonato) (BLOCK, 1985), conforme **Figura 2.1**. Assume-se que o

potencial natural de formação de alicina é aproximadamente 2,5 – 6,6 µg/g de aliina (QUEIROZ, 2010; MAJEWSKI, 2014).

A alicina corresponde a 70 – 80% dos tiosulfatos formados quando o alho é danificado (LAWSON, 1993; AMAGASE *et al.*, 2001). Foi isolada pela primeira vez em 1944, consistindo em um líquido incolor oleoso de intenso odor e com forte propriedade bactericida, até mesmo contra bactérias resistentes à penicilina (CAVALLITO; BAILEY, 1944). Sua fórmula química foi determinada apenas em 1947, mesmo ano em que sua substância precursora, a aliina, foi isolada (BLOCK, 1985; PETROVSKA; CEKOVSKA, 2010).

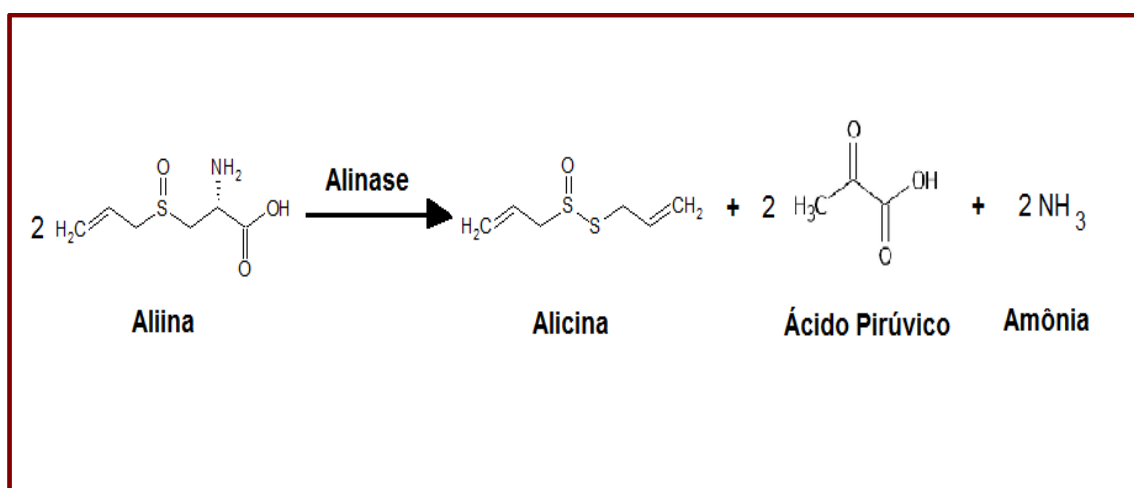


Figura 2.1. Formação da alicina a partir da aliina e alinase. Fonte: adaptado de MENDES, 2008.

Apesar de a alicina ser considerada a principal responsável pelas propriedades benéficas à saúde atribuídas ao alho (TOULOUPAKIS; GHANOTAKIS, 2010), é um composto muito instável que, a temperatura ambiente ou sob aquecimento, rapidamente sofre degradação, dando origem a uma variedade de compostos organosulfurados mais estáveis que, assim como a alicina, também apresentam os vários efeitos salutíferos mencionados anteriormente (KIM *et al.*, 1995; LEE; KIM; LEE, 2003; LANZOTTI, 2006; MAJEWSKI, 2014). Entre esses compostos temos 2-vinil-[4H]-1,3-ditiina, 3-vinil-[4H]-1,2-ditiina, E-ajoeno, Z-ajoeno, mono, di e trissulfetos, cujas estruturas estão representadas na **Figura 2.2**.

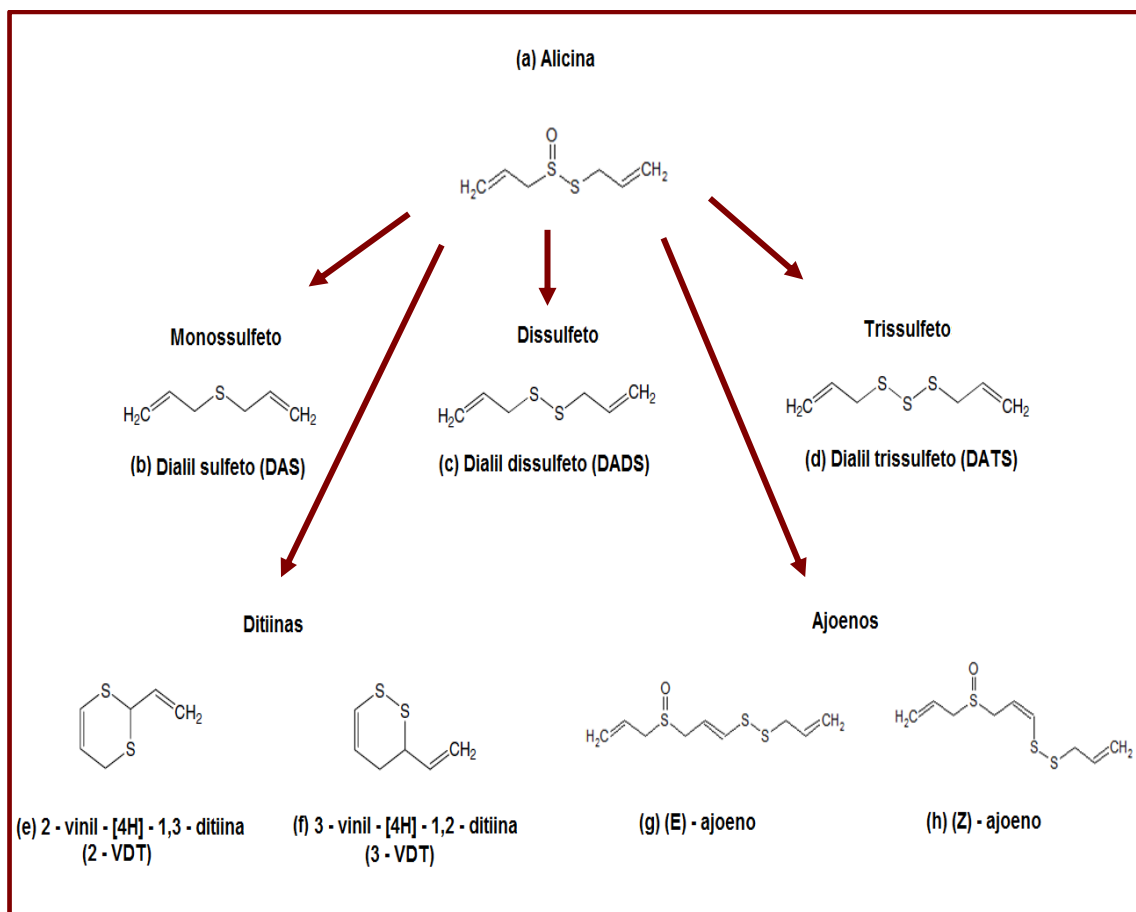


Figura 2.2. Compostos sulfurados formados a partir da degradação da (a) alicina: (b) dialil sulfeto (DAS), (c) dialil dissulfeto (DADS), (d) dialil trissulfeto (DATS), (e) 2-vinil-[4H]-1,3-ditiina (2-VDT), (f) 3-vinil-[4H]-1,2-ditiina (3-VDT), (g) (E)-ajoeno, (h) (Z)-ajoeno.

2.2. Suplementos de alho

Apesar da ingestão regular do alho proporcionar diversos benefícios à saúde, seu sabor e odor característico tornam seu consumo muitas vezes desagradável. Para driblar esses inconvenientes, produtos à base de alho de diferentes marcas estão comercialmente disponíveis, sendo encontrados na forma de pó, óleo ou extrato de alho envelhecido. Esses produtos possuem diferentes composições, variando de acordo com o processo produtivo adotado, mas nenhum deles possui alicina, visto ser um composto altamente instável e que rapidamente se decompõe em outros compostos organosulfurados mais estáveis (AMAGASE *et al.*, 2001; TOULOUPAKIS; GHANOTAKIS, 2010).

Comprimidos e cápsulas de pó de alho são produzidos a partir da desidratação e pulverização de dentes de alho. Sua composição idealmente deveria ser igual ao alho intacto caso o processo necessário para convertê-lo em pó fosse aplicado ao alho sem que fosse necessário cortá-lo. Porém, para melhorar a eficiência da secagem, os dentes de alho são cortados e em seguida secos a aproximadamente 50-60°C. Logo, além da aliina e da enzima alinase, como componentes principais, os comprimidos e cápsulas de alho em pó contém pequena quantidade de alguns compostos sulfurados resultantes da degradação da alicina, formada na etapa de corte (LAWSON, 1993; AMAGASE *et al.*, 2001). Um inconveniente do alho em pó é que a enzima alinase é inativada em $\text{pH} \leq 3$, que é o pH do estômago. Assim, não há formação de alicina e dos compostos sulfurados resultantes de sua degradação, responsáveis pelas propriedades benéficas à saúde atribuídas ao alho. Logo, os produtos à base de pó de alho devem possuir revestimento entérico que se dissolve apenas no intestino, cujo pH é favorável à formação da alicina (LAWSON, 1993; AMAGASE *et al.*, 2001). Porém, essa é uma tecnologia muito imprecisa que gera produtos não padronizados e cuja quantidade de alicina realmente produzida é difícil de ser determinada (MAPS OF WORLD, 2017).

O extrato de alho envelhecido (AGE) é obtido armazenando-se dentes de alho picados ou inteiros em uma solução extratora (15-20% etanol) por longo período de tempo, aproximadamente de 6 a 20 meses (LAWSON, 1993; YU-YAN; LIJUAN, 2001; QIDWAI; ASHFAQ, 2013). Depois de decorrido esse tempo, o líquido obtido é filtrado, concentrado e comercializado na forma de pó ou inserido em cápsulas, possuindo composições muito semelhantes nas duas formas (LAWSON, 1993). O extrato obtido contém principalmente os compostos sulfurados solúveis em água S-alil mercaptocisteína (SAMC) e S-alilcisteína (SAC), que possuem poder antioxidante maior do que o alho fresco ou outros suplementos à base de alho (BOREK, 2001b) e pequena quantidade de compostos sulfurados solúveis em óleo. O SAC está presente em maior quantidade e é utilizado para padronizar os produtos à base de extrato de alho envelhecido (AMAGASE *et al.*, 2001; QIDWAI; ASHFAQ, 2013; MAJEWSKI, 2014). Entre os produtos de alho comercialmente disponíveis, o AGE é o que apresenta resultados mais consistentes quanto às propriedades benéficas à

saúde e o único que possui estudos de toxicidade e segurança (AMAGASE *et al.*, 2001) . Provavelmente isso se deve ao fato de ser o único dos suplementos cujo conteúdo é realmente padronizado, tendo sido já objeto de estudo de centenas de pesquisas científicas (AMAGASE *et al.*, 2001; MAJEWSKI, 2014).

Os produtos à base de óleo de alho são comercializados na forma de cápsulas gelatinosas. A extração do óleo para confecção das cápsulas se dá por duas diferentes técnicas, cada uma delas gerando composições distintas. A maior parte das cápsulas encontradas comercialmente possui óleo de alho extraído por destilação a vapor (KIM *et al.*, 1995), chamado equivocadamente de óleo essencial de alho, pois sua composição consiste exclusivamente de sulfetos gerados pelo aquecimento de seus precursores tiosulfatos, que são formados apenas após o corte ou esmagamento do alho. O óleo de alho extraído por destilação a vapor consiste aproximadamente de 26% de dissulfeto de dialila, 19% de trissulfeto de dialila, 15% de trissulfeto de metil alila, 13% de dissulfeto de metil alila, 8% de tetrassulfeto de dialila, 6% de tetrassulfeto de metil alila, 3% de trissulfeto de dimetila, 3% de monossulfetos e 5% de polissulfetos. As cápsulas de óleo de alho obtido por essa técnica contêm apenas 1% de óleo de alho e 99% de óleo vegetal, principalmente óleo de soja. Essa diluição é feita de forma a se obter cápsulas de tamanho prático para consumo e para estabilizar os sulfetos, que tendem a formar gradualmente polímeros de sulfetos na ausência do solvente (LAWSON, 1993).

A outra forma de extração utilizada para obtenção de óleo de alho para confecção de cápsulas direcionadas a fins relacionados à saúde é a maceração com óleo vegetal. Nessa técnica o alho cortado é amassado em óleo vegetal, sendo o óleo de soja o mais utilizado para essa finalidade. Após homogeneização, os componentes insolúveis em óleo são separados e obtém-se óleo de alho composto de vinil ditiinas (70%), sulfetos (18%) e ajoenos (12%). Os produtos comercialmente disponíveis de óleo de alho macerado apresentam, entretanto, apenas 20% dos compostos organosulfurados contidos nas cápsulas cujo óleo foi obtido por destilação a vapor. Apesar de não se saber com clareza o motivo da diferença de rendimento desses compostos, acredita-se que seja

devido a perdas por volatilização durante o processo e à formação insuficiente de tiosulfatos antes de sua maceração (LAWSON, 1993).

Embora não haja dosagem padrão para o consumo de alho fresco e seus suplementos, em 1988 German Kommission E monograph propôs, mesmo sem embasamento científico, que o consumo de 1 a 2 dentes de alho (correspondendo a aproximadamente 4g de alho) por dia podem proporcionar benefícios à saúde. Em estudos mais recentes, recomenda-se a utilização diária de aproximadamente 0,6 g a 1,2 g de pó de alho desidratado, 1,0 g a 7,2 g de extrato de alho envelhecido e de 2,0 mg a 5,0 mg de óleo de alho (AMAGASE *et al.*, 2001; HELOU; HARRIS, 2007; DRUGS, 2017).

2.3. Análise da composição de alho e de seus produtos por GC e HPLC

Artigos acadêmicos têm relatado o estudo da composição do alho fresco e de seus produtos, bem como a diferença de composição de acordo com o método de extração utilizado para obtenção do óleo de alho, por meio das técnicas cromatográficas GC e HPLC (YU; WU; LIOU, 1989; LEE; KIM; LEE, 2003; ABU-LAFI *et al.*, 2004; CALVO-GÓMEZ; MORALES-LÓPEZ; LÓPEZ, 2004; KIMBARIS *et al.*, 2006). Nesse sentido, Yu, Wu e Liou (1989), determinaram as composições de amostras de óleos de alho extraídos por destilação em água e destilação a vapor, seguidos ou não por extração por solvente, utilizando GC (com detector FID e FPD) e GC-MS. Os autores concluíram que, em todos os casos, os componentes majoritários nos extratos obtidos foram monossulfetos, dissulfetos e trissulfetos, com pequenas quantidades dos compostos cíclicos 2-vinil-4H-1,3-ditiina e 3-vinil-4H-1,2-ditiina.

Abu-Lafi *et al.* (2004) determinaram a composição de frações de alho extraídas por diferentes solventes por GC-MS, sob diferentes condições cromatográficas, e por HPLC. Em todos os casos, 2-vinil-4H-1,3-ditiina foi encontrado como componente majoritário nos extratos e, quando o extrato foi submetido a aquecimento em água fervente, houve decomposição desse composto e de 3-vinil-4H-1,2-ditiina em outros compostos organosulfurados, como dialil trissulfeto e dialil dissulfeto.

Lee, Kim e Lee (2003), utilizaram 4 diferentes técnicas de extração para determinar a composição do alho por GC-MS, obtendo diferentes resultados para cada técnica. Enquanto para destilação a vapor (SD) os autores encontraram 89,77%, 2,43% e 3,89% de dialil dissulfeto, alil sulfeto e dialil trissulfeto, respectivamente, para destilação a vapor seguida de extração por solvente (SDE) encontraram 57,88%, 23,59% e 11,40%, respectivamente. Para a extração em fase sólida, as quantidades obtidas foram 97,77%, 0,17% e 0,10% dos mesmos compostos, respectivamente, e para microextração em fase sólida por *headspace* (HS-SPME), 97,85%, 0,01% e 0,01% dos 3 compostos respectivamente. Apenas para SDE foi encontrada cerca de 0,07% de 2-vinil-4H-1,3-ditiina, para as outras técnicas esse composto não foi detectado.

Calvo-Gómez, Morales-López e López (2004), desenvolveram a hidrodestilação em baixa pressão para extração do óleo de alho e analisaram sua composição por GC-MS, com amostragem por SPME (micro extração em fase sólida). A identificação dos compostos analisados foi realizada comparando seus espectros de massa com os dos compostos autênticos, disponíveis em bases de dados ou em artigos científicos. Dialil dissulfeto foi encontrado como componente majoritário, seguido por dialil trissulfeto e metil alil trissulfeto.

Kimbaris *et al.* (2006) apresentaram um estudo comparativo da composição de óleo de alho obtido pelas técnicas de extração SDE (destilação seguida de extração por solvente), MWHD (hidrodestilação assistida por micro-ondas) e USE (extração por ultra-som). Para cada técnica, 3 solventes foram testados e a composição de cada óleo obtido foi determinada quantitativamente pela técnica GC (com detector FID), e qualitativamente por GC-MS, por comparação dos espectros de massa obtidos com espectros disponíveis em bases de dados. As duas técnicas de análise revelaram que os óleos produzidos pelos três métodos de extração são claramente similares em termos qualitativos, mas mostram grandes diferenças quantitativas. Com a mudança do método de extração, ocorreram mudanças drásticas na porcentagem dos sulfetos majoritários, que diminuíram de 77% para 50,9% e para 9,4%, para os óleos obtidos por SDE, MWHD e USE, respectivamente. Por outro lado, a mudança do método de extração ocasionou um aumento significativo nas ditiinas cíclicas, de

9,5% para 33,66% e para 70,8%, para os óleos obtidos por SDE, MWHD e USE, respectivamente. A partir desses resultados, os autores concluíram que esses compostos cíclicos (2-vinil-4H-1,3-ditiina e 3-vinil-4H-1,2-ditiina) são os componentes majoritários em dentes de alho recém cortados e se decompõem em vários sulfetos quando submetidos a tratamentos térmicos como destilação a vapor, ou mesmo sob as condições da análise por cromatografia gasosa. Logo, o aquecimento é o fator crucial para a diferenciação dos compostos presentes em óleo de alho.

Apesar dos diversos artigos relatando a composição do óleo de alho, quase a totalidade dos trabalhos publicados é referente à análise em amostras de óleo de alho extraído em laboratório, para fins de estudo. Apenas três artigos que investigam a composição do óleo de alho contido em cápsulas gelatinosas comercialmente disponíveis foram encontrados, com dois deles empregando GC e o outro, HPLC (LAWSON; WANG; HUGHES, 1991; YAN; WANG; BARLOW, 1992, YAN; WANG; BARLOW, 1993).

Lawson, Wang e Hugues (1991) determinaram o teor de alicina e de seus produtos de degradação por HPLC-UV, usando seus respectivos coeficientes de extinção, para diversos produtos de alho comercialmente disponíveis, incluindo óleo de alho obtido por destilação a vapor e por maceração, alho em pó e extrato de alho envelhecido (AGE). Para isso, os compostos sulfurados foram extraídos dos produtos de alho utilizando a acetonitrila (extração de sulfetos, ditiinas e ajoenos) e água (tiosulfinaos). Diferentes condições cromatográficas foram necessárias (colunas e eluentes) para separar os diversos compostos obtidos a partir da degradação da alicina. Os produtos à base de óleo de alho destilado a vapor apresentaram composição e concentrações dos compostos sulfurados significativamente diferentes dos produtos à base de óleo de alho macerado, conforme esperado. Os resultados obtidos por HPLC foram comparados com a técnica GC, com duas diferentes colunas. Por HPLC, foi possível detectar (E) e (Z)-ajoenos, penta e hexassulfetos em concentrações muito superiores às obtidas por GC, demonstrando ser uma técnica mais adequada à determinação de compostos termicamente instáveis como os encontrados em produtos à base de alho. A partir do método proposto, os autores verificaram que algumas marcas

de óleo de alho destilado a vapor apresentaram quantidade de óleo de alho muito inferior ao esperado. Conforme os autores, essa diferença não se deve à instabilidade, visto que a composição desses óleos permanece inalterada por, ao menos 2 anos, à temperatura ambiente. O mesmo ocorreu para os óleos macerados. Os produtos à base de alho em pó apresentaram teor de alicina de 4% a 50% do valor esperado para esses produtos, e no único produto à base de extrato de alho envelhecido analisado, não foi detectada a presença de alicina ou de seus produtos de degradação.

Yan, Wang e Barlow (1992) utilizaram a técnica GC-FID para determinar o teor de óleo de alho de produtos comerciais à base de óleo de alho destinados à fins medicinais. Para isso, os sulfetos existentes nas amostras foram separados do óleo vegetal por extração em fase sólida empregando um cartucho C₁₈ e acetonitrila (eluente), e foram identificados e quantificados por GC-FID, utilizando como padrão interno dissulfeto de dipropil. As proporções padrão dos três sulfetos majoritários (dialil dissulfeto, alil metil trissulfeto e dialil trissulfeto) foram determinadas em óleo de alho puro e corrigidas por um fator de resposta relativo, obtido em função do número de carbono efetivo encontrado para cada sulfeto. Um aumento de 650 mg g⁻¹ para 795 mg g⁻¹ do teor total de sulfetos no óleo de alho foi obtido quando este fator foi utilizado. Considerando que, a composição do óleo de alho dos produtos comerciais destinados à saúde é igual à composição de óleo de alho puro, a média da porcentagem em massa desses três sulfetos foi utilizada, então, para determinar o teor de óleo de alho nas cápsulas. Valores muito próximos aos indicados nos rótulos dos produtos comerciais foram encontrados.

Yan, Wang e Barlow (1993) também utilizaram GC-FID, empregando dois diferentes preparos de amostras. Para determinar o teor total de compostos de enxofre, foi realizada a homogeneização de óleo de alho e alho em pó (provenientes de cápsulas comerciais e comprimidos, respectivamente) com água, a temperatura ambiente por 30 minutos, com posterior aquecimento a 100°C, permitindo que os compostos organosulfurados se convertessem totalmente em sulfetos. No segundo preparo de amostra, os produtos comerciais de alho foram homogeneizados com iso-octano e aquecidos a 100°C por 30 min,

para conversão da alicina formada em vinil-ditiinas e sulfetos. Ambos os preparos resultaram em teores de sulfetos muito similares para a formulação à base de óleo de alho. No entanto, para a formulação à base de alho em pó, nenhum sulfeto ou vinil-ditiina foi detectado no preparo com iso-octano, demonstrando que a aliina é o principal composto nessas formulações, conforme esperado. Quando o preparo foi realizado com água, a aliina se converteu em alicina, sendo possível determinar os sulfetos e vinil-ditiinas formados por sua degradação.

Conforme se pode observar, as análises realizadas pelos métodos desenvolvidos utilizando as técnicas cromatográficas GC e HPLC, são demoradas, necessitam de padrões específicos para determinação individual de cada componente e há necessidade da etapa de preparo das amostras que, no caso de amostras de óleo de alho, consiste na separação dos compostos organosulfurados do óleo vegetal. Para driblar esses inconvenientes, a espectroscopia MIR aliada a técnicas quimiométricas, pode ser empregada na avaliação do teor de óleo de alho em cápsulas gelatinosas, em função dos princípios ativos que o compõem, garantindo uma equivalência entre as diversas pesquisas na área clínica, de modo a possibilitar conclusões acertadas e definitivas a respeito da funcionalidade do óleo de alho.

2.4. Espectroscopia no infravermelho

A espectroscopia no infravermelho é baseada na absorção ou emissão de energia radiante na região do espectro eletromagnético que vai de 12800 cm^{-1} a 10 cm^{-1} ($0,78\text{ }\mu\text{m}$ a $15\text{ }\mu\text{m}$) e é dividida em três sub-regiões: infravermelho próximo (12800 cm^{-1} a 4000 cm^{-1}), infravermelho médio (4000 cm^{-1} a 200 cm^{-1}) e infravermelho distante (200 cm^{-1} a 10 cm^{-1}). Essa radiação não é suficientemente energética para provocar transições eletrônicas, no entanto, causa transições de moléculas de estados de energia vibracional ou rotacional para outros, fornecendo informações qualitativas e quantitativas úteis, traduzidas em seus espectros. Para que uma molécula absorva a radiação infravermelha, esta deve possuir frequência igual à frequência de vibração da molécula e deve

sofrer variação no momento de dipolo durante seu movimento rotacional ou vibracional (COATES, 2000). Logo, espécies homonucleares como O₂, N₂ ou Cl₂ não absorvem radiação nesta região (HOLLER; SKOOG; CROUCH, 2009).

2.4.1 Espectroscopia no Infravermelho Médio (MIR)

A espectroscopia MIR é amplamente utilizada na identificação de compostos orgânicos. Os grupos funcionais presentes nas moléculas apresentam picos ou bandas características na região conhecida como frequência de grupos que compreende a faixa de 3600 cm⁻¹ a 1250 cm⁻¹ (PAVIA *et al.*, 2010). Na faixa espectral de aproximadamente 1200 cm⁻¹ a 600 cm⁻¹, conhecida como região de impressão digital, pequenas diferenças na estrutura e constituição de uma molécula dão origem a alterações significativas das bandas de absorção (HOLLER; SKOOG; CROUCH, 2009). Logo, as atribuições das bandas dos grupos funcionais de uma molécula associadas à comparação espectral detalhada na região de impressão digital com espectros de compostos que contenham todos os grupos funcionais encontrados na região de frequência de grupos, torna possível a identificação de compostos desconhecidos (COATES, 2000; HOLLER, SKOOG, CROUCH, 2009). Além de análises qualitativas, a espectroscopia MIR também é aplicada às determinações quantitativas, pois as intensidades das bandas são proporcionais às concentrações dos respectivos grupos funcionais.

Os principais modos de medidas são por transmitância e reflectância (difusa, especular e reflexão total atenuada – ATR). Esse último modo de medida é um dos mais utilizados na atualidade e baseia-se na reflexão total da luz infravermelha quando esta passa de um meio de maior índice de refração (cristal como seleneto de zinco, diamante, germânio, etc.) para outro de menor índice (amostra em análise). Durante o processo de reflexão, o feixe chamado de onda evanescente penetra levemente na amostra, e sofre perda de energia naqueles comprimentos de onda em que o material absorve (KAROUI; DOWNEY; BLECKER, 2010; RODRIGUEZ-SAONA; ALLENDORF, 2011). A intensidade da radiação é atenuada devido às múltiplas reflexões ao longo do comprimento da

amostra, ou seja, ocorre refletância total atenuada e um espectro de superfície é produzido. Esta técnica permite a obtenção de espectros de sólidos, líquidos, pastas, pós e filmes finos com pouco ou nenhum preparo de amostra, de maneira rápida, não destrutiva e diferentemente das análises por transmitância, independe da espessura da amostra.

Dois tipos de espectrômetros são utilizados na espectroscopia no infravermelho médio: os dispersivos e os com transformada de Fourier (FT-MIR), baseados no interferômetro de Michelson (HOLLER; SKOOG; CROUCH, 2009). Os espectrômetros com transformada de Fourier apresentam inúmeras vantagens sobre os dispersivos, tais como aquisição simultânea de todas as frequências, reduzindo sobremaneira o tempo de análise; maior sensibilidade; melhor aproveitamento da potência luminosa; melhor razão sinal/ruído; melhor resolução de comprimentos de onda e etc. (VAN DE VOORT, 1992; GUILLEN; CABO, 1997). Devido a essas inúmeras vantagens técnicas e ainda por serem mais baratos do que os dispersivos, hoje a maioria dos espectrômetros de infravermelho médio são FT-MIR.

2.4.1.1. Aplicações do MIR

A existência da radiação infravermelha foi descoberta e demonstrada em 1800 pelo astrônomo William Herschel e desde o início, a potencialidade da espectroscopia no infravermelho médio (MIR) para a elucidação de estruturas químicas foi comprovada. Os primeiros experimentos mostraram que cada composto possui um espectro único e que, um determinado grupo funcional, presente em diferentes moléculas, exibe bandas de absorção localizadas aproximadamente no mesmo comprimento de onda. Porém, os primeiros instrumentos de infravermelho chegaram aos laboratórios industriais apenas durante a década de 1930, com o desenvolvimento de detectores fotoelétricos que tornaram possível registrar seus espectros.

Ao final da 2ª Guerra Mundial, a espectroscopia IR estava sendo usada em dezenas de indústrias e laboratórios federais nos Estados Unidos e Grã-

Bretanha, bem como em pesquisas sobre análises de misturas orgânicas e inorgânicas complexas em diversas universidades (RABKIN, 1987). Porém, a popularidade da técnica só ocorreu com a chegada de instrumentos usando transformada de Fourier (FTIR - *Fourier Transform Infrared Spectroscopy*), que possuíam inúmeras vantagens sobre os dispersivos, já mencionadas anteriormente (**seção 2.3.1**). Após sua introdução no mercado em meados de 1960, os equipamentos FTIR sofreram intenso desenvolvimento em sua instrumentação e, com a evolução da tecnologia computacional, essa técnica permite que a análise espectral de dados hoje, seja realizada em segundos, uma tarefa que levava horas há algumas décadas (SUN, 2009).

Aliado aos avanços na instrumentação e tecnologia computacional, o desenvolvimento na área da análise multivariada de dados, tornou a técnica espectroscópica FTIR uma ferramenta poderosa e amplamente empregada na análise de alimentos, fármacos, indústrias petroquímicas, setores agrícolas e diversos outros setores industriais. A análise de alimentos por espectroscopia FTIR na região MIR é uma área relativamente recente e que pode ser intensamente beneficiada por essa técnica, visto que os alimentos são constituídos principalmente de lipídios, carboidratos, proteínas e água, compostos que apresentam bandas características de absorção nesta região do espectro eletromagnético.

Nas últimas três décadas o número de pesquisas nessa área tem crescido significativamente, voltadas principalmente ao estudo de interações entre componentes de um alimento, à quantificação de nutrientes e outros componentes específicos, à caracterização de moléculas, à determinação da qualidade de matérias-primas e aditivos utilizados no setor alimentício, à detecção e quantificação de adulterantes (SUN, 2009). Pesquisas na área abrangem uma diversidade de alimentos tais como carnes, peixes, frutas e verduras, cereais, leite e derivados, vinhos, cervejas, ovos, açúcar, mel, café e óleos comestíveis, entre outros (VAN DE VOORT, 1992; KAROUI; DOWNEY; BLECKER, 2010; RODRIGUEZ-SAONA; ALLENDORF, 2011; VONGSVIVUT *et al.*, 2012; CRAIG; FRANCA; OLIVEIRA, 2012; HERNÁNDEZ-MARTÍNEZ *et al.*, 2013; BASSBASI *et al.*, 2014; CAPUANO *et al.*, 2014; ANJOS *et al.*, 2015;

RAHMANIA; SUDJADI; ROHMAN, 2015; CUSTERS *et al.*, 2015; GOK *et al.*, 2015; JAISWAL *et al.*, 2015; LI *et al.*, 2015; NUNES *et al.*, 2016; SANTOS *et al.*, 2016; RÍOS-REINA *et al.*, 2017). Entre esses alimentos, os óleos comestíveis constituem grande parte das pesquisas nessa área, sendo um dos alimentos mais frequentemente falsificados no mundo, conforme podemos verificar no **Gráfico 2.1** (NUNES, 2014).

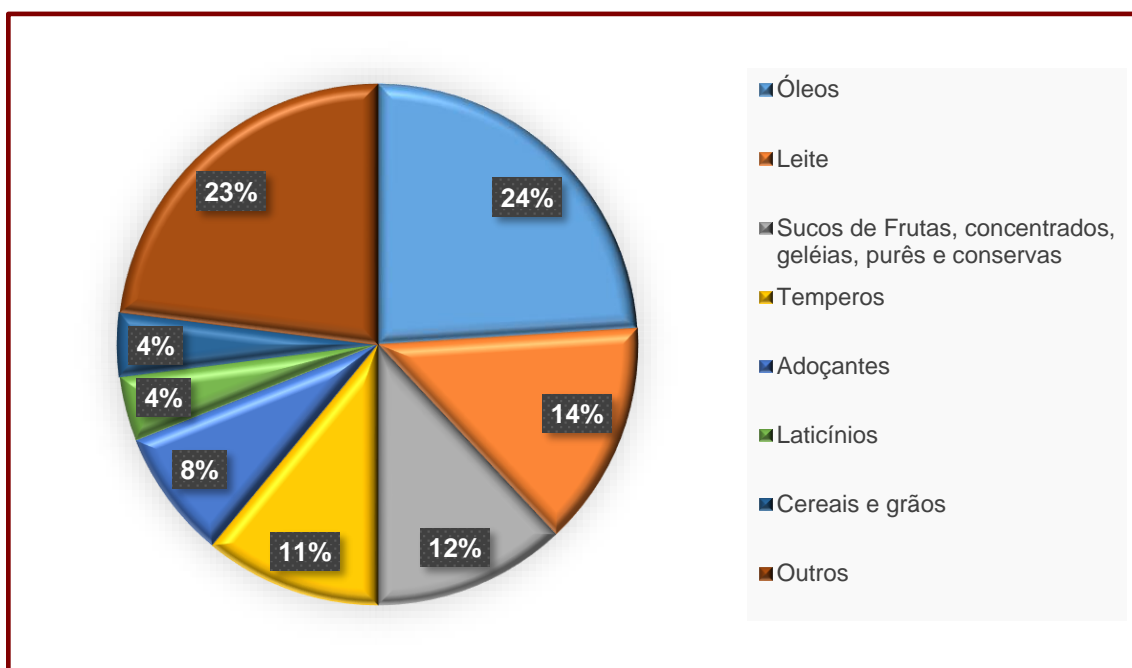


Gráfico 2.1. Porcentagem de casos de fraudes de 1980 a 2010. **Fonte:** adaptado de Nunes, 2014.

Existem vários métodos oficiais para atestar a qualidade de óleos comestíveis, porém, muitos desses métodos são demorados, custosos, usam reagentes perigosos e geram grande quantidade de descartes. Logo, a necessidade de desenvolver metodologias analíticas rápidas, de baixo custo e confiáveis promoveu o uso da espectroscopia MIR como uma ferramenta poderosa para detectar adulterações, atestar autenticidade e determinar parâmetros intrínsecos de qualidade em óleos comestíveis, tais como índices de iodo, acidez, peróxido, anisidina, saponificação e conteúdo de ácidos graxos *trans* (VAN DE VOORT, 1992; GUILLEN; CABO, 1997; KAROUI; DOWNEY; BLECKER, 2010; NUNES, 2014).

Para auxiliar a extração e interpretação de informações úteis obtidas por meio da espectroscopia MIR, técnicas quimiométricas de reconhecimento de

padrões não supervisionadas, supervisionadas e de calibração multivariada são utilizadas, tornando possível a avaliação de classes (amostras adulteradas/ não adulteradas ou autênticas/não autênticas) e quantificação de parâmetros de qualidade. Na próxima seção serão abordadas algumas das principais técnicas quimiométricas utilizadas.

2.5. Análise multivariada dos dados

A espectroscopia MIR fornece informações específicas dos compostos analisados em seus espectros, baseados nas posições, intensidades e formas de suas bandas. Entretanto, para extrair ao máximo informações químicas relevantes da grande quantidade de dados multidimensionais geradas por essa técnica, faz-se necessária a redução da dimensionalidade dos dados por meio de técnicas quimiométricas. Na atualidade, existe uma grande variedade de técnicas quimiométricas disponíveis para extrair informações qualitativas e quantitativas dos dados obtidos com a espectroscopia MIR, que inclui pré-processamento dos dados, técnicas de reconhecimento de padrões, calibração multivariada e seleção de variáveis espectrais (SUN, 2009).

2.5.1. Pré-processamento dos dados

Os espectros oriundos da espectroscopia no infravermelho geralmente são afetados por fenômenos químicos e físicos que causam desvios na linearidade entre a absorvância e a concentração das amostras em análise (SUN, 2009). Para superar essas dificuldades, torna-se necessário submeter os dados a pré-processamentos apropriados, que transformem os dados de maneira a melhor adequá-los à lei de Beer. Dessa forma, em geral, a etapa de pré-processamento de dados espectrais é de fundamental importância, quer seja para realizar análise exploratória dos dados, classificação ou para construção de modelos de calibração multivariada.

Entre os pré-processamentos mais utilizados na espectroscopia MIR temos *Multiplicative Scatter Correction* (MSC) e *Standard Normal Variate* (SNV). Ambos são utilizados com a finalidade de corrigir efeitos do espalhamento da radiação, muito comuns em análises realizadas por reflectância. O pré-processamento MSC, introduzido por Martens *et al.* (1983) e posteriormente elaborado por Geladi *et al.* (1985), corrige os espectros utilizando coeficientes de correção gerados por meio de regressão por mínimos quadrados de uma amostra de referência, que normalmente é a média dos espectros das amostras (SUN, 2009). Diferentemente de MSC, SNV não utiliza sinal de referência para correção dos espectros, ele transforma cada amostra isoladamente das demais, empregando a média e o desvio padrão de cada espectro como fatores de correção (BARNES; DHANOA; LISTER, 1989; SUN, 2009). Em geral, SNV e MSC fornecem resultados muito semelhantes, porém SNV está mais suscetível a ruídos, visto não aplicar regressão por mínimos quadrados para obtenção dos fatores de correção.

Outro pré-processamento muito utilizado é a suavização por filtros Savitzky Golay (SG), aplicado quando se deseja melhorar a relação sinal/ruído (SAVITZKY; GOLAY, 1964). Esse método ajusta os dados a polinômios de baixa ordem, substituindo pontos centrais de janelas de tamanhos previamente definidos pelos pontos estimados pelo polinômio. Dois parâmetros são de fundamental importância para que a suavização SG forneça melhor relação sinal/ruído sem que haja perda de informações. Um deles é o tamanho da janela a ser utilizado, que deve ser escolhido com prudência, pois tamanhos de janelas muito grandes levam à distorção dos dados, enquanto janelas muito pequenas não são capazes de eliminar o ruído de forma satisfatória. Outro parâmetro é a ordem do polinômio que será utilizada para ajuste dos dados. Como a escolha de ambos os fatores é estabelecida pela natureza dos dados, não havendo um padrão pré-determinado a ser seguido, recomenda-se que sejam testados diferentes tamanho de janela e ordens de polinômio e que a combinação desses dois fatores que fornecer melhores resultados seja selecionada (BEEBE; PELL; SEASHOLTZ, 1998; RINNAN; BERG; ENGELSEN, 2009; SUN, 2009).

Métodos para correção de desvios de linha de base também são extensamente aplicados a dados de espectroscopia MIR. Dentre eles, o mais utilizado é a derivação baseada no método de suavização Savitzky Golay (SAVITZKY; GOLAY, 1964). Nesse método, ao invés de substituir o ponto central de uma janela de dados por um ponto estimado pelo polinômio utilizado para ajustá-los, a substituição é feita pela derivada desse ponto. Dependendo da natureza dos dados, pode-se aplicar derivadas de diferentes ordens mas, para os dados espectroscópicos, em geral, derivadas de 1ª e 2ª ordem são suficientes para eliminar deslocamento de linha de base linear (efeito aditivo) - *offset* - e inclinação de linha de base (efeito multiplicativo), respectivamente. Assim como na suavização, a derivação SG exige cautela na escolha de seus parâmetros, para que a resolução de um problema não seja favorecida em detrimento de outro, como por exemplo, escolher um tamanho de janela suficientemente grande para melhorar a razão sinal/ruído, acarretando na perda de picos informativos (BEEBE; PELL; SEASHOLTZ, 1998; RINNAN; BERG; ENGELSEN, 2009; SUN, 2009).

Além dos métodos já citados, empregados para corrigir espalhamento de radiação, corrigir desvios de linha de base e melhorar a razão sinal/ruído, existem ainda aqueles que operam no domínio das variáveis. Nesse sentido, dois tipos de pré-processamentos são utilizados: centrar os dados na média e ponderação das variáveis (*variable weighing*) (BEEBE; PELL; SEASHOLTZ, 1998). A centralização dos dados na média consiste em calcular a média dos valores referentes a cada variável e subtraí-la de cada um dos elementos daquela variável. Desse modo, as coordenadas são movidas para o centro dos dados, permitindo melhor visualização de diferenças nas intensidades relativas das variáveis.

No tocante à ponderação das variáveis, escalonamento e auto escalonamento, apesar de serem amplamente utilizados em dados de diversas origens, não se aplicam a dados espectroscópicos, pois esses métodos atribuem os mesmos pesos a todos as variáveis, fazendo com que todas elas possuam a mesma influência na construção de modelos multivariados (SUN, 2009). Métodos úteis de ponderação de variáveis para dados espectroscópicos são

aqueles baseados no conhecimento *a priori* do sistema químico sob investigação, quando por exemplo se sabe que determinadas regiões do espectro são responsáveis pelos parâmetros que se quer determinar e pesos maiores são atribuídos a elas, ou quando existem regiões que possuem baixa relação sinal/ruído e podem ser eliminadas, sem prejuízo de informação útil (BEEBE; PELL; SEASHOLTZ, 1998).

A seleção de variáveis espectrais também é intensamente realizada por meio de diversos algoritmos existentes na literatura e apresenta-se como uma ferramenta muito útil na construção de modelos bem ajustados, mais parcimoniosos e que permitem uma melhor interpretação dos dados. Alguns dos algoritmos utilizados para essa finalidade são o algoritmo genético (GA), algoritmo das projeções sucessivas (SPA), regressão *stepwise* (STW), mínimos quadrados parciais por intervalo (iPLS), etc. (ARAÚJO *et al.*, 2001; NAES *et al.*, 2002; PONTES *et al.*, 2005; XIAOBO *et al.*, 2010). De uma maneira geral, esses métodos utilizam diferentes estratégias e critérios para escolher as regiões espectrais mais informativas e que levem a menores erros de previsão, quando se trata de modelos de calibração multivariada ou, maiores taxas de classificação correta, quando se trata de técnicas de reconhecimento de padrões supervisionadas. Porém, como cada algoritmo trabalha de forma diferente e foram desenvolvidos para diferentes aplicações, é demasiadamente difícil prever qual deles é mais adequado para um determinado conjunto de dados, portanto, resta ao usuário testar algumas alternativas e concluir qual delas gera melhores resultados.

A etapa de pré-processamento dos dados é imprescindível em uma análise multivariada e a escolha da técnica mais adequada para uma determinada matriz de dados vai determinar o sucesso dessa análise. De uma forma geral, essa etapa transforma/otimiza a matriz de dados de maneira a eliminar ou reduzir informações irrelevantes do ponto de vista químico, possibilitando que etapas seguintes como análise exploratória dos dados, construção de modelos de calibração multivariada ou de classificação, sejam realizadas com êxito.

2.5.2. Técnicas de reconhecimento de padrões

Para análise qualitativa dos dados, existem os métodos de reconhecimento de padrões que têm por finalidade encontrar similaridades e diferenças entre grupos de amostras que foram submetidas a algum tipo de análise. Essas técnicas são classificadas em dois grupos: reconhecimento de padrões não supervisionados e supervisionados. Os métodos de reconhecimento de padrões não supervisionados avaliam a existência de agrupamentos naturais sem utilizar o conhecimento prévio dos membros das classes e os mais conhecidos são Análise de Componentes Principais (*Principal Component Analysis* - PCA) e Análise por Agrupamento Hierárquico (*Hierarchical Cluster Analysis* - HCA). Já os métodos de reconhecimento de padrões supervisionados, através de conjuntos de treinamento com objetos de categorias conhecidos, elaboram modelos matemáticos capazes de identificar amostras desconhecidas. Os mais utilizados são Modelagem Independente por Analogia de Classes (*Software Independent Modelling Class Analogy* - SIMCA), Análise Discriminante Linear (*Linear Discriminant Analysis* - LDA) e Análise Discriminante por Mínimos Quadrados Parciais (*Partial Least Squares-Discriminant Analysis* - PLS-DA) (BRERETON, 2003; BRERETON; LLOYD, 2014; SUN, 2009).

2.5.2.1. Técnicas de reconhecimento de padrões não supervisionadas

As técnicas de reconhecimento de padrões não supervisionadas PCA e HCA prestam-se à análise exploratória dos dados por meio da visualização gráfica do conjunto de dados (BEEBE; PELL; SEASHOLTZ, 1998; CORREIA; FERREIRA, 2007). No entanto, essas duas técnicas possuem distintas metodologias. Em HCA, as amostras são agrupadas baseadas na similaridade entre elas. Essa similaridade é determinada através da medida das distâncias entre as amostras, que podem ser calculadas por diferentes métodos como distância de Mahalanobis ou Euclidiana. A partir dessas medidas, a matriz de dados é reduzida passo a passo a uma dimensão, pela reunião de pares

semelhantes, até que todos os pontos estejam agrupados em uma única classe. Desse modo, obtém-se um gráfico bidimensional de todo o conjunto de dados, chamado de dendrograma, em que o eixo y representa os grupos unidos por ordem decrescente de semelhança e as distâncias entre eles são apresentadas no eixo x. Por meio da análise dos dendrogramas formados utilizando distâncias entre os pontos como uma métrica de semelhança entre amostras, a formação de agrupamentos e padrões naturais é ressaltada, facilitando a interpretação e compreensão dos dados.

No caso da PCA, que é uma técnica multivariada amplamente utilizada para análise exploratória de dados, seu objetivo é representar a variação de conjuntos de dados em um espaço de dimensionalidade reduzida, permitindo a interpretação de conjuntos de dados amplos e complexos (BRO; SMILDE, 2014; PEREIRA *et al.*, 2014). Esse propósito é alcançado pela projeção das amostras em um novo sistema de eixos chamados de fatores ou componentes principais, que são formados a partir de combinações lineares das variáveis originais. As componentes principais são ortogonais entre si, contendo informações distintas e variância em ordem decrescente, ou seja, a primeira componente principal explica a maior variância dos dados, a segunda componente principal explica a segunda maior variância, e assim sucessivamente. Nesse novo sistema de eixos, as coordenadas das amostras relativas aos componentes principais são chamadas de escores e as contribuições dos eixos originais para cada componente principal são denominadas de pesos (cossenos dos ângulos entre os eixos das variáveis e os eixos das componentes principais).

Diferentemente da HCA, que apresenta toda a variação existente no conjunto de dados em seus dendrogramas, PCA apresenta apenas frações da variância total nos gráficos de escores (BEEBE; PELL; SEASHOLTZ, 1998). Em geral, essas frações são suficientes para a interpretação dos dados, desde que a determinação do número de componentes principais a ser utilizado (dimensionalidade inerente ou *rank*) seja adequada, de modo a incluir apenas informações relevantes e filtrar ruídos que podem ser introduzidos quando muitas componentes principais são utilizadas. Já os gráficos dos pesos são muito úteis para determinar que variáveis são mais importantes para a formação

dos agrupamentos, informação essa indisponível nos dendrogramas resultantes da HCA. Logo, o estudo conjunto dos gráficos de escores e pesos permite relacionar a identidade de uma amostra com suas características químicas, que é o objetivo principal dos estudos de reconhecimento de padrões.

De modo geral, HCA e PCA são técnicas eficientes para aumentar a compreensão de conjuntos de dados multivariados, pois a apresentação dos resultados experimentais na forma de gráficos facilita sua interpretação, visto que o ser humano possui poderoso sistema visual de reconhecimento de padrões. Ambas as técnicas permitem avaliar a presença ou ausência de agrupamentos naturais entre as amostras, sem utilizar conhecimento prévio de membros de classes. Além do mais, são úteis para identificar amostras anômalas (*outliers*), ou seja, amostras que apresentam comportamentos distintos das demais e que podem influenciar negativamente a interpretação dos dados (BEEBE; PELL; SEASHOLTZ, 1998).

2.5.2.2. Técnicas de reconhecimento de padrões supervisionadas

As técnicas de reconhecimento de padrões supervisionadas (técnicas de classificação) são utilizadas quando o objetivo é prever a que classes pertencem amostras de identidades/propriedades desconhecidas (BEEBE; PELL; SEASHOLTZ, 1998). Para que isso ocorra, são construídos modelos matemáticos que relacionem propriedades de interesse com medidas químicas suficientemente sensíveis para detectar diferenças entre as amostras. Uma vez estabelecido o modelo, tenta-se determinar a que classe pertence uma amostra desconhecida. Existe uma diversidade de técnicas que usam diferentes metodologias para alcançar esse objetivo e, escolher entre elas, fica a critério do analista. Nesse contexto, nas próximas seções serão abordadas as técnicas Análise Discriminante Linear (LDA) e Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA), bem como os parâmetros utilizados para avaliar a qualidade de um modelo de classificação.

2.5.2.2.1. Parâmetros de desempenho

Embora ainda não exista um consenso a respeito de protocolos de validação e da terminologia a ser utilizada em análises qualitativas multivariadas, alguns parâmetros de desempenho que avaliam a qualidade de um modelo de classificação e sua capacidade de classificar amostras futuras têm sido utilizados com certa frequência nos últimos anos (OLIVERI; DOWNEY, 2012; DRIVELOS *et al.*, 2014; SOUTO *et al.*, 2014; LÓPEZ; CALLAO; RUISÁNCHEZ, 2015; COSTA *et al.*, 2016; FERNANDES *et al.*, 2016; PIERINI *et al.*, 2016). Esses parâmetros são baseados em erros de classificação resultantes da aplicação dos modelos desenvolvidos. A maneira mais usual de apresentar esses erros é por meio de uma matriz de confusão (**Tabela 2.2**), que é uma matriz quadrada de dimensões $G \times G$, onde G é o número de classes, cada entrada n_{gk} representa o número de objetos pertencentes à classe g e atribuídas à classe k e os elementos n_{gg} da diagonal representam os objetos corretamente classificados.

Tabela 2.2. Representação geral de uma matriz de confusão.

		CLASSE PREVISTA			
		CLASSE 1	CLASSE 2	CLASSE ...	CLASSE G
CLASSE VERDADEIRA	CLASSE 1	n_{11}	n_{12}	...	n_{1G}
	CLASSE 2	n_{21}	n_{22}	...	n_{2G}
	CLASSE
	CLASSE G	n_{G1}	n_{G2}	...	n_{GG}

A partir da matriz de confusão pode-se calcular alguns parâmetros de desempenho utilizados para avaliar a qualidade do modelo de classificação (OLIVERI; DOWNEY, 2012; LÓPEZ; CALLAO; RUISÁNCHEZ, 2015). Entre eles, os principais são a exatidão, a sensibilidade que mede a habilidade do modelo

para reconhecer amostras pertencentes às suas classes e a especificidade que mede a habilidade do modelo para rejeitar amostras de outras classes. Esses parâmetros são calculados a partir das **Equações (2.1), (2.2) e (2.3)**:

$$\text{Sensibilidade} = \frac{VP}{VP+FN} \quad (2.1)$$

$$\text{Especificidade} = \frac{VN}{VN+FP} \quad (2.2)$$

$$\text{Exatidão} = \frac{VP+VN}{VP+VN+FP+FN} \quad (2.3)$$

onde VP (verdadeiros positivos) é o número de amostras pertencentes a uma classe e corretamente classificadas nela; VN (verdadeiros negativos) é o número de amostras que não pertencem à classe e não foram classificadas nela; FP (falsos positivos) é o número de amostras que não pertencem à classe e foram incorretamente classificadas nela e FN (falsos negativos) é o número de amostras que pertencem à classe e foram incorretamente classificadas como não pertencentes à ela.

Os parâmetros de desempenho são utilizados tanto na etapa de ajuste do modelo, onde um conjunto de treinamento é utilizado, quanto na etapa de previsão de amostras completamente independentes, ou seja, que não foram utilizadas no desenvolvimento do modelo. Dessa forma, além de avaliar a confiabilidade do modelo, esses parâmetros podem ser utilizados para orientar a escolha de configurações que forneçam sua melhor capacidade preditiva, tais como o melhor pré-processamento, a escolha de variáveis mais informativas, etc.

2.5.2.2.2. Análise discriminante por mínimos quadrados parciais (PLS-DA)

PLS-DA é uma técnica de reconhecimento de padrão supervisionada baseada em uma regressão PLS, cujo objetivo é encontrar a relação entre a resposta instrumental (matriz \mathbf{X}) e propriedades de interesse (matriz \mathbf{Y}).

Diferentemente do algoritmo PLS clássico utilizado em calibração multivariada, onde a matriz \mathbf{Y} é composta por variáveis contínuas, como por exemplo concentrações, em PLS-DA ela é composta por variáveis de categoria que representam classes de amostras. Desse modo, cada coluna da matriz \mathbf{Y} representa uma classe e cada amostra assume o valor 1 ou 0, dependendo se a amostra pertence ou não à classe representada por aquela coluna (BRERETON; LLOYD, 2014). Assim, após construção do modelo com as variáveis latentes (combinações lineares das variáveis originais) que proporcionam melhor descrição dos dados e, portanto, melhor discriminação das classes, PLS-DA retorna valores estimados de \mathbf{y} para cada amostra. Visto que os valores estimados não são exatamente 0 e 1, um *threshold* deve ser escolhido para cada classe de maneira que se possa determinar se uma amostra pertence ou não àquela classe.

Uma maneira de se determinar o *threshold* para cada classe é por meio de curvas ROC (do inglês, *Receiver Operating Characteristics*), calculadas para cada classe separadamente e construídas a partir dos valores de sensibilidade e especificidade obtidos variando-se o *threshold*, conforme **Figura 2.3** (BRERETON, 2009; SUN, 2009; BALLABIO; CONSONNI, 2013). O ponto dessa curva onde se obtém máximas sensibilidade e especificidade, ou seja, o ponto que se localizar no canto superior esquerdo, deve ser escolhido como o *threshold* mais adequado para atribuir uma amostra a determinada classe ou não: se o valor de \mathbf{y} encontrado for superior ao *threshold*, a amostra pertence àquela classe, caso contrário, não pertence. Enquanto uma classificação perfeita produz um ponto da curva no canto superior esquerdo, uma classificação randômica fornece pontos ao longo de uma linha diagonal que vai do canto inferior esquerdo ao canto superior direito, indicando que o método não possui poder discriminante.

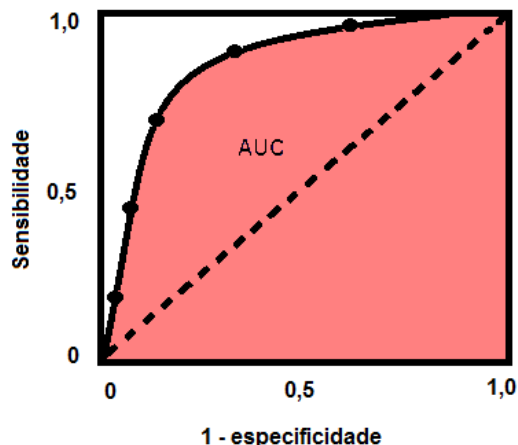


Figura 2.3. Exemplo de curva ROC. Pontos na curva correspondem a valores de sensibilidade e especificidade obtidas variando-se o *threshold*. A linha pontilhada indica classificação randômica.

Além dos parâmetros de desempenho já abordados (seção 2.4.2.2.1), a qualidade de um modelo também pode ser avaliada pela AUC (*Area under the Curve*), que é simplesmente a área abaixo da curva ROC (**Figura 2.3**) e possui valores entre 0 e 1: quanto maior o valor de AUC, melhor o modelo. Valores em torno de 0,5 indicam classificação randômica (BRERETON, 2009).

A técnica PLS-DA é amplamente utilizada para classificação de diferentes amostras analisadas por diferentes técnicas analíticas, conforme relatado em inúmeros artigos na atualidade (BOTELHO *et al.*, 2015; LENHARDT *et al.*, 2015; NKANSAH *et al.*, 2015; BORRÀS *et al.*, 2016; PEREIRA *et al.*, 2016; VALDERRAMA; VALDERRAMA, 2016; SANTOS *et al.*, 2017).

2.5.2.2.3. Análise discriminante linear (LDA)

A técnica LDA, proposta originalmente por Fisher (FISHER, 1936), é um dos métodos de reconhecimento de padrões supervisionados mais utilizados. Fundamenta-se nos cálculos de funções discriminantes lineares, com o objetivo de alcançar máxima variância entre as classes e mínima variância entre os elementos da mesma classe, através das combinações lineares das variáveis

originais. Determinadas as funções discriminantes, as fronteiras lineares de separação entre as classes são calculadas de modo que se situem igualmente distantes dos centroides de pares de classes. Assim, uma amostra desconhecida é classificada como pertencente a uma determinada classe se o quadrado da distância de Mahalanobis entre ela e o centroide daquela classe for menor do que a distância ao centroide das demais classes (MASSART *et al.*, 1997).

Essa técnica pode ser considerada como um método de redução de dimensionalidade, visto que seu objetivo é encontrar uma direção no espaço multivariado que alcance a separação máxima entre as classes avaliadas (MASSART *et al.*, 1997; NAES *et al.*, 2002). No entanto, este método é comprometido por problemas de colinearidade e pode ser utilizado apenas em dados de baixa dimensão (BRERETON, 2003). Desse modo, a utilização desse método em dados provenientes da espectroscopia no infravermelho, por exemplo, que em geral possuem um número de variáveis muito superior ao número de amostras, só é possível através de sua combinação com técnicas de seleção de variáveis e/ou redução de dimensionalidade.

Nesse contexto, diferentes algoritmos para seleção de variáveis informativas vêm sendo desenvolvidos ao longo dos anos, entre eles, o algoritmo da projeções sucessivas (SPA). Esse algoritmo foi inicialmente desenvolvido para selecionar variáveis informativas para calibração multivariada, quando utilizada regressão linear múltipla (MLR) (ARAÚJO *et al.*, 2001; GALVÃO *et al.*, 2001). Sua filosofia consiste em buscar um pequeno subconjunto representativo de variáveis espectrais com ênfase na minimização da colinearidade, de modo a se obter o menor RMSEV (raiz quadrada do erro médio quadrático para um conjunto de validação) a partir do modelo MLR construído com as variáveis selecionadas.

Posteriormente, o algoritmo foi modificado para seleção de variáveis com fins de classificação, de modo a minimizar o risco médio de uma classificação incorreta por LDA, método onde a colinearidade entre as variáveis representa um grave problema (PONTES *et al.*, 2005). Assim como no SPA-MLR, o SPA-LDA baseia-se em uma sequência de operações de projeções de vetores aplicadas às colunas de uma matriz de treinamento, gerando cadeias de

variáveis minimamente colineares. Uma vez construídas, o SPA escolherá a cadeia de variáveis que apresentar o menor risco médio de classificação incorreta de um conjunto de validação pela LDA, calculado pela função de custo:

$$G = \frac{1}{N_{\text{val}}} \sum_{n=1}^{N_{\text{val}}} g_n \quad (2.3)$$

onde N_{val} é o número de misturas de validação e g_n (risco de uma classificação incorreta do objeto \mathbf{x}_{val} da n -ésima amostra de validação) é definido como:

$$g_n = \frac{r^2(\mathbf{x}_{\text{val},n}, \boldsymbol{\mu}_{I_n})}{\min_{I_j \neq I_n} r^2(\mathbf{x}_{\text{val},n}, \boldsymbol{\mu}_{I_j})} \quad (2.4)$$

Na **Equação 2.4**, o numerador $r^2(\mathbf{x}_{\text{val}}, \boldsymbol{\mu}_{I_n})$ é o quadrado da distância de Mahalanobis entre o objeto \mathbf{x}_{val} (com índice de classe I_n) e a média de sua classe ($\boldsymbol{\mu}_{I_n}$). O denominador da **Equação 2.4** corresponde ao quadrado da distância de Mahalanobis entre o objeto \mathbf{x}_{val} e o centro da classe errada mais próxima. Logo, quanto menor o valor de g_n mais perto o objeto \mathbf{x}_{val} deverá estar do centro da sua verdadeira classe e distante dos centros das demais classes, ou seja, minimizando G , uma melhor separação das amostras de acordo com suas verdadeiras classes será alcançada (SOARES *et al.*, 2013).

Para evitar a necessidade de utilizar um conjunto de teste separado para a validação, Soares *et al.* (2014) modificaram o cálculo do risco médio de uma classificação incorreta, que passou a ser calculado utilizando as próprias misturas de treinamento, conforme a equação:

$$g_n = \frac{r^2(\mathbf{x}_{\text{trein},n}, \boldsymbol{\mu}_{I_n})}{\min_{I_j \neq I_n} r^2(\mathbf{x}_{\text{trein},n}, \boldsymbol{\mu}_{I_j})} \quad (2.5)$$

Desse modo, para evitar sobreajuste do modelo, visto que é construído e validado utilizando as mesmas amostras, o cálculo da função de custo foi alterado para:

$$G = \frac{1}{N_{\text{trein}} - L - C} \sum_{n=1}^{N_{\text{trein}}} g_n \quad (2.6)$$

Onde N_{trein} é o número de amostras de treinamento, C é o número de classes envolvidas no problema e L é o número de variáveis da cadeia que está sob avaliação (SOARES *et al.*, 2014).

O algoritmo SPA-LDA tem sido aplicado na classificação de diversos tipos de amostras, tais como óleos vegetais e diesel (PONTES *et al.*, 2005), solos (PONTES *et al.*, 2009), óleos vegetais comestíveis (GAMBARRA-NETO *et al.*, 2009; SOARES *et al.*, 2010), cigarros (MOREIRA *et al.*, 2009), cafés (SOUTO *et al.*, 2014; SOUTO *et al.*, 2015), misturas de biodiesel/diesel (COSTA *et al.*, 2016; PONTES *et al.*, 2011), cervejas (GHASEMI-VARNAMKHAZI *et al.*, 2012), vinhos brancos (AZCARATE *et al.*, 2015), tintas de canetas azuis (SILVA *et al.*, 2013), etanol combustível (SILVA *et al.*, 2012), chás (DINIZ *et al.*, 2014; DINIZ *et al.*, 2015; DINIZ *et al.*, 2016), própolis (PIERINI *et al.*, 2016), meles (DOMÍNGUEZ *et al.*, 2014), sucos cítricos (FREITAS *et al.*, 2013) etc. O êxito alcançado na classificação dessa variedade de amostras, analisadas por diferentes técnicas analíticas, demonstram a potencialidade dessa ferramenta.

2.5.3. Calibração multivariada

A calibração é o processo no qual é estabelecida uma relação matemática entre medidas instrumentais de determinadas amostras e parâmetros de interesse, tais como concentração, densidade, viscosidade, etc. Desse modo, pode-se prever os valores desses parâmetros em amostras futuras, utilizando-se de suas medidas instrumentais e da relação matemática (ou modelo) estabelecida *a priori*. No caso de dados espectrais, geralmente são construídos

modelos de calibração lineares que relacionam os valores de absorvância de uma série de padrões com suas concentrações, permitindo, dessa forma, que concentrações de amostras desconhecidas sejam previstas.

De modo geral, a calibração pode ser univariada ou multivariada. Em termos espectroscópicos, uma calibração é dita univariada quando relaciona a absorvância medida em um único número de onda com as concentrações dos padrões. Entretanto, a calibração univariada só é útil quando apenas uma espécie, o analito, contribui para o sinal analítico. Do contrário, a separação dos interferentes deverá ser realizada, levando a procedimentos geralmente complicados e caros. Já na calibração multivariada, múltiplos comprimentos de onda são utilizados na construção de um modelo que relaciona a absorvância com a concentração. Essa metodologia possibilita a determinação simultânea de mais de uma espécie química, reduzindo o tempo de análise, proporciona melhora da precisão nas predições e permite a detecção de interferentes. Outra grande vantagem, denominada de vantagem de 1ª ordem, é que, quando utilizados os métodos inversos de calibração multivariada, ou seja, métodos em que a variável dependente é a concentração (ao contrário da lei de Lambert Beer em que a absorvância é uma função da concentração), ela pode ser realizada na presença de interferentes, mesmo que estes não sejam incluídos como variáveis adicionais no modelo. Para isso, basta que também estejam presentes nos padrões utilizados na construção do modelo, dessa forma, os interferentes serão modelados implicitamente (BEEBE; PELL; SEASHOLTZ, 1998; SAURINA; TAULER, 2000; SUN, 2009).

Quando se deseja realizar calibração multivariada em dados espectroscópicos, os dados são organizados conforme **Figura 2.4**, onde \mathbf{X} ($i \times j$) corresponde à matriz formada pelos espectros de “i” amostras e “j” números de onda e \mathbf{Y} ($i \times k$) é a matriz composta pelas concentrações (ou outros parâmetros de interesse) de “k” componentes, presentes nas “i” amostras.

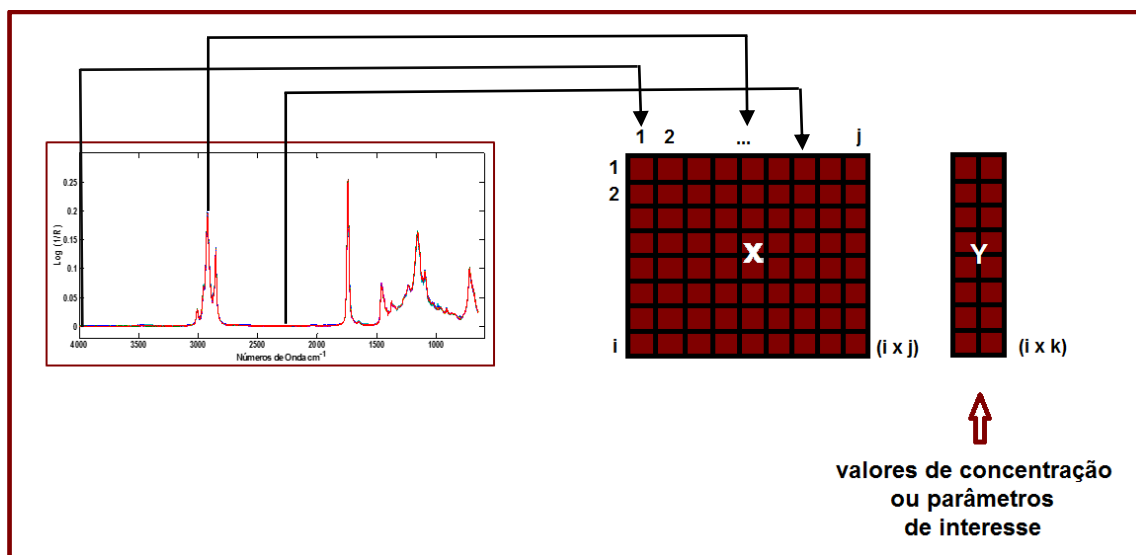


Figura 2.4. Organização dos dados em calibração multivariada.

Três etapas compõem o processo de calibração multivariada, são elas: a calibração, validação e previsão. Na etapa de calibração é construído um modelo matemático que relaciona o sinal analítico a propriedade(s) de interesse. Em seguida, esse modelo é avaliado na etapa de validação usando algumas ferramentas estatísticas de diagnóstico (tratados na seção 2.4.3.4) e, por fim, na etapa de previsão, esse modelo validado é aplicado a amostras desconhecidas para prever suas propriedades (por exemplo, concentração).

Na atualidade, existem diversas técnicas de calibração multivariada. Entre elas, as mais utilizadas no âmbito da química analítica são a Regressão linear múltipla (MLR) e as técnicas baseadas em redução de variáveis Regressão por componentes principais (PCR) e Regressão por mínimos quadrados parciais (PLS). As três técnicas pressupõem uma relação linear entre o sinal analítico e as propriedades de interesse e seus fundamentos serão discutidos a seguir.

2.5.3.1. Regressão linear múltipla (MLR)

A regressão linear múltipla faz parte dos métodos de calibração multivariada classificados como métodos inversos (BEEBE; PELL; SEASHOLTZ,

1998). Logo, em termos espectroscópicos, a construção do modelo utiliza o inverso da lei de Lambert - Beer, onde a concentração é expressa como uma função da absorvância e é escrita conforme a **Equação 2.5**:

$$\mathbf{Y} = \mathbf{X}\mathbf{B}_{\text{MLR}} + \mathbf{E} \quad (2.5)$$

em que as matrizes \mathbf{Y} ($i \times k$) e \mathbf{X} ($i \times j$) contém, respectivamente, as concentrações e os espectros das amostras selecionadas para a etapa de calibração, a matriz \mathbf{E} ($i \times k$) representa o resíduo não modelado em \mathbf{Y} , e \mathbf{B}_{MLR} é a matriz desconhecida dos coeficientes de regressão. Esses, são estimados na etapa de calibração com base no método dos mínimos quadrados, a partir da **Equação 2.6**:

$$\mathbf{B}_{\text{RLM}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \quad (2.6)$$

onde os índices sobrescritos -1 e T representam, respectivamente, a inversão e transposição de uma matriz ou vetor. Uma vez que a matriz \mathbf{B}_{MLR} foi obtida, é utilizada, na etapa de previsão, para estimar as concentrações de amostras desconhecidas, a partir de seus espectros, conforme **Equação 2.7**:

$$\hat{\mathbf{Y}}_{\text{pred}} = \mathbf{X}^T_{\text{pred}} \hat{\mathbf{B}} \quad (2.7)$$

Entretanto, a MLR apresenta algumas limitações. A obtenção dos coeficientes de regressão (\mathbf{B}_{MLR}) na **Equação 2.6** requer a inversão da matriz ($\mathbf{X}^T\mathbf{X}$) e essa operação algébrica só é possível, quando os dados obedecem a algumas condições. A primeira condição especifica que o número de amostras de calibração deve ser maior ou igual ao número de variáveis ($i > j$), caso contrário, o sistema torna-se indeterminado visto que o modelo consiste na resolução de um sistema de equações lineares simultâneas. A outra condição é que as variáveis (colunas de \mathbf{X}) devem ser idealmente vetores linearmente independentes, ou seja, não devem apresentar correlação entre si. Se esta condição não for obedecida, a matriz ($\mathbf{X}^T\mathbf{X}$) pode não apresentar inversa. Essas

condições impossibilitam a aplicação da MLR a dados com alta dimensionalidade como os dados espectroscópicos, sem que uma seleção prévia de variáveis seja realizada (PASQUINI, 2003; GOMES, 2012).

Vários algoritmos foram desenvolvidos com a finalidade de selecionar variáveis mais informativas e, desse modo, contornar as restrições da MLR, tais como Algoritmo Genético (GA), *Stepwise* (STW), Algoritmo das projeções sucessivas (SPA), entre outros (XIAOBO *et al.*, 2010). O SPA foi particularmente desenvolvido para seleção de variáveis espectrais para determinação de concentrações por MLR (ARAÚJO *et al.*, 2001) e, ao contrário do GA que possui natureza estocástica, seus resultados são sempre reproduzíveis. Neste trabalho foi utilizado o algoritmo SPA-MLR, que será descrito na próxima seção.

2.5.3.1.1. SPA-MLR

O SPA é um método de seleção de variáveis por passos, desenvolvido para selecionar variáveis espectrais cujo conteúdo de informação seja minimamente redundante. Foi primeiramente aplicado a dados espectrofotométricos no UV-VIS para determinação simultânea de complexos de Co^{2+} , Cu^{2+} , Mn^{2+} , Ni^{2+} e Zn^{2+} com 4-(2-piridilazo) resorcinol em amostras contendo os analitos na faixa de 0,02 a 0,5 mg l^{-1} , cujos espectros apresentavam alta colinearidade (ARAÚJO *et al.*, 2001). O SPA-MLR apresentou valores de RMSEP comparáveis aos alcançados quando utilizado o GA, com a vantagem de ser realizado em menos tempo (aproximadamente 200 vezes inferior) e de ser um método determinístico. As variáveis selecionadas pelo SPA também proporcionaram melhora na capacidade preditiva de modelos PCR e PLS.

O algoritmo SPA opera na matriz de respostas instrumentais \mathbf{X}_{cal} ($M_{\text{cal}} \times J$), cujas linhas e colunas correspondem a M_{cal} amostras de calibração e J variáveis espectrais, respectivamente. Partindo de uma dada variável inicial $k(0)$, o SPA determina qual das variáveis remanescentes apresenta a maior projeção em um subespaço ortogonal a $k(0)$. Essa variável, denominada de x_1 contém, então, a maior quantidade de informação a respeito do sistema em análise, não

incluída em \mathbf{x}_0 , e será usada na próxima iteração como a nova variável de referência. Logo, a cada iteração o SPA incorpora uma nova variável que possua mínima colinearidade com a anterior, até que um número específico N de variáveis seja alcançado, gerando $J \times [(M_{cal} - A) + 1]$ pares de conjuntos (k(0), N), onde:

- J = número total de variáveis espectrais;
- A = número de analitos;
- M_{cal} = número de misturas de calibração;
- k(0) = o vetor inicial de partida;
- N = número de variáveis espectrais a serem selecionadas;
- Possibilidades para k(0): 1, 2, ..., J;
- Possibilidades para N: A, A+1, ..., M_{cal}

Para escolher o melhor par (k(0), N), modelos MLR são construídos com cada um deles e comparados em termos da raiz quadrada do erro médio quadrático de previsão para o conjunto de validação (RMSEV), o qual é calculado por:

$$RMSECV = \sqrt{\frac{1}{M_{cal}} \sum_{i=1}^{M_{cal}} [\hat{Y}_{cal}(i) - Y_{cal}(i)]^2} \quad (2.8)$$

se for utilizado como método de validação interna a validação cruzada, onde \mathbf{Y}_{cal} e $\hat{\mathbf{Y}}_{cal}$ são, respectivamente, as concentrações verdadeiras e estimadas dos analitos na i-ésima amostra, ou

$$RMSEP = \sqrt{\frac{1}{M_{teste}} \sum_{i=1}^{M_{teste}} [\hat{Y}_{teste}(i) - Y_{teste}(i)]^2} \quad (2.9)$$

se um conjunto de teste não utilizado na etapa de calibração for empregado na validação (validação externa), onde $\mathbf{Y}_{\text{teste}}$ e $\hat{\mathbf{Y}}_{\text{teste}}$ são, respectivamente, as concentrações verdadeiras e estimadas dos analitos na i -ésima amostra. Por fim, o algoritmo seleciona o conjunto de variáveis cujo modelo MLR levou ao menor RMSEV.

Após sua apresentação, diversos artigos sobre modificações e novas aplicações do SPA foram publicados, demonstrando sua robustez e versatilidade (GALVÃO *et al.*, 2001, 2007, 2008; BREITKREITZ *et al.*, 2003; FILHO *et al.*, 2005; HONORATO *et al.*, 2005; PISTONESI *et al.*, 2010; INÁCIO; MOURA; LIMA, 2011; VASCONCELOS *et al.*, 2012; XU *et al.*, 2012; GOMES *et al.*, 2013; SOARES *et al.*, 2013; INSAUSTI *et al.*, 2013; WANG; XIE, 2014; LIU *et al.*, 2015; LI *et al.*, 2016; GONÇALVES *et al.*, 2016). Em 2013, diante do interesse da comunidade analítica, um artigo de revisão foi publicado apresentando as características do SPA tanto no âmbito da calibração multivariada como da classificação, com exemplos demonstrativos para ambos os casos. Algumas modificações do algoritmo também foram abordadas na ocasião (SOARES *et al.*, 2013).

2.5.3.2. Regressão por componentes principais (PCR)

A Regressão em Componentes Principais (*Principal Component Regression – PCR*) é um método de calibração multivariada que realiza uma regressão múltipla inversa relacionando os escores de uma matriz de dados \mathbf{X} com a propriedade de interesse, \mathbf{y} (BEEBE; PELL; SEASHOLTZ, 1998). Para que isso ocorra, em primeiro lugar, a matriz \mathbf{X} deve ser decomposta em componentes principais conforme **Equação 2.10**:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} = \sum_{a=1}^A t_a p_a^T + \mathbf{E} \quad (2.10)$$

Uma vez escolhido o número A de componentes principais ideal para descrever a matriz \mathbf{X} , esta pode ser representada por sua matriz de escores \mathbf{T} :

$$\mathbf{T} = \mathbf{X}\mathbf{P} \quad (2.11)$$

onde \mathbf{T} e \mathbf{P} são as matrizes de escores e pesos, respectivamente, obtidas a partir da matriz \mathbf{X} . A matriz \mathbf{Y} pode, então, ser relacionada à matriz de escores \mathbf{T} :

$$\mathbf{Y} = \mathbf{T}\mathbf{B} + \mathbf{E} \quad (2.12)$$

onde \mathbf{B} é a matriz dos coeficientes de regressão e é calculada pelo método dos mínimos quadrados, conforme **Equação 2.13**, desde que os valores das concentrações do conjunto de calibração sejam conhecidos:

$$\hat{\mathbf{B}} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{Y} \quad (2.13)$$

Visto que os escores são ortogonais, a PCR não sofre com problemas de colinearidade como MLR e, portanto, o termo $\mathbf{T}^T\mathbf{T}$ pode ser facilmente invertido e a **Equação 2.13** resolvida.

Uma vez estabelecido o modelo da calibração, este é utilizado para prever as concentrações de novas amostras a partir dos espectros registrados. Esse procedimento é realizado calculando-se os escores \mathbf{T}^* das amostras de previsão \mathbf{X}^* (cujas concentrações se quer prever) a partir da matriz de pesos \mathbf{P} calculada na etapa de calibração, para A componentes ótimos:

$$\mathbf{T}^* = \mathbf{X}^*\mathbf{P} \quad (2.14)$$

Logo, combinando-se a matriz dos coeficientes de regressão \mathbf{B} , também calculada na etapa de calibração, com a matriz de escores \mathbf{T}^* das amostras de previsão, determina-se a propriedade $\hat{\mathbf{Y}}$ das novas amostras:

$$\hat{\mathbf{Y}} = \mathbf{T}^* \mathbf{B} \quad (2.15)$$

As etapas envolvidas nesse processo estão representadas na **Figura 2.5**.

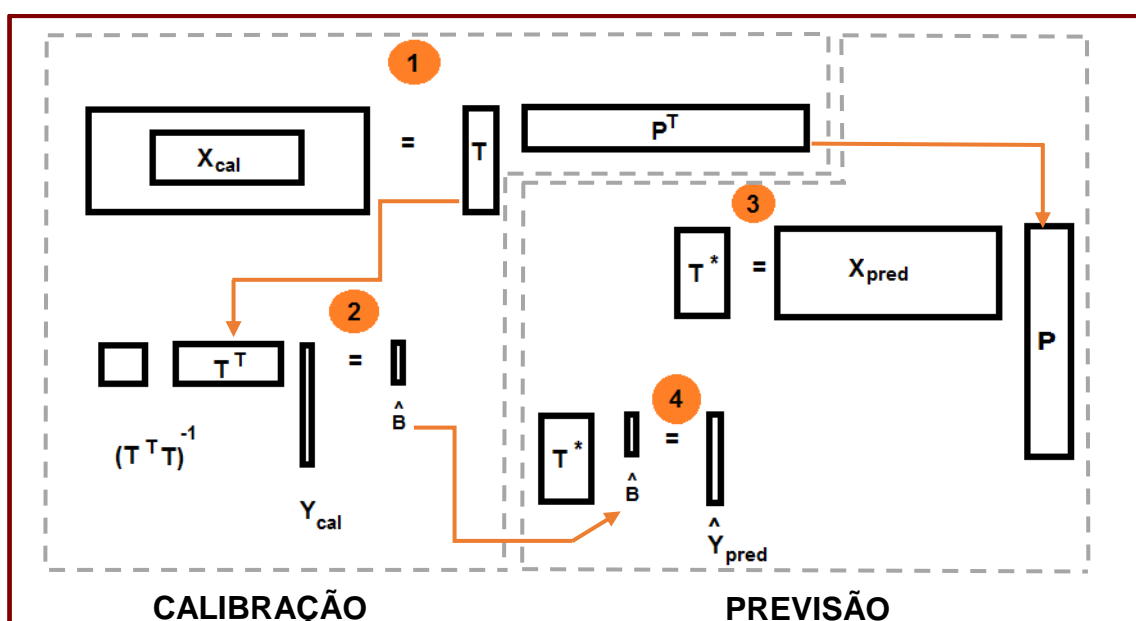


Figura 2.5. Ilustração de calibração PCR e previsão de novas amostras. (1) Decomposição de matriz de dados X em matrizes de escores e pesos. (2) Cálculo do vetor de regressão. (3) Projeção dos espectros de predição no espaço da calibração. (4) Previsão de propriedade de interesse das novas amostras. *Fonte:* adaptado de SUN, 2009.

Geralmente, os componentes principais selecionados para descrever \mathbf{X} são aqueles que apresentam maior variância, pois o subespaço definido por eles captura a maior parte da variação nos dados. Entretanto, como PCR utiliza apenas a matriz \mathbf{X} na determinação das combinações lineares das variáveis originais, às vezes os componentes de maior variância não são importantes para prever \mathbf{Y} , estando as informações necessárias para isso localizadas nas componentes de variância mais baixa.

2.5.3.3. Regressão por mínimos quadrados parciais (PLS)

A técnica de regressão por mínimos quadrados parciais (PLS) foi inicialmente desenvolvida por Herman Wold em 1975. No início da década de 80, Svante Wold e Harald Martens fizeram algumas modificações para melhor adequá-la a dados complexos de ciências e tecnologia, surgindo como uma poderosa ferramenta em diversas áreas, em particular, a química (WOLD, 2001; WOLD; SJÖSTRÖM; ERIKSSON, 2001).

Assim como em PCR, PLS utiliza a técnica de análise de componentes principais para a redução da dimensionalidade do conjunto de dados para posterior correlação entre os espectros (matriz \mathbf{X}) e as propriedades de interesse (matriz \mathbf{Y}). Entretanto, diferente de PCR, PLS utiliza as informações contidas nas matrizes \mathbf{X} e \mathbf{Y} para o cálculo dos escores e pesos, de maneira que a maior parte da variância de ambas as matrizes consegue ser explicada. Cada componente, denominado em PLS de variáveis latentes ou fatores, é obtido de forma a maximizar a covariância entre \mathbf{X} e \mathbf{Y} . Para que isso ocorra, as matrizes \mathbf{X} e \mathbf{Y} são decompostas por PCA, conforme as **Equações 2.16 e 2.17**:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (2.16)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \quad (2.17)$$

onde \mathbf{T} , \mathbf{P} e \mathbf{E} são as matrizes de escores, pesos e resíduos, respectivamente, para \mathbf{X} , e \mathbf{U} , \mathbf{Q} e \mathbf{F} são as matrizes de escores, pesos e resíduos, respectivamente, para \mathbf{Y} . Os valores dos escores \mathbf{T} e \mathbf{U} nas matrizes são alterados até que o melhor modelo linear seja estabelecido entre eles e assim, a maior covariância entre \mathbf{X} e \mathbf{Y} seja alcançada.

O modelo resultante relaciona linearmente os escores da matriz \mathbf{X} (\mathbf{T}) com os escores da matriz \mathbf{Y} (\mathbf{U}), conforme **Equações 2.18 e 2.19**:

$$\mathbf{U} = \mathbf{BT} + \mathbf{G} \quad (2.18)$$

$$\mathbf{Y} = \mathbf{BTQ}^T + \mathbf{H} \quad (2.19)$$

em que **B**, **G** e **H** são as matrizes dos coeficientes de regressão, resíduos dos escores e resíduos de concentração, respectivamente.

O número ótimo de variáveis latentes utilizado no modelo PLS é geralmente determinado por validação cruzada (BEEBE; PELL; SEASHOLTZ, 1998). Neste procedimento, o conjunto de validação é dividido em subgrupos, onde um deles é deixado de fora da construção do modelo de calibração e seu parâmetro de interesse é previsto a partir do modelo construído com os demais subgrupos, empregando um número crescente de variáveis latentes. Este procedimento é repetido até que todos os subgrupos tenham sido testados e, o número de variáveis latentes ideal para construção do modelo será aquele a partir do qual não exista variação significativa no valor de RMSECV (raiz quadrada do erro médio quadrático de validação cruzada). Quando uma amostra por vez é retirada do conjunto de calibração, o método é denominado de *leave-one-out* (SOUZA *et al.*, 2013).

Quando o objetivo for a determinação de mais de um analito, duas possibilidades de PLS podem ser utilizadas: PLS1 em que, para cada analito, um modelo diferente é construído; ou PLS2 para determinação simultânea de todos os analitos, a partir da construção de um modelo único empregando todas as colunas de **Y**. A aplicação do método PLS2 deve ser realizada em situações nas quais as propriedades de interesse estão correlacionadas. Quando não, deve-se dar preferência à regressão empregando PLS1 para cada coluna da matriz **Y**, gerando um modelo para a previsão de cada analito.

Embora teoricamente PLS deva gerar resultados melhores que PCR devido ao seu compromisso entre a explicação da variância em **X** e a obtenção da maior correlação com **Y**, não se pode afirmar que um método seja superior ao outro, já que ambos, em geral, apresentam resultados muito semelhantes. Entretanto, em PLS a maior parte da informação útil para prever os parâmetros de interesse está contida nas primeiras variáveis latentes, o que nem sempre ocorre em PCR, como mencionado na seção **2.5.3.2**.

2.5.3.4. Validação de modelos de calibração multivariada

O objetivo de uma calibração é calcular os parâmetros de uma equação, permitindo que propriedades de amostras desconhecidas sejam previstas no futuro. Uma vez que um modelo de calibração foi construído, a qualidade deste pode ser avaliada utilizando determinados parâmetros estatísticos como a exatidão, ajuste, erros sistemáticos (*bias*), limite de detecção e quantificação, etc, com o intuito de averiguar se apresenta uma performance adequada para as condições nas quais será aplicado.

Para avaliar a exatidão de um modelo, em geral, utiliza-se o RMSEV que pode ser calculado pela **Equação 2.8** (RMSECV) ou **Equação 2.9** (RMSEP), dependendo se a validação do modelo foi realizada por validação cruzada ou utilizando um conjunto de validação externo, que não fez parte da construção do modelo. Ambas as equações medem o desvio entre os valores reais do parâmetros de interesse e os valores estimados pelo modelo e incorporam erros aleatórios e sistemáticos.

Outra parâmetro utilizado para avaliar a exatidão de um método é o RPD (do inglês, *Residual Prediction Deviation* ou *Relative Predictive Determinant*), que é definido como o quociente entre o desvio padrão dos valores das amostras e o RMSECV ou RMSEP, dependendo do procedimento de validação. Segundo William e Sobering (1993), um valor de RPD maior que 2,5 é aceitável, enquanto um valor maior que 5,0 reflete exatidão satisfatória para controle de qualidade e valores maiores que 10,0 são um indicativo de modelos de calibração de excelente qualidade.

A existência de erros sistemáticos nos modelos afeta a precisão e a exatidão das medidas e pode ser avaliada pela **Equação 2.20**:

$$bias = \frac{\sum(y_i - \hat{y}_i)}{m} \quad (2.20)$$

onde y_i é o valor de referência, \hat{y}_i é o valor predito pelo modelo e m é a quantidade de objetos de calibração. Devido à existência de erros aleatórios, o somatório no numerador da **Equação 2.20** que, idealmente deveria ser zero, deve estar próximo a zero para que a existência de erros sistemáticos seja excluída. Conforme a norma E1655-05 da ASTM (do inglês, *American Society for Testing and Materials*), para avaliar se o *bias* incluso no modelo é significativo, deve-se realizar um teste t, por meio da fórmula:

$$t = \frac{|bias|\sqrt{n}}{SDV} \quad (2.21)$$

onde n é a quantidade de objetos da validação e SDV é o desvio padrão dos erros de validação (do inglês, *Standard Deviation of Validation*), e é estimado por:

$$SDV = \sqrt{\frac{\sum[(y - \hat{y}) - bias]^2}{n-1}} \quad (2.22)$$

Caso o valor de t calculado seja menor que o seu valor crítico para $n-1$ graus de liberdade, com 95% de confiança, o *bias* incluído no modelo pode ser considerado insignificante e desprezado.

Outros parâmetros muito utilizados para avaliar a qualidade de um modelo são o coeficiente de correlação (r), coeficiente de determinação (R^2), coeficiente angular (*slope*) e o intercepto ou coeficiente linear (*offset*). Esses parâmetros são obtidos ajustando-se uma reta, por meio de regressão por mínimos quadrados, entre o valor previsto e o valor de referência. Quando há concordância entre eles, os valores das amostras devem estar próximos a uma reta com inclinação de 45° , intercepto próximo a zero (*bias*), e coeficiente de determinação próximo à unidade, indicando um bom ajuste do modelo.

O limite de detecção e o limite de quantificação de um procedimento analítico, expressam as menores quantidades da espécie de interesse que podem ser detectadas e determinadas quantitativamente, respectivamente. Uma maneira de determiná-los consiste em prever, utilizando o modelo de calibração desenvolvido, o parâmetro de interesse das amostras que não contêm a espécie

em investigação e das que contêm a maior proporção calibrada desta espécie. A média do desvio padrão das previsões é então multiplicado por 10/3 ou 10, para determinação do limite de detecção e de quantificação, respectivamente (LUTZ *et al.*, 2014).

2.5.3.4.1. Ensaio de Recuperação

A determinação da exatidão de um método, definida como a concordância entre o valor real de um analito na amostra e o previsto por um método analítico, pode ser realizada por 3 diferentes métodos: baseada na análise de material de referência certificado; na comparação dos resultados obtidos com o método proposto com os resultados obtidos por um método de referência; e no uso de ensaios de recuperação (EURACHEM GUIDE, 2014). Entretanto, a abrangência limitada de matrizes e analitos, bem como o alto custo dos materiais de referência certificados, restringem seu uso e nem sempre existem métodos de referência validados para a determinação que se almeja. Logo, os ensaios de recuperação constituem um método muito utilizado e eficaz para a determinação da exatidão e validação de métodos analíticos.

O estudo da recuperação consiste na fortificação da amostra sob investigação com o analito de interesse, seguida pela determinação da concentração do analito adicionado, pelo método proposto. Calcula-se, então, a quantidade percentual recuperada, usando a **Equação 2.23**:

$$REC\% = \left(\frac{Y_{adic_{prev}}}{Y_{adic_{ref}}} \right) \times 100 \quad (2.23)$$

onde $Y_{adic_{prev}}$ é a concentração do analito prevista pelo método, e $Y_{adic_{ref}}$ é a concentração do analito adicionada à amostra. Testes de significância podem ser realizados, utilizando o teste t de *Student*, de acordo com a **Equação 2.24**:

$$t = \frac{|(100-REC)|\sqrt{n}}{S_{REC}} \quad (2.24)$$

onde REC é a média das recuperações obtidas para n repetições; 100 é a recuperação percentual desejada; S_{REC} é o desvio padrão das recuperações. Assim, se o valor de t obtido estiver dentro do intervalo estabelecido pelo valor tabelado para (n-1) graus de liberdade, em dado nível de significância, então o método será considerado exato (OLIVIERI, 2015).

2.5.4. Resolução multivariada de curvas com mínimos quadrados alternantes (MCR-ALS)

A Resolução Multivariada de Curvas com Mínimos Quadrados Alternantes (MCR-ALS – *Multivariate Curve Resolution – Alternating Least Squares*) é um método quimiométrico de processamento de sinais, incluído na família de técnicas de análise de fatores. Seu objetivo principal é a resolução de misturas com fins qualitativos e/ou quantitativos, por meio da determinação das contribuições dos componentes puros presentes na mistura, a partir de uma única informação que é a matriz original de dados contendo as medidas das misturas (JUAN; JAUMOT; TAULER, 2014; MARÇO *et al.*, 2014).

Tratando-se de aplicações espectroscópicas, a resolução de misturas pela técnica MCR-ALS é realizada por meio da decomposição de uma matriz de dados experimentais **D** em um modelo bilinear, composto pela multiplicação de duas matrizes, conforme **Equação 2.25**:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (2.25)$$

onde a matriz **C** representa os perfis de concentração dos compostos presentes na amostra, a matriz **S^T** representa os perfis espectrais desses compostos puros

e **E** a matriz dos resíduos, que contém a variação dos dados não explicada pelo modelo (GOICOECHEA; OLIVIERI; TAULER, 2010; JAUMOT; JUAN; TAULER, 2015). A **Equação 2.25** pode ser representada pelo esquema mostrado na **Figura 2.6**, o qual permite uma descrição visual da decomposição de dados por MCR-ALS considerando como exemplo um sistema composto por dois componentes.

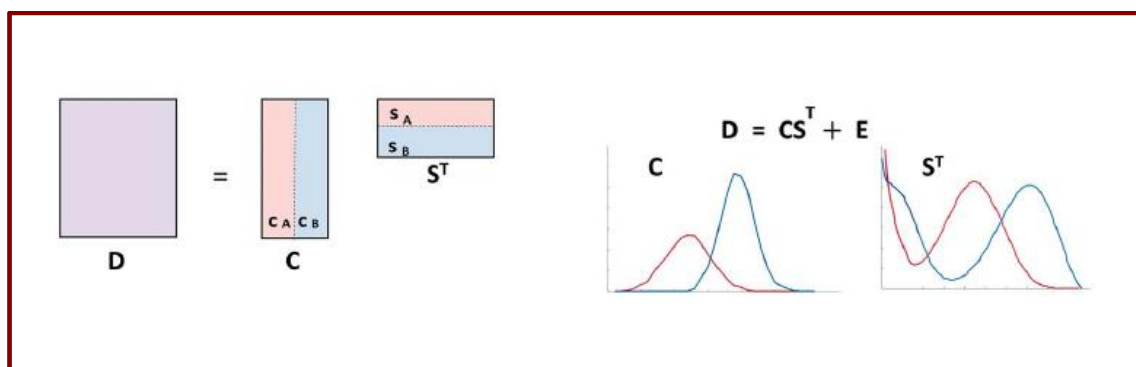


Figura 2.6. Esquema de decomposição de dados por MCR-ALS para um sistema de dois componentes. Fonte: JUAN; JAUMOT; TAULER, 2014.

Para que o MCR-ALS resolva a **Equação 2.25**, além da matriz dos dados originais, o número de componentes presentes, ou posto (*rank*), deve ser fornecido. Caso não se tenha conhecimento, o posto pode ser estimado por meio de ferramentas de análise exploratória não supervisionada que, no caso de dados de primeira ordem (dados em que se tem um vetor de resposta instrumental para cada amostra), pode ser PCA ou Decomposição em Valores Singulares (SVD – *Singular Values Decomposition*) (JUAN; JAUMOT; TAULER, 2014; MARÇO *et al.*, 2014).

Para que o processo de otimização das matrizes **C** e **S^T** tenha início, é necessário ainda fornecer ao MCR-ALS uma estimativa inicial de uma delas. Entretanto, nem sempre esta informação está disponível e, portanto, deve-se recorrer a métodos apropriados para essa finalidade. Entre esses métodos, destacam-se a Análise de Fatores Evolucionários (EFA – *Evolving Factor Analysis*) e métodos baseados na aproximação da variável pura (do inglês, *pure*) (JUAN; TAULER, 2006; GOICOECHEA; OLIVIERI; TAULER, 2010; JUAN; JAUMOT; TAULER, 2014; MARÇO *et al.*, 2014).

Uma vez que a estimativa inicial é fornecida, dá-se início à etapa de otimização iterativa por ALS, em que, a cada ciclo as matrizes \mathbf{C} e \mathbf{S}^T são calculadas, uma matriz \mathbf{D}^* é reconstruída a partir do produto dessas e comparada à matriz \mathbf{D} original. O processo de otimização iterativo continua até que o critério de convergência seja atingido, ou seja, quando a variação de resultados entre iterações consecutivas apresentar valor abaixo de um valor limite predeterminado, ou quando um certo número de iterações for excedido (GOICOECHEA; OLIVIERI; TAULER, 2010; JAYARAMAN *et al.*, 2012).

Para avaliar a qualidade dos resultados obtidos por MCR-ALS, são utilizados alguns parâmetros que comparam a matriz \mathbf{D}^* obtida a partir das estimativas de \mathbf{C} e \mathbf{S}^T com a matriz original \mathbf{D} . São eles: a percentagem de falta de ajuste (%LOF – *Lack of Fit*) e a percentagem de variância explicada (R^2) (JAYARAMAN *et al.*, 2012; JAUMOT; JUAN; TAULER, 2015), determinados pelas **Equações 2.26** e **2.27**, respectivamente:

$$\%LOF = 100 \times \sqrt{\frac{\sum e_{ij}^2}{\sum d_{ij}^2}} \quad (2.26)$$

$$R^2 = 100 \times \left(1 - \frac{\sum e_{ij}^2}{\sum d_{ij}^2}\right) \quad (2.27)$$

em que o resíduo e_{ij} é igual a $d_{ij} - d_{ij}^*$ (d_{ij} é um elemento da matriz \mathbf{D} e é o mesmo elemento na matriz reconstruída \mathbf{D}^*). Logo, espera-se que baixos valores de %LOF e valores de R^2 próximos a 1 sejam obtidos.

2.5.4.1. Ambiguidade rotacional e restrições

Conforme explicado anteriormente, a técnica MCR-ALS busca soluções para a **Equação 2.25** de forma iterativa, otimizando as estimativas iniciais de \mathbf{C} e \mathbf{S}^T de forma a ajustarem-se ao conjunto de dados experimentais \mathbf{D} . Entretanto, diferentes soluções (e combinações lineares delas) que se ajustem aos dados experimentais podem ser encontradas e, embora sejam equivalentes do ponto

de vista matemático, possuem significados físicos completamente diferentes. Ou seja, existem perfis espectrais e de concentração com diferentes formatos, capazes de reproduzir a matriz de dados experimentais com a mesma qualidade de ajuste. Essa propriedade é denominada de ambiguidade rotacional e é um grave problema em MCR-ALS (JAUMOT; TAULER, 2010). Para contornar essa dificuldade, o processo de otimização deve ser guiado pela imposição de restrições adequadas, forçando os perfis em \mathbf{C} e/ou \mathbf{S}^T a obedecer a certas propriedades químicas ou matemáticas, a partir de conhecimento a priori do sistema em estudo. Desse modo, a partir da aplicação dessas restrições, pode-se minimizar ou eliminar a ambiguidade rotacional, de maneira que soluções únicas sejam encontradas. A seguir, algumas das principais restrições disponíveis empregadas na etapa de otimização da análise MCR-ALS estão descritas.

A não negatividade é uma restrição que força com que os perfis sejam formados por valores positivos e pode ser implementada por meio de determinados algoritmos ou substituindo os valores negativos por zeros. É utilizada para perfis de concentração e, no caso de perfis espectrais, não pode ser utilizada quando pré-processamentos como SNV e derivação forem empregados (JUAN; JAUMOT; TAULER, 2014).

Em diversos tipos de sinais instrumentais os perfis de concentração têm uma forma unimodal, apresentando um único máximo. É o caso de perfis de eluição, perfis de concentração em sistemas baseados em reações e em sinais eletroquímicos. Por outro lado, os perfis espectrais geralmente não são unimodais. Sempre que a forma dos perfis for unimodal, pode ser aplicada a restrição de unimodalidade, evitando a formação de máximos secundários (JUAN; JAUMOT; TAULER, 2014).

A restrição para sistema fechado (*closure*) está relacionada a equações de balanço de massa. É aplicada quando a soma das concentrações de todas as espécies que participam da reação, ou a soma de algumas delas, é considerada constante em cada fase da reação (JUAN; JAUMOT; TAULER, 2014).

A **Figura 2.7** apresenta perfis antes e após a aplicação das restrições de não negatividade, unimodalidade e sistema fechado (*closure*).

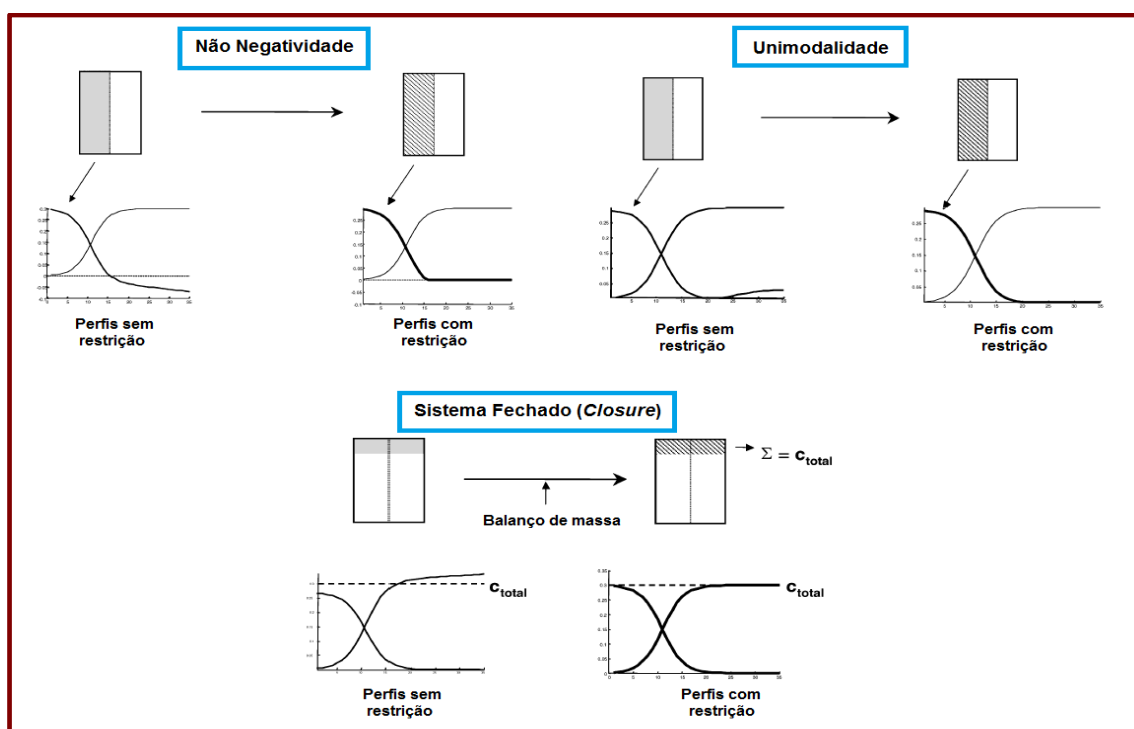


Figura 2.7. Perfis antes e após a aplicação das restrições de não negatividade, unimodalidade e sistema fechado (*closure*).

Embora as restrições acima descritas sejam muito úteis para a resolução da **Equação 2.25**, elas são utilizadas para obtenção de informações qualitativas, uma vez que os perfis de concentração são obtidos em unidades arbitrárias. Para que se obtenha informações quantitativas, é necessário aplicar a restrição de correlação, que proporciona a obtenção de perfis de concentração em unidades reais. A restrição de correlação foi proposta por ANTUNES *et al.*(2002) para análise quantitativa de íons metálicos em misturas binárias e quaternárias, utilizando dados voltamétricos. Os resultados obtidos demonstraram boa concordância com os obtidos por outras técnicas bem estabelecidas de calibração multivariada, como PLS.

Essa restrição é baseada na construção de um modelo de calibração interno que relaciona os valores conhecidos de concentração das amostras de calibração aos valores obtidos pelo algoritmo ALS para essas amostras. Dessa forma, essa restrição redimensiona os valores encontrados pelo algoritmo para a unidade real de concentração das amostras de calibração.

Em cada iteração, é realizada uma regressão entre os valores de concentração relativa calculados pelo MCR-ALS (C_{cal}^{ALS}) e os respectivos valores de referência (C_{cal}^{ref}) para os componentes da amostra, gerando um modelo descrito pela **Equação 2.28** (LYNDGAARD; VAN DEN BERG; JUAN, 2013; OLIVEIRA *et al.*, 2014):

$$C_{cal}^{ALS} = bC_{cal}^{ref} + b_0 \quad (2.28)$$

Obtidos os parâmetros do modelo (b e b_0), pode-se prever os valores das concentrações das amostras de teste através da **Equação 2.29**, em que um vetor com essas concentrações é obtido (\hat{C}_{test}) a partir dos valores de concentração calculados pelo MCR-ALS para essas amostras (C_{test}^{ALS}). Dessa forma, consegue-se obter os perfis de concentração dos componentes em unidades reais (LYNDGAARD; VAN DEN BERG; JUAN, 2013; OLIVEIRA *et al.*, 2014).

$$\hat{C}_{test} = \frac{C_{test}^{ALS} - b_0}{b} \quad (2.29)$$

A técnica MCR-ALS aliada à restrição de correlação tem possibilitado a determinação simultânea de diferentes analitos em amostras de alta complexidade na área farmacêutica, na agricultura, entre outras (AZZOUZ; TAULER, 2008; RICHARDS *et al.*, 2008; GOICOECHEA; OLIVIERI; TAULER, 2010; JAUMOT *et al.*, 2013; LYNDGAARD; VAN DEN BERG; JUAN, 2013; KOCH *et al.*, 2014; OLIVEIRA *et al.*, 2014; PARASTAR; SHAYE, 2015). A quantificação de dados de primeira ordem por MCR-ALS apresenta uma grande vantagem em relação aos métodos de calibração multivariada convencionais, conhecida como vantagem de segunda ordem: não é necessário conhecer e nem incluir no modelo de calibração, interferentes que porventura existam nas amostras desconhecidas (BOOKSH; KOWALSKI, 1994; MARÇO *et al.*, 2014). Ou seja, é possível quantificar um analito na presença de interferentes, mesmo

que esses não estejam presentes na etapa de calibração (PARASTAR; SHAYE, 2015).

2.5.4.2. Matrizes aumentadas

Uma outra maneira de reduzir a ocorrência de ambiguidades rotacionais e auxiliar na busca por uma solução mais robusta, é a aplicação de MCR-ALS utilizando matrizes aumentadas. Essa estratégia utiliza várias matrizes simultaneamente, que possuam algum tipo de informação em comum (perfis espectrais ou perfis de concentração), enriquecendo o leque de informações úteis e melhorando o entendimento de processos complexos (JUAN; TAULER, 2006; JAUMOT; JUAN; TAULER, 2015).

As configurações para as matrizes aumentadas podem ser de três tipos: aumentadas pelas linhas (*row-wise*), pelas colunas (*column-wise*) e por ambas (*row and column-wise*), conforme **Figura 2.8**.

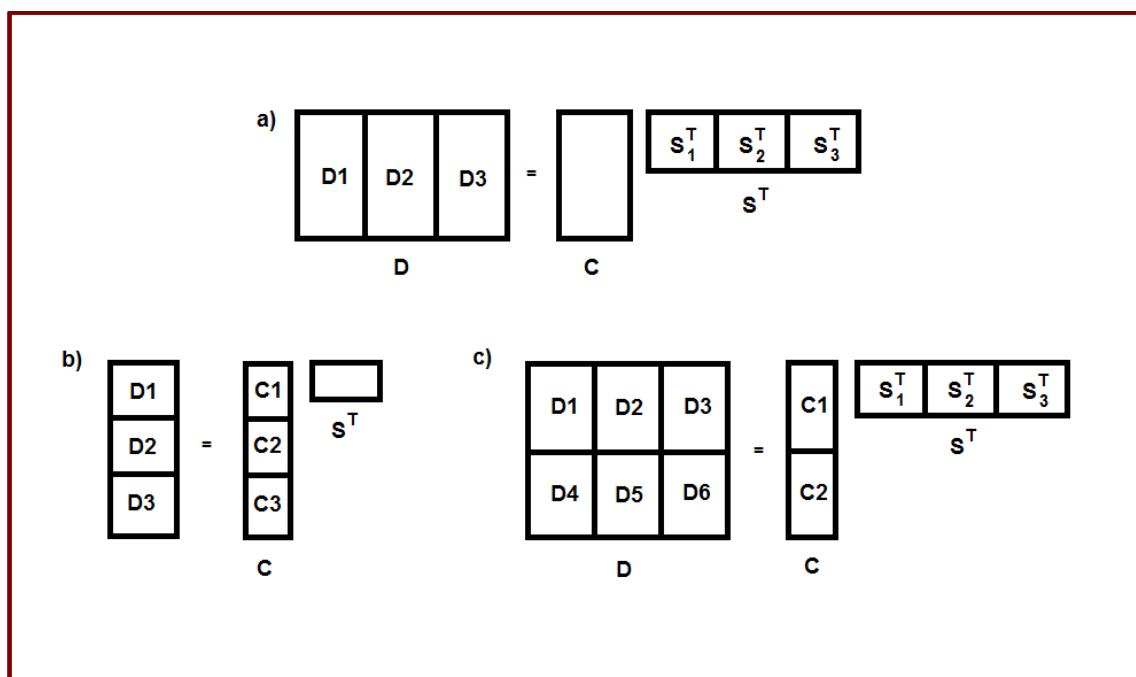


Figura 2.8. Modelo bilinear de MCR-ALS aumentado (a) pelas linhas, (b) pelas colunas e (c) pelas linhas e pelas colunas ao mesmo tempo. *Fonte:* JUAN; TAULER, 2006

As matrizes são aumentadas pelas linhas (*row-wise*) quando um determinado sistema é monitorado por mais de uma técnica instrumental. Nesse caso, os perfis de concentração são iguais e o que muda são os espectros puros obtidos para cada técnica. Dessa forma, matrizes de dados individuais relacionadas a cada técnica são organizadas e analisadas umas ao lado das outras, como demonstrado na **Figura 2.8(a)**.

Quando os dados de vários experimentos em condições diferentes (pH, bateladas, tempo, etc.), são analisados pela mesma técnica instrumental, a matriz de espectros obtida é comum para todos os experimentos e os perfis de concentração podem ser diferentes. Nessa situação, as matrizes são aumentadas na direção das colunas (*column-wise*) conforme **Figura 2.8(b)**, considerando um exemplo com três experimentos analisados pela mesma técnica. A análise simultânea de diferentes misturas dos mesmos compostos, em diferentes condições físicas ou químicas, permite a extração de informações sobre as individualidades dos sistemas, auxiliando na tarefa de resolução da **Equação 2.25**. Essa configuração permite a utilização da restrição correspondência das espécies, onde é empregada uma matriz indicando a presença ou ausência de componentes em cada uma das submatrizes **C**.

Já quando diferentes sistemas químicos são monitorados por diferentes técnicas instrumentais e em diferentes condições experimentais, temos a configuração ilustrada na **Figura 2.8(c)**, onde as matrizes são aumentadas em ambas as direções (*row and column-wise*).

2.6. Espectroscopia MIR aplicada à análise de óleos comestíveis

A espectroscopia FTIR vem se demonstrando uma técnica poderosa e eficaz na indústria alimentícia, devido a características como rapidez, permitindo a análise de grande número de amostras em pouco tempo, simplicidade, ser uma técnica não destrutiva com pouca ou nenhuma necessidade de preparação de amostras e, por isso, menor utilização de solventes perigosos em relação aos métodos analíticos tradicionais, tornando-a uma ferramenta “verde” e atrativa

(NURRULHIDAYAH *et al.*, 2011; ROHMAN; MAN, 2011a, 2011b; NENADIS *et al.*, 2013).

Fatores que contribuíram para a vasta utilização dessa técnica foram o desenvolvimento de acessórios de amostragem (RODRIGUEZ-SAONA; ALLENDORF, 2011; ROHMAN; MAN, 2011b) como acessórios ATR, que viabilizam a análise de sólidos, líquidos, semi-sólidos e filmes finos sem necessidade de preparo das amostras, e sua combinação com as técnicas de análise multivariada dos dados, que permitem extrair informações químicas relevantes, reduzindo a dimensionalidade dos dados obtidos nas análises.

Na atualidade, encontra-se na literatura grande quantidade de trabalhos que utilizam espectroscopia FTIR em conjunto com técnicas quimiométricas para avaliar a qualidade de óleos comestíveis. Grande parte das pesquisas nesta área envolve a detecção e quantificação de adulterações de óleos de elevado valor nutricional com óleos de menor valor comercial. O azeite de oliva extra virgem é um dos produtos que mais sofre adulterações por fabricantes, com o intuito de aumentar seus lucros, acarretando na redução das propriedades benéficas à saúde, que são atribuídas à sua composição rica em ácido oléico e compostos fenólicos (LERMA-GARCÍA *et al.*, 2010). Diversas pesquisas relatam que a espectroscopia FT-MIR associada a técnicas quimiométricas é uma poderosa ferramenta para detectar essas adulterações.

Nesse sentido, Rohman e Man (2010) desenvolveram metodologia para quantificar a adulteração de azeite de oliva extra virgem com óleo de palma, por meio da técnica FT-MIR com medidas de ATR, associado à modelos de calibração multivariada PLS e PCR. A análise discriminante foi utilizada para classificar amostras puras de azeite de oliva extra virgem e adulteradas com óleo de palma e outros óleos vegetais (óleo de girassol, óleo de milho e canola). Para isso restringiram a região espectral utilizada nos cálculos à faixa de 1500 cm^{-1} a 1000 cm^{-1} e realizaram dois pré-processamentos para avaliar qual deles forneceria melhores resultados de RMSEP e porcentagens de classificações corretas. Os melhores resultados foram obtidos com os espectros derivativos de 1ª ordem, com valores de RMSECV de 0,285 e 0,373 para PLS e PCR,

respectivamente, R^2 de 0,998 para ambos os modelos e 100% de classificação correta (ROHMAN; MAN, 2010).

A eficiência da espectroscopia FT-MIR associada à análise multivariada dos dados também foi demonstrada na determinação do teor de azeite de oliva extra virgem em misturas com óleos de soja e de milho (ROHMAN; MAN, 2011b). Análises das misturas ternárias em diferentes concentrações foram realizadas por ATR. Três regiões que apresentaram pequenas diferenças entre os espectros devido à composição dos óleos nas misturas foram utilizadas separadamente e combinadas para avaliar qual região seria mais informativa, baseado nos valores de R^2 obtidos com a construção de modelos PLS e PCR. Seleccionadas as regiões mais informativas, o melhor pré-processamento foi definido de acordo com os maiores valores de R^2 e menores valores de RMSEC. Com os parâmetros otimizados, os modelos foram então utilizados para prever as concentrações de amostras de previsão, obtendo-se valor de RMSEP de 0,975 para óleo de oliva extra virgem e R^2 igual a 0,999, sendo os melhores resultados obtidos com PLS.

Adultrações de azeite de oliva extra virgem com óleo de avelã, girassol, milho e soja foram analisadas com FT-MIR no modo absorvância e quantificadas por MLR precedida de seleção das variáveis informativas através do algoritmo *stepwise*. A região seleccionada para construção dos modelos para as 4 combinações binárias dos óleos foi a região de transição de ligações duplas, indicando que os ácidos graxos insaturados são de grande importância para quantificação das misturas. A metodologia desenvolvida foi capaz de quantificar os adulterantes com erros médios de previsão inferiores a 2% e com limite de detecção inferior a 5% (LERMA-GARCÍA *et al.*, 2010).

Adultrações de óleo de côco virgem, um alimento funcional emergente que apresenta atividades biológicas antivirais e antimicrobianas, além de possuir um sabor singular, também têm sido investigadas por meio da espectroscopia FT-MIR e quimiometria. Misturas de óleo de côco virgem com óleo de palma (ROHMAN; MAN, 2009a) e com óleo de milho e de girassol (ROHMAN; MAN, 2011c) foram analisadas por espectroscopia FT-MIR com acessório ATR e métodos multivariados de classificação (Análise Discriminante - DA) e

quantificação (PLS). Na adulteração com óleo de palma, foram selecionadas para construção dos modelos, regiões de frequências onde variações nos perfis espectrais dos óleos foram observadas e na adulteração com óleo de milho e de girassol, regiões do espectro foram selecionadas de acordo com sua capacidade de fornecer menos ou nenhum erro de classificação entre amostras puras de óleo de côco virgem e amostras adulteradas. No caso da calibração multivariada, as regiões foram escolhidas com base nos menores valores de RMSECV e maiores valores de R^2 encontrados. Em ambos os trabalhos, 100% das amostras foram classificadas corretamente. Os teores de adulterantes foram determinados com sucesso, obtendo-se valores de R^2 e de RMSEP iguais a 0,996 e 0,953 para adulteração com óleo de palma e para adulteração com óleo de milho e de girassol os valores de RMSEP foram 0,994% e 1,060%, respectivamente, e o R^2 encontrado foi 0,998 para ambos.

A espectroscopia FT-MIR usando acessório ATR e associada a técnicas quimiométricas também foi utilizada para quantificar óleo de côco virgem em misturas ternárias com azeite de oliva e óleo de palma (ROHMAN; MAN, 2011b). As regiões espectrais utilizadas, bem como o pré-processamento mais adequado, foram selecionadas baseadas nos valores de R^2 e RMSEC obtidos com a construção de modelos PLS e PCR. Determinados esses parâmetros, concentrações de amostras de previsão foram calculadas, e os melhores resultados foram obtidos com o modelo PLS, onde valores de 0,200 e 0,999 foram encontrados para RMSEP e R^2 , respectivamente.

Óleo de linhaça e de nim (ou amargosa) também foram avaliados por meio da espectroscopia FT-MIR e análise multivariada dos dados. Adulterações desses dois óleos com um óleo vegetal (de identidade não fornecida) e azeite de oliva extra virgem em diferentes concentrações foram adequadamente preparadas e analisadas por FT-MIR com acessório ATR. Modelos PLS foram construídos selecionando as regiões que apresentavam diferenças em seus espectros de acordo com a composição das amostras adulteradas, resultando em R^2 superior a 0,998 em todos os casos. Amostras de teste foram então analisadas e suas concentrações previstas. Os modelos construídos foram capazes de determinar corretamente as concentrações dos dois óleos naturais

(óleo de nim e de linhaça) com RMSEP de 3,63% e 4,56% para adulterações com óleo vegetal e azeite de oliva extra virgem, respectivamente (ELZEY; POLLARD; FAKAYODE, 2016).

O desempenho da espectroscopia FT-MIR aliado às técnicas quimiométricas também foi avaliado na autenticação do óleo de *Nigella sativa* que é utilizado como condimento e como remédio natural para asma, hipertensão, diabetes, tosse, bronquite, dor de cabeça, febre e etc. Para isso, inicialmente foi realizada PCA utilizando a composição de ácidos graxos deste e de outros óleos vegetais para identificar qual seria um possível adulterante. De acordo com o gráfico dos escores, determinou-se que o óleo de semente de uva teria a composição mais parecida com o óleo de *Nigella sativa*, e então, misturas binárias destes óleos foram preparadas e analisadas na região 4000 cm^{-1} a 650 cm^{-1} , por ATR. Tomando como referência os valores de R^2 e RMSECV, foram selecionadas três regiões do espectro para construção do modelo de calibração PLS, que determinou com sucesso os teores do adulterante óleo de semente de uva, obtendo-se RMSEP de 2,37% (v/v) (NURRULHIDAYAH *et al.*, 2011).

Utilizando também espectroscopia FT-MIR, com acessório ATR e calibração por PLS, foi possível determinar quantitativamente os teores dos adulterantes óleo de girassol e de soja em óleo de Argan. Para isso, foram preparadas e analisadas misturas binárias do óleo de Argan com os adulterantes na proporção de 0% a 30% (m/m). Diferentes pré-processamentos e regiões espectrais foram testadas para construção dos modelos PLS1 a fim de quantificar os adulterantes. Utilizando como critério baixos valores de RMSECV, as melhores regiões espectrais e pré-processamentos foram selecionados e, após construção dos modelos, suas capacidades preditivas foram testadas com conjuntos de validação externas. Os valores de R^2 obtidos foram 0,990 e 0,995 para adulteração com óleo de girassol e óleo de soja, respectivamente. Pequenas diferenças foram encontradas entre os valores reais e os valores previstos dos adulterantes, fornecendo RMSEP de 0,81 e 0,68 para adulteração com óleo de girassol e óleo de soja, respectivamente, indicando que a metodologia desenvolvida é eficiente para monitorar a adulteração de óleo de Argan com esses óleos (OUSSAMA; ELABADI; DEVOS, 2012).

Adulterações de óleo de gergelim com 2% a 20% (m/m) dos óleos de colza, soja, palma e amendoim também foram avaliadas utilizando espectroscopia FT-MIR e as técnicas de modelagem de classes SIMCA e PLSCM (*partial least squares class model*). Nenhum preparo prévio das amostras foi necessário, e os espectros foram medidos em uma célula de KBr desmontável, sem nenhuma adição de solvente. Diferentes pré-processamentos foram realizados e os melhores resultados de sensibilidade e especificidade - critérios escolhidos para avaliar a eficiência do modelo - foram obtidos com os espectros derivativos de 2ª ordem, onde todas as amostras com adulterações acima de 3% (m/m) foram corretamente classificadas (DENG *et al.*, 2012).

Banha de porco também vem sendo investigada como adulterante em óleos vegetais, pois a mistura entre eles pode ser utilizada eficientemente na produção de margarinas e gorduras vegetais, diminuindo os custos de produção destes alimentos. Porém, diferenciar banha de porco de outras gorduras ou óleos baseado apenas nos perfis de ácidos graxos não é uma tarefa simples, pois a semelhança de composição entre eles gera padrões muito semelhantes de ácidos graxos, sendo necessário o desenvolvimento de outras metodologias rápidas e sensíveis para detectar e quantificar esse adulterante. Neste sentido, foi desenvolvida metodologia baseada em espectroscopia FT-MIR e análise quimiométrica para detectar e quantificar banha de porco em óleos vegetais (ROHMAN *et al.*, 2011). Misturas dos óleos de canola, milho, girassol, soja e azeite de oliva extra virgem com banha de porco foram preparadas e analisadas por ATR. Foram construídos modelos de calibração PLS utilizando toda a região espectral e apenas a região onde variações espectrais entre as amostras foram observadas, sendo o melhor modelo obtido com a região de 1500 cm^{-1} a 1000 cm^{-1} . Esta região também foi utilizada posteriormente para classificação das amostras em dois grupos, utilizando DA: óleos vegetais puros e óleos vegetais adulterados com banha de porco. Pré-processamentos foram avaliados e, tanto a quantificação do adulterante como a classificação foram efetivos, conseguindo detectar a presença do adulterante em nível de 1% (v/v), com desvios padrão relativos de 2,32 a 5,68% e classificando corretamente praticamente 100% das amostras analisadas, com exceção de uma amostra de óleo de soja.

Espectroscopia FT-MIR por ATR combinada à PCA e análise de clusters (CA) também foi utilizada para diferenciar banha de porco de outras gorduras e óleos comestíveis (MAN; ROHMAN; MANSOR, 2011). Os autores selecionaram 16 frequências onde apareciam picos referentes às vibrações dos grupos funcionais dos ácidos graxos presentes na composição dos óleos e gorduras/ analisados. Aplicando PCA a esses dados, verificaram que as principais frequências para diferenciar banha de outras gorduras e óleos comestíveis são $2.852,8 \text{ cm}^{-1}$, seguida por $2.922,0 \text{ cm}^{-1}$ e $1.464,7 \text{ cm}^{-1}$, permitindo classificar banha de porco em sua classe por meio da CA, utilizando distância euclidiana.

Não apenas os óleos comestíveis de origem vegetal vêm sendo pesquisados, mas também óleos de origem animal como os produtos à base de óleo de peixe que têm sido comercializados como alimentos funcionais ou como medicamentos. Um exemplo de óleo de peixe bastante difundido no mercado é o óleo de fígado de bacalhau, cuja produção é de aproximadamente dez mil toneladas por ano (ROHMAN; MAN, 2011a) e confere diversos benefícios à saúde. Porém, por ser vendido a preços elevados é importante garantir a autenticidade e identificar possíveis adulterantes adicionados para aumentar lucros dos fornecedores e/ou comerciantes, garantindo a qualidade do produto a seus consumidores. Deste modo, a espectroscopia FT-MIR usando ATR, combinada com métodos quimiométricos de classificação (DA) e calibração multivariada (PLS), foi empregada com sucesso para classificar e quantificar amostras de óleo de fígado de bacalhau adulteradas com gorduras de origem animal (ROHMAN; MAN, 2009b; ROHMAN; MAN, 2011a).

Vários outros trabalhos que associam FT-MIR e quimiometria para análise de óleos comestíveis estão disponíveis na literatura científica, cuja abordagem vai além de adulterações por óleos de preços inferiores. A determinação da estabilidade térmica e o monitoramento do processo de oxidação são temas de grande interesse na área. Nesse sentido, análises de amostras de óleo de jojoba submetidas a diferentes temperaturas foram analisadas pela técnica FT-MIR, utilizando ATR, para estudar o processo de degradação oxidativo desses óleos, com o auxílio das técnicas quimiométricas PCA e MCR-ALS. Utilizando PCA foi possível diferenciar amostras de jojoba de acordo com sua origem e também de

acordo com as diferentes condições oxidativas aplicadas a essas amostras. Já MCR-ALS, através da obtenção de seus espectros puros aplicando apenas a restrição de não negatividade, permitiu identificar com sucesso as espécies produzidas ou degradadas durante o tratamento térmico e obter seus perfis de concentrações relativas (DRÉAU *et al.*, 2009b).

MCR-ALS associada à técnica FT-MIR também foi utilizada para estudar o processo de oxidação térmico ao longo do tempo de sete tipos de óleos vegetais, contendo diferentes proporções de triacilgliceróis insaturados. Quatro famílias de produtos, correspondentes aos produtos formados e degradados ao longo do processo oxidativo, foram descritas pelo MCR-ALS. Os perfis de concentração obtidos forneceram informações importantes sobre a habilidade de vários óleos em suportar a oxidação térmica e mostraram que todos os óleos estudados apresentam o mesmo processo de envelhecimento. Os tempos de indução do processo de oxidação para cada óleo, revelados em MCR-ALS a partir do aparecimento de produtos de oxidação, apresentaram boa concordância com os tempos de indução determinados pelo método dos índices oxidativos espectroscópicos, propostos pelos autores. A partir dos resultados obtidos os autores concluíram que MCR-ALS associado à técnica FT-MIR, pode ser usado para determinar a contribuição dos compostos químicos no fenômeno de oxidação e para aprofundar a comparação entre óleos de acordo com sua estabilidade (DRÉAU *et al.*, 2009a).

As técnicas FT-MIR e FT-Raman foram utilizadas para estudar o processo de oxidação térmico de seis tipos de óleos vegetais. Para isso, amostras desses óleos foram aquecidas a 160°C, sob agitação constante, durante 9 horas, nas quais a cada meia hora uma alíquota era retirada e analisada por ambas as técnicas. Uma matriz aumentada foi organizada com os dados dos óleos provenientes das duas técnicas e cálculos MCR-ALS foram aplicados a fim de investigar a oxidação desses óleos. Embora uma resolução completa dos dados sem ambiguidade rotacional não tenha sido alcançada, MCR-ALS foi capaz de descrever as alterações ocorridas durante o processo oxidativo dos óleos com um modelo de cinco componentes. Bandas devido à formação de compostos de carbonila, à formação de duplas ligações *trans* e sistemas conjugados, típicos

da oxidação, foram identificadas e, em conjunto com seus perfis de concentração, foi possível verificar diferenças entre o comportamento dos óleos durante o processo de oxidação que, na maioria dos casos pôde ser associada à diferenças na composição de seus ácidos graxos (MUIK *et al.*, 2007).

Wójcicki *et al.* (2015) avaliaram o processo de oxidação do óleo de oliva extra-virgem, óleo de girassol e óleo de colza por meio das técnicas FT-MIR (com acessório ATR) e NIR (por transmitância). Para isso, amostras desses óleos foram analisadas por essas técnicas antes e após serem submetidas a aquecimento a 60°C por 15 dias. As amostras submetidas a aquecimento foram analisadas a cada 3 dias e, para cada amostra, o índice de peróxido foi medido pelo método oficial em cada condição experimental. Análises PCA permitiram a visualização de pequenas mudanças ocorridas durante a oxidação, revelando diferenças espectrais entre os óleos de girassol e de colza não oxidados e oxidados. A técnica MCR-ALS foi aplicada a todo o conjunto de dados simultaneamente (todos os tipos de óleos e as duas regiões do infravermelho), conduzindo à resolução das matrizes espectrais nas regiões MIR e NIR em três componentes, correspondendo aos produtos formados pela oxidação, às estruturas saturadas e insaturadas. Para os óleos de girassol e de colza, bons resultados de correlação foram encontrados entre os valores de peróxido e as contribuições relativas dos produtos de oxidação, estimadas por MCR-ALS. Modelos PLS foram construídos com sucesso para determinação do índice de peróxido, resultando em valores satisfatórios para os modelos construídos para NIR e MIR separadamente e para as duas regiões em conjunto.

Diferentes técnicas quimiométricas aplicadas a dados provenientes da região MIR têm sido demonstradas no estudo qualitativo e quantitativo do processo de oxidação de óleos e outros parâmetros de qualidade como triacilgliceróis, ácidos graxos, compostos polares totais, valor de carbonila, dienos e trienos conjugados, ácidos graxos *trans*, índice de gordura sólida e índice de peróxido e acidez (VAN DE VOORT *et al.*, 1996; VAN DE VOORT *et al.*, 1994; VAN DE VOORT; ISMAIT; SEDMAN, 1995; VLACHOS *et al.*, 2006; MAGGIO *et al.*, 2009; MOROS *et al.*, 2009; DUPUY *et al.*, 2010; NENADIS *et al.*, 2013; COSTA FILHO, 2014; CHEN *et al.*, 2015; TALPUR *et al.*, 2015; JIANG *et*

al., 2016). Além disso, a classificação de azeite de oliva segundo sua origem geográfica (TAPP; DEFERNEZ; KEMSLEY, 2003; BENDINI *et al.*, 2007; DUPUY *et al.*, 2010; SINELLI *et al.*, 2010) e de outros óleos conforme sua origem botânica (YANG; IRUDAYARAJ; PARADKAR, 2005) também é explorada pela combinação de métodos quimiométricos e espectroscopia MIR. Esses estudos vêm demonstrando a versatilidade e eficácia da técnica espectroscópica FT-MIR associada à análise multivariada dos dados para averiguação da qualidade de óleos comestíveis. Sua potencialidade pode ser de grande utilidade em indústrias e agências de inspeção, substituindo métodos clássicos, demorados e dispendiosos por metodologias simples, rápidas, geralmente não destrutivas, que utilizam poucos solventes e que podem determinar simultaneamente diversos parâmetros.

Apesar da vasta literatura a respeito da aplicação da FT-MIR com quimiometria a óleos comestíveis, apenas parcialmente descrita neste capítulo, não foram encontrados artigos que tratavam da aplicação da técnica em óleo de alho ou de produtos relacionados.

CAPÍTULO 3.

EXPERIMENTAL



Capítulo 3. EXPERIMENTAL

3.1. Amostras e reagentes

Cápsulas gelatinosas de óleo de alho de 11 marcas e 26 lotes distintos foram adquiridas em farmácias nas cidades de João Pessoa – PB – e Recife - PE. Óleos de soja de 4 diferentes marcas comerciais e o padrão de óleo de alho, necessários para o desenvolvimento dos modelos de classificação e calibração multivariada, foram adquiridos em supermercados de Recife - PE - e na Sigma Aldrich, respectivamente. Os reagentes utilizados para limpeza, n-hexano (99,0%) e acetona (99,8%) foram adquiridos na MERCK.

3.1.1. Preparação das misturas binárias de treinamento ou calibração

As misturas binárias para construção dos modelos de classificação e calibração foram preparadas a partir do padrão de óleo de alho da Sigma Aldrich e das 4 diferentes marcas de óleo de soja adquiridas. A faixa de concentração utilizada foi escolhida em função da concentração usual (1% de óleo de alho em óleo vegetal) das cápsulas informada na literatura (AMAGASE *et al.*, 2001; MAJEWSKI, 2014), e dos valores de concentração de óleo de alho observados em alguns rótulos das cápsulas comerciais adquiridas.

As misturas foram preparadas por pesagens de alíquotas do padrão de óleo de alho e de óleo de soja, de maneira a se obter as seguintes frações mássicas de óleo de alho: 0,00%, 0,25%; 0,50%; 0,75%; 1,00%; 1,25%; 1,50% e 1,75% (m/m). A fração mássica 0,00% corresponde às amostras de óleo de soja de diferentes marcas. Para as frações mássicas de 0,25% a 1,75% foram preparadas 4 amostras com cada uma das 4 marcas de óleo de soja, com exceção da fração 1,00%, em que mais misturas foram preparadas. No total, 123 misturas de treinamento ou calibração foram preparadas e a quantidade de misturas referentes à cada concentração encontra-se na **Tabela 3.1**.

Tabela 3.1. Quantidades de misturas para cada fração mássica (m/m) com cada marca de óleo de soja.

Concentração de Óleo de alho (% m/m)	Óleo de soja marca 1	Óleo de soja marca 2	Óleo de soja marca 3	Óleo de soja marca 4	Total de misturas
0,00%	2	1	2	2	7
0,25%	4	4	4	4	16
0,50%	4	4	4	4	16
0,75%	4	4	4	4	16
1,00%	4	4	6	6	20
1,25%	4	4	4	4	16
1,50%	4	4	4	4	16
1,75%	4	4	4	4	16

Todas as 123 misturas foram preparadas com auxílio de pipeta automática de 10 µL e as massas dos componentes de cada mistura foram pesadas em um vidro de relógio, em balança analítica calibrada RBC. Cada mistura foi agitada durante 3 minutos para garantir sua homogeneidade.

3.1.2. Preparação dos ensaios de recuperação

Os ensaios de recuperação necessários para determinação da exatidão da metodologia proposta de determinação quantitativa de óleo de alho nas cápsulas, foram preparados nos níveis de fortificação 0,50%, 1,10% e 1,70% (m/m). As fortificações foram realizadas a 3 marcas de cápsulas, por pesagens em balança analítica, de massas do padrão de óleo de alho suficientes para se obter aproximadamente cada um dos 3 níveis de concentração. Para cada marca, 3 fortificações foram realizadas para cada um dos 3 níveis, resultando em um total de 27 ensaios. A quantidade de ensaios para cada uma das 3

marcas encontra-se indicada na **Tabela 3.2** e na **Figura 3.1** temos uma ilustração da preparação dos ensaios em cada nível de fortificação, para maior entendimento do procedimento.

Tabela 3.2. Quantidade de ensaios de recuperação para cada marca de cápsula e cada nível de fortificação.

	Nível 0,50%	Nível 1,10%	Nível 1,70%	Total de ensaios por marca
Cápsula (marca 1)	3	3	3	9
Cápsula (marca 2)	3	3	3	9
Cápsula (marca 3)	3	3	3	9
Total de ensaios por nível de fortificação	9	9	9	-

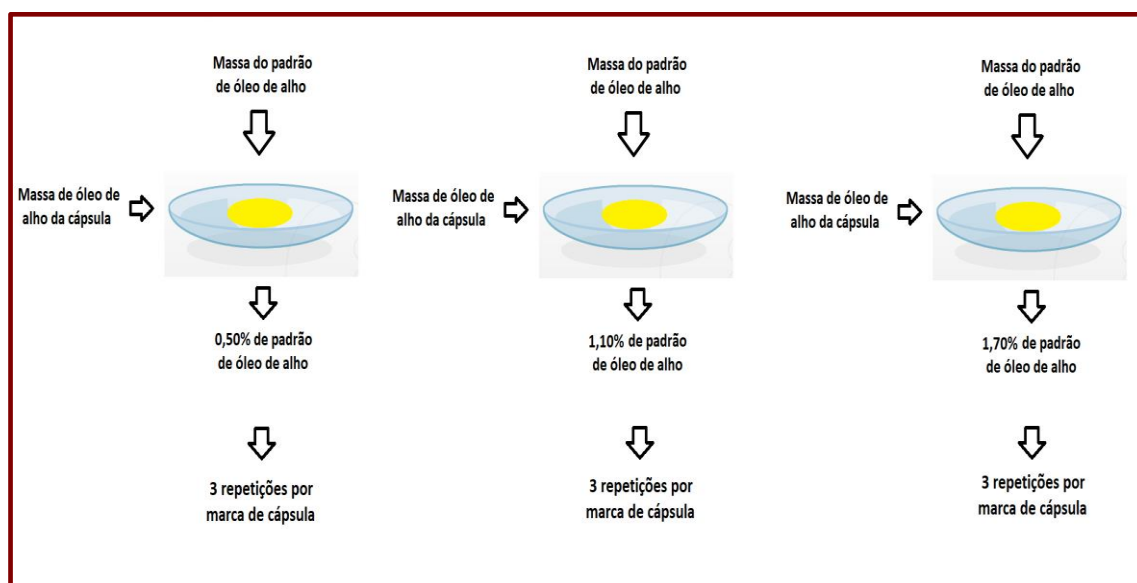


Figura 3.1. Ilustração da preparação dos ensaios de recuperação.

Todas os 27 ensaios foram preparados com auxílio de pipeta automática de 10 μL . As massas dos componentes foram pesadas em vidros de relógio, em balança analítica calibrada RBC, permanecendo sob agitação constante durante 3 minutos, para garantir sua homogeneidade.

3.1.3. Cápsulas de óleo de alho

As quantidades de marcas, lotes e amostras por concentração (m/m de óleo de alho), das 11 marcas de cápsulas adquiridas, utilizadas nas etapas de previsão da classificação e da calibração, encontram-se discriminadas na **Tabela 3.3**.

Tabela 3.3. Quantidades de cápsulas, marcas e lotes de acordo com suas concentrações (m/m).

Concentrações	Quantidade de marcas	Quantidade de lotes	Quantidade de amostras
0,20% (m/m)	6	12	70
0,30% (m/m)	1	1	4
1,04% (m/m)	2	10	45
1,50% (m/m)	2	3	17
Total	11	26	136

As concentrações das cápsulas indicadas na **Tabela 3.3** foram fornecidas nos rótulos das cápsulas e, em alguns casos, quando esta informação não estava disponível, foram fornecidas diretamente pelos fabricantes. Todas as marcas adquiridas possuem óleo de alho extraído por destilação a vapor.

3.2. Aquisição dos espectros FT-IR

Os espectros das 123 misturas e 136 cápsulas foram obtidos na região do infravermelho médio, de 4000 cm^{-1} a 650 cm^{-1} , como uma média de 16 varreduras e com resolução de 4 cm^{-1} , utilizando um espectrômetro FT-IR

Spectrum 400 da Perkin Elmer (**Figura 3.2**), com o acessório de reflexão total atenuada UATR (Universal Accessory Total Reflectance) e sob temperatura controlada a 23 ± 2 ° C, resultando em uma matriz de dimensão (259 x 3351).



Figura 3.2. Espectrômetro FT-IR Spectrum 400 da Perkin Elmer com o acessório UATR. Fonte: Própria autora.

Antes do registro de cada espectro, o acessório UATR era limpo com algodão embebido primeiramente em n-hexano e depois em acetona, e aguardava-se 3 minutos para garantir que estivesse seco. Obtinha-se, então, o *background* do ar antes de cada nova leitura das misturas e/ou cápsulas, comparando-o com o *background* obtido anteriormente, para verificar a eficácia da etapa de limpeza.

A leitura de cada mistura descrita na **Tabela 3.1** foi realizada em triplicata, imediatamente após seu preparo, depositando-se um pequeno volume no acessório UATR (**Figura 3.2**) com o auxílio de pipetas de Pasteur. As médias dessas replicatas foram utilizadas para a construção dos modelos de classificação e calibração.

Os espectros das 136 cápsulas comerciais de óleo de alho (**Tabela 3.3**) foram registrados retirando-se um pequeno volume do interior de cada cápsula com a ajuda de seringas descartáveis e depositando-o sob o acessório UATR. Todas as análises foram realizadas em triplicata e seus espectros médios

compuseram o conjunto de teste utilizado na classificação e na calibração multivariada.

Para os ensaios de recuperação, os espectros foram registrados para cada amostra antes e imediatamente após a fortificação em cada nível descrito na **Tabela 3.2**. Assim como no registro dos espectros das cápsulas e misturas, utilizadas para construção dos modelos de classificação e calibração, a leitura dos ensaios de recuperação também foi realizada em triplicata, seguindo os mesmos parâmetros de leitura e procedimentos de limpeza descritos anteriormente.

3.3. Procedimento quimiométrico

3.3.1. Classificação

As 123 misturas discriminadas na **Tabela 3.1**, foram divididas em três classes de acordo com o teor de óleo de alho, de modo que cada classe englobasse uma das concentrações encontradas nas cápsulas comerciais adquiridas. Logo, a classe 1 corresponde às misturas com concentrações de 0,00% a 0,50% (dosagem baixa); classe 2 àquelas com concentrações de 0,75% a 1,25% (dosagem média, usual); e classe 3 às misturas de 1,50% a 1,75% (dosagem alta) de óleo de alho. As quantidades de misturas por classe encontram-se descritas na **Tabela 3.4**.

As cápsulas gelatinosas de óleo de alho da **Tabela 3.3** também foram divididas em classes, de acordo com as concentrações indicadas em seus rótulos, e foram utilizadas no conjunto de teste. As quantidades de lotes e amostras de cápsulas para cada uma das três classes são indicadas na **Tabela 3.5**.

Tabela 3.4. Quantidades de misturas de treinamento por classe.

	Concentrações (m/m)	Total de misturas por concentração	Total de misturas por classe
Classe 1	0,00%	7	39
	0,25%	16	
	0,50%	16	
Classe 2	0,75%	16	52
	1,00%	20	
	1,25%	16	
Classe 3	1,50%	16	32
	1,75%	16	

Tabela 3.5. Número de marcas, lotes e quantidades de amostras comerciais para cada classe.

Classes	Marcas	Lotes	Amostras
Classe 1 (0,00% a 0,50%)	7	13	74
Classe 2 (0,75% a 1,25%)	2	10	45
Classe 3 (1,50% a 1,75%)	2	3	17

O maior número de amostras referentes à classe 1 deve-se à maior variedade de marcas de cápsulas comerciais encontradas nas farmácias, cuja concentração situa-se nessa faixa.

As misturas da **Tabela 3.4** compuseram o conjunto de treinamento e foram utilizadas na seleção dos números de onda mais informativos, na construção dos modelos de classificação PLS-DA e SPA-LDA, e na validação interna empregando validação cruzada.

A escolha do número ótimo de variáveis latentes para os modelos PLS-DA foi feita empregando validação cruzada, com a escolha dos grupos pela técnica *venetian blinds* (BALLABIO; CONSONNI, 2013). Foram escolhidos 5 grupos de validação, permitindo que mais misturas fossem utilizadas para testar o modelo e evitando, assim, que a habilidade preditiva do modelo fosse superestimada. A classificação das amostras foi realizada utilizando *thresholds* para os valores de y , determinados como sendo o valor em que o número de falsos positivos e falsos negativos é minimizado. Esses valores foram encontrados por meio de gráficos de sensibilidade e especificidade em função da variação do *threshold*, baseados nas curvas ROC que são calculadas para cada classe separadamente. As cápsulas gelatinosas de óleo de alho foram, então, classificadas como pertencentes a uma determinada classe se o valor calculado de y fosse superior ao *threshold* selecionado para cada classe.

Para a construção dos modelos SPA-LDA, as variáveis foram escolhidas de acordo com a função de custo dada na **Equação (2.6)**, de modo a minimizar o risco médio de uma classificação incorreta por LDA.

A capacidade preditiva dos modelos foi avaliada de acordo os parâmetros de sensibilidade, especificidade e exatidão, descritos na **seção 2.5.2.2.1**.

3.3.2. Calibração Multivariada

Os espectros das 123 misturas preparadas a partir do padrão de óleo de alho e óleos de soja comerciais (**Tabela 3.1**) foram utilizados para construção dos modelos de calibração multivariada PLS, SPA-MLR e MCR-ALS com restrição de correlação. As mesmas misturas foram utilizadas para validação dos modelos PLS e SPA-MLR, por validação cruzada. Para avaliar a capacidade preditiva dos modelos, foram utilizados os espectros das 136 cápsulas gelatinosas de óleo de alho (**Tabela 3.3**).

A qualidade dos modelos PLS e SPA-MLR para quantificação de óleo de alho, foi avaliada a partir do coeficiente de correlação, coeficiente de determinação, RMSECV e RMSEP, tomando como referência os valores de

concentração das misturas preparadas e as concentrações indicadas nos rótulos das cápsulas. Para MCR-ALS, além desses parâmetros, foram utilizados ainda a porcentagem de falta de ajuste do modelo, a porcentagem de variância explicada, os limites de detecção e de quantificação, o teste t para avaliar a existência de erros sistemáticos, o RPD e a correlação entre os espectros puros de óleo de alho e óleo de soja obtidos pelo algoritmo, com seus respectivos espectros puros originais, obtidos a partir da leitura do padrão de óleo de alho e das amostras de óleo de soja comerciais.

A exatidão do método MCR-ALS desenvolvido foi avaliada a partir das porcentagens de recuperação obtidas para os ensaios realizados em 3 níveis de fortificação, com 3 réplicas por amostra em cada nível (**Tabela 3.2**). A porcentagem de recuperação para as amostras utilizadas nos ensaios foi calculada pela **Equação 2.23 (seção 2.5.3.4.1)**, a partir da concentração obtida pelo MCR-ALS, para o sinal resultante da subtração do espectro após a fortificação pelo espectro da mesma amostra, antes de sua fortificação, conforme ilustrado na **Figura 3.3**.

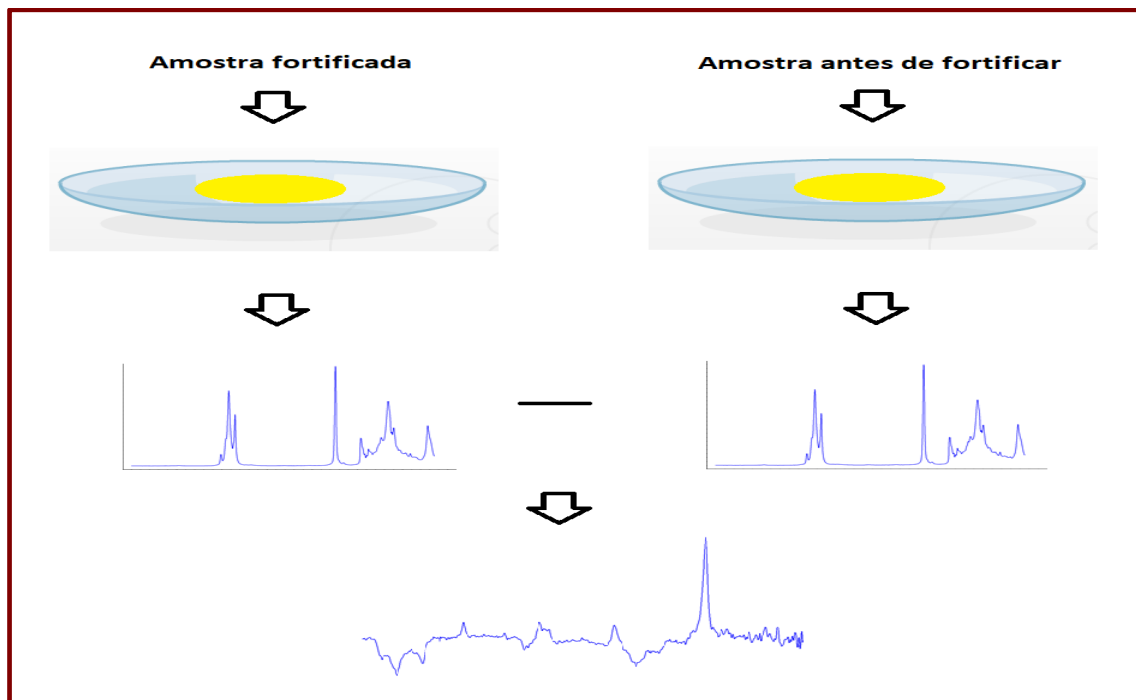


Figura 3.3. Ilustração da obtenção dos dados para a determinação da exatidão do método por MCR-ALS, desenvolvido para quantificação do teor de óleo de alho em cápsulas gelatinosas.

A partir das médias das recuperações para cada nível de fortificação, foi realizado o teste t de *Student* e comparados os valores de t calculados (**Equação 2.24**) com o valor crítico tabelado para 95% de confiança (ou 0,025 de significância) e (n-1) graus de liberdade. Todos os parâmetros avaliados foram calculados conforme descrito no **Capítulo 2**.

3.3.3. Tratamento dos dados

O pré-processamento dos dados por SNV, MSC, derivação Savitzky-Golay e a análise exploratória dos dados por PCA, foram realizados utilizando o programa Unscrambler® v9.7 (CAMO S.A.). Os cálculos SPA-LDA foram realizados em Matlab® 2010^a e PLS-DA usando *Classification Toolbox v 3.1*, desenvolvido pelo grupo de pesquisa *Milano Chemometrics and QSAR Research Group* e disponível publicamente em <http://michem.disat.unimib.it/chm/>.

Os cálculos de calibração multivariada usando PLS foram realizados empregando Unscrambler® v9.7 (CAMO AS, Oslo, Noruega) e os cálculos SPA-MLR e MCR-ALS foram realizados em Matlab® 2010^a, usando as interfaces gráficas SPA-GUI e MCR-ALS 2.0 GUI, disponíveis publicamente em <http://www.ele.ita.br/~kawakami/spa/> e <https://mcrals.wordpress.com/>, respectivamente.

CAPÍTULO 4.

RESULTADOS E DISCUSSÃO



Capítulo 4. RESULTADOS E DISCUSSÃO

4.1. Análise espectral

A **Figura 4.1** apresenta os espectros obtidos com o acessório UATR para o padrão de óleo de alho e para a média das 4 marcas de óleo de soja, na região de 4000 cm^{-1} a 650 cm^{-1} .

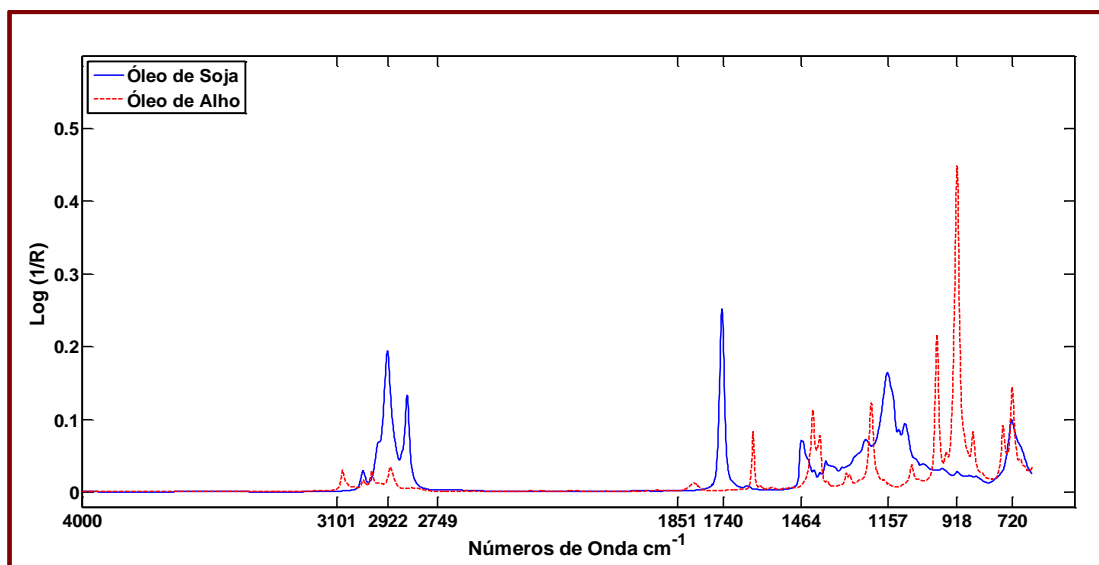


Figura 4.1. Espectros obtidos do padrão de óleo de alho e da média das 4 marcas de óleo de soja na região de 4000 cm^{-1} a 650 cm^{-1} .

As regiões de 4000 cm^{-1} a 3101 cm^{-1} e 2749 cm^{-1} a 1851 cm^{-1} não apresentam informação relevante. No espectro do óleo de soja, na região entre 3000 cm^{-1} a 2800 cm^{-1} , a absorção se dá devido a estiramento assimétrico e simétrico da ligação $-\text{CH}_2$. A intensa banda em torno de 1740 cm^{-1} é devido à vibração de estiramento da ligação $\text{C}=\text{O}$ dos grupos carboxílicos. Entre 1400 cm^{-1} e 1300 cm^{-1} , absorções são originadas pela deformação assimétrica de $-\text{CH}_2$ e deformação simétrica de grupos CH_3 . Em aproximadamente 1157 cm^{-1} , uma banda intensa é atribuída ao estiramento da ligação $-\text{C}-\text{O}$ e em aproximadamente 720 cm^{-1} , verifica-se uma banda devido às vibrações *rocking* e *wagging* de $-\text{CH}_2-$.

No espectro do óleo de alho, na região de 3100 cm^{-1} a 2800 cm^{-1} temos bandas fracas atribuídas ao estiramento assimétrico e simétrico de $=\text{CH}_2$ e estiramento de $-\text{CH}_2$. Em aproximadamente em 1634 cm^{-1} , verifica-se uma

banda atribuída ao estiramento da ligação C=C e entre 1400 cm^{-1} a 1200 cm^{-1} bandas originadas pelos estiramentos de $-\text{CH}_2$ e $\text{CH}_2=\text{CH}-$. No espectro do padrão de óleo de alho, percebem-se picos de alta intensidade que não estão presentes no espectro do óleo de soja, situados na região entre 1000 cm^{-1} e 900 cm^{-1} . Esses picos referem-se à deformação de $=\text{CH}_2$ dos grupos vinil (NISHIMURA; WIJAYA; MIZUTANI, 1988; DEVLIN; KALASINSKY; LEVIN, 1989) e ao estiramento C-S-C (WANG *et al.*, 2011; KHOSHTINAT *et al.*, 2017), presentes nos sulfetos e vinil ditiinas, os componentes majoritários do óleo de alho, formados pela decomposição da alicina. Em torno de 720 cm^{-1} , o padrão de óleo de alho apresenta uma banda correspondendo ao estiramento da ligação C-S, de intensidade proporcional à banda existente nessa mesma região no espectro do óleo de soja.

Os espectros obtidos para as 123 misturas nas diferentes concentrações e para as 136 cápsulas, na região de 4000 cm^{-1} a 650 cm^{-1} , encontram-se na **Figura 4.2**.

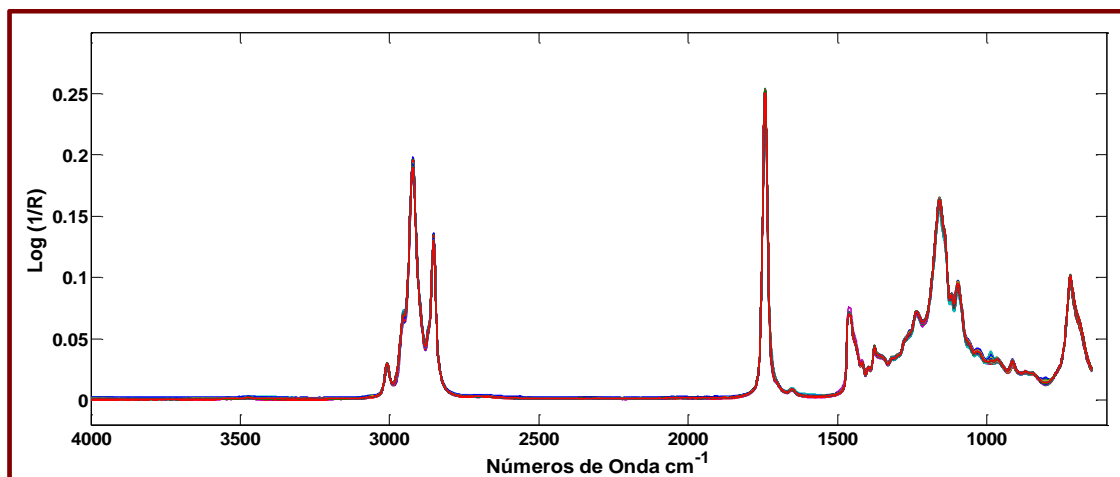


Figura 4.2. Espectros obtidos das 123 misturas em diferentes concentrações e das 136 cápsulas na região de 4000 cm^{-1} a 650 cm^{-1} .

Conforme se pode verificar, os espectros das misturas e cápsulas (**Figura 4.2**) assemelham-se muito ao espectro do óleo de soja (**Figura 4.1**), visto esses produtos e misturas serem constituídos por mais de 98% de óleo de soja. Logo, as bandas aparentes nesses espectros possuem as mesmas atribuições do espectro do óleo de soja e, na região de 1000 cm^{-1} a 900 cm^{-1} temos as bandas

de deformação de $=\text{CH}_2$ dos grupos vinil e de estiramento de C-S-C, presentes no espectro do óleo de alho. Assim como nos espectros do padrão de óleo de alho e do óleo de soja, as regiões compreendidas entre 4000 cm^{-1} e 3101 cm^{-1} e entre 2749 cm^{-1} a 1851 cm^{-1} , não apresentam bandas e, portanto, não contribuem com informações relevantes à análise dos dados. Logo, essas regiões não devem ser consideradas nos cálculos. A **Figura 4.3** apresenta os espectros das 123 misturas de treinamento e 136 amostras de cápsulas gelatinosas de óleo de alho sem essas regiões.

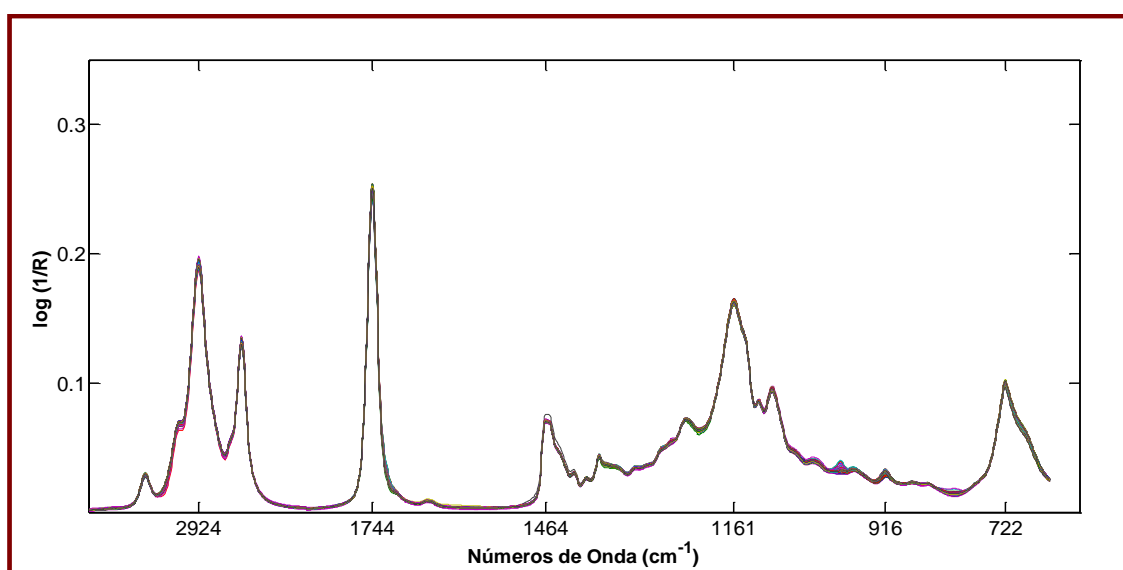


Figura 4.3. Espectros das 123 misturas de treinamento e 136 amostras de cápsulas gelatinosas de óleo de alho nas regiões compreendidas entre 3100 cm^{-1} e 2750 cm^{-1} e entre 1850 cm^{-1} a 650 cm^{-1} .

Analisando a região espectral de trabalho selecionada, percebe-se inclinação de linha de base e leve espalhamento dos dados, provavelmente devido à presença de micro bolhas originadas durante agitação das misturas. Os espectros foram então submetidos aos pré-processamentos SNV e MSC para corrigir o efeito de espalhamento da radiação. Para corrigir a inclinação da linha de base, foi realizada 1ª derivada Savitzky Golay, com polinômio de 2ª ordem e janela de 11 pontos, após testar algumas possibilidades de tamanho de janela. Os espectros pré-processados por SNV, MSC e derivação Savitzky Golay encontram-se na **Figura 4.4**.

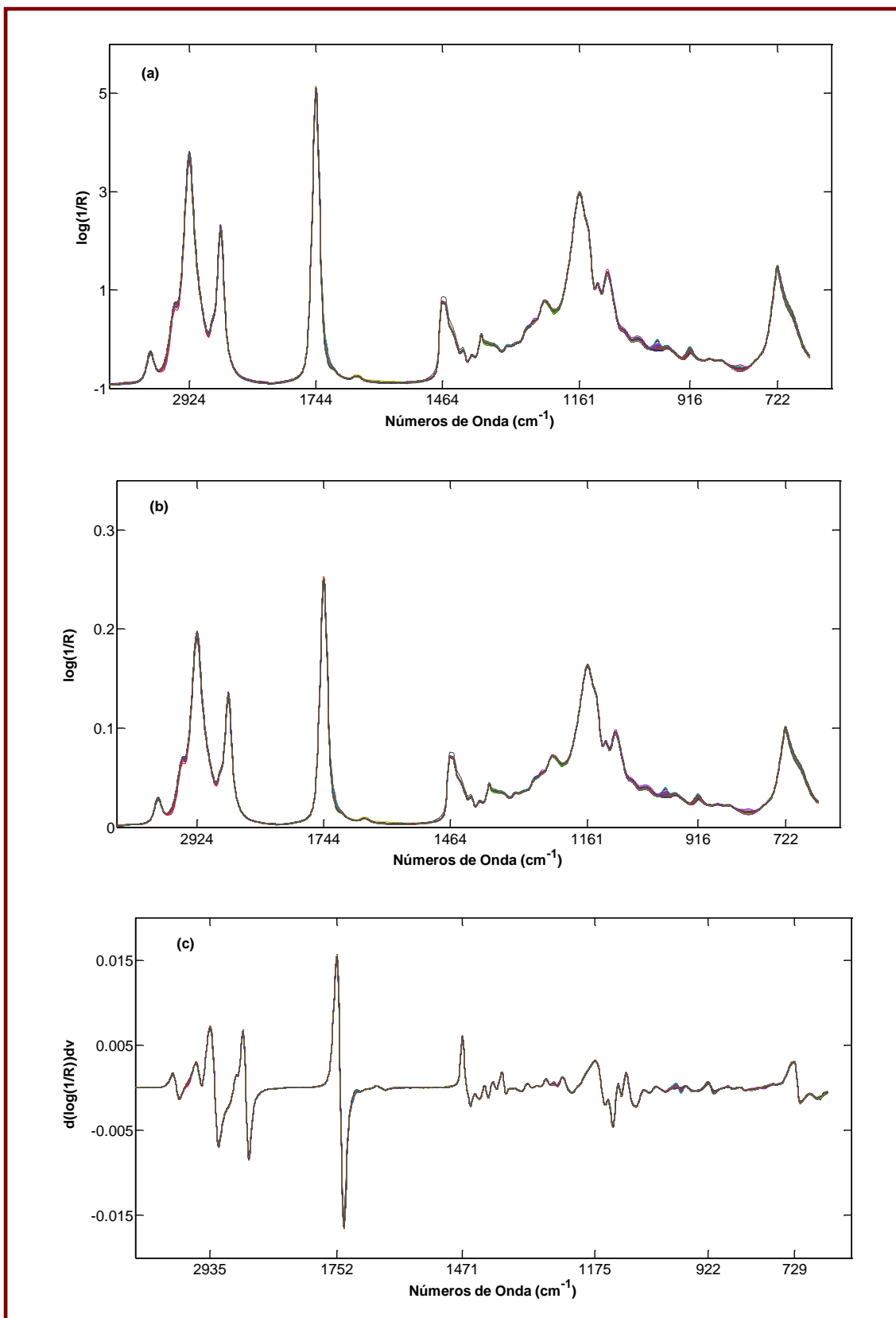


Figura 4.4. Espectros das misturas de treinamento e cápsulas gelatinosas de óleo de alho pré-processadas por (a) SNV, (b) MSC e (c) 1ª derivada Savitzky Golay, com polinômio de 2ª ordem e janela de 11 pontos.

Os espectros pré-processados por SNV e MSC aparentemente não apresentaram alterações significativas em relação aos espectros brutos. Já a 1ª derivada Savitzky Golay com polinômio de 2ª ordem e janela de 11 pontos eliminou a inclinação de linha de base. A escolha do pré-processamento mais adequado ao conjunto de dados será realizada avaliando o desempenho da predição das amostras de cápsulas a partir dos modelos de classificação construídos com cada um deles.

4.2. Análise de componentes principais (PCA)

A fim de observar a tendência natural de separação dos agrupamentos de acordo com suas frações mássicas, foi realizada a análise de componentes principais (PCA), que permite, através da redução da dimensionalidade dos dados, uma análise exploratória sem perda de informação. A PCA foi aplicada aos dados pré-processados por SNV, MSC e 1ª derivada com polinômio de 2ª ordem e janela de 11 pontos e seus gráficos de escores e pesos são apresentados a seguir (**Figuras 4.5 a 4.7**).

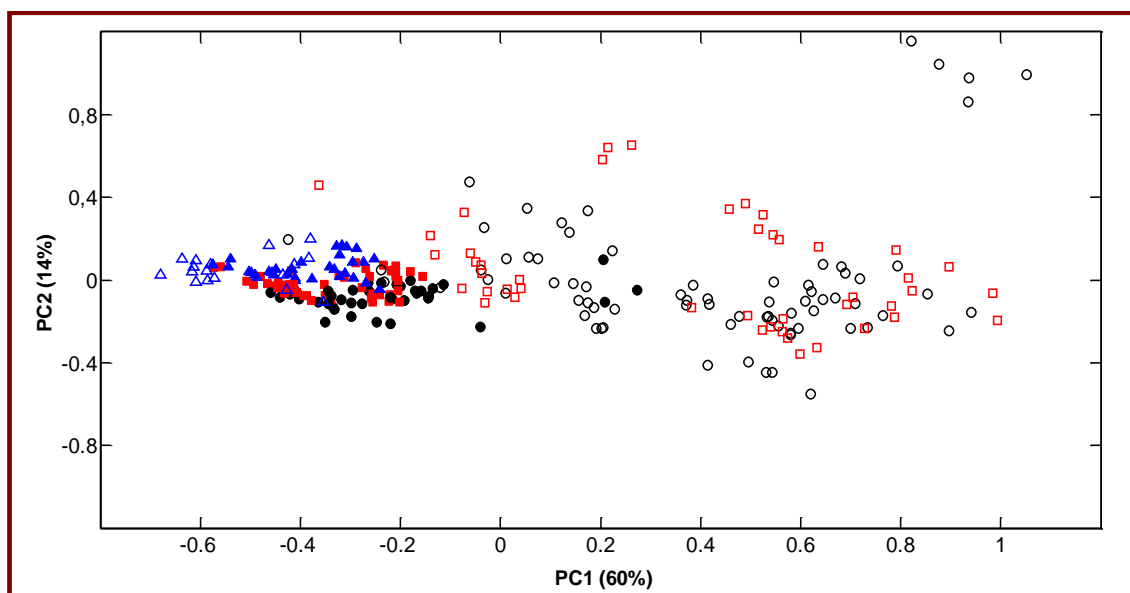


Figura 4.5. Gráfico dos escores de PC1 x PC2 para as 123 misturas de treinamento e 136 amostras de cápsulas de óleo de alho pré-processadas por SNV, utilizando toda a faixa espectral. (●) Misturas de treinamento da classe 1; (■) Misturas de treinamento da classe 2; (▲) Misturas de treinamento da classe 3; (○) Amostras da classe 1; (□) Amostras da classe 2; (△) Amostras da classe 3.

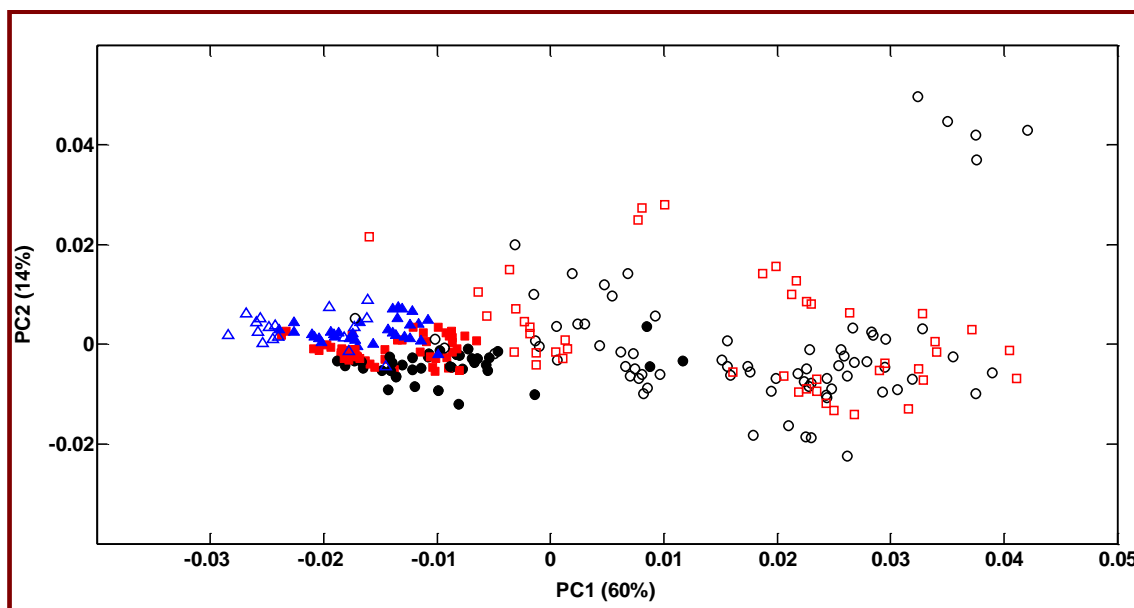


Figura 4.6. Gráfico dos escores de PC1 x PC2 para as 123 misturas de treinamento e 136 amostras de cápsulas de óleo de alho pré-processadas por MSC utilizando toda a faixa espectral. (●) Misturas de treinamento da classe 1; (■) Misturas de treinamento da classe 2; (▲) Misturas de treinamento da classe 3; (○) Amostras da classe 1; (□) Amostras da classe 2; (△) Amostras da classe 3.

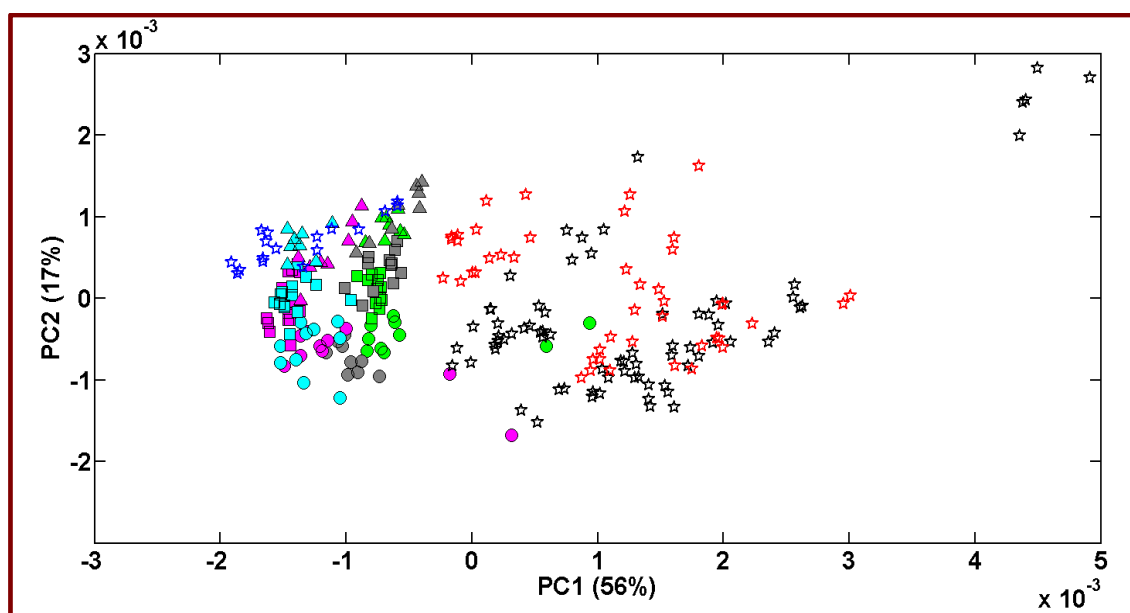


Figura 4.7. Gráfico dos escores de PC1 x PC2 para as 123 misturas de treinamento e 136 amostras de cápsulas de óleo de alho pré-processadas por 1ª derivada Savitzky Golay com polinômio de 2ª ordem e janela de 11 pontos, utilizando toda a faixa espectral. Misturas de treinamento da classe 1 - círculos; Misturas de treinamento da classe 2 - quadrados; Misturas de treinamento da classe 3 - triângulos; (☆) Amostras da classe 1; (★) Amostras da classe 2; (☆☆) Amostras da classe 3. Óleo de soja da marca 1 – verde, Óleo de soja da marca 2 – cinza, Óleo de soja da marca 3 – magenta, Óleo de soja da marca 4 – ciano.

As PCAs realizadas nos dados pré-processados por SNV e MSC apresentaram basicamente os mesmos resultados. Foram necessárias 8 componentes principais para explicar aproximadamente 95% da variância dos dados, com PC1 e PC2 explicando 60% e 14% da variância, respectivamente. Conforme gráficos dos escores (**Figura 4.5** e **Figura 4.6**) percebem-se agrupamentos das misturas de treinamento em relação às suas frações mássicas ao longo de PC2, porém com certo grau de sobreposição. O mesmo comportamento foi observado no gráfico dos escores para os dados pré-processados com 1ª derivada (**Figura 4.7**), porém, aparentemente há uma separação levemente melhor entre as misturas de treinamento em relação às suas classes, com PC1 e PC2 explicando 56% e 17% da variância dos dados, respectivamente, sendo necessárias 7 componentes principais para explicar aproximadamente 91% da variância total dos dados. Analisando os gráficos dos pesos de PC2 (**Figura 4.8 (b), (d) e (f)**), para os três pré-processamentos investigados, percebe-se que a região da carbonila (em torno de 1740 cm^{-1}), relacionada ao óleo de soja, e a região entre 1000 cm^{-1} e 900 cm^{-1} , relacionada à vibrações de deformação de $=\text{CH}_2$ e estiramento de C-S-C, dos componentes do óleo de alho, contribuem para a distribuição das misturas e amostras ao longo de PC2.

Em relação às amostras de cápsulas de óleo de alho, apenas aquelas pertencentes à classe 3 agruparam-se próximas à sua classe, independente do pré-processamento utilizado. As demais cápsulas encontram-se separadas das misturas de treinamento, ao longo de PC1. Analisando os gráficos dos pesos de PC1 (**Figura 4.8 (a), (c) e (e)**), percebe-se que, para os três pré-processamentos, as variáveis que mais contribuem para a distribuição das misturas e amostras em PC1 estão relacionadas ao espectro do óleo de soja. De fato, a região em torno de 1740 cm^{-1} , relacionada às vibrações de estiramento da ligação C=O dos grupos carboxílicos do óleo de soja, é uma das mais importantes para os 3 pré-processamentos (**Figura 4.8 (a), (c) e (e)**). Para SNV e MSC (**Figura 4.8 (a) e (c)**), a região ao redor de 2921 cm^{-1} e 2852 cm^{-1} , relacionadas à composição do óleo de soja (conforme descrito na **seção 4.1**), também contribuem para a distribuição ao longo de PC1. Logo, constata-se que PC1 explica a variância dos dados quanto à composição do óleo de soja, o que pode ser percebido na **Figura**

4.7, onde há uma ligeira separação das misturas de treinamento ao longo de PC1, de acordo com as marcas de óleo de soja utilizadas em seus preparos.

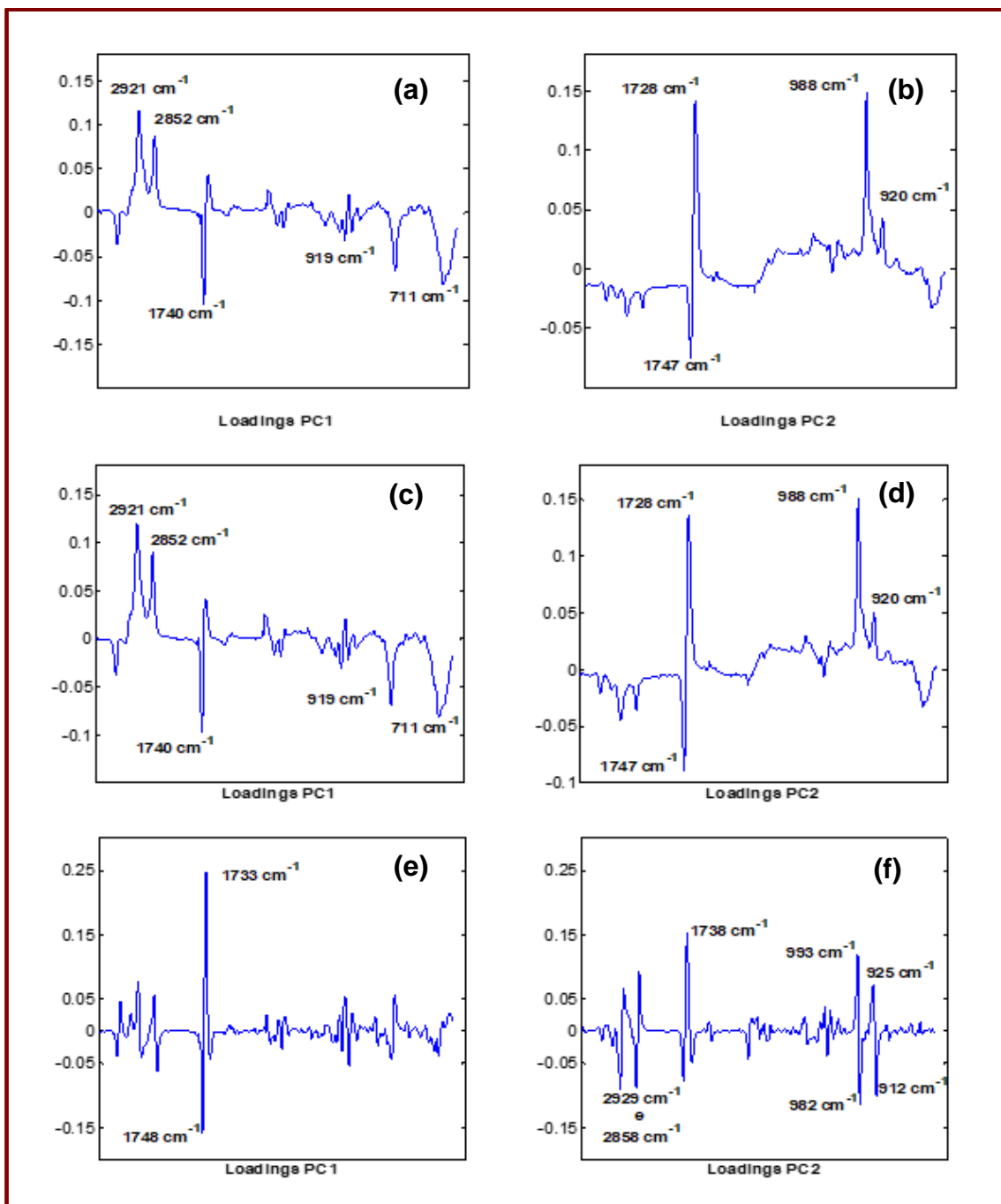


Figura 4.8. Gráfico dos pesos para PC1 e PC2 para as 123 misturas de treinamento e 136 amostras de cápsulas de óleo de alho pré-processadas por SNV (a) e (b), por MSC (c) e (d), e por 1ª derivada Savitzky Golay com polinômio de 2ª ordem e janela de 11 pontos (e) e (f), utilizando toda a faixa espectral.

Levando em consideração as informações extraídas dos gráficos dos pesos e dos escores, uma PCA dos dados levando em consideração apenas a região de 1000 cm^{-1} a 900 cm^{-1} foi realizada. Como a PCA para os dados pré-processados por SNV e MSC na região de 1000 cm^{-1} a 900 cm^{-1} apresentaram comportamentos muito semelhantes, apenas os resultados referentes à MSC e 1ª derivada serão apresentados (**Figuras 4.9 (a) e (b)**).

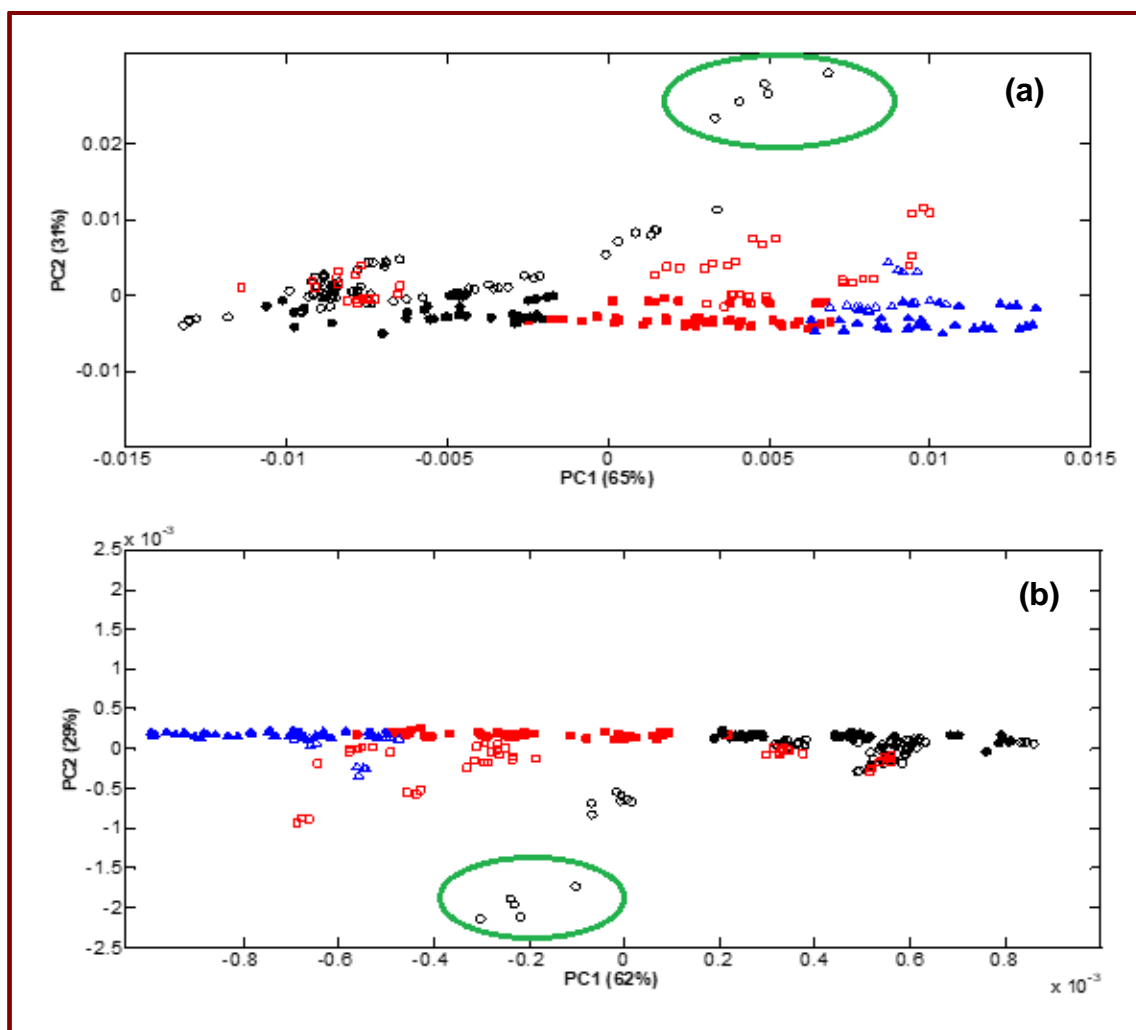


Figura 4.9. Gráfico dos escores de PC1 x PC2 para as 123 misturas de treinamento e 136 amostras de cápsulas de óleo de alho pré-processadas por (a) MSC e (b) 1ª derivada Savitzky Golay com polinômio de 2ª ordem e janela de 11 pontos, na região de 1000 cm^{-1} a 900 cm^{-1} . (●) Misturas de treinamento da classe 1; (■) Misturas de treinamento da classe 2; (▲) Misturas de treinamento da classe 3; (○) Amostras da classe 1; (□) Amostras da classe 2; (△) Amostras da classe 3. Amostras circuladas em verde apresentam comportamento diferente das demais.

As PCAs realizadas na região de 1000 cm^{-1} a 900 cm^{-1} , para os dados pré-processados por MSC e 1ª derivada, apresentaram agrupamentos melhor definidos em função do teor de óleo de alho e ao longo da PC1 (**Figura 4.9 (a)** e

(b)), diferentemente das PCAs realizadas em toda a região espectral, que mostraram uma tendência das classes agruparem-se ao longo da PC2 (**Figuras 4.5 a 4.7**). Para os dados pré-processados por MSC, a PC2 separa as misturas de treinamento em função da marca de soja utilizada em seu preparo, comportamento percebido na **Figura 4.9 (a)**. Como se pode verificar nos gráficos dos pesos (**Figura 4.10**), a região de 1000 cm^{-1} a 980 cm^{-1} possui maior intensidade em PC2 e, portanto, é a região responsável pela separação das misturas em relação à marca de óleo de soja utilizada. Como PC2 também separa algumas marcas de cápsulas das misturas, provavelmente a composição dos óleos de soja utilizados nas cápsulas e nas misturas difere entre si, originando essa distribuição ao longo de PC2. Para os dados pré-processados por 1ª derivada, apesar de não ser perceptível a diferença entre as marcas de óleo de soja utilizadas nas misturas, também ocorre a separação de algumas marcas de cápsulas ao longo de PC2. Em relação às cápsulas, para os três pré-processamentos testados, algumas amostras não se agruparam junto às misturas supostamente pertencentes às suas classes, fornecendo indícios de que possuem teor de óleo de alho distinto do indicado em seus rótulos.

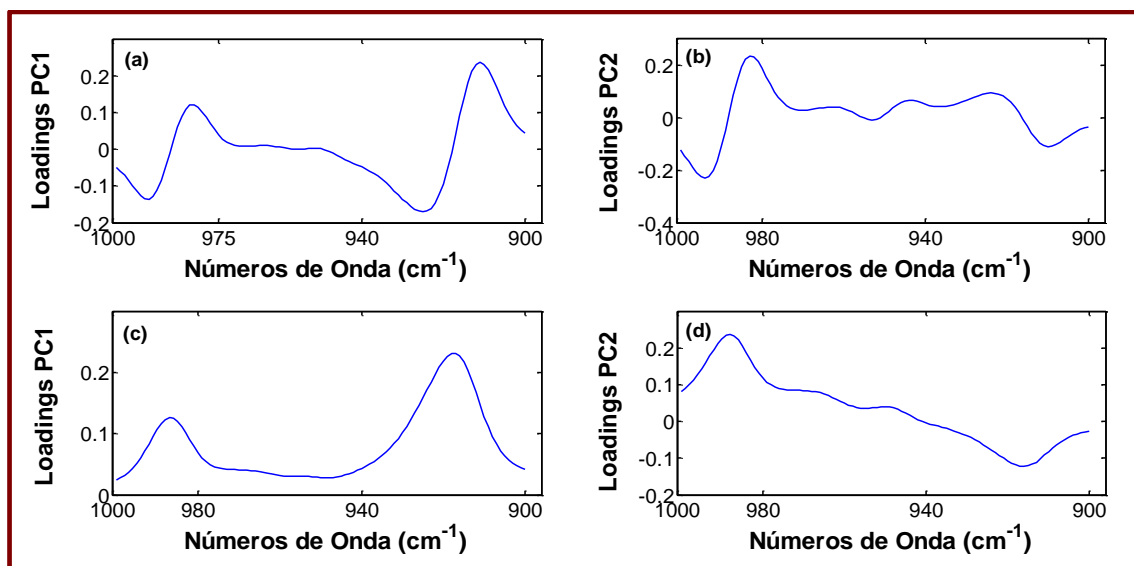


Figura 4.10. Gráfico dos pesos para PC1 e PC2 para as 123 misturas de treinamento e 136 amostras de cápsulas de óleo de alho pré-processadas por 1ª derivada Savitzky Golay com polinômio de 2ª ordem e janela de 11 pontos (a) e (b) e por MSC (c) e (d), na região espectral de 1000 cm^{-1} a 900 cm^{-1} .

Ainda em relação às cápsulas, na **Figura 4.9 (a) e (b)**, um grupo de amostras (circuladas em verde) pertencentes a uma determinada marca situa-se distante das demais. Seu comportamento distinto das demais marcas também é percebido no gráfico de influência e no próprio perfil espectral, cuja banda na região de 1000 cm^{-1} a 945 cm^{-1} apresenta intensidade pronunciada em relação às demais (**Figura 4.11**).

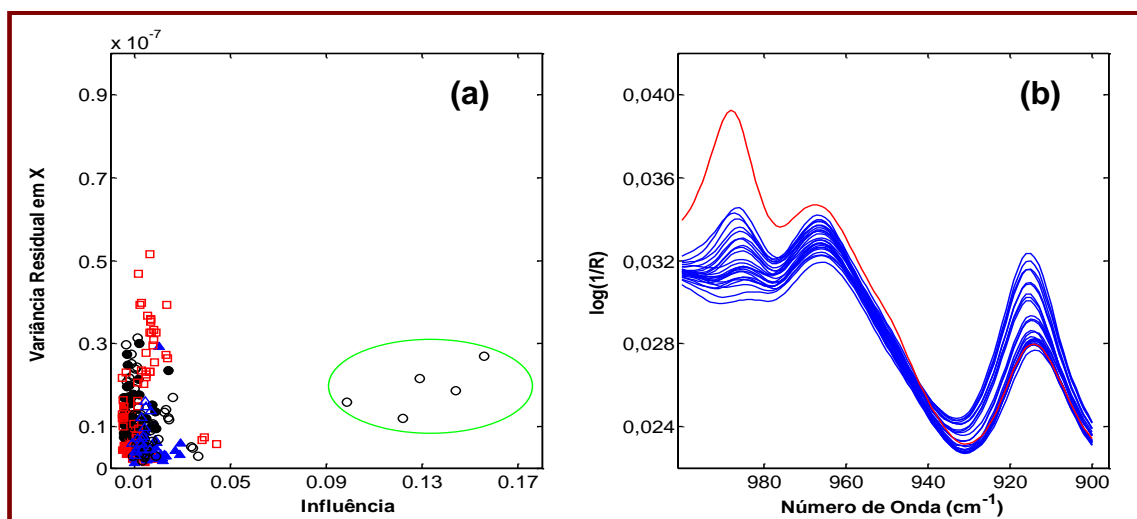


Figura 4.11. Gráfico de Influência (a) e médias dos espectros dos lotes de cápsulas (b) para os dados pré-processados por MSC, na região de 1000 cm^{-1} a 900 cm^{-1} . (●) Misturas de treinamento da classe 1; (■) Misturas de treinamento da classe 2; (▲) Misturas de treinamento da classe 3; (○) Amostras da classe 1; (□) Amostras da classe 2; (△) Amostras da classe 3. Amostras circuladas em verde em (a) correspondem ao grupo de cápsulas separadas das demais cápsulas na **Figura 4.9**. Em (b) a curva em vermelho refere-se à média dos espectros das cápsulas circuladas em verde em (a).

Ao se investigar possíveis causas para esse comportamento, verificou-se que a análise dessas amostras foi realizada no mês em que sua validade expirava e, provavelmente, essas diferenças em seus espectros são devido às alterações ocasionadas pela oxidação dessas amostras. De fato, comparando o espectro médio dessas amostras em toda a região espectral com os espectros médios de amostras do mesmo lote, 1 ano após seu vencimento, e com o espectro médio de amostras da mesma marca e de lote distinto, verifica-se algumas alterações em regiões atribuídas ao processo de oxidação do óleo de soja (MUIK et al., 2007; WÓJCICKI et al., 2015), conforme se pode observar na **Figura 4.12**.

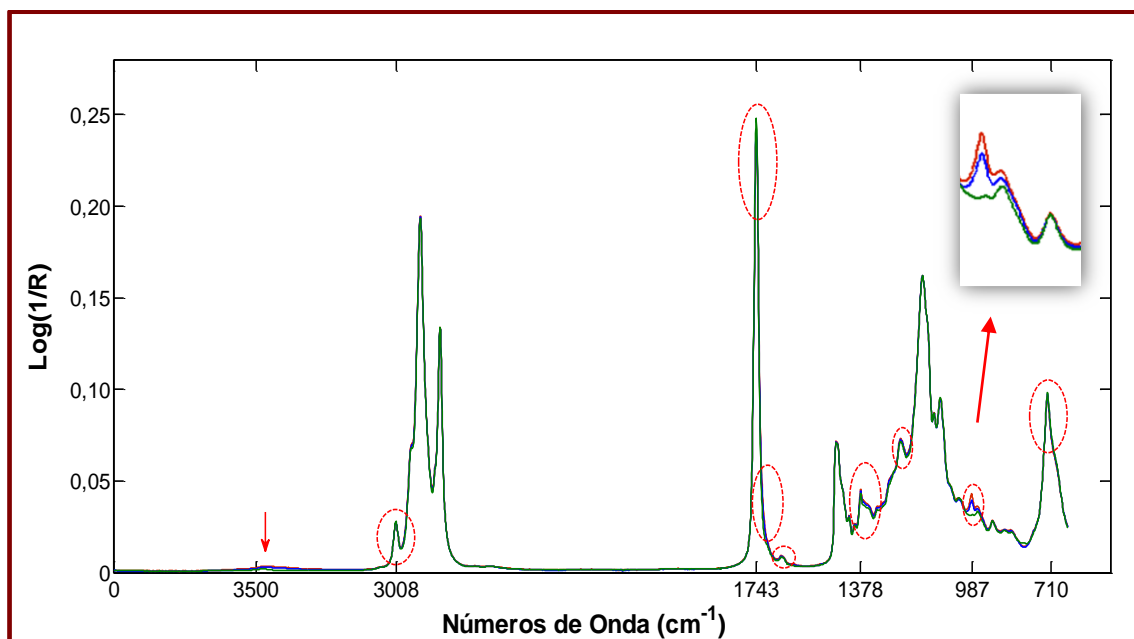


Figura 4.12. Espectros médios das amostras cujas análises foram realizadas no mês de seu vencimento (em azul), das amostras do mesmo lote 1 ano após sua data vencimento (em vermelho) e das amostras de mesma marca e lote diferente, dentro do prazo de validade (em verde). Círculos e setas em vermelho estão indicando regiões afetadas pela oxidação das amostras.

Na **Figura 4.12**, observa-se diminuição na intensidade da banda em 1743 cm^{-1} , atribuída à redução de grupos carbonila de ésteres, com consequente aumento da intensidade da banda em aproximadamente 3500 cm^{-1} e alargamento da banda em 1728 cm^{-1} , atribuídos à vibração de estiramento de grupos OH em álcoois e à presença de cetonas e aldeídos, respectivamente, formados durante o processo de oxidação. Em 710 cm^{-1} e 3008 cm^{-1} , há uma diminuição na intensidade das bandas devido à perda de duplas ligações *cis*. Como consequência, em 968 cm^{-1} e 986 cm^{-1} , ocorre aumento pronunciado das intensidades dessas bandas, originadas pela formação de duplas ligações *trans*, decorrentes do processo de oxidação. Outras regiões também apresentam aumento nas intensidades das bandas devido ao processo de oxidação, como as regiões entre 1000 cm^{-1} e 1400 cm^{-1} e 1660 cm^{-1} a 1735 cm^{-1} , atribuídas à formação de compostos de carbonila.

Visto que a região de aproximadamente 1000 cm^{-1} a 945 cm^{-1} também está relacionada ao óleo de soja e é fortemente afetada pelo processo de oxidação, conforme os perfis espectrais das amostras de cápsulas analisadas após a data de validade, conclui-se que a região mais seletiva e indicada para a

distribuição das amostras de acordo com o teor de óleo de alho é a região de 940 cm^{-1} a 900 cm^{-1} . De fato, nessa região, o espectro de óleo de alho apresenta uma banda estreita e muito intensa (**Figura 4.1**), que não é afetada pela oxidação, e que é atribuída à deformação de $=\text{CH}_2$ dos grupos vinil e ao estiramento C-S-C, presentes nos componentes majoritários do óleo de alho, conforme mencionado na **seção 4.1**.

De acordo com a análise exploratória dos dados, percebe-se que os pré-processamentos utilizados geram resultados muito semelhantes entre si e que uma melhor separação entre as classes é alcançada quando a PCA se restringe à região espectral de 1000 cm^{-1} a 900 cm^{-1} . Verifica-se ainda que algumas cápsulas provavelmente possuem teor de óleo de alho diferente do indicado em seus rótulos. Para comprovar essas observações, as técnicas de classificação PLS-DA e SPA-LDA foram aplicadas aos dados pré-processados por SNV, MSC e 1ª derivada, utilizando toda a região espectral, e os resultados principais serão apresentados a seguir.

4.3. Classificação de cápsulas gelatinosas de óleo de alho quanto ao seu teor

Técnicas de reconhecimento de padrão supervisionadas foram aplicadas aos dados com a finalidade de desenvolver um método *screening* para avaliar cápsulas de óleo de alho quanto ao seu teor. Para isso, foram utilizados os algoritmos PLS-DA e SPA-MLR, a fim de averiguar os resultados obtidos na análise exploratória dos dados por PCA, onde a região de 940 cm^{-1} a 900 cm^{-1} foi avaliada como a mais informativa para a distribuição das misturas e amostras de cápsulas quanto ao teor de óleo de alho. Os principais resultados serão apresentados a seguir.

4.3.1. PLS-DA

Para cada um dos três pré-processamentos investigados foi construído um modelo PLS1-DA, para as três classes descritas na seção **3.3.1**, utilizando todo o espectro. Entretanto, como os resultados obtidos foram muito

semelhantes, apenas os resultados com os dados pré-processados por MSC serão apresentados.

Para os dados pré-processados por MSC, 5 variáveis latentes foram selecionadas em função da menor taxa de erro obtida por validação cruzada (**Figura 4.13(a)**). Os *thresholds*, determinados para cada classe individualmente, encontram-se na **Figura 4.13(b), (c) e (d)**.

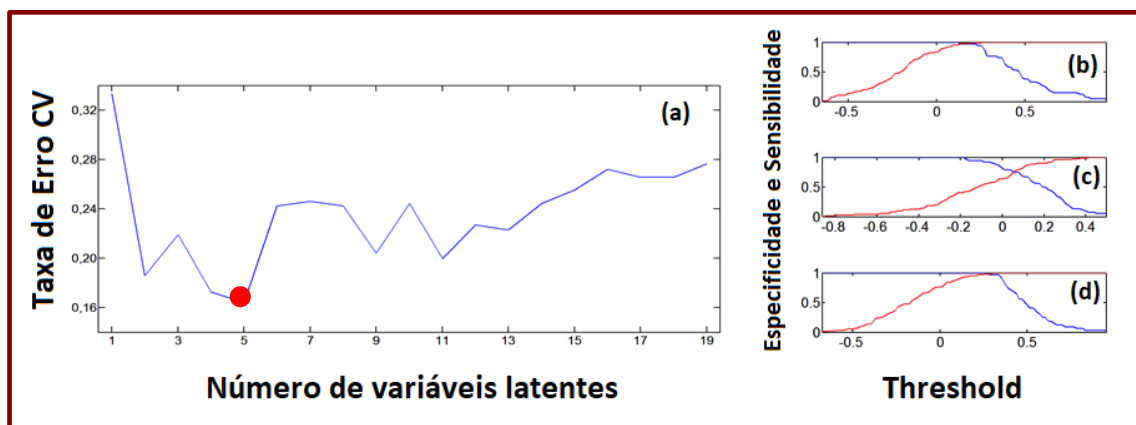


Figura 4.13. Gráfico (a) da taxa de classificações erradas em função do número de variáveis latentes utilizado na construção do modelo para os dados pré-processados por MSC, em toda a região espectral, e de sensibilidade e especificidade em função da variação do *threshold* selecionado para (b) classe 1, (c) classe 2 e (d) classe 3. Sensibilidade em azul e especificidade em vermelho.

As respostas de \mathbf{y} calculadas para as misturas de treinamento e para as amostras de teste são apresentadas na **Figura 4.14**. Conforme se observa na **Figura 4.14**, tanto misturas de treinamento como amostras de teste foram classificadas em classes não condizentes com o teor de óleo de alho indicado nos rótulos das cápsulas ou preparados, no caso das misturas de treinamento.

De fato, analisando o gráfico dos escores (**Figura 4.15**), percebemos que as misturas de treinamento e amostras de teste apresentam distribuição semelhante à exibida quando realizada PCA. Algumas marcas de cápsulas localizam-se distantes das misturas pertencentes às classes às quais deveriam pertencer, segundo as informações contidas em seus rótulos, principalmente as amostras pertencentes à classe 2.

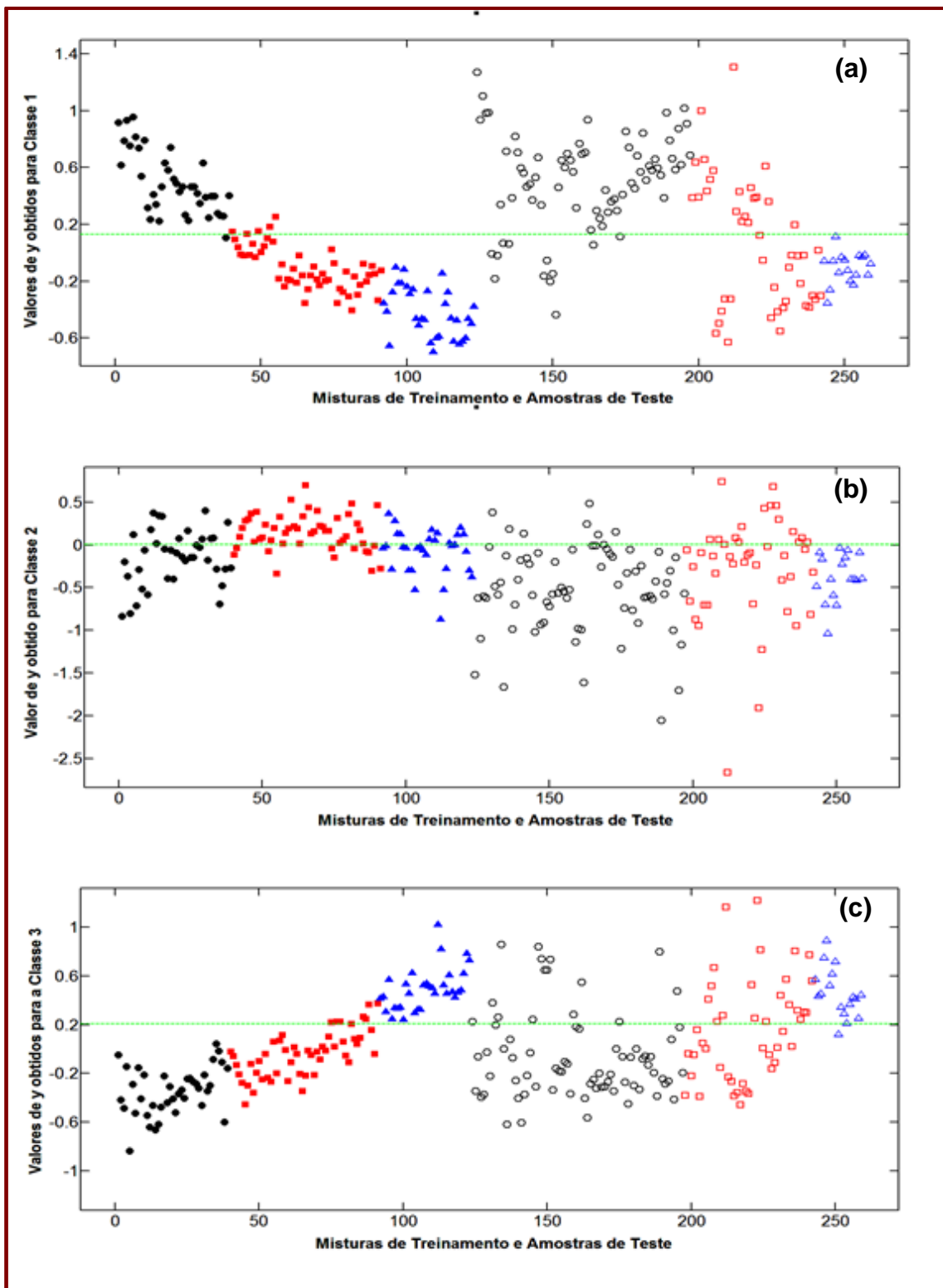


Figura 4.14. Respostas calculadas para as classes (a) 1, (b) 2 e (c) 3, versus misturas de treinamento e amostras de teste, para os dados pré-processados por MSC em toda a região espectral. (●) Misturas de treinamento da classe 1; (■) Misturas de treinamento da classe 2; (▲) Misturas de treinamento da classe 3; (○) Amostras da classe 1; (□) Amostras da classe 2; (△) Amostras da classe 3. A linha verde pontilhada indica o valor do *threshold* para cada classe.

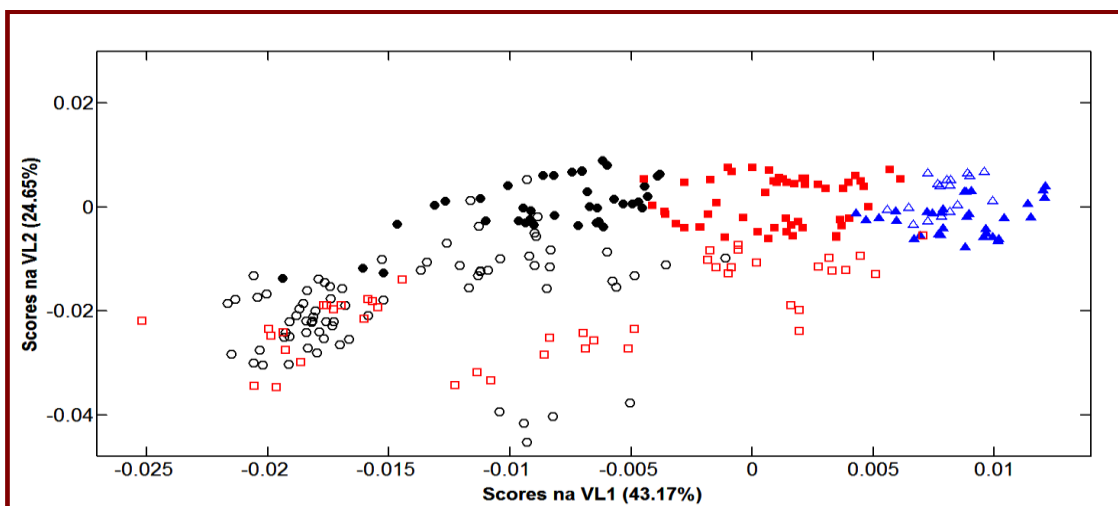


Figura 4.15. Gráfico dos escores das misturas de treinamento e amostras de teste nas duas primeiras variáveis latentes, obtidos pelo modelo PLS-DA, para os dados pré-processados por MSC em toda a região espectral. (●) Misturas de treinamento da classe 1; (■) Misturas de treinamento da classe 2; (▲) Misturas de treinamento da classe 3; (○) Amostras da classe 1; (□) Amostras da classe 2; (△) Amostras da classe 3.

A matriz de confusão obtida a partir da classificação PLS-DA, para as amostras de cápsulas de óleo de alho pré-processadas por MSC, utilizando toda a região espectral, é apresentada na **Tabela 4.1**.

Tabela 4.1. Matriz de confusão para as amostras de cápsulas de óleo de alho, obtida a partir do modelo PLS-DA construído para os dados pré-processados por MSC, em toda a região espectral.

	Número de amostras	CLASSE 1	CLASSE 2	CLASSE 3
CLASSE 1	74	61	5	8
CLASSE 2	45	17	8	20
CLASSE 3	17	0	1	16

Com base na **Tabela 4.1**, verifica-se que a maior taxa de acertos ocorreu nas amostras pertencentes à classe 3, comportamento já esperado, visto a distribuição dessas amostras no espaço reduzido, definido pelas duas primeiras variáveis latentes selecionadas pelo modelo (**Figura 4.15**). Para as amostras da classe 2, verifica-se que das 45 amostras de sua classe, 20 foram classificadas

como pertencentes à classe 3. Provavelmente isso ocorre devido à existência de uma banda de alta intensidade, nos espectros dessas amostras, na região de 1000 cm^{-1} a 945 cm^{-1} , que é atribuída não apenas ao óleo de alho, mas também está relacionada à composição do óleo de soja e é afetada por sua oxidação, conforme discutido na **seção 4.2**.

A exatidão e os valores de sensibilidade e especificidade obtidos para as amostras de teste a partir do modelo PLS-DA, são apresentados na **Tabela 4.2**.

Tabela 4.2. Valores de exatidão, especificidade e sensibilidade obtidos para as amostras de cápsulas de óleo de alho, a partir do modelo PLS-DA construído para os dados pré-processados por MSC, em toda a região espectral.

	EXATIDÃO	ESPECIFICIDADE	SENSIBILIDADE
CLASSE 1	73%	0,73	0,82
CLASSE 2		0,93	0,18
CLASSE 3		0,76	0,94

Analisando a **Tabela 4.2**, percebe-se que apesar dos valores de sensibilidade e especificidade serem satisfatórios para as classes 1 e 3, o modelo não é capaz de reconhecer as amostras pertencentes à classe 2, demonstrado no baixo valor de sensibilidade para esta classe. Comportamentos e resultados semelhantes foram obtidos para os dados pré-processados por SNV e 1ª derivada.

Com base nessas observações, conclui-se que os modelos PLS-DA construídos para os dados pré-processados por SNV, MSC e 1ª derivada com *full spectrum*, não são adequados para classificar as amostras quanto ao teor de óleo de alho contido nas cápsulas. Acredita-se que, assim como na análise exploratória dos dados (**seção 4.2**), o fato de utilizar toda a região espectral na realização dos cálculos, incluindo informações não relacionadas exclusivamente à propriedade de interesse, é a razão do desempenho insatisfatório dos modelos. Para contornar esse inconveniente, modelos de classificação utilizando seleção

de variáveis informativas foram construídos e serão apresentados na próxima seção.

4.3.2. SPA-LDA

Modelos LDA foram construídos para os três pré-processamentos investigados, com as variáveis selecionadas (**Figura 4.16**) pelo SPA em função do custo de validação.

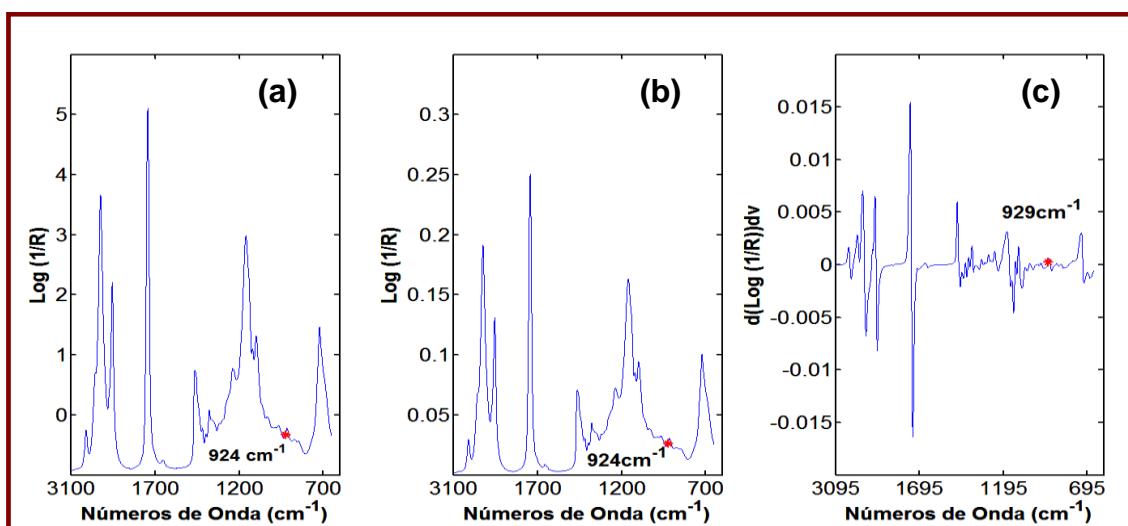


Figura 4.16. Variáveis selecionadas pelo algoritmo SPA-MLR, para os dados pré-processados por (a) SNV, (b) MSC e (c) 1ª derivada.

Como se pode observar na **Figura 4.16**, apenas uma variável foi selecionada para cada pré-processamento, sempre localizada na região onde a banda existente é atribuída ao estiramento C-S-C e à deformação de $=\text{CH}_2$ dos grupos vinil, presentes nos componentes majoritários do óleo de alho. Como resultado, apenas 4, 5 e 6 amostras foram classificadas erradamente na etapa de validação, fornecendo taxas de classificação corretas de 96,7%, 95,9% e 95,1% para os dados pré-processados por 1ª derivada, MSC e SNV, respectivamente.

Os modelos LDA foram, então, aplicados à classificação das cápsulas gelatinosas de óleo de alho, gerando a matriz de confusão apresentada na **Tabela 4.3** e os parâmetros de desempenho na **Tabela 4.4**.

Tabela 4.3. Matriz de confusão para o conjunto de teste, obtida a partir dos modelos SPA-LDA, para os três pré-processamentos utilizados.

Classes	NÚMERO DE AMOSTRAS	SNV			MSC			1ª DERIVADA		
		1	2	3	1	2	3	1	2	3
Classe 1	74	74	-	-	74	-	-	74	-	-
Classe 2	45	17	28	-	17	27	1	17	25	3
Classe 3	17	-	9	8	-	6	11	-	7	10

Tabela 4.4. Parâmetros de desempenho da classificação do conjunto de teste, obtidos a partir dos modelos SPA-LDA.

Classes	SNV			MSC			1ª DERIVADA		
	1	2	3	1	2	3	1	2	3
Exatidão	86,8%			88,2%			86,8%		
Especificidade	73%	89%	100%	73%	93%	99%	73%	92%	97%
Sensibilidade	100%	62%	47%	100%	60%	65%	100%	56%	59%

Analisando as **Tabelas 4.3 e 4.4**, verifica-se que os três pré-processamentos resultaram em desempenhos muito semelhantes. Logo, optou-se por discutir apenas os resultados referentes ao pré-processamento MSC.

Com base na **Tabela 4.4**, verifica-se que bons valores de sensibilidade e especificidade foram encontrados para a classe 1, com 100% das amostras pertencentes à essa classe classificadas de acordo com seus rótulos.

A classe 2 é constituída de duas diferentes marcas de cápsulas, cada uma contendo 5 lotes de fabricação. Entre as 45 amostras do total dessa classe (**Tabela 3.5**), uma foi atribuída à classe 3 e todas as 17 amostras que compõem três lotes de uma determinada marca foram classificadas como classe 1, apesar de seus rótulos indicarem que contêm 1% de óleo de alho. Das 20 amostras dessa classe que foram classificadas como pertencentes à classe 3 quando utilizado o algoritmo PLS-DA, 19 foram classificadas pelo SPA-MLR, cuja variável foi selecionada na região de 940 cm^{-1} a 900 cm^{-1} , como pertencentes à classe 2, conforme o indicado em seus rótulos.

A classe 3, composta por três lotes e duas diferentes marcas de cápsulas (**Tabela 3.5**), apresenta em seus rótulos teor de óleo de alho igual a 1,50%. Das 17 amostras dessa classe, 6 cápsulas foram classificadas como classe 2, sendo 4 delas o total de amostras de um específico lote.

Analisando a **Tabela 4.4**, verifica-se que os modelos apresentam ótima capacidade para rejeitar amostras não pertencentes às classes 2 e 3, demonstrada em seus altos valores de especificidade (93% e 99%, respectivamente). Quanto à sensibilidade, embora os valores encontrados para essas classes não sejam equiparados aos de especificidade, acredita-se que eles não refletem a realidade, pois, conforme observado no gráfico de escores (**Figura 4.9**) da PCA, essas amostras situam-se mais próximas às misturas de treinamento pertencentes às classes em que foram classificadas. Outro fator que corrobora para essa conclusão é que seus espectros na região que compreende a variável selecionada pelo SPA-LDA, e que é atribuída aos componentes majoritários do óleo de alho, apresentam perfis (intensidades) semelhantes às misturas das classes em que foram classificadas (**Figura 4.17**), diferentemente do que ocorreu com a classe 1, cujas amostras apresentam perfis espectrais semelhantes às misturas de sua classe. Logo, acredita-se que essas amostras realmente pertençam às classes nas quais foram atribuídas e não às classes referentes às concentrações indicadas em seus rótulos.

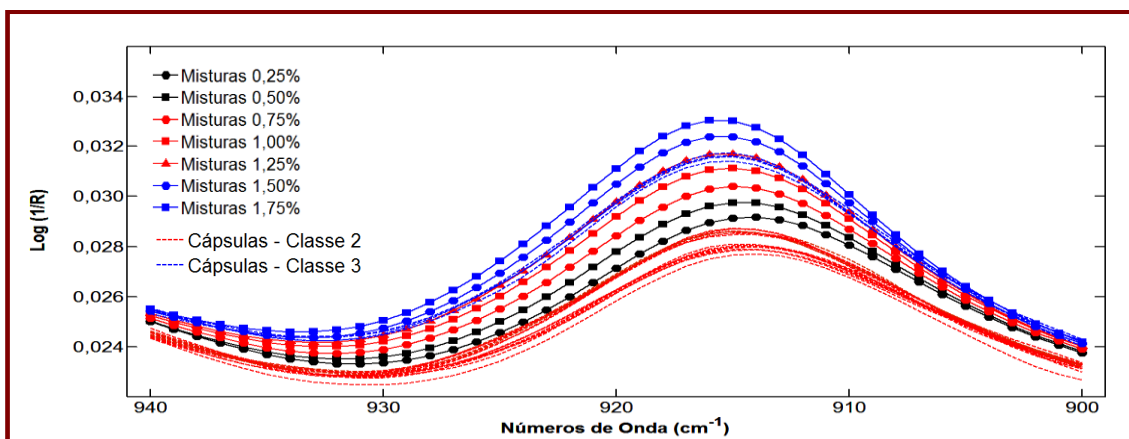


Figura 4.17. Espectros médios das misturas de treinamento a cada concentração, e das cápsulas classificadas em classes não condizentes com seus rótulos.

Para ilustrar os resultados da classificação das cápsulas gelatinosas de óleo de alho obtidos com o modelo SPA-LDA para os dados pré-processados por MSC, a **Figura 4.18** apresenta um gráfico das cápsulas e misturas de treinamento com suas respectivas intensidades na variável selecionada pelo SPA-LDA (924 cm^{-1}). Para demonstrar a separação entre a classe 1 e a classe 2, foi calculada a média das intensidades de cada classe na variável selecionada pelo SPA-LDA, e a média entre as duas classes foi usada como *threshold* (VARMUZA, KURT; FILZMOSER, 2009). O mesmo procedimento foi adotado para determinar o *threshold* entre as classes 2 e 3.

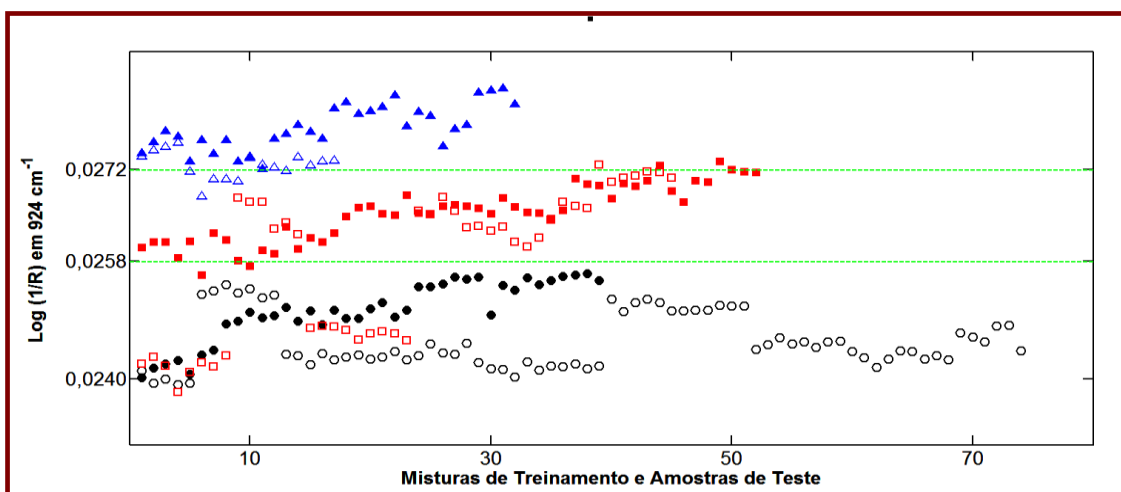


Figura 4.18. Gráfico das misturas de treinamento e amostras de teste e suas intensidades na variável selecionada pelo SPA-LDA, para os dados pré-processados por MSC. Os *thresholds* utilizados para separar as classes são indicados por linhas pontilhadas verdes. (●) Misturas de treinamento da classe 1; (■) Misturas de treinamento da classe 2; (▲) Misturas de treinamento da classe 3; (○) Amostras da classe 1; (□) Amostras da classe 2; (△) Amostras da classe 3.

Analisando a **Figura 4.18**, pode-se perceber a separação entre as três classes alcançada com SPA-LDA, demonstrando que a variável selecionada apresenta poder de discriminação significativa e fundamental para construção do modelo. Essa conclusão ratifica a informação extraída na análise exploratória dos dados por PCA, cuja região mais informativa do espectro foi diagnosticada como sendo a região de 940 cm^{-1} a 900 cm^{-1} . Assim como demonstrado na **Tabela 4.3**, na **Figura 4.18** também se pode observar algumas cápsulas situadas em classes diferentes das quais deveriam pertencer, principalmente cápsulas da classe 2, pertencentes a uma determinada marca.

A partir dos resultados apresentados, conclui-se que a maioria das cápsulas classificadas em classes distintas do indicado em seus rótulos realmente deve pertencer às classes nas quais foram classificadas. Portanto, o método proposto por SPA-MLR, é uma ferramenta simples e eficaz, que pode ser útil para a análise *screening* de cápsulas gelatinosas de óleo de alho quanto à sua concentração, visto que não há nenhuma metodologia definida para essa finalidade.

4.4. Quantificação do teor de óleo de alho em cápsulas gelatinosas

4.4.1. Modelos PLS1

Modelos utilizando toda a região espectral foram construídos para os dados pré-processados por SNV, MSC e 1ª derivada Savitzky Golay, com polinômio de segunda ordem e janela de 11 pontos. Uma rápida investigação de gráficos de influência *versus* resíduos foi realizada e nenhum *outlier* foi detectado. Visto que os resultados obtidos para os dados pré-processados por SNV mostraram-se praticamente idênticos aos obtidos por MSC, serão mostrados apenas os resultados provenientes deste pré-processamento e da 1ª derivada.

O número de variáveis latentes (VL) a ser incluído nos modelos (**Figura 4.19**) foi escolhido em função do RMSECV, conforme descrito na **seção 2.5.3.3**.

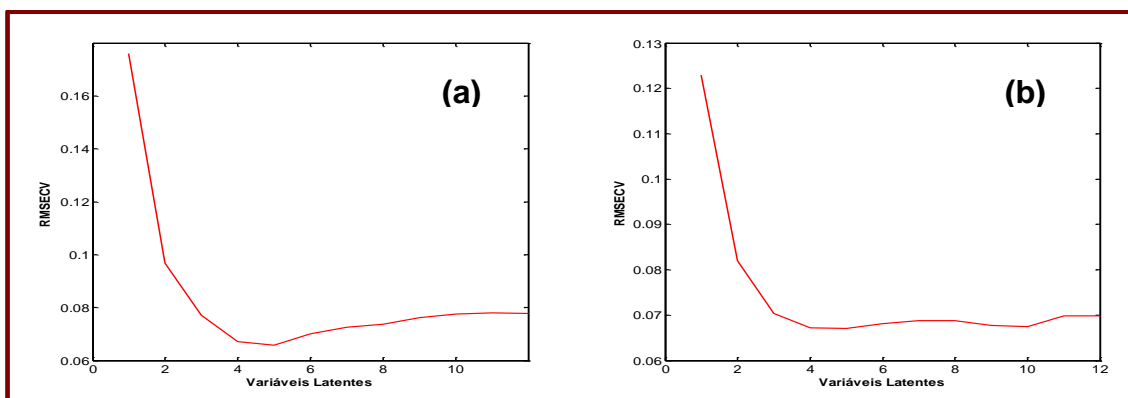


Figura 4.19. Gráficos de RMSECV em função do número de variáveis latentes utilizado nos modelos, para os dados pré-processados por 1ª derivada (a) e por MSC (b), em toda a região espectral.

Para os dados pré-processados por 1ª derivada e MSC, os modelos foram construídos com 3 e 2 VL, respectivamente (**Figuras 4.20(a) e 4.20(b)**). Com 1ª derivada, o modelo resultante explica 42%, 24% e 13% da variância em **X** e 90%, 7% e 1% da variância em **Y**, com as 3 VL escolhidas. Com MSC, o modelo explica 42% e 27% da variância em **X** e 95% e 3% da variância em **Y**, com as 2 VL escolhidas.

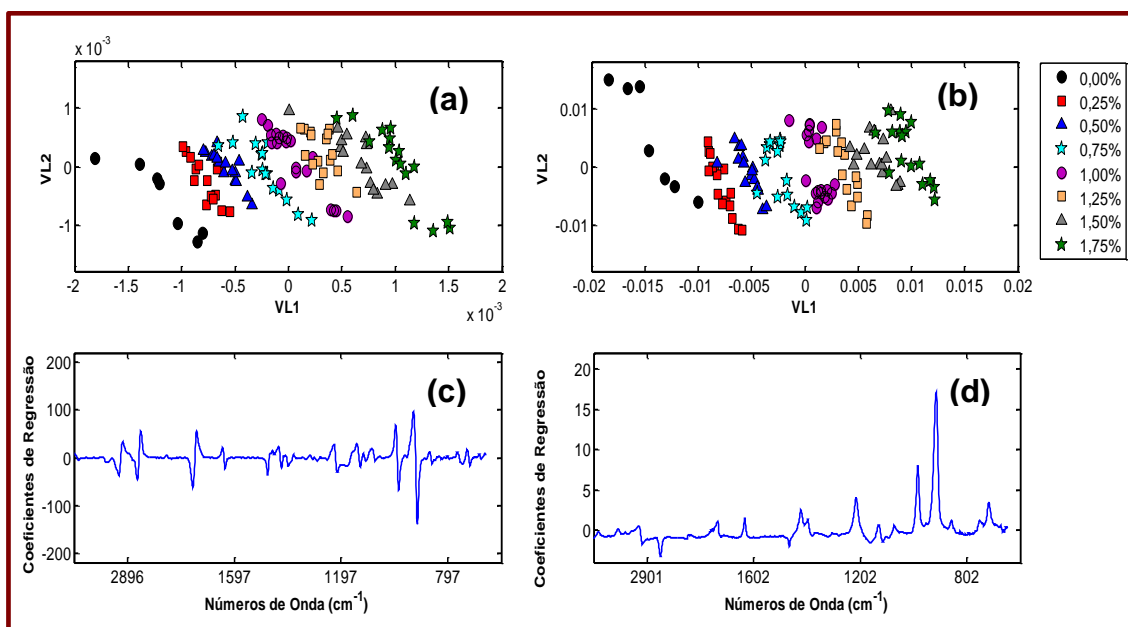


Figura 4.20. Gráficos de escores (a) e (b) e de coeficientes de regressão (c) e (d) obtidos pelos modelos PLS1, aplicados aos dados pré-processados por 1ª derivada (a) e (c) e por MSC (b) e (d), em toda a região espectral.

Como esperado, visto as análises realizadas na etapa de classificação, nos gráficos dos coeficientes de regressão (**Figuras 4.20(c) e 4.20(d)**), verifica-

se que a região mais importante para modelar **X** e **Y** é a de 940 cm^{-1} a 900 cm^{-1} . De fato, como já mencionado anteriormente, nessa região localizam-se as bandas atribuídas ao estiramento C-S-C e à deformação de $=\text{CH}_2$ dos grupos vinil (**Figura 3.2**), presentes nos componentes majoritários do óleo de alho.

Nos gráficos dos escores **Figuras 4.20(a)** e **4.20(b)**, percebe-se claramente a distribuição dos grupos ao longo da VL 1, de acordo com suas concentrações. Já a VL 2, em ambos os modelos, distingue as marcas de óleo de soja utilizadas nas misturas, com efeito um pouco mais pronunciado para os dados pré-processados por MSC, conforme se pode verificar nas **Figuras 4.21(a)** e **4.21(b)**.

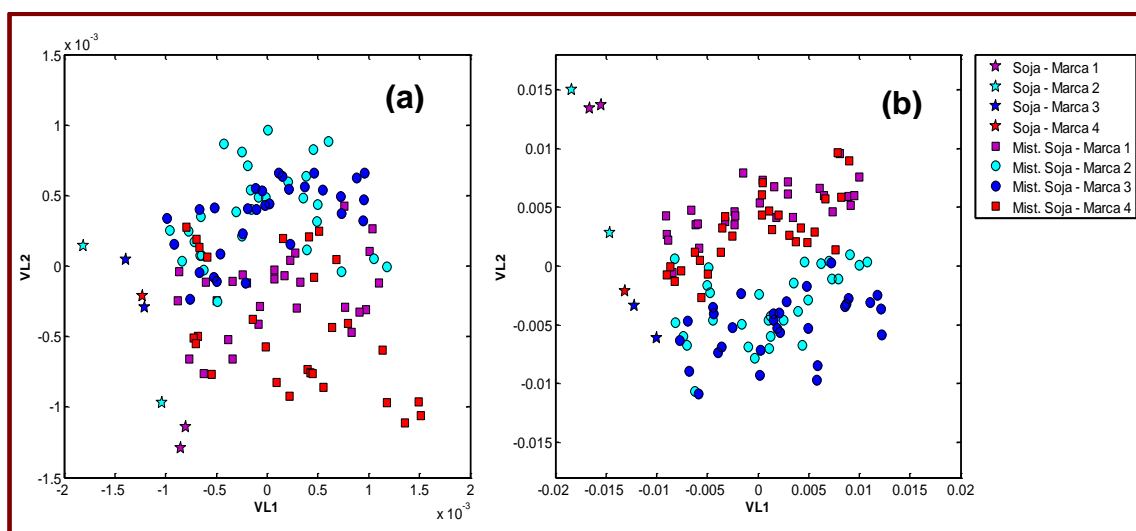


Figura 4.21. Gráficos de escores obtidos pelos modelos PLS1, aplicados aos dados pré-processados por (a) 1ª derivada e (b) MSC, em toda a região espectral, com VL 2 explicando a variância dos dados em relação às marcas de soja utilizadas nas misturas.

Quando os modelos foram construídos utilizando a região de 940 cm^{-1} a 900 cm^{-1} , apenas 1 VL (**Figura 4.22(c)**) foi suficiente para explicar a variância dos dados no modelo construído com os dados pré-processados por 1ª derivada, explicando 98% da variância em **X** e em **Y**. Para os dados pré-processados por MSC, 2VL (**Figura 4.22(d)**) foram necessárias para explicar 99% e 97% da variância em **X** e em **Y**, respectivamente. Analisando esses resultados, percebe-se que, ao se restringir a região de trabalho, a 1ª VL em ambos os modelos revelou-se capaz de explicar praticamente toda a variância dos dados em **X** e **Y**,

ao contrário do que ocorria quando toda a região espectral foi incluída nos modelos. Além disso, percebe-se que a 2ª VL não mais explica a variância dos dados relativa às marcas de soja utilizadas na preparação das misturas de calibração (Figuras 4.23(a) e 4.23(b)).

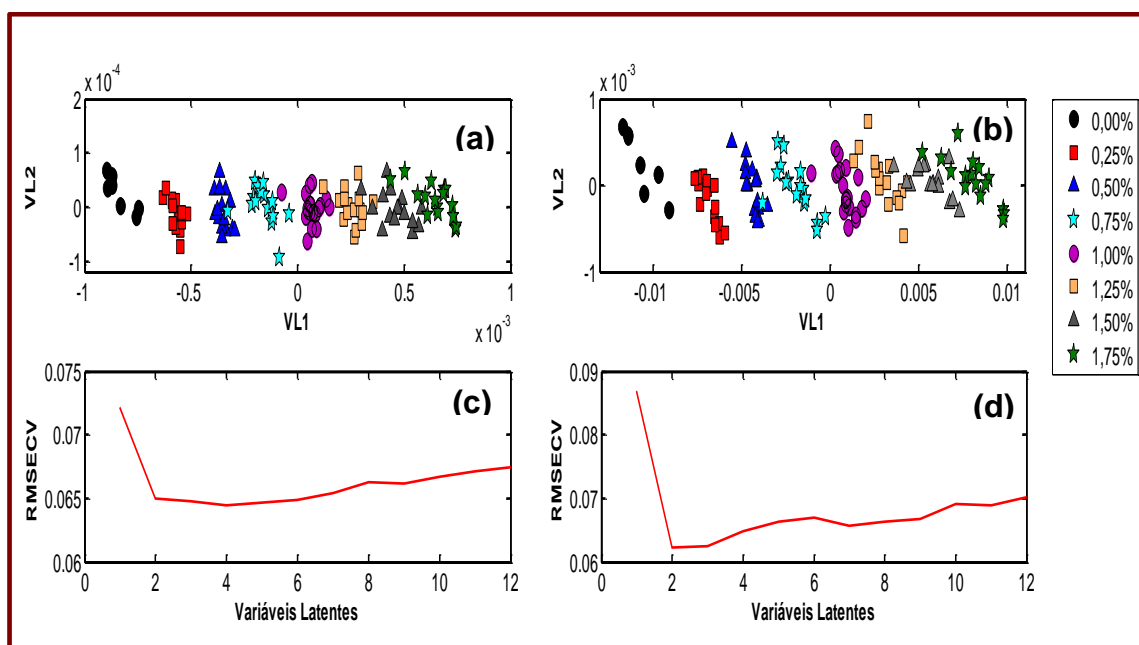


Figura 4.22. Gráficos de escores (a) e (b); e dos RMSECV em função do número de variáveis latentes (c) e (d); obtidos pelos modelos PLS1, aplicados aos dados pré-processados por 1ª derivada (a) e (b) e por MSC (b) e (d), na região de 940 cm^{-1} a 900 cm^{-1} .

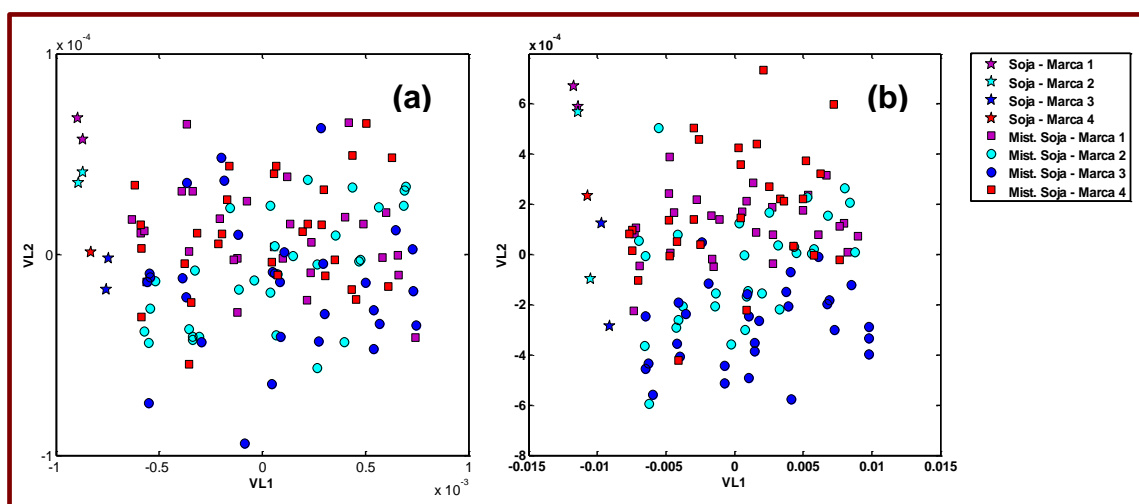


Figura 4.23. Gráficos de escores obtidos pelos modelos PLS1, aplicados aos dados pré-processados por (a) 1ª derivada e (b) MSC, na região de 940 cm^{-1} a 900 cm^{-1} , com as marcas dos óleos de soja utilizadas nas misturas diferenciados por cor.

A Figura 4.24 mostra os gráficos das concentrações medidas versus concentrações previstas para os modelos de calibração construídos para ambos

os pré-processamentos e regiões espectrais, com seus respectivos resultados na **Tabela 4.5**.

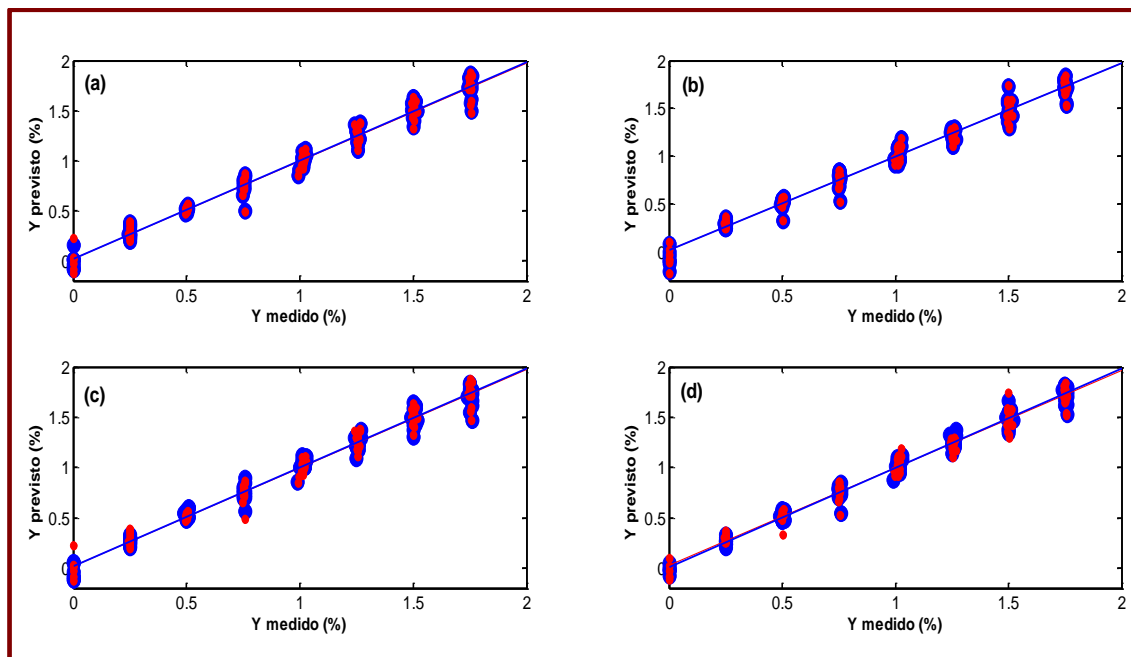


Figura 4.24. Gráficos das concentrações medidas versus as concentrações previstas na construção dos modelos de calibração PLS1, em toda a região espectral (a) e (b); e na região de 940 cm^{-1} a 900 cm^{-1} (c) e (d), para os pré-processados por 1^{a} derivada (a) e (c) e por MSC (b) e (d). Círculos azuis são os valores previstos na calibração e os vermelhos na validação.

Na **Figura 4.24**, percebe-se que os modelos construídos com ambos os pré-processamentos e regiões espectrais, estão bem ajustados aos dados, com valores de coeficiente de determinação entre 0,98 e 0,99 (**Tabela 4.5**). Não se observa a existência de desvios sistemáticos da linha ideal e tampouco a existência de *outliers* com valores de concentração previstos distantes dos seus valores de referência.

Tabela 4.5. Resultados da validação dos modelos PLS para os dados pré-processados por MSC e 1^{a} derivada, em toda a região espectral e na região de 940 cm^{-1} a 900 cm^{-1} , para a faixa de 0,00% a 1,75% m/m.

	Toda região espectral			Região de 940 cm^{-1} a 900 cm^{-1}		
	r	R ²	RMSECV	r	R ²	RMSECV
1^a Derivada	0,9907	0,9793	0,08	0,9907	0,9818	0,07
MSC	0,9881	0,9766	0,08	0,9931	0,9865	0,06

Pode-se verificar na **Tabela 4.5** que todos os modelos apresentaram valores satisfatórios de coeficiente de correlação, de determinação e RMSECV. Como os resultados no geral, são muito semelhantes entre si, a previsão das amostras comerciais de cápsulas de óleo de alho foi realizada com todos os 4 modelos construídos.

Uma vez validados os modelos, realizou-se a previsão das 136 amostras comerciais de cápsulas de óleo de alho, tomando como valores de referência os valores indicados em seus rótulos, visto não haver uma metodologia oficial estabelecida para determinar o teor de óleo de alho nas cápsulas. Na **Figura 4.25** estão os gráficos das concentrações de referência versus concentrações previstas das amostras de previsão, para cada modelo de calibração construído, com seus respectivos resultados na **Tabela 4.6**.

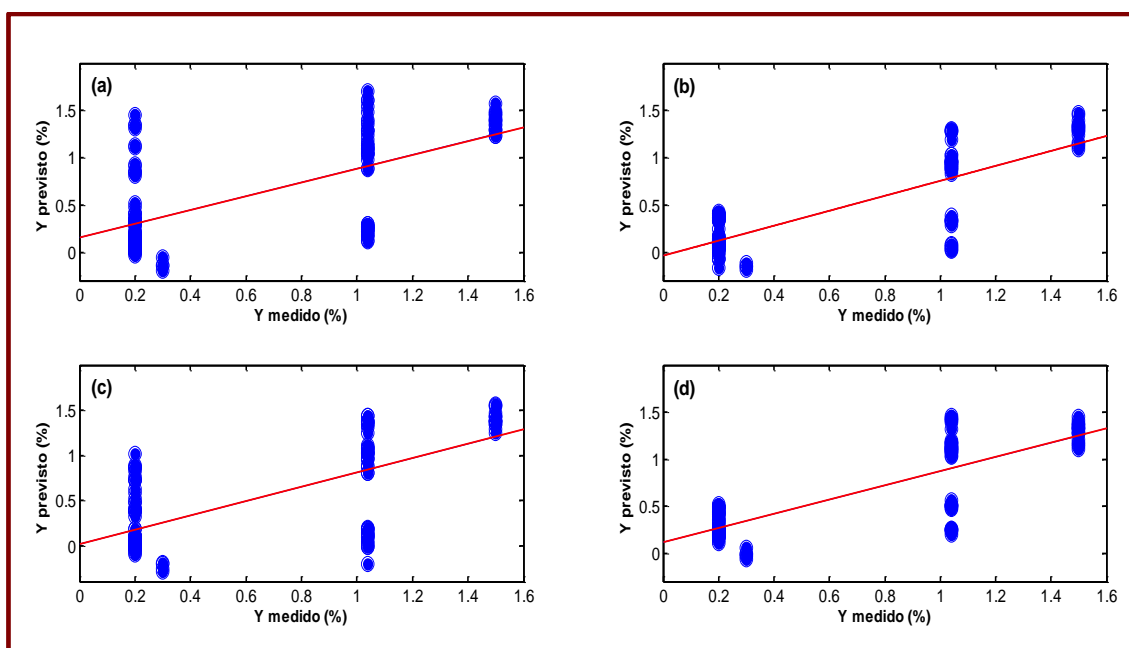


Figura 4.25. Gráficos das concentrações medidas versus as concentrações previstas na etapa de previsão, em toda a região espectral (a) e (c); e na região de 940 cm^{-1} a 900 cm^{-1} (b) e (d). Dados pré-processados por 1^{a} derivada (a) e (b) e por MSC (c) e (d).

Tabela 4.6. Resultados da previsão das amostras de cápsulas de óleo de alho, a partir dos modelos PLS construídos para os dados pré-processados por MSC e 1^{a} derivada, em toda a região espectral e na região de 940 cm^{-1} a 900 cm^{-1} , na faixa de 0,00% a 1,75% m/m.

	Toda região espectral			Região de 940 cm^{-1} a 900 cm^{-1}		
	r	R ²	RMSEP	r	R ²	RMSEP
1ª Derivada	0,6636	0,4404	0,43	0,8123	0,6600	0,35
MSC	0,7069	0,4997	0,43	0,8186	0,6517	0,29

Conforme podemos verificar na **Figura 4.25** e **Tabela 4.6**, as previsões das concentrações das cápsulas comerciais apresentaram valores discordantes dos informados em seus rótulos, principalmente quando utilizada toda a região espectral (**Figura 4.25(a)** e **Figura 4.25(c)**), com coeficientes de determinação muito baixos e RMSEPs elevados.

Apesar de não ser possível estar certo com relação às concentrações reais das cápsulas comerciais, devido à ausência de métodos oficiais, algumas interpretações podem ser realizadas. Para as previsões realizadas com os modelos construídos em toda a região espectral, uma marca apresentou valores de concentração negativos, e outros 2 lotes pertencentes a uma determinada marca apresentaram concentrações muito inferiores aos seus rótulos, porém, consistentes com seus perfis espectrais. Ainda 3 lotes, referentes a 3 diferentes marcas, apresentaram teor de óleo de alho muito superior ao valor esperado. Esses lotes apresentam uma banda pronunciada na região entre 1000 cm^{-1} e 945 cm^{-1} , com intensidade elevada em relação aos demais lotes e às misturas de calibração. Talvez isso ocorra devido ao processo de oxidação das amostras, conforme discutido anteriormente (**seção 4.2**), à diferenças na composição do óleo vegetal utilizado, ou ao próprio processo de produção/embalagem do óleo de alho. FARAG et al. (2017) detectaram por CG-MS, em uma marca de cápsula de alho em pó, a presença de dietil ftalato, composto tóxico comumente encontrado em medicamentos de venda livre na China, provenientes de embalagens plásticas ou de revestimentos gastrorresistentes. Logo, o perfil diferenciado dessas cápsulas na região entre 1000 cm^{-1} e 945 cm^{-1} , pode ser devido à adição de algum composto que não está presente nos demais lotes de cápsulas e nas misturas de calibração preparadas, e que absorve nessa região do infravermelho.

Para a região espectral de 940 cm^{-1} a 900 cm^{-1} , onde os perfis espectrais de todas as cápsulas e misturas de calibração são semelhantes, houve melhora significativa nos resultados das previsões, avaliada por meio de testes F a 95% de confiança e com 136 graus de liberdade. Essa melhora é evidenciada pelos valores de RMSEP, coeficiente de correlação e de determinação observados na **Tabela 4.6**, principalmente para os dados pré-processados por MSC. Entretanto,

as previsões não apresentaram exatidão satisfatória, tomando como referência os valores declarados nos rótulos das cápsulas.

Acredita-se que algumas amostras devem estar, de fato, em desacordo com as informações contidas nos rótulos, tomadas como referência. Todavia, a fim de se confirmar esta possibilidade, outras técnicas de calibração, apresentadas a seguir, também foram empregadas, utilizando seleção de variáveis espectrais (SPA-MLR) e resolução de curvas multivariadas (MCR), para comparação de seus resultados. Por fim, a exatidão da metodologia por MCR-ALS foi avaliada por ensaios de recuperação.

4.4.2. Modelos SPA-MLR

Assim como em PLS, apenas os resultados referentes aos pré-processamentos 1ª derivada e MSC serão apresentados. Quando empregada a 1ª derivada, para as duas regiões investigadas, apenas duas variáveis foram selecionadas pelo SPA-MLR (**Figura 4.26**), em função do RMSECV.

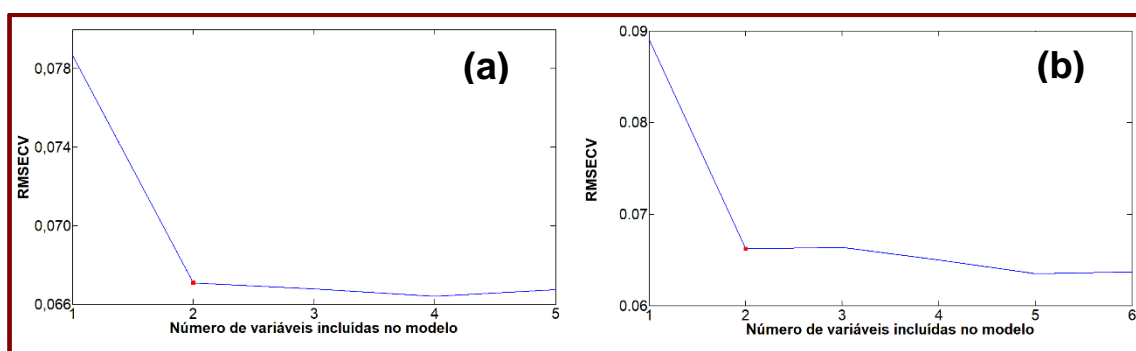


Figura 4.26. Número de variáveis selecionadas pelo SPA-MLR com os dados pré-processados por 1ª derivada (polinômio de 2ª ordem com janela de 11 pontos) (a) utilizando toda a região espectral e (b) a região de 940 cm^{-1} a 900 cm^{-1} .

Em ambos os casos, as variáveis foram selecionadas (**Figura 4.27** e **Figura 4.28**) na banda de maior intensidade no espectro do óleo de alho (**Figura 4.1**), atribuída à vibrações de estiramento e deformação de ligações, contidas nos componentes majoritários do óleo de alho.

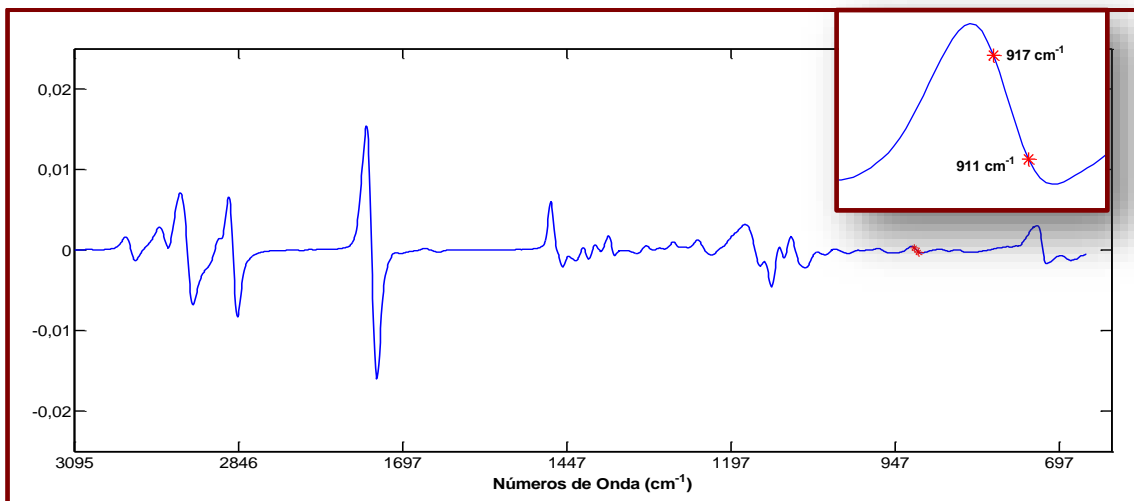


Figura 4.27. Variáveis selecionadas pelo SPA-MLR, utilizando toda a região espectral, com os dados pré-processados por 1ª derivada (polinômio de 2ª ordem com janela de 11 pontos).

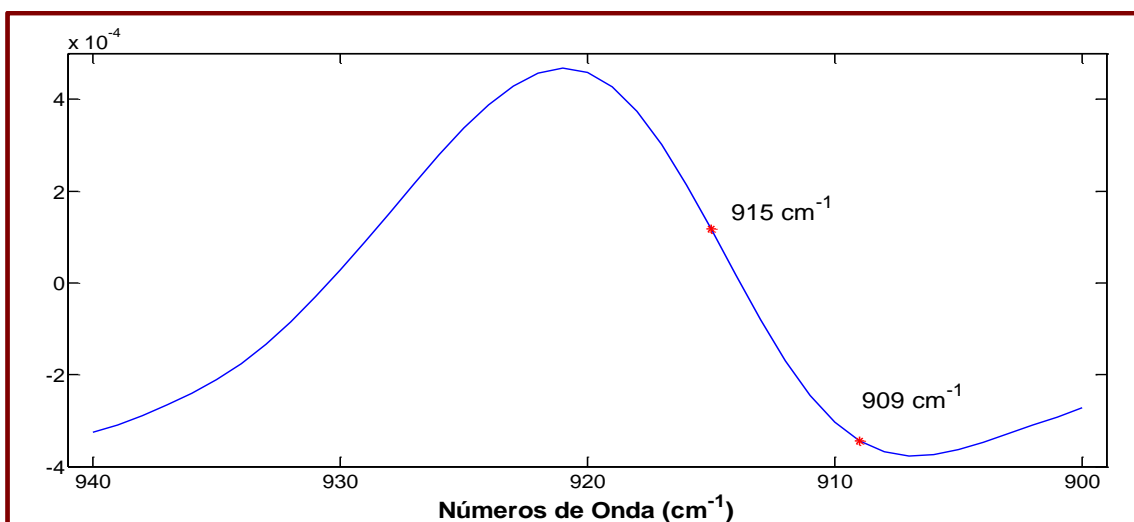


Figura 4.28. Variáveis selecionadas pelo SPA-MLR, utilizando a região espectral de 940 cm^{-1} a 900 cm^{-1} , com os dados pré-processados por 1ª derivada (polinômio de 2ª ordem com janela de 11 pontos).

Na **Figura 4.29** temos os gráficos das concentrações de referência versus as concentrações previstas na construção dos modelos de calibração usando validação cruzada completa, para os modelos construído com as 123 misturas de calibração, em toda a região espectral (**Figura 4.29(a)**) e para a região de 940 cm^{-1} a 900 cm^{-1} (**Figura 4.29(b)**).

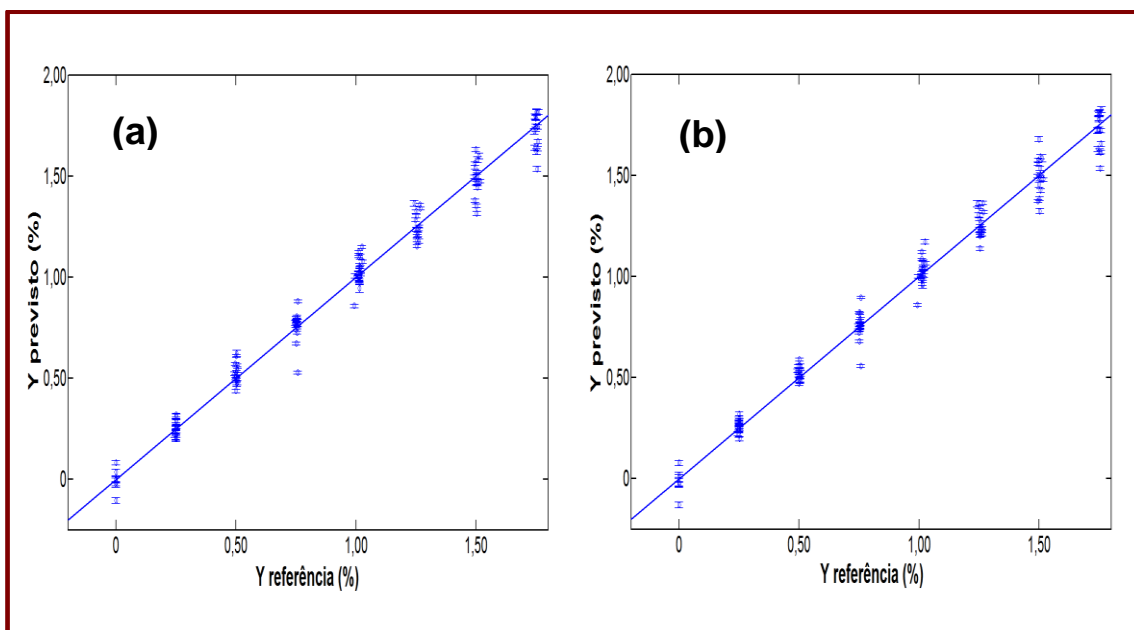


Figura 4.29. Gráfico das concentrações de referência versus as concentrações previstas na construção do modelo de calibração SPA-MLR, usando validação cruzada completa, para os dados pré-processados por 1ª derivada em (a) toda a região espectral e (b) na região de 940 cm^{-1} a 900 cm^{-1} .

Ambos os modelos obtiveram 0,07 de RMSECV, com coeficientes de correlação iguais a 0,9920 e 0,9922, e coeficientes de determinação iguais a 0,9841 e 0,9845, para os modelos construídos em toda a região espectral e na região de 940 cm^{-1} a 900 cm^{-1} , respectivamente.

Para os dados pré-processados por MSC, apesar do SPA-MLR ter selecionado 5 variáveis (**Figura 4.30** e **Figura 4.31**), resultados semelhantes aos obtidos com 1ª derivada foram obtidos. Ambos os modelos obtiveram 0,07 de RMSECV, com coeficientes de correlação iguais a 0,9929 e 0,9921, e coeficientes de determinação iguais a 0,9858 e 0,9842, para os modelos construídos em toda a região espectral e na região de 940 cm^{-1} a 900 cm^{-1} , respectivamente.

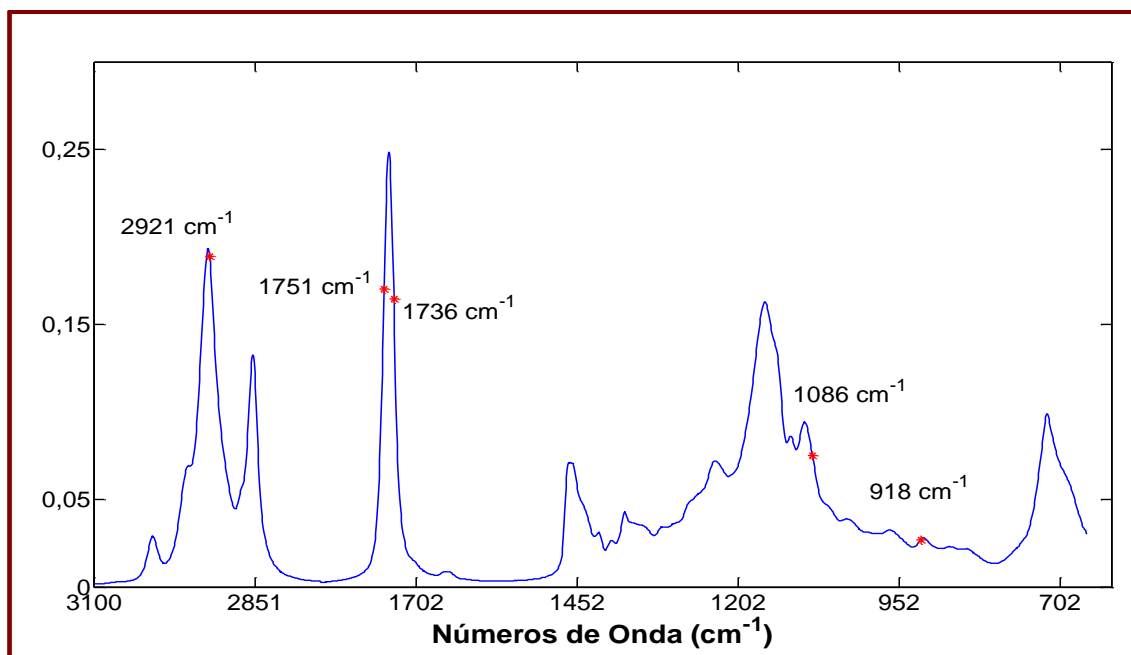


Figura 4.30. Variáveis selecionadas pelo SPA-MLR, utilizando toda a região espectral, com os dados pré-processados por MSC.

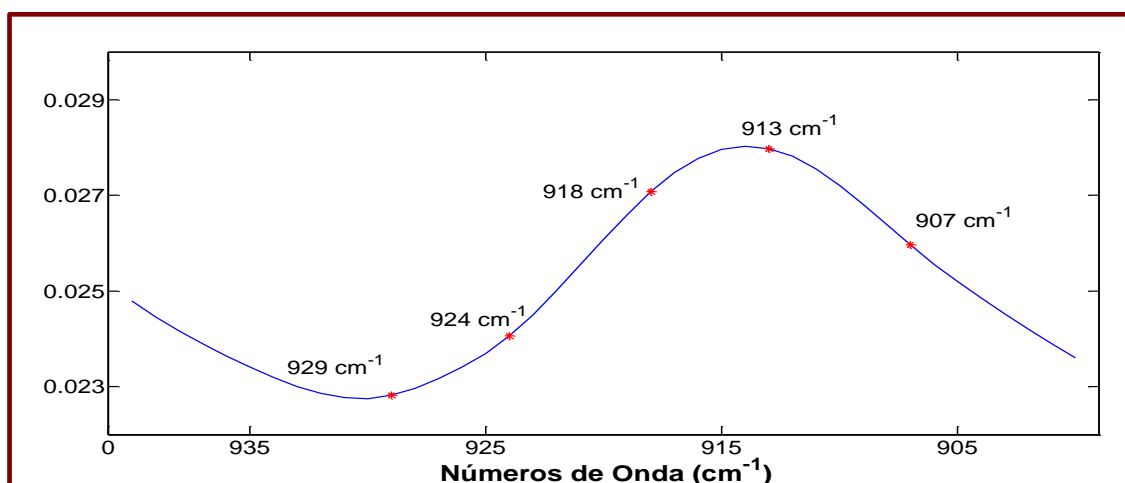


Figura 4.31. Variáveis selecionadas pelo SPA-MLR, utilizando a região espectral de 940 cm^{-1} a 900 cm^{-1} , com os dados pré-processados por MSC.

Os gráficos das concentrações previstas pelo SPA-MLR dos dados pré-processados por MSC versus as concentrações de referência, para ambas as regiões espectrais, na etapa de validação, encontram-se na **Figura 4.32**. Assim como em PLS, os modelos construídos por SPA-MLR em ambas as regiões espectrais, estão bem ajustados aos dados e não apresentam *outliers*.

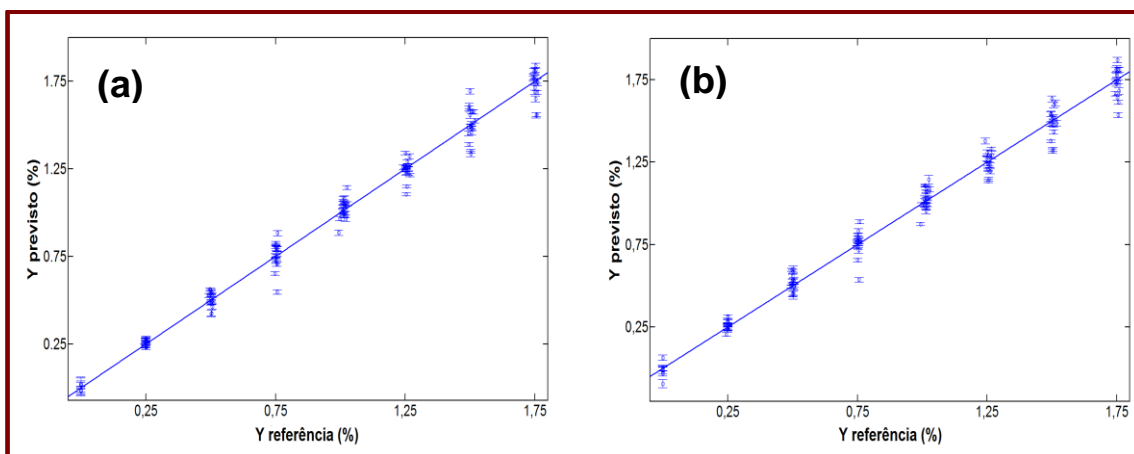


Figura 4.32. Gráfico das concentrações de referência versus as concentrações previstas na construção do modelo de calibração SPA-MLR, para os dados pré-processados por MSC em (a) toda a região espectral e (b) na região de 940 cm^{-1} a 900 cm^{-1} .

Ambos os pré-processamentos e regiões espectrais apresentaram resultados semelhantes e, na **Figura 4.33**, verifica-se que todos os modelos apresentam resíduos em Y distribuídos aleatoriamente em torno de zero.

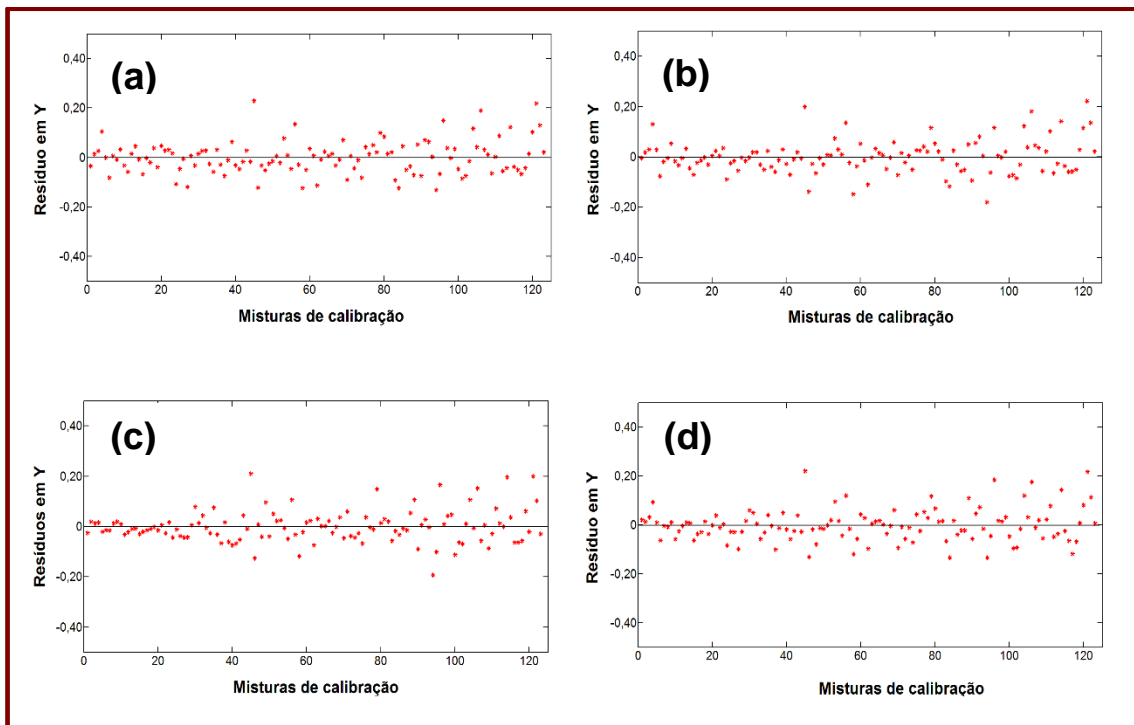


Figura 4.33. Gráficos dos resíduos das misturas de calibração, para os dados pré-processados por 1ª derivada Savitzky Golay com polinômio de 2ª ordem e janela de 11 pontos (a) e (b); e por MSC (c) e (d), em toda a região espectral (a) e (c) e na região 940 cm^{-1} a 900 cm^{-1} (b) e (d).

As previsões das concentrações das 136 cápsulas comerciais de óleo de alho foram realizadas empregando os 4 modelos construídos. Os gráficos das concentrações de referência versus as concentrações previstas encontram-se nas **Figura 4.34** e **Figura 4.35**.

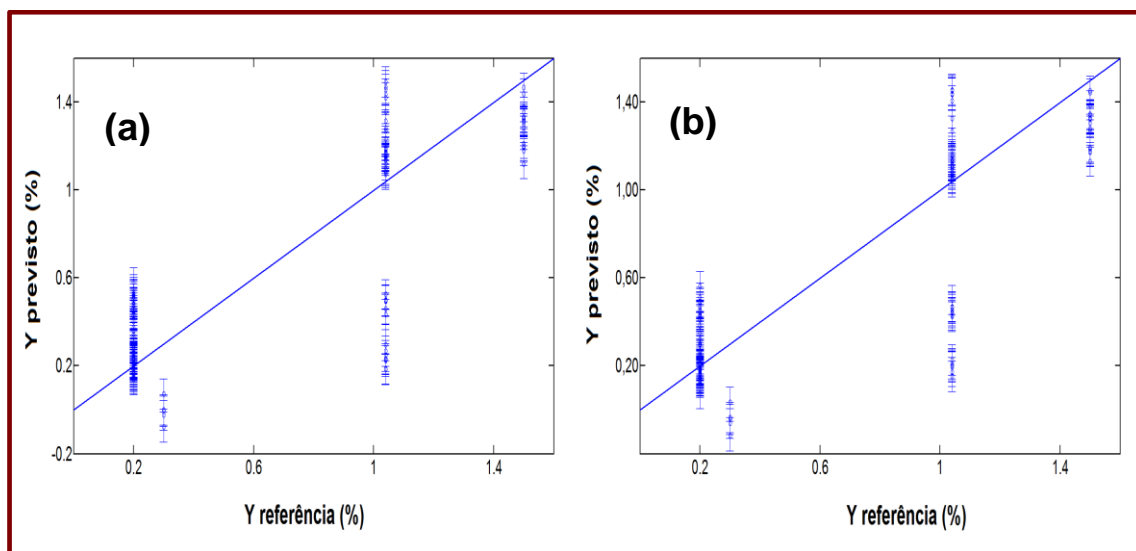


Figura 4.34. Gráfico das concentrações de referência versus as concentrações previstas das 136 amostras de cápsulas comerciais de óleo de alho obtidas pelo modelo SPA-MLR, para os dados pré-processados por 1ª derivada em (a) toda a região espectral e (b) na região de 940 cm^{-1} a 900 cm^{-1} .

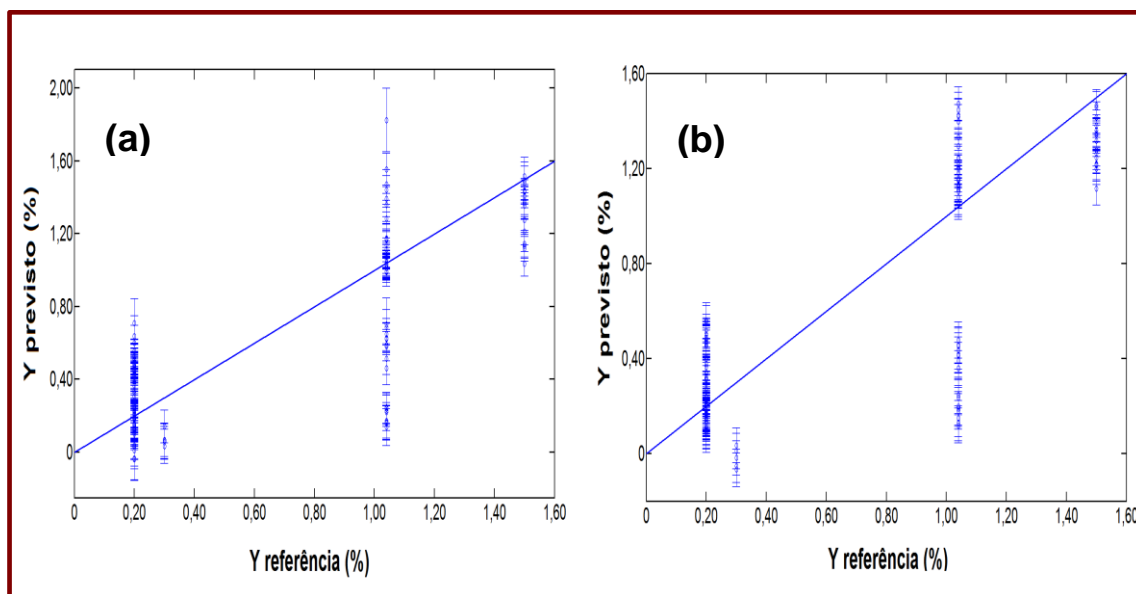


Figura 4.35. Gráfico das concentrações de referência versus as concentrações previstas das 136 amostras de cápsulas comerciais de óleo de alho, obtidas para os modelos SPA-MLR, para os dados pré-processados por MSC em (a) toda a região espectral e (b) na região de 940 cm^{-1} a 900 cm^{-1} .

Na **Tabela 4.7** encontram-se os resultados obtidos para as previsões realizadas e na **Figura 4.36**, os gráficos dos resíduos obtidos para as previsões das 136 amostras comerciais de cápsulas de óleo de alho.

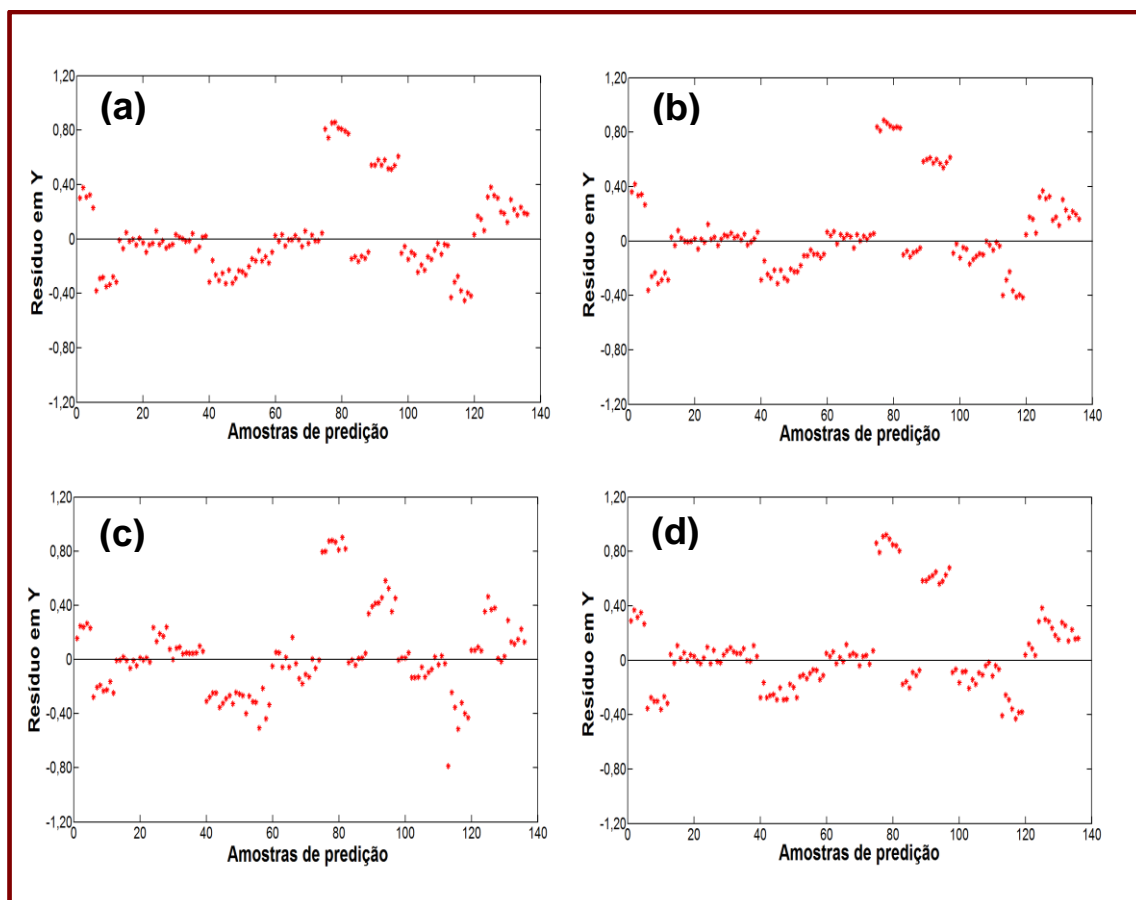


Figura 4.36. Gráficos dos resíduos das 136 cápsulas de óleo de alho, para os dados pré-processados por 1ª derivada Savitzky Golay com polinômio de 2ª ordem e janela de 11 pontos (a) e (b); e por MSC (c) e (d), em toda a região espectral (a) e (c) e na região de 940 cm^{-1} a 900 cm^{-1} (b) e (d).

Tabela 4.7. Resultados da previsão das amostras comerciais de cápsulas de óleo de alho, obtidas pelos modelos SPA-MLR, para os dados pré-processados por 1ª derivada e por MSC, utilizando toda a região espectral e a região de 940 cm^{-1} a 900 cm^{-1} , na faixa de 0,00% a 1,75% m/m.

		Previsão		
		r	R ²	RMSEP
Toda a região espectral	1ª Derivada(2)	0,8061	0,6498	0,30
	MSC(5)	0,8026	0,6441	0,31
Região de 940 cm^{-1} a 900 cm^{-1}	1ª Derivada(2)	0,8098	0,6558	0,30
	MSC(5)	0,8022	0,6436	0,31

Em todos os casos, os resultados obtidos foram muito semelhantes para os dados pré-processados por 1ª derivada e MSC, conforme se pode verificar na **Tabela 4.7** e nos gráficos de resíduos apresentados na **Figura 4.36**, onde apenas algumas cápsulas apresentam resíduos próximos a zero. Por meio de testes F ($n=136$ e $\alpha=5\%$), constatou-se que, realmente, os valores de RMSEP obtidos pelos modelos construídos com ambos os pré-processamentos e em ambas as regiões espectrais, não apresentam diferença estatisticamente significativa.

Em relação aos resultados obtidos com PLS (**Tabela 4.6**) para os dados pré-processados por 1ª derivada, SPA-MLR empregando apenas 2 variáveis, apresentou melhora significativa, constatada por testes F ($n=136$ e $\alpha=5\%$) realizados entre os RMSEP obtidos para ambos os modelos, nas duas regiões espectrais de análise. Para os dados pré-processados por MSC, os valores de RMSEP obtidos pelos modelos PLS e SPA-MLR, na região de 940 cm^{-1} a 900 cm^{-1} , não apresentam diferença. Entretanto, utilizando toda a região espectral, SPA-MLR apresenta valor de RMSEP melhor do que o obtido por PLS nessa mesma região. Dos 3 lotes de cápsulas que haviam apresentado valores de concentração, por PLS, muito superiores aos indicados em seus rótulos, 2 apresentaram concentrações muito próximas aos seus rótulos, quando previstas por SPA-MLR. Para o outro lote, houve redução da concentração, apesar de ainda apresentar-se elevada. Quanto às demais marcas e lotes de cápsulas, as concentrações obtidas por SPA-MLR apresentaram-se semelhantes aos obtidos por PLS, na região de 940 cm^{-1} a 900 cm^{-1} .

Pode-se concluir que, os modelos SPA-MLR construídos com os dados pré-processados por 1ª derivada e por MSC, em ambas as regiões espectrais de análise (*full spectrum* e de 940 cm^{-1} a 900 cm^{-1}), não apresentam diferença entre si, e as concentrações encontradas para algumas marcas de cápsulas de óleo de alho, provavelmente discordam das concentrações indicadas em seus rótulos.

4.4.3. Modelos MCR-ALS

As técnicas de calibração multivariada utilizadas até aqui, apresentaram valores de concentração das cápsulas comerciais discordantes de seus rótulos, apesar da validação dos modelos de calibração apresentarem valores satisfatórios de parâmetros de desempenho como RMSECV, coeficiente de correlação e de determinação. Visto que as misturas de calibração preparadas a partir do padrão de óleo de alho e diferentes marcas de óleo de soja, podem não representar toda a complexidade das cápsulas comerciais, optou-se por utilizar MCR-ALS devido à sua característica de não precisar incluir na calibração, interferentes que porventura existam nas amostras desconhecidas, conhecida como vantagem de segunda ordem (BOOKSH; KOWALSKI, 1994; MARÇO et al., 2014; PARASTAR; SHAYE, 2015).

Para construção do modelo MCR-ALS, os dados foram organizados na forma de matriz aumentada, visto que essa configuração permite uma melhora significativa dos resultados fornecidos por MCR-ALS, podendo reduzir a ocorrência de ambiguidades rotacionais (JAUMOT et al., 2013; DE JUAN; JAUMOT; TAULER, 2014). Como os dados foram analisados pela mesma técnica instrumental (FT-MIR), a configuração adotada foi a de matriz aumentada na direção da coluna (*column wise*), conforme **Figura 4.37**.

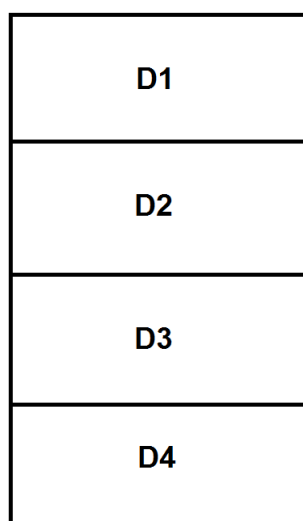


Figura 4.37. Configuração de matriz aumentada utilizada na análise MCR-ALS.

Na **Figura 4.37**, **D1** corresponde à matriz espectral dos óleos de soja, **D2** à matriz das misturas de calibração e **D3** à matriz das cápsulas comerciais de óleo de alho.

A matriz **D4** corresponde à matriz de dados para os ensaios de recuperação, empregada com a finalidade de determinar a exatidão do método. Foram utilizados ensaios com níveis de concentração de 0,50%, 1,10% e 1,70% do padrão de óleo de alho. A matriz **D4** é composta pelos espectros das cápsulas comerciais fortificadas com determinadas quantidades do padrão de óleo alho, subtraídos pelos seus respectivos espectros, antes do acréscimo do padrão (**Tabela 3.2**). Após a subtração, apenas a região de 940 cm^{-1} a 900 cm^{-1} apresenta sinal, as demais regiões apresentam apenas ruídos, visto praticamente não haver alteração nessas regiões com o acréscimo do padrão de óleo de alho. Logo, o modelo MCR-ALS foi construído apenas nessa região.

Na execução do MCR-ALS, como o número de componentes nas misturas de calibração é conhecido a priori, não foi necessária a utilização das técnicas PCA ou SVD, e foi fixado como sendo igual a 2. As estimativas iniciais dos espectros foram constituídas pelos espectros puros dos dois componentes das misturas de calibração: óleo de alho e óleo de soja. No caso de óleo de soja, o espectro fornecido como estimativa inicial foi uma média dos espectros das 4 marcas de soja utilizadas. As restrições utilizadas para auxiliar na resolução, para os dados pré-processados por 1ª derivada, foram não negatividade nas concentrações, seletividade (indicando, na matriz **isp**, quais componentes estão presentes em cada matriz) e restrição de correlação para obtenção das concentrações em unidades reais.

A matriz **Csel**, utilizada para relacionar as concentrações conhecidas das misturas de calibração aos valores obtidos pelo algoritmo ALS para essas amostras, é composta pelas concentrações conhecidas do óleo de alho nas misturas de calibração (C1 e C2) e as linhas referentes às amostras de previsão (C3 e C4) são preenchidas pela expressão NaN (do inglês, *Not a Number*). Para os dados pré-processados por MSC, além da seletividade e da restrição de

correlação, a não negatividade foi aplicada tanto às concentrações, quanto aos dados espectrais. A **Figura 4.38** ilustra as etapas seguidas no MCR-ALS.

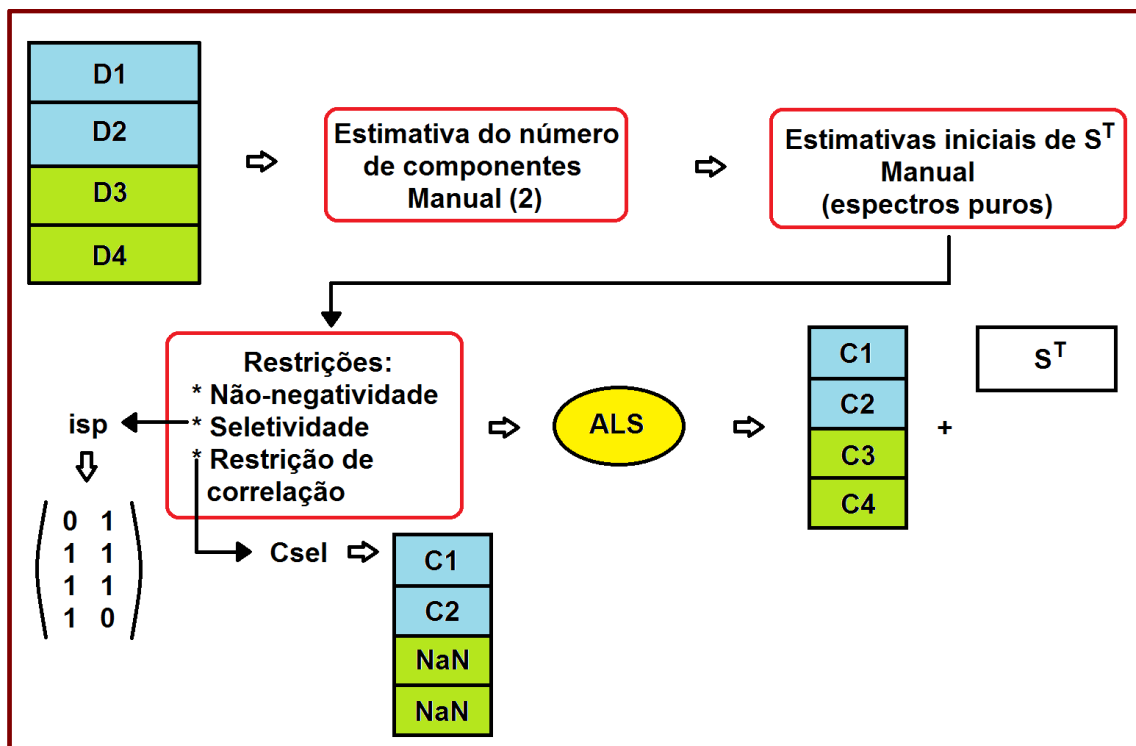


Figura 4.38. Ilustração das etapas da análise MCR-ALS.

Seguindo os passos ilustrados na **Figura 4.38**, obteve-se, para os dados pré-processados por 1ª derivada, falta de ajuste igual a 3,63% e porcentagem de variância explicada igual a 99,87%. Com a matriz S^T , realizou-se a comparação entre os espectros puros estimados pelo MCR-ALS e os espectros experimentais do padrão de óleo de alho e do óleo de soja, adquiridos por FTIR (com acessório ATR), obtendo-se uma correlação de 0,8914 entre os espectros de óleo de alho, e correlação de 0,9997 entre os espectros do óleo de soja. Considerando-se que o teor de óleo de alho é cerca de 100 vezes inferior à concentração de óleo de soja, a correlação encontrada entre os espectros de óleo de alho pode ser considerada satisfatória. Os espectros recuperados e os espectros experimentais para o padrão de óleo de alho e para o óleo de soja, encontram-se na **Figura 4.39**.

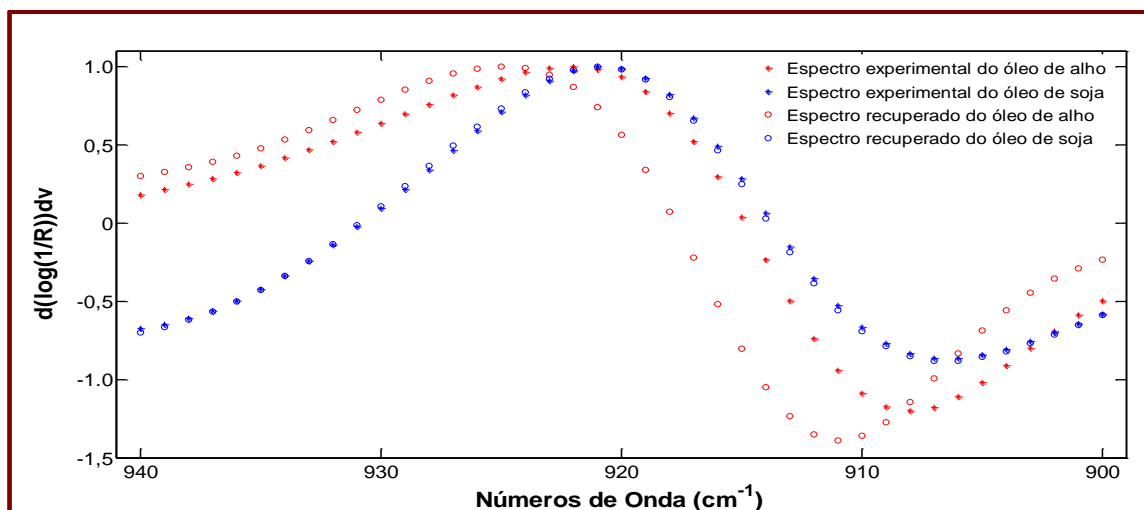


Figura 4.39. Espectros recuperados e experimentais do padrão de óleo de alho e do óleo de soja.

A **Figura 4.40** ilustra o resultado da regressão linear feita entre as concentrações estimadas pelo MCR-ALS e as concentrações de referência para as misturas de calibração.

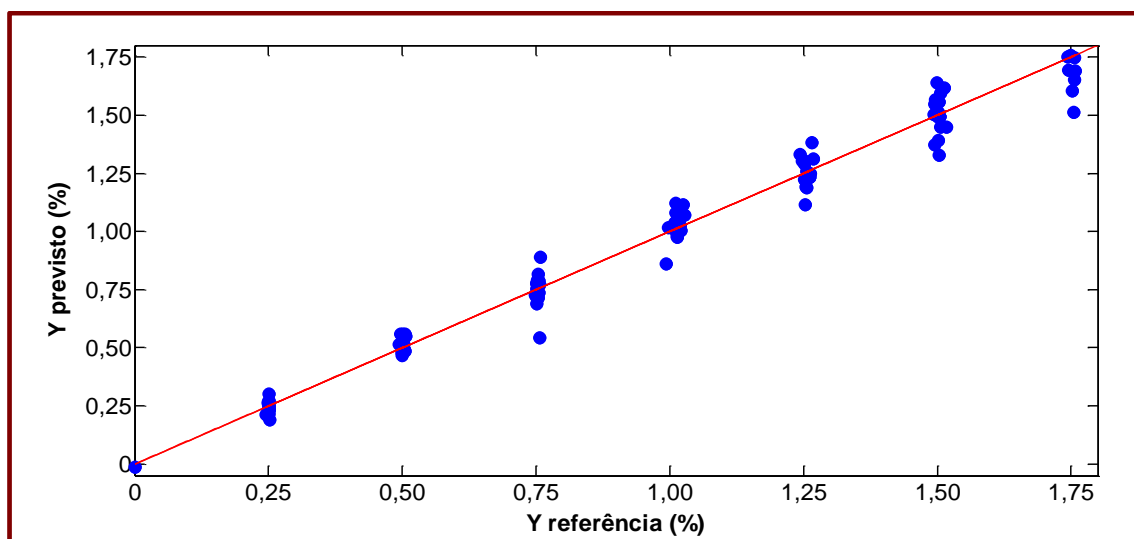


Figura 4.40. Gráfico da regressão linear entre as concentrações previstas pelo MCR-ALS e as concentrações de referência para as misturas de calibração.

A regressão linear entre os perfis de intensidades das misturas de calibração (gerado pelo MCR-ALS) e os valores de concentrações conhecidos, resultou em bons valores de R^2 e de r , iguais a 0,9866 e 0,9933, respectivamente

e RMSEC igual a 0,06. Os limites de detecção e de quantificação calculados foram 0,09% e 0,27%, respectivamente. O RPD apresentou valor igual a 8,89, revelando exatidão satisfatória para utilização do modelo em controle de qualidade, e a avaliação de erros sistemáticos (*bias*) apresentou valor de *t* calculado inferior ao *t* crítico, para 122 graus de liberdade a 95% de confiança. Logo, pode-se concluir que o modelo não apresenta erros sistemáticos.

Para as cápsulas comerciais de óleo de alho e para as recuperações, MCR-ALS obteve valores de R^2 e de *r* iguais a 0,7389 e 0,8596, respectivamente, e RMSEP igual a 0,21. A regressão linear entre as concentrações estimadas pelo MCR-ALS e as concentrações de referência para as cápsulas de óleo de alho e para as recuperações encontram-se na **Figura 4.41**. Duas marcas apresentaram concentrações negativas, sendo uma delas a mesma que apresentou valor negativo por PLS e SPA-MLR. A outra, é uma das marcas que apresenta uma banda pronunciada entre 1000 cm^{-1} e 945 cm^{-1} . Porém, na região de 940 cm^{-1} a 900 cm^{-1} , essa marca de cápsula apresenta a banda com intensidade praticamente igual ao óleo de soja puro, e provavelmente, sua concentração deve estar abaixo do limite de detecção da técnica.

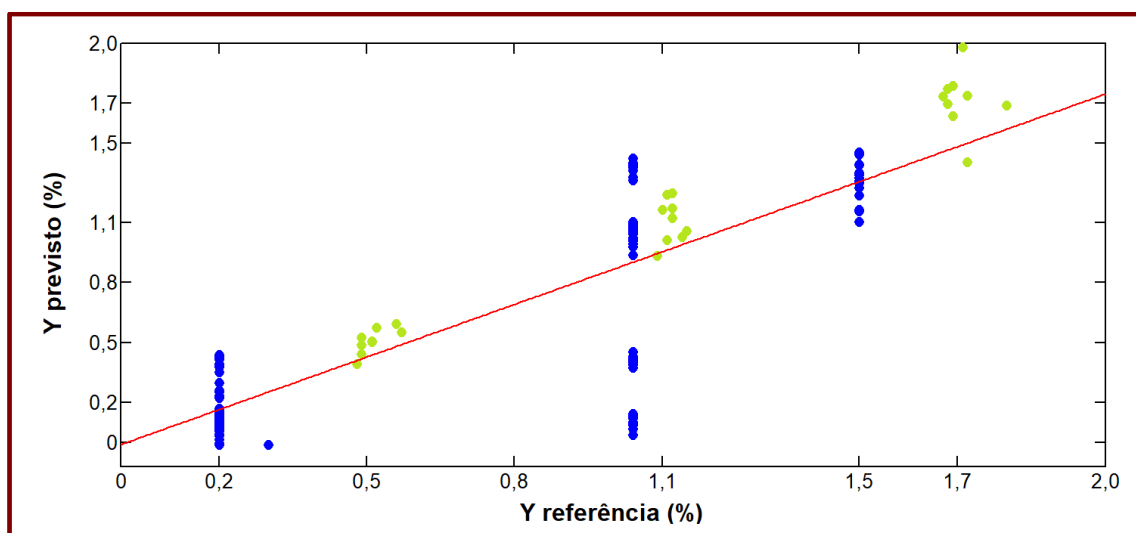


Figura 4.41. Gráfico da regressão linear entre as concentrações previstas pelo MCR-ALS e as concentrações de referência para as cápsulas comerciais de óleo de alho (em azul) e para as recuperações (em verde).

Em relação ao estudo de recuperação, para os níveis de concentração empregados nos ensaios de 0,50%, 1,10% e 1,70%, obteve-se índices de

recuperação iguais a 99%, 98% e 99%, respectivamente. Realizando-se teste t de *Student* (**Equação 2.24**) a 95% de confiança, com 26 graus de liberdade, obteve-se valor de t calculado (0,289) inferior ao t crítico (2,056), revelando, portanto, boa exatidão do método. Para os dados pré-processados por MSC, apesar de bons resultados de calibração, e resultados semelhantes aos encontrados por 1ª derivada na previsão das cápsulas comerciais, os ensaios de recuperação foram insatisfatórios. Provavelmente isso se deve ao fato que os espectros resultantes da subtração dos sinais (sinal da cápsula com padrão, subtraído pelo sinal da cápsula antes da adição do padrão de óleo de alho), quando submetido ao pré-processamento MSC, mostrar-se muito ruidoso. Portanto, seus resultados não serão abordados.

Após a avaliação de exatidão da metodologia proposta por MCR-ALS, confirma-se que várias marcas apresentaram concentrações em desacordo com o declarado em seus rótulos, e que entre lotes de mesma marca há diferenças significativas de concentração.

CAPÍTULO 5.

CONCLUSÃO



Capítulo 5. CONCLUSÃO

Nesta tese, foram desenvolvidos métodos para classificar cápsulas gelatinosas de óleo de alho quanto ao seu teor e determinar suas concentrações. A análise exploratória por PCA, revelou a importância da região espectral de 940 cm^{-1} a 900 cm^{-1} para o agrupamento das misturas e amostras de cápsulas quanto ao teor de óleo de alho.

Os modelos de classificação e de calibração construídos utilizando diferentes técnicas quimiométricas, no geral, apresentaram os melhores resultados quando utilizada apenas a região espectral indicada por PCA (940 cm^{-1} a 900 cm^{-1}), onde a banda existente é referente à vibrações de estiramento da ligação C-S-C e deformação de $=\text{CH}_2$, presentes nos componentes majoritários do óleo de alho, independente do pré-processamento utilizado.

O método proposto de classificação por SPA-LDA apresentou valores satisfatórios de sensibilidade, especificidade e exatidão, demonstrando que pode ser utilizado como um método “*screening*”. As informações obtidas por esse método podem ser confirmadas posteriormente pelo método também proposto de determinação quantitativa do teor de óleo de alho em cápsulas gelatinosas, empregando MCR-ALS, cuja exatidão, obtida por ensaios de recuperação, resultou em valores de 98% a 99%.

Os métodos aqui propostos destacam-se por serem simples, rápidos, não destrutivos, sem etapas de preparo de amostras e sem geração de resíduos, além de serem de fácil implementação.

Através dos resultados, foi possível constatar que várias marcas, e, em alguns casos, diferentes lotes de uma mesma marca, apresentam concentrações em discordância com os rótulos. Justifica-se, assim a controvérsia existente a respeito da funcionalidade do óleo de alho, indicando a necessidade de controle dos órgãos de fiscalização junto aos fabricantes, para garantir a qualidade desses produtos ao consumidor.

5.2. Propostas futuras

As propostas futuras deste trabalho são:

- ✓ Aplicar o algoritmo de classificação empregando MCR-ALS, que está sendo desenvolvido por alunos do LAQA (Laboratório de Automação em Química Analítica – UFPB), para classificar as amostras de cápsulas quanto ao teor de óleo de alho;
- ✓ Avaliar a possibilidade de concatenar as regiões do infravermelho médio e próximo para o desenvolvimento de método de quantificação do teor de óleo de alho em cápsulas gelatinosas;
- ✓ Desenvolver método para quantificação de outros produtos derivados do alho, como comprimidos e extrato de alho envelhecido, em função de seus compostos bioativos, empregando espectroscopia no infravermelho médio e métodos de calibração multivariada.

REFERÊNCIAS BIBLIOGRÁFICAS

ABU-LAFI, S.; DEMBICKI, J. W.; GOLDSHLAG, P.; HANUŠ, L. O.; DEMBITSKY, V. M. The use of the “Cryogenic” GC/MS and on-column injection for study of organosulfur compounds of the *Allium sativum*. **Journal of Food Composition and Analysis**, v. 17, p. 235–245, 2004.

AGARWAL, K. C. Therapeutic actions of garlic constituents. **Medicinal Research Reviews**, v. 16, n. 1, p. 111–124, 1996.

ALLICIN FACTS, Garlic supplements. Disponível em <http://www.allicinfacts.com/garlic_supplement/> Acesso em 22 de fevereiro de 2017.

AMAGASE, H., PETESCH, B. L., MATSUURA, H., KASUGA, S., & ITAKURA, Y. Recent Advances on the Nutritional Effects Associated with the Use of Garlic as a Supplement: Intake of garlic and its bioactive components. **American Society for Nutritional Sciences**, p. 955S–962S, 2001.

ANJOS, O.; CAMPOS, M. G.; RUIZ, P. C.; ANTUNES, P. Application of FTIR-ATR spectroscopy to the quantification of sugar in honey. **Food Chemistry**, v. 169, p. 218–223, 2015.

ANTUNES, M. C.; SIMÃO, J. E. J.; DUARTE, A. C.; TAULER, R. Multivariate curve resolution of overlapping voltammetric peaks: quantitative analysis of binary and quaternary metal mixtures. **The Analyst**, v. 127, n. 6, p. 809–817, 2002.

ARAÚJO, M. C. U.; SALDANHA, T. C. B.; GALVÃO, R. K. H.; YONEYAMA, T.; CHAME, H. C.; VISANI, V. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. **Chemometrics and Intelligent Laboratory Systems**, v. 57, p. 65–73, 2001.

GOMES, A. A.; GALVÃO, R. K. H.; ARAÚJO, M. C. U.; VÉRAS, G.; SILVA, E. C. The successive projections algorithm for interval selection in PLS. **Microchemical Journal**, v. 110, p. 202–208, 2013.

ASTM E1655-05. Standard Practices for Infrared Multivariate Quantitative Analysis. **ASTM International**, West Conshohocken, Pennsylvania, USA, 2012.

AVIELLO, G.; ABENAVOLI, L.; BORRELLI, F.; CAPASSO, R.; IZZO, A. A.; LEMBO, F.; ROMANO, B.; CAPASSO, F. Garlic: Empiricism or science? **Natural Product Communications**, v. 4, n. 12, p. 1785–1796, 2009.

AZCARATE, S. M.; GOMES, A. A.; ALCARAZ, M. R.; UGULINO, M. C. A.; CAMIÑA, J. M.; GOICOECHEA, H. C. Modeling excitation-emission fluorescence matrices with pattern recognition algorithms for classification of Argentine white wines according grape variety. **Food Chemistry**, v. 184, p. 214–219, 2015.

AZZOUZ, T.; TAULER, R. Application of multivariate curve resolution alternating least squares (MCR-ALS) to the quantitative analysis of pharmaceutical and agricultural samples. **Talanta**, v. 74, n. 5, p. 1201–1210, 2008.

BALLABIO, D.; CONSONNI, V. Classification tools in chemistry. Part 1: linear models. PLS-DA. **Analytical Methods**, v. 5, n. 16, p. 3790–3798, 2013.

BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. **Applied Spectroscopy**, v. 43, n. 5, p. 772–777, 1989.

BASSBASI, M.; PLATIKANOV, S.; TAULER, R.; OUSSAMA, A. FTIR-ATR determination of solid non fat (SNF) in raw milk using PLS and SVM chemometric methods. **Food Chemistry**, v. 146, p. 250–254, 2014.

BEEBE, K. R.; PELL, R. J.; SEASHOLTZ, M. B. **Chemometrics: A Practical Guide**. New York: John Wiley Y Sons, 1998.

BENDINI, A.; CERRETANI, L.; VIRGILIO, F. di; BELLONI, P.; BONOLI-CARBOGNIN, M.; LERCKER, G. Preliminary evaluation of the application of the ftir spectroscopy to control the geographic origin and quality of virgin olive oils. **Journal of Food Quality**, v. 30, p. 424–437, 2007.

BLOCK, E. The chemistry of garlic and onions. **Scientific American**, v. 252, n. 3, p. 114–119, 1985.

BOOKSH, K. S.; KOWALSKI, B. R. Theory of analytical chemistry. **Analytical Chemistry**, v. 66, n. 15, p. 782A–791A, 1994.

BOREK, C. Recent Advances on the Nutritional Effects Associated with the Use of Garlic as a Supplement: Antioxidant Health Effects of Aged Garlic Extract. **The Journal of Nutrition**, p. 1010S–1015S, 2001.

BOREK, C. Significance of Garlic and Its Constituents in Cancer and Cardiovascular Disease: Garlic Reduces Dementia and Heart Disease Risk. **The Journal of Nutrition**, p. 810S–812S, 2006.

BORRÀS, E.; FERRÉ, J.; BOQUÉ, R.; MESTRES, M.; ACEÑA, L.; CALVO, A.; BUSTO, O. Olive oil sensory defects classification with data fusion of instrumental techniques and multivariate analysis (PLS-DA). **Food Chemistry**, v. 203, p. 314–322, 2016.

BOTELHO, B. G.; REIS, N.; OLIVEIRA, L. S.; SENA, M. M. Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA. **Food Chemistry**, v. 181, p. 31–37, 2015.

BRASIL. AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA (ANVISA). Resolução RDC no 270, de 22 de setembro de 2005. **Regulamento técnico para óleos vegetais, gorduras vegetais e creme vegetal**. 2005.

BRASIL. Ministério da Agricultura e da Pecuária e Abastecimento (MAPA). Instrução Normativa nº 1, de 30 de janeiro de 2012. **Regulamento Técnico do Azeite de Oliva e do Óleo de Bagaço de Oliva**. 2012.

BREITKREITZ, M. C.; RAIMUNDO, JR, I. M.; ROHWEDDER, J. J. R.; PASQUINI, C.; DANTAS FILHO, H. A.; JOSÉ, G. E.; ARAÚJO, M. C. U. Determination of total sulfur in diesel fuel employing NIR spectroscopy and

multivariate calibration. **The Analyst**, v. 128, n. 9, p. 1204–1207, 2003.

BRERETON, R. G. **Chemometrics: Data Analysis for the Laboratory and Chemical Plant**. John Wiley & Sons Ltd, 2003.

BRERETON, R. G. **Chemometrics for Pattern Recognition**. Chichester, UK: John Wiley & Sons, Ltd, 2009.

BRERETON, R. G.; LLOYD, G. R. Partial least squares discriminant analysis: Taking the magic away. **Journal of Chemometrics**, v. 28, p. 213–225, 2014.

BRO, R.; SMILDE, A. K. Principal component analysis. **Analytical Methods**, v. 6, p. 2812–2831, 2014.

BUTT, M. S.; SULTAN, M. T.; BUTT, M. S.; IQBAL, J. Garlic: Nature's Protection Against Physiological Threats. **Critical Reviews in Food Science and Nutrition**, v. 49, n. 6, p. 538–551, 2009.

CALVO-GÓMEZ, O.; MORALES-LÓPEZ, J.; LÓPEZ, M. G. Solid-phase microextraction–gas chromatographic–mass spectrometric analysis of garlic oil obtained by hydrodistillation. **Journal of Chromatography A**, v. 1036, n. 1, p. 91–93, 2004.

CAPUANO, E.; VAN DER VEER, G.; BOERRIGTER-EENLING, R.; ELGERSMA, A.; RADEMAKER, J.; STERIAN, A.; VAN RUTH, S. M. Verification of fresh grass feeding, pasture grazing and organic farming by cows farm milk fatty acid profile. **Food Chemistry**, v. 164, p. 234–241, 2014.

CAVALLITO, C.; BAILEY, J. Allicin, the Antibacterial Principle of *Allium sativum*. I. Isolation, Physical Properties and Antibacterial Action. **J. Am. Chem. Soc.**, v. 66, n. 11, p. 1950–1951, 1944.

CHEN, J. Y.; ZHANG, H.; MA, J.; TUCHIYA, T.; MIAO, Y. Determination of the Degree of Degradation of Frying Rapeseed Oil Using Fourier-Transform Infrared Spectroscopy Combined with Partial Least-Squares Regression. **International Journal of Analytical Chemistry**, v. 2015, p. 1–6, 2015.

CLARKE, T. C.; BLACK, L. I.; STUSSMAN, B. J.; BARNES, P. M.; NAHIN, R. L. Trends in the use of complementary health approaches among adults: United States, 2002-2012. **National health statistics reports**, n. 79, p. 1–16, 2015.

COATES, J. Interpretation of Infrared Spectra , A Practical Approach. Ed. R.A. Meyers, 2000.

CORREIA, P. R. M.; FERREIRA, M. M. C. Reconhecimento de padrões por métodos não supervisionados: Explorando procedimentos quimiométricos para tratamento de dados analíticos. **Química Nova**, v. 30, n. 2, p. 481–487, 2007.

COSTA, G. B.; FERNANDES, D. D. S.; ALMEIDA, V. E.; MAIA, M. S.; ARAÚJO, M. C. U.; VÉRAS, G.; DINIZ, P. H. G. D. Identification of biodiesel feedstock in biodiesel/diesel blends using digital images and chemometric methods. **Anal. Methods**, v. 8, n. 24, p. 4949–4954, 2016.

CRAIG, A. P.; FRANCA, A. S.; OLIVEIRA, L. S. Evaluation of the potential of

FTIR and chemometrics for separation between defective and non-defective coffees. **Food Chemistry**, v. 132, n. 3, p. 1368–1374, 2012.

CUSTERS, D.; CAUWENBERGH, T.; BOTHY, J. L.; COURSELLE, P.; DE BEER, J. O.; APERS, S.; DECONINCK, E. ATR-FTIR spectroscopy and chemometrics: An interesting tool to discriminate and characterize counterfeit medicines. **Journal of Pharmaceutical and Biomedical Analysis**, v. 112, p. 181–189, 2015.

DENG, D. H.; XU, L.; YE, Z. H.; CUI, H. F.; CAI, C. B.; YU, X. P. FTIR spectroscopy and chemometric class modeling techniques for authentication of Chinese sesame oil. **JAOCS, Journal of the American Oil Chemists' Society**, v. 89, n. 6, p. 1003–1009, 2012.

DEVLIN, M. T.; KALASINSKY, V. F.; LEVIN, I. W. SOLID AND LIQUID PHASE RAMAN AND INFRARED. **Journal of Molecular Structure**, v. 213, p. 35–50, 1989.

DINIZ, P. H. G. D.; BARBOSA, M. F.; MILANEZ, K. D. T. de M.; PISTONESI, M. F.; ARAÚJO, M. C. U. Using UV-Vis spectroscopy for simultaneous geographical and varietal classification of tea infusions simulating a home-made tea cup. **Food Chemistry**, v. 192, p. 374–379, 2016.

DINIZ, P. H. G. D.; GOMES, A. A.; PISTONESI, M. F.; BAND, B. S. F.; ARAÚJO, M. C. U. Simultaneous Classification of Teas According to Their Varieties and Geographical Origins by Using NIR Spectroscopy and SPA-LDA. **Food Analytical Methods**, v. 7, n. 8, p. 1712–1718, 2014.

DINIZ, P. H. G. D.; PISTONESI, M. F.; ALVAREZ, M. B.; BAND, B. S. F.; ARAÚJO, M. C. U. Simplified tea classification based on a reduced chemical composition profile via successive projections algorithm linear discriminant analysis (SPA-LDA). **Journal of Food Composition and Analysis**, v. 39, p. 103–110, 2015.

DOMÍNGUEZ, M. A.; DINIZ, P. H. G. D.; DI NEZIO, M. S.; ARAÚJO, M. C. U.; CENTURIÓN, M. E. Geographical origin classification of Argentinean honeys using a digital image-based flow-batch system. **Microchemical Journal**, v. 112, p. 104–108, 2014.

DRÉAU, Y.; DUPUY, N.; ARTAUD, J.; OLLIVIER, D.; KISTER, J. Infrared study of aging of edible oils by oxidative spectroscopic index and MCR-ALS chemometric method. **Talanta**, v. 77, n. 5, p. 1748–1756, 2009a.

DRÉAU, Y.; DUPUY, N.; GAYDOU, V.; JOACHIM, J.; KISTER, J. Study of jojoba oil aging by FTIR. **Analytica Chimica Acta**, v. 642, p. 163–170, 2009b.

DRIVELOS, S. A.; HIGGINS, K.; KALIVAS, J. H.; HAROUTOUNIAN, S. A.; GEORGIU, C. A. Data fusion for food authentication. Combining rare earth elements and trace metals to discriminate “*fava Santorinis*” from other yellow split peas using chemometric tools. **Food Chemistry**, v. 165, p. 316–322, 2014.

DRUGS, Garlic. Disponível em <<https://www.drugs.com/npp/garlic.html>> Acesso em 22 de fevereiro de 2017.

DUPUY, N.; GALTIER, O.; OLLIVIER, D.; VANLOOT, P.; ARTAUD, J. Analytica Chimica Acta Comparison between NIR , MIR , concatenated NIR and MIR analysis and hierarchical PLS model . Application to virgin olive oil analysis. **Analytica Chimica Acta**, v. 666, p. 23–31, 2010.

ELZEY, B.; POLLARD, D.; FAKAYODE, S. O. Determination of adulterated neem and flaxseed oil compositions by FTIR spectroscopy and multivariate regression analysis. **Food Control**, v. 68, p. 303–309, 2016.

FARAG, M. A.; ALI, S. E.; HODAYA, R. H.; EL-SEEDI, H. R.; SULTANI, H. N.; LAUB, A.; EISSA, T. F.; ABOU-ZAID, F. O. F.; WESSJOHANN, L. A. Phytochemical Profiles and Antimicrobial Activities of *Allium cepa* Red cv. and *A. sativum* Subjected to Different Drying Methods: A Comparative MS-Based Metabolomics. **Molecules**, v. 22, n. 5, p. 1–18, 2017.

FERNANDES, D. D. .; ALMEIDA, V. E.; PINTO, L.; VÉRAS, G.; HARROP GALVÃO, R. K.; GOMES, A. A.; UGULINO ARAÚJO, M. C. The successive projections algorithm for interval selection in partial least squares discriminant analysis. **Analytical Methods**, v. 8, n. 41, 2016.

FILHO, H. A. D.; SOUZA, Ê. S. O. N.; VISANI, V.; BARROS, S. R. R. C. Simultaneous spectrometric determination of Cu^{2+} , Mn^{2+} and Zn^{2+} in polivitaminic/polimineral drug using SPA and GA algorithms for variable selection. **J. Braz. Chem. Soc.**, v. 16, n. 1, p. 58–61, 2005.

FILHO, P. A. C. Developing a rapid and sensitive method for determination of trans-fatty acids in edible oils using middle-infrared spectroscopy. **Food Chemistry**, v. 158, p. 1–7, 2014.

FISHER, R. A. The use of multiple measurements in taxonomic problems. **Annals of Eugenics**, v. 7, p. 179–188, 1936.

FREITAS, S. K. B.; NASCIMENTO, E. C. L. D.; DIONÍZIO, A. G. G.; GOMES, A. D. A.; ARAÚJO, M. C. U. D.; GALVÃO, R. K. H. A flow-batch analyzer using a low cost aquarium pump for classification of citrus juice with respect to brand. **Talanta**, v. 107, p. 45–48, 2013.

FUNDAÇÃO OSWALDO CRUZ, Pesquisa nacional de saúde. Disponível em <<http://www.pns.fiocruz.br/>>. Acesso em 12 de fevereiro de 2017.

GALVÃO, R. K. H.; ARAÚJO, M. C. U.; FRAGOSO, W. D.; SILVA, E. C.; JOSÉ, G. E.; SOARES, S. F. C.; PAIVA, H. M. A variable elimination method to improve the parsimony of MLR models using the successive projections algorithm. **Chemometrics and Intelligent Laboratory Systems**, v. 92, n. 1, p. 83–91, 2008.

GALVÃO, R. K. H.; ARAÚJO, M. C. U.; SILVA, E. C.; JOSÉ, G. E.; SOARES, S. F. C.; PAIVA, H. M. Cross-validation for the selection of spectral variables using the successive projections algorithm. **Journal of the Brazilian Chemical Society**, v. 18, n. 8, p. 1580–1584, 2007.

GALVÃO, R. K. H.; PIMENTEL, M. F.; ARAUJO, M. C. U.; YONEYAMA, T.; VISANI, V. Aspects of the successive projections algorithm for variable selection

in multivariate calibration applied to plasma emission spectrometry. **Analytica Chimica Acta**, v. 443, p. 107–115, 2001.

GAMBARRA-NETO, F. F.; MARINO, G.; ARAÚJO, M. C. U.; GALVÃO, R. K. H.; PONTES, M. J. C.; MEDEIROS, E. P. de; LIMA, R. S. Classification of edible vegetable oils using square wave voltammetry with multivariate data analysis. **Talanta**, v. 77, n. 5, p. 1660–1666, 2009.

GHASEMI-VARNAMKHAZI, M.; MOHTASEBI, S. S.; RODRIGUEZ-MENDEZ, M. L.; GOMES, A. A.; ARAÚJO, M. C. U.; GALVÃO, R. K. H. Screening analysis of beer ageing using near infrared spectroscopy and the Successive Projections Algorithm for variable selection. **Talanta**, v. 89, p. 286–291, 2012.

GOICOECHEA, H. C.; OLIVIERI, A. C.; TAULER, R. Application of the correlation constrained multivariate curve resolution alternating least-squares method for analyte quantitation in the presence of unexpected interferences using first-order instrumental data. **Analyst**, v. 135, n. 3, p. 636–642, 2010.

GOK, S.; SEVERCAN, M.; GOORMAGHTIGH, E.; KANDEMIR, I.; SEVERCAN, F. Differentiation of Anatolian honey samples from different botanical origins by ATR-FTIR spectroscopy using multivariate analysis. **Food Chemistry**, v. 170, p. 234–240, 2015.

GOMES, A. D. A. **Algoritmo das Projeções Sucessivas aplicado à seleção de variáveis em regressão PLS**. Dissertação de Mestrado. UNIVERSIDADE FEDERAL DA PARAÍBA, João Pessoa, p. 120, 2012.

GONÇALVES, M. I. S.; VILAR, W. T. S.; MEDEIROS, E. P.; PONTES, M. J. C. An analytical method for determination of quality parameters in cotton plumes by digital image and chemometrics. **Computers and Electronics in Agriculture**, v. 123, p. 89–94, 2016.

GUERCIO, V.; GALEONE, C.; TURATI, F.; LA VECCHIA, C. Gastric cancer and allium vegetable intake: a critical review of the experimental and epidemiologic evidence. **Nutrition and cancer**, v. 66, n. 5, p. 757–773, 2014.

GUILLEN, M. D.; CABO, N. Infrared Spectroscopy in the Study of Edible Oils and Fats. **Journal of the Science of Food and Agriculture**, v. 75, p. 1–11, 1997.

HASSAN, H. T. Ajoene (natural garlic compound): A new anti-leukaemia agent for AML therapy. **Leukemia Research**, v. 28, n. 7, p. 667–671, 2004.

HELOU, L.; HARRIS, I. M. **Herbal Products**. 2^a ed. Totowa, NJ: Humana Press, 2007.

HERNÁNDEZ-MARTÍNEZ, M.; GALLARDO-VELÁZQUEZ, T.; OSORIO-REVILLA, G.; ALMARAZ-ABARCA, N.; PONCE-MENDOZA, A.; VÁSQUEZ-MURRIETA, M. S. Prediction of total fat, fatty acid composition and nutritional parameters in fish fillets using MID-FTIR spectroscopy and chemometrics. **LWT - Food Science and Technology**, v. 52, n. 1, p. 12–20, 2013.

HOLLER, J. F.; SKOOG, D. A.; CROUCH, S. R. **Princípios de Análise Instrumental**. 6^a Edição ed. Bookman, 2009.

HONORATO, F. A.; GALVÃO, R. K. H.; PIMENTEL, M. F.; DE BARROS NETO, B.; UGULINO ARAÚJO, M. C.; RODRIGUES DE CARVALHO, F. Robust modeling for multivariate calibration transfer by the successive projections algorithm. **Chemometrics and Intelligent Laboratory Systems**, v. 76, n. 1, p. 65–72, 2005.

ILIC, D.; NIKOLIC, V.; NIKOLIC, L.; STANKOVIC, M.; STANOJEVIC, L.; CAKIC, M. Allicin and related compounds: Biosynthesis, synthesis and pharmacological activity. **Facta universitatis - series: Physics, Chemistry and Technology**, v. 9, n. 1, p. 9–20, 2011.

INÁCIO, M. R. C.; DE MOURA, M. D. F. V.; DE LIMA, K. M. G. Classification and determination of total protein in milk powder using near infrared reflectance spectrometry and the successive projections algorithm for variable selection. **Vibrational Spectroscopy**, v. 57, n. 2, p. 342–345, 2011.

INSAUSTI, M.; ROMANO, C.; PISTONESI, M. F.; BAND, B. S. F. Simultaneous determination of quality parameters in biodiesel/diesel blends using synchronous fluorescence and multivariate analysis. **Microchemical Journal**, v. 108, p. 32–37, 2013.

JAISWAL, P.; JHA, S. N.; BORAH, A.; GAUTAM, A.; GREWAL, M. K.; JINDAL, G. Detection and quantification of soymilk in cow-buffalo milk using Attenuated Total Reflectance Fourier Transform Infrared spectroscopy (ATR-FTIR). **Food Chemistry**, v. 168, p. 41–47, 2015.

JAUMOT, J.; DE JUAN, A.; TAULER, R. MCR-ALS GUI 2.0: New features and applications. **Chemometrics and Intelligent Laboratory Systems**, v. 140, p. 1–12, 2015.

JAUMOT, J.; IGNE, B.; ANDERSON, C. A.; DRENNEN, J. K.; DE JUAN, A. Blending process modeling and control by multivariate curve resolution. **Talanta**, v. 117, p. 492–504, 2013.

JAUMOT, J.; TAULER, R. MCR-BANDS: A user friendly MATLAB program for the evaluation of rotation ambiguities in Multivariate Curve Resolution. **Chemometrics and Intelligent Laboratory Systems**, v. 103, n. 2, p. 96–107, 2010.

JAYARAMAN, A.; MAS, S.; TAULER, R.; DE JUAN, A. Study of the photodegradation of 2-bromophenol under UV and sunlight by spectroscopic, chromatographic and chemometric techniques. **Journal of Chromatography B**, v. 910, p. 138–148, 2012.

JIANG, X.; LI, S.; XIANG, G.; LI, Q.; FAN, L.; HE, L.; GU, K. Determination of the acid values of edible oils via FTIR spectroscopy based on the O-H stretching band. **Food Chemistry**, v. 212, p. 585–589, 2016.

JUAN, A.; JAUMOT, J.; TAULER, R. Multivariate Curve Resolution (MCR). Solving the mixture analysis problem. **Analytical Methods**, v. 6, n. 14, p. 4964–4976, 2014.

JUAN, A.; TAULER, R. Multivariate Curve Resolution (MCR) from 2000: Progress

in Concepts and Applications. **Critical Reviews in Analytical Chemistry**, v. 36, n. 3–4, p. 163–176, 2006.

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.

KHOSHTINAT, K.; BARZEGAR, M.; SAHARI, M. A.; HAMIDI, Z. Encapsulation of Iranian garlic oil with β -cyclodextrin: Optimization and its characterization. **Journal of Agricultural Science and Technology**, v. 19, n. 1, p. 97–111, 2017.

KIM, S. M.; WU, C. M.; KUBOTA, K.; KOBAYASHIL, A.; KOBAYASHI, A. Effect of Soybean Oil on Garlic Volatile Compounds Isolated by Distillation. **Journal of Agricultural and Food Chemistry**, v. 43, n. 2, p. 449–452, 1995.

KIMBARIS, A. C.; SIATIS, N. G.; DAFERERA, D. J.; TARANTILIS, P. A.; PAPPAS, C. S.; POLISSIOU, M. G. Comparison of distillation and ultrasound-assisted extraction methods for the isolation of sensitive aroma compounds from garlic (*Allium sativum*). **Ultrasonics Sonochemistry**, v. 13, n. 1, p. 54–60, 2006.

KOCH, C.; POSCH, A. E.; GOICOECHEA, H. C.; HERWIG, C.; LENDL, B. Multi-analyte quantification in bioprocesses by Fourier-transform-infrared spectroscopy by partial least squares regression and multivariate curve resolution. **Analytica Chimica Acta**, v. 807, p. 103–110, 2014.

KREST, I.; KEUSGEN, M. Quality of herbal remedies from *Allium sativum*: Differences between alliinase from garlic powder and fresh garlic. **Planta Medica**, v. 65, n. 2, p. 139–143, 1999.

LANZOTTI, V. The analysis of onion and garlic. **Journal of Chromatography A**, v. 1112, n. 1–2, p. 3–22, 2006.

LAWSON, L. D. Bioactive Organosulfur Compounds of Garlic and Garlic Products. In: *Human Medicinal Agents from Plants*, Washington, DC. **Anais...** Washington, DC: American Chemical Society, 1993.

LAWSON, L.; WANG, Z.-Y.; HUGHES, B. Identification and HPLC Quantitation of the Sulfides and Dialk(en)yl Thiosulfinates in Commercial Garlic Products. **Planta Medica**, v. 57, n. 4, p. 363–370, 1991.

LEE, S.-N.; KIM, N.-S.; LEE, D.-S. Comparative study of extraction techniques for determination of garlic flavor components by gas chromatography-mass spectrometry. **Analytical and bioanalytical chemistry**, v. 377, n. 4, p. 749–756, 2003.

LENHARDT, L.; BRO, R.; ZEKOVIĆ, I.; DRAMIĆANIN, T.; DRAMIĆANIN, M. D. Fluorescence spectroscopy coupled with PARAFAC and PLS DA for characterization and classification of honey. **Food Chemistry**, v. 175, p. 284–291, 2015.

LERMA-GARCÍA, M. J.; RAMIS-RAMOS, G.; HERRERO-MARTÍNEZ, J. M.; SIMÓ-ALFONSO, E. F. Authentication of extra virgin olive oils by Fourier-transform infrared spectroscopy. **Food Chemistry**, v. 118, n. 1, p. 78–83, 2010.

LI, X.; SUN, C.; LUO, L.; HE, Y. Determination of tea polyphenols content by infrared spectroscopy coupled with iPLS and random frog techniques. **Computers and Electronics in Agriculture**, v. 112, p. 28–35, 2015.

LI, Z.; WANG, J.; XIONG, Y.; LI, Z.; FENG, S. The determination of the fatty acid content of sea buckthorn seed oil using near infrared spectroscopy and variable selection methods for multivariate calibration. **Vibrational Spectroscopy**, v. 84, p. 24–29, 2016.

LIU, K.; CHEN, X.; LI, L.; CHEN, H.; RUAN, X.; LIU, W. A consensus successive projections algorithm - multiple linear regression method for analyzing near infrared spectra. **Analytica Chimica Acta**, v. 858, n. 1, p. 16–23, 2015.

LÓPEZ, M. I.; CALLAO, M. P.; RUISÁNCHEZ, I. A tutorial on the validation of qualitative methods: From the univariate to the multivariate approach. **Analytica Chimica Acta**, v. 891, p. 62–72, 2015.

LYNDGAARD, L. B.; VAN DEN BERG, F.; DE JUAN, A. Quantification of paracetamol through tablet blister packages by Raman spectroscopy and multivariate curve resolution-alternating least squares. **Chemometrics and Intelligent Laboratory Systems**, v. 125, p. 58–66, 2013.

LUTZ, O. M. D.; BONN, G. K.; RODE, B. M.; HUCK, C. W. Reproducible quantification of ethanol in gasoline via a customized mobile near-infrared spectrometer. **Analytica Chimica Acta**, v. 826, n. 1, p. 61–68, 2014.

MAGGIO, R. M.; KAUFMAN, T. S.; DEL CARLO, M.; CERRETANI, L.; BENDINI, A.; CICHELLI, A.; COMPAGNONE, D. Monitoring of fatty acid composition in virgin olive oil by Fourier transformed infrared spectroscopy coupled with partial least squares. **Food Chemistry**, v. 114, n. 4, p. 1549–1554, 2009.

MAJEWSKI, M. Allium Sativum: Facts and Myths Regarding Human Health. **Rocz Panstw Zakl Hig**, v. 65, n. 1, p. 1–8, 2014.

MAN, Y. B. C.; ROHMAN, A.; MANSOR, T. S. T. Differentiation of lard from other edible fats and oils by means of Fourier transform infrared spectroscopy and chemometrics. **JAOCs, Journal of the American Oil Chemists' Society**, v. 88, n. 2, p. 187–192, 2011.

MAPS OF WORLD, Top 10 garlic producing countries. Disponível em <<http://www.mapsofworld.com/world-top-ten/garlic-producing-countries.html>>. Acesso em 12 de fevereiro de 2017.

MARÇO, P. H.; VALDERRAMA, P.; ALEXANDRINO, G. L.; POPPI, R. J.; TAULER, R. Resolução multivariada de curvas com mínimos quadrados alternantes: descrição, funcionamento e aplicações. **Química Nova**, v. 37, n. 9, p. 1525–1532, 2014.

MASSART, D. L.; VANDEGINSTE, B. G. .; BUYDENS, L. M. C.; DE JONG, S.; LEWI, P. J.; SMEYERS-VERBEKE, J. **Handbook of Qualimetrics: Part A**. Amsterdam: Elsevier Science, 1997.

MENDES, P. **Estudo do Teor de Alicina em Alho**. 2008. Instituto Politécnico de Bragança, 2008.

MOREIRA, E. D. T.; PONTES, M. J. C.; GALVÃO, R. K. H.; ARAÚJO, M. C. U. Near infrared reflectance spectrometry classification of cigarettes using the successive projections algorithm for variable selection. **Talanta**, v. 79, p. 1260–1264, 2009.

MOROS, J.; ROTH, M.; GARRIGUES, S.; GUARDIA, M. de la. Preliminary studies about thermal degradation of edible oils through attenuated total reflectance mid-infrared spectrometry. **Food Chemistry**, v. 114, n. 4, p. 1529–1536, 2009.

MUIK, B.; LENDL, B.; MOLINA-DIAZ, A.; VALCARCEL, M.; AYORA-CAÑADA, M. J. Two-dimensional correlation spectroscopy and multivariate curve resolution for the study of lipid oxidation in edible oils monitored by FTIR and FT-Raman spectroscopy. **Analytica Chimica Acta**, v. 593, n. 1, p. 54–67, 2007.

NAES, T.; ISAKSSON, T.; FEARN, T.; DAVIES, T. **A user Friendly guide to Multivariate Calibration and Classification**. Chichester, UK: NIR Publications, 2002.

NENADIS, N.; TSIKOURAS, I.; XENIKAKIS, P.; TSIMIDOU, M. Z. Fourier transform mid-infrared spectroscopy evaluation of early stages of virgin olive oil autoxidation. **European Journal of Lipid Science and Technology**, v. 115, n. 5, p. 526–534, 2013.

NISHIMURA, H.; WIJAYA, C. H.; MIZUTANI, J. Volatile Flavor Components and Antithrombotic Agents : Vinylidithiins from *Allium victorialis* L . v. 1630, n. 2, p. 563–566, 1988.

NKANSAH, K.; ADEDIPE, O.; DAWSON-ANDOH, B.; ATTA-OBENG, E.; SLAHOR, J.; OSBORN, L. Determination of concentration of ACQ wood preservative components by UV-Visible spectroscopy coupled with multivariate data analysis. **Chemometrics and Intelligent Laboratory Systems**, v. 147, p. 157–166, 2015.

NUNES, C. A. Vibrational spectroscopy and chemometrics to assess authenticity, adulteration and intrinsic quality parameters of edible oils and fats. **Food Research International**, v. 60, p. 255–261, 2014.

NUNES, K. M.; ANDRADE, M. V. O.; SANTOS FILHO, A. M. P.; LASMAR, M. C.; SENA, M. M. Detection and characterisation of frauds in bovine meat in natura by non-meat ingredient additions using data fusion of chemical parameters and ATR-FTIR spectroscopy. **Food Chemistry**, v. 205, p. 14–22, 2016.

NURRULHIDAYAH, A. F.; CHE MAN, Y. B.; AL-KAHTANI, H. A.; ROHMAN, A. Application of FTIR spectroscopy coupled with chemometrics for authentication of *Nigella sativa* seed oil. **Spectroscopy**, v. 25, n. 5, p. 243–250, 2011.

OLIVEIRA, R. R.; LIMA, K. M. G.; TAULER, R.; DE JUAN, A. Application of correlation constrained multivariate curve resolution alternating least-squares methods for determination of compounds of interest in biodiesel blends using NIR and UV-visible spectroscopic data. **Talanta**, v. 125, p. 233–241, 2014.

OLIVERI, P.; DOWNEY, G. Multivariate class modeling for the verification of

food-authenticity claims. **Trends in Analytical Chemistry**, v. 35, p. 74–86, 2012.

OLIVIERI, A. C. Practical guidelines for reporting results in single- and multi-component analytical calibration: A tutorial. **Analytica Chimica Acta**, v. 868, p. 10–22, 2015.

OUSSAMA, A.; ELABADI, F.; DEVOS, O. Analysis of Argan Oil Adulteration Using Infrared Spectroscopy. **Spectroscopy Letters**, v. 45, n. 6, p. 458–463, 2012.

PARASTAR, H.; SHAYE, H. Comparative study of partial least squares and multivariate curve resolution for simultaneous spectrophotometric determination of pharmaceuticals in environmental samples. **RSC Adv.**, v. 5, n. 86, p. 70017–70024, 2015.

PASQUINI, C. Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. **Journal of the Brazilian Chemical Society**, v. 14, n. 2, p. 198–219, 2003.

PAVIA, D. L.; LAMPMAN, G. M.; KRIZ, G. S.; VYVYAN, J. R. **Introdução a Espectroscopia**. 4ª Edição ed. CENGAGE Learning, 2010.

PEREIRA, H. V.; AMADOR, V. S.; SENA, M. M.; AUGUSTI, R.; PICCIN, E. Paper spray mass spectrometry and PLS-DA improved by variable selection for the forensic discrimination of beers. **Analytica Chimica Acta**, v. 940, p. 104–112, 2016.

PEREIRA, P. C. S.; FREITAS, C. F.; CHAVES, C. S.; ESTEVÃO, B. M.; PELLOSI, D. S.; TESSARO, A. L.; BATISTELA, V. R.; SCARMINIO, I. S.; CAETANO, W.; HIOKA, N. Chemometry in undergraduate chemistry courses: a proposal for the use of multivariate analysis in the determination of pK_a. **Química Nova**, v. 37, n. 8, p. 1417–1425, 2014.

PETROVSKA, B. B.; CEKOVSKA, S. Extracts from the history and medical properties of garlic. **Pharmacognosy reviews**, v. 4, n. 7, p. 106–110, 2010.

PIERINI, G. D.; FERNANDES, D. D. S.; DINIZ, P. H. G. D.; ARAÚJO, M. C. U.; DI NEZIO, M. S.; CENTURIÓN, M. E. A digital image-based traceability tool of the geographical origins of Argentine propolis. **Microchemical Journal**, v. 128, p. 62–67, 2016.

PISTONESI, M. F.; DI NEZIO, M. S.; CENTURIÓN, M. E.; LISTA, A. G.; FRAGOSO, W. D.; PONTES, M. J. C.; ARAÚJO, M. C. U.; BAND, B. S. F. Simultaneous determination of hydroquinone, resorcinol, phenol, m-cresol and p-cresol in untreated air samples using spectrofluorimetry and a custom multiple linear regression-successive projection algorithm. **Talanta**, v. 83, n. 2, p. 320–323, 2010.

PONTES, M. J. C.; GALVÃO, R. K. H.; ARAÚJO, M. C. U.; MOREIRA, P. N. T.; NETO, O. D. P.; EMÍDIO, G.; SALDANHA, T. C. B. The successive projections algorithm for spectral variable selection in classification problems. **Chemometrics and Intelligent Laboratory Systems**, v. 78, n. 1, p. 11–18, 2005.

PONTES, M. J. C. **Algoritmo das Projeções Sucessivas para seleção de variáveis espectrais em problemas de classificação**, 2009. Tese de Doutorado. UNIVERSIDADE FEDERAL DA PARAÍBA, João Pessoa, 2009.

PONTES, M. J. C.; CORTEZ, J.; GALVÃO, R. K. H.; PASQUINI, C.; ARAÚJO, M. C. U.; COELHO, R. M.; CHIBA, M. K.; ABREU, M. F.; MADARI, B. E. Classification of Brazilian soils by using LIBS and variable selection in the wavelet domain. **Analytica Chimica Acta**, v. 642, p. 12–18, 2009.

PONTES, M. J. C.; PEREIRA, C. F.; PIMENTEL, M. F.; VASCONCELOS, F. V. C.; SILVA, A. G. B. Screening analysis to detect adulteration in diesel/biodiesel blends using near infrared spectrometry and multivariate classification. **Talanta**, v. 85, n. 4, p. 2159–2165, 2011.

PORTAL BRASIL, Levantamento analisa hábitos alimentares dos brasileiros, 2014. Disponível em <<http://www.brasil.gov.br/saude/2014/12/levantamento-analisa-habitos-alimentares-dos-brasileiros>>. Acesso em 12 de fevereiro de 2017.

QIDWAI, W.; ASHFAQ, T. Role of Garlic Usage in Cardiovascular Disease Prevention: An Evidence-Based Approach. **Evidence-Based Complementary and Alternative Medicine**, p. 1–9, 2013.

QUEIROZ, Y. S. De. **Efeito do processamento do alho (*Allium sativum* L.) sobre os seus compostos bioativos e potencial antioxidante in vitro e in vivo**. Tese de doutorado, 2010. Universidade de São Paulo Faculdade de Saúde Pública, 2010.

RABKIN, Y. M. Technological Innovation In Science: The Adoption of Infrared Spectroscopy by Chemists. **Isis**, v. 78, p. 31–54, 1987.

RAHMAN, K. Garlic and aging: new insights into an old remedy. **Ageing Research Reviews**, v. 2, n. 1, p. 39–56, 2003.

RAHMAN, M. S. Allicin and other functional active components in garlic: Health benefits and bioavailability. **International Journal of Food Properties**, v. 10, n. 2, p. 245–268, 2007.

RAHMANIA, H.; SUDJADI; ROHMAN, A. The employment of FTIR spectroscopy in combination with chemometrics for analysis of rat meat in meatball formulation. **Meat Science**, v. 100, p. 301–305, 2015.

RATTI, C.; ARAYA-FARIAS, M.; MENDEZ-LAGUNAS, L.; MAKHLOUF, J. Drying of Garlic (*Allium sativum*) and Its Effect on Allicin Retention. **Drying Technology**, v. 25, n. 2, p. 349–356, 2007.

REHMAN, F.; SAMYA, M. International Journal of Pharma and Bio Sciences ISSN. **International Journal of Pharma and Bio Sciences**, v. 3, n. 3, p. 381–390, 2012.

REINHART, K. M.; COLEMAN, C. I.; TEEVAN, C.; VACHHANI, P.; WHITE, C. M. Effects of garlic on blood pressure in patients with and without systolic hypertension: A meta-analysis. **The Annals of Pharmacotherapy**, v. 42, n. 12, p. 1766–1771, 2008.

REVISTA EXAME, ABIAD divulga pesquisa inédita sobre consumo de Suplementos Alimentares no País, 2016. Disponível em <<http://exame.abril.com.br/negocios/dino/abiad-divulga-pesquisa-inedita-sobre-consumo-de-suplementos-alimentares-no-pais-dino89090369131/>> Acesso em 12 de fevereiro de 2017.

RICHARDS, S. E.; BECKER, E.; TAULER, R.; WALMSLEY, A. D. A novel approach to the quantification of industrial mixtures from the Vinyl Acetate Monomer (VAM) process using Near Infrared spectroscopic data and a Quantitative Self Modeling Curve Resolution (SMCR) methodology. **Chemometrics and Intelligent Laboratory Systems**, v. 94, n. 1, p. 9–18, 2008.

RINNAN, Å.; BERG, F. Van Den; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. **TrAC Trends in Analytical Chemistry**, v. 28, n. 10, p. 1201–1222, 2009.

RÍOS-REINA, R.; CALLEJÓN, R. M.; OLIVER-POZO, C.; AMIGO, J. M.; GARCÍA-GONZÁLEZ, D. L. ATR-FTIR as a potential tool for controlling high quality vinegar categories. **Food Control**, v. 78, p. 230–237, 2017.

RODRIGUEZ-SAONA, L. E.; ALLENDORF, M. E. Use of FTIR for Rapid Authentication and Detection of Adulteration of Food. **Annual Review of Food Science and Technology**, v. 2, n. 1, p. 467–483, 2011.

ROHMAN, A.; MAN, Y. B. C. Monitoring of Virgin Coconut Oil (Vco) Adulteration With Palm Oil Using Fourier Transform Infrared Spectroscopy. **Journal of Food Lipids**, v. 16, n. 4, p. 618–628, 2009a.

ROHMAN, A.; MAN, Y. B. C. Analysis of Cod-Liver Oil Adulteration Using Fourier Transform Infrared (FTIR) Spectroscopy. **Journal of the American Oil Chemists' Society**, v. 86, n. 12, p. 1149–1153, 2009b.

ROHMAN, A.; MAN, Y. B. C. Fourier transform infrared (FTIR) spectroscopy for analysis of extra virgin olive oil adulterated with palm oil. **Food Research International**, v. 43, n. 3, p. 886–892, 2010.

ROHMAN, A.; MAN, Y. B. C. Analysis of chicken fat as adulterant in cod liver oil using Fourier transform infrared (FTIR) spectroscopy and chemometrics. **CyTA - Journal of Food**, v. 9, n. 3, p. 187–191, 2011a.

ROHMAN, A.; MAN, Y. B. C. Potential Use of FTIR-ATR Spectroscopic Method for Determination of Virgin Coconut Oil and Extra Virgin Olive Oil in Ternary Mixture Systems. **Food Analytical Methods**, v. 4, n. 2, p. 155–162, 2011b.

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, 2011c.

ROHMAN, A.; MAN, Y. B. C.; HASHIM, P.; ISMAIL, A. FTIR spectroscopy combined with chemometrics for analysis of lard adulteration in some vegetable oils. **CyTA - Journal of Food**, v. 9, n. 2, p. 96–101, 2011.

ROSE, P.; WHITEMAN, M.; MOORE, P. K.; ZHU, Y. Z. Bioactive S-alk(en)yl cysteine sulfoxide metabolites in the genus *Allium*: the chemistry of potential

therapeutic agents. **Natural Product Reports**, v. 22, n. 3, p. 351–368, 2005.

SANTOS, V. H. J. M.; RAMOS, A. S.; PIRES, J. P.; ENGELMANN, P. de M.; LOUREGA, R. V.; KETZER, J. M. M.; RODRIGUES, L. F. Discriminant analysis of biodiesel fuel blends based on combined data from Fourier Transform Infrared Spectroscopy and stable carbon isotope analysis. **Chemometrics and Intelligent Laboratory Systems**, v. 161, p. 70–78, 2017.

SAURINA, J.; TAULER, R. Strategies for solving matrix effects in the analysis of triphenyltin in sea-water samples by three-way multivariate curve resolution. **The Analyst**, v. 125, n. 11, p. 2038–2043, 2000.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627–1639, 1964.

SILVA, A. C.; PONTES, L. F. B. L.; PIMENTEL, M. F.; PONTES, M. J. C. Detection of adulteration in hydrated ethyl alcohol fuel using infrared spectroscopy and supervised pattern recognition methods. **Talanta**, v. 93, p. 129–134, 2012.

SILVA, C. S.; BORBA, F. de S. L.; PIMENTEL, M. F.; PONTES, M. J. C.; HONORATO, R. S.; PASQUINI, C. Classification of blue pen ink using infrared spectroscopy and linear discriminant analysis. **Microchemical Journal**, v. 109, p. 122–127, 2013.

SINELLI, N.; CASALE, M.; DI, V.; OLIVERI, P.; BASSI, D.; TURA, D.; CASIRAGHI, E. Varietal discrimination of extra virgin olive oils by near and mid infrared spectroscopy. **Food Research International journal**, v. 43, p. 2126–2131, 2010.

SINGH, U.; KUMAR, S.; DHAKAL, S. Future prospect of garlic usage in clinical practice of hyperlipidemia : A review. **International Journal of Herbal Medicine**, v. 3, n. 2, p. 38–43, 2015.

SOARES, A. D. S.; GALVÃO, R. K. H.; ARAÚJO, M. C. U.; SOARES, S. F. C.; PINTO, L. A. Multi-core computation in chemometrics: Case studies of voltammetric and NIR spectrometric analyses. **Journal of the Brazilian Chemical Society**, v. 21, n. 9, p. 1626–1634, 2010.

SOARES, S. F. C.; GALVÃO, R. K. H.; PONTES, M. J. C.; ARAÚJO, M. C. U. A New Validation Criterion for Guiding the Selection of Variables by the Successive Projections Algorithm in Classification Problems. **Journal of the Brazilian Chemical Society**, v. 25, n. 1, p. 176–181, 2014.

SOARES, S. F. C.; GOMES, A. A.; ARAUJO, M. C. U.; FILHO, A. R. G.; GALVÃO, R. K. H. The successive projections algorithm. **TrAC - Trends in Analytical Chemistry**, v. 42, p. 84–98, 2013.

SOUTO, U. T. de C. P.; BARBOSA, M. F.; DANTAS, H. V.; PONTES, A. S.; LYRA, W. da S.; DINIZ, P. H. G. D.; ARAÚJO, M. C. U.; SILVA, E. C. Screening for Coffee Adulteration Using Digital Images and SPA-LDA. **Food Analytical Methods**, v. 8, p. 1515–1521, 2014.

SOUTO, U. T. de C. P.; BARBOSA, M. F.; DANTAS, H. V.; PONTES, A. S.; LYRA, W. da S.; DINIZ, P. H. G. D.; ARAÚJO, M. C. U.; SILVA, E. C. Identification of adulteration in ground roasted coffees using UV-Vis spectroscopy and SPALDA. **LWT - Food Science and Technology**, v. 63, p. 1037–1041, 2015.

SOUZA, A. M. de; BREITKREITZ, M. C.; FILGUEIRAS, P. R.; ROHWEDDER, J. J. R.; POPPI, R. J. Experimento didático de quimiometria para calibração multivariada na determinação de paracetamol em comprimidos comerciais utilizando espectroscopia no infravermelho próximo: um tutorial, parte II. **Química Nova**, v. 36, n. 7, p. 1057–1065, 2013.

SUN, D. W. (ed.). **Infrared Spectroscopy for Food Quality Analysis and Control**. 1ª edição ed. Elsevier, 2009.

TALPUR, M. Y.; HASSAN, S. S.; SHERAZI, S. T. H.; MAHESAR, S. A.; KARA, H.; KANDHRO, A. A. A simplified FTIR chemometric method for simultaneous determination of four oxidation parameters of frying canola oil. **Spectrochimica acta part a: molecular and biomolecular spectroscopy**, v. 149, p. 656–661, 2015.

TAPP, H. S.; DEFERNEZ, M.; KEMSLEY, E. K. FTIR Spectroscopy and Multivariate Analysis Can Distinguish the Geographic Origin of Extra Virgin Olive Oils. **J. Agric. Food Chem.**, v. 51, p. 6110–6115, 2003.

TEIXEIRA DOS SANTOS, C. A.; PÁSCOA, R. N. M. J.; PORTO, P. A. L. S.; CERDEIRA, A. L.; LOPES, J. A. Application of Fourier-transform infrared spectroscopy for the determination of chloride and sulfate in wines. **LWT - Food Science and Technology**, v. 67, p. 181–186, 2016.

TOULOUPAKIS, E.; GHANOTAKIS, D. F. Nutraceutical Use of Garlic Sulfur-Containing Compounds. **Bio-Farms for Nutraceuticals: Functional Food and Safety Control by Biosensors**, v. 698, p. 110–121, 2010.

VALDERRAMA, L.; VALDERRAMA, P. Nondestructive identification of blue pen inks for documentoscopy purpose using iPhone and digital image analysis including an approach for interval confidence estimation in PLS-DA models validation. **Chemometrics and Intelligent Laboratory Systems**, v. 156, p. 188–195, 2016.

VAN DE VOORT, F. R. Fourier transform infrared spectroscopy applied to food analysis. **Food Research International**, v. 25, n. 5, p. 397–403, 1992.

VAN DE VOORT, F. R. V.; ISMAIL, A. A.; SEDMAN, J.; DUBOIS, J.; NICODEMO, T. The determination of peroxide value by fourier transform infrared spectroscopy. **Journal of the American Oil Chemists' Society**, v. 71, n. 9, p. 921–926, 1994.

VAN DE VOORT, F. R. V.; ISMAIT, A. A.; SEDMAN, J. A rapid, automated method for the determination of cis and trans content of fats and oils by Fourier Transform Infrared Spectroscopy. **JAOCs, Journal of the American Oil Chemists' Society**, v. 72, n. 8, p. 873–880, 1995.

VAN DE VOORT, F. R. V.; MEMON, K. P.; SEDMAN, J.; ISMAIL, A. A.

Determination of Solid Fat Index by Fourier Transform Infrared Spectroscopy. **JAACS, Journal of the American Oil Chemists' Society**, v. 73, n. 4, p. 411–416, 1996.

VARMUZA, KURT; FILZMOSE, P. **Introduction to Multivariate Statistical Analysis in Chemometrics**. CRC Press, 2009. ISBN 9781420059496.

VASCONCELOS, F. V. C.; SOUZA, P. F. B.; PIMENTEL, M. F.; PONTES, M. J. C.; PEREIRA, C. F. Using near-infrared overtone regions to determine biodiesel content and adulteration of diesel/biodiesel blends with vegetable oils. **Analytica Chimica Acta**, v. 716, p. 101–107, 2012.

VLACHOS, N.; SKOPELITIS, Y.; PSAROUDAKI, M.; KONSTANTINIDOU, V.; CHATZILAZAROU, A.; TEGOU, E. Applications of Fourier transform-infrared spectroscopy to edible oils. **Analytica Chimica Acta**, v. 573–574, p. 459–465, 2006.

VONGSVIVUT, J.; HERAUD, P.; ZHANG, W.; KRALOVEC, J. A.; MCNAUGHTON, D.; BARROW, C. J. Quantitative determination of fatty acid compositions in micro-encapsulated fish-oil supplements using Fourier transform infrared (FTIR) spectroscopy. **Food Chemistry**, v. 135, n. 2, p. 603–609, 2012.

WANG, A.; XIE, L. Technology using near infrared spectroscopic and multivariate analysis to determine the soluble solids content of citrus fruit. **Journal of Food Engineering**, v. 143, p. 17–24, 2014.

WANG, J.; CAO, Y.; SUN, B.; WANG, C. Physicochemical and release characterisation of garlic oil- β -cyclodextrin inclusion complexes. **Food Chemistry**, v. 127, n. 4, p. 1680–1685, 2011.

WILLIAMS, P.; SOBERING, D. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. **Journal of Near Infrared Spectroscopy**, v. 1, n. 1, p. 25–32, 1993.

WOLD, S. Personal memories of the early PLS development. **Chemometrics and Intelligent Laboratory Systems**, v. 58, n. 2, p. 83–84, 2001.

WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 58, n. 2, p. 109–130, 2001.

XIAOBO, Z.; JIEWEN, Z.; POVEY, M. J. W.; HOLMES, M.; HANPIN, M. Variables selection methods in near-infrared spectroscopy. **Analytica Chimica Acta**, v. 667, n. 1–2, p. 14–32, 2010.

XU, H.; QI, B.; SUN, T.; FU, X.; YING, Y. Variable selection in visible and near-infrared spectra: Application to on-line determination of sugar content in pears. **Journal of Food Engineering**, v. 109, n. 1, p. 142–147, 2012.

YAN, X.; WANG, Z.; BARLOW, P. Quantitative estimation of garlic oil content in garlic oil based health products. **Food Chemistry**, v. 45, n. 2, p. 135–139, 1992.

YAN, X.; WANG, Z.; BARLOW, P. Quantitative determination and profiling of total sulphur compounds in garlic health products using a simple GC procedure. **Food**

Chemistry, v. 47, n. 3, p. 289–294, 1993.

YANG, H.; IRUDAYARAJ, J.; PARADKAR, M. M. Food Chemistry Discriminant analysis of edible oils and fats by FTIR, FT-NIR and FT-Raman spectroscopy. **Food Chemistry**, v. 93, p. 25–32, 2005.

YU-YAN, Y. .; LIJUAN, L. . Recent Advances on the Nutritional Effects Associated with the Use of Garlic as a Supplement. **Journal of Nutrition**, p. 1118–1119, 2001.

YU, T. H.; WU, C. M.; LIOU, Y. C. Volatile compounds from garlic. **Journal of Agricultural and Food Chemistry**, v. 37, p. 725–730, 1989.

ANEXO – Artigo publicado no periódico *Microchemical Journal*, referente à classificação das cápsulas de óleo de alho quanto ao seu teor



Screening analysis of garlic-oil capsules by infrared spectroscopy and chemometrics☆



Valeria Visani^a, José Marcelino Souza Netto^b, Ricardo Saldanha Honorato^c,
Mário César Ugulino de Araújo^a, Fernanda Araújo Honorato^{b,*}

^a Universidade Federal da Paraíba, Departamento de Química, João Pessoa, PB 58051-900, Brazil

^b Universidade Federal de Pernambuco, Departamento de Engenharia Química, Recife, PE 50740-521, Brazil

^c Polícia Federal, Recife, PE 50030-230, Brazil

ARTICLE INFO

Article history:

Received 15 November 2016

Received in revised form 14 March 2017

Accepted 10 April 2017

Available online 12 April 2017

Keywords:

Garlic oil

Medium infrared spectroscopy

Chemometrics

Pattern recognition methods

ABSTRACT

The search for a healthier life has led to an increase in the consumption of natural products with certain functional and/or nutritional properties while avoiding the collateral and harmful effects of medicines. In this context, garlic oil, in the form of gelatinous capsules, contains various substances considered beneficial to health, to reduce cholesterol and blood pressure, in the treatment of influenza and diabetes and to prevent the development of tumors and cardiovascular diseases. Currently, garlic oil is classified by ANVISA (National Health Surveillance Agency) as a “new food”, denomination given to food or substances without historical consumption in the country, or foods with substances already consumed and that may be added or used at much higher levels than currently observed in food. While there is no established law ensuring safety and quality of garlic oil commercialized in capsules, the development of simple and rapid methods of analysis to verify the authenticity of these products is of great importance. In this paper, we propose a screening analysis method to do this, using spectroscopy in the mid-infrared region and supervised pattern recognition methods, aimed at classifying garlic oil capsules by concentration levels. For this purpose, spectra of 136 samples from 11 brands of commercial capsules and 123 solutions of garlic oil in soybean oil (0.00%, 0.25%, 0.50%, 0.75%, 1.00%, 1.25%, 1.50% and 1.75% (m/m)) were obtained using an attenuated total reflectance accessory. The spectra obtained were subjected to different preprocessing (Standard Normal Variate - SNV, Multiplicative Scatter Correction - MSC, and derivation by Savitzky Golay filters - SG) and uninformative spectral regions were eliminated. The supervised pattern recognition methods Partial Least Squares Discriminant Analysis (PLS-DA) and Linear Discriminant Analysis (LDA) with variable selection by the Successive Algorithm Projections (SPA) were used to build models capable for classifying commercial capsules into three classes: class 1 (low dosage) from 0.00% to 0.50%, class 2 (medium dosage, usual) from 0.75% to 1.25% and class 3 (high dosage) from 1.50% to 1.75%. The PLS-DA provided, in the validation stage, a correct classification rate of 81.3% against 95.9% of the SPA-LDA method. In the SPA-LDA classification, only the 924 cm⁻¹ variable was selected. This is where a band from the deformation of the vinyl groups of the sulfides and vinyl dithiols, main constituents of garlic oil, was observed. For the 136 capsules tested a rate of 88.2% agreement with the information contained on their labels was obtained. The high number of capsules classified in the wrong classes suggests that their labels had presented incorrect information about their concentration levels. The screening methodology proposed in this work is a simple, fast and effective tool that can be useful to investigate the content of garlic oil gelatinous capsules.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The quality of modern life is affected by many diverse factors, such as a diet poor in nutrients and high in fats harmful to health, lack of

physical activities, work overload, and stress. This has facilitated the emergence of many diseases, since the body is vulnerable. Because of this, there has been a considerable increase in the demand for natural products with specific functional and/or nutritional properties. These can bring about health benefits, in some cases eliminating the need of medicine and its collateral effects. Among these products, garlic oil marketed in the form of gelatinous capsules is one of the most popular and has many beneficial health properties. The major ones include reducing LDL cholesterol and blood pressure, acting as an antiviral agent

☆ Selected paper from the 18th Brazilian Meeting on Analytical Chemistry, 18–21 September, 2016, Brazil.

* Corresponding author.

E-mail address: fah@ufpe.br (F.A. Honorato).