



Universidade Federal da Paraíba
Centro de Ciências Exatas e da Natureza
Departamento de Química
Programa de Pós-Graduação em Química

Dissertação de Mestrado

**Desenvolvimento de metodologias analíticas verdes para detecção de
adulterantes em leite em pó caprino**

José Leonardo Alves Ferreira

João Pessoa – PB – Brasil

Outubro/2024



Universidade Federal da Paraíba
Centro de Ciências Exatas e da Natureza
Departamento de Química
Programa de Pós-Graduação em Química

**Desenvolvimento de metodologias analíticas verdes para detecção de
adulterantes em leite em pó caprino**

José Leonardo Alves Ferreira*

Dissertação apresentada ao Programa de Pós-Graduação em Química da Universidade Federal da Paraíba, como parte das exigências para obtenção do título de mestre em Química/ Área de concentração: Química Analítica e Quimiometria.

Orientador: **Prof. Dr. Luciano Farias de Almeida**

Coorientador: **Dr. David Douglas de Sousa Fernandes**

***Bolsista CNPQ**

João Pessoa – PB – Brasil

Outubro/2024

Catálogo na publicação
Seção de Catalogação e Classificação

F383d Ferreira, José Leonardo Alves.

Desenvolvimento de metodologias analíticas verdes para detecção de adulterantes em leite em pó caprino / José Leonardo Alves Ferreira. - João Pessoa, 2024.
126 f. : il.

Orientação: Luciano Farias de Almeida.

Coorientação: David Douglas de Sousa Fernandes.
Dissertação (Mestrado) - UFPB/CCEN.

1. Leite de cabra - Adulteração. 2. Química analítica verde. 3. Espectroscopia NIR. 4. Espectroscopia RAMAN. 5. Quimiometria. I. Almeida, Luciano Farias de. II. Fernandes, David Douglas de Sousa. III. Título.

UFPB/BC

CDU 637.12'639(043)

“Desenvolvimento de metodologias analíticas verdes para detecção de adulterantes em leite em pó caprino”.

Dissertação de Mestrado apresentada pelo discente José Leonardo Alves Ferreira e aprovada pela banca examinadora em 23 de outubro de 2024.

Documento assinado digitalmente
 **LUCIANO FARIAS DE ALMEIDA**
Data: 25/10/2024 18:39:12-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Luciano Farias de Almeida
DQ/UFPB
Orientador/Presidente

Documento assinado digitalmente
 **DAVID DOUGLAS DE SOUSA FERNANDES**
Data: 24/10/2024 20:39:59-0300
Verifique em <https://validar.iti.gov.br>

Dr. David Douglas de Sousa Fernandes
UEPB/CG-PB
Co-Orientador

Documento assinado digitalmente
 **JOSE LICARION PINTO SEGUNDO NETO**
Data: 24/10/2024 14:12:40-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. José Licarion Pinto Segundo Neto
UERJ/Rio de Janeiro-RJ
Examinador

Documento assinado digitalmente
 **MARCIO JOSE COELHO DE PONTES**
Data: 24/10/2024 15:14:11-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Márcio José Coelho de Pontes
DQ/UFPB
Examinador

*“É necessário sempre acreditar que o sonho é possível
Que o céu é o limite e você, truta, é imbatível”
(Afro-X e Racionais MC's)*

AGRADECIMENTOS

Aos meus pais, Ana Martha e Flávio, que independente das dificuldades eu os amo muito, pois sempre foram meus exemplos, minha base, meu alicerce na vida, obrigado por tudo. Especialmente aos meus avós (Maria Alberice, Olívio, Josefa e Francisco) e principalmente à minha Tia Maria Helena, queria que todos estivessem comigo presenciando essa conquista.

Aos meus amigos que tenho para minha vida (Gabrihel, Bruno, Rogério, Ramon, Gabriel dos Santos, Cleideane) e a minha companheira, Kyedja, obrigado por todo apoio em situações inimagináveis que já passamos.

Ao meu professor Diego Bizerra, que foi minha maior inspiração para os caminhos da Química. Agradeço por todos conselhos, ensinamentos e direcionamentos durante minha formação escolar.

Ao meu orientador na graduação, Prof. Dr. Anderson Simões, por sempre me incentivar e apoiar, e pelos conselhos e ensinamentos profissionais e pessoais.

Ao meu orientador, Prof. Dr. Luciano Farias, pela orientação e apoio.

Agradecimento muito especial ao meu amigo e coorientador Dr. David Douglas. Suas orientações foram essenciais para o desenvolvimento deste trabalho e de meu desenvolvimento pessoal e profissional. Agradeço pelo apoio constante e pela paciência em cada etapa do processo. Sou imensamente grato por ter tido a oportunidade de trabalhar com o senhor.

Aos membros das bancas Prof. José Germano, Prof. Márcio Coelho e Prof. José Licarion, pelas contribuições que colaboraram e enriqueceram o trabalho.

A minha amiga e supervisora Amanda Cecília, que chegou no término, mas saiba que sua ajuda foi imprescindível para a conclusão do trabalho

Aos amigos e técnicos Wellington Lyra e Renato Navarro por todo apoio durante as análises experimentais.

A todos professores e colegas de curso e de laboratório (LAQA), que sempre me ajudaram e apoiaram com o necessário, obrigado por todo apoio.

Às instituições UFPB, UEPB campina grande e UFOB pelo apoio inestimável e pela colaboração ao longo deste projeto.

Ao PPGQ/UFPB, CNPQ e FAPESQ, pelo fomento da pesquisa.

Agradeço a todos, peço desculpas se esqueci de alguém, são tantas pessoas, mas saibam que sou muito agradecido a todos, e que essas palavras não são nem um sopro do real sentimento de gratidão que tenho por cada um.

SUMÁRIO

1 INTRODUÇÃO.....	18
2 OBJETIVOS.....	21
2.1 Objetivo Geral.....	21
2.2 Objetivos Específicos.....	21
3 FUNDAMENTAÇÃO TEÓRICA.....	23
3.1 Leite.....	23
3.1.1 Leite caprino.....	24
3.1.2 Leite em pó.....	26
3.1.3 Leite caprino x Leite bovino.....	29
3.1.4 Alergia à Proteína do Leite de Vaca (APLV).....	33
3.2 Controle de qualidade.....	35
3.2.1 Fraude alimentar.....	36
3.3 Metodologias analíticas para autenticidade do leite caprino.....	38
3.4 Espectroscopia Vibracional.....	41
3.4.1 Espectroscopia no Infravermelho Próximo - NIR.....	43
3.4.2 Espectroscopia RAMAN.....	46
3.5 Instrumentação Portátil.....	48
3.6 Quimiometria.....	50
3.6.1 Pré-processamento dos dados.....	51
3.6.2 Técnicas de reconhecimento de padrões.....	53
3.6.3 Análise por componentes principais.....	53
3.6.4 Mínimos quadrados parciais de uma classe (OC-PLS).....	57
3.6.5 Modelagem independente e flexível por analogia de classes direcionada pelos dados (DD-SIMCA).....	60
3.6.6 Procrustes cross validation - PCV.....	63
3.6.7 Métricas de Desempenho para modelos de classificação.....	66
4 METODOLOGIA.....	70
4.1 Aquisição das Amostras.....	70
4.2 NIR.....	71
4.2.1 Aquisição dos Espectros.....	71
4.2.2 Seleção de amostras.....	72
4.3 RAMAN.....	72
4.3.1 Aquisição dos Espectros.....	72
4.3.2 Seleção de amostras.....	73
4.4 Procedimento Quimiométrico e Software.....	74
4.4.1 Tratamento dos espectros NIR.....	74
4.4.2 Tratamento dos espectros RAMAN.....	75

5 RESULTADOS E DISCUSSÃO.....	77
5.1 NIR.....	77
5.1.1 Análise Exploratória.....	77
5.1.2 Autenticação do Leite em pó caprino.....	82
5.2 RAMAN.....	86
5.2.1 Análise Exploratória.....	86
5.2.2 Autenticação do Leite em pó caprino.....	92
6 CONCLUSÕES E PERSPECTIVAS.....	100
6.1 Conclusões.....	100
6.2 Perspectivas.....	101
REFERÊNCIAS.....	103
APÊNDICE.....	121

LISTA DE ILUSTRAÇÕES

Quadro 1. Classificação do leite de cabra quanto ao teor de gordura.....	25
Figura 1. Processo de Spray-dryer.....	28
Figura 2. Esquema representativo do percurso do leite durante a produção do leite em pó....	29
Figura 3. Estruturas da galactose, glicose e lactose.....	29
Figura 4. Ilustração da estrutura química de um triglicerídeo.....	31
Quadro 2. Tipos de Ácidos graxos e suas principais características.....	31
Figura 5. Classificação das reações adversas aos alimentos.....	34
Figura 6. Tipos de fraude em alimentos.....	37
Quadro 3. Tipos de Adulterações em alimentos.....	37
Quadro 4. Metodologias utilizadas na autenticação de leites e seus derivados.....	40
Figura 7. Diagrama de energia dos modelos osciladores harmônico e anarmônico.....	44
Figura 8. Vibrações moleculares mais comuns encontradas na espectroscopia infravermelho.....	45
Figura 9. Principais bandas analíticas e posições de pico relativas para absorções proeminentes no infravermelho próximo.....	45
Figura 10. Representação esquemática dos tipos de espalhamentos observados. Os níveis energéticos S_0 e S_1 correspondem ao nível fundamental e ao excitado da molécula, respectivamente. O estado virtual (E_{vir}) é representado pelas linhas tracejadas.....	47
Figura 11. Ilustração idealizada do método de PCA para duas variáveis aleatórias x_1 e x_2 (a), e da centragem na média dos dados (b).....	54
Figura 12. Representação gráfica das PCs, com indicação dos pesos de PC1 e dos escores da amostra B.....	55
Figura 13. Gráfico do modelo OC-PLS na etapa de predição da classe-alvo.....	59
Figura 14. Gráfico do modelo OC-PLS na etapa de predição da classe não-alvo.....	60
Figura 15. Gráfico do modelo DD-SIMCA na etapa de treinamento.....	62
Figura 16. Gráfico do modelo DD-SIMCA na etapa de teste.....	63
Figura 17. Ilustração do conceito de conjunto de pseudo-validação.....	65

Figura 18. Representação gráfica das métricas de avaliação de desempenho do modelo de classificação DD-SIMCA (a).....	67
Figura 19. Esquema ilustrativo do processo de aquisição das amostras.....	71
Figura 20. Esquema ilustrativo do processo de aquisição dos Espectros NIR.....	72
Figura 21. Esquema ilustrativo do processo de aquisição dos Espectros RAMAN.....	73
Figura 22. Espectros NIR de leites caprino (em vermelho) e de misturas caprino + adulterantes (em verde): espectro bruto.....	77
Figura 23. Espectros NIR médios das classes: Leite caprino puro (Vermelho), Leite caprino adulterados (Verde): Espectros brutos cortados a partir de 1660 nm (a) e suavizados com filtro Savitzky – Golay 21 pontos acoplado com (b) OFF, (c) MSC e (d) SNV; aplicando filtro derivativos de 1° e 2° ordens Savitzky– Golay com janelas de 21 (e) e 25 (f) pontos.....	78
Figura 24. Amostras de Leites caprino (em vermelho) e de misturas caprino + adulterantes (em verde): Gráficos de escores (PC1 x PC2) para os dados espectrais cortados a partir de 1660 nm (a) e suavizados com filtro Savitzky – Golay 21 pontos acoplado com (b) OFF, (c) MSC e (d) SNV; aplicando filtro derivativos de 1° e 2° ordens Savitzky – Golay com janelas de 21 (e) e 25 (f) pontos, respectivamente.....	79
Figura 25. Amostras de Leites caprino (em vermelho) e de misturas caprino + adulterantes (em verde): Gráficos de escores (PC1 x PC3) para os dados espectrais cortados a partir de 1660 nm (a) e suavizados com filtro Savitzky – Golay 21 pontos acoplado com (b) OFF, (c) MSC e (d) SNV; aplicando filtro derivativos de 1° e 2° ordens Savitzky – Golay com janelas de 21 (e) e 25 (f) pontos, respectivamente.....	80
Figura 26. Gráficos dos pesos, PC1 (Preto), PC2 (Vermelho) e PC3 (Azul): para os dados espectrais cortados a partir de 1660 nm (a) e suavizados com filtro Savitzky – Golay 21 pontos acoplado com (b) OFF, (c) MSC e (d) SNV; aplicando filtro derivativos de 1° e 2° ordens Savitzky– Golay com janelas de 21 (e) e 25 (f) pontos, respectivamente.....	81
Figura 27. Gráficos de área de aceitação para o conjunto de treinamento (a) e teste (b) empregando DD-SIMCA.....	83
Figura 28. Espectros NIR SGS-21p-OFF: Amostra extrema (Azul) e médio do conjunto de treinamento restante (Vermelho).....	84
Figura 29. Gráficos de área de aceitação para o conjunto de teste, rotulando os tipos de adulterantes por cor.....	84
Figura 30. Gráficos de área de aceitação para o conjunto de treinamento (a) e teste (c) usando algoritmo OC- PLS. Magnificação da região destacada por um retângulo no canto inferior direito (b) e esquerdo (d).....	85

Figura 31. Gráficos de área de aceitação para o conjunto de teste usando algoritmo OC- PLS (a). Magnificação da região destacada por um retângulo no canto inferior esquerdo (b).....	86
Figura 32. Espectros RAMAN médios das classes: Leite caprino puro (Azul), Leite caprino adulterado com leite bovino (Rosa) e Leite bovino puro (Verde).....	87
Figura 33. Espectros RAMAN pré processados com SGD 25p: Leite caprino puro (Azul) e Leite caprino adulterado com leite bovino (Rosa).....	88
Figura 34. Gráficos de escores para os dados corrigidos com OFF (a), MSC (b), SNV (c) e aplicando filtro derivativo de 1° ordem Savitzky – Golay com janelas de 25 pontos (d), respectivamente.....	89
Figura 35. Gráficos de escores para os dados Brutos (b) e corrigidos aplicando o filtro derivativo de 1° ordem Savitzky – Golay com janelas de 25 pontos (b), respectivamente, rotulados com teores.....	89
Figura 36. Gráficos de escores (PC1 vs. PC3) para os dados corrigidos com SGD para janelas de 25 pontos.....	90
Figura 37. Gráfico dos pesos da PCA obtidos para PC1 (linha preta), PC2 (linha vermelha) e PC3(linha preta), a partir dos espectros Raman de amostras puras de leite de cabra em pó Brutos (a) e pré-processadas com OFF (b), MSC (c), SNV (d) e a SGD com um polinômio de segunda ordem e uma janela de 25 pontos (e).....	91
Figura 38. Gráficos de área de aceitação para o conjunto de treinamento (a, b) e teste (c, d) empregando DD-SIMCA, abordagem 1.....	96
Figura 39. Gráficos de área de aceitação para o conjunto de treinamento (a, b) e teste (c, d) empregando DD-SIMCA, abordagem 2.....	97
Figura 40. Gráficos de área de aceitação para o conjunto de treinamento (a, b) e teste (c, d) empregando DD-SIMCA, abordagem 3.....	98
Figura A1. Gráficos de escores (PC1 x PC2) para os dados espectrais cortados a partir de 1660 nm (a), suavizados com filtro Savitzky – Golay 21 pontos acoplado com (b) OFF, (c) MSC e (d) SNV; e aplicando filtro derivativos de 1° e 2° ordens Savitzky – Golay com janelas de 21 (e) e 25 (f) pontos, respectivamente. Gráficos à esquerda: amostras de Leites caprino (em vermelho) e de misturas caprino e amido (em alaranjado), composto lácteo (em verde), extrato de soja (em azul) e leite de vaca (em amarelo). Gráficos à direita seguem a legenda das cores por teor de adulteração.....	121
Figura A2. Gráficos de área de aceitação para o conjunto de treinamento (a, c) e teste (b, d) empregando DD-SIMCA, abordagens 1 (a, b) e 2 (c, d), pré-processamento OFF.....	123

Figura A3. Gráficos de área de aceitação para o conjunto de treinamento (a, c) e teste (b, d) empregando DD-SIMCA, abordagem 3, para $\alpha = 0,01$ (a, b) e $\alpha = 0,05$ (c, d), pré-processamento OFF.....124

Figura A4. Gráficos de área de aceitação para o conjunto de treinamento (a, c) e teste (b, d) empregando DD-SIMCA, abordagem 2, para $\alpha = 0,01$ (a, b) e $\alpha = 0,05$ (c, d), pré-processamento SGD1_25p.....125

LISTA DE TABELAS

Tabela 1. Percentuais dos componentes: Leite de cabra x Leite de vaca.....	32
Tabela 2. Faixas espectrais para as diferentes regiões do infravermelho.....	42
Tabela 3. Níveis de teores de adulteração utilizados.....	70
Tabela 4. Quantidade de amostras por nível de adulteração em cada experimento.....	71
Tabela 5. Diferentes estratégias de particionamento de amostras para construir e validar os modelos DD-SIMCA.....	73
Tabela 6. Resultados em termos de sensibilidade e especificidade para os modelos DD-SIMCA e OC-PLS nos conjuntos de treinamento e teste.....	82
Tabela 7. Resultados obtidos por DD-SIMCA ($\alpha = 0,01$) para a autenticação do leite de cabra em pó usando espectroscopia RAMAN.....	94
Tabela 8. Resultados obtidos por DD-SIMCA ($\alpha = 0,05$) para a autenticação do leite de cabra em pó usando espectroscopia RAMAN.....	95

LISTA DE ABREVIATURAS E SIGLAS

ACC - Acurácia.

ACR - Resíduos Absolutos Centralizados.

APLV - Alergia à Proteína do Leite de vaca.

DD-SIMCA - Modelagem Independente e Flexível por Analogia de Classes direcionado pelos dados (*Data-Driven – Soft Independent Modelling of Class Analogy*).

DoF - Graus de liberdade (*Degrees of freedom*).

EFF - Eficiência.

FIR - Infravermelho Distante (*Far Infrared*).

FN - Falsos Negativos (*False Negatives*).

FP - Falsos Positivos (*False Positives*).

GA - Algoritmo Genético (*Genetic Algorithm*).

HCA - Análise de Agrupamento Hierárquico (*Hierarchical Clustering Analysis*).

IBGE - Instituto Brasileiro de Geografia e Estatística.

IgE - Imunoglobulina E.

KS - Algoritmo Kennard-Stone (*Kennard-Stone Algorithm*).

LBC - Correção de linha de base Linear (*Linear Baseline Correction*).

LOO - *Leave-one-out*.

LV - Variáveis latentes (*Latents Variables*).

MAPA - Ministério da Agricultura, Pecuária e Abastecimento.

MIR - Infravermelho Médio (*Mid Infrared*).

MSC - Correção Multiplicativa de Espalhamento (*Multiplicative Scatter Correction*).

NIR - Infravermelho próximo (*Near Infrared*).

NIRS - Espectroscopia no Infravermelho Próximo (*Near Infrared Spectroscopy*).

OC-PLS - Mínimos Quadrados Parciais de uma Classe (*One-class Partial Least Squares*).

OD - Distância ortogonal (*Orthogonal Distance*).

ODS - Objetivos de Desenvolvimento Sustentável.

OFF - Correção de linha de base Offset (*Baseline correction Offset*).

ONU - Organização das Nações Unidas.

PC - Componentes Principais (*Principal Component*).

PCA - Análise de Componentes Principais (*Principal Component Analysis*).

- PCV** - Validação Cruzada Procrustes (*Procrustes Cross Validation*).
- PLS** - Mínimos Quadrados Parciais (*Partial Least Squares*).
- PLSR** - Regressão por Mínimos Quadrados Parciais (*Partial Least Squares Regression*).
- PTFE** - Politetrafluoretuleno.
- PV** - Pseudo Validação.
- SD** - Distância dos Escores (*Scores Distance*).
- SEN** - Sensibilidade (*Sensibility*).
- SPC** - Especificidade (*Specificity*).
- SG** - Savitzky Golay.
- SGD** - Derivada Savitzky Golay (*Savitzky Golay Derivative*).
- SGS** - Suavização por filtro Savitzky Golay (*Savitzky Golay Smoothing*).
- SNV** - Variação Normal Padrão (*Standard Normal Variate*).
- SPA** - Algoritmo das Projeções Sucessivas (*Successive Projections Algorithm*).
- TN** - Verdadeiros Negativos (*True Negatives*).
- TP** - Verdadeiros Positivos (*True Positives*).

RESUMO

O leite de cabra é considerado um alimento seguro, capaz de produzir efeitos metabólicos e/ou fisiológicos, que podem ser benéficos à saúde. Atualmente, o consumo mundial de leite de cabra e seus derivados têm aumentado devido ao fato deste alimento apresentar características singulares quando comparado ao leite de vaca, tais com baixa alergenicidade, alta digestibilidade e propriedades funcionais. Do ponto de vista econômico, o leite de cabra possui maior valor comercial do que o leite de vaca e diversos estudos apontam que a adulteração por adição de leite de vaca ao leite caprino é frequente. Nesse cenário, políticas de controle de qualidade de alimentos são importantes para garantir a ingestão segura de alimentos e impedir prejuízos financeiros e à saúde dos consumidores. Portanto, neste estudo objetivou-se desenvolver de novas metodologias analíticas baseadas em NIR, RAMAN e ferramentas quimiométricas para o controle de qualidade de leite de cabra, a fim de obter análises rápidas, não invasivas e não destrutivas, alinhadas com os Princípios da Química Verde. No primeiro experimento foram adquiridos espectros NIR de um total de 348 amostras, objetivando classificar o leite de cabra puro, quando a presença dos adulterantes: amido, extrato de soja, composto lácteo e leite bovino. Considerando os dados NIR, a correção de linha de base offset em conjunto com suavização Savitzky-Golay com 21 pontos levaram aos melhores resultados para os modelos construídos (DD-SIMCA e OC-PLS). A autenticação do leite caprino em pó usando estes dados resultou em valores de sensibilidade para os conjuntos de treinamento e teste de 97,5, 100%, não obstante da especificidade no teste de 83,9% utilizando DD-SIMCA, já para os modelos usando OC-PLS os valores obtidos foram 97,5, 100 e 95,1%, respectivamente, para as mesmas figuras de mérito avaliadas. Em outra mão, o segundo experimento seguiu com a aquisição de espectros RAMAN de 83 amostras, cujas 67 adulteradas apenas com leite bovino, também com o intuito de classificar a matriz quanto à adulteração. Os dados RAMAN, no qual foram utilizadas 3 estratégias diferentes de validação, os resultados para os modelos DD-SIMCA em termos da sensibilidade e especificidade foram todos 100%, exceto para a terceira abordagem com o algoritmo KS que apresentou um valor de 98%, para o conjunto teste. Em todas as abordagens, os melhores resultados foram obtidos utilizando os pré-processamentos primeira derivada de Savitzky-Golay com polinômio de segunda ordem e janela de 25 pontos, e MSC. Como consequência, ao autenticar o leite em pó quanto a presença de adulterantes fraudulentos, os produtores podem garantir que os consumidores estejam desfrutando de um produto seguro e de qualidade. Assim, a metodologia proposta pode contribuir para a segurança do consumidor e garantir a conformidade regulatória.

Palavras-chave: Leite de cabra, Adulteração, Química Analítica Verde, Espectroscopia NIR, Espectroscopia RAMAN, Quimiometria.

ABSTRACT

Goat milk is considered a safe food, capable of producing metabolic and/or physiological effects that can be beneficial to health. Currently, the global consumption of goat milk and its derivatives has been increasing due to the fact that this food presents unique characteristics when compared to cow milk, such as low allergenicity, high digestibility, and functional properties. From an economic point of view, goat milk has a higher commercial value than cow milk, and several studies point out that adulteration by adding cow milk to goat milk is frequent. In this context, food quality control policies are important to ensure the safe consumption of food and prevent financial losses and harm to consumers' health. Therefore, this study aimed to develop new analytical methodologies based on NIR, RAMAN, and chemometric tools for the quality control of goat milk, in order to obtain fast, non-invasive, and non-destructive analyses, aligned with the principles of Green Chemistry. In the first experiment, NIR spectra were acquired from a total of 348 samples, aiming to classify pure goat milk when adulterants such as starch, soy extract, dairy compound, and cow milk were present. Considering the NIR data, baseline offset correction combined with Savitzky-Golay smoothing with 21 points yielded the best results for the models built (DD-SIMCA and OC-PLS). The authentication of powdered goat milk using these data resulted in sensitivity values for the training and test sets of 97.5% and 100%, respectively, despite the specificity in the test being 83.9% using DD-SIMCA. For the models using OC-PLS, the values obtained were 97.5%, 100%, and 95.1%, respectively, for the same performance metrics evaluated. On the other hand, the second experiment proceeded with the acquisition of RAMAN spectra from 83 samples, of which 67 were adulterated only with cow milk, also aiming to classify the matrix regarding adulteration. The RAMAN data, in which three different validation strategies were used, showed that the results for the DD-SIMCA models in terms of sensitivity and specificity were all 100%, except for the third approach with the KS algorithm, which showed a value of 98% for the test set. In all approaches, the best results were obtained using the Savitzky-Golay first derivative with a second-order polynomial and a window of 25 points, along with MSC. As a result, by authenticating powdered milk for the presence of fraudulent adulterants, producers can ensure that consumers are enjoying a safe and high-quality product. Thus, the proposed methodology can contribute to consumer safety and ensure regulatory compliance.

Keywords: Goat milk, Adulteration, Green Analytical Chemistry, NIR Spectroscopy, RAMAN Spectroscopy, Chemometrics.

Capítulo 1

Introdução

1 INTRODUÇÃO

O leite de cabra é um alimento que proporciona diversos benefícios para a saúde humana, pois é rico em elementos essenciais como proteínas, carboidratos, gorduras, vitaminas e minerais. Além disso, possui características que o tornam hipoalergênico e de fácil digestão da gordura, promovendo assim melhorias nas funções fisiológicas e nutricionais, especialmente para crianças e idosos. Essas qualidades tornam o leite de cabra uma opção mais adequada para pessoas com alergia a outros tipos de leite (Souza et al., 2019). Esta matriz também se destaca pela versatilidade na produção de uma ampla gama de produtos derivados, no entanto, apenas o leite de cabra de excelente qualidade pode ser utilizado na fabricação de produtos de alto valor agregado. O leite de qualidade deve possuir atributos nutricionais, sanitários e sensoriais que atendam às expectativas dos consumidores para garantir a produção de produtos satisfatórios (Ribeiro, 2008; Delgado Júnior; Siqueira; Stock, 2020).

O controle de qualidade alimentar é essencial para garantir a segurança, a qualidade e a conformidade dos alimentos. Ele abrange a prevenção de contaminantes que possam representar riscos à saúde dos consumidores, assegura que os alimentos atendam aos padrões nutricionais necessários e mantenham uma qualidade sensorial adequada. Ademais, o controle de qualidade alimentar promove a conformidade com regulamentações, previne fraudes, otimiza processos de produção e contribui para a redução do desperdício. Ao garantir esses aspectos, ele protege a saúde pública, fortalece a reputação das marcas e incentiva a inovação na indústria alimentícia, promovendo um ambiente de consumo seguro e confiável para todos (Ménard et al., 2024).

No setor alimentício, a prática de adulterar produtos para aumentar o rendimento e maximizar lucros é uma realidade tanto no Brasil quanto internacionalmente, mesmo com regulamentações rigorosas, certificações de qualidade e inspeções sistemáticas. No contexto do leite líquido, a adição de água para aumentar o volume do produto distribuído tem sido historicamente uma das fraudes mais comuns, contudo, essa não é a única forma de adulteração do leite com o objetivo de obter benefícios econômicos; práticas como a adição de reconstituintes, conservantes e neutralizantes também são frequentes (Uberti & Pinto, 2022; Ménard et al., 2024). É importante destacar que, no caso de derivados como pó, queijos, requeijões, iogurtes, entre outros, as características específicas para cada tipo de derivado dificultam a detecção de adulterações através de métodos mais grosseiros, como a

inspeção visual. Por exemplo, é desafiador diferenciar visualmente amostras de leite em pó caprino do bovino, o que seria possível apenas com uso de métodos padrão de análise, como um ensaio químico enzimático específico, ou instrumentalmente. Logo, o desenvolvimento e implementação de novos métodos de detecção mais sensíveis e conclusivos são cruciais para assegurar a autenticidade desses produtos (Dias *et al.*, 2020; Panciere; Ribeiro, 2021; Ménard *et al.*, 2024).

Neste cenário, diversas metodologias têm sido desenvolvidas para identificar e quantificar adulteração em leite de cabra e seus produtos derivados. Isso é necessário devido aos impactos econômicos e, principalmente, aos potenciais riscos à saúde do consumidor, especialmente considerando a natureza tóxica de certos contaminantes (Moore; Spink; Lipp, 2012; Kamal; Karoui, 2015; Pereira *et al.*, 2020). Isto posto, muitas das metodologias empregadas para assegurar a autenticidade do leite de cabra têm sido fundamentadas em técnicas convencionais (ensaios físico-químicos), sensoriais, instrumentais (como cromatografia eletroforética) e utilizando ferramentas da biologia molecular (ensaios imunológicos) (Song; Xue; Han, 2011; Pesic *et al.*, 2011; Kamal; Karoui, 2015; Vieitez *et al.*, 2016; Di Pinto *et al.*, 2017). Entretanto, essas metodologias requerem o uso de equipamentos e/ou reagentes de alto custo tanto na aquisição quanto na manutenção, demandam mais tempo de análise, além de consumir mais reagentes e gerar mais resíduos prejudiciais ao meio ambiente, contrariando tanto os 12 princípios da Química Verde, quanto os 17 Objetivos de Desenvolvimento Sustentável (ODS) recomendados pela ONU.

Neste contexto, o objetivo deste trabalho foi desenvolver uma nova metodologia analítica rápida, não invasiva e não destrutiva baseada nas espectroscopias NIR e RAMAN associadas à ferramentas quimiométricas para detecção de adulteração em leite caprino em pó industrializado. É esperado que o emprego dessa metodologia promova uma redução significativa da geração de resíduos ambientais, minimizando ou eliminando a necessidade de uso de reagentes químicos e de pré-tratamento das amostras, alinhando-se aos Princípios da Química Verde.

Capítulo 2

Objetivos

2 OBJETIVOS

2.1 OBJETIVO GERAL

Desenvolver uma metodologia analítica simples, rápida e não destrutiva baseada na obtenção de *fingerprints* espectrais do NIR e do RAMAN, associadas à ferramentas quimiométricas para detecção de adulteração em amostras de leite caprino em pó industrializado.

2.2 OBJETIVOS ESPECÍFICOS

- Realizar análises exploratórias dos *fingerprints* das amostras, em termos espectrais e por meio dos escores e pesos (PCA);
- Construir modelos de autenticação para o leite de cabra em pó utilizando as ferramentas DD-SIMCA e OC-PLS;
- Empregar diferentes estratégias para validação dos modelos construídos a partir de uma pequena quantidade de amostras;
- Avaliar o desempenho dos modelos de classificação construídos por meio de diversos tipos de pré -processamentos;
- Avaliar o desempenho dos modelos de classificação construídos utilizando as figuras de mérito como ferramenta de análise.

Capítulo 3

Fundamentação teórica

3 FUNDAMENTAÇÃO TEÓRICA

3.1 LEITE

O leite de várias espécies animais têm historicamente desempenhado um papel significativo, e em algumas áreas, cruciais, como na alimentação humana (Chen *et al.*, 2021). O leite é um dos alimentos mais consumidos globalmente, e seu consumo, tanto em termos de qualidade quanto de quantidade, está influenciado por diversos fatores, como a dieta do consumidor, questões de saúde, nível de renda, disponibilidade e aspectos culturais. Em países em desenvolvimento, por exemplo, o consumo de leite costuma ser *in natura*, pois o processamento do leite requer infraestrutura e investimentos, recursos financeiros e técnicos que frequentemente são limitados nesses países (Scaquette, 2021). Apesar do considerável aumento na disponibilidade e oferta de alimentos nas últimas décadas nos países desenvolvidos, o leite e seus derivados permanecem como componentes essenciais da dieta em nações menos desenvolvidas ou em desenvolvimento, no qual são vitais para a sobrevivência humana.

Composto principalmente por água, lipídios, proteínas, lactose, vitaminas e minerais, as quantidades variam conforme a espécie do animal e o período de lactação. Adicionalmente, fatores genéticos, fisiológicos, nutricionais e ambientais exercem uma significativa influência nas variações observadas (Pietrzak-Fiećko; Kamelska-Sadowska, 2020; Suh, 2022). Com odor e sabor distintos, o leite tem na gordura um dos principais componentes, sendo um importante indicador de sua qualidade (Gu *et al.*, 2021; Silva, 2023).

Os lipídios constituem cerca de 3 a 5% da composição do leite e são responsáveis por conferir características sensoriais ao alimento e seus derivados (Clarke *et al.*, 2019; Liu *et al.*, 2020). Já as proteínas presentes no leite contém aminoácidos essenciais que são indispensáveis e benéficos para a saúde humana, além de estarem associados às características organolépticas deste (Zhang *et al.*, 2022).

O leite pode ser ofertado de diferentes formas: líquida, em pó, integral, desnatada, pasteurizada e ultraprocessada. O que abre portas para a produção e comercialização de uma variedade de seus derivados, como queijos, coalhada, iogurte, manteigas, sorvetes, leite condensado, creme de leite, bebidas fermentadas e não fermentadas. Os produtos derivados do leite oferecem uma ampla gama de benefícios aos consumidores, desempenhando papéis essenciais no crescimento, desenvolvimento e imunidade humanos, não obstante do papel

econômico agregando valor de mercado em seus produtos mais refinados (Gu, *et al.*, 2021; Liu; Cai; Zhang, 2021).

Preservar as propriedades originais do leite cru é fundamental para assegurar a segurança, higiene e qualidade dos seus derivados (Zhiqiang, 2021). Tratando-se de um alimento tão importante, o cuidado é tão crucial quanto, pois um fornecimento adequado e de alta qualidade é fundamental para o desenvolvimento da indústria leiteira moderna. Para tanto, garantir um produto de qualidade se faz necessário ter o controle processual para cada atividade a ser executada.

3.1.1 Leite caprino

A caprinocultura foi iniciada no Brasil pelos colonizadores portugueses em torno de 1535, quando também foram introduzidos os primeiros animais domésticos. Até o início dos anos 1970, o rebanho caprino brasileiro consistia predominantemente de animais sem raça definida. Em 1975, ocorreu a primeira importação de caprinos leiteiros para o país. Na década de 80, grandes empresários ingressaram no setor, realizando investimentos significativos que impulsionaram seu desenvolvimento (Maia, 1994; Delgado Júnior; Siqueira; Stock, 2020). Durante os anos 1990, programas estaduais foram estabelecidos para a compra e distribuição de leite de cabra a crianças carentes no Rio Grande do Norte e na Paraíba (Fonseca; Bruschi, 2009; Delgado Junior, Siqueira e Stock 2020).

A cabra é valorizada por sua capacidade de fornecer carne, pele e, especialmente, leite, além de diversos produtos lácteos. Apesar disso, elas representam aproximadamente 2% da produção global de leite. Contudo, estudos indicam um aumento no consumo até 2030 (Foroutan *et al.*, 2019; Nayik *et al.*, 2021; Namonje-Kapembwa; Chiwawa; Sitko, 2022). Tal aumento pode estar associado a características peculiares, como a capacidade de atingir a maturidade precoce, fácil adaptação a diferentes climas e dietas, e sua lucratividade (Namonje-Kapembwa; Chiwawa; Sitko, 2022).

No Brasil, a caprinocultura leiteira tem se destacado como uma atividade econômica em expansão, sendo amplamente praticada na região nordeste, abrigando cerca de 84,56% do rebanho nacional de cabras ordenhadas. Esta região lidera a produção de leite de cabra no país, seguida pelas regiões sudeste, sul, centro-oeste e norte (IBGE, 2021a; IBGE, 2021b; Reis, 2023). A produção do leite de cabra apontou um crescimento significativo de 52% entre 2000 e 2017, superando o aumento de 35% do leite de vaca durante o mesmo período (Reis, 2023). Contudo, apesar do crescimento significativo registrado, ainda existe um amplo

potencial a ser explorado no setor, tanto no aumento da produção quanto na melhoria da comercialização.

No que diz respeito às formas de beneficiamento e classificação (**Quadro 1**) do leite de cabra quanto ao teor de gordura, a Instrução Normativa 37 estabelece que:

6.2. Durante o beneficiamento do leite de cabra, poderão ser realizadas operações de acerto do teor de gordura através de equipamento acoplado ao circuito de beneficiamento, visando a oferta ao consumo de produtos padronizados, desnatados ou semi-desnatados, segundo parâmetros de qualidade fixados neste Regulamento (Brasil, 2000, p.5).

Quadro 1: Classificação do leite de cabra quanto ao teor de gordura.

Classificação	Parâmetros
Integral	Quando não houver qualquer alteração do teor de gordura contido na matéria-prima
Padronizado	Quando o teor de gordura, expresso em % m/m, for acertado para 3%
Semidesnatado	Quando o teor de gordura, expresso em % m/m, for acertado para o intervalo de 0,6 e 2,9%
Desnatado	Quando o teor de gordura, expresso em % m/m, não superar o limite máximo 0,5%

Fonte: adaptado de Brasil, 2000.

Para além dos critérios de classificação, se o leite for aromatizado ou se destinar ao comércio na forma congelada, é obrigatória a inclusão das expressões "aromatizado" ou "congelado" no final da denominação de venda do produto (Brasil, 2000). A versatilidade do leite de cabra, faz com que o mesmo também seja utilizado na fabricação de uma ampla gama de produtos. No entanto, apenas o leite de cabra de excelente qualidade pode ser transformado em produtos de alto valor agregado que atendam às expectativas dos consumidores em termos de atributos nutricionais, sanitários e sensoriais (Delgado Junior; Stock; Siqueira, 2020; Silva, 2023).

Quando comparado com leite bovino, o mesmo agrega um maior valor biológico, especialmente por conter lipídios compostos por partículas pequenas e médias, contendo um maior número de triacilgliceróis insaturados, menor peso molecular e glicoproteínas que contribuem para uma baixa alergenicidade, fácil digestão e absorção (Liu; Cai; Zhang, 2021; Oliveira *et al.*, 2021; Zhao; Wang; Mao, 2022; Silva, 2023). Os triacilgliceróis constituem a

estrutura fundamental dos glóbulos de gordura do leite de cabra, sendo a principal fonte de energia para o desenvolvimento físico natural humano (Yuan *et al.*, 2021; Zhang; Jia; Shi, 2023).

O leite de cabra um promissor substituto na produção de fórmulas infantis, oferecendo diversos benefícios à saúde tanto quanto o leite bovino. Além disso, contém componentes que podem contribuir para a proteção contra infecções, como hidrolisados de caseína que melhoram a resistência à insulina, antioxidantes, entre outros (Cakir; Tunali-akbay, 2021; Oliveira *et al.*, 2021).

3.1.2 Leite em pó

O leite em pó começou a ser produzido com o intuito de prolongar a durabilidade do produto nas prateleiras, reduzir os custos com transporte e armazenamento devido à diminuição do volume, e fornecer um produto com maior concentração de sólidos lácteos. Obtido pela desidratação dos leites em geral, sendo integral, desnatado ou parcialmente desnatado, o leite em pó, é utilizado como ingrediente por indústrias produtoras de sorvetes, confeitaria, fórmulas infantis e outros alimentos, mediante processos tecnológicos apropriados adequando-o para consumo humano (Cruz *et al.*, 2019). Vale salientar que exceto as alterações decorrentes do processo tecnológico, a matriz supramencionada deve manter as proporções relativas de proteínas, açúcares, gorduras e outras substâncias minerais (Brasil, 2018).

Em relação ao teor de gordura, comercialmente o leite em pó é categorizado da seguinte forma: aquele com teor mínimo de 26% é designado como leite em pó integral; teor entre 1,5% e 26% é considerado leite em pó parcialmente desnatado; teor máximo de 1,5% caracteriza o leite em pó desnatado; e teor entre 14% e 16% define o leite em pó semidesnatado. Quanto à umectabilidade e dispersabilidade, o leite em pó pode ser classificado como instantâneo ou não. O processamento do leite em pó inclui várias etapas: recebimento da matéria-prima, resfriamento, padronização, pasteurização, concentração, secagem e embalagem (Brasil, 2018; Ruis, 2022).

Para ser aceito no estabelecimento, o leite cru deve estar refrigerado a 7,0°C, podendo ser aceito excepcionalmente até 9,0°C. Após o recebimento, o produto deve ser armazenado sob refrigeração a 4,0°C e passar por testes de controle de qualidade para avaliar se atende aos parâmetros físico-químicos, microbiológicos e sensoriais necessários. Atendendo aos parâmetros, o leite é filtrado, a fim de eliminar as impurezas maiores, e armazenado em

tanques de resfriamento até o processo de clarificação, na qual o leite é aquecido a 45°C para facilitar a remoção de partículas de impurezas menores, que são separadas por meio de centrifugação (Clarificador). Já na fase de padronização, o intuito principal é garantir que o produto tenha uma composição química específica, reintegrando-o na quantidade desejada e garantindo uniformidade (Brasil, 2018; Ruis, 2022).

A etapa de pasteurização visa eliminar as formas vegetativas de microrganismos patogênicos e inativar enzimas. Geralmente, são utilizadas três condições de aquecimento com variações no tempo e temperatura: baixa temperatura (74°C por 30 segundos), média temperatura (80 a 100°C por 1 a 2 minutos) e alta temperatura (120 a 135°C por 1 a 2 minutos). Após isso, o leite é concentrado até alcançar 50% dos sólidos totais. Em seguida, ocorre a remoção parcial da água, retirando até 85% da umidade do leite, etapa de concentração, que ajuda a reduzir os custos na fase subsequente de secagem, onde os 15% restantes serão eliminados (Fernandes, 2009; Cruz *et al.*, 2019; Ruis, 2022).

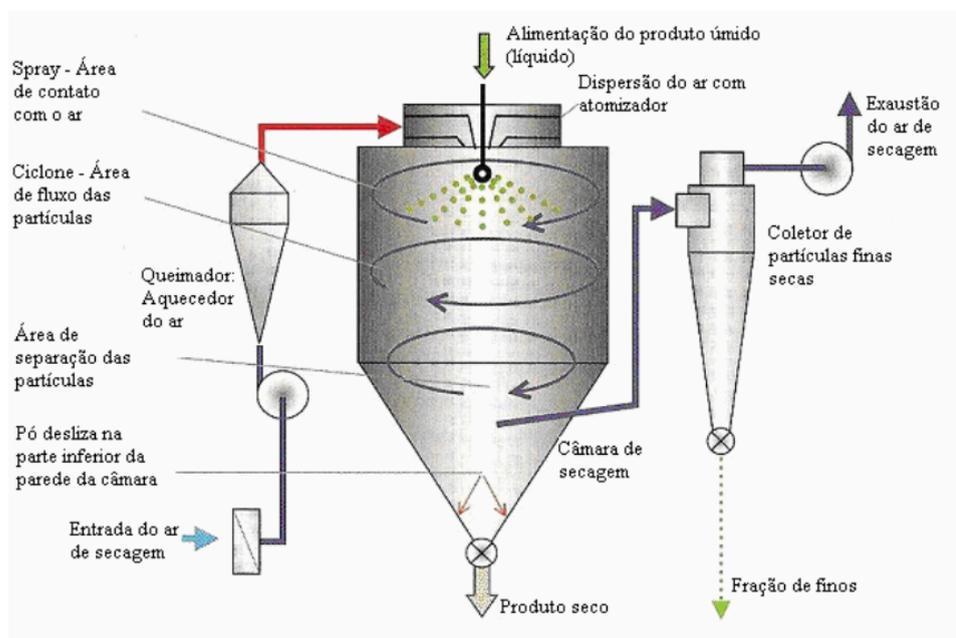
A evaporação realizada durante a etapa de concentração, pode utilizar evaporadores operando sob vácuo, visando evitar a desnaturação das proteínas, ou evaporadores com trocadores de calor posicionados verticalmente, que separam o líquido do vapor formado por centrifugação. Outrossim, o emprego da técnica de filtração por membrana, que separa as partículas com base no tamanho, pode ser uma opção viável (Carié *et al.*, 2009; Ruis, 2022).

Sequencialmente, com o intuito de produzir um produto estável, com baixa umidade e mínimas mudanças sensoriais, chegamos ao processo de secagem, este envolve a remoção da água por meio do calor e pode ser realizado por dois métodos distintos: através de rolos ou *spray dryer*. No método com rolos, utiliza-se um ou mais cilindros rotativos aquecidos internamente com vapor a temperaturas entre 130°C e 150°C, que secam o leite líquido na superfície do rolo durante a rotação. A camada seca é então removida e moída para formar um pó fino, que é posteriormente embalado (Fernandes, 2009; Ruis, 2022).

De maneira análoga, o método *spray dryer* (Figura 1), atua como método de secagem do leite. Atualmente, essa técnica é amplamente adotada devido à sua eficiência e economia, ao uso de equipamentos simples e à capacidade de produção em larga escala, além de provocar menos modificações no produto final (Cruz *et al.*, 2019). A palavra inglês *spray* remete à pulverizar, enquanto a palavra *dryer* advém de secador, logo a grosso modo, o nome da técnica representa seu fundamento, ou seja, secar ao pulverizar a matriz, e assim transformar líquidos em pó.

O leite é pulverizado dentro de uma câmara onde há um fluxo controlado de ar quente. A temperatura do ar na câmara varia de 180 a 220°C e, ao entrar em contato com as gotículas de leite, provoca a rápida evaporação da água, resultando na formação de pó (Anandharamkrishnan et al., 2015; Ruis, 2022). A escolha da temperatura depende da natureza do líquido a ser atomizado, da velocidade desejada de secagem e das propriedades finais que se deseja obter no produto em pó, garantindo que a secagem não afete o produto. As partículas secas são então coletadas na parte inferior da torre de secagem, onde são separadas do ar por métodos como ciclones ou filtros.

Figura 1. Processo de Spray-dryer.



Fonte: Madeira, 2009.

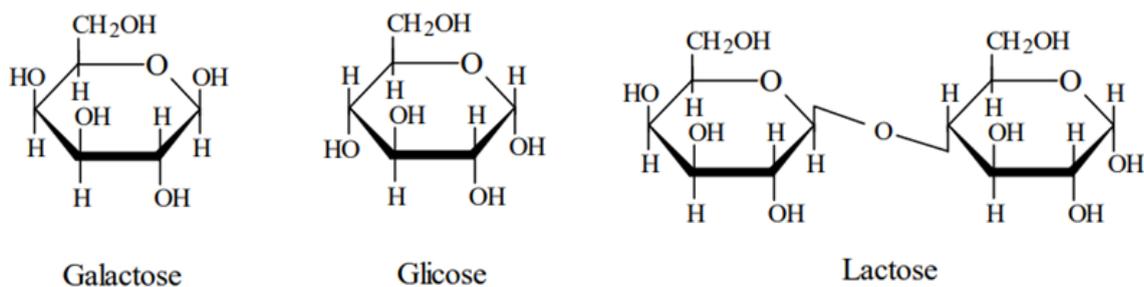
Após o processamento, o leite em pó é envasado em embalagens estéreis para garantir sua integridade. Essas embalagens atuam como uma barreira física contra oxigênio e luz, preservando as propriedades do alimento. O leite em pó desnatado pode ser armazenado por até três anos, enquanto o leite integral tem uma validade máxima de seis meses. Isso se deve à oxidação da gordura presente no pó durante o armazenamento, que pode resultar em deterioração e alteração do sabor (Ruis, 2022). A Figura 2 apresenta um esquema representativo do percurso do leite durante a produção do leite em pó:

Figura 2. Esquema representativo do percurso do leite durante a produção do leite em pó.

Fonte: Autoria própria, 2024.

3.1.3 Leite caprino x Leite bovino

O leite de cabra apresenta quantidades de nutrientes semelhantes ao leite de vaca, contudo com algumas diferenças. Por exemplo, existe uma diferença em valores percentuais no teor médio aproximado de lactose entre o leite caprino (4,5%) e o bovino (4,7%). A lactose (**Figura 3**) é o principal carboidrato presente no leite dos mamíferos e em seus derivados. É um dissacarídeo, composto por glicose e galactose, que desempenha um papel crucial na absorção de diversos sais minerais, como cálcio, magnésio e zinco. Esse açúcar é fundamental em todos os processos tecnológicos que envolvem a acidificação do leite, como na produção de iogurtes e queijos maturados. Além disso, a lactose desempenha um papel crucial na síntese do leite, atuando como fator osmótico ao atrair água para as células epiteliais da glândula mamária, controlando assim o volume de leite produzido (Cruz *et al.*, 2016; Delgado Junior, Siqueira; Stock, 2020; Barbosa, 2023).

Figura 3. Estruturas da galactose, glicose e lactose

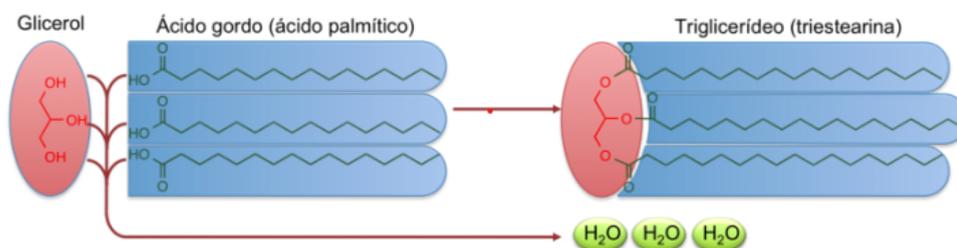
Fonte: adaptada de Carminatti, 2001.

No caso das proteínas, o teor total de proteína no leite de cabra geralmente é inferior ao do leite de vaca, sendo aproximadamente 3,1% e 3,4%, respectivamente. As proteínas do leite incluem caseínas, proteínas solúveis como proteases, peptonas, albumina sérica e imunoglobulinas. No entanto, as imunoglobulinas estão presentes em baixas concentrações e têm pouca influência na produção de derivados lácteos. As caseínas são as proteínas coaguláveis do leite, compostas por frações de: a) α S caseína (α S1 e α S2), b) β -caseína, c) κ -caseína e d) γ -caseína. É importante ressaltar que a caseína é crucial na formação dos queijos e desempenha um papel essencial na estabilidade térmica do leite, contudo são principais agentes alergênicos, cabendo destaque para a α S1, que em diversos estudos é apontada como principal causadora (Verruck *et al.*, 2019; Delgado Junior, Siqueira; Stock, 2020; Do Egito *et al.*, 2023).

As micelas de caseína no leite de cabra são principalmente compostas pela β -caseína, que representa 55% do total de caseínas, enquanto no leite de vaca esse percentual não chega a 40%. Os leites apresentam valores próximos para quantidades de α S2-caseína e κ -caseína, porém quanto a fração α S1-caseína é possível observar uma discrepância significativa no valor para o leite caprino, que é muito menor em comparação ao leite de vaca (Delgado Junior, Siqueira; Stock, 2020; Roy *et al.*, 2020; Souza, 2024). Segundo estudos, essa quantidade reduzida pode estar relacionada ao menor potencial alergênico do leite de cabra, tornando-o mais facilmente tolerável para crianças do que o leite de vaca (Pimenta *et al.*, 2021; Neto *et al.*, 2021).

Um outro componente presente em ambos os leites é a gordura. A gordura do leite é o componente mais variável em sua concentração, requerendo uma interpretação cuidadosa. Cerca de 98% da gordura do leite é composta por triglicerídeos, enquanto os 2% restantes são compostos por fosfolipídios, esteróis, vitaminas lipossolúveis (A, D, E e K) e pequenas quantidades de ácidos graxos livres (Delgado Junior, Siqueira; Stock, 2020; Ramos Neto *et al.*, 2021). Os triglicerídeos (**Figura 4**) são compostos por três ácidos graxos (iguais ou diferentes) esterificados em uma molécula de glicerol.

Figura 4. Ilustração da estrutura química de um triglicerídeo.



Fonte: Adaptado de Oliveira, 2015.

Esses triglicerídeos são dispersos no leite na forma de glóbulos de gordura de diferentes tamanhos, o que influencia significativamente diversos processos tecnológicos, como o desnatado do leite para a obtenção de creme e manteiga. Os ácidos graxos são moléculas compostas por uma cadeia longa de átomos de carbono, com um grupo carboxila (-COOH) em uma extremidade. Esses compostos são classificados por três grupos, conforme ilustra o **Quadro 2**.

Quadro 2. Tipos de Ácidos graxos e suas principais características.

Tipos de Ácidos Graxos	Principais características
Cadeia curta	Possuem até 8 átomos de carbono
Cadeia média	Possuem de 10 a 14 átomos de carbono
Cadeia longa	Possuem mais de 16 átomos de carbono

Fonte: adaptado de Cruz *et al.*, 2016.

A gordura desempenha um papel crucial no sabor dos alimentos e é fundamental para definir características de textura e consistência. No leite de cabra, o teor de gordura é aproximadamente 4,1%, enquanto no leite de vaca a média é de 3,6%, mais uma diferença entre os valores percentuais. No entanto, a maior diferença está na composição dos ácidos graxos que compõem os triglicerídeos, os quais exercem uma influência significativa sobre o sabor do produto (Cruz *et al.*, 2016; Delgado Junior; Siqueira; Stock, 2020; Silva, 2023).

O leite de cabra contém gorduras compostas principalmente por ácidos graxos de cadeias médias e curtas. Os glóbulos de gordura são menores em comparação com o leite de vaca, o que facilita uma absorção mais rápida pelo organismo humano, resultando em maior

digestibilidade (Aragão *et al.*, 2022). Isso reduz a quantidade de resíduos no intestino, ajudando a evitar fermentação, má digestão, constipação, entre outros problemas. Em contrapartida, esses ácidos são os principais responsáveis pelo sabor e aroma característicos dos produtos derivados do leite caprino, o que segundo estudos, talvez esses elementos, como sabor e odor forte sejam parte das razões pelas quais alguns consumidores rejeitam o leite de cabra (Cruz *et al.*, 2016; Delgado Junior; Siqueira; Stock, 2020). A **Tabela 1** apresenta a comparação dos leites bovino e caprino em termos de macronutrientes. Os valores apresentados são estimados por médias, podendo variar em função de raça, alimentação, períodos de lactação, entre outros fatores:

Tabela 1. Percentuais dos componentes: Leite de cabra x Leite de Vaca

Parâmetro	Leite de Cabra	Leite de Vaca
Gordura (%)	4,1	3,6
Lactose (%)	4,5	4,7
Proteínas (%)	3,5	3,4
Caseína (%)	2,4	3,0
α S1 caseína (%)	5,6	40,9
α S2 caseína (%)	19,2	10,7
β -caseína (%)	54,8	36,3
κ -caseína (%)	20,4	11,9

Fonte: adaptada de Park *et al.*, 2007; Amigo; Fontecha, 2011.

Por fim, mas não menos importante, as vitaminas e minerais. No leite de cabra os minerais representam uma pequena quantidade quando comparados com os outros componentes, contudo como todos os outros já mencionados, os teores podem variar por diferentes fatores como raça, período de lactação, estação do ano, composição da dieta, entre outros. O cálcio (Ca) e o fósforo (P) são abundantemente encontrados no leite e desempenham um papel crucial na estrutura das micelas de caseína. Além desses minerais, o leite também contém sódio, potássio e cloro, que são essenciais para regular o volume de água no leite e complementar os efeitos da lactose. O leite de cabra também pode conter minerais como magnésio, ferro, cobre e zinco. (Delgado Junior; Siqueira; Stock, 2020; Aragão, 2022; Teixeira, 2022).

Para as vitaminas, temos, mesmo em pequenas quantidades, a presença das vitaminas: A, B1, B2, B6, B12, D, E e K. Vale salientar que quando comparado ao leite de vaca, a vitamina A é a que apresenta maior diferença a favor do leite de cabra. Segundo Pereira *et al.* (2022), o leite de vaca mais próximo da maturação apresenta uma tonalidade levemente

amarelada devido à presença de carotenoides, em contraste, ao leite de cabra que geralmente não contém β -caroteno, mas contém mais retinol (vitamina A). Dessa forma, a cor do leite de cabra é um fator a ser evidenciado, por apresentar uma coloração branca característica devido à baixa concentração de betacaroteno, diferente do leite de vaca que é amarelo e apresenta uma elevada quantidade de betacaroteno e xantofilas.

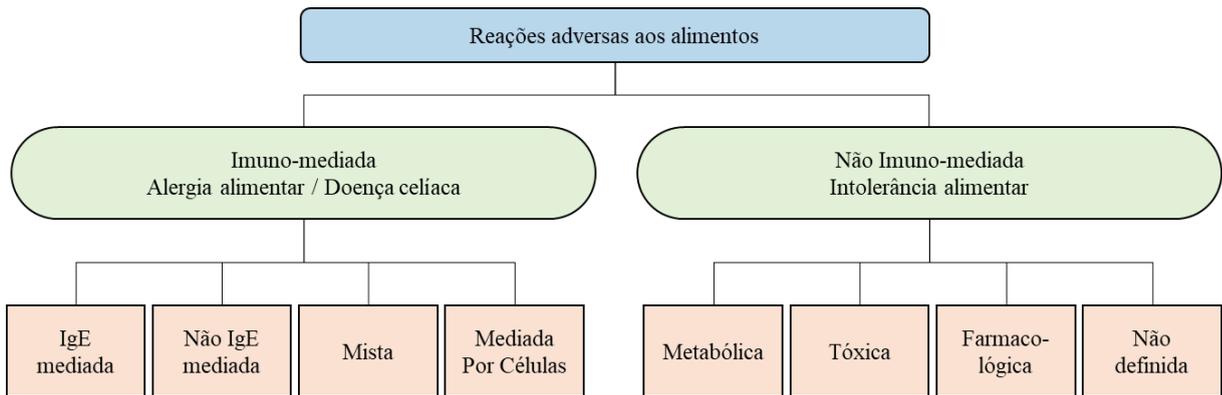
Em resumo, os leites desses animais possuem características físico-químicas semelhantes, como viscosidade e odor característicos. No entanto, distinguem-se em alguns pontos como a diferença no tamanho dos glóbulos de gordura e nos teores da proteína caseína, especialmente da porção α S1. Então como já mencionado, essas características conferem ao leite de cabra vantagens significativas para o consumo, destacando a maior digestibilidade e menor propensão a causar alergias em comparação ao leite de vaca.

3.1.4 Alergia à Proteína do Leite de Vaca (APLV)

A alergia à proteína do leite de vaca (APLV) é definida como uma reação adversa mediada pelo sistema imunológico às proteínas presentes no leite de vaca, sendo uma das alergias alimentares mais comuns na infância, atingindo cerca de 2% a 3% das crianças com menos de três anos. Em geral, alimentos como leite de vaca, ovo e trigo são frequentemente associados à alergia alimentar em crianças, enquanto amendoim, nozes e frutos do mar são mais comuns entre adultos (Fiquene *et al.*, 2023).

Este distúrbio imunológico apresenta uma diversidade de sintomas e sinais que frequentemente surgem em bebês e podem diminuir até os 6 anos de idade, ocorrendo quando o sistema imunológico reage de maneira anormal às proteínas presentes no leite de vaca, podendo resultar em sintomas variados que afetam o trato gastrointestinal, a pele e o sistema respiratório. Então, se não tratada de forma adequada, pode resultar em deficiências nutricionais subsequentes, logo se faz necessário uma dieta livre de leite, o que pode ser estressante para pais e familiares (Edwards; Younus, 2019).

Nesse contexto, cabe destacar um outro tipo de reação corporal adversa relacionada à ingestão de leite que é a intolerância à lactose. É crucial definir a APLV para distingui-la claramente da Intolerância à Lactose, pois certa confusão atrapalha não apenas os cuidadores de pacientes pediátricos, mas também dentro do nosso meio acadêmico e profissional causando riscos à saúde dos indivíduos (Arruda *et al.*, 2023). A **Figura 5** ilustra um esquema de como são classificadas as reações adversas aos alimentos.

Figura 5. Classificação das reações adversas aos alimentos.

Fonte: Autoria própria, 2024.

A intolerância à lactose ocorre devido a uma capacidade diminuída de digerir o açúcar lactose. Por outro lado, a Alergia à Proteína do Leite de Vaca é uma resposta alérgica que pode ser classificada como mediada por Imunoglobulina E (IgE), não mediada por IgE ou mista. Desse modo, crianças com intolerância à lactose têm dificuldade em digerir a lactose, mas normalmente toleram as proteínas do leite de vaca, contudo crianças com APLV têm reações alérgicas às proteínas do leite de vaca, mas geralmente podem tolerar a lactose (Di Costanzo *et al.*, 2021; Arruda *et al.*, 2023).

Segundo Di Costanzo *et al* (2021), a reação alérgica mediada por IgE ocorre quando o organismo se sensibiliza a proteínas alimentares específicas, levando à produção de anticorpos IgE direcionados contra essas proteínas. Esses anticorpos se ligam aos receptores das células do sistema imunológico. Com repetidos contatos com o alimento, ocorre a ativação das células que carregam IgE, resultando na liberação de substâncias inflamatórias. Isso desencadeia manifestações clínicas diversas, como urticária aguda, angioedema, broncoespasmo e possivelmente anafilaxia. Dito isto, em termos de riscos, para a Intolerância depende muito do nível ingerido, entretanto no caso da APLV a aversão é crítica e a concentração mínima pode causar grande estrago.

Uma dieta de exclusão de leite de vaca e seus derivados é essencial para o diagnóstico, confirmado através do teste de provocação oral, e também para o tratamento da condição. A princípio, o leite de outros mamíferos, como cabra e ovelha, assim como fórmulas parcialmente hidrolisadas e fórmulas poliméricas isentas de lactose, não devem ser recomendados para crianças com alergia à proteína do leite de vaca (APLV). Contudo, estudos apontam que o baixo teor de caseínas ofensivas (como a α S1), e polimorfismos genéticos ocorrentes na lactação das cabras, caracterizam o leite deste animal como menos propenso a

causar alergias e aumenta a sua digestibilidade, o que reforça seu poder como substituto numa dieta sem leite bovino (Pereira *et al.*, 2022).

3.2 CONTROLE DE QUALIDADE

Na indústria alimentícia, o processo de mudanças e inovações tecnológicas têm alterado os desejos e necessidades dos consumidores, forçando as organizações a adotarem novas práticas para atender às exigências e expectativas dos clientes em relação aos produtos. Portanto, seguir boas práticas de fabricação é fundamental para garantir a produção de um produto final de qualidade e seguro, que atenda aos padrões exigidos pelo mercado e cumpra os requisitos estabelecidos pela legislação atual (Gobis; Campanatti, 2012; Vanzella; Santos, 2015; Da Silva, Da Silva Andrade; De Andrade, 2024).

O controle de qualidade garante o bom funcionamento e a qualidade dos serviços prestados, norteando a realização de boas práticas. Indivíduos mal intencionados, muitas vezes visando lucro, podem fraudar processos da produção com produtos ou alterações químicas, físicas e biológicas. Para assegurar a qualidade do produto, é essencial implementar controle rigoroso em cada etapa do processo. A utilização de manuais de procedimentos operacionais padrão é indispensável para orientar todas as atividades a serem realizadas (Vanzella; Santos, 2015).

Como resultado, as empresas buscam certificações para competir efetivamente não só no mercado nacional, mas também no internacional. A qualidade não é mais apenas um diferencial competitivo, mas uma condição essencial para a permanência no mercado. Recentemente, a qualidade se destacou como um dos principais focos das organizações. Com a influência da globalização, as empresas passaram a avaliar sua qualidade conforme padrões internacionais. Isso requer das estruturas produtivas a adoção de padronizações, registros e o uso eficaz de ferramentas (Vanzella; Santos, 2015; Da Silva; Da Silva Andrade; De Andrade, 2024).

Normalmente, o controle de qualidade de produtos alimentícios utiliza técnicas analíticas clássicas, como ensaios volumétricos e gravimétricos. Estes métodos são conhecidos por demandarem um tempo considerável para análise, além de requererem preparação de amostras e outras etapas subsequentes. Eles também exigem um nível significativo de conhecimento e treinamento por parte dos profissionais envolvidos, e frequentemente utilizam grandes volumes de solventes tóxicos, o que pode ter um impacto

adverso no meio ambiente (Galuszka; Migaszewski; Namiesnik, 2013; Da Silva; Da Silva Andrade; De Andrade, 2024).

Com isso, grandes empresas estão cada vez mais em busca de métodos simples, baratos e eficazes para executar a análise de matrizes complexas, como alimentos, visando otimizar a produção sem sacrificar a agilidade. Considerando a relevância da gestão da qualidade para o desempenho das organizações, estudos salientam a importância do panorama das inovações tecnológicas em desenvolvimento, focadas na aplicação em produtos alimentícios (Thomé *et al.*, 2021; Silva, 2023).

3.2.1 Fraude alimentar

O crime alimentar ocorre quando a escala e o impacto potencial da atividade são substanciais, envolvendo operações criminosas que podem abranger regiões, países ou até mesmo nível internacional, representando um risco significativo para a segurança pública, além de possíveis perdas financeiras substanciais para consumidores e empresas. As oportunidades para fraudes alimentares e adulterações são amplas, dadas as grandes quantidades de alimentos produzidos, exportados e importados, bem como o alto valor agregado de certos produtos específicos (Oliveira; Moraes; Coelho, 2021).

Em termos gerais, as fraudes alimentares consistem em uma ação que provoca intencionalmente uma discrepância entre as alegações sobre o produto alimentar e suas características reais, seja por fornecer informações deliberadamente falsas ou por omissão de forma intencional de ações que deveriam ter sido feitas. Como a maioria dos produtos alimentares é fabricada e comercializada em conformidade com regulamentos e exigências específicas, a fraude alimentar também ocorre quando algum aspecto da produção infringe esses requisitos ou normas (Morin; Lees, 2018).

As fraudes alimentares podem ser feitas de diferentes formas e podem dividir-se em grupos principais (**Figura 6**), entre eles estão: Falsificação, Rotulagem errada, Mercado cinza e as diferentes formas de Adulteração (**Quadro 3**) que podem ocorrer em produtos alimentares, sendo conjunto das adulterações o tipo das fraudes mais frequentes e o foco deste trabalho.

Figura 6. Tipos de fraude em alimentos.



Fonte: Adaptado de Morin; Lees, 2018.

Quadro 3. Tipos de Adultração em alimentos.

Tipos de Adultração	Principais características
Substituição	Ação de trocar um nutriente, ingrediente ou parte de um alimento (geralmente de alto valor) por outro nutriente, ingrediente ou parte de alimento (geralmente de valor inferior).
Diluição	Processo de misturar um ingrediente líquido (soluto) de alto valor com um líquido de valor inferior.
Melhoria não autorizada	Adição de compostos desconhecidos e não declarados aos produtos alimentares para aprimorar suas características de qualidade.
Ocultação	Disfarçar a baixa qualidade de ingredientes ou produtos alimentares.

Fonte: Adaptado de Morin; Lees, 2018.

De modo geral a Adultração é descrita como ação que envolve a adição intencional de uma substância ou elemento estrangeiro de qualidade inferior, geralmente com o objetivo de preparar o produto para venda, substituindo ingredientes mais valiosos por outros menos

valiosos ou inertes (Morin; Lees, 2018). Tais atitudes podem ocorrer com o objetivo de aumentar seu lucro econômico, sem a intenção inicial de causar danos à saúde ou prejuízo à empresa, entretanto representam um risco para a saúde do consumidor, tornando-se um grave problema de saúde pública ao afetar a segurança alimentar. Portanto, são necessárias medidas rigorosas de inspeção por parte dos órgãos públicos para detectar qualquer tipo de fraude. Essas entidades devem aplicar penalidades aos responsáveis por esses crimes e monitorar de perto esses produtos (Spinky e Moyer, 2011; Neves, 2021; Oliveira; Moraes; Coelho, 2021).

Destarte, a qualidade dos produtos alimentares representa uma das principais preocupações tanto no âmbito profissional quanto científico. Portanto, torna-se essencial avaliar a viabilidade de técnicas mais rigorosas para ampliar o entendimento sobre a qualidade desses alimentos. O conhecimento da composição química é crucial para garantir a adequação ao processamento e consumo humano. Assim, o estudo da qualidade química do leite é fundamental para atender às crescentes demandas dos consumidores atuais (Verruck *et al.*, 2019; Ducatti *et al.*, 2023).

Fraudes relacionadas ao leite estão comumente enquadradas como Fraudes por Adulteração econômica, no qual ingredientes mais baratos são adicionados a produtos de maior valor. Estas têm sido registradas globalmente, comprometendo a qualidade e a segurança deste alimento para a população. Nesses casos, os adulterantes comuns incluem adição de substâncias para aumentar o volume, conservantes para prolongar a vida útil do produto, e outros que visam mascarar a acidez resultante da fermentação microbiana (Faria *et al.*, 2022; Silva, 2023).

Assim, a prática comum de adulterar o leite de cabra misturando-o com diferentes proporções de leite de vaca ocorre devido ao maior valor comercial do leite de cabra e seus derivados em comparação ao leite de vaca. Para identificar adulterações causadas pela adição de leite bovino ao leite de cabra, é possível detectar divergências nos parâmetros estabelecidos pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA). Os teores de gordura, proteínas totais, lactose anidra, sólidos não gordurosos e totais podem ser analisados utilizando técnicas de absorção de radiação no infravermelho médio, sensores de condutividade e/ou ensaios enzimáticos (MAPA, 2018; Silva, 2023).

3.3 METODOLOGIAS ANALÍTICAS PARA AUTENTICIDADE DO LEITE CAPRINO

A autenticidade e a detecção de adulteração do leite e produtos lácteos têm sido tradicionalmente realizadas por métodos analíticos convencionais, como os físico-químicos,

sensoriais e cromatográficos (Finete *et al.*, 2013; Kamal; Karoui, 2015; Vieitez *et al.*, 2016). Adicionalmente, outros estudos também mencionaram a eficácia de técnicas eletroforéticas (Cartoni *et al.*, 1999; Pesic *et al.*, 2011) e de ensaios imunológicos (Song; Xue; Han, 2011), com o mesmo intuito de identificar e/ou quantificar a adulteração, obtendo elevados coeficientes de correlação entre os resultados das análises e a concentração de leite bovino detectado no leite de cabra. Para leite em pó, homogeneizar a amostra, reconstituir de acordo com as instruções do fabricante e proceder com análises análogas para leite fluido (MAPA, 2022).

Além destes, é possível citar os estudos de López-Calleja *et al.* (2004), Dabrowska *et al.* (2010), Rodrigues *et al.* (2012) e Di Pinto *et al.* (2017), que empregam técnicas da biologia molecular promissoras na identificação e quantificação de adulterantes em leite e seus derivados. Especificamente, os métodos baseados na amplificação de DNA por PCR (*Polymerase Chain Reaction*) são destacados nesse contexto. No primeiro estudo, foi realizada a quantificação da adulteração de leite por adição de leite bovino, utilizando a amplificação de regiões altamente conservadas do DNA mitocondrial. Já nos estudos subsequentes, focou-se na identificação dessas adulterações.

Contudo, considerando a importância da adulteração na matriz alimentar do leite e as limitações das técnicas convencionais (como alto custo, grande consumo de reagentes e necessidade de operadores qualificados), há diversos esforços voltados para o desenvolvimento de métodos analíticos mais rápidos e econômicos, capazes de detectar, identificar e quantificar esses adulterantes. Dito isto, o uso de práticas não destrutivas e rápidas vem ganhando espaço dentre as metodologias modernas (Kamal; Karoui, 2015; Handford; Campbell; Elliott, 2016; Silva, 2023).

Dentro do contexto das metodologias não destrutivas, técnicas espectroscópicas como Espectroscopia de Fluorescência, Infravermelho Próximo (NIR, do inglês, *Near infrared spectroscopy*) e Infravermelho Médio (MIR, do inglês, *Mid-infrared spectroscopy*), Ressonância Magnética Nuclear (NMR, do inglês, *Nuclear Magnetic Resonance*) e imagens digitais, combinadas com quimiometria e outras técnicas alternativas, têm sido opções extensivamente investigadas na detecção de adulterantes no leite (Dias *et al.*, 2009; Santos; Pereira Filho, 2013; Kamal; Karoui, 2015; Lohumi *et al.*, 2015; Brandão; Anjos; Bell, 2017; Chen *et al.*, 2017; Genis *et al.*, 2020; Ullah *et al.*, 2020; Perreira, 2020; Silva, 2023).

Ainda no âmbito das metodologias não destrutivas, vale destacar que a espectroscopia Raman foi empregada apenas em estudos sobre a adulteração de leite de cabra com leite de

vaca fluidos. Li *et al.* (2023), Batesttin, *et al.* (2022) e Yaman (2020) utilizaram regressão por Mínimos Quadrados Parciais (PLS) para quantificar leite de vaca em leite de cabra. Li *et al.* (2023) e Yaman (2020) estudaram apenas um número limitado de marcas de cada tipo de leite ($n = 3$ e $n = 5$, respectivamente), enquanto Batesttin *et al.* (2022) não revelaram essa informação. Por outro lado, Yaman (2020) utilizou o SIMCA para detectar a adulteração do leite de cabra, mas não obteve resultados significativos de classificação. O **Quadro 4**, apresenta um resumo das metodologias supracitadas:

Quadro 4. Metodologias utilizadas na autenticação de leites e seus derivados.

Metodologias	Produto analisado	Adulterantes	Referência
Eletroforese capilar de zona	Leite e queijo caprino	Leite Bovino	Cartoni et al., (1999)
PCR	Leite caprino e ovino	Leite Bovino	López-Calleja et al. (2004)
Língua eletrônica	Leite caprino	Leite Bovino	Dias et al., (2009)
PCR em tempo real	Leite caprino	Leite Bovino	Dabrowska et al. (2010)
Eletroforese em gel de poliacrilamida nativa	Leite caprino e ovino	Leite Bovino	Pesic et al., (2011)
Ensaio imunoenzimático (ELISA)	Leite caprino	Leite Bovino	Song et al., (2011)
PCR duplex	Leite caprino	Leite Bovino	Rodrigues et al. (2012)
Biureto, Bradford, lowry, markwell, kjeldahl	Leite Bovino UHT (amostras comerciais)	Melamina, ureia e sulfato de amônio	Finete et al, (2013)
Imagens Digitais	Leite Bovino	Água, soro de leite, peróxido de hidrogênio, urina e leites sintéticos	Santos e Pereira Filho, (2013)
Cromatografia gasosa	Leite caprino	Leite Bovino	Vieitez et al, (2016)
Espectroscopia de Fluorescência resolvida no tempo	Leite em pó caprino	Leite em pó Bovino	Brandão, Anjos e Bell, (2017)
PCR em tempo real	Leite caprino fluido e em pó, derivados de leite caprino (queijos e iogurte)	Leite Bovino e ovino	Di Pinto et al. (2017)
Espectroscopia de Fluorescência sincrônica	Leite caprino, ovino e bubalino	Leite Bovino	Genis et al., (2020)

Espectroscopia de Fluorescência “front-face”	Leite Bubalino	Leite Bovino	Ullah et al., (2020)
Espectroscopia no Infravermelho Próximo	Leite caprino	Leite Bovino	Pereira (2020)
Espectroscopia RAMAN	Leite caprino	Leite bovino	Yaman et al (2022)

Fonte: Adaptado de Pereira, 2020.

Diante do exposto, é evidente a variedade de técnicas disponíveis para abordar o problema das fraudes no leite caprino. Considerando fatores como custo, acessibilidade, rapidez na medição e adequação para amostras sólidas, este estudo investigou o desempenho das técnicas vibracionais para enfrentar essa questão. Além de utilizar a Espectroscopia no Infravermelho Próximo, que obteve bons resultados para leite caprino líquido (Pereira, 2020), esta pesquisa também empregou a Espectroscopia Raman, uma técnica que, até o momento, não foi reportada na literatura para a detecção de adulterações em leite em pó.

3.4 ESPECTROSCOPIA VIBRACIONAL

No âmbito das técnicas espectroscópicas, a espectroscopia vibracional estuda a interação da radiação eletromagnética com os movimentos vibracionais das moléculas, abrangendo técnicas no infravermelho e Raman (Sala, 2008).

Por volta de 1800, o astrônomo e compositor alemão, Sir William Herschel (1738 – 1822), ao investigar quais cores eram responsáveis pelo aquecimento dos objetos, acabou por descobrir a radiação infravermelha. Ao perceber que a luz solar continha cores e também funcionava como uma fonte de calor, Herschel, utilizando um prisma para decompor as cores e termômetros, observou que a temperatura aumentava à medida que o termômetro se deslocava do violeta para o vermelho no espectro visível. Vale ressaltar, que o cientista notou que para além do vermelho a temperatura ainda aumentava, ocorrendo em uma radiação invisível ao olho humano (Burns; Ciurczak, 2007).

A região do Infravermelho muito abrangente, logo a técnica de espectroscopia no Infravermelho pode ser dividida em espectroscopia NIR (Near Infrared, infravermelho próximo), MIR (Mid Infrared, infravermelho médio) e FIR (Far Infrared, infravermelho distante). A **Tabela 2** apresenta as diferentes regiões baseando-se nos valores de comprimento de onda (nm) e número de onda (cm^{-1}), o que as subdivide em regiões próximas, médias e distantes.

Tabela 2. Faixas espectrais para as diferentes regiões do infravermelho

Região	Comprimento de onda (nm)	Número de onda (cm ⁻¹)
Próximo	780 - 2500	12800 - 4000
Médio	2500 - 50000	4000 - 200
Distante	50000 - 1000000	200 - 10

Fonte: Adaptado de Neves, 2021.

Além do IR, a espectroscopia Raman também envolve estudo das vibrações moleculares. O título da técnica é uma homenagem a Sir Chandrasekhara Venkata Raman (1888 - 1970), que em 1928 observou e descreveu o efeito de espalhamento inelástico da luz em líquidos, a técnica de espectroscopia Raman recebeu este nome. Com o efeito já previsto matematicamente por Smekal em 1923, e a descoberta do efeito Compton em 1923, Raman começou a considerar a possibilidade de um fenômeno semelhante de espalhamento inelástico na região do espectro visível. Ele conduziu seus experimentos utilizando apenas um telescópio, luz solar e a visão humana como detector (Raman, 1929).

Para garantir que apenas a radiação de maior energia alcançasse a amostra, um filtro azul foi colocado entre o recipiente que continha o líquido transparente purificado (amostra) e o sistema de lentes. Dessa forma, a luz solar, ao ser focalizada pelas lentes, foi filtrada antes de interagir com a amostra. Como resultado, foi observado um desvio na frequência da luz devido ao espalhamento, o que confirmou a existência de um novo fenômeno: o efeito Raman (Raman, 1929). O mesmo foi reconhecido por suas contribuições em 1930, quando recebeu o Prêmio Nobel de Física.

A natureza da fonte de luz (seja monocromática ou policromática) e da própria amostra, são fatores imprescindíveis na compreensão da forma como a radiação vai interagir com a amostra, podendo ocorrer de várias maneiras. Os efeitos dessas interações na espectroscopia vibracional: absorção (ou absorvância) e reflexão (ou reflectância) podem ser observadas nas técnicas NIR, MIR e FIR ao utilizar fontes policromáticas, entretanto a interação da matéria com a radiação ocorre monocromaticamente. Por outro lado, a espectroscopia Raman é caracterizada exclusivamente pelo efeito de espalhamento inelástico ao empregar radiação monocromática (Neves, 2021). Em termos gerais, conhecer os tipos de transições que ocorrem nos níveis vibracionais, a faixa de radiação eletromagnética, o tipo de fonte empregada, além do tipo de matriz analisada, são critérios chave para a seleção da técnica, informações importantes que serão discutidas em detalhes nas Seções 3.4.1 e 3.4.2.

Originalmente utilizada predominantemente para análises qualitativas, a espectroscopia vibracional ganhou bastante espaço na análise quantitativa, crescendo rapidamente a partir dos anos 1970. Nesta década que suas aplicações experimentaram um crescimento significativo, impulsionado pelos avanços em instrumentação, computação e pelo desenvolvimento de ferramentas estatísticas para análise de dados químicos, conhecidas como Quimiometria. Após a década de 1980, tornou-se incomum encontrar trabalhos na literatura que não utilizassem algoritmos para análise de dados multivariados (Pasquini, 2018; Neves, 2021; Pereira, 2022).

Nos últimos anos, as técnicas vibracionais se estabeleceram como ferramentas fundamentais em pesquisas acadêmicas, controle de qualidade em indústrias e órgãos governamentais. Elas encontram aplicações variadas em áreas como química (Yan; Siesler, 2018; Schuchardt; Siesler, 2017), agricultura (Dos Santos *et al.*, 2013), alimentos (Lohumi *et al.*, 2015), meio ambiente (Käppler *et al.*, 2016; Lu; Haes; Forbes, 2018), forense (Silva; Braz; Pimentel, 2019), farmacêutica (Deidda *et al.*, 2019) e petróleo (Dantas; Alves; Poppi, 2017; Hackley e Lünsdorf, 2018).

Ademais, para aplicações na área de alimentos, destacam-se diversas, incluindo autenticação (Lohumi *et al.*, 2015; Valand *et al.*, 2018), classificação e quantificação (Deconinck *et al.*, 2018), detecção de adulterações (Santana; Borges Neto; Poppi, 2019; Rodriguez; Rolandelli; Buera, 2019), controle de qualidade (Ringsted; Siesler; Engelsen, 2017), determinação de parâmetros nutricionais (Neves; Poppi; Siesler, 2019), umidade (Czaja *et al.*, 2018), discriminação de variedades e cultivares (Xu *et al.*, 2016), perfis lipídicos (Casale *et al.*, 2012), embalagens (Haghighi *et al.*, 2019) e processos industriais (Tahir *et al.*, 2019; Grassi; Alamprese, 2018).

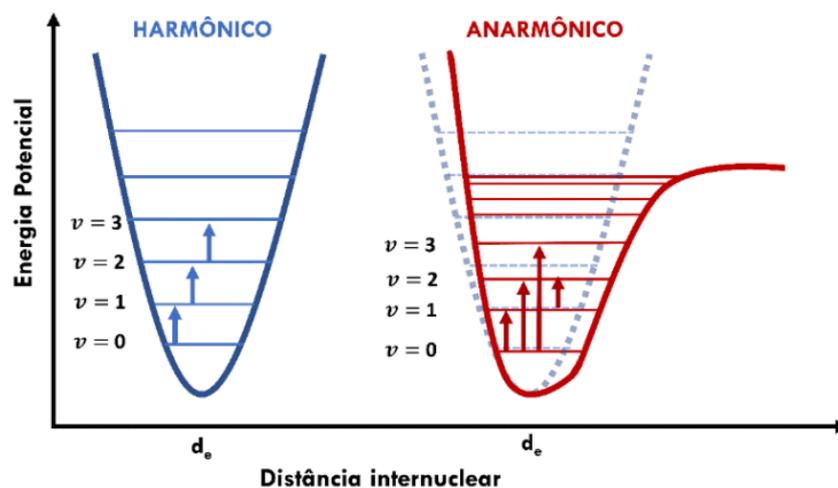
3.4.1 Espectroscopia no Infravermelho Próximo - NIR

Para compreender as transições possíveis inicialmente, consideramos uma molécula como um oscilador harmônico, cujo diagrama de energia é uma parábola com níveis energéticos igualmente espaçados (Burns; Ciurezak, 2007). O comportamento do oscilador harmônico segue a Lei de Hooke, cujo o movimento dos átomos em ligações químicas é comparável ao movimento harmônico de dois corpos ligados por uma mola, que tende a restaurar as massas às suas posições originais (Schrader, 2008; Britannica, 2024). A solução final do modelo harmônico leva em conta as equações da mecânica quântica para definir energeticamente cada nível, onde apenas as transições com $\Delta v = \pm 1$ são permitidas. Portanto,

espera-se observar um único pico de absorção para uma determinada vibração molecular (Skoog; Holler; Crouch, 2018). Desta forma, um único pico de absorção deveria ser observado para uma dada vibração molecular.

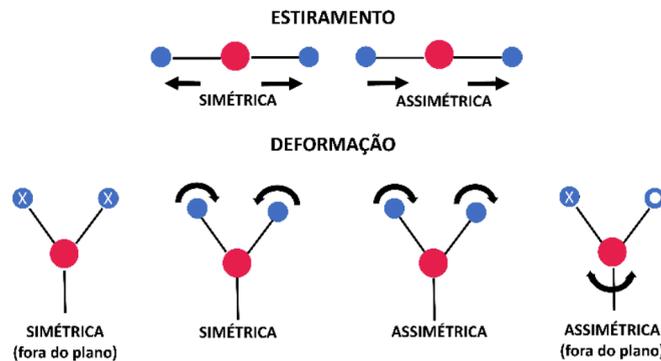
Transições com $\Delta\nu = \pm 2, \pm 3$ (sobretens) e outros sobretens não deveriam ser observadas nos espectros. No entanto, em moléculas que possuem diferenças significativas nas massas dos átomos na ligação química, a vibração não pode ser adequadamente descrita por um oscilador harmônico. O fenômeno de anarmonicidade elétrica e mecânica resulta em desvios onde os níveis energéticos não são igualmente espaçados. À medida que a energia potencial entre os átomos aumenta, ocorre uma diminuição nos níveis quânticos vibracionais, permitindo transições do tipo $\Delta\nu = \pm 2$ ou ± 3 . O diagrama de energia de um oscilador anarmônico se apresenta como uma parábola distorcida (**Figura 7**) (Workman; Weyer, 2007).

Figura 7. Diagrama de energia dos modelos osciladores harmônico e anarmônico.



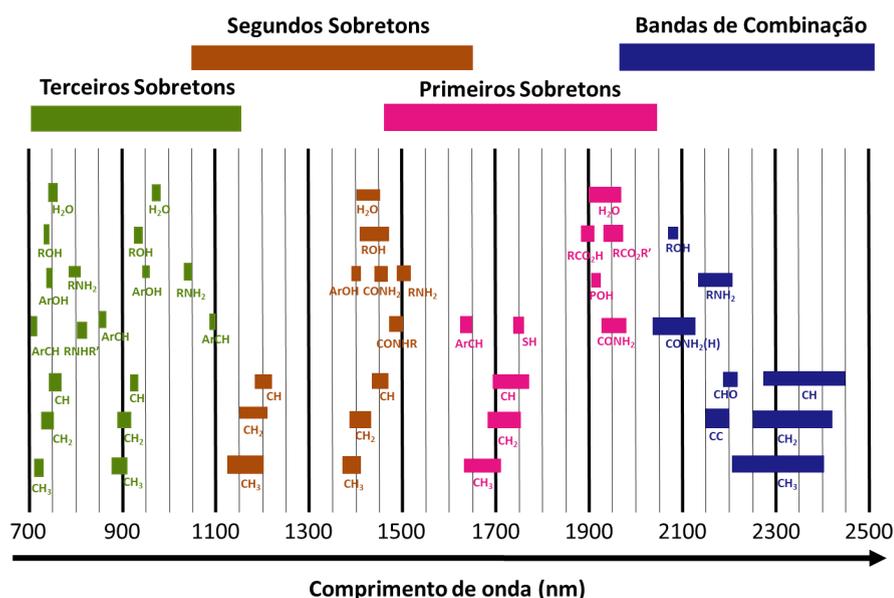
Fonte: Neves, 2021.

As vibrações moleculares podem ser categorizadas em dois grupos principais: estiramentos e deformações angulares, ilustrados na **Figura 8**. Estiramentos envolvem variações na distância interatômica ao longo do eixo da ligação entre dois átomos, podendo ocorrer de maneira simétrica e/ou assimétrica. Deformações angulares são caracterizadas pela variação nos ângulos entre duas ligações químicas, podendo ocorrer dentro do plano e/ou fora dele, também de forma simétrica e/ou assimétrica (Stuart, 2004; Workman; Weyer, 2007).

Figura 8. Vibrações moleculares mais comuns encontradas na espectroscopia infravermelho.

Fonte: Neves, 2021.

É na faixa do Infravermelho Próximo (780 - 2500 nm, **Figura 9**), que são observadas as vibrações de sobretons de primeiro, segundo e terceiro ordem das bandas fundamentais, típicas do infravermelho médio, além das bandas de combinação. Os sobretons ocorrem devido a transições com $\Delta\nu = \pm 2$ ou ± 3 , sendo que suas intensidades são geralmente menores que as das transições fundamentais. As bandas de combinação surgem, quando excitam-se simultaneamente dois modos vibracionais e são predominantes na região de 1900 – 2500 nm. Esse fenômeno resulta da absorção de energia por duas ligações em vez de uma, e a frequência da banda de combinação é a soma ou diferença das frequências dos dois modos fundamentais. Também são observadas as *hot bands*, cujas transições originárias de níveis vibracionais excitados são permitidas (Nakamoto, 2009).

Figura 9. Principais bandas analíticas e posições de pico relativas para absorções proeminentes no NIR.

Fonte: Adaptado de Raypah *et al.*, 2022.

Devido à alta anarmonicidade e ao grande momento dipolar, as principais bandas de absorção na região do infravermelho próximo estão relacionadas aos sobretons e às bandas de combinação das transições vibracionais dos grupos C-H, N-H, O-H e S-H (Williams; Norris, 1987). Os espectros NIR geralmente apresentam bandas largas de absorção com considerável sobreposição. No entanto, essas características não limitam seu uso em aplicações qualitativas e quantitativas. Com o auxílio de ferramentas quimiométricas, as possibilidades de aplicação da espectroscopia vibracional de modo geral têm crescido significativamente, como mencionado anteriormente.

A reflectância difusa é uma das técnicas mais empregadas para análise de sólidos na região do infravermelho próximo. Devido à superfície irregular da amostra, o feixe de radiação incidente se dispersa em várias direções ao ser refletido, resultando em difusão. Esse método analítico possibilita a análise direta de amostras sólidas na superfície ou em contato direto com elas. Antes de cada análise, é essencial realizar a medição do espectro de referência ou branco. Para isso, utiliza-se um material que dispersa e não absorve a radiação, como BaSO₄ e Spectralon®. Alguns equipamentos de bancada estão equipados com um acessório chamado esfera de integração, que é uma esfera oca revestida com material altamente refletivo. Esse acessório concentra a radiação no detector, melhorando a relação sinal/ruído nos espectros obtidos. As vantagens dessa metodologia incluem a capacidade de previsão rápida, ser ambientalmente amigável, além de ser não destrutiva e não invasiva (Prieto *et al.*, 2017; Pereira, 2022; Silva, 2023).

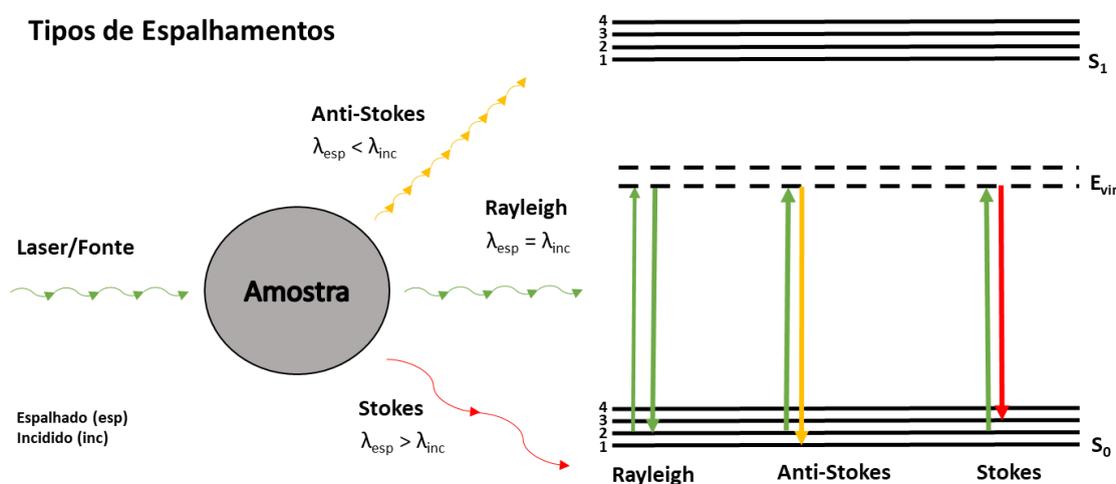
3.4.2 Espectroscopia RAMAN

Existem dois tipos de espalhamento que podem ocorrer quando a radiação eletromagnética incide sobre a amostra: o elástico e o inelástico. O primeiro, conhecido como espalhamento Rayleigh, envolve fótons que possuem a mesma energia que o fóton incidente. Este tipo de espalhamento não fornece informações sobre a composição molecular ou estrutura da amostra. O segundo tipo é denominado espalhamento Raman e pode ser subdividido em Stokes e Anti-Stokes. A nomenclatura "Stokes" é uma homenagem a Sir George Gabriel Stokes, pioneiro no estudo da fluorescência (Long, 2002).

Quando a radiação eletromagnética incide sobre a amostra com energia menor do que a necessária para absorção, cria-se um estado virtual de energia na molécula. Esse estado é chamado assim por não ter solução na equação de Schrödinger (Schrödinger, 1926). Posteriormente, a molécula pode retornar ao seu estado fundamental ou para um estado

vibracional diferente do fundamental, a diferença de energia entre os estados vibracionais inicial e final determina qual tipo de espalhamento ocorrerá. No espalhamento Rayleigh os fótons espalhados têm energia igual aos incidentes ($\lambda_{\text{esp}} = \lambda_{\text{inc}}$), enquanto no Raman Stokes, os fótons espalhados têm energia menor que os incidentes ($\lambda_{\text{esp}} > \lambda_{\text{inc}}$), e por último no Anti-Stokes, os fótons têm energia maior que os incidentes ($\lambda_{\text{esp}} < \lambda_{\text{inc}}$) (Figura 10)(Long, 2002).

Figura 10. Representação esquemática dos tipos de espalhamentos observados. Os níveis energéticos S_0 e S_1 correspondem ao nível fundamental e ao excitado da molécula, respectivamente. O estado virtual (E_{vir}) é representado pelas linhas tracejadas.



Fonte: Adaptado de Neves, 2021.

O espectro Raman exibe um conjunto de modos vibracionais permitidos para uma molécula, no qual o número de bandas, suas frequências e as intensidades estão associadas às vibrações das ligações químicas. As intensidades são determinadas pela distribuição populacional das moléculas nos estados fundamental e excitado, seguindo a distribuição de Boltzmann (Nakamoto, 2007). À medida que o nível energético vibracional diminui, aumenta a população de moléculas nesse nível. Logo, o nível fundamental vibracional contém mais moléculas do que os primeiros níveis vibracionais excitados no estado eletrônico fundamental. Com um maior número de moléculas, há uma maior interação com a radiação eletromagnética incidente, resultando em espectros com maiores intensidades.

Os espectros do ramo Stokes são geralmente mais intensos do que os espectros do ramo Anti-Stokes. Por consequência, o ramo Stokes é preferencialmente utilizado em aplicações analíticas, incluindo na área de alimentos (Qin; Chao; Kim, 2013; Lohumi *et al.*, 2015). Dado o exposto, é visível o potencial dessa técnica, uma vez que permite a obtenção de

informações sobre a estrutura molecular, os níveis de energia, as ligações químicas, bem como a identificação e a quantificação de elementos químicos e moléculas (Sala, 2008).

Como uma das técnicas ópticas emergentes na literatura, a espectroscopia Raman é aplicada em análises de componentes moleculares de vários materiais. Ao ser baseada em vibrações essa técnica gera espectros de emissão com bandas correspondentes às vibrações das ligações moleculares presentes na amostra, tais espectros podem ser comparados a um "Fingerprint" (Impressão digital) da molécula, oferecendo percepções que outras técnicas ópticas não conseguem proporcionar (Silveira, 2008, Rodrigues Júnior, 2015).

De maneira geral, cada impressão digital da substância fornece indícios acerca da composição molecular e caracteriza variações estruturais interatômicas das moléculas nos materiais analisados, por meio do modo vibracional característico do composto. Os espectros Raman são obtidos irradiando-se uma amostra com uma fonte de radiação eletromagnética no espectro visível ou infravermelho próximo, e durante o processo, o espectro da radiação espalhada de forma inelástica é medido usando um espectrômetro apropriado. Geralmente, essa radiação Raman corresponde a uma fração pequena, tipicamente entre 10^{-9} e 10^{-6} do espalhamento total observado, sendo que a maior parte do espalhamento é elástico. Por esse motivo, é necessário uma fonte de excitação monocromática intensa e um detector sensível para detectar eficientemente o espalhamento Raman (Angel *et al.*, 1995; Sócrates, 2004; Rodrigues Júnior, 2015).

Com espectros de bandas estreitas e de difícil interpretação, cabe salientar a importância da extração de informações, logo a quantidade e a qualidade destas obtidas de um espectro dependem de vários fatores. Dentre eles tem-se: características da fonte de radiação utilizada para excitar as amostras, como potência e comprimento de onda do laser; a forma como o sinal é coletado da amostra após a interação, como o tempo de aquisição, e a resolução temporal ou espectral do instrumento de dispersão (espectrógrafo) também são críticos; Por fim, o tipo e a sensibilidade do detector do sinal, juntamente com as propriedades físicas e o coeficiente de espalhamento da amostra em análise, completam os fatores determinantes para a análise espectroscópica (Alcântara Junior, 2002; Afseth; Wold; Segtnan, 2006). Por muito tempo limitada à pesquisa acadêmica, a técnica de espectroscopia Raman está gradualmente emergindo como uma inovação na área de controle de qualidade, o que corrobora com o objetivo do trabalho.

3.5 INSTRUMENTAÇÃO PORTÁTIL

Embora os instrumentos e equipamentos de grande escala sejam conhecidos por suas diversas aplicações e elevada sensibilidade na detecção, eles não atendem às exigências de portabilidade que a tecnologia moderna demanda (Wang *et al.*, 2023). Por exemplo, métodos analíticos tradicionais, como cromatografia, espectrometria de massa e ressonância magnética, são reconhecidos por sua alta sensibilidade e especificidade. No entanto, esses métodos frequentemente apresentam desvantagens significativas, como a necessidade de tempo considerável para os resultados, equipamentos caros, uso de reagentes e operadores com treinamentos específicos para a execução dos testes. Essas características incapacitam esses métodos para testes *in situ*, onde a capacidade de tomar decisões rápidas é crucial para a obtenção de soluções. Em resposta a esses desafios, os equipamentos portáteis, com sua flexibilidade e capacidade de adaptação, têm se mostrado mais adequados para atender às exigências de análise e inspeção no local (Xing *et al.* 2021; Jiang *et al.*, 2024).

O principal objetivo da miniaturização de instrumentos é garantir que o desempenho do equipamento não seja comprometido, isto é, que os resultados obtidos sejam comparáveis aos de dispositivos fixos, tanto em termos qualitativos, quanto quantitativos. A inclusão desses instrumentos enriquece o repertório de ferramentas disponíveis, amplia as possibilidades de medição e contribui para a solução de problemas complexos, impulsionando o progresso da ciência e da tecnologia. Além disso, a integração desses dispositivos em celulares, tornam essas aplicações ainda mais acessíveis e versáteis (Reinig *et al.*, 2018; Gullifa *et al.*, 2023).

A diversidade destes dispositivos, incluindo scanners, drones e equipamentos analíticos miniaturizados, destaca-se como essencial em uma ampla gama de aplicações (Xing *et al.*, 2021). Como exemplo é possível citar os espectrômetros NIR (infravermelho próximo) miniaturizados e portáteis, que têm expandido significativamente as aplicações da tecnologia de infravermelho próximo, tornando-se cada vez mais comuns em análises em campo. Uma das principais vantagens para instrumentos portáteis é o peso, pois enquanto os espectrômetros NIR de bancada pesam cerca de quilos, os espectrômetros NIR portáteis são significativamente mais leves, com aproximadamente 100 g, facilitando o manuseio e deslocamento (Gullifa *et al.*, 2023).

Inicialmente desenvolvidos para uso militar, esses dispositivos foram rapidamente adotados pela indústria para a verificação rápida da qualidade em diversas situações. Hoje, os espectrômetros NIR portáteis são amplamente reconhecidos por sua eficácia em áreas como: fiscalização e vigilância de produtos alimentícios (Goi *et al.*, 2022; Coombs *et al.*, 2021; Yao

et al., 2021), monitoramento e diagnósticos médicos (Vera *et al.*, 2022), análises forense (Fonseca *et al.*, 2022), inspeção de combustíveis (Tosato *et al.*, 2020 ; Wan *et al.*, 2021), entre outras, oferecendo uma ferramenta valiosa para avaliações rápidas e precisas em uma variedade de contextos (Gullifa *et al.*, 2023). Outrossim, tal equipamento pode ser manuseado de forma eficiente por usuários treinados, mesmo que não sejam cientistas especializados (Pomerantsev e Rodionova, 2021).

3.6 QUIMIOMETRIA

Os estudos envolvendo a área iniciaram na década de 1970 com a introdução do termo por Svante Wold, um renomado estatístico e quimiometrista sueco, marcando o uso pioneiro de métodos estatísticos avançados como análise de componentes principais (PCA) e regressão multivariada na química analítica. Essas técnicas foram desenvolvidas para lidar com dados complexos e multidimensionais, permitindo uma análise mais precisa e interpretativa dos resultados experimentais (Brereton *et al.*, 2017; Silva, 2023).

Ao longo dos anos 1980, a quimiometria ganhou proeminência internacional devido à crescente complexidade dos problemas analíticos na química e à necessidade de abordagens mais sofisticadas. Eventos como a implementação do título *Chemometrics* pela revista *Analytical Chemistry* em 1980, e o lançamento de dois jornais exclusivamente dedicados ao tema em 1987, contribuíram significativamente para a divulgação e consolidação do campo em menos de duas décadas (Ferreira, 2015). Desde então, a quimiometria continua a evoluir, sendo amplamente aplicada em áreas como análise de alimentos, farmacêutica e ambiental, impulsionada pela integração de novas tecnologias e métodos estatísticos avançados. No Brasil, a quimiometria começou a ser aplicada há pouco mais de 30 anos. (Brereton *et al.*, 2017; Silva, 2023).

A quimiometria possui a capacidade de otimizar o número de experimentos, estimar interações entre variáveis e selecionar condições experimentais ideais. Por isso, esse conjunto de técnicas vem ganhando destaque em um número crescente de publicações, devido à sua versatilidade em superar limitações analíticas e fornecer informações por meio da medição de sinais (Tarapoulouzi; Ortone; Cinti, 2022).

O estudo quimiométrico engloba uma variedade de métodos com aplicações diversas, incluindo análise exploratória de dados, calibração multivariada, planejamento e otimização experimental, e reconhecimento de padrões. A escolha do método quimiométrico adequado depende das informações que se deseja extrair do conjunto de dados em questão. Neste

trabalho, foram utilizados a análise exploratória dos dados e reconhecimento de padrões, em conjunto com pré-processamentos.

3.6.1 Pré-processamento dos dados

As informações relevantes dos dados gerados pelas técnicas de espectroscopia vibracional geralmente estão distribuídas por todo o espectro e não são prontamente acessíveis para a determinação analítica, representando um grande desafio para as análises e correlações químicas (Brasca; Goicoechea; Culzoni, 2018; Pasquini *et al.*, 2018; Pereira, 2020).

Nas análises químicas de dados espectrais, advindos de técnicas vibracionais, as variações físicas da amostra, como tamanho, distribuição de partículas, densidade, formato e superfície, são fatores cruciais que influenciam as medições. Essas variações podem resultar em espectros com variações não correlacionadas às características químicas das amostras, comprometendo a precisão dos modelos desenvolvidos. Portanto, é essencial aplicar pré-processamentos para minimizar tais efeitos físicos e, assim, realçar o efeito químico de forma mais significativa (Huang; Romero-Torres; Moshgbar, 2010; Pereira, 2020).

O pré-processamento dos dados visa reduzir tanto a variabilidade aleatória quanto às fontes de variação indesejáveis não relacionadas à propriedade de interesse. Nesta fase, o objetivo é melhorar as condições da matriz de dados para explorar mais eficazmente as informações disponíveis (Rinnan; Van den berg; Engelsen, 2009). Em análises de dados espectroscópicos, especialmente nos espectros NIR, observam-se variações sistemáticas indesejáveis influenciadas por características físicas da amostra, como diferenças no tamanho, distribuição de partículas e variações na densidade. Consequentemente, essa variação precisa ser eliminada ou reduzida, pois a falta de correlação com as características químicas da amostra pode comprometer a capacidade preditiva dos modelos desenvolvidos (Tibola *et al.*, 2018; Huang; Romero-Torres; Moshgbar, 2010).

Diversas técnicas de pré-processamento de espectros são empregadas para mitigar o impacto das características físicas das amostras, destacando-se as derivada Savitzky-Golay (do inglês, *Savitzky-Golay Derivative* - SGD), normalização padrão por variação (do inglês, *Standard Normal Variate* - SNV), suavização por filtro Savitzky-Golay (do inglês, *Savitzky-Golay Smoothing* - SGS), correção multiplicativa de sinal (do inglês, *Multiplicative Scatter Correction* - MSC), correção linear da linha de base (do inglês, *Linear Baseline Correction* - LBC) e ajuste de offset (*Baseline Offset* - OFF) (Ely; Thommes; Carvajal, 2008;

Huang; Romero-Torres; Moshgbar, 2010; Pereira, 2020, Nóbrega, 2021). Vale salientar que os métodos de pré-processamento podem ser empregados nas amostras (linhas) e nas variáveis (colunas) da matriz de dados (Ferreira, 2015; Nóbrega, 2021).

O filtro de suavização Savitzky-Golay é uma técnica projetada para reduzir o ruído nos dados enquanto preserva características importantes, aumentando a relação sinal/ruído presente no conjunto de dados antes da modelagem. As derivadas são ferramentas matemáticas eficazes para corrigir o desvio da linha de base nos espectros, removendo deslocamentos constantes e lineares, o que melhora a resolução dos sinais originais. A primeira derivada é usada para eliminar desvios aditivos da linha-base, enquanto a segunda derivada também corrige efeitos multiplicativos (Savitzky; Golay, 1964; Huang; Romero-Torres; Moshgbar, 2010; Pereira, 2020; Nóbrega, 2021).

Um outro método de pré-processamento utilizado, este para mitigar os efeitos multiplicativos do espalhamento e compensações de linha de base é o MSC. A normalização MSC é alcançada centralizando cada espectro na média da absorvância para cada comprimento de onda e corrigindo o espectro usando a inclinação do ajuste linear (Huang; Romero-Torres; Moshgbar, 2010; Manley; Baeten, 2018). Adicionalmente, temos o SNV, que é similar ao MSC no sentido de que ambos corrigem efeitos multiplicativos e interferências de espalhamento. No entanto, diferentemente do MSC, o SNV dispensa a necessidade de um espectro médio de referência (Oliveri; Simonetti, 2016; Manley; Baeten, 2018).

Referente às variáveis (colunas), temos a centralização dos dados na média, que é amplamente empregada em análises espectrais. Nesse método de pré-processamento, o valor médio é calculado para cada variável e subtraído de cada elemento correspondente na respectiva coluna, o que desloca a origem natural das coordenadas para a média multivariada dos dados. Do ponto de vista estatístico, a centralização visa evitar que os pontos mais distantes do centro dos dados tenham mais influência do que aqueles que estão mais próximos (Tibola *et al.*, 2018; Nóbrega, 2021).

Em estudos sobre adulteração de leite e derivados os autores frequentemente usam uma ou várias ferramentas de pré-processamento em combinação, tendo em vista as características complementares das mesmas, ou comparam a aplicação de diferentes métodos para alcançarem melhores índices de precisão do modelo (Huang, *et al.*, 2015; Lim *et al.*, 2016; Wu *et al.*, 2016; Mabood *et al.*, 2017a; Mabood *et al.*, 2017b; Karunathilaka *et al.*, 2018). Dito isto, os métodos de pré-processamento, tanto isolados quanto combinados,

aplicados nas amostras deste estudo incluíram: correção linear da linha de base e ajuste de offset, SNV, MSC, SGS, e SGD. Além disso, as variáveis foram centralizadas na média.

3.6.2 Técnicas de reconhecimento de padrões

Os métodos computacionais utilizados para identificar e interpretar padrões complexos em dados são conhecidos como técnicas de reconhecimento de padrões. Essas técnicas são fundamentais em áreas como aprendizado de máquina e análise de dados, onde podem ser divididas em duas categorias principais: métodos supervisionados e métodos não supervisionados. Na análise não supervisionada, o objetivo é identificar similaridades e diferenças entre amostras em um conjunto de dados total, sem a necessidade de rotular amostras ou objetos (Ferreira, 2015; Nóbrega, 2021). Já as técnicas supervisionadas envolvem o treinamento de modelos com dados rotulados para previsão ou classificação, enquanto as não supervisionadas exploram estruturas e padrões nos dados sem a necessidade de rótulos prévios.

Algoritmos como Análise de Componentes Principais (PCA) e Análise de Agrupamento Hierárquico (HCA) foram desenvolvidos para facilitar a interpretação dos resultados, incluindo a geração de gráficos, visando extrair informações contidas em conjuntos de dados multivariados. Essas abordagens são cruciais para extrair percepções valiosas de conjuntos de dados complexos em diversas aplicações, como reconhecimento de imagens, processamento de linguagem natural e análise científica.

Com o objetivo de classificar novas amostras em uma das classes conhecidas (alvo) com base nos padrões aprendidos durante o treinamento do modelo, os métodos de classificação supervisionada, utilizam as informações de cada amostra com classes pré-estabelecidas para construir modelos de classificação (Berrueta; Alonso-Salces; Héberger, 2007; Tibola *et al.*, 2018). Nesse contexto, para este estudo, foram utilizados métodos não supervisionados (PCA), além de métodos supervisionados divididos em modelos de classe, como a Modelagem Independente e Flexível por Analogia de Classes direcionado pelos dados (Data-Driven – Soft Independent Modelling of Class Analogy - DD-SIMCA) e Mínimos quadrados parciais de uma classe (OC-PLS).

3.6.3 Análise por componentes principais

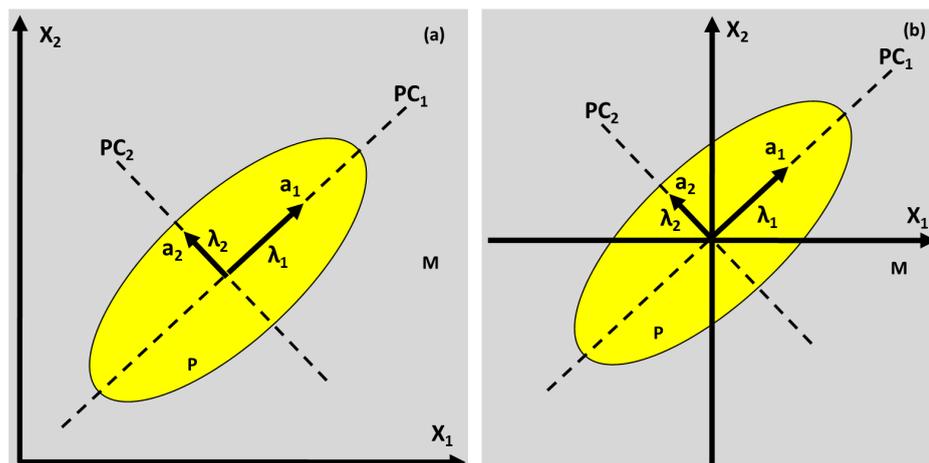
A Análise de Componentes Principais (PCA) é uma técnica não supervisionada (linear), amplamente aplicada em sistemas multissensoriais, que consiste numa abordagem

exploratória usada para reduzir a dimensionalidade dos dados originais, facilitando a avaliação das relações entre amostras semelhantes (Harrou *et al.*, 2015; Moufid *et al.*, 2022). A PCA projeta os dados multivariados em um espaço de dimensão menor, reduzindo a dimensionalidade do conjunto de dados enquanto preserva a maior parte da variância presente nos dados originais (Ferreira, 2015; Silva *et al.*, 2017).

A ferramenta transforma as variáveis originais, que podem estar correlacionadas, em variáveis linearmente não correlacionadas, ortogonais e independentes entre si, conhecidas como componentes principais (PCs). Esta transformação é orientada para maximizar a variância nas primeiras componentes principais. É importante notar que a primeira componente principal (PC1) explica a máxima variância dos dados originais, enquanto a segunda componente (PC2) é ortogonal à PC1 e captura a maior variância restante. As componentes subsequentes não estão correlacionadas com as anteriores, sendo calculadas em ordem decrescente de importância (Granato *et al.*, 2018; Shah; Chauhan; Chaudhury, 2020).

De maneira ilustrativa a PCA identifica as direções principais de variabilidade nos dados e estabelece um novo sistema de coordenadas por meio de rotações ótimas (PC1 e PC2). Os autovetores a_1 e a_2 determinam os eixos deste sistema com autovalores λ_1 e λ_2 (em que $\lambda_1 \geq \lambda_2$), que representam a variância dos dados no novo sistema de coordenadas (Figura 11) (Niculescu; Andrei, 2016; Akulich *et al.*, 2022).

Figura 11. Ilustração idealizada do método de PCA para duas variáveis aleatórias x_1 e x_2 (a), e da centragem na média dos dados (b).



Fonte: Adaptado de Niculescu; Andrei, 2016.

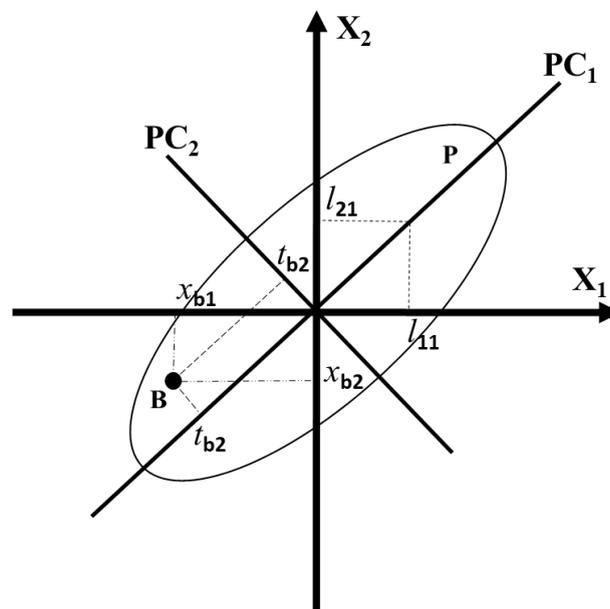
Como representado na Figura 11, o espaço transformado que pode ser imaginado como um elipsoide de dimensão “P”, em que o maior eixo do elipsoide (PC_2) indica a direção de máxima variância nos dados, sendo “M” representante do número de dimensões

originais, logo $P < M$. Dado um conjunto de N amostras e uma matriz de dados X de dimensões $N \times M$, onde cada $X_i \in \mathbb{R}^M$ para $i = 1, \dots, N$ e $j = 1, \dots, M$, os dados devem ser padronizados de modo a serem centrados em torno de zero. Isso é feito subtraindo a média de cada variável dos pontos de dados individuais em cada dimensão específica (Akulich et al., 2022).

Matematicamente, o algoritmo PCA realiza a decomposição da matriz de dados original ou pré-processada, denotada como matriz X , em duas matrizes menores: os escores (Scores - T) e os pesos (Loadings - P). Os escores representam as projeções ortogonais das amostras ao longo de cada PC (Figura 12, t_{b1} , t_{b2}), enquanto os pesos indicam a contribuição de cada variável original em relação à sua respectiva PC (Figura 12, l_{11} , l_{21}). Além disso, há uma matriz de resíduos (E) que captura a variância não explicada pelo modelo (Eq.1) (Wold et al., 1987; Ferreira et al., 1999; Milani et al., 2020; Nóbrega, 2021; Silva, 2023).

$$X = T P^T + E \quad \text{Eq. (1)}$$

Figura 12. Representação gráfica das PCs, com indicação dos pesos de PC1 e dos escores da amostra B.



Fonte: Adaptado de Ferreira, 2015.

Na PCA, graficamente, diversas ferramentas de diagnóstico são empregadas para explorar conjuntos de dados. Isso inclui o gráfico de resíduos, útil para determinar o número ideal de PCs, os gráficos de escores, que revelam possíveis agrupamentos entre as amostras e

detectam amostras anômalas, e o gráfico de pesos, que facilita a identificação de variáveis influentes na formação de agrupamentos nos gráficos de escores, além de destacar variáveis anômalas (Souza; Poppi, 2012; Allegreta *et al.*, 2020).

O modelo PCA é o primeiro termo na Eq.1 que explica os dados usando o número selecionado de PCs, A. A variância residual relativa (R_A):

$$R_A = \frac{|SS_{err}|}{|SS_{tot}|} = \frac{|\sum_{i=1}^i \sum_{j=1}^j e_{ij}^2|}{|\sum_{i=1}^i \sum_{j=1}^j x_{ij}^2|} \quad \text{Eq. (2)}$$

É utilizada como uma métrica para avaliar o desempenho do modelo PCA. A variância explicada complementar é calculada como $1 - R_A$. Esses valores podem ser obtidos tanto para o modelo PCA completo, fornecendo a variância residual total e a variância total explicada, quanto para a contribuição individual de cada componente principal (PC). Normalmente, os valores são apresentados em um gráfico em função do número de PCs.

A relação entre o modelo PCA e cada objeto pode ser descrita por duas distâncias: a distância ortogonal e a distância no espaço dos escores. A distância ortogonal (OD; do inglês orthogonal distance), v_i , é a distância euclidiana ao quadrado entre um ponto dos dados, correspondente ao objeto, e o espaço das PCs calculado no espaço das variáveis originais (Eq.3).

$$v_i = \sum_{j=1}^j e_{ij}^2 \quad \text{Eq. (3)}$$

A distância de escore (SD; do inglês score distance), h_i , É calculada no subespaço dos escores de PCA como a distância de Mahalanobis quadrada (T^2) entre a projeção do ponto e a origem do subespaço (Eq.4)

$$h_i = \sum_{a=1}^A \frac{t_{ia}^2}{\lambda_a} \quad \text{Eq. (4)}$$

Ambas as distâncias são bem aproximadas pelas distribuições qui quadrado escalonadas:

$$N_v \frac{v}{v_0} \alpha X^2(N_v), N_h \frac{h}{h_0} \alpha X^2(N_h) \quad \text{Eq. (5)}$$

onde v_0 e h_0 são os fatores de escala, enquanto N_v e N_h representam os graus de liberdade (DoF, do inglês *degrees of freedom*). Os fatores de escala e os graus de liberdade são parâmetros da distribuição que são desconhecidos e são estimados por meio de uma abordagem orientada aos dados. Para um conjunto de dados regular (sem outliers), as estimativas são baseadas nos valores padrão de média e variância calculados para as ODs e SDs. (Rodionova; Kucheryavskiy; Pomerantsev, 2021).

3.6.4 Mínimos quadrados parciais de uma classe (OC-PLS)

O OC-PLS é um método de modelagem de classe que utiliza um modelo de mínimos quadrados parciais para correlacionar as medições instrumentais com um vetor composto de valores 1. Diferentemente do SIMCA, que estima os componentes principais apenas com base na variância das variáveis independentes, o OC-PLS ajusta os componentes (LVs) levando em consideração simultaneamente a variância explicada nas variáveis independentes e dependentes, além da compactação das amostras dentro da mesma classe (Xu et al., 2014; Miaw, 2018).

O número de LVs é estimado por meio de validação cruzada. Ao construir um modelo OC-PLS são calculadas duas medidas de distância: o valor de Hotelling (T^2 , Eq.7), que se baseia na distância dos escores (SD), e os resíduos absolutos centralizados (ACR, Eq.6).

$$ACR = |1 - \hat{y}_j - \hat{\mu}_e| \quad \text{Eq. (6)}$$

\hat{y}_j - resposta ajustada do objeto

$\hat{\mu}_e$ - média dos erros de treinamento

$$T^2 = \sum_{i=1}^K \frac{(t_i - \bar{t}_i)^2}{S_{t,i}^2} \quad \text{Eq. (7)}$$

\bar{t}_i - Média da i -ésima LV
 S_{t,i^2} - Variância da amostra da i -ésima LV
 K - Número de LVs significativas

O ACR pode assumir uma distribuição normal com média 0. O desvio padrão do modelo residual pode ser estimado por validação cruzada, de acordo com:

$$\widehat{\sigma}_e = \sqrt{\frac{\sum_{i=1}^N (1 - \hat{y}_i - \widehat{\mu}_{e_i})^2}{N - 1}} \quad \text{Eq. (8)}$$

Em que N é o número total de objetos omitidos durante a validação cruzada e \hat{y}_i é a resposta prevista do i -ésimo objeto omitido. Dado um nível de confiança, α , os limites de confiança superiores (UCLs) para ACR e T^2 podem ser calculados, conforme as Equações abaixo:

$$ACR_{UCL} = Z_{\alpha/2} \times \widehat{\sigma}_e \quad \text{Eq. (9)}$$

$Z_{\alpha/2}$ - Ponto crítico superior da distribuição normal padrão

$$T_{UCL}^2 = \frac{(n^2 - 1)K}{n(n - K)} F_{\alpha(K, n-K)} \quad \text{Eq. (10)}$$

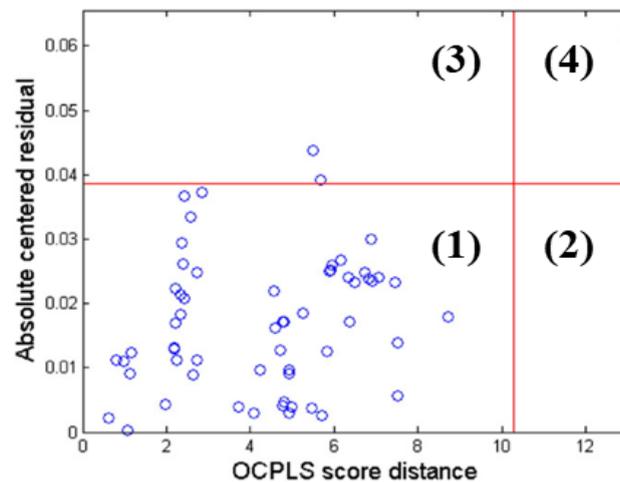
$F_{\alpha(K, n-K)}$ - Ponto crítico superior da distribuição F com $(K, n - K)$ graus de liberdade.

A SD mede a distância de um objeto em relação ao centro da classe no espaço das LVs significativas de OC-PLS, enquanto o ACR avalia a dispersão dos resíduos na projeção sobre o vetor dos coeficientes de regressão OC-PLS. Com base nos valores de SD e ACR, as amostras podem ser classificadas em quatro categorias: amostras regulares (pequena SD e pequeno ACR), pontos de alavancagem bons (grande SD e pequeno ACR), outliers de classe (pequena SD e grande ACR), e pontos de alavancagem ruins (grande SD e grande ACR).

Dependendo da definição do problema, pontos de alavancagem bons, outliers de classe e pontos de alavancagem ruins podem ser identificados como três tipos distintos de outliers. Os limites para SD e ACR são estimados com base em níveis de confiança pré-estabelecidos (Xu et al., 2014; Miaw, 2018).

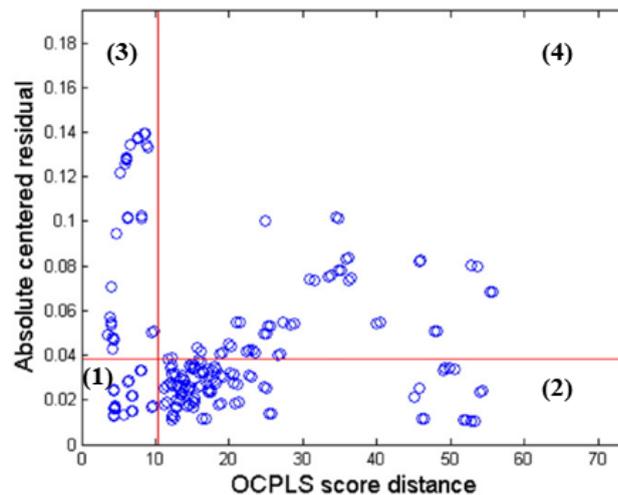
De acordo com as Eqs. (9) e (10), um gráfico de diagnóstico de outliers pode ser plotado como mostrado na Figura 13. Um objeto pode se enquadrar em uma das quatro categorias. Como o ACR é uma medida da distância de um objeto em relação ao modelo de classe OC-PLS, um objeto com uma SD pequena, mas um ACR grande, é chamado de outlier de classe. Para controle estatístico multivariado da qualidade (MSQC), objetos que caem nas regiões 2, 3 e 4 podem ser detectados como diferentes tipos de outliers (Xu et al., 2013).

Figura 13. Gráfico do modelo OC-PLS na etapa de predição da classe-alvo.



Fonte: Xu et al., 2013.

A Figura 14 demonstra os resultados da previsão de objetos. A SD OC-PLS é útil para detectar objetos adulterados, indicando que usar apenas o resíduo da resposta prevista é insuficiente para o diagnóstico de outliers.

Figura 14. Gráfico do modelo OC-PLS na etapa de predição da classe não-alvo.

Fonte: Xu *et al.*, 2013.

O OC-PLS, que considera tanto o resíduo das respostas previstas quanto a SD, mostra-se promissor como um classificador de uma única classe. A correlação entre o OC-PLS e alguns métodos existentes sugere que ele é uma ferramenta eficaz para o MSQC. Quando aplicado à detecção não direcionada de adulteração, o OC-PLS demonstrou a eficácia de utilizar ambas as medidas de distância mencionadas (Xu *et al.*, 2013).

3.6.5 Modelagem independente e flexível por analogia de classes direcionada pelos dados (DD-SIMCA)

O modelo SIMCA, proposto por Wold, é empregado para classificar amostras em conjuntos de dados de alta dimensionalidade (Pontes, 2006; Ferreira, 2015). O SIMCA orientado aos dados (*Data Driven SIMCA*) é uma versão modificada do SIMCA original, com foco na construção de uma área de aceitação confiável para cada classe. Para isso, é necessário realizar uma análise estatística detalhada das distâncias dos escores e das distâncias ortogonais. Os valores de distância devem ser tratados como dados experimentais, e suas distribuições são estimadas por um método baseado em dados. Utilizando a distribuição qui-quadrado, é possível estabelecer uma área de decisão com dois níveis, incluindo limiares extremos e outliers, tanto para conjuntos de dados normais quanto para aqueles que contêm outliers (Pomerantsev & Rodionova, 2014a, 2014b; Zontov *et al.*, 2017).

Este modelo também é um classificador do tipo *one class*, que é bastante utilizado como ferramenta de reconhecimento de padrões para autenticação de alimentos (Nóbrega, 2021), fármacos (Silva, 2020), entre outros. Neste método, apenas a classe "alvo" é modelada,

e amostras que não se enquadram nessa classe são consideradas extremas, podendo indicar adulteração ou falsificação, dependendo da análise dos dados (Nóbrega, 2021).

Em um conjunto de dados a classe alvo é construída com base nas propriedades de suas amostras representativas, conhecidas como impressões digitais. Essas propriedades são sinais analíticos gerados por equipamentos como espectroscopia, cromatografia ou outras técnicas analíticas. Os resultados dessas propriedades formam uma Matriz Principal (X_0), que representa um conjunto de dados das amostras autênticas, devidamente pré-processadas e centralizadas na média.

De maneira igual ao SIMCA, seu precursor, o modelo DD-SIMCA possui várias etapas, sendo realizado primeiramente a decomposição da matriz de dados de treinamento da classe autêntica utilizando a PCA (Eq. 1) (Rodionova; Titova; Pomerantsev, 2016). O número de PCs influencia diretamente na complexidade e na qualidade da classificação. Aumentar o número de PCs aumenta a variância explicada nos dados de treinamento, porém, a inclusão excessiva de PCs pode introduzir ruído e afetar negativamente a capacidade preditiva do modelo. Dito isto, a complexidade do modelo geralmente é determinada pelo princípio da parcimônia, ou seja, escolhendo o melhor resultado com o menor número de PCs possível (Rodionova *et al.*, 2018).

Já na segunda etapa, são estimadas as distâncias dos escores (do inglês: *score distance* - SD), representado por h_i , e a distância ortogonal (do inglês: *orthogonal distance* - OD), representado por v_i , para todas as amostras pertencentes ao conjunto de treinamento resultantes da PCA, seguindo as Eq.11 e 12 (Zontov *et al.*, 2017).

$$h_i = t_i^t (T^T T)^{-1} t_i \quad \text{Eq. (11)}$$

$$v_i = \sum_{j=1}^j e_{ij}^2 \quad \text{Eq. (12)}$$

A SD representa a localização de uma amostra dentro do espaço definido pelas componentes principais, enquanto a OD é calculada como a soma dos quadrados dos resíduos associados ao modelo PCA, refletindo a distância da amostra em relação ao espaço que define a classe alvo. A combinação de SD e OD é descrita em termos da variável c , conhecida como

distância global (c) (Eq. 6), cujos valores seguem uma distribuição qui-quadrado (Eq.13) (Rodionova; Titova; Pomerantsev, 2016; Zontov, 2017).

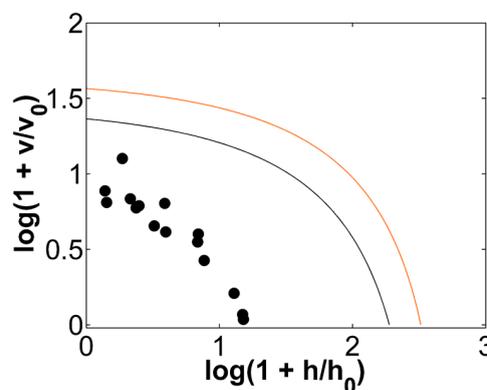
$$c = N_h \frac{h}{h_0} + N_v \frac{v}{v_0} aX^2(N_v + N_h) \quad \text{Eq. (13)}$$

Os parâmetros h_0 e v_0 são fatores de normalização das distâncias dos escores e ortogonal, enquanto N_h e N_v representam os valores para os graus de liberdade para SD e OD, respectivamente (Rodionova; Titova; Pomerantsev, 2016). Finalizada a segunda etapa, na terceira é estabelecida a área de aceitação ou limite para a classe alvo com base no valor do erro tipo I (α), que representa a taxa de falsa rejeição quando amostras da classe alvo são erroneamente consideradas não pertencentes a ela. A delimitação dessa área de aceitação segue a regra de decisão que é determinada pela condição $c \leq c_{\text{crit}}$ (Eq.14), pois se uma amostra tiver uma medida de distância global (c) menor que a distância crítica (c_{crit}), ela é classificada como pertencente à classe alvo (Eq. 7)(Neves; Poppi, 2020).

$$c \leq c_{\text{crit}}(\alpha) = X^{-2}(1 - \alpha, N_h + N_v) \quad \text{Eq. (14)}$$

O modelo construído é visualizado por meio de uma área de aceitação no gráfico (Figura 15) que tem como eixos coordenados: a distância ortogonal e a distância dos *scores* (gráfico de aceitação), determinado pelo valor α . Cada amostra do conjunto de treinamento é posicionada nas novas coordenadas do gráfico de aceitação, sendo classificada como pertencente (Dentro da área) ou não à classe alvo (Fora da área) com base nessa visualização (Máquina *et al.*, 2021).

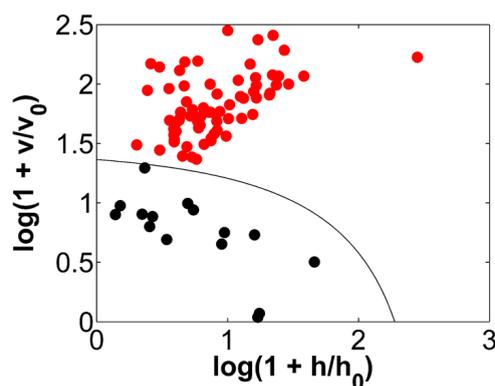
Figura 15. Gráfico do modelo DD-SIMCA na etapa de treinamento.



Fonte: Autoria própria, 2024.

Por fim, temos a última etapa que consiste na fase de teste do conjunto de dados (**Figura 16**). Nessa etapa, novas amostras são usadas para avaliar o modelo desenvolvido, independentemente de pertencerem ou não à classe alvo. Para classificar as amostras previstas, define-se uma área de aceitação (*threshold*), representada por uma linha (-), que separa as amostras da classe de interesse (●) das amostras da classe de não interesse (●) (Achuate, 2023). Dessa forma, amostras desconhecidas são inseridas no modelo e classificadas conforme a sua classe, ou seja, se pertencem ou não à classe de interesse. O resultado dessa avaliação gera o gráfico de aceitação correspondente. Quando se utiliza uma classe alternativa, existe a possibilidade de calcular o erro tipo II (β), que representa a taxa de falsa aceitação quando amostras não autênticas são erroneamente classificadas pertencentes a classe alvo (Neves & Poppi, 2020).

Figura 16. Gráfico do modelo DD-SIMCA na etapa de teste.



Fonte: Autoria própria, 2024.

De maneira geral, o DD-SIMCA possui várias vantagens distintas, pois ao adotar uma abordagem baseada em dados avalia as distribuições de distâncias e estabelece uma associação com níveis de significância extremos e discrepantes, além de proporcionar métodos de estimativa dupla tanto para dados normais quanto para dados contaminados, o que garante a robustez do modelo (Pomerantsev & Rodionova, 2014b; Achuate, 2023).

3.6.6 Procrustes cross validation - PCV

Visando resolver alguns problemas enfrentados pela estratégia de validação cruzada, Kucheryavskiy *et al.* (2020) propuseram a utilização de um conjunto de teste “simulado”, baseado em um conjunto de calibração, mas com a adição de ruído aleatório. Neste estudo, é apresentado uma abordagem alternativa, que se baseia na medição da variação entre modelos

de validação cruzada locais (representada por ângulos entre subespaços de variáveis latentes) e na incorporação dessa variação ao conjunto de calibração, criando assim um novo conjunto de dados — o conjunto pseudo-validação (PV). Esta abordagem foi denominada de validação cruzada Procrustes.

Segundo os autores, o conjunto pseudo-validação pode ser utilizado de maneira semelhante a um conjunto de teste independente, permitindo a validação do modelo global. Os resultados de desempenho do modelo são equivalentes aos obtidos com a validação cruzada convencional, contudo não há necessidade de repetir as iterações de validação cruzada mais de uma vez. Outrossim, um aspecto importante da abordagem proposta é a capacidade de avaliar a qualidade da validação cruzada ao analisar a distribuição dos novos valores e compará-la com a distribuição dos valores do conjunto de calibração (Kucheryavskiy *et al.*, 2020).

Primeiramente, um modelo PCA global é criado usando o conjunto de calibração \mathbf{X} (**Figura 17a**). A linha preta espessa representa o espaço das componentes principais S , que neste caso consiste em apenas uma componente principal. Em seguida, para cada iteração de validação cruzada k , o conjunto de calibração é dividido em $\tilde{\mathbf{X}}_k$ e \mathbf{X}_k , e usando $\tilde{\mathbf{X}}_k$ um modelo local é criado (**Figura 17b**). A diferença entre o modelo local S_k e o modelo global S pode ser medida como um ângulo β entre as componentes principais correspondentes. Consideramos esse ângulo como uma estimativa do erro de amostragem — a incerteza nos parâmetros do modelo devido à variação entre amostras (Kucheryavskiy *et al.*, 2020).

Esse erro pode ser introduzido no conjunto de calibração se o objeto de \mathbf{X}_k for rotacionado no sentido contrário ao erro ($-\beta$, **Figura 17c**). O objeto rotacionado faz parte do novo conjunto chamado de conjunto de pseudo-validação (\mathbf{X}_{pv}). Para completar o conjunto de pseudo-validação, é preciso realizar todas as iterações no procedimento de validação cruzada. No caso de validação cruzada completa, cada objeto é rotacionado para seu próprio ângulo, conforme mostrado na **Figura 17c**. Para a validação cruzada segmentada, objetos do mesmo segmento são rotacionados pelo mesmo ângulo.

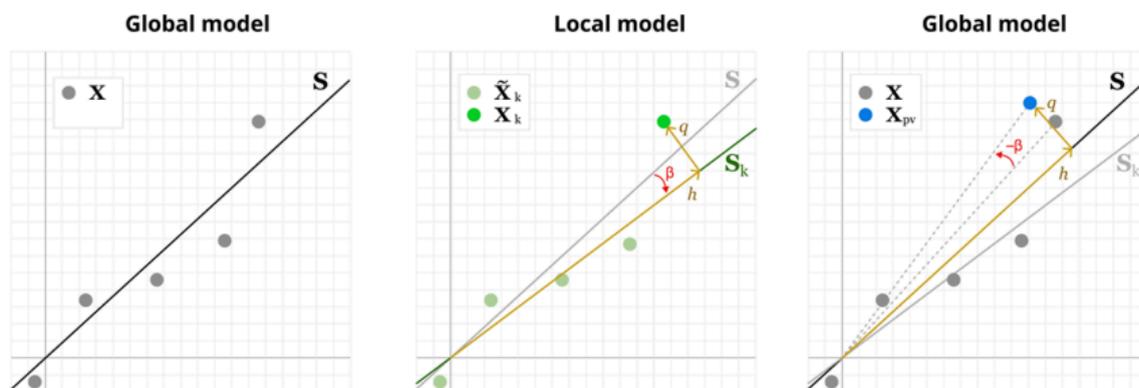
O conceito do método, ilustrado na **Figura 17**, pode ser generalizado da seguinte forma. Para um procedimento de validação cruzada com um número de segmentos K e um número de componentes A , em cada etapa $k = 1, \dots, K$, execute o seguinte:

1. Exclua o k -ésimo segmento dos dados para criar um conjunto de calibração local $\tilde{\mathbf{X}}_k$.

2. Use \tilde{X}_k para desenvolver um modelo local, representado pela matriz de cargas P_k .
3. Calcule a matriz de rotação $J \times J R_k$, tal que $P_k = R_k P$. A matriz mantém o comprimento dos vetores no espaço original e os ângulos entre eles inalterados.
4. Aplique a matriz ao segmento excluído X_k : $X_{pvk} = X_k R_k^T$.

Finalmente, todas as matrizes X_{pvk} são combinadas para formar o conjunto de pseudo-validação X_{pv} : $X_{pv} = (X_{pv1}, X_{pv2}, \dots, X_{pvK})^T$.

Figura 17. Ilustração do conceito de conjunto de pseudo-validação.



Fonte: Kucheryavskiy *et al.*, 2020.

Procrustes, também conhecido como "*Procrustes*", era um personagem mitológico que vivia na serra de Elêusis. Em sua casa, ele possuía uma cama de ferro feita sob medida, à qual convidava todos os viajantes para se deitarem. Se os hóspedes fossem muito altos, ele amputava o excesso para ajustá-los ao tamanho da cama; já aqueles de estatura baixa eram esticados até alcançarem o comprimento necessário (Brandão, 1987). A palavra significa "o esticador", em referência ao castigo que o referido malfeitor aplicava às suas vítimas. Pode-se imaginar que, em cada etapa, busca-se uma maneira de ajustar o modelo local atual ao "leito de Procrustes" do modelo global, daí o nome (Kucheryavskiy *et al.*, 2020).

Os resultados obtidos pelos autores confirmaram que o conjunto de pseudo-validação, criado com a abordagem de validação cruzada Procrustes proposta, é uma ferramenta eficiente para validar modelos PCA/SIMCA durante a etapa de otimização. Por um lado, ele pode ser utilizado de forma análoga a um conjunto de teste independente, oferecendo um conjunto completo de resultados com base na validação de um único modelo, incluindo escores, distâncias e variância explicada/residual, que pode ser considerado uma alternativa para um número de amostras reduzido. Por outro lado, os autores destacam que estatísticas de

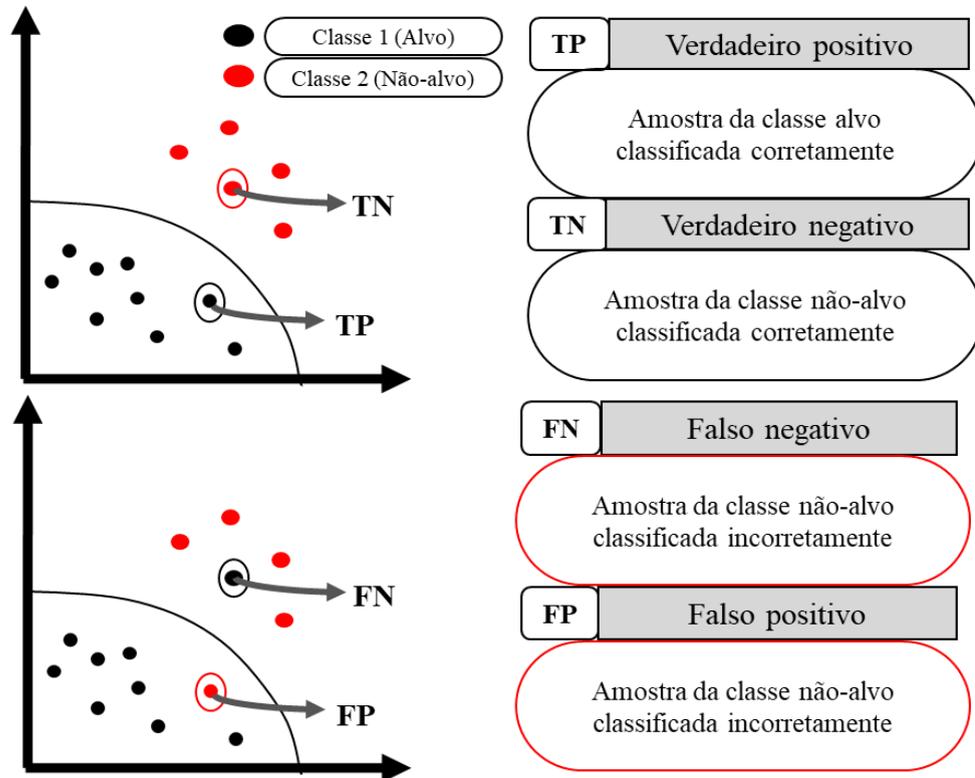
desempenho obtidas com o conjunto de pseudo-validação (como o número de objetos classificados como extremos) são comparáveis às obtidas através da validação cruzada convencional, ou seja, não há uma melhoria no resultado. Entretanto, resulta em resultados reprodutíveis e comparáveis, facilitando a adoção da abordagem proposta.

3.6.7 Métricas de Desempenho para modelos de classificação

Para avaliar o desempenho dos modelos de classificação, é comum na literatura utilizar parâmetros como Sensibilidade, Especificidade e Taxa de Classificação Correta (Acurácia). Geralmente, para obter esses valores, os resultados dos métodos de classificação multivariada são apresentados em uma tabela de confusão, contingência ou matriz de classificação. Nessa tabela, são apresentados os resultados das classificações, comparando a classe modelada com a classe não modelada. Normalmente, as amostras da classe verdadeira ou rotulada são representadas pelas linhas na tabela, enquanto as colunas da matriz de confusão mostram os resultados das amostras atribuídas pelo classificador do modelo.

Os parâmetros de desempenho são definidos com base nas probabilidades que surgem a partir de quatro cenários possíveis: (i) amostras da classe alvo corretamente atribuídas à classe alvo (verdadeiro positivo - TP, do inglês *true positive*), (ii) amostras da classe não alvo corretamente atribuídas à classe não alvo (verdadeiro negativo - TN, do inglês *true negative*), (iii) amostras da classe alvo incorretamente atribuídas à classe não alvo (falso negativo - FN, do inglês *false negative*), e (iv) amostras da classe não alvo incorretamente atribuídas à classe alvo (falso positivo - FP, do inglês *false positive*). A **Figura 18**, apresenta uma representação gráfica da disposição das amostras com relação aos cenários possíveis: TP (verdadeiro positivo), TN (verdadeiro negativo), FP (falso positivo) e FN (falso negativo).

Figura 18. Representação gráfica das métricas de avaliação de desempenho do modelo de classificação DD-SIMCA (a).



Fonte: Autoria própria, 2024.

A sensibilidade (SEN), por exemplo, reflete a eficácia do modelo em classificar corretamente as amostras da classe alvo e pode ser descrita como a taxa de verdadeiros positivos. Para o conjunto de treinamento e teste, a sensibilidade é calculada conforme:

$$SEN = \frac{TP}{TP + FN} \times 100 \quad \text{Eq. (17)}$$

No caso da especificidade (SPC), é avaliado a capacidade do modelo de identificar corretamente as amostras da classe não alvo (verdadeiros negativos). Vale lembrar que a especificidade do modelo DD-SIMCA só pode ser calculada com base nas amostras da classe não alvo, e, portanto, não há valores de especificidade para os resultados de treinamento (Zontov *et al.*, 2017). Sendo expressa, matematicamente por:

$$SPC = \frac{TN}{FP + TN} \times 100 \quad \text{Eq. (18)}$$

Para avaliar o desempenho global dos modelos qualitativos, a acurácia (ACC) é o parâmetro estatístico utilizado. Ela é calculada dividindo o número de amostras classificadas corretamente (independentemente da classe, TP e TN) pelo número total de amostras, ou seja, a soma das quantidades de todos os quatro cenários (Santana *et al.*, 2020).

$$ACC = \frac{TP + TN}{TP + FN + FP + TN} \times 100 \quad \text{Eq. (19)}$$

Um parâmetro que resume a eficiência (EFF) do modelo é frequentemente calculado como a média geométrica dos valores de sensibilidade e especificidade. Esse parâmetro pode variar de 0, quando a sensibilidade ou a especificidade é nula, até 1, quando ambos os parâmetros atingem o valor máximo de 1 (Oliveri & Downey, 2012).

$$EFF = \sqrt{\frac{TP \times TN}{(TP + FN)(FP + TN)}} \times 100 \quad \text{Eq. (20)}$$

Capítulo 4

Metodologia

4 METODOLOGIA

4.1 AQUISIÇÃO DAS AMOSTRAS

Foram adquiridas amostras de leite caprino em pó industrializado de 10 lotes diferentes, juntamente com quatro tipos de possíveis adulterantes: Leite em pó bovino, composto lácteo, amido de milho e extratos de soja. Todas as amostras foram transportadas ao laboratório e armazenadas em temperatura ambiente. Os 10 lotes de leite caprino em pó foram misturados aleatoriamente em combinações binárias e ternárias para atingir 60 amostras. Em seguida, um total de 288 misturas foram preparadas adicionando diferentes 9 proporções dos adulterantes supramencionados (**Tabela 3**), sendo preparadas em duplicatas autênticas e homogêneas em tubos falcon 50 mL por inversões sucessivas. Para todas as amostras a massa total foi de 5 g.

Tabela 3. Níveis de teores de adulteração utilizados.

%	Adulterante (g)	Leite de cabra (g)
1	0,05	4,95
3	0,15	4,85
5	0,25	4,75
7	0,35	4,65
10	0,50	4,50
15	0,75	4,25
20	1,00	4,00
25	1,25	3,75
30	1,50	3,50

Fonte: Autoria própria, 2024.

Para evitar a deterioração das amostras, estas foram então armazenadas em sacos plásticos, lacradas e colocadas em dessecador a vácuo em temperatura ambiente até as análises (**Figura 19**). Posteriormente, as amostras foram dispostas em placas de Petri lisas de poliestireno (60 × 15 mm), numa fina camada de aproximadamente um terço da placa, para aquisição dos espectros NIR (Experimento 1) e RAMAN (Experimento 2). Vale salientar que, enquanto o experimento 1 seguiu o estudo com os quatro adulterantes mencionados (288 amostras no total), o experimento 2 teve o número de amostras reduzido (**Tabela 4**), pois

neste utilizou-se apenas um único adulterante, especificamente o leite bovino em pó (68 adulteradas + 15 puras = 83 amostras totais).

Tabela 4. Quantidade de amostras por nível de adulteração em cada experimento.

<i>Experimentos</i>	1	2
<i>Níveis (%)</i>	Quant.	Quant.
1	32	8
3	32	7
5	32	8
7	32	7
10	32	8
15	32	7
20	32	7
25	32	8
30	32	8
<i>Total</i>	288 amostras	68 amostras

Fonte: Autoria própria, 2024.

Figura 19. Esquema ilustrativo do processo de aquisição das amostras



Fonte: Autoria própria, 2024.

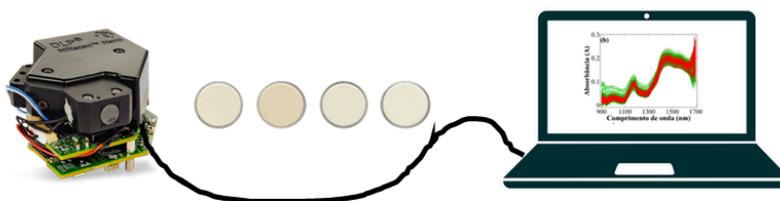
4.2 NIR

4.2.1 Aquisição dos Espectros

Para o experimento 1, os espectros NIR foram adquiridos na faixa de 900 a 1700 nm usando um instrumento NIR portátil, modelo DLP NIRscan Nano Evaluation Module (Texas Instruments, Dallas, TX) acoplado a um notebook (**Figura 20**). O registro dos espectros de absorvância foi realizado usando o software integrado, em resolução digital de 228 e integrando 32 varreduras com transformada de Hadamard. Todas as medições foram feitas a

uma temperatura ambiente, com a limpeza da placa de petri sendo realizada entre uma medida e outra. Um espectro do branco foi adquirido para compensar variações ambientais entre cada amostra, utilizando pó de politetrafluoretileno (PTFE) como referência. Os espectros médios das duplicatas foram utilizados para construção dos modelos quimiométricos.

Figura 20. Esquema ilustrativo do processo de aquisição dos Espectros NIR.



Fonte: Autoria própria, 2024.

4.2.2 Seleção de amostras

As amostras da classe alvo foram divididas em conjuntos de treinamento (40 amostras) e teste (20 amostras), sendo selecionadas através do algoritmo Kennard-Stone a partir da interface Data Hand Gui, em Matlab® versão R2010a (Galvão *et al.*, 2005). Posteriormente gerou-se um conjunto de “teste completo”, combinando as amostras da classe não-alvo (288 amostras) com as 20 amostras da classe alvo selecionadas pelo algoritmo supramencionado.

4.3 RAMAN

4.3.1 Aquisição dos Espectros

Para o experimento 2, os espectros de dispersão Raman foram adquiridos utilizando um espectrômetro Raman portátil, modelo Mira DS (Metrohm®), que mede 68 mm de altura e 88 mm de largura externamente e pesa 705 gramas, acoplado a um notebook (**Figura 21**). As medições foram registradas com o software Mira Cal DS, cobrindo o intervalo espectral de 2300 a 400 cm^{-1} . O laser opera dentro de uma faixa de potência de 5 a 500 mW, classificado como Classe 3B, oferecendo uma resolução espectral de 8 – 10 cm^{-1} , emitindo em um comprimento de onda de 785 nm, com um tempo de integração de 2500 ms. Todas as medições foram feitas a uma temperatura ambiente, com a limpeza da placa de petri sendo realizada entre uma medida e outra. Um espectro do branco foi adquirido para compensar

variações ambientais entre cada amostra, utilizando pó de politetrafluoretileno (PTFE) como referência. Os espectros médios das duplicatas foram utilizados para construção dos modelos quimiométricos.

Figura 21. Esquema ilustrativo do processo de aquisição dos Espectros RAMAN.



Fonte: Autoria própria, 2024.

4.3.2 Seleção de amostras

O experimento 2 seguiu três abordagens para a seleção de amostras, uma vez que para a validação, construção e avaliação dos modelos DD-SIMCA foram utilizadas diferentes estratégias, conforme descrito na **Tabela 5**.

Tabela 5. Diferentes estratégias de particionamento de amostras para construir e validar os modelos DD-SIMCA

Figuras de mérito	Conjunto de treino	Conjunto de teste	
	Sensibilidade	Sensibilidade	Especificidade
Abordagem 1	15 Amostras alvo	-	68 Amostras não-alvo
Abordagem 2	15 Amostras alvo	15 Amostras alvo PCV	68 Amostras não-alvo
Abordagem 3	10 Amostras alvo (KS)	5 Amostra alvo (KS)	68 Amostras não-alvo

PCV: Algoritmo procrustes cross-validation; KS: Algoritmo de particionamento de amostras Kennard-Stone.

Fonte: Autoria própria, 2024.

Posteriormente gerou-se um conjunto de “teste completo”, combinando as amostras da classe não-alvo (68 amostras) com as amostras da classe alvo selecionadas de acordo com a **Tabela 5**.

4.4 PROCEDIMENTO QUIMIOMÉTRICO E SOFTWARE

Considerando que a estratégia adotada neste estudo consistiu em utilizar técnicas de reconhecimento de padrões em que apenas a classe alvo é modelada, simulando situações reais em que não é possível caracterizar a classe de amostras adulteradas, foram consideradas apenas duas classes para fins de classificação. A classe alvo foi composta por amostras puras de leite caprino, enquanto a outra classe foi formada por amostras adulteradas, ou seja, as misturas.

Para o tratamento dos dados obtidos em ambos os experimentos, foram empregadas diversas ferramentas quimiométricas, incluindo PCA, DD-SIMCA e OC-PLS. A modelagem desses métodos foi realizada com os *softwares* The Unscrambler® versão 9.7, Matlab® versão R2010a e Matlab® 2018b. A PCA foi aplicada para auxiliar na identificação de possíveis outliers, das quais nenhuma foi detectada. Já o DD-SIMCA e o OC-PLS foram utilizados para desenvolver modelos de classificação das classes pura e adulterada com base nos espectros obtidos.

4.4.1 Tratamento dos espectros NIR

Inicialmente, o intervalo espectral entre 1660 e 1700 nm, que não contém informações relevantes no infravermelho próximo, foi eliminado. Além disso, para lidar com problemas decorrentes de variações sistemáticas na linha de base, foram aplicadas técnicas diferentes de pré-processamento e combinações destas, ao intervalo espectral informativo de 900 a 1660 nm: (a) Suavização com filtro Savitzky-Golay com uma janela de 21 pontos (SGS-21p), SGS-21p acoplada com (b) correção de deslocamento de linha de base (OFF), (c) correção multiplicativa de dispersão (MSC), (d) transformação de variância normal padrão (SNV); e por fim (e) Derivadas de Savitzky-Golay com um polinômio de segunda ordem (Primeira com janela de 25 pontos e Segunda com 25 pontos) (SGD).

O número adequado de PCs para o DD-SIMCA e de variáveis latentes (LVs) para o OC-PLS foi determinado ao examinar os valores de sensibilidade do modelo no conjunto de treinamento. A avaliação de desempenho dos modelos concentrou-se na sensibilidade, especificidade e eficiência. Todos os procedimentos quimiométricos foram realizados utilizando os *softwares* The Unscrambler® versão 9.7 e Matlab® 2010a. PCA e KS foram

executados utilizando rotinas desenvolvidas internamente, enquanto o DD-SIMCA foi encontrado em <https://github.com/yzontov/dd-simca>.

4.4.2 Tratamento dos espectros RAMAN

Inicialmente, o intervalo espectral entre 2300 e 1800 cm^{-1} , que não contém deslocamentos Raman, foi eliminado. Além disso, para lidar com problemas decorrentes de variações sistemáticas na linha de base, foram aplicadas quatro técnicas diferentes de pré-processamento ao intervalo espectral informativo de 1800 a 400 cm^{-1} : (a) correção de deslocamento de linha de base offset (OFF), (b) correção multiplicativa de dispersão (MSC), (c) transformação de variância normal padrão (SNV) e (d) derivada primeira de Savitzky-Golay com um polinômio de segunda ordem e uma janela de 25 pontos (SGD).

O estudo explorou três estratégias distintas de validação para a construção e avaliação da capacidade preditiva dos modelos DD-SIMCA, conforme descrito na **Tabela 5**. Em seguida, os modelos DD-SIMCA foram desenvolvidos empregando níveis de significância para detecção de outliers de 0,01 e 0,05, com regiões de aceitação determinadas pelo modo robusto da distribuição qui-quadrado. A otimização do número de componentes principais (PCs) no DD-SIMCA foi orientada pelo critério de alcançar o equilíbrio ideal de casos verdadeiramente positivos em ambos os conjuntos de treinamento e teste.

A avaliação de desempenho dos modelos concentrou-se na sensibilidade, especificidade e eficiência. Todos os procedimentos quimiométricos foram realizados utilizando o Matlab® 2018b (Mathworks Inc.). PCA e KS foram executados utilizando rotinas desenvolvidas internamente, enquanto o DD-SIMCA foi encontrado em <https://github.com/yzontov/dd-simca> e PCV em <https://github.com/svkucheryavski/pcv>.

Capítulo 5

Resultados e discussão

5 RESULTADOS E DISCUSSÃO

5.1 NIR

5.1.1 Análise Exploratória

Os espectros das 348 amostras de leite caprino puro e das misturas possuem considerável sobreposição entre ambas as classes, como pode ser observado na **Figura 22**. Uma banda proeminente e ampla em torno de 1450 nm é atribuída ao primeiro harmônico da vibração de estiramento O-H. É possível observar ainda a presença de bandas de absorção em cerca de 1600 nm, que são atribuídas ao primeiro harmônico da extensão livre N-H/cominação amida II, referentes a presença de proteínas. Além disso, pequenos picos entre 1100 e 1200 nm correspondem ao segundo sobretom harmônico do estiramento C-H atribuído à gordura (Núñez-Sánchez *et al.*, 2016; Dupont *et al.*, 2018; Pereira *et al.*, 2020).

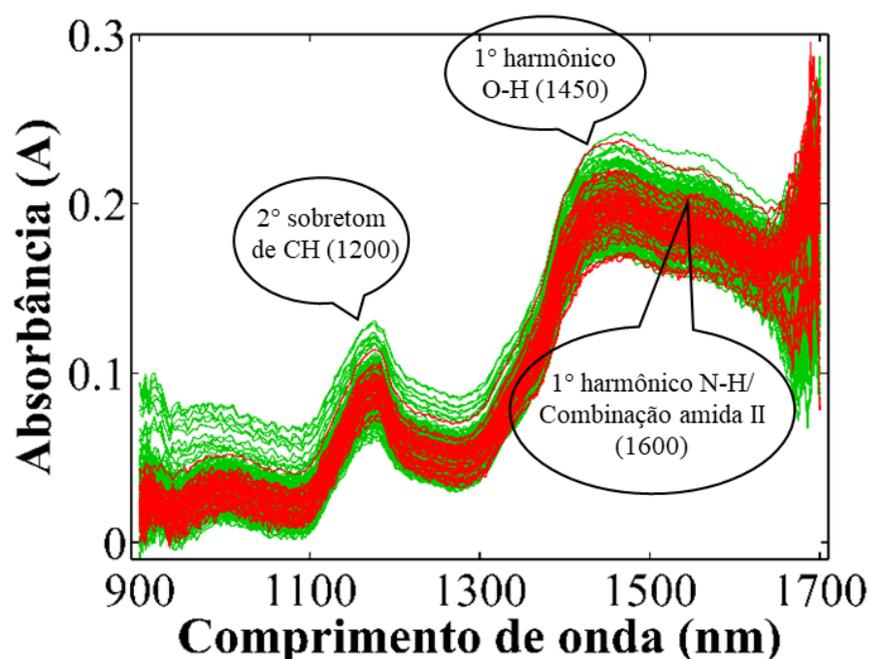


Figura 22. Espectros NIR de leites caprino (em vermelho) e de misturas caprino + adulterantes (em verde): espectro bruto.

Os espectros apresentam saturação espectral do detector do instrumento, ruído e deslocamento sistemático na linha de base, tais problemas podem ser atribuídos ao espalhamento de luz causado pelos tamanhos de partículas presentes nas amostras de leite caprino em pó e nos adulterantes. Assim sendo, a região saturada (1660 - 1700) foi removida para evitar interferências na construção dos modelos de classificação. Por fim, os espectros

foram corrigidos por meio dos seguintes pré-processamentos: correção de base offset em conjunto com suavização Savitzky-Golay com 21 pontos (SGS-21p-OFF), correção de dispersão multiplicativa em conjunto com suavização Savitzky-Golay com 21 pontos (SGS-21p+MSC), transformação variável normal padrão em conjunto com suavização Savitzky-Golay com 21 pontos (SGS-21p+SNV), além das técnicas de primeira e segunda derivada Savitzky-Golay utilizando polinômios de segunda ordem e janelas com tamanhos de 21 (SGD-21p) e 25 (SGD2-25p) pontos, conforme ilustrado na **Figura 23**, por meio dos espectros médios.

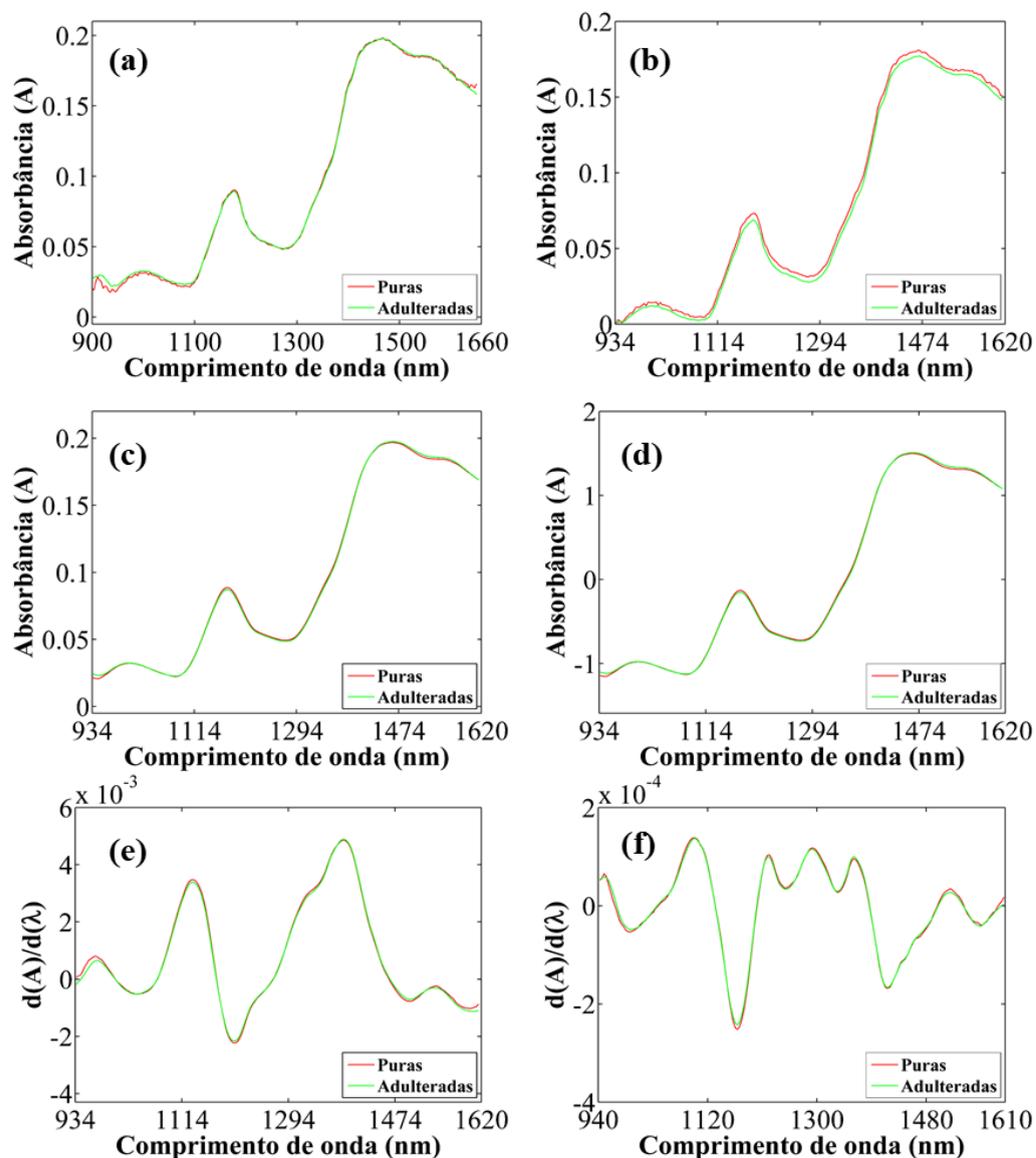


Figura 23. Espectros NIR médios das classes: Leite caprino puro (Vermelho), Leite caprino adulterados (Verde): Espectros brutos cortados a partir de 1660 nm (a) e suavizados com filtro Savitzky – Golay 21 pontos acoplado com (b) OFF, (c) MSC e (d) SNV; aplicando filtro derivativos de 1º e 2º ordens Savitzky – Golay com janelas de 21 (e) e 25 (f) pontos.

Após o pré-processamento dos espectros, uma análise exploratória dos dados foi realizada utilizando a PCA. Os resultados gráficos bidimensionais são apresentados através dos escores de PC1 e PC2 como pode ser visto nas **Figura 24** para os dados cortados e para os seguintes pré-processamentos SGS-21p-OFF, SGS-21p+MSC, SGS-21p+SNV, SGD-21p e SGD2-25p, respectivamente.

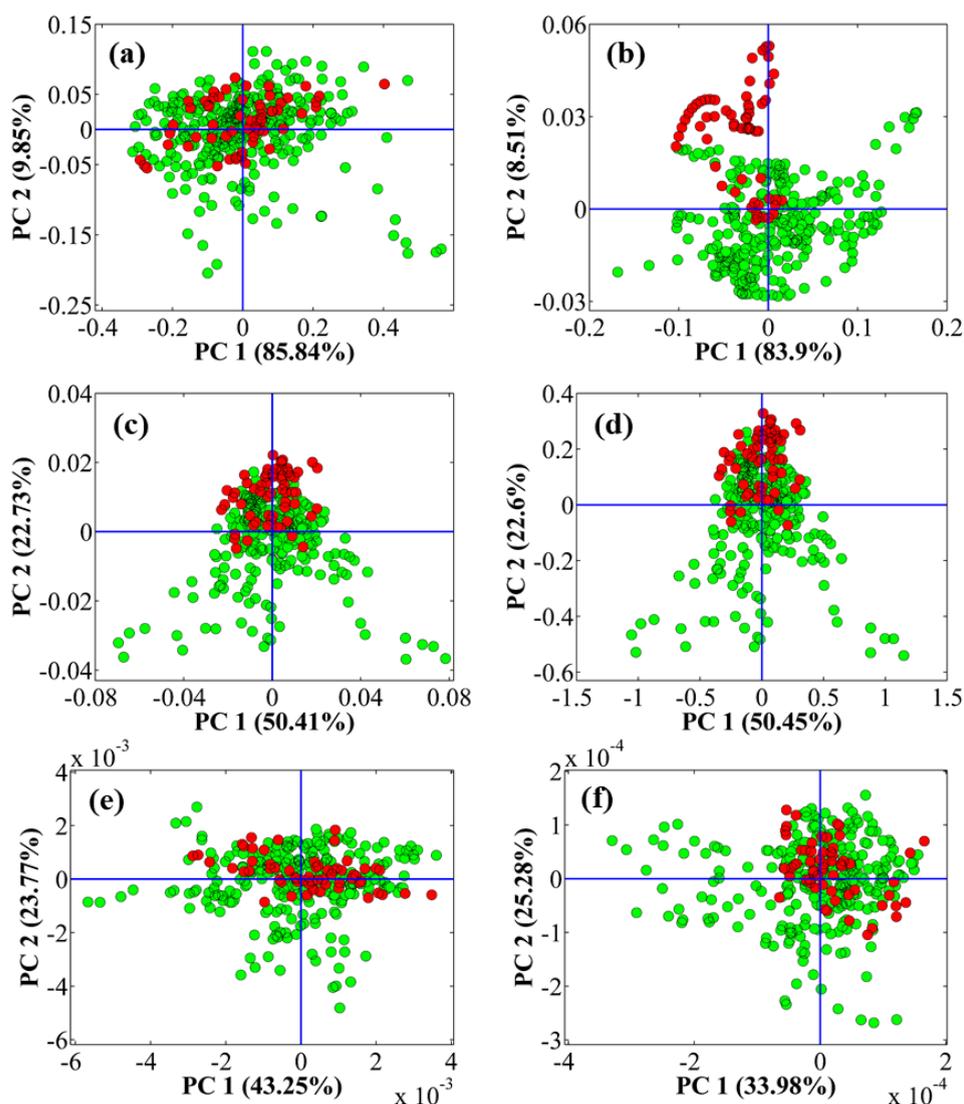


Figura 24. Amostras de Leites caprino (em vermelho) e de misturas caprino + adulterantes (em verde): Gráficos de escores (PC1 x PC2) para os dados espectrais cortados a partir de 1660 nm (a) e suavizados com filtro Savitzky – Golay 21 pontos acoplado com (b) OFF, (c) MSC e (d) SNV; aplicando filtro derivativos de 1° e 2° ordens Savitzky – Golay com janelas de 21 (e) e 25 (f) pontos, respectivamente.

Analisando os gráficos de escores obtidos por meio de PCA acima, observa-se forte sobreposição entre as amostras de leite de cabra puro e as demais adulteradas. Contudo, vale destacar a existência de uma tendência de separação para o pré-processamento SGS-21p-OFF. Na **Figura 24b**, o gráfico de PCA utilizando os dados SGS-21p-OFF, destaca-se que mesmo a

PC1 apresentando a maior variância explicada a tendência de separação ocorre ao longo de PC2 entre as classes. Para ampliar essa discussão, gráficos dos escores de PC1 e PC2, rotulando as amostras por adulterante e por teor, são apresentados na **Figura A1** do apêndice.

Por apresentarem valores consideravelmente altos de variância explicada, analisou-se o potencial da PC3 (**Figura 25**) na visualização de agrupamentos naturais entre as classes, entretanto não houveram contribuições significativas para classificação em nenhuma técnica de pré-processamento.

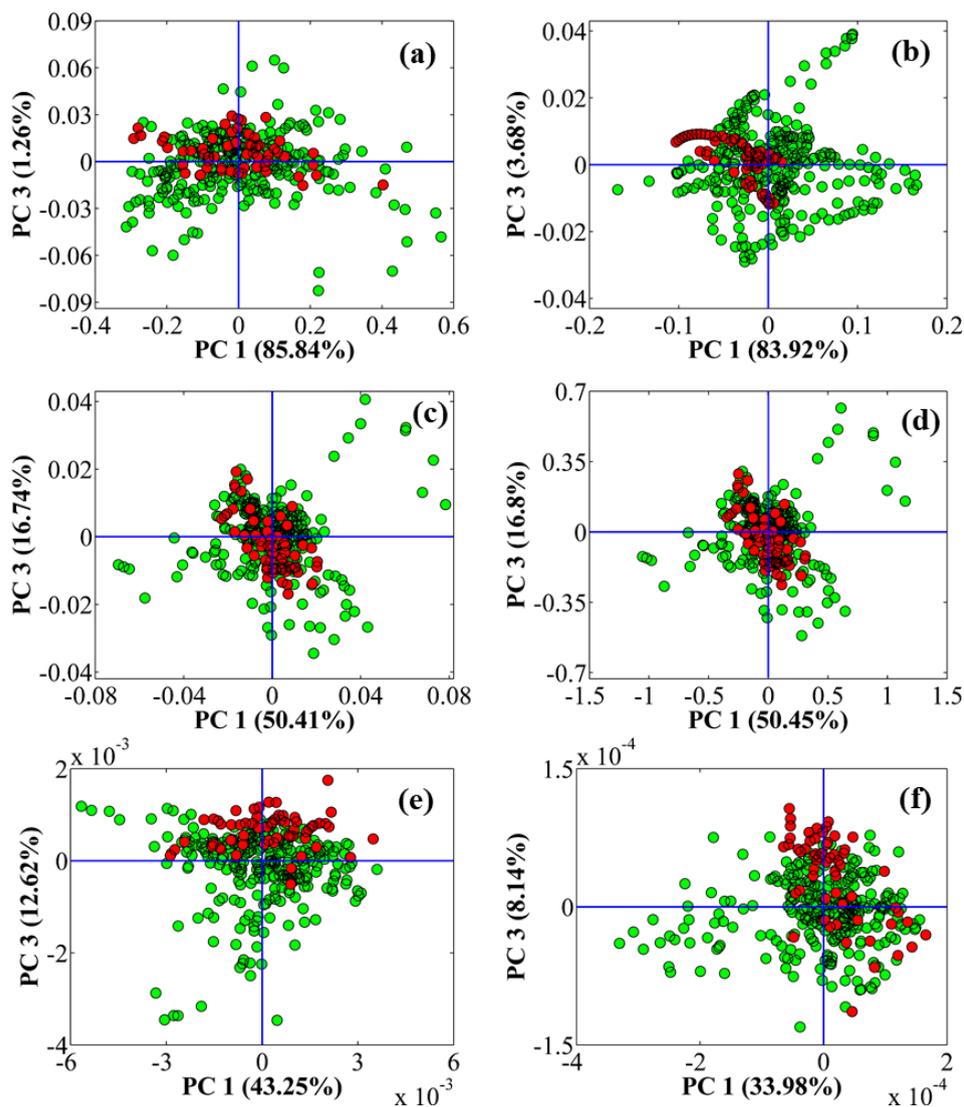


Figura 25. Amostras de Leites caprino (em vermelho) e de misturas caprino + adulterantes (em verde): Gráficos de escores (PC1 x PC3) para os dados espectrais cortados a partir de 1660 nm (a) e suavizados com filtro Savitzky – Golay 21 pontos acoplado com (b) OFF, (c) MSC e (d) SNV; aplicando filtro derivativos de 1° e 2° ordens Savitzky – Golay com janelas de 21 (e) e 25 (f) pontos, respectivamente.

Uma análise adicional do gráfico de pesos, utilizando espectros NIR (**Figura 26**), confirma a importância das bandas associadas aos compostos macromoleculares

(especialmente gorduras, proteínas e carboidratos) na análise das classes e na possível contribuição para a separação. As linhas preta, vermelha e azul correspondem aos pesos de PC1, PC2 e PC3, respectivamente, obtidos para os dados cortados (**Figura 26a**), SGS-21p-OFF (**Figura 26b**), SGS-21p+MSC (**Figura 26c**), SGS-21p+SNV (**Figura 26d**), SGD-21p (**Figura 26e**) e SGD2-25p (**Figura 26f**). Essas três primeiras PCs respondem por 97,2%; 96,1%; 89,9%; 81,8%; 79,6% e 67,4% da variância total explicada dos dados, respectivamente.

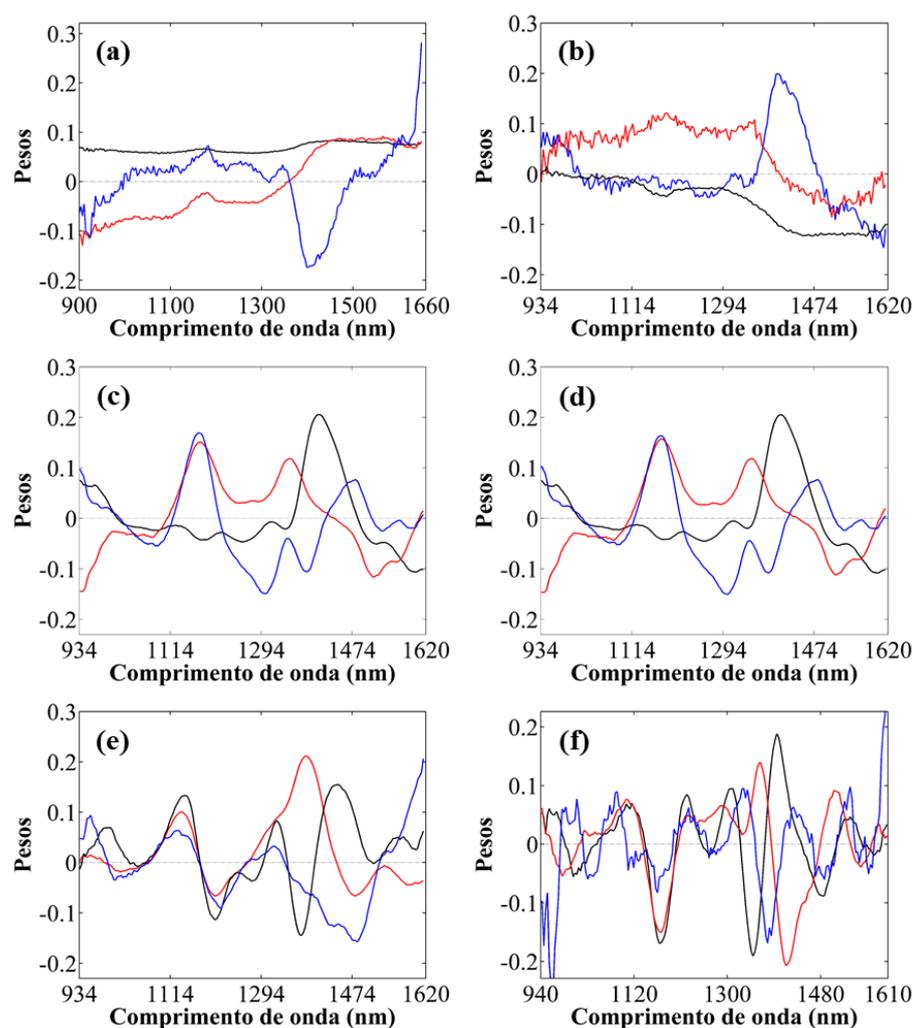


Figura 26. Gráficos dos pesos, PC1 (Preto), PC2 (Vermelho) e PC3 (Azul): para os dados espectrais cortados a partir de 1660 nm (a) e suavizados com filtro Savitzky – Golay 21 pontos acoplado com (b) OFF, (c) MSC e (d) SNV; aplicando filtro derivativos de 1° e 2° ordens Savitzky – Golay com janelas de 21 (e) e 25 (f) pontos, respectivamente.

De maneira geral, os gráficos de escores PCA para o pré-processamento SGS-21p-OFF ressaltou um potencial de separação entre amostras de leite de cabra puro e aquelas adulteradas, já as outras técnicas de pré-processamentos exibiram sobreposições

significativas entre as classes estudadas. Ademais, é observado que o tipo de pré-processamento aplicado nos dados influenciou na dispersão das amostras em torno das duas primeiras componentes principais. Contudo, a aplicação da PCA, como uma ferramenta não supervisionada, fornece apenas uma análise qualitativa prévia, servindo como uma triagem. Dessa forma, técnicas de reconhecimento de padrões supervisionadas como a modelagem *one class*, DD-SIMCA e OC-PLS, especificamente, foram empregadas para fins de autenticação, conforme discutido na seção seguinte.

5.1.2 Autenticação do Leite em pó caprino

A **Tabela 6** apresenta os resultados obtidos com DD-SIMCA e OC-PLS para identificação de adulteração de leite caprino em termos de sensibilidade e especificidade usando os diferentes pré-processamentos.

Tabela 6. Resultados em termos de sensibilidade e especificidade para os modelos DD-SIMCA e OC-PLS nos conjuntos de treinamento e teste.

Modelos	Pré Processamentos	LVs/PCs	Treinamento	Teste		Eficiência (%)
			Sensib. (%)	Sensib.(%)	Espec. (%)	
OC-PLS	Bruto	11	95,0	100	29,3	65,3
	SGS_21p_OFF	10	97,5	100	95,1	97,5
	SGS_21p_MSC	11	97,5	100	22,3	60,1
	SGS_21p_SNV	10	100	100	29,3	66,4
	SGD_21p	10	95,0	100	41,1	73,0
	SGD2_25p	8	100	100	31,0	67,7
DD-SIMCA	Bruto	7	95,0	95,0	71,4	86,4
	SGS_21p_OFF	2	97,5	100	83,6	93,4
	SGS_21p_MSC	9	97,5	100	70,0	88,0
	SGS_21p_SNV	9	97,5	100	70,7	88,3
	SGD_21p	8	97,5	90,0	73,3	86,3
	SGD2_25p	6	97,5	100	55,0	81,2

De acordo com os resultados apresentados na **Tabela 6**, o algoritmo DD-SIMCA obteve melhor desempenho em termos de sensibilidade para os modelos construídos com SGS-21p-OFF, com $\alpha = 0,05$. A modelagem alcançou 97,5% para o conjunto de treinamento e 100% no conjunto de teste, evidenciando o reduzido número de erros nas classes alvo (amostras puras), enquanto a especificidade do conjunto de teste alcançou apenas 83.6%.

Os gráficos de área de aceitação e de extremos obtidos com algoritmo DD-SIMCA nos conjuntos de treinamento e teste usando o conjunto de dados pré-processados através da configuração SGS-21p-OFF são apresentados na **Figura 27**. A linha verde representa a borda

entre as amostras regulares e extremas; a linha vermelha é a fronteira da área de aceitação. Os círculos vermelhos e verdes representam amostras de leite caprino puras e adulteradas, respectivamente. O quadrado laranja representa as amostras consideradas extremas, evidenciando que o DD-SIMCA é sensível para identificar amostras anômalas, confirmando a robustez/viabilidade desse tipo de modelagem.

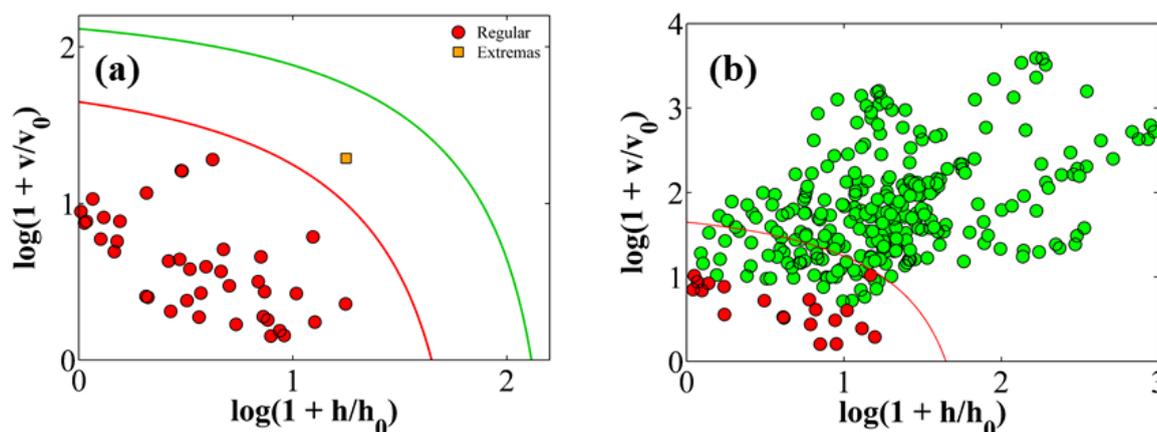


Figura 27. Gráficos de área de aceitação para o conjunto de treinamento (a) e teste (b) empregando DD-SIMCA

O rastreamento da amostra extrema no caso do modelo, e das amostras errôneas para o teste foi realizado, a fim de compreender a causa dos erros. Em primeiro momento, foi analisada a amostra extrema presente no treinamento do modelo, esta é referente ao espectro de resposta de uma mistura secundária de lotes do leite de cabra puro, lote 01 e 02, especificamente. Isto posto, não é possível atribuir motivação química a presença dessa amostra, uma vez que, baseando-se no conjunto de treinamento de 40 amostras há a presença de ambos os lotes puros e em outras misturas. Contudo, ao analisar o espectro dessa amostra específica é notória a presença de ruídos em bandas cruciais para a análise, estes ruídos podem ter surgido de diversas formas, como instabilidade na corrente elétrica no equipamento. Como ilustrado na **Figura 28**, os ruídos em torno da banda entre 1100 e 1200 configuram uma diferença considerável do espectro dessa amostra extrema quando comparado com o espectro médio das 39 amostras restantes, causando este deslocamento para fora da área de aceitação da classe.

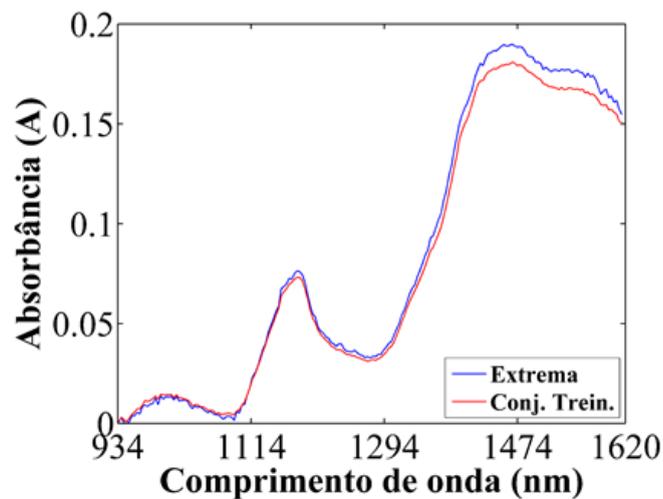


Figura 28. Espectros NIR SGS-21p-OFF: Amostra extrema (Azul) e médio do conjunto de treinamento (Vermelho).

No caso dos erros de classificação no teste do modelo, temos 16,4% referentes a 47 amostras das 288 pertencentes à classe não-alvo. Dentre estas, 25 amostras possuem baixo teor (1 a 5%) de adulteração, independente da classe do adulterante. Para médio teor (7 a 15%) um total de 12 amostras, e para alto teor (20 a 30%) surpreendentemente 10 amostras, o que leva a uma discussão sobre o tipo do adulterante. Levando em consideração cada adulterante, dentre as 47 amostras classificadas erroneamente, 26 amostras são envolvendo o leite de vaca, 10 amostras envolvendo amido, 5 e 6 para composto lácteo e extrato de soja, respectivamente. O número alto de amostras referentes a adulteração com leite de vaca, mostra como a semelhança entre os leites atrapalha o processo de classificação até nos níveis mais altos de adulteração, enquanto para os outros adulterantes os erros giraram em torno dos teores mais baixos. A **Figura 29** mostra os resultados do conjunto teste do modelo DD-SIMCA rotulando com cores cada tipo de adulterante.

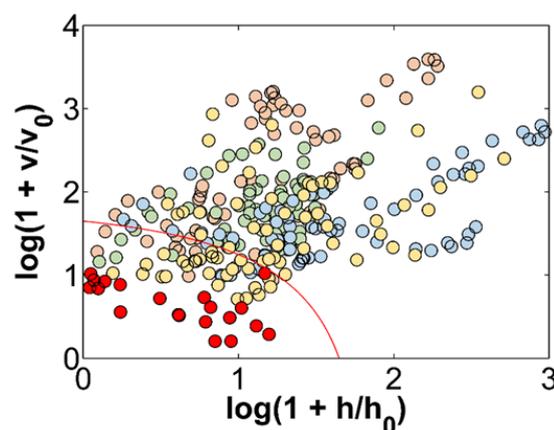


Figura 29. Gráficos de área de aceitação para o conjunto de teste, rotulando os tipos de adulterantes por cor.

Partindo para análise do modelo OC-PLS (**Figura 30**), também construído com SGS-21p-OFF, que na autenticação das amostras do leite caprino puro mostrou-se mais eficaz. O desempenho em termos de sensibilidade para os modelos construídos alcançou 97,5% para o conjunto de treinamento e 100% no conjunto de teste, evidenciando o reduzido número de erros nas classes alvo (amostras puras), enquanto a especificidade do conjunto de teste alcançou 95.1%.

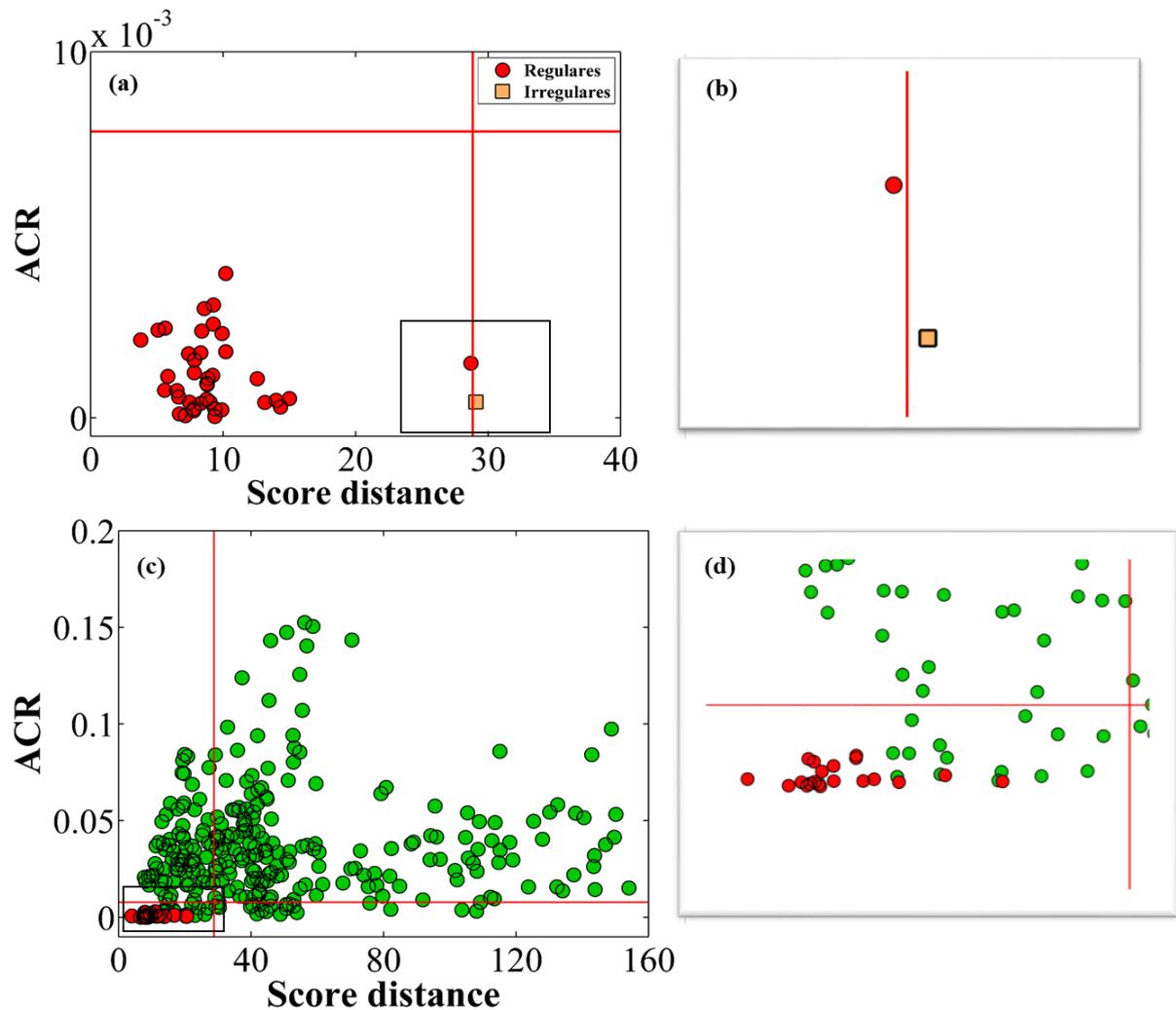


Figura 30. Gráficos de área de aceitação para o conjunto de treinamento (a) e teste (c) usando algoritmo OC-PLS. Magnificação da região destacada por um retângulo no canto inferior direito (b) e esquerdo (d).

Como ilustrado nos gráficos para o treinamento do modelo (**Figura 30**), apenas uma amostra também foi projetada fora da área de aceitação (quadrado laranja), que de maneira análoga ao modelo DD-SIMCA trata-se da amostra referente a mistura secundária dos lotes 01 e 02, já discutida. Sobre os erros de classificação no teste do modelo, temos apenas 4,9% referentes a 14 amostras das 288 pertencentes à classe não-alvo. Analisando independente da

classe do adulterante, tem-se 4 amostras para baixo teor (1 a 5%) e 5 amostras tanto para médio teor (7 a 15%), quanto para alto teor (20 a 30%). Analisando por composição de cada adulterante, repete-se a majoritariedade da quantidade de erros provenientes da adulteração com leite de vaca com 6 amostras, com a surpresa de que os 3 erros provenientes da adulteração por composto lácteo foram em nível médio e alto, mais uma vez mostrando como a semelhança entre os leites, e agora o composto lácteo nesta modelagem, atrapalham o processo de classificação até nos níveis mais altos. Para os outros adulterantes, amido (1 amostra), e extrato de soja (5 amostras), os erros continuaram girando em torno dos teores mais baixos. A **Figura 31** mostra os resultados do conjunto teste do modelo OC-PLS rotulando com cores cada tipo de adulterante.

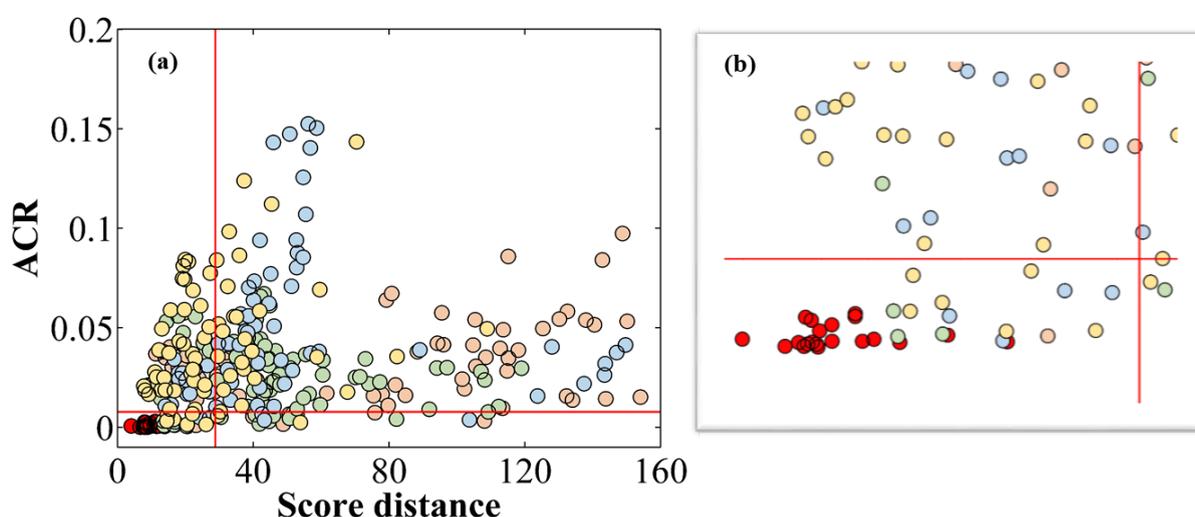


Figura 31. Gráficos de área de aceitação para o conjunto de teste (a) usando algoritmo OC-PLS. Magnificação da região destacada por um retângulo no canto inferior esquerdo (b).

5.2 RAMAN

5.2.1 Análise Exploratória

A **Figura 32** mostra uma comparação dos espectros Raman médios adquiridos a partir de amostras individuais de leite de cabra em pó puro (em azul), leite de vaca em pó puro (em verde) e leite de cabra em pó adulterado com leite de vaca (em roxo). Como observado, não houve deslocamentos nos picos típicos dos espectros Raman na faixa de 1800 a 400 cm^{-1} , e as bandas estavam coerentes com os resultados de estudos anteriores (Batesttin *et al.*, 2022; Li *et al.*, 2023; Yaman, 2020; Yazgan *et al.*, 2020).

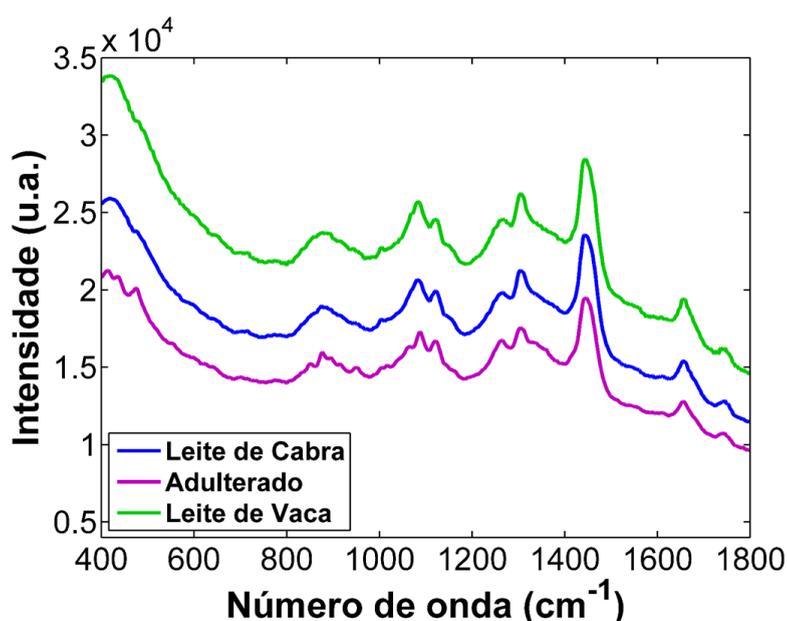


Figura 32. Espectros RAMAN médios das classes: Leite caprino puro (Azul), Leite caprino adulterado com leite bovino (Rosa) e Leite bovino puro (Verde).

Para facilitar a visualização das principais diferenças entre o leite de cabra puro (em azul) e o leite de cabra em pó adulterado com leite de vaca (em rosa), os espectros médios derivados foram mostrados na **Figura 33**. De fato, a espectroscopia Raman revelou características espectrais distintas indicativas de vários constituintes moleculares dentro das amostras analisadas. Notavelmente, bandas características foram observadas na faixa de $445\text{-}455\text{ cm}^{-1}$, atribuídas ao ácido láctico (Yaman, 2020). Além disso, picos proeminentes apareceram aproximadamente a 871 cm^{-1} , indicando a presença de aminoácidos. A análise adicional revelou um pico distinto a 1154 cm^{-1} , característico dos carotenoides, uma característica ausente nas amostras de leite de cabra puro. Outros picos de carotenoides também são visualizados ao redor de 1005 e 1551 cm^{-1} (Batestin *et al.*, 2022; Yaman, 2020). Uma banda a 1084 cm^{-1} foi atribuída à lactose (Batestin *et al.*, 2022). Outros deslocamentos Raman distintos na região espectral que se estende de 1000 a 1120 cm^{-1} , juntamente com a presença de outras bandas centradas ao redor de 1300 cm^{-1} e 1748 cm^{-1} , forneceram evidências de constituintes de ácidos graxos (Batestin *et al.*, 2022). A característica espectral centrada em $1655\text{-}1665\text{ cm}^{-1}$ foi associada a Amida I, provavelmente originada de proteínas (Batestin *et al.*, 2022; Li *et al.*, 2023). Além disso, bandas em $1440\text{-}1450\text{ cm}^{-1}$ foram indicativas da presença de carboidratos e ácidos graxos (especificamente colesterol a 1442 cm^{-1}) nas amostras analisadas (Batestin *et al.*, 2022; Li *et al.*, 2023; Yaman, 2020).

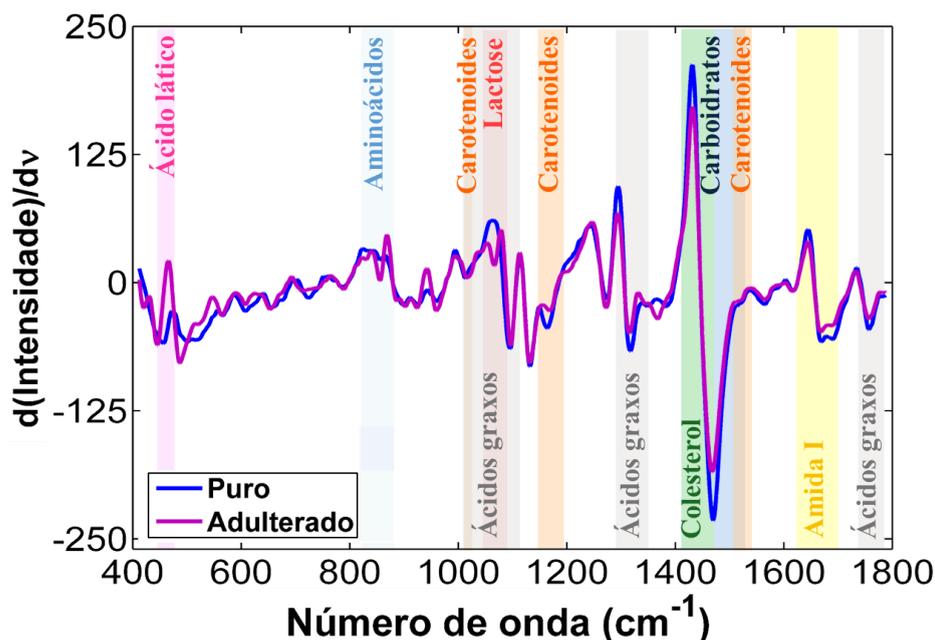


Figura 33: Espectros RAMAN pré processados com SGD 25p: Leite caprino puro (Azul) e Leite caprino adulterado com leite bovino (Rosa).

Os espectros foram corrigidos por meio dos seguintes pré-processamentos: correção de linha de base por Offset MSC, SNV e aplicação de filtro derivativo de 1º ordem Savitzky–Golay com janelas de 25 pontos. Após o pré-processamento dos espectros, uma análise exploratória dos dados foi realizada utilizando PCA para ilustrar as tendências de agrupamento e separação entre as amostras analisadas, a fim de verificar a presença ou ausência de leite de vaca no leite de cabra em pó. Os resultados gráficos bidimensionais apresentados na **Figura 34**, referentes aos escores de PC1 e PC2, ilustram se há tendência de separação entre amostras de leite de cabra puro (círculos azuis) e adulterado (círculos roxos). Analisando os gráficos nota-se sobreposição entre as amostras de leite de cabra puro e adulteradas. Ainda é observado que o tipo de pré-processamento aplicado nos dados influenciou na dispersão das amostras em torno das duas primeiras PCs.

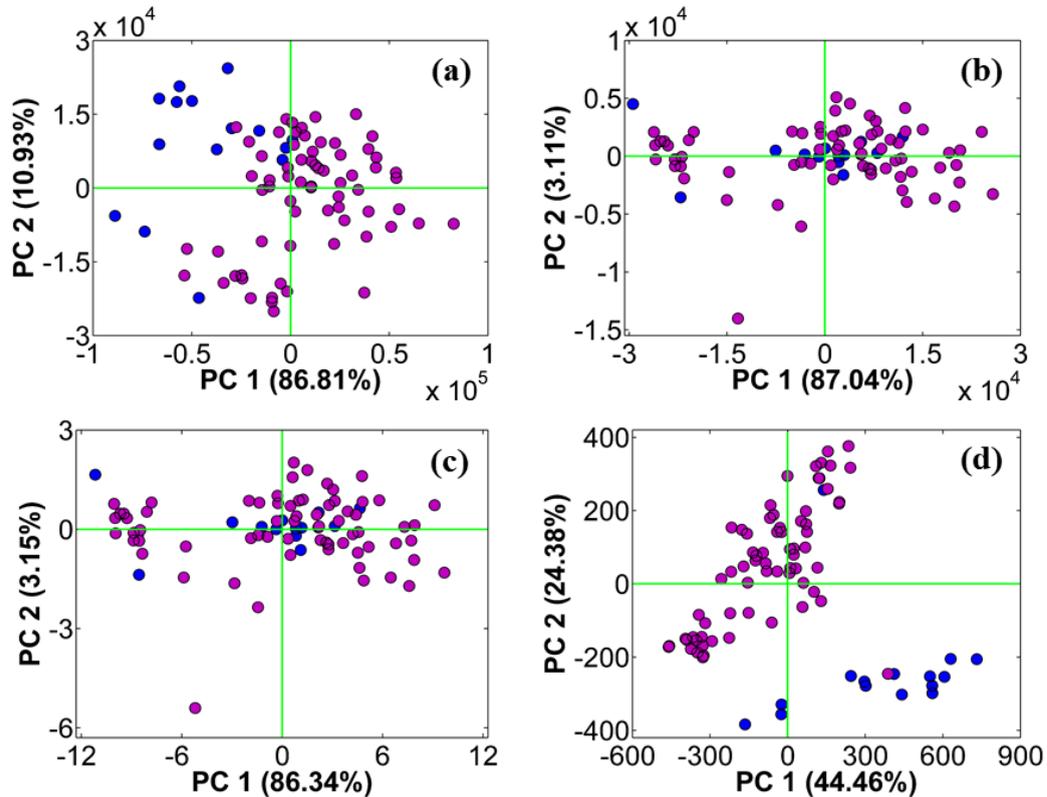


Figura 34: Gráficos de escores para os dados corrigidos com OFF (a), MSC (b), SNV (c) e aplicando filtro derivativo de 1° ordem Savitzky – Golay com janelas de 25 pontos (d), respectivamente.

Os escores dos espectros Raman brutos (**Figura 35a**) mostram uma sutil tendência de separação entre amostras puras e adulteradas, com agrupamentos predominantemente em valores negativos de PC1 e positivos de PC2, explicando variâncias de 97,97% e 1,44%, respectivamente. Por outro lado, uma clara tendência de separação é observada ao aplicar SGD como técnica de pré-processamento.

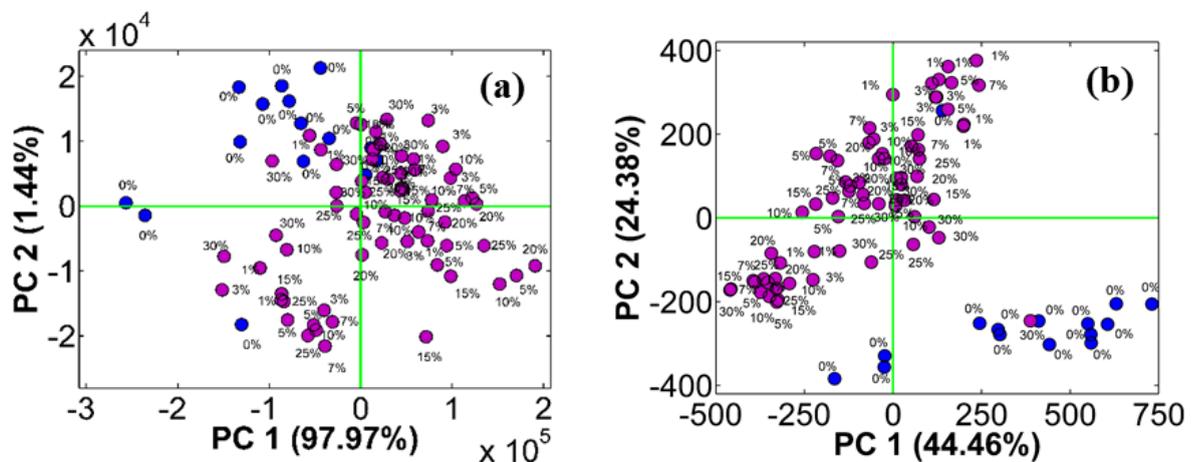


Figura 35: Gráficos de escores para os dados Brutos (a) e corrigidos aplicando o filtro derivativo de 1° ordem Savitzky – Golay com janelas de 25 pontos (b), respectivamente, rotulados com teores.

Observando as **Figuras 35a e 35b**, seria esperado que amostras contendo baixas concentrações de adulterante se comportassem de maneira mais semelhante às amostras não adulteradas. No entanto, o leite de vaca apresenta uma composição química complexa que, apesar de possuir relativa semelhança com o leite de cabra, introduz na matriz uma maior variabilidade. Essa complexidade e semelhança impedem uma relação linear direta em termos de nível de adulteração. Ao contrário de substâncias puras (como glicose ou melamina, por exemplo), nas quais a correlação direta entre o nível de adulteração e seu impacto é mais previsível.

Ainda sobre a discussão da **Figura 35b**, é possível identificar uma amostra adulterada agrupada na classe das puras, enquanto uma amostra pura está agrupada na classe das adulteradas. Ao analisar este cenário, percebeu-se que houve um possível erro de rotulagem uma vez que, dificilmente uma amostra com 30% de adulteração seria confundida com uma amostra pura. Outro fato que reforça esse argumento seria a alternância de apenas uma única amostra agrupada erroneamente em cada classe.

Não obstante, mesmo o filtro derivativo com 25 pontos demonstrando-se promissor, este apresentou apenas 68,84% da variância explicada dos dados, logo analisou-se o potencial da PC3 (**Figura 36**), que contém cerca de 8,83%, na visualização de agrupamentos naturais entre as classes. O gráfico de escores apresentado abaixo elucidou uma sutil tendência de separação, com as amostras da classe-alvo mais uma vez deslocadas para valores mais positivos em PC1, entretanto a adição da PC3 não contribuiu significativamente para além das informações já obtidas.

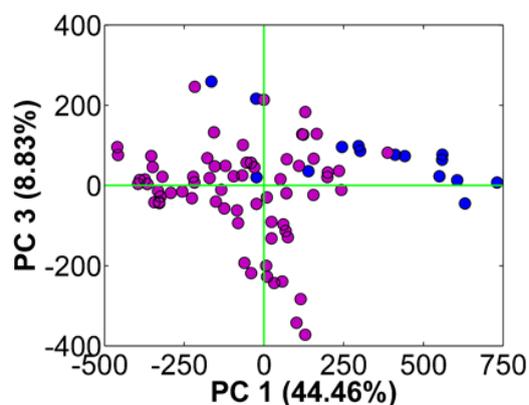


Figura 36: Gráficos de escores (PC1 vs. PC3) para os dados corrigidos com SGD para janelas de 25 pontos.

Uma outra ferramenta analisada foram os gráficos dos pesos. Estes, por sua vez, tem como objetivo mostrar como as diferentes bandas espectrais (ou regiões) contribuem para

cada componente principal, ou seja, qual ou quais regiões do espectro carregam maior informação sobre a classe ou substância estudada. A **Figura 37** ilustra o comportamento dos pesos referentes às três primeiras PCs que carregam a maior parte da variância ou informação contida nos dados. Cada gráfico é referente a um pré tratamento utilizado. Com base nas regiões de alta intensidade, sejam positivas ou negativas, é possível confirmar a importância das bandas associadas aos compostos macromoleculares (especialmente gorduras, proteínas, carboidratos e carotenoides) na contribuição para essa separação.

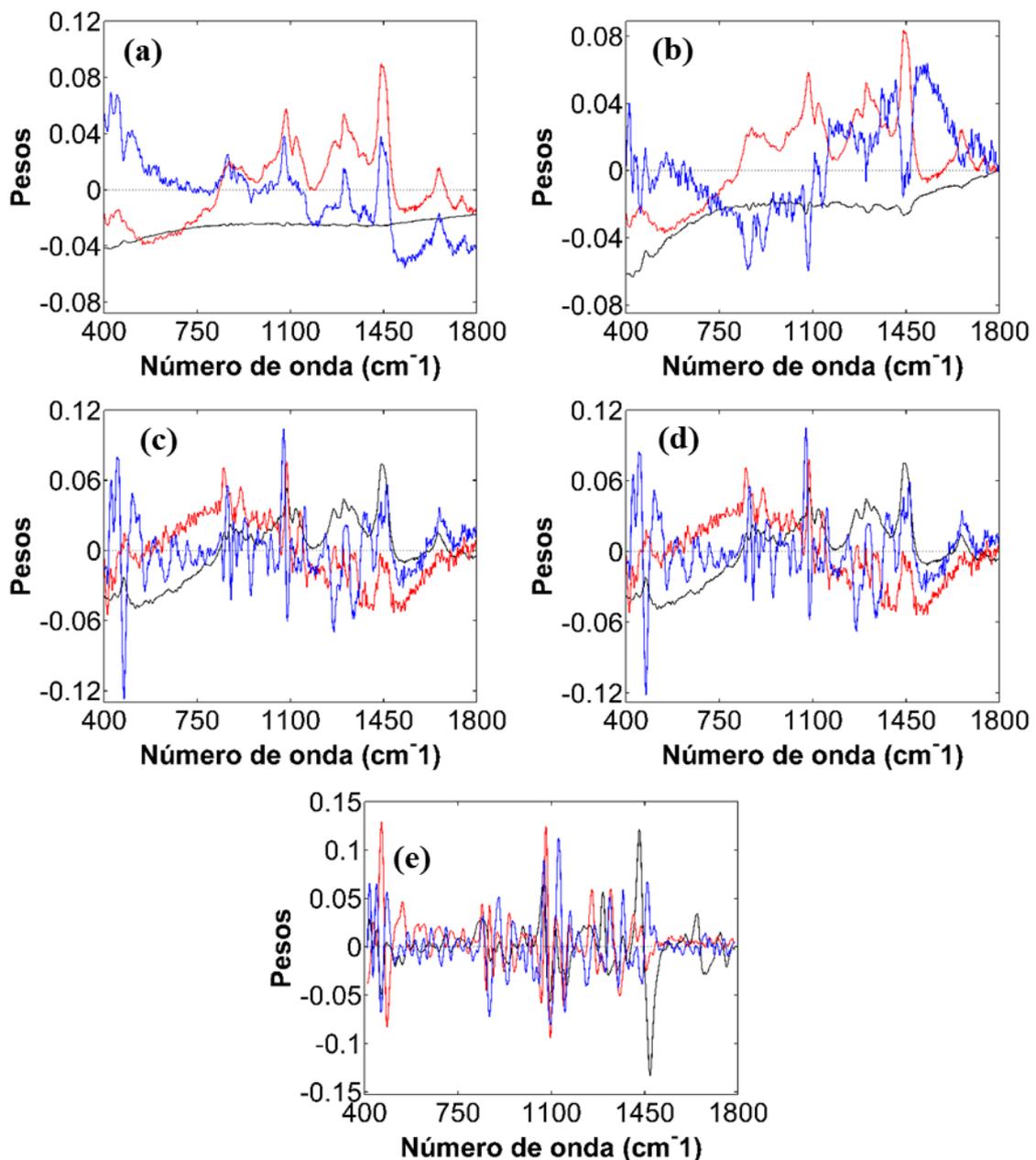


Figura 37: Gráfico dos pesos da PCA obtidos para PC1 (linha preta), PC2 (linha vermelha) e PC3 (linha azul), a partir dos espectros Raman de amostras puras de leite de cabra em pó Brutos (a) e pré-processadas com OFF (b), MSC (c), SNV (d) e a SGD com um polinômio de segunda ordem e uma janela de 25 pontos (e).

As linhas preta, vermelha e azul correspondem aos pesos de PC1, PC2 e PC3, respectivamente, obtidos para os dados brutos (**Figura 37a**), OFF (**Figura 37b**), MSC (**Figura 37c**), SNV (**Figura 37d**) e 1ª SGD (**Figura 37e**). Essas três primeiras PCs respondem por 97,7%; 98,3%; 92,7%; 92,1% e 77,7% da variância total explicada dos dados, respectivamente. É perceptível em todos os gráficos picos muito intensos em torno de 1450, portanto independente de qual seja o pré processamento, essa região referente a presença de carboidratos e ácidos graxos, será crucial para a diferenciação das classes. Cabe salientar a importância de outras regiões presentes em todos os gráficos, como a banda em torno de 1100, que envolvem a presença da lactose, gorduras e carotenoides.

Em resumo, os gráficos de escores PCA para as técnicas de pré-processamento OFF e SGD revelam uma distinção quase completa entre amostras de leite de cabra puro e aquelas adulteradas com leite de vaca, já os pré-processamentos MSC e SNV exibiram sobreposições significativas entre as classes estudadas. No entanto, a PCA atua apenas como uma ferramenta não supervisionada para análise de triagem. Por isso, a DD-SIMCA foi empregada para fins de autenticação, conforme discutido na seção seguinte.

5.2.2 Autenticação do Leite em pó caprino

De maneira análoga ao experimento 1, para fins de autenticação na análise de alimentos, também optou-se por utilizar classificadores de uma única classe (do inglês, one-class classifiers - OCC) (Araújo, *et al.*, 2023; Rodionova; Titova; Pomerantsev, 2016). Os procedimentos para delimitação do *threshold* dos modelos de classificação e avaliação de amostras desconhecidas seguiram em concordância com o experimento 1.

Todavia, nesse cenário, a representatividade da classe alvo é o primeiro passo crítico para construir um modelo sem sobreajuste ou subajuste (overfitting e underfitting, respectivamente), especialmente ao lidar com um número reduzido de amostras alvo. Portanto, o desafio é definir uma estratégia de amostragem adequada que garanta a representatividade estatística, considerando a natureza complexa das amostras de leite comercial (seja leite de cabra ou de vaca), uma vez que as operações unitárias de tratamentos térmicos e separação e padronização de gordura na indústria de laticínios tornam mais difícil avaliar as diferenças entre amostras puras e adulteradas. Como este trabalho compreende apenas 15 amostras alvo (ou seja, leite de cabra puro), testou-se três diferentes estratégias de validação para construir e avaliar a capacidade preditiva dos modelos DD-SIMCA, conforme detalhado na **Tabela 5**.

Na Abordagem 1, foi avaliada apenas a sensibilidade do conjunto de treinamento (composto por 15 amostras de leite de cabra puro) em reconhecer amostras alvo e a especificidade do conjunto de teste (composto por 68 amostras de leite de cabra adulterado) em projetar amostras não-alvo para fora da área de aceitação calculada. A Abordagem 2 utilizou, além das amostras da Abordagem 1, um novo conjunto de pseudo-validação, simulando um novo conjunto de objetos retirados das amostras do conjunto de treinamento, usando o algoritmo PCV (Pomerantsev & Rodionova, 2021). Isso permite avaliar a sensibilidade na etapa de previsão (ou seja, incluindo as novas amostras PCV no conjunto de teste), uma vez que nenhuma amostra é removida do conjunto de treinamento original e, conseqüentemente, a robustez do modelo desenvolvido.

Finalmente, na Abordagem 3, o algoritmo KS (Kennard & Stone, 1969) foi usado para dividir as 15 amostras alvo em conjuntos de treinamento e teste, sendo 10 e 5, respectivamente. As amostras alvo usadas no conjunto de teste permitem avaliar a sensibilidade na etapa de previsão, juntamente com a representatividade estatística e a qualidade dos espectros Raman obtidos. Por último, todas as 68 amostras de leite de cabra adulteradas foram usadas nas três abordagens para avaliar a especificidade dos modelos desenvolvidos. Nas **Tabelas 7 e 8** são apresentados os resultados obtidos pelo DD-SIMCA nos níveis de significância de 0,01 e 0,05 para as três estratégias de validação na autenticação do leite de cabra usando espectroscopia Raman.

Tabela 7: Resultados obtidos por DD-SIMCA ($\alpha = 0,01$) para a autenticação do leite de cabra em pó usando espectroscopia RAMAN.

		Abordagem 1		Abordagem 2			Abordagem 3		
		Treino	Teste	Treino	Teste		Treino	Teste	
		Alvo	Não-alvo	Alvo	Alvo	Não-alvo	Alvo	Alvo	Não-alvo
(OFF)									
PCs		6		6			4		
Matriz de confusão	Alvo	15	2	15	15	2	10	3	2
	Não alvo	-	66	-	-	66	-	2	66
	Sensibilidade	100%	-	100%	100%		100%	60%	
	Especificidade	-	97%	-	97%		-	97%	
	Eficiência	98%		99%			83%		
(MSC)									
PCs		6		6			2		
Matriz de confusão	Alvo	15	-	15	15	-	10	5	19
	Não alvo	-	68	-	-	68	-	-	49
	Sensibilidade	100%	-	100%	100%		100%	100%	
	Especificidade	-	100%	-	100%		-	72%	
	Eficiência	100%		100%			90%		
(SNV)									
PCs		6		6			2		
Matriz de confusão	Alvo	15	-	15	15	-	10	5	19
	Não alvo	-	68	-	-	68	-	-	49
	Sensibilidade	100%	-	100%	100%		100%	100%	
	Especificidade	-	100%	-	100%		-	72%	
	Eficiência	100%		100%			90%		
(SGD)									
PCs		3		3			3		
Matriz de confusão	Alvo	15	-	15	13	-	10	5	-
	Não alvo	-	68	-	2	68	-	-	68
	Sensibilidade	100%	-	100%	87%		100%	100%	
	Especificidade	-	100%	-	100%		-	100%	
	Eficiência	100%		95%			100%		

Tabela 8: Resultados obtidos por DD-SIMCA ($\alpha = 0,05$) para a autenticação do leite de cabra usando espectroscopia RAMAN.

		Abordagem 1		Abordagem 2			Abordagem 3		
		Treino	Teste	Treino	Teste		Treino	Teste	
		Alvo	Não-alvo	Alvo	Alvo	Não-alvo	Alvo	Alvo	Não-alvo
(OFF)									
PCs		6		6			4		
Matriz de confusão	Alvo	15	-	15	15	-	10	2	-
	Não-alvo	-	68	-	-	68	-	3	68
Sensibilidade		100%	-	100%	100%		100%	40%	
Especificidade		-	100%	-	100%		-	100%	
Eficiência		100%		100%			74%		
(MSC)									
PCs		5		5			6		
Matriz de confusão	Alvo	15	-	15	15	-	10	5	5
	Não-alvo	-	68	-	-	68	-	-	63
Sensitivity		100%	-	100%	100%		100%	100%	
Especificidade		-	100%	-	100%		-	93%	
Eficiência		100%		100%			97%		
(SNV)									
PCs		7		7			7		
Matriz de confusão	Alvo	15	-	15	15	-	10	5	8
	Não-alvo	-	68	-	-	68	-	-	60
Sensitivity		100%	-	100%	100%		100%	100%	
Especificidade		-	100%	-	100%		-	88%	
Eficiência		100%		100%			96%		
(SGD)									
PCs		7		7			5		
Matriz de confusão	Alvo	15	-	15	14	-	10	5	1
	Não-alvo	-	68	-	1	68	-	-	67
Sensitivity		100%	-	100%	93%		100%	100%	
Especificidade		-	100%	-	100%		-	98%	
Eficiência		100%		98%			99%		

Para a Abordagem 1, todas as amostras alvo e não-alvo foram corretamente projetadas dentro e fora das áreas de aceitação, respectivamente, independentemente da técnica de pré-processamento e do nível de significância. Isso implica sensibilidades, especificidades e eficiências de 100% para todos os modelos desenvolvidos, utilizando entre 3 e 7 PCs para construir os modelos DD-SIMCA, exceto para o OFF com $\alpha = 0,01$ (cuja discussão encontra-se no apêndice). No entanto, utilizando o princípio da parcimônia, o modelo DD-SIMCA que usou espectros Raman pré-processados com derivada Savitzky-Golay e MSC alcançou os melhores resultados com eficiências de 100%, empregando apenas 3 e 5 PCs para $\alpha = 0,01$ e $\alpha = 0,05$, respectivamente, como ilustrado na **Figura 38**.

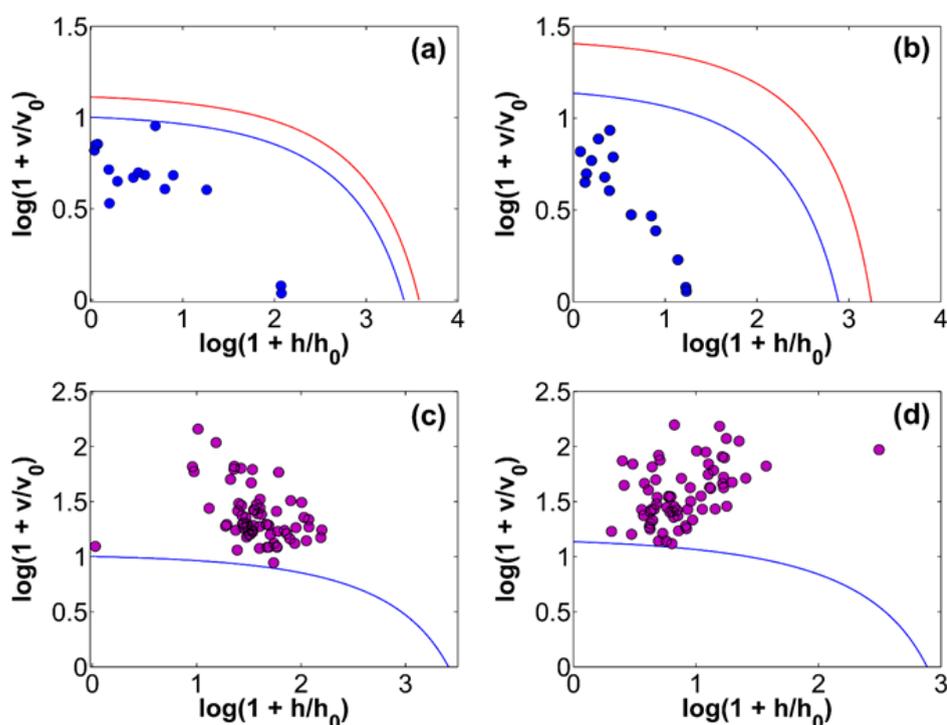


Figura 38: Gráficos de área de aceitação para o conjunto de treinamento (a, b) e teste (c, d) empregando DD-SIMCA, abordagem 1.

Ao usar novas amostras de pseudo-validação PCV na Abordagem 2, os modelos MSC e SNV alcançaram 100% de sensibilidade tanto nos conjuntos de treinamento quanto de teste para os dois níveis de significância (0,01 e 0,05). Para o modelo OFF com $\alpha = 0,01$, duas amostras não-alvo foram incorretamente projetadas como amostras alvo (cuja discussão encontra-se no apêndice). Além disso, os modelos SGD projetaram fora da área de aceitação no conjunto de teste duas amostras alvo para $\alpha = 0,01$ e uma amostra alvo para $\alpha = 0,05$ (cuja discussão encontra-se no apêndice). Apesar disso, todos os modelos desenvolvidos

alcançaram eficiências superiores a 95% para ambos os níveis de significância. Em outras palavras, a sensibilidade da abordagem PCV parece muito razoável, com um ótimo número de PCs, em torno de 3 e 7, o que está de acordo com o princípio dos Parentes Mais Próximos (Rodionova & Pomerantsev, 2020). Isso ocorre porque o conjunto PCV alternativo pode ser usado de forma independente como um verdadeiro conjunto de teste, uma vez que exibe uma variabilidade diferente em comparação com os dados de treinamento originais. A **Figura 39** ilustra o desempenho ótimo dos modelos DD-SIMCA usando espectros Raman pré-processados com MSC e empregando 6 e 5 PCs para $\alpha = 0,01$ e $\alpha = 0,05$, respectivamente, com sensibilidades, especificidades e eficiências de 100%.

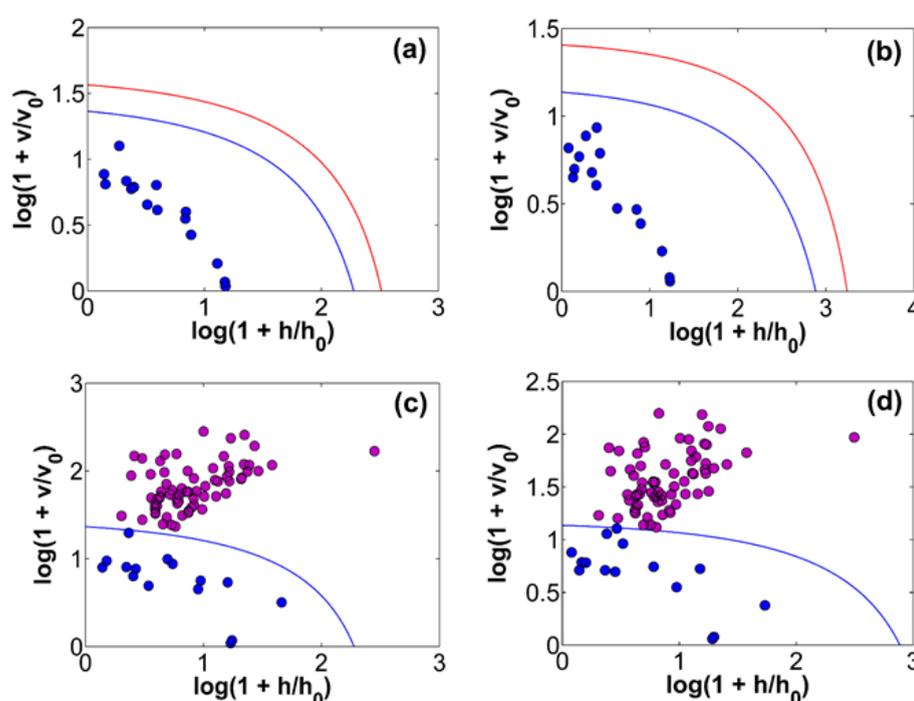


Figura 39: Gráficos de área de aceitação para o conjunto de treinamento (a, b) e teste (c, d) empregando DD-SIMCA, abordagem 2.

Finalmente, ao remover até cinco amostras da classe alvo para compor o conjunto de teste na Abordagem 3, as 10 amostras alvo restantes não resultaram em mudanças dramáticas nos modelos desenvolvidos. De fato, todos os modelos reconheceram todas as amostras alvo tanto nos conjuntos de treinamento quanto de teste, exceto para o OFF, que projetou fora da área de aceitação no conjunto de teste duas amostras alvo para $\alpha = 0,01$ e três amostras alvo para $\alpha = 0,05$ (cuja discussão encontra-se no apêndice). Especificamente, os modelos DD-SIMCA que utilizaram espectros Raman pré-processados com SGD alcançaram os melhores resultados com eficiências de 100% e 99%, empregando 3 e 5 PCs para $\alpha = 0,01$ e α

= 0,05, respectivamente, conforme ilustrado na **Figura 40**, com apenas uma amostra não-alvo projetada incorretamente dentro da área de aceitação para o modelo utilizando 5 PCs e $\alpha = 0,05$. A presença desse erro pode ser atribuída a um possível problema de rotulagem discutido na **Seção 5.2.1**.

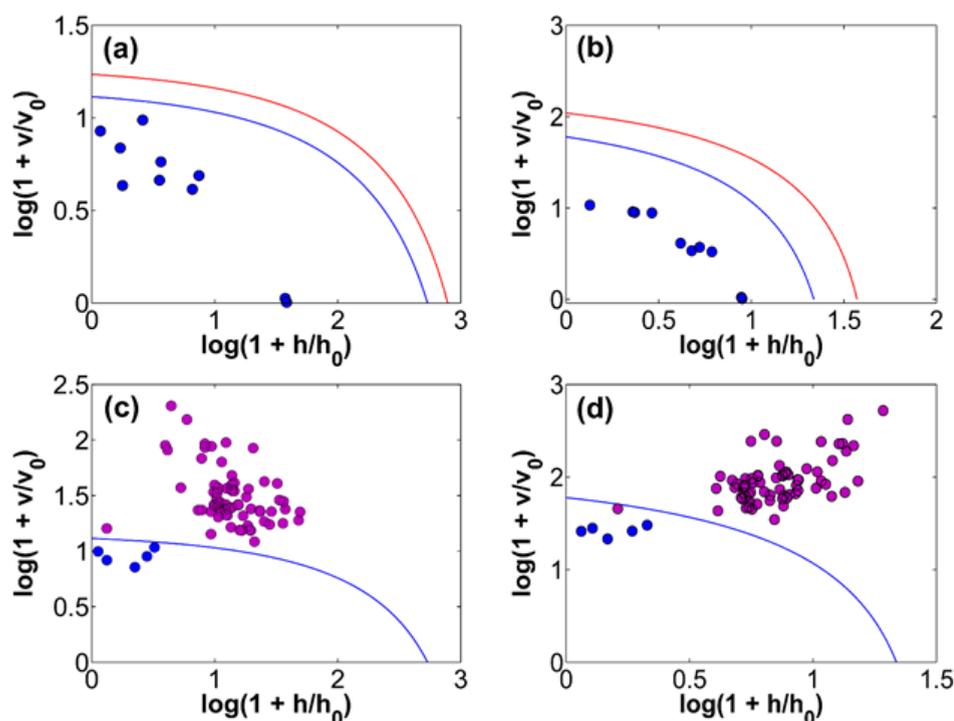


Figura 40: Gráficos de área de aceitação para o conjunto de treinamento (a, b) e teste (c, d) empregando DD-SIMCA, abordagem 3.

Apesar do pequeno tamanho do conjunto de treinamento, esses resultados foram compatíveis com as Abordagens 1 e 2, particularmente no caso de $\alpha = 0,01$, que utilizou o mesmo número de PCs da Abordagem 1, demonstrando a representatividade estatística das informações analíticas contidas nos espectros Raman registrados.

Dado o exposto, o método proposto se destaca como uma alternativa promissora para a autenticação de leite em pó caprino, seguindo os Princípios da Química Verde (Ballesteros-Vivas *et al.*, 2021), ao utilizar um instrumento portátil com preço razoável em conjunto com uma análise rápida, não invasiva e não destrutiva. Ademais, o método com resultados validados incentiva as agências reguladoras a adotá-lo oficialmente no futuro como um método de rotina usado na autenticação de leite de cabra em pó para resolução da problemática levantada, esta que não só leva em consideração fraudes ilegais motivadas economicamente, como também levanta preocupações de saúde pública.

Capítulo 6

Conclusões e perspectivas

6 CONCLUSÕES E PERSPECTIVAS

6.1 CONCLUSÕES

Este trabalho demonstrou a viabilidade da utilização da espectroscopia NIR empregando instrumentos miniaturizado para a análise *in situ* como ferramenta na identificação de adulteração de leite caprino por adição de adulterantes como: leite bovino em pó, extrato de soja, amido de milho e composto lácteo. Para isso, diferentes técnicas de pré-processamento para dados NIR, juntamente com os modelos OC-PLS e DD-SIMCA, isoladamente, foram utilizados. Os modelos obtiveram taxas de sensibilidades e especificidades relativamente altas, destacando o resultado alcançado empregando Xs-21p-OFF. Dessa forma, a espectroscopia NIR se mostrou uma alternativa excelente para identificar leite de cabra em pó, mesmo na presença de possíveis interferências.

Para os dados RAMAN, o estudo investigou a eficácia dos modelos DD-SIMCA na verificação da autenticidade do leite de cabra em pó puro alcançando um desempenho ótimo nos níveis de significância testados ($\alpha = 0,01$ e $\alpha = 0,05$). A seleção das técnicas de pré-processamento mostrou-se crucial, com MSC e a primeira derivada de Savitzky-Golay produzindo os resultados mais favoráveis. Apesar de uma classe-alvo limitada, composta por apenas 15 amostras, a avaliação de três estratégias de validação destacou a robustez dos espectros Raman no fornecimento de informações analíticas relevantes para a autenticação do leite de cabra em pó puro. Os modelos construídos com $\alpha = 0,01$ demonstraram consistentemente alta sensibilidade (100% tanto nos conjuntos de treinamento quanto de teste) e especificidade (100% no conjunto de teste), superando aqueles construídos com $\alpha = 0,05$ e maiores números de componentes principais. Esses resultados ressaltam a eficácia da espectroscopia Raman portátil combinada com DD-SIMCA na verificação precisa da autenticidade do leite de cabra em pó, alinhando-se aos princípios da análise de alimentos sustentáveis.

Por fim, o método proposto pode ser utilizado no futuro, por pequenos produtores e cooperativas, como um método de rotina eficaz, rápido, barato e não destrutivo para análises *in situ* de amostras de leite de cabra em pó quanto a possíveis adulterações, particularmente valiosa em cenários onde caracterizar o tipo e/ou natureza dos adulterantes é impraticável. Outrossim, pode ser utilizado para nortear as decisões dos órgãos reguladores na prevenção direta desse tipo de problemática, que não só leva em consideração fraudes ilegais motivadas economicamente, como também levanta preocupações de saúde pública.

6.2 PERSPECTIVAS

Embora o método proposto tenha proporcionado resultados satisfatórios, sugerimos as seguintes propostas para dar continuidade no desenvolvimento deste trabalho:

- Propor modelos de calibração com os resultados obtidos pelas espectroscopias NIR e RAMAN.
- Avaliar o desempenho dos resultados utilizando a fusão de dados.

Capítulo 7

Referências bibliográficas

REFERÊNCIAS

- Achuate, A. J. et al. **Aplicação de espectrometria de massas com ionização por eletrospray e métodos quimiométricos para classificar espécies de Miconia spp.** 2023. Dissertação (Mestrado em Química) - Instituto de Química, Universidade Federal de Uberlândia, Uberlândia.
- Afseth, N. K.; Wold, J. P.; Segtnan, V. H. The potencial spectroscopy for characterisation of the acid unsaturation of Salmon. **Analytica Chimica Acta**, v. 572, p. 85 – 92, 2006.
- Akulich, F. et al. Explainable predictive modeling for limited spectral data. **Chemometrics and Intelligent Laboratory Systems**, v. 225, p. 104572, 2022.
- Alcantara Junior, P. **Espectroscopia molecular**. 2002. Disponível em: <https://www.researchgate.net/profile/Petrus-Alcantara-Jr/publication/267222966_ESPECTROSCOPIA_MOLECULAR/links/558c133908ae591c19d9e888/ESPECTROSCOPIA-MOLECULAR.pdf>. Acesso em: ago, 2024.
- Allegretta, I. et al. Macro-classification of meteorites by portable energy dispersive x-ray fluorescence spectroscopy (ped-xrf), principal component analysis (PCA) and machine learning algorithms. **Talanta**, v.212, p. 120785, 2020.
- Amigo, L.; Fontecha, J. **Goat Milk**. 2. Ed. Elsevier, 2011.
- Anandharamakrishnan, C. et al. **Spray Drying Techniques For Food Ingredient Encapsulation**. John Wiley & Sons, 2015.
- Angel, S. M.; Carrabba M.; Cooney T. F.; The Utilization Of Diode Lasers For Raman Spectroscopy. **Spectrochim Acta A Mol Spectrosc.** v. 51, p. 1779-1799. 1995.
- Aragão, S. R. M. et al. Aspectos Nutricionais E Sensoriais Do Leite De Cabra E Seus Derivados. **Ciência e Tecnologia de Alimentos: O Avanço Da Ciência No Brasil**. v. 2, p. 47-62, 2022.
- Arlot, S.; Celisse, A. A Survey of cross-validation procedures for model selection. **Statistics Survey**, v.4, p. 40-79, 2010.
- Araújo, T.K.L. et al. Authentication of the shelf-life and decaffeination process of instant coffee samples using UV–vis and NIR spectral fingerprints. **Food Control**, v. 155, p. 110098, 2024.
- Arruda, N. F. S. et al. Diagnóstico da alergia à proteína do leite de vaca (aplv) e os seus desafios: uma revisão integrativa da literatura. **Revista Da Faculdade De Ciências Médicas Da Paraíba**, v. 1, n. 2, 2023.
- Ballesteros-Vivas, D. et al. Green food analysis: current trends and perspectives. **Current Opinion In Green And Sustainable Chemistry**, v. 31, p. 100522, 2021.

Barbosa, A. L. P. B. **Desenvolvimento de pão de queijo sem lactose e análise da influência deste carboidrato em sua composição**. 2023. Trabalho de conclusão de curso (Bacharelado em Engenharia de alimentos) - Centro de Ciências da Natureza, Universidade Federal de São Carlos, Buri-SP.

Barros Neto, B.; Scarminio, I. S.; Bruns, R. E. 25 anos de quimiometria no Brasil. **Química Nova**, v. 29, p. 1401-1406, 2006.

Batestin, C. et al. High resolution raman spectroscopy of raw and UHT bovine and goat milk. **Measurement: Food**, v. 6, p. 100029, 2022.

Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. Chemometrics: a practical guide. **(No Title)**, 1998.

Berrueta, L. A.; Alonso-Salces, R. M.; Héberger, K. Supervised pattern recognition in food analysis. **Journal of Chromatography A**, v. 1158, p. 196–214, 2007.

Borra, S.; Di Ciaccio, A. Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods. **Computational statistics & data analysis**, v. 54, n. 12, p. 2976-2989, 2010.

Brandão, J. de S. Mitologia Grega (Vols. I, II e III). **Petrópolis-RJ: Vozes**, 1987.

Brandão, M. P.; Dos Anjos, V. C.; Bell, M. J. V. Time resolved fluorescence of cow and goat milk powder. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 171, p. 193-199, 2017.

Brasca, R.; Goicoechea, H. C.; Culzoni, M. J. Multiway Calibration Approaches for Quality Control of Food Samples. In: **Food Safety and Preservation**. Academic Press, 2018. p. 143-165.

Brasil. Ministério da Agricultura, Pecuária e Abastecimento. **Instrução Normativa n. 53 de 1º de outubro de 2018**. In: Diário Oficial da União, Brasília, 02 de outubro de 2018. Disponível em: <https://www.in.gov.br/materia/-/asset_publisher/Kujrw0TZC2Mb/content/id/45374372/do1-2018-10-16-instrucao-normativan-53-de-1-de-outubro-de-2018-45374042>. Acesso em: 07 ago. 2022.

Brasil. **Instrução Normativa nº 37, de 31 de outubro de 2000**. Diário Oficial da União, Brasília, DF, 31 out. 2000.

Brasil. Instrução Normativa Nº 76, de 26 de Novembro de 2018. **Diário Oficial da União**, p. 10, 2018.

Brereton, R. G. et al. Chemometrics in analytical chemistry—part I: history, experimental design and data analysis tools. **Analytical and Bioanalytical Chemistry**, v. 409, p. 5891-5899, 2017.

Brereton, R. G. **Chemometrics: data driven extraction for science**. John Wiley & Sons, 2018.

Brereton, R. G.; Lloyd, G. R. Partial least squares discriminant analysis: taking the magic away. **Journal of Chemometrics**, v. 28, n. 4, p. 213-225, 2014.

Britannica, The Editors of Encyclopaedia. "Hooke's law". **Encyclopedia Britannica**, 13 Aug. 2024, Disponível: <<https://www.britannica.com/science/Hookes-law>>. Acesso em: 15 Ago. 2024.

Burns, D. A.; Ciurczak, E. W. (Ed.). **Handbook of near-infrared analysis**. CRC press, 2007.

Cakir, B.; Tunali-Akbay, T. Potential anticarcinogenic effect of goat milk-derived bioactive peptides on HCT-116 human colorectal carcinoma cell line. **Analytical biochemistry**, v. 622, p. 114166, 2021.

Carić, M. et al. Technology of evaporators, membrane processing and dryers. **Dairy Powders and Concentrated Products**, p. 99-148, 2009.

CARTONI, G. et al. Determination of cows' milk in goats' milk and cheese by capillary electrophoresis of the whey protein fractions. **Journal of Chromatography A**, v. 846, n. 1-2, p. 135-141, 1999.

Casale, M. et al. Characterisation of PDO olive oil Chianti Classico by non-selective (UV-visible, NIR and MIR spectroscopy) and selective (fatty acid composition) analytical techniques. **Analytica chimica acta**, v. 712, p. 56-63, 2012.

Chen, H. et al. Detection of melamine adulteration in milk by near-infrared spectroscopy and one-class partial least squares. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**. v. 173, p. 832-836, 2017.

Chen, X. et al. Lipase-catalyzed modification of milk fat: A promising way to alter flavor notes of goat milk products. **Lwt**, v. 145, p. 111286, 2021.

Clarke, H. J. et al. Development of a headspace solid-phase microextraction gas chromatography mass spectrometry method for the quantification of volatiles associated with lipid oxidation in whole milk powder using response surface methodology. **Food chemistry**. v. 292, p. 75-80, 2019.

Coombs, C. E. O.; Liddle, R. R.; González, L. A. Portable vibrational spectroscopic methods can discriminate between grass-fed and grain-fed beef. **Journal of Near Infrared Spectroscopy**, v. 29, n. 6, p. 321-329, 2021.

Cruz, A. G. da et al. **Microbiologia, higiene e controle de qualidade no processamento de leites e derivados**. Elsevier, 2019.

Cruz, A. G. et al. **Química, bioquímica, análise sensorial e nutrição no processamento de leite e derivados**. Elsevier, 2016.

Czaja, T. et al. Determining moisture content in pasta by vibrational spectroscopy. **Talanta**, v. 178, p. 294-298, 2018.

Dąbrowska, A. et al. Quality of UHT goat's milk in Poland evaluated by real-time PCR. **Small Ruminant Research**, v. 94, n. 1-3, p. 32-37, 2010.

Dantas, W. C.; Alves, J. C. L.; Poppi, R. J. MCR-ALS with correlation constraint and Raman spectroscopy for identification and quantification of biofuels and adulterants in petroleum diesel. **Chemometrics and Intelligent Laboratory Systems**, v. 169, p. 116-121, 2017.

Da Silva, W. B.; Da Silva Andrade, H. M. L.; De Andrade, L. P. Técnicas inovadoras aplicadas ao controle de qualidade de alimentos: Uma revisão sistemática. **Diversitas Journal**, v. 9, n. 1, p. 426-438, 2024.

Dec, A T. S et al. Validação cruzada na aprendizagem de redes neurais convolucionais para triagem de câncer cervical. **Open science research XV**, p. 546-561, 2024.

Deconinck, E. et al. Detection and identification of multiple adulterants in plant food supplements using attenuated total reflectance—Infrared spectroscopy. **Journal of pharmaceutical and biomedical analysis**, v. 152, p. 111-119, 2018.

Deidda, R. et al. Vibrational spectroscopy in analysis of pharmaceuticals: Critical review of innovative portable and handheld NIR and Raman spectrophotometers. **TrAC Trends in Analytical Chemistry**, v. 114, p. 251-259, 2019.

Delgado Júnior, I. J.; Siqueira, K. B.; Stock, L. A. **Produção, composição e processamento de leite de cabra no Brasil**. Embrapa, 2020.

Di Costanzo, M. et al. Lactose Intolerance in Pediatric Patients and Common Misunderstandings About Cow's Milk Allergy. **Pediatric Annals**, v. 50, n. 4, p. e178-e185, 2021.

Di Pinto, A. et al. DNA-based approach for species identification of goat-milk products. **Food chemistry**, v. 229, p. 93-97, 2017.

Dias, L. A. et al. An electronic tongue taste evaluation: Identification of goat milk adulteration with bovine milk. **Sensors and Actuators B: Chemical**, v. 136, n. 1, p. 209-217, 2009.

Dias, V. H. C. et al. Métodos oficiais empregados para autenticidade, controle da qualidade e detecção de fraudes em leite cru refrigerado no Brasil. **Brazilian Journal of Development**, v. 6, n. 2, p. 7129-7137, 2020.

Do Egito, A. S. et al. **Variação anual na composição do leite de cabra produzido nos estados da Paraíba e Pernambuco**. Embrapa, 2023.

Dos Santos, C. A. T. et al. A review on the applications of portable near-infrared spectrometers in the agro-food industry. **Applied spectroscopy**, v. 67, n. 11, p. 1215-1233, 2013.

Dos Santos, P. M.; Pereira-Filho, E. R. Digital image analysis—an alternative tool for monitoring milk authenticity. **Analytical Methods**, v. 5, n. 15, p. 3669-3674, 2013.

Ducatti, G.C. et al. Aplicação da ressonância magnética como técnica de avaliação da composição de leite bovino. **Tekhne e Logos**, v. 14, n. 2, p. 28-37, 2023.

Dupont, D.; Croguennec, T.; Pochet, S.. Milk proteins-analytical methods. **Reference module in food science**, p. np, 2018.

Edwards, C. W.; Younus, M. A. Cow milk allergy. In: **StatPearls** [Internet]. Treasure Island (FL): StatPearls Publishing; 2019.

El Khoury, Y.; Hellwig, P. Far infrared spectroscopy of hydrogen bonding collective motions in complex molecular systems. **Chemical communications**, v. 53, n. 60, p. 8389-8399, 2017.

Ely, D. R.; Thommes, M.; Carvajal, M. T. Analysis of the effects of particle size and densification on NIR spectra. **Colloids and Surfaces A: Physicochemical and Engineering Aspects**, v. 331, n. 1-2, p. 63-67, 2008.

Faria, I. D. L. et al. Determination of formaldehyde in bovine milk by micellar electrokinetic chromatography with diode array detection. **LWT**, v. 163, p. 113473, 2022.

Fernandes, D. D. S. et al. The successive projections algorithm for interval selection in partial least squares discriminant analysis. **Analytical Methods**, v. 41, p. 1-9, 2016.

Fernandes, D. D. S. **Novas estratégias para seleção de variáveis por intervalos em problemas de classificação**. 2016. Tese (Doutorado em Química) - Programa de Pós Graduação em Química, Universidade Federal da Paraíba, João Pessoa-PB.

Fernandes, R.. 2. Concentrated and Dried Milk Products. In: **Microbiology Handbook**. p. 21-35, 2009.

Ferreira, M. M. C. et al. Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, v. 22, p. 724- 731, 1999.

Ferreira, M. M. C. **Quimiometria: conceitos, métodos e aplicações**. Editora da UNICAMP, 2015.

Finete, V. L. M. et al. Is it possible to screen for milk or whey protein adulteration with melamine, urea and ammonium sulphate, combining Kjeldahl and classical spectrophotometric methods. **Food chemistry**, v. 141, n. 4, p. 3649-3655, 2013.

Fiquene, L. P. et al. Alergia à proteína do leite de vaca: uma revisão de literatura. **Revista Ibero-Americana de Humanidades, Ciências e Educação**, v. 10, n. 4, p. 2704-2710, 2024.

Fonseca, J.F.; Bruschi, J. H. A caprinocultura leiteira no Brasil – uma visão histórica. Belo Horizonte: Caprileite, 2009.

- Fonseca, A. C.S. et al. Hierarchical classification models and Handheld NIR spectrometer to human blood stains identification on different floor tiles. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 267, p. 120533, 2022.
- Foroutan, A. et al. Chemical composition of commercial cow's milk. **Journal of agricultural and food chemistry**, v. 67, n. 17, p. 4897-4914, 2019.
- Gałuszka, A.; Migaszewski, Z.; Namieśnik, J. The 12 principles of green analytical chemistry and the SIGNIFICANCE mnemonic of green analytical practices. **TrAC Trends in Analytical Chemistry**, v. 50, p. 78-84, 2013.
- Galvão, R. K. H. et al. A method for calibration and validation subset partitioning. **Talanta**, v. 67, p. 736-740, 2005.
- Genis, D. O. et al. Development of synchronous fluorescence method for identification of cow, goat, ewe and buffalo milk species. **Food Control**, v. 108, p. 106808, 2020.
- Giuliano, B. M. et al. Interstellar ice analogs: H₂O ice mixtures with CH₃OH and NH₃ in the far-IR region. **Astronomy & Astrophysics**, v. 592, p. A81, 2016.
- Gobis, M. A.; Campanatti, R. Os benefícios da aplicação de ferramentas de gestão de qualidade dentro das indústrias do setor alimentício. **Revista Hórus**, v. 7, n. 01, p. 26-40, 2012.
- Goi, A. et al. Handheld near-infrared spectrometer allows on-line prediction of beef quality traits. **Meat Science**, v. 184, p. 108694, 2022.
- Gomes, A. A. et al. Variable selection in the chemometric treatment of food data: A tutorial review. **Food chemistry**, v. 370, p. 131072, 2022.
- Gomes, A. A. et al. Pattern recognition techniques in food quality and authenticity: A guide on how to process multivariate data in food analysis. **TrAC Trends in Analytical Chemistry**, v. 164, p. 117105, 2023.
- Granato, D. et al. Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. **Trends in Food Science & Technology**, v. 72, p. 83-90, 2018.
- Grassi, S.; Alamprese, C. Advances in NIR spectroscopy applied to process analytical technology in food industries. **Current Opinion in Food Science**, v. 22, p. 17-21, 2018.
- Gu, F. et al. Multi-omics revealed the effects of rumen-protected methionine on the nutrient profile of milk in dairy cows. **Food Research International**, v. 149, p. 110682, 2021.
- Gullifa, G. et al. Portable NIR spectroscopy: The route to green analytical chemistry. **Frontiers in Chemistry**, v. 11, p. 1214825, 2023.

Hackley, P. C.; Lünsdorf, N. K.. Application of Raman spectroscopy as thermal maturity probe in shale petroleum systems: insights from natural and artificial maturation series. **Energy & Fuels**, v. 32, n. 11, p. 11190-11202, 2018.

Haghighi, H. et al. Comprehensive characterization of active chitosan-gelatin blend films enriched with different essential oils. **Food Hydrocolloids**, v. 95, p. 33-42, 2019.

Handford, C. E.; Campbell, K.; Elliott, C. T. Impacts of milk fraud on food safety and nutrition with special emphasis on developing countries. **Comprehensive Reviews in Food Science and Food Safety**, v. 15, n. 1, p. 130-142, 2016.

Harrou, F.; Kadri, F.; Chaabane, S.; Tahon, C.; Sun, Y. Improved principal component analysis for anomaly detection: Application to an emergency department. **Computers & Industrial Engineering**, v. 88, p. 63-77, 2015.

Huang, J.; Romero-Torres, S.; Moshgbar, M. Raman: practical considerations in data pre-treatment for Nir and Raman spectroscopy. **American pharmaceutical review**, v. 13, n. 6, p. 116, 2010.

Huang, Y. et al. Distribution assessment and quantification of counterfeit melamine in powdered milk by NIR imaging methods. **Food Chemistry**, v. 177, p. 174-181, 2015.

James, G et al. **An Introduction to Statistical Learning**. Springer, 2013.

Jiang, Y. et al. Research progress on portable electrochemical sensors for detection of mycotoxins in food and environmental samples. **Chemical Engineering Journal**, p. 149860, 2024.

Kaliappan, J. et al. Impact of cross-validation on machine learning models for early detection of intrauterine fetal demise. **Diagnostics**, v. 13, n. 10, p. 1962, 2023.

Kamal, M.; Karoui, R.. Analytical methods coupled with chemometric tools for determining the authenticity and detecting the adulteration of dairy products: A review. **Trends in Food Science & Technology**, v. 46, n. 1, p. 27-48, 2015.

Käppler, A. et al. Analysis of environmental microplastics by vibrational microspectroscopy: FTIR, Raman or both?. **Analytical and bioanalytical chemistry**, v. 408, p. 8377-8391, 2016.

Karunathilaka, S. R. et al. Non-targeted NIR spectroscopy and SIMCA classification for commercial milk powder authentication: A study using eleven potential adulterants. **Heliyon**, v. 4, n. 9, 2018.

Kennard, R. W.; Stone, L. A. Computer aided design of experiments. **Technometrics**, v. 11, p. 137-148. 1969.

Krepper, G. et al. Determination of fat content in chicken hamburgers using NIR spectroscopy and the Successive Projections Algorithm for interval selection in PLS regression (iSPA-PLS). **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v.189, p. 300-306, 2018.

Kucheryavskiy, S. et al. Procrustes cross-validation—a bridge between cross-validation and independent validation sets. **Analytical chemistry**, v. 92, n. 17, p. 11842-11850, 2020.

Li, J.Y. et al. Quantification of cow milk in adulterated goat milk by HPLC-MS/MS using N-acetylglucosamine as a reliable biomarker of cow milk. **Journal of Food Composition and Analysis**, v. 123, p. 105583, 2023.

Lim, J. et al. Detection of melamine in milk powders using near-infrared hyperspectral imaging combined with regression coefficient of partial least square regression model. **Talanta**, v. 151, p. 183-191, 2016.

Liu, Y.; Cai, J.; Zhang, F. Functional comparison of breast milk, cow milk and goat milk based on changes in the intestinal flora of mice. **Lwt**, v. 150, p. 111976, 2021.

Liu, Z.; Li, C.; Pryce, J.; Rochfort, S. Comprehensive characterization of bovine milk lipids: Triglycerides. **Acs Omega**, v. 5, n. 21, p. 12573-12582, 2020. DOI: 10.1021/acsomega.0c01841.

Lohumi, S. et al. A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. **Trends in Food Science & Technology**, v. 46, n. 1, p. 85-98, 2015.

Long, D. A. **The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules**. Wiley, 2002.

Lopez-Calleja, I. et al. Rapid detection of cows' milk in sheeps' and goats' milk by a species-specific polymerase chain reaction technique. **Journal of Dairy Science**, v. 87, n. 9, p. 2839-2845, 2004.

Lu, G.; Haes, A. J.; Forbes, T. Z. Detection and identification of solids, surfaces, and solutions of uranium using vibrational spectroscopy. **Coordination chemistry reviews**, v. 374, p. 314-344, 2018.

Luca, S. et al. Characterization of the effects of different roasting conditions on coffee samples of different geographical origins by HPLC-DAD, NIR and chemometrics. **Microchemical Journal**, v. 129, p. 348–361, 2016.

Mabood, F. et al. Development of new NIR-spectroscopy method combined with multivariate analysis for detection of adulteration in camel milk with goat milk. **Food Chemistry**, v. 221, p. 746-750, 2017a.

Mabood, F. et al. FT-NIRS coupled with chemometric methods as a rapid alternative tool for the detection & quantification of cow milk adulteration in camel milk samples. **Vibrational Spectroscopy**, v. 92, p. 245-250, 2017b.

Madeira, A. N.. **Otimização do processo de spray drying pelo uso de pré-desumidificadores no ar de entrada**. 2009. Dissertação (Mestrado em Engenharia Mecânica) - Universidade de Taubaté, Taubaté-SP.

Maia, M. da S. **Considerações sobre a caprinocultura no Brasil**. Rio branco: Embrapa-CPAF-acre, 1994.

Manley, M.; Baeten, V. Spectroscopic technique: Near infrared (NIR) spectroscopy. **Modern techniques for food authentication**, p. 51-102, 2018.

Marquetti, I. et al. Partial least square with discriminant analysis and near infrared spectroscopy for evaluation of geographic and genotypic origin of arabica coffee. **Computers and Electronics in Agriculture**, v. 121, p. 313–319, 2016.

Máquina, A. D. V. et al. Determination of Adulteration of the B10 Blend of Diesel and Crambe Biodiesel Using Proton Nuclear Magnetic Resonance (1 H NMR) Spectroscopy with a Data Driven Soft Independent Modeling of Class Analogy (DD-SIMCA) Model. **Analytical Letters**, v. 54, n. 5, p. 790-801, 2021.

Ménard, H. et al. Fraude alimentar no Brasil: panorama e impactos para leite e derivados dos últimos 25 anos. **Observatório de la economía latinoamericana**, v. 22, n. 4, p. e3991-e3991, 2024.

Miaw, C. S. W. **Detecção de fraudes em néctares de frutas: análises espectroscópicas aliadas a métodos de calibração e classificação multivariada**. 2018. Tese (Doutorado em Ciência de Alimentos) - Faculdade de Farmácia, Universidade Federal de Minas Gerais, Belo Horizonte-MG.

Milani, M. I. et al. Authentication of roasted and ground coffee samples containing multiple adulterants using NMR and a chemometric approach. **Food Control**, v. 112, p. 107104, 2020.

Moore, J. C.; Spink, J.; Lipp, M. Development and application of a database of food ingredient fraud and economically motivated adulteration from 1980 to 2010. **Journal of food science**, v. 77, n. 4, p. R118-R126, 2012.

Morin, J. F.; Lees, M. **Food integrity handbook. A guide to food authenticity issues and analytical solutions**, 2018.

Moufid, M. et al. Pollution parameters evaluation of wastewater collected at different treatment stages from wastewater treatment plant based on E-nose and E-tongue systems combined with chemometric techniques. **Chemometrics and Intelligent Laboratory Systems**, v. 227, p. 104593, 2022.

Nakamoto, K.. **Infrared and Raman spectra of inorganic and coordination compounds, part B: applications in coordination, organometallic, and bioinorganic chemistry**. John Wiley & Sons, 2009.

Namonje-Kapembwa, T.; Chiwawa, H.; Sitko, N. Analysis of Goat Production and Marketing Among Smallholder Farmers Zambia. **Small Ruminant Research**, v. 208, p. 106620, 2022.

Nayik, G. A. et al. Recent insights into processing approaches and potential health benefits of goat milk and its products: a review. **Frontiers in nutrition**, v. 8, p. 789117, 2021.

Neves, M. D. G.; Poppi, R. J.; Siesler, H. W. Rapid determination of nutritional parameters of pasta/sauce blends by handheld near-infrared spectroscopy. **Molecules**, v. 24, n. 11, p. 2029, 2019.

Neves, M. D. G. **Avaliação de diferentes métodos quimiométricos para autenticação, classificação e quantificação em matrizes alimentícias empregando espectroscopia vibracional**. 2021. Tese (Doutorado em Ciências) - Instituto de Química, Universidade Estadual de Campinas, Campinas-SP.

Neves, M. G.; Poppi, R. J. Authentication and identification of adulterants in virgin coconut oil using ATR/FTIR in tandem with DD-SIMCA one class modeling. **Talanta**, v. 219, p. 121-338, 2020.

Niculescu, B. M.; Andrei, G. Principal component analysis as a tool for enhanced well log interpretation. **Rev. Roum. GÉOPHYSIQUE**, v. 60, p. 49-61, 2016.

Nóbrega, R. O. da et al. **Classificação de cafés solúveis usando espectroscopia NIR e quimiometria**. 2021. Dissertação (Mestrado em Química) - Programa de Pós Graduação em Química, Universidade Federal da Paraíba, João Pessoa-PB.

Núñez-Sánchez, N. et al. Near Infrared Spectroscopy (NIRS) for the determination of the milk fat fatty acid profile of goats. **Food Chemistry**, v. 190, p. 244-252, 2016.

Oliveira, A. P. D. et al. Symbiotic goat milk ice cream with umbu fortified with autochthonous goat cheese lactic acid bacteria. **LWT**, v. 141, p. 110888, 2021.

Oliveira, E. Triglicerídeos. **Revista de Ciência Elementar**, v. 3, n. 2, 2015.

Oliveira, S. R. B.; Moraes, L. D. S.; Coelho, C. P. Fraudes em alimentos industrializados. **PubSaúde**, v. 5, p. 115, 2021.

Oliveri, P.; Simonetti, R. Chemometrics for food authenticity applications. In: Advances in food authenticity testing. **Woodhead Publishing**, p. 701-728, 2016.

Panciere, B. M.; Ribeiro, L. F. Detecção e ocorrência de fraudes no leite fluido ou derivados. **Revista GeTeC**, v. 10, n. 27, 2021.

Pasquini, C. Near infrared spectroscopy: A mature analytical technique with new perspectives—A review. **Analytica chimica acta**, v. 1026, p. 8-36, 2018.

Park, Y. W. et al. Physico-chemical characteristics of goat and sheep milk. **Small ruminant research**, v. 68, n. 1-2, p. 88-113, 2007.

Pereira, E. V. S. et al. Goat milk authentication by one-class classification of digital image-based fingerprint signatures: Detection of adulteration with cow milk. **Microchemical Journal**, v. 180, p. 107640, 2022.

Pereira, E. V. S. **Novas estratégias analíticas baseadas em espectroscopia no infravermelho próximo e imagens digitais para identificação e quantificação de adulterações em leite caprino**. 2020. Tese (Doutorado em Química) - Programa de pós graduação em Química, Universidade Federal da Paraíba, João Pessoa.

Pereira, E. V. S. et al. Simultaneous determination of goat milk adulteration with cow milk and their fat and protein contents using NIR spectroscopy and PLS algorithms. **Lwt**, v. 127, p. 109427, 2020.

Pesic, M. et al. Qualitative and quantitative analysis of bovine milk adulteration in caprine and ovine milks using native-PAGE. **Food Chemistry**, v. 125, n. 4, p. 1443-1449, 2011.

Pietrzak-Fiećko, R.; Kamelska-Sadowska, A. M. The comparison of nutritional value of human milk with other mammals' milk. **Nutrients**, v. 12, n. 5, p. 1404, 2020.

Pimenta, J. L. L.A. et al. Fatores inerentes ao consumo de leite de cabra e derivados no Brasil. **Research, Society and Development**, v. 10, n. 12, p. e53101220175-e53101220175, 2021.

Pomerantsev, A. L.; Rodionova, O. Y.. Procrustes Cross-Validation of short datasets in PCA context. **Talanta**, v. 226, p. 122104, 2021.

Pomerantsev, A. L.; Rodionova, O. Y.. On the type II error in SIMCA method. **Journal of Chemometrics**, v. 28, n. 6, p. 518-522, 2014a.

Pomerantsev, A. L.; Rodionova, O. Y. Concept and role of extreme objects in PCA/SIMCA. **Journal of Chemometrics**, v. 28, n. 5, p. 429-438, 2014b.

PONTES, M. J. C. **Algoritmo das projeções sucessivas para a seleção de variáveis espectrais em problemas de classificação**. 2009. Tese (Doutorado em Química) - Curso de Pós Graduação em Química, Universidade Federal da Paraíba, João Pessoa-PB.

Pontes, M. J. C. et al. Classification of distilled alcoholic beverages and verification of adulteration by near infrared spectrometry. **Food research international**, v. 39, n. 2, p. 182-189, 2006.

Prieto, Nuria et al. A review of the principles and applications of near-infrared spectroscopy to characterize meat, fat, and meat products. **Applied spectroscopy**, v. 71, n. 7, p. 1403-1426, 2017.

Qin, J.; Chao, K.; Kim, M. S. Simultaneous detection of multiple adulterants in dry milk using macro-scale Raman chemical imaging. **Food chemistry**, v. 138, n. 2-3, p. 998-1007, 2013.

Raman, C. V. Part II.—The Raman effect. Investigation of molecular structure by light scattering. **Transactions of the Faraday Society**, v. 25, p. 781-792, 1929.

Ramos Neto, K. X. C. et al. Leite de cabra: qualidade x instrução normativa nº 37/2000 do MAPA. **Revista Extensão em Foco Palotina**, n. 22, p. 51-61, 2021.

Raypah, M. E. et al. Near-infrared spectroscopy as a potential COVID-19 early detection method: A review and future perspective. **Sensors**, v. 22, n. 12, p. 4391, 2022.

Reis, A. C. et al. **Caprinocultura leiteira no brasil**. 2023. Monografia (Bacharelado em Zootecnia) - Instituto Federal de Educação, Ciência e Tecnologia Goiano–Campus Rio Verde, GO.

Ribeiro, A. C. O melhoramento animal e a qualidade do leite dos caprinos no Brasil. **Anais do VII Simpósio Brasileiro de Melhoramento Animal**. São Carlos, 2008.

Rinnan, Å.; Van Den Berg, F.; 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.

Ringsted, T.; Siesler, H. W.; Engelsen, S. B. Monitoring the staling of wheat bread using 2D MIR-NIR correlation spectroscopy. **Journal of Cereal Science**, v. 75, p. 92-99, 2017.

Rodionova, O. Y.; Balyklova, K. S.; Titova, A. V.; Pomerantsev, A. L. Application of NIR spectroscopy and chemometrics for revealing of the ‘high quality fakes’ among the medicines. **Forensic Chemistry**, v. 8, p. 82–89, 2018.

Rodionova, O. Y.; Kucheryavskiy, S.; Pomerantsev, A. L. Efficient tools for principal component analysis of complex data— a tutorial. **Chemometrics and Intelligent Laboratory Systems**, v. 213, p. 104304, 2021.

Rodionova, O. Y.; Pomerantsev, A. L. Chemometric tools for food fraud detection: The role of target class in non-targeted analysis. **Food chemistry**, v. 317, p. 126448, 2020.

Rodionova, O. Y.; Titova, A. V.; Pomerantsev, A. L. Discriminant analysis is an inappropriate method of authentication. **TrAC Trends in Analytical Chemistry**, v. 78, p. 17-22, 2016.

Rodrigues Júnior, P. H. **Utilização de FT-Raman e ferramentas quimiométricas na determinação rápida de fraude por maltodextrina e hidrólise da lactose em leite em pó**. 2015. Dissertação (Mestrado em Ciência de Alimentos) - Programa de Pós-Graduação em Ciência e Tecnologia de Alimentos, Universidade de Viçosa, Minas Gerais.

Rodrigues, N. P. A. et al. Milk adulteration: Detection of bovine milk in bulk goat milk produced by smallholders in northeastern Brazil by a duplex PCR assay. **Journal of dairy science**, v. 95, n. 5, p. 2749-2752, 2012.

Rodríguez, S. D.; Rolandelli, G.; Buera, M. P. Detection of quinoa flour adulteration by means of FT-MIR spectroscopy combined with chemometric methods. **Food chemistry**, v. 274, p. 392-401, 2019.

Roy, D. et al. Composition, structure, and digestive dynamics of milk from different species—A review. **Frontiers in Nutrition**, v. 7, p. 577759, 2020.

Ruis, M.. **Quantificação e caracterização da diversidade genética de esporulados isolados da matriz leite em pó produzido e comercializado no Rio Grande do Sul**. 2022.

Dissertação (Mestrado em Ciência e Tecnologia de Alimentos) - Universidade Estadual do Rio Grande do Sul, Unidade Encantado. Encantado.

Sala O. **Fundamentos da Espectroscopia Raman e no Infravermelho**. Unesp, 2008.

Santana, F. B.; Borges Neto, W. ; Poppi, R. J. Random forest as one-class classifier and infrared spectroscopy for food adulteration detection. **Food chemistry**, v. 293, p. 323-332, 2019.

Santana, F. B. et al. Experimento didático de quimiometria para classificação de óleos vegetais comestíveis por espectroscopia no infravermelho médio combinado com análise discriminante por mínimos quadrados parciais: um tutorial, parte V. **Química Nova**, v. 43, n. 3, p. 371-381, 2020.

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.

Scaquette, B. L.; Silva, J. **Leite: reflexões sobre as verdades e os mitos do leite**. 2021. Trabalho de Conclusão de Curso (Tecnólogo em Alimentos) - Faculdade de Tecnologia de Marília Estudante Rafael Almeida Camarinha - FATEC, Marília -SP.

Schrader, B. (Ed.). **Infrared and Raman spectroscopy: methods and applications**. John Wiley & Sons, 2008.

Schrödinger, E. **Phys. Rev.**,v. 28, p. 1049, 1926.

Schuchardt, P.; Siesler, H. W. Real-time analysis of the polymerization kinetics of 1, 4-butanediol and 4, 4'-diphenylmethanediisocyanate by fiber-coupled Fourier transform infrared spectroscopy. **Analytical and bioanalytical chemistry**, v. 409, p. 833-839, 2017.

Shah, A.; Chauhan, Y.; Chaudhury, B. Análise de componentes principais baseada na construção e avaliação do índice de criptomoea. **Expert Systems with Applications**, v. 163, p. 113-796, 2020.

Silva, S. F. **Análise de triagem para confirmação do estado de conservação de insulina humana NPH injetável**. 2020. Dissertação (Mestrado em Química) - Programa de Pós Graduação em Química, Universidade Federal da Paraíba.

Silva, C. S.; Braz, A.; Pimentel, M. F.. Vibrational spectroscopy and chemometrics in forensic chemistry: critical review, current trends and challenges. **Journal of the Brazilian Chemical Society**, v. 30, n. 11, p. 2259-2290, 2019.

Silva, C. S.; Pimentel, M. F.; Amigo, J. M.; Honorato, R. S.; Pasquini, C. Detecting semen stains on fabrics using near infrared hyperspectral images and multivariate models. **TrAC Trends in Analytical Chemistry**, v. 55, p. 23-25, 2017

Silva, W. M. **Determinação simultânea de adulteração em leite de cabra por adição de leite de vaca e teor de lipídios usando NIR portátil e algoritmos PLS**. 2023. Dissertação

(Mestrado em Ciências Farmacêuticas) Programa de Pós Graduação em Ciências Farmacêuticas, Universidade Estadual da Paraíba. Campina Grande-PB.

Silveira, F. L. **Determinação da concentração de gorduras saturadas e insaturadas em alimentos comerciais por espectroscopia Raman dispersiva**. 2008. Dissertação (Mestrado em Engenharia Biomédica) - Programa de pós graduação da CAPES, Universidade do Vale do Paraíba.

Skoog, D.A.; Holler, F. J.; Crouch, S. R., **Principles of instrumental analysis**; Cengage Learning, Ed.; Bookman, 2018.

Spink, J.; Moyer, D. C. Defining the public health threat of food fraud. **Journal of food science**, v. 76, n. 9, p. R157-R163, 2011.

Socrates G. **“Infrared and Raman Characteristic Group Frequencies: Tables and Charts - γ edition”** John Wiley & Sons. 2004.

Song, H.; Xue, H.; Han, Y. Detection of cow’s milk in Shaanxi goat’s milk with an ELISA assay. **Food Control**, v. 22, n. 6, p. 883-887, 2011.

Sousa, E. R.; Frota, C. S.; Costa, C. H. C.; Silva, G. S.; Sampaio, D. G.; Avaliação da oxidação e de parâmetros de qualidade do óleo de Babacu por espectroscopia no infravermelho médio com transformada de Fourier (FTIR) e calibração multivariada. **Revista Virtual de Química**, v. 11, p. 849-865, 2019.

Souza, A. M.; Poppi, R.J. Experimento didático de quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: um tutorial, parte I. **Química Nova**, v. 35, p. 223- 229, 2012.

Souza, E. R. A. Fatores que afetam a ocorrência de Leite Instável Não Ácido (LINA) e impactos na indústria: 2024. Monografia (Bacharelado em Zootecnia) - Curso De Graduação Em Zootecnia, Universidade Federal Rural da Amazônia, Belém-PA

Stuart, B. H. **Infrared spectroscopy: fundamentals and applications**. John Wiley & Sons, 2004.

Suh, J. H. Critical review: Metabolomics in dairy science–Evaluation of milk and milk product quality. **Food Research International**, v. 154, p. 110984, 2022.

Tahir, Haroon Elrasheid et al. Recent progress in rapid analyses of vitamins, phenolic, and volatile compounds in foods using vibrational spectroscopy combined with chemometrics: A review. **Food Analytical Methods**, v. 12, p. 2361-2382, 2019.

Tarapoulouzi, M.; Ortone, V.; Cinti, S. Heavy metals detection at chemometrics-powered electrochemical (bio) sensors. **Talanta**, v. 244, p. 123410, 2022.

Teixeira, J. L. P. **Ensaio de digestão in vitro para estimativa de bioacessibilidade e biodisponibilidade de minerais e avaliação da viabilidade do emprego de NIR e quimiometria para detecção de adulterações em leite de cabra e derivados**. 2022. Tese

(Doutorado em Ciência de Alimentos) Faculdade de Engenharia de Alimentos, Universidade Estadual de Campinas, Campinas-SP

Thomé, B. R. et al. Gestão da qualidade nas agroindústrias de suínos de Francisco Beltrão–Paraná. **Revista Espacios**, v. 38, n. 21, p. 9, 2017.

Tibola, C. S. et al. **Espectroscopia no Infravermelho próximo para avaliar indicadores de qualidade tecnológica e contaminantes em grãos**. Embrapa, 2018.

Tosato, F. et al. Análise de amostras de combustíveis por fotometria, NIR portátil e RMN de ¹H - uma comparação com os resultados encontrados por técnicas normatizadas. **Química Nova**, v. 43, p. 155-167, 2020.

Uberti, A.; Pinto A.T. **O Leite e suas principais fraudes**. Editora Científica Digital. 2022.

Ullah, R. et al. Potentiality of using front face fluorescence spectroscopy for quantitative analysis of cow milk adulteration in buffalo milk. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 225, p. 117518, 2020.

Valand, R. et al. A review of Fourier Transform Infrared (FTIR) spectroscopy used in food adulteration and authenticity investigations. **Food Additives & Contaminants: Part A**, v. 37, n. 1, p. 19-38, 2020.

Vanzella, E.; Santos, W. S. O controle de qualidade, por meio das ferramentas BPF e APPCC, em uma linha de produção de uma indústria de alimentos. **Destarte**, v. 5, n. 2, p. 76-90, 2015.

Veloso, L. T. **Um estudo comparativo de técnicas de validação cruzada aplicadas a modelos para dados desbalanceados**. 2022. Dissertação (Mestrado em Ciências) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo-SP.

Vera, D. A. et al. A Monte Carlo study of near infrared light propagation in the human head with lesions—a time-resolved approach. **Biomedical Physics & Engineering Express**, v. 8, n. 3, p. 035005, 2022.

Verruck, S.; Dantas, A.; Prudencio, E. S. Functionality of the components from goat's milk, recent advances for functional dairy products development and its implications on human health. **Journal of functional foods**, v. 52, p. 243-257, 2019.

Vieitez, I. et al. Composition of fatty acids and triglycerides in goat cheeses and study of the triglyceride composition of goat milk and cow milk blends. **Journal of Food Composition and Analysis**, v. 48, p. 95-101, 2016.

Wan, S. K. et al. Quick measurement method of condensation point of diesel based on temperature-compensation model [基于温度修正模型的柴油凝点快速检测方法]. **Guang Pu Xue Yu Guang Pu Fen Xi Spectroscopy Spectr. Anal**, v. 41, p. 3111-3116, 2021.

Wang, X. et al. Quantitative detection of malachite green in sediment by a time-resolved immunofluorescence method combined with a portable 3D printing equipment platform. **Science of the Total Environment**, v. 855, p. 158897, 2023.

Williams, P.; Norris, K.. **Near-infrared technology in the agricultural and food industries.** Cereals & Grains Assn, 1987.

Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. **Chemometrics and Intelligent Laboratory Systems**, v. 2, n. 1-3, p. 37-52, 1987.

Workman Jr, J.; Weyer, L.. **Practical guide to interpretive near-infrared spectroscopy.** CRC press, 2007.

Wu, T. et al. Identification and quantitation of melamine in milk by near-infrared spectroscopy and chemometrics. **Journal of Spectroscopy**, v. 2016, n. 1, p. 6184987, 2016.

Xing, R. et al. Application of portable measuring equipment in the field of industrial measurement. In: **Journal of Physics: Conference Series.** IOP Publishing, 2021. p. 012153.

Xu, L. et al. A MATLAB toolbox for class modeling using one-class partial least squares (OCPLS) classifiers. **Chemometrics and Intelligent Laboratory Systems**, v. 139, p. 58-63, 2014.

Xu, L. et al. Enhanced specificity for detection of frauds by fusion of multi-class and one-class partial least squares discriminant analysis: geographical origins of Chinese shiitake mushroom. **Food analytical methods**, v. 9, p. 451-458, 2016.

Xu, L. et al. One-class partial least squares (OCPLS) classifier. **Chemometrics and Intelligent Laboratory Systems**, v. 126, p. 1-5, 2013.

Yaman, H.. A rapid method for detection adulteration in goat milk by using vibrational spectroscopy in combination with chemometric methods. **Journal of Food Science and Technology**, v. 57, n. 8, p. 3091-3098, 2020.

Yan, H.; Siesler, H. W. Identification performance of different types of handheld near-infrared (NIR) spectrometers for the recycling of polymer commodities. **Applied spectroscopy**, v. 72, n. 9, p. 1362-1370, 2018.

Yao, S.; Aykas, D. P.; Rodriguez-Saona, L. Rapid authentication of potato chip oil by vibrational spectroscopy combined with pattern recognition analysis. **Foods**, v. 10, n. 1, p. 42, 2020.

Yazgan, N. et al. Discrimination of milk species using Raman spectroscopy coupled with partial least squares discriminant analysis in raw and pasteurized milk. **Journal of the science of food and agriculture**, v. 100, n. 13, 2020.

Yuan, T. et al. Medium-and long-chain triacylglycerols composition in preterm and full-term human milk across different lactation stages. **Lwt**, v. 142, p. 110907, 2021.

Zhang, K. et al. Identifying muscle hemorrhage in rat cadavers with advanced decomposition by FTIR microspectroscopy combined with chemometrics. **Legal Medicine**, v. 47, p. 101-748, 2020.

Zhang, H. et al. Milk lipids characterization in relation to different heat treatments using lipidomics. **Food Research International**, v. 157, p. 111345, 2022.

Zhang, R.; Jia, W.; Shi, L.. A comprehensive review on the development of foodomics-based approaches to evaluate the quality degradation of different food products. **Food Reviews International**, v. 39, n. 8, p. 5563-5582, 2023.

Zhao, L.; Wang, J.; Mao, X. Composition and interfacial properties play key roles in different lipid digestion between goat and cow milk fat globules in vitro. **Food chemistry**, v. 374, p. 131538, 2022.

Zhiqiang, H. Analysis on the Development Status and Prospects of My Country's Dairy Industry. **China Dairy**, v. 21, p. 20-23, 2021.

Zontov, Y. V. et al. DD-SIMCA – A MATLAB GUI tool for data driven SIMCA approach. **Chemometrics and Intelligent Laboratory Systems**, v. 167, p. 23–28, 2017.

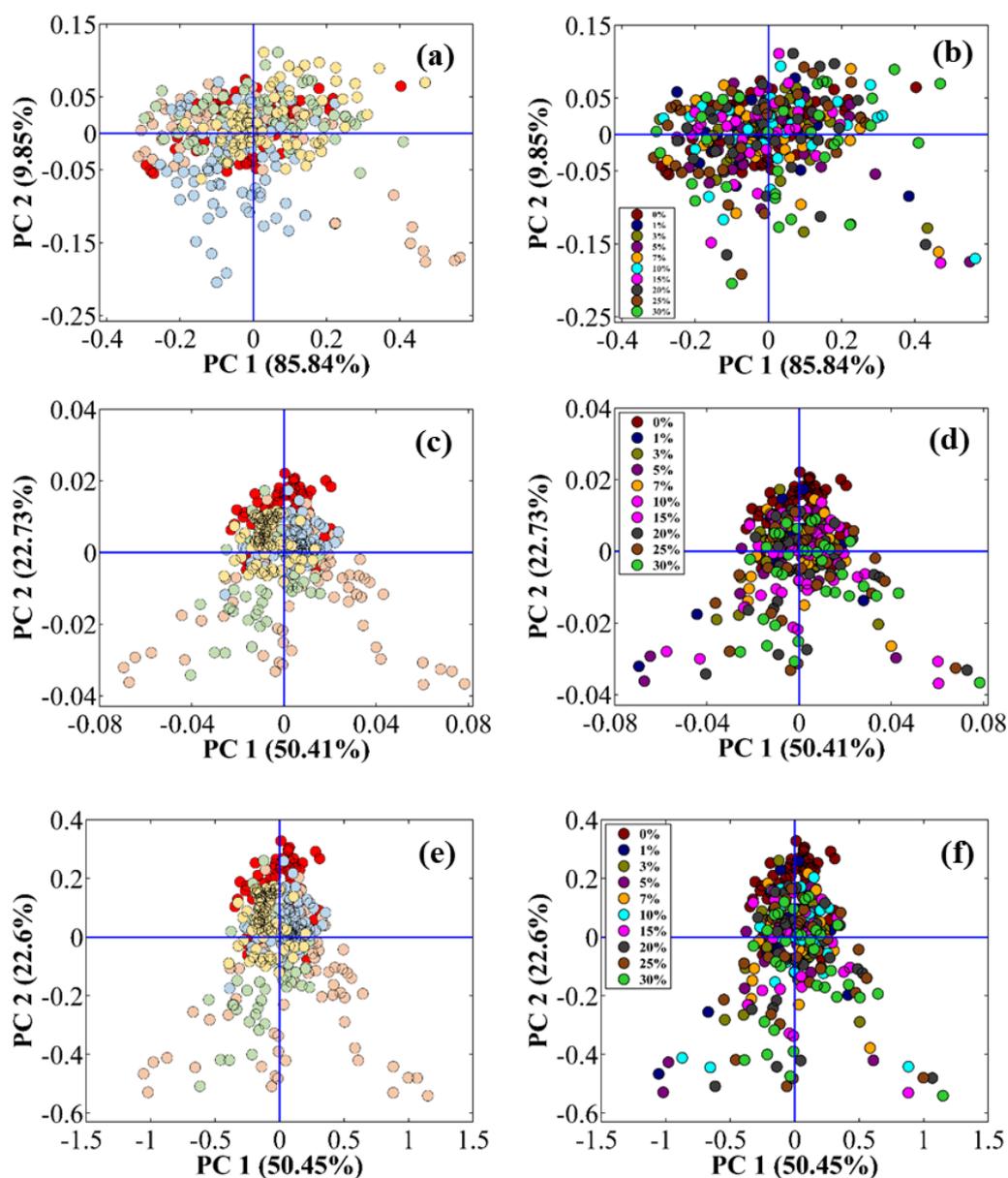
Capítulo 8

Apêndice

APÊNDICE

1. NIR

Abaixo seguem alguns gráficos de escores de PCA (**Figura A1**), referente a espectroscopia NIR, porém com uma visão diferente ao segregar as amostras de duas formas, por adulterante e por teor. Por meio dos gráficos, fica claro a existência de sutis formações de grupos, tanto por teor de adulteração quanto por tipo de adulterante.



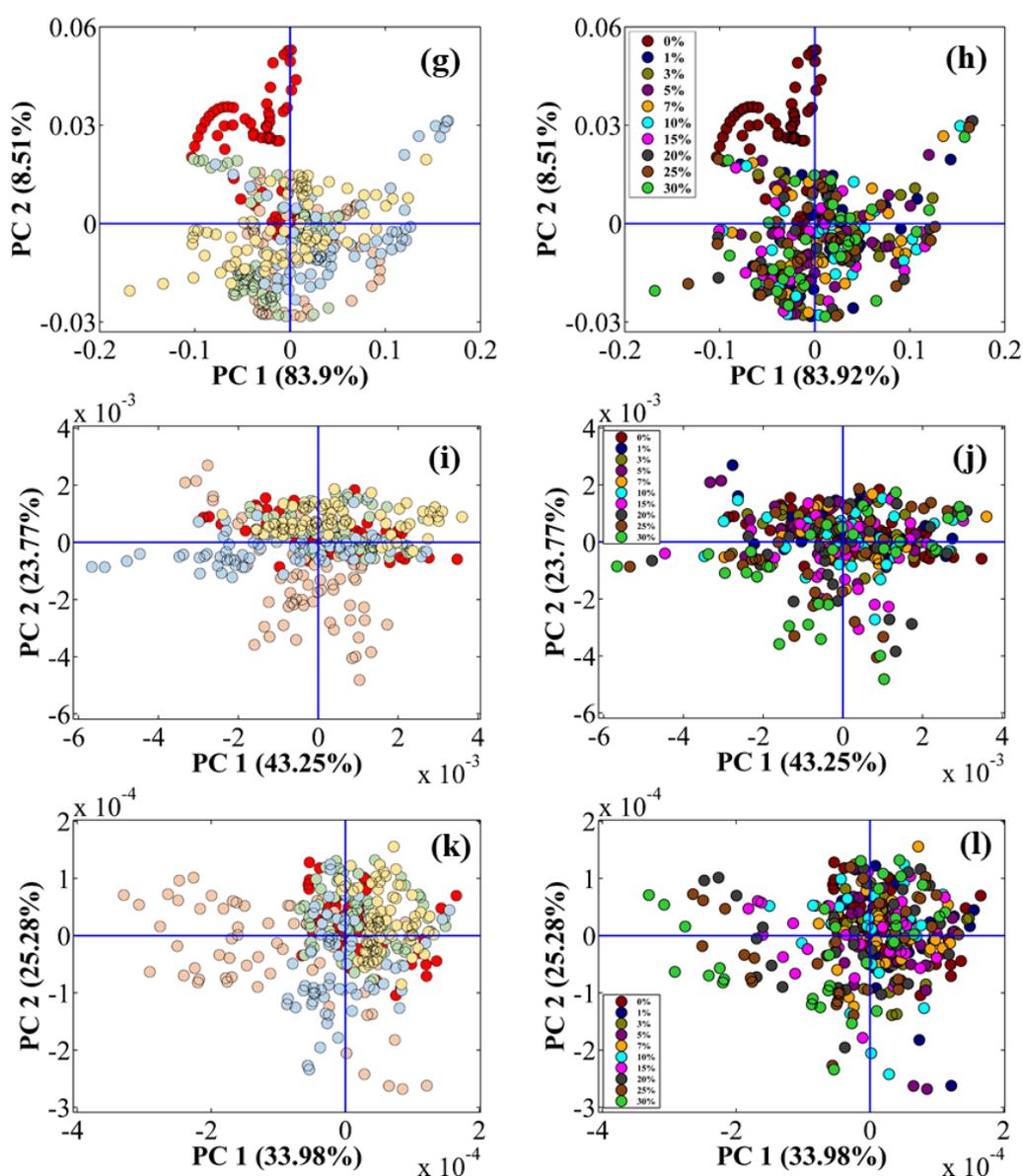


Figura A1. Gráficos de escores (PC1 x PC2) para os dados espectrais cortados a partir de 1660 nm (a), suavizados com filtro Savitzky – Golay 21 pontos acoplado com (b) OFF, (c) MSC e (d) SNV; e aplicando filtro derivativos de 1° e 2° ordens Savitzky – Golay com janelas de 21 (e) e 25 (f) pontos, respectivamente. Gráficos à esquerda: amostras de Leites caprino (em vermelho) e de misturas caprino e amido (em alaranjado), composto lácteo (em verde), extrato de soja (em azul) e leite de vaca (em amarelo). Gráficos à direita seguem a legenda das cores por teor de adulteração.

Como discutido anteriormente, é esperado que amostras com baixos teores de adulteração sejam confundidas com a classe alvo. Conseqüentemente, amostras com maior teor se distanciam mais das outras, e faz sentido, pois conforme o teor aumenta, mais distante da classe-alvo essas amostras ficam. Já no caso do tipo de adulterante, a sobreposição entre as classes não é esperada, porém pode ocorrer devido a semelhança na composição química do tipo de adulterante com a matriz adulterada, como é o caso da sobreposição entre as amostras de leites caprino e bovino. Por outro lado, é perceptível em todos os gráficos, com exceção

dos resultados apresentados nas **Figuras A1a e A1g**, que amostras adulteradas com amido (alaranjado) encontram-se mais distantes das demais, uma vez que diferem mais em termos de composição. Em cenários assim, é possível atribuir um maior poder de influência no agrupamento à composição química do adulterante específico do que ao teor de adulteração. Contudo, não é possível generalizar, pois nessa análise de triagem o único pré-processamento que conseguiu de fato separar a classe alvo das demais foi o Xs-21p-OFF.

2. RAMAN

A **Figura A2** ilustra os desempenhos dos modelos DD-SIMCA, na Abordagem 1, 2 e 3, usando espectros Raman pré-processados com OFF, nos quais foram encontrados erros de classificação.

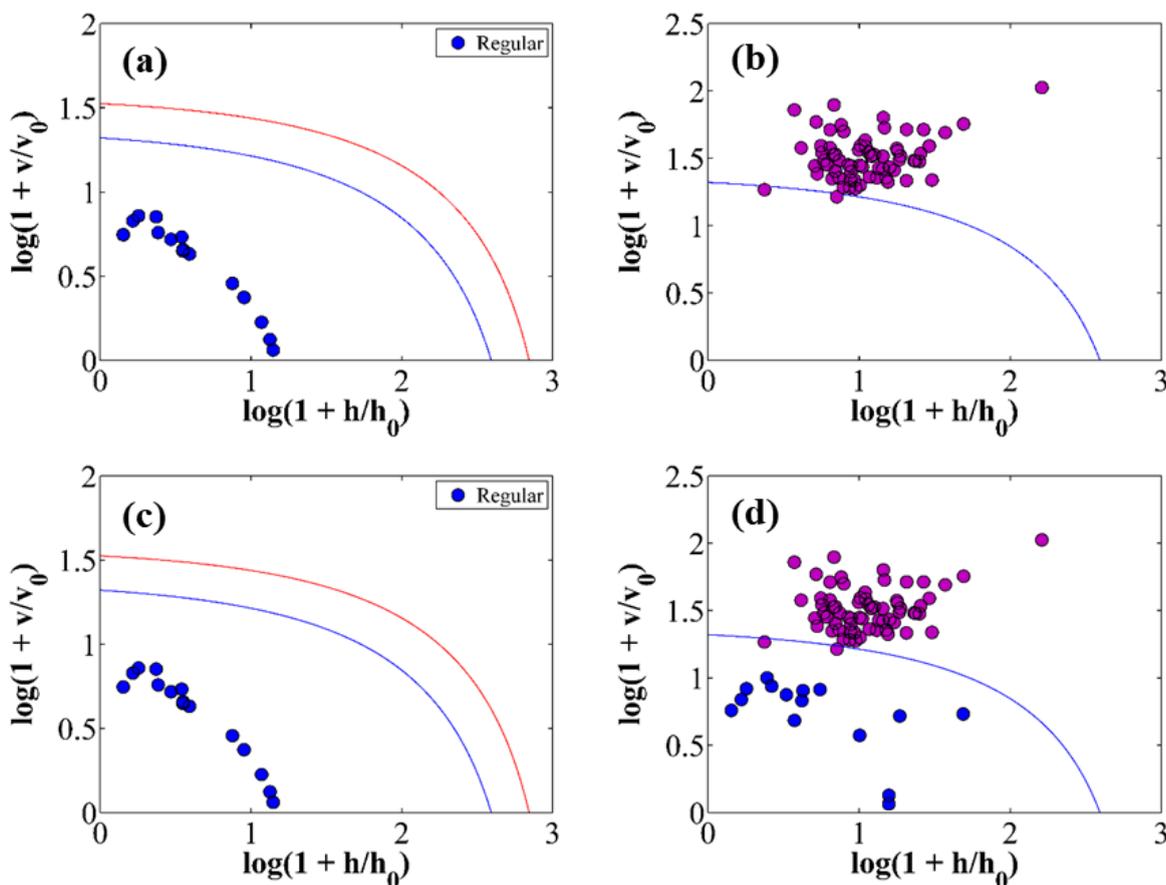


Figura A2. Gráficos de área de aceitação para o conjunto de treinamento (a, c) e teste (b, d) empregando DD-SIMCA, abordagens 1 (a, b) e 2 (c, d), pré-processamento OFF.

Nas **Figuras A2b e A2d** (Abordagem 1 e 2, respectivamente, com pré-processamento OFF, empregando 6 PCs para $\alpha = 0,01$), é possível identificar duas amostras adulteradas classificadas como pertencentes à classe das puras, sendo elas as amostras rotuladas como 20

e 50 no conjunto de teste. A amostra 20 é referente a um baixo teor de adulteração (1%), logo é esperado que tal amostra seja confundida com a classe alvo. No caso da amostra 50, houve um possível erro de rotulagem, o qual foi discutido na **Seção 5.2.1**.

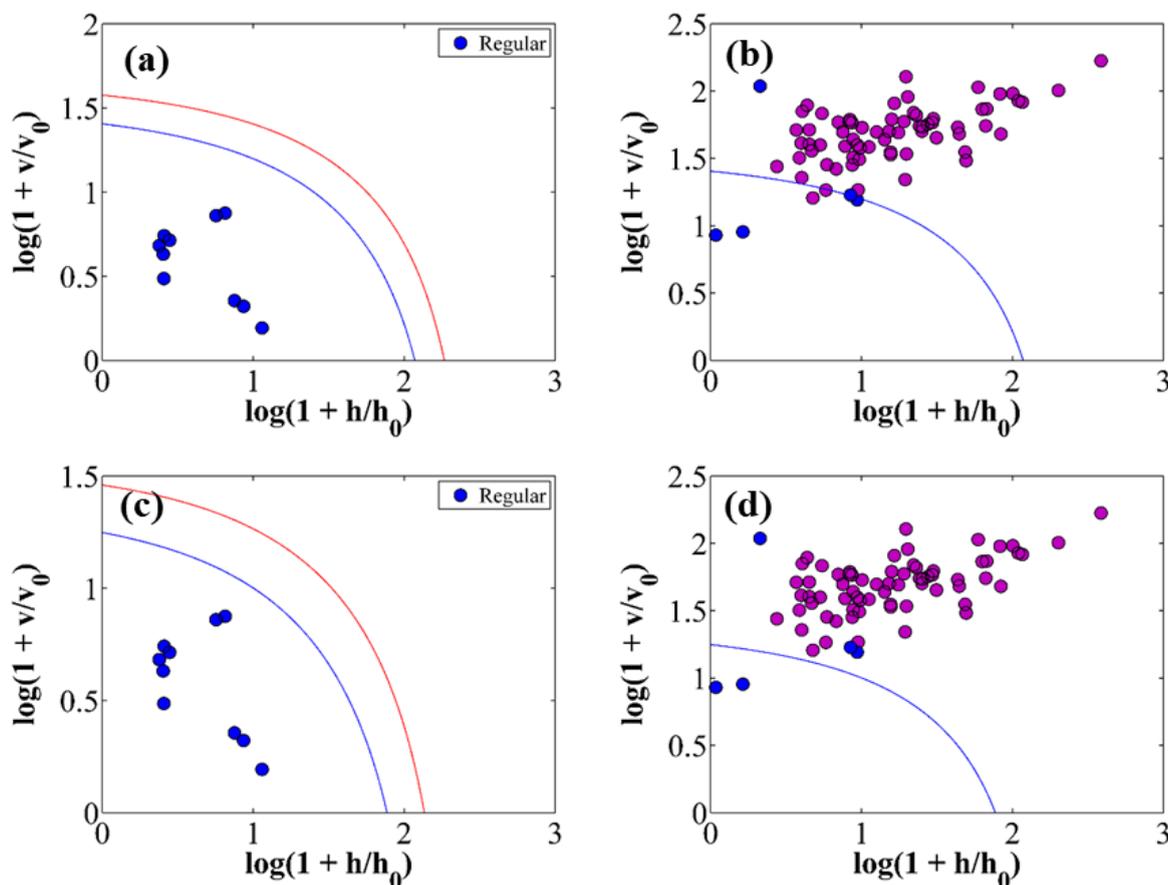


Figura A3. Gráficos de área de aceitação para o conjunto de treinamento (a, c) e teste (b, d) empregando DD-SIMCA, abordagem 3, para $\alpha = 0,01$ (a, b) e $\alpha = 0,05$ (c, d), pré-processamento OFF.

Na **Figura A3b** (Abordagem 3 com pré-processamento OFF, empregando 4 PCs para $\alpha = 0,01$), foram identificadas duas amostras puras (70 e 73) classificadas erroneamente, sendo projetadas fora da área de aceitação. A amostra 70 apresentou um espectro bastante ruidoso em certas bandas com esse pré tratamento, o que pode ter jogado-a para tão distante da área de aceitação. A amostra 73, apesar de ser classificada incorretamente, manteve-se próxima à área de aceitação da classe. Além disso, outras duas amostras da classe não-alvo (21 e 50) foram classificadas como puras. De maneira análoga à discussão anterior, a amostra 21 possui um baixo teor de adulteração (1%) e encontra-se próxima ao limite da área de aceitação, enquanto a amostra 50 refere-se ao possível erro de rotulagem.

Por fim, a **Figura A3d** (Abordagem 3 com pré-processamento OFF, empregando 4 PCs para $\alpha = 0,05$), aponta a presença de três erros de classificação (amostras 70, 71 e 73). A

amostra 71, apesar de ser classificada incorretamente, também manteve-se próxima à área de aceitação da classe. E no caso das demais, ambas já foram discutidas.

Para o último caso de rastreamento de amostras, a **Figura A4**, ilustra os desempenhos dos modelos DD-SIMCA, na Abordagem 2, usando espectros Raman pré-processados com SGD, no qual foram encontrados dois erros de classificação para $\alpha = 0,01$ e um erro para $\alpha = 0,05$.

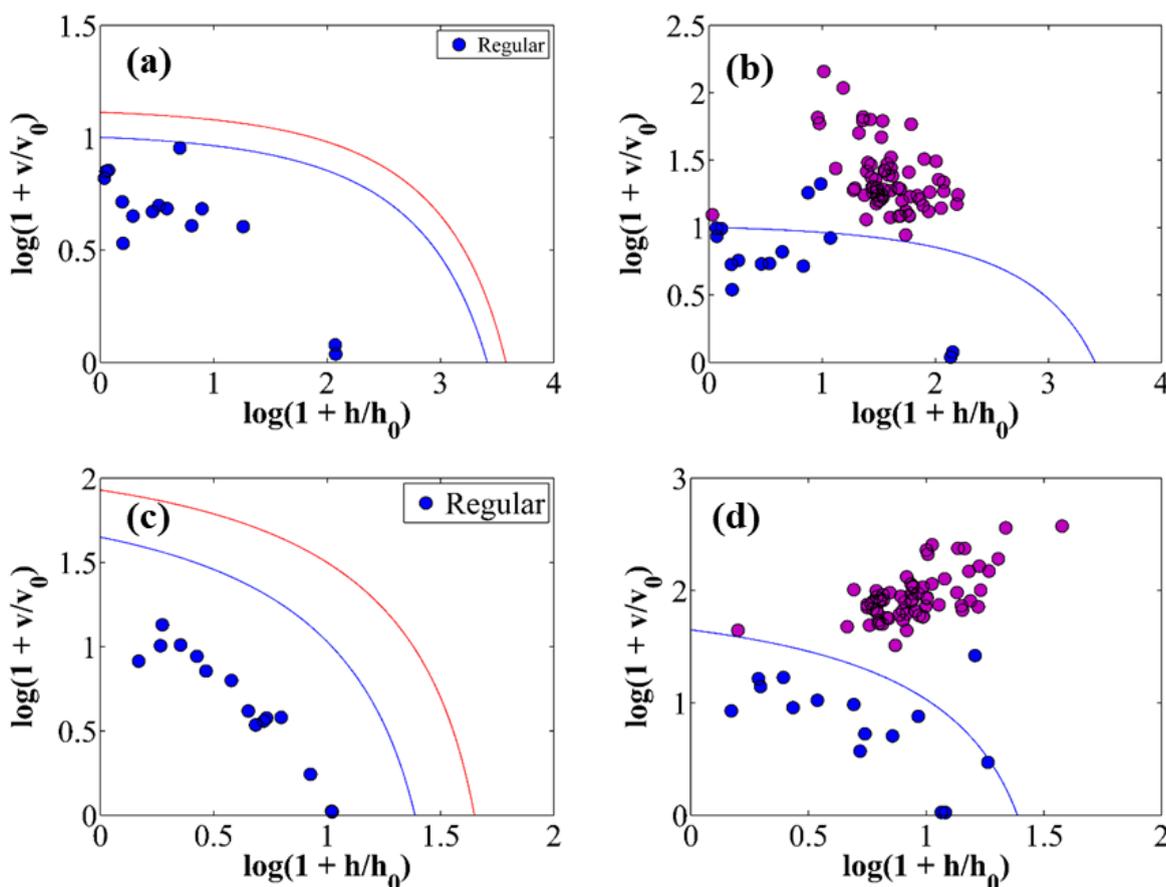


Figura A4. Gráficos de área de aceitação para o conjunto de treinamento (a, c) e teste (b, d) empregando DD-SIMCA, abordagem 2, para $\alpha = 0,01$ (a, b) e $\alpha = 0,05$ (c, d), pré-processamento SGD1_25p.

Na **Figura A4b** (empregando 3 PCs para $\alpha = 0,01$), foram identificadas duas amostras puras (70 e 72) classificadas erroneamente, sendo projetadas fora da área de aceitação. Ambas amostras são referentes a mixes do lote 01, sendo a amostra 70 é referente a rotação da amostra 2, esta que já no treinamento ficou bem próxima ao limite da área de aceitação, quase sendo considerada uma amostra extrema. enquanto a amostra 72 é referente a rotação da amostra 4, que no treinamento não apresentou problemas, porém apresentou um espectro ruidoso em bandas cruciais neste pré-processamento. Isto posto, é possível considerar estes fatores influentes na causa destes erros. Já na **Figura A3d** (empregando 7 PCs para $\alpha = 0,05$),

apenas a amostra 72 foi classificada incorretamente, assim justifica-se o acerto da amostra 70, uma vez que com o aumento do número de PCs e da significância, conseqüentemente a área de aceitação aumentou e a disposição das amostras também, de forma a abarcar a amostra 70, contudo ainda não incluindo a amostra ruidosa. Vale salientar, que com o uso de diferente abordagens e pré-processamentos, ocorre a mudança nos conjuntos de treinamento e teste, assim nem sempre as amostras que possuem a mesma numeração serão as mesmas nas diferentes abordagens, e vice-versa, amostras com numerações distintas podem ser a mesma nos diferentes casos.

3. Artigo da Dissertação

Food Control 167 (2025) 110800



Contents lists available at ScienceDirect

Food Control

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



Raman spectroscopy-based authentication of powder goat milk adulteration with cow milk

José Leonardo Alves Ferreira^a, Luciano Farias de Almeida^a, Simone da Silva Simões^b, Paulo Henrique Gonçalves Dias Diniz^{c,*}, David Douglas de Sousa Fernandes^{b,**}

^a Departamento de Química, Centro de Ciências Exatas e da Natureza, Universidade Federal da Paraíba, CEP 58051-970, João Pessoa, Paraíba, Brazil

^b Departamento de Química, Centro de Ciências e Tecnologia, Universidade Estadual da Paraíba, CEP 58429-500, Campina Grande, Paraíba, Brazil

^c Programa de Pós-Graduação em Química Pura e Aplicada, Universidade Federal do Oeste da Bahia, CEP 47810-059, Barreiras, Bahia, Brazil