

SIMONE MELO VIEIRA

**DETECÇÃO E CLASSIFICAÇÃO DE SOROS
LÁCTEOS EM LEITES UHT E *IN NATURA*
EMPREGANDO INFRAVERMELHO MÉDIO COM
TRANSFORMADA DE FOURIER E ANÁLISES
MULTIVARIADAS**

Belo Horizonte, MG.

2016

SIMONE MELO VIEIRA

**DETECÇÃO E CLASSIFICAÇÃO DE SOROS
LÁCTEOS EM LEITES UHT E *IN NATURA*
EMPREGANDO INFRAVERMELHO MÉDIO COM
TRANSFORMADA DE FOURIER E ANÁLISES
MULTIVARIADAS**

Tese apresentada ao Programa de Pós-Graduação em Ciências de Alimentos da Faculdade de Farmácia da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do grau de Doutor em Ciências de Alimentos

Área de concentração: Ciência de Alimentos

Orientador: Prof. Dr. Leandro Soares Oliveira

Belo Horizonte, MG.

2016

Vieira, Simone Melo.

V658d Detecção e classificação de soros lácteos em leites UHT e *in natura* empregando infravermelho médio com transformada de Fourier e análises multivariadas / Simone Melo Vieira. – 2016.

162 f. : il.

Orientador: Leandro Soares Oliveira.

Tese (doutorado) – Universidade Federal de Minas Gerais, Faculdade de Farmácia, Programa de Pós-Graduação em Ciência de Alimentos.

1. Leite – Análise – Teses. 2. Leite – Adulteração – Inspeção – Teses. 3. Fourier, Espectroscopia de infravermelho por transformada de – Teses. 4. Quimiometria – Teses. I. Oliveira, Leandro Soares. II. Universidade Federal de Minas Gerais. Faculdade de Farmácia. III. Título.

CDD: 637.1

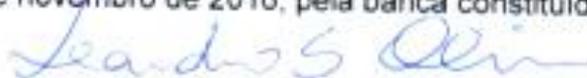
FOLHA DE APROVAÇÃO

**DETECÇÃO E CLASSIFICAÇÃO DE SOROS LÁCTEOS EM LEITE UHT E CRU
EMPREGANDO INFRAVERMELHO MÉDIO COM TRANSFORMADA DE FOURIER
E ANÁLISES MULTIVARIADAS**

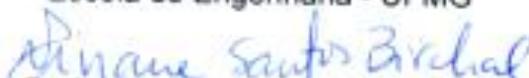
SIMONE MELO VIEIRA

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em CIÊNCIA DE ALIMENTOS, como requisito para obtenção do grau de Doutor em CIÊNCIA DE ALIMENTOS, área de concentração CIÊNCIA DE ALIMENTOS.

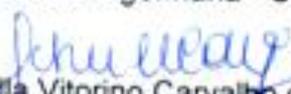
Aprovada em 28 de novembro de 2016, pela banca constituída pelos membros:



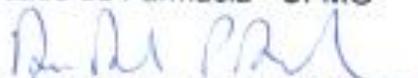
Prof. Dr. Leandro Soares de Oliveira (Orientador)
Escola de Engenharia - UFMG



Profa. Dra. Viviane Santos Birchall
Escola de Engenharia - UFMG



Profa. Dra. Scheilla Vitorino Carvalho de Souza Ferreira
Faculdade de Farmácia - UFMG



Profa. Dra. Ana Amélia Paolucci Almeida
Universidade FUMEC



Prof. Dr. Bruno Gonçalves Botelho
Centro Universitário de Belo Horizonte - UniBH

Belo Horizonte, 28 de novembro de 2016.

AGRADECIMENTOS

Agradeço a Deus pela oportunidade e conquista,

E também:

Ao meu orientador, coorientadora e equipe de pesquisa,

Aos colaboradores: Prof. Dr. Waldomiro Borges, Dra. Letícia Souza, Prof. Dr. Anselmo de Oliveira; Dra. Mariana Godinho e Sr. Wellington Lima Junior,

Às queridas filhas, filho, genros, nora e netos,

Aos meus demais familiares,

Aos amigos,

E finalmente,

Às instituições de apoio FAPEMIG e IFTM – Campus Uberlândia.

*Fé,
Âncora para os momentos
difícies e combustível
imprescindível para a
prosperidade...*

RESUMO

O leite é um alimento que está sujeito a adulterações que podem alterar suas características nutricionais, de forma a promover comprometimento da saúde e lesar economicamente o consumidor. Entre os adulterantes frequentemente encontrados no leite, grande importância é dada à detecção de soros provenientes do processamento de derivados, tais como queijos e manteigas. Esse tipo de adulteração, proibida pela legislação brasileira e internacional, corresponde a uma fraude econômica que visa aumentar o volume de leite comercializado. O monitoramento da presença do soro no leite requer a realização de análises complexas e de elevado custo que podem ser substituídas por análises alternativas capazes de fornecer resultados rápidos e precisos sem a demanda de tratamento prévio das amostras e uso de reagentes. Nesse contexto, a pesquisa utilizou técnica de espectroscopia no infravermelho médio dotada de transformada de Fourier e modelos quimiométricos para propor metodologia capaz de detectar e classificar soros de origem láctea em leites cru e UHT integral (*Ultra High Temperature*). A partir dos dados espectroscópicos foram desenvolvidos os modelos quimiométricos PLS e PLS-DA. O modelo preditivo quantitativo PLS foi capaz de detectar e quantificar diferentes teores de soro de queijo Minas Frescal em leite cru com RMSEP (erro quadrático médio, % p/p) igual a 0,180. O modelo preditivo qualitativo PLS-DA foi capaz de classificar soros de queijos e manteigas que podem ser utilizados em adulterações de leite UHT com erro médio de classificação para o modelo de previsão igual a 0,1914. A classificação dos diferentes soros em leite UHT por meio do modelo PLS-DA corresponde a uma informação que pode servir de apoio às ações de fiscalização para o rastreamento da origem da substância adulterante. Os modelos desenvolvidos apresentaram alta capacidade de reconhecimento de novas amostras e foram validados por meio de figuras de méritos e considerados aptos para serem utilizados como análises alternativas rápidas e precisas para controle de fraudes em leite cru e UHT. Os modelos comprovaram que o FTIR-MIR juntamente com modelos quimiométricos são capazes de fornecer relevantes informações norteadoras de pesquisas e metodologias para detecção de adulterações em leite com soros lácteos.

Palavras-chave: fraudes, espectroscopia, leite e quimiometria.

ABSTRACT

Milk is a food susceptible to adulterations that can compromise its nutritional characteristics, and therefore compromise the health and economically prejudice the consumers. Among the adulterants frequently found in milk, great importance is given to the detection of whey, result of the derivatives processing such as cheeses and butters, added. This type of adulteration, prohibited by Brazilian and international legislation, corresponds to an economic fraud aimed to increasing the volume of milk sold. The inspection of the presence of whey in milk requires complex and costly analyzes that can be replaced by alternative analyzes capable of delivering fast and accurate results without demanding pre-treatment of the samples and use of reagents. In this context, the research used a medium infrared spectroscopy technique with Fourier transform and chemiometric models to propose a methodology capable of detecting and classifying of milk whey origin in raw and whole UHT milks (Ultra High Temperature). From the spectroscopic data, the chemiometric models PLS and PLS-DA were developed. The PLS quantitative predictive model was able to detect and quantify different whey contents of *Minas Frescal* cheese in raw milk with RMSEP (Root Mean Square Error of Prediction, % p/p) equal to 0.180. The qualitative predictive model PLS-DA was able to classify cheeses and butters whey that can be used in adulterations of UHT milk with an average error of classification for the model of prediction equal to 0.1914. The classification of the different whey in UHT milk by means of PLS-DA model corresponds to information that can be used to support surveillance actions to trace the origin of the adulterant substance. The quantitative and qualitative predictive models were validated utilizing figures of merit and were considered suitable for the purpose as fast and accurate alternative analyzes to control frauds in raw and UHT milks. The models proved that the FTIR-MIR together with chemometric models are able to provide relevant information guiding research and methodologies for detecting milk adulterations.

Keywords: fraud, spectroscopy, milk and chemiometrics.

LISTA DE FIGURAS

FIGURA 1: Representação simplificada da espectroscopia no infravermelho.....	17
FIGURA 2: Diagrama esquemático de um espectrofotômetro com transformada de Fourier.....	21
FIGURA 3: Esquema ilustrativo para dispositivo ATR.....	24
FIGURA 4: Diagrama simplificado de um processo de análise quimiométrica.....	32
FIGURA 5: Organização dos dados em uma matriz.....	33
FIGURA 6: Esquema simplificado da calibração multivariada.....	37
FIGURA 7: Representação da matriz X em uma PCA	39
FIGURA 8: Representação geométrica de duas PCs em espaço multidimensional com três variáveis (3D)	40
FIGURA 9: Decomposição das matrizes X e Y no modelo PLS	42
FIGURA 10: Representação geométrica do PLS	44
FIGURA 11: Representação geométrica dos modelos PLS-DA e OPLS-DA	49
FIGURA 12: Gráfico da função cosseno de três amostras em um espaço tridimensional	54
FIGURA 13: Crioscópio eletrônico PZL 7000.	59
FIGURA 14: PHmetro modelo 100 da PHTEK	59
FIGURA 15: Termolactodensímetro	60
FIGURA 16: Conjunto de espectrofotômetros da Perkin	62
FIGURA 17: Acessório de ATR	63
FIGURA 18: Amostras de leite cru adulteradas com soro de queijo Minas Frescal	80
FIGURA 19: Espectros das amostras adulteradas com cada tipo de soro lácteo utilizados na construção do modelo PCA e HCA.....	87
FIGURA 20: Espectros das amostras de leite adulteradas artificialmente com soros lácteos do modelo PCA e HCA.....	87
FIGURA 21: Conjunto contendo um espectro de cada tipo de soro adulterante no modelo PCA e HCA.....	88
FIGURA 22: Absorção das amostras na região de 900 a 1500 cm^{-1}	90
FIGURA 23: Absorção das amostras na região entre 1620 a 1650 cm^{-1}	91
FIGURA 24: Absorção das amostras na região de 1720 a 1760 cm^{-1}	92
FIGURA 25: Absorção das amostras na região de 2850 a 2920 cm^{-1}	93

FIGURA 26: Absorção das amostras na região de 3160 a3500 cm^{-1}	94
FIGURA 27: PCAs definidas pelo modelo e escores das amostras	95
FIGURA 28: Variância acumulada em todos os PC (%).....	95
FIGURA 29: Variância explicada por todos os PC (%).....	96
FIGURA 30: Dendrograma dos diferentes soros lácteos.....	97
FIGURA 31: Espectros de amostras de soros de queijos e de leite (puros).....	98
FIGURA 32: Conjunto de espectros das amostras de leite adulteradas com diferentes soros de queijos	99
FIGURA 33: Bandas de absorção das amostras de leite adulteradas com soros de queijos na região de 900-1600 cm^{-1}	100
FIGURA 34: Bandas de absorção das amostras de leite adulteradas com soros de queijos nas regiões de 1700 a 1820 cm^{-1} (A) e 2800 a 3000 cm^{-1} (B)	101
FIGURA 35: Gráficos de classificação das amostras de leite adulteradas com diferentes soros de queijos	103
FIGURA 36: Espectro do leite puro utilizado nas adulterações com leiteiro	106
FIGURA 37: Espectros puros dos diferentes leiteiros utilizados nas adulterações.....	107
FIGURA 38: Espectros de cada tipo de leiteiro adulterante do modelo preliminar PLS-DA	108
FIGURA 39: Absorção na região de 2820 a3000 cm^{-1}	109
FIGURA 40: Escores das duas primeiras CPs para modelo PCA.....	110
FIGURA 41: Modelo PLS-DA para classificação das amostras de leite UHT adulteradas com diferentes leiteiros.....	111
FIGURA 42: Escores das duas primeiras PCs para modelo com leite puro	113
FIGURA 43: Classificação das amostras adulteradas com leiteiro.....	114
FIGURA 44: Espectro do leite cru utilizado nas adulterações com soro Minas Frescal	118
FIGURA 45: Conjunto espectros de leite cru adulterado com diferentes concentrações de soro Minas Frescal	119
FIGURA 46: Região de absorção de 950 a 1450 cm^{-1}	120
FIGURA 47: Região de absorção de 1600 a 1650 cm^{-1}	121
FIGURA 48: Região de absorção 1720-1750 cm^{-1} ;e de 2840 a 2940 cm^{-1}	122
FIGURA 49: Região de absorção de 3150-3500 cm^{-1}	123
FIGURA 50: Variância explicada pelo modelo definida pelos valores de RMSEC E RMSECV	125

FIGURA 51: Variância explicada pelo modelo em função do número de variáveis latentes	125
FIGURA 52: Valores de <i>leverage</i> e Q residual das amostras do grupo de calibração com intervalo de confiança de 95%.....	126
FIGURA 53: Relação entre o erro médio de Previsão e o número de Variáveis Latentes	127
FIGURA 54: Correlação linear do modelo PLS para amostras de calibração e de previsão	128
FIGURA 55: Conjunto dos espectros das amostras do modelo PLS-DA contendo adulterações de soros de queijos e leiteiros.	132
FIGURA 56: Região de absorção referente à faixa de 950 a 1580 cm^{-1}	132
FIGURA 57: Região de absorção referente à faixa de 1700 a 1780 cm^{-1}	134
FIGURA 58: Região de absorção referente à faixa de 2850 a 3000 cm^{-1}	135
FIGURA 59: Região de absorção referente à faixa de 3200 e 3400 cm^{-1}	136
FIGURA 60: Gráficos das classes reconhecidas pelo modelo PLS-DA para adulterações de leite com soros de queijos e leiteiros.....	139
FIGURA 61: Modelo de gráfico VIP-scores para cada classe do modelo PLS-DA(B).142	

LISTA DE TABELAS

TABELA 1: Projeção de produção, consumo e exportação de leite do Brasil no período de 2015 a 2025. Leite em milhões de litros (Projeção).....	7
TABELA 2 Variação na composição do leite bovino, do soro "doce" e do soro "ácido"	11
TABELA 3: Composição média do leiteiro e do leite desnatado	12
TABELA 4: Amostras de leite UHT e pasteurizado analisadas pelo PCFL no período de 2009 a 2012.....	13
TABELA 5: Divisões da região do infravermelho	19
TABELA 6: Faixas de absorção de diferentes grupos funcionais	20
TABELA 7: Regiões de absorção do leite de vaca no infravermelho médio.....	27
TABELA 8: Possíveis resultados para classificação de amostras	55
TABELA 9: Figuras de mérito para modelos qualitativos	56
TABELA 10: Concentrações de soro de queijo em leite UHT(amostras A, B e C)	76
TABELA 11: Concentrações de leiteiros em leite UHT (amostras A, B e C).....	78
TABELA 12: Crioscopia (H°) das diferentes concentrações de adulterações de leite UHT integral	83
TABELA 13: Densidade e pH amostras A,B e C de UHT integral	85
TABELA 14: Composição dos principais minerais do soro doce e do soro ácido.....	89
TABELA 15: Sensibilidade e seletividade do modelo PLS-DA para soros de queijos .	104
TABELA 16: Sensibilidade e seletividade do modelo PLS-DA para leiteiros.....	112
TABELA 17: Sensibilidade e seletividade para modelo PLS-DA (leiteiros e leite puro)	116
TABELA 18: Valores das Figuras de Mérito (FOM) para modelo PLS	130
TABELA 19: Grupos funcionais que maior influencia exerceram sob o modelo para definição das classes de acordo com os gráficos VIP-escores	143
TABELA 20 - Parâmetros de mérito para validação do modelo PLS-DA contendo amostras do conjunto de treinamento	144
TABELA 21: Parâmetros de mérito para validação do modelo PLS-DA contendo amostras do conjunto teste (predição)	145

LISTA DE ABREVIATURAS

ABNT	Associação Brasileira de Normas Técnicas
ANVISA	Agência Nacional de Vigilância Sanitária
ASTM	Sociedade Americana para Testes de Materiais (<i>American Society for Testing and Materials</i>)
CART	Árvore de Classificação e Regressão (<i>Classification And Regression Tree</i>)
CLAE	Cromatografia Líquida de Alta Eficiência
CMP	Caseinomacropéptido
DA	Análise Discriminante (<i>Discriminant Analysis</i>)
DICLA	Divisão de Acreditação de Laboratórios
DIPOA	Departamento de Inspeção de Produtos de Origem Animal
DPR	Desvio padrão relativo
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
ETE	Estação de Tratamento de Efluentes
EU	União Europeia
EFR	Taxa de eficiência (<i>Efficiency Rate</i>)
FA	Análise de Fatores (<i>Factor. Analysis</i>)
FN	Falsos Negativos
FNR	Taxa de Falsos Negativos (<i>False Negative Rate</i>)
FP	Falsos Positivos
FPR	Taxa de Falsos Positivos (<i>False Positive Rate</i>)
FOM	Figuras de Mérito (<i>Figures of Merit</i>)
FTIR	Infravermelho com Transformada de Fourier
FTIR-MIR	Infravermelho Médio com Transformada de Fourier
GMP	Glicomacropéptido
HCA	Análise Hierárquica de Agrupamentos (<i>Hierarchical Cluster Analysis</i>)
HPLC	Cromatografia Líquida de Alto Desempenho (<i>High performance liquid chromatography</i>)
IgE	Imunoglobulina E
INMETRO	Instituto Nacional de Metrologia, Qualidade e Tecnologia
IUPAC	União Internacional de Química Pura e Aplicada

	<i>(International Union of Pure and Applied Chemistry)</i>
KNN	K-ésimo Vizinho mais Próximo (<i>Kth Nearest Neighbour</i>)
LDA	Análise Linear Discriminante (<i>Linear Discriminant Analysis</i>)
LAMAM	Laboratório de Ensaios em Alimentos
MAPA	Ministério da Agricultura, Pecuária e Abastecimento
MPF	Ministério Público Federal
MATLAB	<i>Matrix Laboratory</i>
MCR	Resolução Multivariada de Curvas (<i>Multivariate Curve Resolution</i>)
MCR-ALS	Resolução Multivariada de Curvas com Mínimos Quadrados Alternados (<i>Multivariate Curve Resolution-Alternating Least Squares</i>)
MIR	Infravermelho Médio (<i>Mid Infrared</i>)
MLR	Regressão linear múltipla (<i>Multiple Linear Regression</i>)
MSC	Correção de Espalhamento de Luz (<i>Multiple Scatter Correction</i>)
MSE	Erro Quadrático Médio (<i>Mean Square Error</i>)
NCFPD	Centro Nacional de Produção de Alimentos e Defesa
NAS	Sinal Analítico Líquido (<i>Net Analyte Signal</i>)
NILPALS	Algoritmo iterativo não-linear de mínimos quadrados parciais (<i>Non-linear Iterative Partial Least Squares</i>)
NIR	Infravermelho Próximo (<i>Near Infrared</i>)
OPLS-DA	Análise Discriminante Ortogonal por Mínimos Quadrados Parciais (<i>Orthogonal Partial Least Squares Analysis Discriminant</i>)
PAMVet	Programa de Análise de Resíduo de Medicamentos Veterinários em Produtos de Origem Animal
PC	Componente Principal (<i>Principal Component</i>)
pH	Potencial hidrogeniônico
PCR	Principal Componente de Regressão (<i>Principal Component Regression</i>)
PCFL	Programa Nacional de Combate a Fraude de Leite
PLS	Regressão por Mínimos Quadrados Parciais (<i>Partial Least</i>

	<i>Squares)</i>
PQI	Padrão de Identidade e Qualidade
PRESS	Soma dos quadrados dos desvios dos erros previstos (<i>Predective Residual Error Sum of Squares</i>)
QDA	Análise Quadrática Discriminante (<i>Quadratic Discriminant Analysis</i>)
RIISPOA	Regulamento de Inspeção Industrial e Sanitária de Produtos de Origem Animal
RMSE	Raiz Quadrada do Erro Quadrático Médio (<i>Root Mean Squares Error</i>)
RMSEC	Raiz Quadrada do Erro Quadrático Médio de Calibração (<i>Root Mean Squares Errors of Calibration</i>)
RMSECV	Raiz Quadrada do Erro Quadrático Médio de validação Cruzada (<i>Root Mean Squares Errors of Cross Validation</i>)
RMSEP	Raiz Quadrada do Erro Quadrático Médio de Previsão (<i>Root Mean Squares Errors of Prediction</i>)
RP	Correlação de Previsão (<i>Correlation Prediction</i>)
RVC	Coefficiente de Validação Cruzada (<i>Cross Validation coefficient</i>)
SENAI	Serviço Nacional de Aprendizagem Industrial
SIMCA	Modelagem Independente e Flexível por Analogia de Classes (<i>Soft Independent Modelling of Class Analogies</i>)
SPA-LDA	Algoritmo das Progressões Sucessivas- Análise Discriminante Linear (<i>Successive Projections Algorithm-Linear Discriminant Analysis</i>)
UFG	Universidade Federal de Goiás
UFU	Universidade Federal de Uberlândia
UHT	Estrema Alta Temperatura (<i>Ultra Righ Temperature</i>)
USA	Estados Unidos da America (<i>United States of America</i>)
VIP scores	Importância das Variáveis na Projeção dos escores (<i>Variable Importance in Projection</i>)

SUMÁRIO

RESUMO	vii
LISTA DE FIGURAS.....	ix
LISTA DE TABELAS	xii
LISTA DE ABREVIATURAS	xiii
SUMÁRIO	xvi
1 INTRODUÇÃO.....	1
1.1 Justificativa.....	1
1.2 Objetivos	5
1.2.1 Objetivo Geral	5
1.2.2 Objetivos específicos	5
2 REVISÃO DE LITERATURA	6
2.1 Características de leites fluidos.....	6
2.2 Adulteração em leites fluidos por adição de soro.....	8
2.3 Técnicas analíticas para detecção de adulterações com soro	14
2.4 Princípios da espectroscopia no infravermelho.....	17
2.4.1 Equipamentos de infravermelho	21
2.4.2 Aplicação da espectroscopia do infravermelho médio em análises de alimentos e de leite	24
2.5 Análises Multivariadas.....	29
2.5.1. Quimiometria.....	30
2.5.2. Métodos de classificação não supervisionados	37
2.5.3. Métodos de classificação supervisionados	45
2.5.4. Validação de modelos quimiométricos.....	50
3 METODOLOGIA.....	57
3.1 Testes preliminares.....	57
3.1.1 Análise de Crioscopia	57
3.2 Análise espectroscópica	61

3.2.1 FTIR-MIR	61
3.2.2 Acessório de ATR	63
3.2.3 Obtenção dos espectros	63
3.2.4 Interpretação dos espectros.....	64
3.3 Modelagem.....	65
3.3.1 Pré-tratamentos dos espectros para construção dos modelos	66
3.3.2 Obtenção dos modelos	67
3.3.3 Modelo PCA.....	68
3.3.4 Modelo HCA.....	69
3.3.5 Modelo PLS	69
3.3.6 Modelo PLS-DA	70
3.3.7 Detecção de amostras anômalas.....	71
3.3.8 Interpretação dos resultados.....	71
3.4 Modelos Preliminares	74
3.5 Modelo preditivo PLS para detecção de adulteração de leite cru com soro em leite fluido	79
3.6 Modelo preditivo PLS-DA para classificação de adulterações com soro de queijos e leitelhos	81
4 RESULTADOS E DISCUSSÃO DOS TESTES E MODELOS PRELIMINARES.....	83
4.1 Teste de crioscopia.....	83
4.2 Teste da qualidade do leite.....	84
4.2.1 Resultados de densidade e acidez	84
4.2.1 Resultados referentes ao perfil espectral.....	85
4.3 Modelo para avaliar a existência de diferenças entre soros de origem láctea: modelo exploratório PCA e HCA.....	86
4.3.1 Resultados referentes ao perfil espectral.....	86
4.3.2 Resultados referentes aos modelos PCA e HCA	94
4.4 Modelo para avaliar a possibilidade de classificar soros de queijos em adulterações de leite UHT por meio do FTIR/MIR e modelo PLS-DA.....	97
4.4.1 Resultados referentes ao perfil espectral.....	97
4.4.2 Resultados referentes ao modelo contendo amostras de soros de queijos	102

4.5 Modelo para avaliar a possibilidade de classificar soros de manteiga (leitelho) em adulterações de leite UHT por meio do modelo PLS-DA.....	106
4.5.1 Resultados referentes ao perfil espectral.....	106
4.5.2 Resultados referentes ao modelo contendo amostras de leiteiro.....	109
4.5.3 Resultados referentes ao modelo contendo amostras de leiteiro e amostra de leite puro.....	113
5 RESULTADOS E DISCUSSÃO REFERENTES À TÉCNICA FTIR/MIR E MODELO PREDITIVO PLS PARA DETECÇÃO DE ADULTERAÇÃO DE LEITE CRU COM SORO DE QUEIJO MINAS FRESCAL.....	118
5.1 Resultados referentes ao perfil espectral.....	118
5.2 Resultados referentes ao modelo PLS.....	123
6 RESULTADOS E DISCUSSÃO REFERENTES À TÉCNICA FTIR/MIR E MODELO CLASSIFICATÓRIO PLS-DA PARA DETECÇÃO DE ADULTERAÇÃO DE LEITE UHT COM SOROS DE MANTEIGA E DE QUEIJOS.....	131
6.1 Resultados referentes ao perfil espectral.....	131
6.2 Resultados referentes ao modelo PLS-DA.....	136
7 CONCLUSÕES.....	147
7.1 Conclusões gerais.....	147
7.2 Conclusões finais.....	148
REFERÊNCIAS BIBLIOGRÁFICAS.....	149

1 INTRODUÇÃO

1.1 Justificativa

De acordo com o Departamento de Agricultura dos Estados Unidos (*United States Department of Agriculture - USDA*), o Brasil ocupou a quinta posição no *ranking* mundial de produção de leite em 2014, atrás da União Europeia, Índia, Estados Unidos e China com uma produção de 35,2 bilhões de litros. A produção brasileira vem aumentando significativamente, não apenas para o suprimento do mercado interno, mas também para exportação (POLL *et al.*, 2013). Tanto a expansão da produção interna como a exportação de lácteos requer grande preocupação com a qualidade, sobretudo, no que diz respeito à inocuidade e aos direitos dos consumidores.

O leite é um alimento que possui alto valor nutricional, componente básico da dieta dos seres humanos, principalmente de crianças e idosos (MARTIN, 2011). Além de ser altamente perecível o leite pode apresentar substâncias que não fazem parte da sua composição natural decorrentes de adulterações. Este fato, pode comprometer a qualidade do leite e seu caráter nutritivo, promovendo o comprometimento da saúde humana e danos econômicos ao consumidor. As substâncias adulterantes são proibidas pela legislação brasileira e internacional (BRASIL, 2011; BRASIL, 1990; BRASIL, 1952; CODEX ALIMENTARIUS, 2008; CODEX ALIMENTARIUS, 2004; JOHNSON, 2014; UE, 2002).

Dados do Programa Nacional de Combate à Fraude do Leite, desenvolvido pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA) para controle da qualidade do leite demonstraram em 2013 que 8,3% do total de amostras de leite pasteurizado, UHT e em pó analisadas se encontraram fora dos padrões regulamentares (MAPA, 2013). No Brasil, a Instrução Normativa nº 62 do MAPA, prevê a obrigatoriedade da realização, pelas indústrias e pelos órgãos fiscalizadores, de análises que atestem a qualidade do leite (BRASIL, 2011).

As adulterações em leite podem ocorrer de maneira não proposital ou intencionalmente, configurando-se como fraudes. Adulterações, realizadas com o objetivo de aumentar o ganho econômico, são denominadas fraudes econômicas e

objetivam aumentar o volume de leite com a adição de água ou de soros de origem láctea (ALMEIDA *et al.*, 2012; CARVALHO *et al.*, 2007; MENDES *et al.*, 2010). Esse tipo de conduta lesa o consumidor que paga por um produto com certas características e, no entanto, adquire outro de menor qualidade nutricional.

Fraudes econômicas resultantes da adição de água no leite são monitoradas nas plataformas de recepção em laticínios por meio da análise de crioscopia. Entretanto, essa análise rápida e simples, não é eficiente quando são adicionadas outras substâncias no leite juntamente com a água. As fraudes para aumento do volume praticadas com a adição de soros de origem láctea oriundos do processo de fabricação de queijos também não são detectadas por meio da crioscopia e requerem, para essa finalidade, a utilização de análises mais complexas.

A técnica oficial para constatação da presença de soro de queijo em leite tem como base a identificação e quantificação do glicomacropéptido (GMP). Esse fragmento proteico da caseína é resultante da ação enzimática do coalho utilizado na produção de queijos e é liberado no soro. A detecção do GMP é feita por meio de Cromatografia Líquida de Alto Desempenho (CLAE ou HPLC, do inglês *High performance liquid chromatography*), (BRASIL, 2006^a; BRASIL, 2006^b). Essa técnica apresenta alta sensibilidade e seletividade, porém, requer a utilização de reagentes de alto custo, analistas treinados e equipamento sofisticado. Entretanto, apesar da precisão analítica, pesquisadores afirmam que a cromatografia pode proporcionar resultados falso-positivos quando estiverem presentes no leite analisado glicomacropéptidos (GMP) decorrentes da ação proteolítica de bactérias (CARVALHO *et al.*, 2007; KENNEDY *et al.*; 1998; MAGALHÃES, 2008; OLIVEIRA, 2010).

Outro aspecto relevante em relação à detecção do soro em leite fluido por meio da identificação do GMP corresponde à limitação da técnica quando a fraude é realizada com soros lácteos que não possuem esse fragmento peptídico. Soros provenientes de processos não enzimáticos, como por exemplo, os soros de manteigas (leitelhos), também podem ser utilizados em fraudes econômicas, e nesses casos, novas técnicas analíticas devem ser empregadas (FURTADO, 2010; RIGUEIRA, 2006).

As fraudes em leites fluidos podem, portanto, ser realizadas com diferentes tipos de soros lácteos, cujas características promovem alterações mais ou menos intensas no leite. As especificidades desses soros de origem láctea necessitam ser melhor estudadas e definidas para o desenvolvimento de técnicas eficientes para detecção.

Como alternativa para substituição de procedimentos analíticos complexos e de alto custo, a espectroscopia no infravermelho vem sendo aplicada a alimentos, inclusive ao leite, por propiciar a obtenção de resultados qualitativos e quantitativos relevantes para o controle de qualidade de alimentos. Segundo França e Oliveira (2011), a espectroscopia no infravermelho corresponde a uma análise rápida, precisa, não destrutiva e que não requer a utilização de reagentes para preparo das amostras, sendo capaz de proporcionar uma identificação dos diferentes tipos de substâncias presentes no alimento.

A técnica espectroscópica baseia-se na geração de picos de absorção correspondentes às frequências de vibrações das ligações intramoleculares presentes na amostra. Quando se incide a radiação infravermelha sobre uma molécula, a absorção da radiação promove vibrações de estiramento das ligações existentes entre os átomos que compõem a molécula, ocasionando faixas de absorção específicas. Cada substância apresenta picos de absorção relacionados com os grupos químicos funcionais presentes na sua composição. Por meio da análise das características dos picos de absorção que compõem um espectro, pode-se identificar uma amostra, constatar a presença de adulterantes ou contaminantes e, constatar alterações promovidas pelo processamento (FRANÇA E OLIVEIRA, 2011).

Espectrofotômetros modernos dotados de transformada de Fourier e acessórios específicos de reflexão permitem a obtenção de espectros de diferentes tipos de amostras de forma direta. Nessas análises, pequena quantidade da amostra é introduzida diretamente em um dispositivo de reflectância atenuada composto por um cristal de seleneto de zinco que corresponde à célula mais utilizada nos equipamentos de infravermelho para amostras líquidas (AMENDOLA, 2007; HOLLER *et al.*, 2009; LYNCH *et al.*, 2006; MORGANO, 2005; PINTO, 2010; RAMOS, 2006; RODRIGUEZ-SAONA E ALLENDORF, 2011).

O resultado espectral gerado pela espectroscopia no infravermelho apresenta um grande conjunto de dados espectrais que não evidenciam prontamente as informações desejadas, sendo necessária a utilização de técnicas quimiométricas multivariadas que auxiliam a quantificação e a caracterização dos constituintes que estão envolvidos nas interações complexas dos componentes das amostras (PINTO, 2010). Métodos quimiométricos, como Análise de Componente Principal (PCA), Análise Hierárquica de Agrupamentos (HCA), Regressão por Mínimos Quadrados Parciais (PLS) e Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA), são

utilizados para calibração, validação e comprovação de análises de alimentos (DANTAS E LIMA, 2010; FRANÇA E OLIVEIRA, 2011; PATACA, 2006; RECHE E FRANCO, 2009; SOUZA *et al.*, 2011;).

Os métodos PCA e HCA correspondem a modelos exploratórios capazes de caracterizar amostras sem que existam informações prévias sobre elas (MORGADO, 2005). Já os modelos PLS e PLS-DA são modelos preditivos que utilizam o mesmo algoritmo e são aplicados em determinação de concentrações e classificação de amostras que requerem informações iniciais sobre as amostras analisadas. O PLS possui excelente poder de previsão e quantificação de concentrações. Já o PLS-DA é utilizado para classificação de amostras por meio da redução do número de variáveis. Esses dois modelos quimiométricos processam grande quantidade de dados e seus cruzamentos de forma precisa e rápida. Em função dessas características, podem ser amplamente aplicados em análises quimiométricas espectroscópicas (NURRULHIDAYAH *et al.*, 2013; BERRUETA, 2007; FRANÇA E OLIVEIRA, 2011).

O aprimoramento da espectroscopia no infravermelho assim como dos métodos quimiométricos têm propiciado o desenvolvimento de novas metodologias para análises de alimentos, sobretudo, relacionadas ao controle de qualidade e garantia da inocuidade. Métodos preditivos como o PLS e o PLS-DA são também denominados de supervisionados e permitem prever novas amostras, ou seja, são capazes de configurar modelos para quantificar e classificar amostras desconhecidas, com base na existência de dados referenciais gerados por amostras previamente tratadas quimiometricamente (PARISOTTO *et al.*, 2007).

A grande quantidade de dados gerados de forma rápida por meio da análise espectroscópica atrelados à especificidade e precisão dos métodos quimiométricos têm propiciado o desenvolvimento de modelos robustos e específicos para detecção de diferentes tipos de adulterações em alimentos, inclusive, em leites (CARVALHO, 2013; LYNCH *et al.*, 2006; RODRIGUEZ-SAONA E ALLENDORF, 2011). Nesse contexto, visando ampliar a possibilidade de detecção de adulterações de leite com soros de origem láctea, por meio de análise preditiva, propôs-se a utilização dos modelos PLS e PLS-DA.

A detecção da adulteração por meio dos modelos PLS e PLS-DA pode ser realizada, independentemente da confirmação da presença do GMP, marcador utilizado como referência pelas análises oficiais para detecção de soro no leite

(CARVALHO, 2007). Nos modelos PLS e PLS-DA a capacidade preditiva é baseada nos dados gerados pelo conjunto de calibração (BOTELHO *et al.*, 2013).

Os modelos propostos no presente trabalho foram validados por figuras de mérito e podem vir a ser configurados como métodos analíticos multivariados alternativos que dispensam o preparo prévio das amostras e a utilização de reagentes de alto custo e que geram resíduos no meio ambiente. Correspondem a potenciais e eficientes ferramentas analíticas que podem vir a ser utilizadas em procedimentos de rotina para detecção de fraudes com soros lácteos em leites cru e UHT.

1.2 Objetivos

1.2.1 Objetivo Geral

Este trabalho objetiva avaliar a aplicabilidade da técnica espectroscópica no infravermelho médio com transformada de Fourier (FTIR-MIR) juntamente com técnicas de análises multivariadas para classificação (quanto ao tipo) e detecção (presença ou ausência) de diferentes tipos de soros lácteos que podem ser utilizados em adulterações de leites UHT integral e cru.

1.2.2 Objetivos específicos

- Avaliar a eficiência do teste de crioscopia para detecção em leite fluido de fraudes econômicas com soros de origem láctea.
- Realizar estudos do perfil espectral das amostras de leite UHT integral adulteradas com diferentes soros lácteos.
- Avaliar a existência de diferenças significativas entre os soros lácteos empregados nas adulterações do leite UHT integral por meio do FTIR-MIR e modelos exploratórios PCA e HCA.
- Desenvolver metodologia de análise para quantificação do soro de queijo Minas Frescal em leite cru adulterado, por meio do FTIR-MIR e modelo PLS.
- Desenvolver metodologia de análise para detecção e classificação de diferentes soros lácteos utilizados em adulterações de leite UHT integral, por meio do FTIR-MIR e modelo PLS-DA.

2 REVISÃO DE LITERATURA

2.1 Características de leites fluidos

O leite corresponde a uma mistura homogênea de diferentes substâncias (lactose, glicérides, proteínas, sais minerais, vitaminas e enzimas) que se encontram dispersas em meio aquoso, sob forma de emulsão, suspensão ou dissolvidas. Os elementos sólidos representam aproximadamente 12 a 13% do leite, enquanto a água aproximadamente 87%. Entre os elementos sólidos, as micelas de caseína (proteína) e os glóbulos de gordura são responsáveis pela maior parte das características físicas (estrutura e cor) encontradas nos produtos lácteos. O leite é um alimento que possui alto valor nutricional, constituindo componente básico da dieta dos seres humanos, principalmente de crianças e idosos (FERREIRA, 2007; MARTIN, 2011).

Variações na composição do leite podem ocorrer em função de diferentes fatores: estágio de lactação, raça, alimentação, temperatura ambiente, intervalo entre as ordenhas, manejo de ordenha e estado de saúde do animal (ORDONEZ, 2005). Entretanto, essas variações não podem ultrapassar os limites mínimos estabelecidos por lei para leite integral, correspondentes a: 3% para teor de gordura; 2,9% para teor de proteínas; 8,4% para teores de sólidos não gordurosos (BRASIL, 2011). A proporção adequada da composição físico-química e microbiológica é que define a qualidade do leite, refletindo as características sensoriais que levam à sua aceitação. A distribuição e as interações das substâncias componentes do leite são determinantes para a sua estrutura, propriedades funcionais e para o processamento (OLIVEIRA, 2011).

Os leites fluidos disponíveis no mercado para consumo direto podem ter maior ou menor vida útil de acordo com o tratamento térmico ao qual é submetido. Segundo Ordonez (2005), os leites fluidos comercializados para consumo direto que possuem menor vida útil são aqueles submetidos ao tratamento térmico de pasteurização. Os leites pasteurizados normalmente comercializados em embalagens plásticas são classificados como integral, semi desnatado ou desnatado conforme o percentual de gordura presente. Os de vida útil mais longa são aqueles submetidos a tratamentos térmicos mais drásticos, destacando-se, comercialmente, o leite UHT (*Ultra High*

Temperature), que também recebe a mesma classificação em relação ao teor de gordura.

O leite cru ou *in natura* corresponde a leite fluido, tem sua comercialização proibida no território brasileiro para consumo direto, sendo destinado para obtenção de leite pasteurizado ou UHT e para transformação em derivados lácteos. Esse leite, oriundo das propriedades rurais, antes de ser destinado para elaboração de derivados lácteos, é obrigatoriamente submetido a análises rotineiras para avaliação dos parâmetros físico-químicos que levam à detecção de adulterações (BRASIL, 2011).

O Brasil vem aumentando significativamente a produção de leite bovino não apenas para o suprimento do mercado interno, mas também para exportação. Situa-se como um dos maiores produtores de leite do mundo, tendo produzido, em 2014, cerca de 35 bilhões de litros (EMBRAPA, 2015). O potencial brasileiro para exportação de lácteos requer grande preocupação com a qualidade, sobretudo, no que diz respeito à inocuidade e aos direitos dos consumidores. Na Tabela 1, apresenta-se dados de produção, consumo e exportação de leite, em 2015, e projeções para 2025, demonstrando o grande potencial da atividade leiteira brasileira.

TABELA 1: Projeção de produção, consumo e exportação de leite do Brasil no período de 2015 a 2025. Leite em milhões de litros (Projeção)

Ano	Produção	Consumo	Exportação
2017	39363	39650	556
2018	40827	40653	580
2019	41416	41656	604
2020	42859	42659	628
2021	43439	43663	652
2022	44879	44666	676
2023	45457	45669	700
2024	46896	46672	724
2025	47474	47675	748

Fonte: BRASIL, 2015.

A demanda nacional e internacional de leite e derivados está atrelada à qualidade do leite e sua acessibilidade econômica. Apesar do custo de produção do leite no Brasil ser um dos menores do mundo, a sanidade dos rebanhos e a qualidade

intrínseca do leite requer o monitoramento por meio de análises que possam diagnosticar problemas que comprometem o consumo e a produção de derivados (CARVALHO *et al.*, 2005).

2.2 Adulteração em leites fluidos por adição de soro

Tanto o leite cru como o industrializado podem apresentar em sua composição substâncias estranhas que comprometem sua qualidade, seu caráter nutritivo e que podem promover o comprometimento da saúde humana. Essas substâncias que, não fazem parte da composição natural do leite, são denominadas de adulterantes e são proibidas pela legislação brasileira (BRASIL, 2011). As alterações promovidas no leite ocasionam o não enquadramento nos Padrões de Identidade e Qualidade (PQI) e podem ocorrer de forma não proposital ou intencionalmente (FURTADO, 2010; BRANDÃO, 2008).

As adulterações de leites fluidos configuram-se como fraudes e são normalmente realizadas com o objetivo de aumentar o ganho econômico ou de prolongar a vida útil em função de sua grande perecibilidade. De acordo com pesquisadores do Centro Nacional de Proteção de Alimentos e Defesa (NCFPD) da Universidade de Minesota, USA, a adulteração motivada economicamente (EMA - *Economically Motivated Adulteration*) corresponde à venda de um alimento de baixo padrão com a finalidade de ganho econômico. Os tipos mais comuns de EMA incluem a substituição intencional de um ingrediente autêntico, por um mais barato; a diluição com água ou com outras substâncias líquidas; o aprimoramento do sabor utilizando substâncias ilícitas ou não aprovadas; e a substituição de componente de uma espécie por outro de outra espécie (JOHNSON, 2014). Segundo Furtado (2010), as principais adulterações que ocorrem no leite estão relacionadas à utilização das seguintes substâncias: água, soro de queijo, leiteiro, reconstituintes, alcalinos, conservantes, álcool, misturas pré-balanceadas e leites de espécies distintas.

As fraudes econômicas realizadas em leite, normalmente, não comprometem a saúde de seres humanos saudáveis, mas comprometem o valor nutricional do leite, que é reduzido em função da adição de água ou soros de origem láctea. O Código de Defesa do Consumidor menciona que se o consumidor adquirir um alimento

apresentando vícios de qualidade que os torne impróprio ao consumo ou que lhe diminua o valor, tem o direito de solicitar a substituição do produto por outro da mesma espécie (isento de adulteração) ou a restituição imediata da quantia paga. Entretanto, nem sempre a constatação da fraude é percebida pelo consumidor, cabendo aos órgãos fiscalizadores, o monitoramento da qualidade do leite comercializado por meio de análises laboratoriais específicas e até mesmo complexas (BRANDÃO 2008; JOHNSON, 2014; MOORE *et al.*, 2012; RODRIGUES *et al.*, 2012).

A utilização de soros lácteos em fraudes econômicas em substituição à adição de água ocorre em função da grande disponibilidade desses subprodutos gerados durante os processos de elaboração de derivados, sem ônus e em grande quantidade (CARVALHO *et al.*, 2007). Os soros lácteos provenientes do processamento de manteigas e queijos possuem em sua composição grande quantidade de água e componentes característicos do leite que dificultam a detecção da fraude. Soros de origem láctea, a princípio, não são tóxicos para seres humanos saudáveis. Não proporcionam malefícios à saúde daqueles cujos organismos funcionem normalmente, mas, podem ocasionar problemas para grupos específicos que apresentam intolerância à lactose, alergia a determinadas proteínas e aditivos que podem estar presentes nos soros (CUCU *et al.*, 2012; SOUZA *et al.*, 2011; TEIXEIRA E FONSECA, 2008).

A legislação brasileira e também as de outros países proíbem a adição de soro no leite para consumo e requerem um controle de qualidade efetivo para conter essa prática. A utilização de soro é permitida para elaboração e comercialização de bebidas lácteas fermentadas ou não fermentadas constituídas basicamente de leite e soros lácteos, desde que, esses produtos obedeçam aos PQIs definidos pela legislação. O soro também pode ser utilizado legalmente na composição de outros alimentos, sobretudo, na forma desidratada e sua presença deve ser informada no rótulo dos produtos (LERAYER *et al.*, 2002; BRASIL, 2006^a).

Os soros lácteos derivados de processos de coagulação do leite são denominados de soro de leite e possuem diferentes características que se encontram no projeto da instrução normativa para definição dos padrões de qualidade dos soros (BRASIL, 2013). De acordo com a normativa proposta, o soro de leite é definido como líquido obtido a partir da coagulação do leite destinado a fabricação de queijos, caseína ou produtos lácteos similares.

O soro de leite proveniente do processamento de queijos é também denominado de lactossoro. Contém a metade do extrato seco do leite e é composto,

primordialmente, por água. Nele estão presentes: lactose, minerais, vitaminas hidrossolúveis e 20% das proteínas do leite. É classificado como soro doce (pH 6-7) ou soro ácido (pH<5)(SISO, 1996). O soro doce é proveniente de queijos obtidos por meio de coagulação enzimática (pH 6,4) e o soro ácido é resultante do processo de elaboração de queijos onde ocorre a coagulação ácida do leite (pH próximo a 4,5) (ORDONEZ, 2005). Do volume de leite destinado à fabricação de queijos, 75 a 85% resultam em soro (ZADOW, 1997).

Existem importantes diferenças entre a composição dos sólidos não gordurosos do soro de queijo e a dos sólidos não gordurosos do leite, sobretudo, no que se refere ao tipo de proteína. O soro de leite bovino contém aproximadamente 0,8% de proteínas de baixo peso molecular, sendo, as predominantes, a beta-lactoglobulina (50-57%), a alfa-lactoalbumina (25%), a soro albumina bovina (10%), imunoglobulinas, lactoferrinas e glicomacropéptídeos (8%) (AIMUTIS, 2004)

Apesar das características benéficas das proteínas do soro, a beta-lactoglobulina, a alfa-lactoalbumina e fragmentos de caseína podem ocasionar alergia em crianças. O processo alérgico é causado pelas imunoglobulinas E (IgE), em resposta à presença destas proteínas consideradas antigênicas pelo sistema imunológico das crianças. A IgE induz a liberação de substâncias vasoativas por alguns tipos de células que causam problemas relacionados ao sistema respiratório, gastrointestinais e cardiovasculares (CUCU *et al.*, 2012; DERMONTT *et al.* 2016; LAM *et al.*, 2008).

O componente sólido presente no soro de queijo em maior concentração é a lactose, cujo percentual está em torno de 4,9%. Durante a elaboração de queijos, quase toda a lactose do leite é carregada para o soro e menos de 1% permanece na massa (ORDONEZ, 2005). A utilização do soro em fraudes de leite eleva consideravelmente a concentração de lactose ocasionando maior predisposição ao comprometimento da saúde de pessoas com intolerância à lactose. A intolerância é decorrente da dificuldade do organismo em digerir a lactose devido à diminuição ou ausência de lactase, enzima responsável pela degradação da lactose. Os indivíduos intolerantes ao ingerirem maior concentração de lactose por meio de leites fraudados com soro, tornam-se mais suscetíveis aos sintomas da doença.

Outro problema de saúde atribuído à presença indesejável de soro de queijo no leite está relacionado ao carregamento de outras substâncias com características alergênicas e tóxicológicas. Resíduos de determinados aditivos usados na fabricação

de alguns tipos de queijo como nitratos e nitritos são carregados para o leite, quando esse tipo de soro é utilizado em adulterações fraudulentas. Consumidores sensíveis a esses aditivos e a outros que também podem ser utilizados em queijos podem ter a saúde comprometida ao ingerir leite adulterado (PERRY, 2004).

Outros autores mencionam, ainda, que a composição dos soros de queijos varia também em função da qualidade do leite utilizado no processamento de derivados. Leite com qualidade inadequada tanto sob o aspecto químico como microbiológico, ocasiona soro com qualidade também comprometida e, conseqüentemente, ao ser utilizado como adulterante, ocasionará também o comprometimento da qualidade do leite ao qual foi adicionado (PINTO, 2010; TEIXEIRA E FONSECA, 2008). Na Tabela 2, são apresentados os limites de variação dos componentes do leite bovino e dos soros de queijos doce e ácido.

TABELA 2 Variação na composição do leite bovino, do soro "doce" e do soro "ácido"

Componentes	Leite bovino	Soro doce	Soro ácido
Lactose	3,80 – 5,30%	4,00 - 5,00%	4,00 - 5,00%
Proteínas	2,30 – 4,40%	-	-
Proteínas solúveis		0,60 - 0,80%	0,60 - 0,70%
Sais minerais	0,53 -0,80%	0,40 – 0,60%	0,70 – 0,80%
Lipídeos	2,04 – 5,50%	0,20 – 0,40%	0,05 – 0,10%
E.S.T.	10,30 – 15,50%	5,30 – 6,60%	5,30 – 6,00%
Água	85,50 – 88,70%	94,70 – 93,40%	94,70 – 94,00%
Acidez Dornic	13,00 – 17,00°D	14,00 – 16,00°D	43,00 – 48,00°D
pH	6,60 – 6,80	5,90 -6,50	4,30 – 4,60

Fonte: PEREIRA *et al.*, 2001; CHAVERON *et al.*, 1979.

Outro tipo de soro lácteo que pode ser utilizado em adulteração de leite, denomina-se leitelho e corresponde ao soro resultante da bateção do creme para obtenção da manteiga (GONZALES *et al.*, 2009). O leitelho é rico em proteínas, sais minerais e sobretudo em fosfolipídios (CONWAY *et al.*, 2014; RIGUEIRA, 2006). Sua composição é muito semelhante à do leite desnatado, uma vez que, o componente retirado do creme para fabricação da manteiga é a gordura que também está ausente no leite desnatado.

A agitação do creme durante a etapa de bateção provoca a separação de duas fases distintas: uma fase aquosa chamada de leitelho e outra mais concentrada (sólida) correspondente à gordura (manteiga). Essa separação acontece como resultado da desestabilização mecânica inicial da emulsão (água/óleo) do creme utilizado como matéria-prima. Durante a bateção do creme o contato com o ar e as colisões físicas repetitivas promovem alterações na membrana dos glóbulos de gordura que levam à coalescências dos glóbulos (agregação) e conseqüentemente, separação das fases (MULVANY, 1970).

O leitelho é composto basicamente por água, lactose, proteínas, gorduras e sais minerais. Os componentes presentes em maior quantidade no leitelho além da água são a lactose e as proteínas. As proteínas predominantes no leitelho correspondem às caseínas, porém, também estão presentes, as lipoproteínas oriundas da membrana dos glóbulos de gordura (CONWAY *et al.*, 2014).

Assim como o soro de queijo, o leitelho pode apresentar diferentes valores de pH conforme o grau de maturação do creme utilizado para a fabricação da manteiga e sua composição também sofre variação em função do processo de diluição do creme para a confecção da manteiga (RIGUEIRA, 2006; SODINI *et al.*, 2006). Antes do creme ser submetido à bateção faz se necessária a sua diluição com leite desnatado ou com água. Esse procedimento relacionado ao rendimento final da manteiga e ao processo de agitação durante a bateção ocasiona alterações na composição e características sensorias do leitelho. Na Tabela 3, apresenta-se a composição média do leitelho e do leite desnatado para efeito comparativo.

TABELA 3: Composição média do leitelho e do leite desnatado

Componentes	Leitelho (%)	Leite desnatado (%)
Água	90,400	90,700
Lactose	4,010	4,800
Gordura	0,510	0,070
Proteínas	3,470	3,430
Sais Minerais	0,750	0,750

Fonte: Adaptado de GILLE, 2011.

Apesar da similaridade da composição dos soros de origem látea com a composição do leite, particularidades existentes na composição desses soros tendem a

afetar a estabilidade e a relação entre a proporção dos componentes do leite em situações de fraude. Quando se mistura soro ao leite ocorrem alterações nas propriedades nutricionais e físicas possíveis de serem detectadas por diferentes procedimentos analíticos.

Desde 2003, o Programa Nacional de Combate à Fraude no Leite (PCFL), do MAPA é responsável pela realização de análises para verificação da ocorrência de fraudes em amostras de leite UHT, pasteurizado e em pó. Essas análises correspondem à avaliação dos teores de gordura, proteínas, lactose e sais minerais com a definição do extrato seco total e desengordurado. Análises de acidez, de crioscopia, de densidade, de neutralizantes de acidez e reconstituintes de densidade também são realizadas, assim como avaliação das características sensoriais. De acordo com relatório do PCFL a quantidade de amostras analisadas e não conformes, entre os anos de 2009 e 2012, estão apresentadas na Tabela 4. Por meio da tabela observou-se o percentual de amostras cujos resultados analíticos encontraram-se fora dos padrões normativos estabelecidos para análises físico-químicas rotineiras (matéria gorda, densidade relativa, acidez titulável, extrato seco desengordurado, índice crioscópico e teor de proteínas).

TABELA 4: Amostras de leite UHT e pasteurizado analisadas pelo PCFL no período de 2009 a 2012

	2009	2010	2011	2012
Leite UHT				
Amostras analisadas	1034	1722	1994	1553
Amostras fora dos padrões	129	189	263	175
Percentual de amostras fora dos padrões	12,5	11,0	13,2	11,3
Leite Pasteurizado				
Amostras analisadas	1071	1652	1453	1296
Amostras fora dos padrões	213	239	225	162
Percentual de amostras fora dos padrões	19,9	14,5	15,5	12,5

Fonte: MAPA, 2013.

O percentual médio das amostras não conformes para leite UHT durante esse período correspondeu a 12% e para leite pasteurizado 15,35%. Apesar do decréscimo no número de amostras adulteradas, observa-se que o problema ainda persiste.

Visando garantir a segurança alimentar por meio do direcionamento das ações dos órgãos de fiscalização e da promoção do fortalecimento de procedimentos preventivos e de combate à fraudes, em 2008, o MAPA e da Agência Nacional de Vigilância Sanitária (ANVISA), criaram o Centro Integrado do Monitoramento da Qualidade do leite (Cquali-leite). Essa medida foi implantada em decorrência das fraudes em leite UHT detectadas e comprovadas pela operação “Ouro branco” em 2007 (MPF, 2007).

Em fevereiro de 2014, 104 amostras de Leite cru refrigerado foram coletadas e analisadas no laboratório credenciado do SENAI na cidade de Chapecó, Santa Catarina, Brasil. Os resultados das análises da maioria das amostras apontaram irregularidades em relação aos parâmetros: lactose, acidez, índice crioscópico, extrato seco e índice de caseinomacro peptídeo (CMP), sugerindo a ocorrência de fraude com soro ou de má conservação da matéria prima (MAPA, 2014). Várias pesquisas já foram e ainda são realizadas para verificar a ocorrência desse tipo de adulteração nos diferentes tipos de leites comercializados (ALMEIDA *et al.*, 2012; OANCEA, 2009; SANTOS *et al.*, 2013).

2. 3 Técnicas analíticas para detecção de adulterações com soro

Diferentes técnicas analíticas podem ser utilizadas para comprovação da presença de soro de queijo em leite. As análises rotineiras de crioscopia e densidade são consideradas adequadas para a detecção de fraudes com água. Segundo Cortez *et al.*, (2010) a análise de crioscopia pode apresentar resultados falso-negativos para a presença de soro no leite de acordo com a quantidade adicionada.

O índice crioscópico ou crioscopia é a medida da depressão do ponto de congelamento do leite em relação ao da água. A temperatura de congelamento do leite é mais baixa do que da água devido ao efeito das substâncias que se encontram dissolvidas no meio, principalmente, da lactose e dos sais minerais. Quando se adiciona água no leite, o ponto de congelamento aumenta em direção ao ponto de

congelamento da água (0°C) e o contrário ocorre, quando se adiciona substâncias sólidas no leite. Índices crioscópicos com valores abaixo do limite mínimo para crioscopia também indicam possibilidade de adulteração relacionada ao excesso de substâncias sólidas dispersas no meio aquoso, que por vezes, são acrescentadas juntamente com a água em fraudes econômicas com a finalidade de aumentar a densidade (BÜTTEL *et al.*, 2008; GIOMBELLI *et al.*, 2011; MENDES *et al.* 2010).

Convencionalmente, a unidade de medida adotada para temperatura nos equipamentos de crioscopia corresponde a graus Hortvet (°H) e a conversão para graus Celsius pode ser realizada, multiplicando-se o valor encontrado, por 1,0356. O valor máximo para crioscopia do leite aceito pela legislação brasileira é -0,512°C (-0,530 °H) e a faixa de crioscopia considerada para leite não adulterado corresponde a valores entre -0,512°C e -0,531°C (-0,530°H e -0,550°H) (PEREIRA *et al.*, 2001).

Essa técnica é considerada eficiente para detecção de água no leite, porém, não detecta a presença de soro no leite com a mesma eficiência. A adição de soros lácteos e de leite desnatado para aumento do volume, não alteram significativamente, a densidade do leite e o índice crioscópico. Segundo Carvalho *et al* (2007), em função da maior semelhança da composição físico-química dos soros lácteos com a do leite, essas adulterações não são facilmente detectáveis por meio das técnicas analíticas de rotina (densidade e crioscopia).

Cortez *et al.* (2010) em pesquisa para verificar alterações promovidas por quantidades progressivas de água, soro de queijo, soro fisiológico e soro glicosado em leite pasteurizado por meio de análise de densidade, acidez titulável, crioscopia, composição (lactose, proteína, gordura e minerais) e análise sensorial, constataram que a crioscopia aumenta significativamente com a adição gradativa de água no leite. Por outro lado, observaram que a adição gradativa de soro no leite não acarretou alterações significativas nos resultados para crioscopia, mesmo, com a adição de 30% de soro. Os valores crioscópicos das amostras fraudadas com soro variaram de -0,5350°H para adulteração com 1% até -0,5378°H para a adição de 30% de soro, comprovando que a crioscopia não é capaz de detectar esse tipo de fraude, uma vez que todos os valores enquadraram-se na faixa de normalidade considerada pela legislação. Os autores que trabalharam com adulterações artificiais realizadas em laboratório, atribuíram a não observação de alterações significativas com a adição de soro, com a presença dos componentes naturais solúveis (lactose, proteínas solúveis e sais minerais) comuns ao leite e ao soro de queijo. Concluíram ainda que, a presença

de água e de soro no leite pasteurizado, pode ser detectada sensorialmente quando o percentual adicionado corresponder a 25% e a 45% respectivamente.

Métodos analíticos laboratoriais específicos como: proporção cisteína/ cistina, microscopia eletrônica, teor total de sulfidril, comparação da composição de aminoácidos e determinação da relação fósforo caseínico com nitrogênio proteico, podem ser utilizados para detecção de soro no leite. Gradativamente, esses métodos que exigem excessivo trabalho, longo tempo de análise e apresentam baixa sensibilidade foram substituídos (CARVALHO *et al.*, 2007).

Técnicas sensíveis e eficientes como a eletroforese capilar e a HPLC ou CLAE, têm sido recomendadas pelos órgãos fiscalizadores para detecção de soro em leite. Diferentes técnicas cromatográficas já foram testadas para detecção desse tipo de fraude: exclusão molecular, fase reversa e troca catiônica. Contudo, de acordo com a IN. 68 de 2006, a HPLC-GF e a técnica espectrofotométrica da ninidrina ácida correspondem às técnicas oficiais utilizadas para detecção desse tipo de adulteração. Elas se baseiam na detecção do CMP e do GMP (ácido siálico livre), respectivamente, formados a partir da clivagem da *k*-caseína pela ação do coalho (quimosina) durante a etapa de coagulação do leite (BRASIL, 2006^a; BRASIL, 2006^b; BRASIL, 1991; MAGALHÃES, 2008; OLIVEIRA, 2010). Segundo Amiot *et al.* (1991), o termo CMP é utilizado para fragmentos de *k*-caseínas que não possuem carboidratos, diferentemente, do GMP que se encontra ligado a um açúcar aminado, o ácido siálico (NEELIMA *et al.*, 2013).

Entretanto, bactérias proteolíticas psicrotróficas presentes no leite de qualidade inferior ou armazenado inadequadamente, podem promover aumento nos teores de GMP e CMP. Conseqüentemente, um leite que possua contagem elevada de bactérias proteolíticas psicrotróficas poderá obter resultado positivo para fraude com soro, sem que na verdade, tenha sido adulterado com essa substância (FUKUDA, 1994; PRATA E PRATA, 2012). Diferentes técnicas alternativas, que possibilitam a detecção de fraudes em leite por adição de soros de origem láctea, vem sendo testadas. Métodos seguros, rápidos e que dispensem o pré-preparo da mesma, como a espectroscopia no infravermelho, podem contribuir significativamente para o avanço dos procedimentos para controle de qualidade do leite e detecção de fraudes (ALMEIDA *et al.*, 2012; BOTELHO *et al.*, 2015; CARVALHO *et al.*, 2007; CASSOLI *et al.*, 2011; FUKUDA, 1994; KARTHEEK *et al.*, 2011; OLIVEIRA, 2010; RODRIGUES *et al.*, 2012; RODRIGUEZ-SAONA E ALLENDORF, 2011; SANTOS, 2013; ZHANG *et al.*, 2014).

2.4 Princípios da espectroscopia no infravermelho

A espectroscopia no infravermelho é uma técnica analítica óptica que envolve interações eletromagnéticas que promovem vibrações moleculares e cuja aplicação se dá na identificação de compostos orgânicos e inorgânicos, análise de misturas complexas, controle de qualidade de diferentes produtos, em estudos de polímeros, de semicondutores, de transporte de moléculas bioativas em tecidos vivos dentre outras aplicações, desde que, as substâncias analisadas possuam ligações covalentes (BARBOSA, 2007).

A radiação no infravermelho faz com que átomos e grupos de átomos de compostos orgânicos vibrem com amplitude aumentada ao redor das ligações covalentes que os ligam. A absorção das radiações promove alterações nos modos rotacionais e vibracionais das moléculas. As alterações vibracionais convertem-se em energia de vibração molecular que configuram bandas de absorção correlacionadas aos grupos funcionais presentes em uma determinada substância. Essas bandas de absorção resultam na configuração de um perfil espectral único para cada tipo de substância, assim como, numa impressão digital capaz de identificar substâncias presentes em diferentes tipos de amostras (BARBOSA, 2007; FRANÇA E OLIVEIRA, 2011; HOLLER *et al.*, 2009; SOUYEURT *et al.* 2006).

Um equipamento denominado de espectrofotômetro emite a radiação infravermelha que é absorvida parcialmente por uma determinada substância absorvente. Na Figura 1, são representadas as etapas principais da espectroscopia no infravermelho.



FIGURA 1: Representação simplificada da espectroscopia no infravermelho.

Fonte: Adaptado de FREITAS, 2006.

A quantidade de radiação absorvida pode ser medida por meio da transmitância ou da absorvância. As equações 01 e 02 expressam o cálculo da transmitância, em que P_0 corresponde à radiação incidente e P à radiação transmitida.

$$T = \frac{P}{P_0} \quad \text{Eq. 01}$$

$$\%T = 100 * T \quad \text{Eq. 02}$$

A Absorvância é calculada pelas Equações 03, 04 e 05 que correlacionam os valores de transmitância com absorvância.

$$A = \log_{10} \frac{P_0}{P} \quad \text{Eq. 03}$$

$$A = \log_{10} \frac{1}{T} \quad \text{Eq. 04}$$

$$A = \log_{10} \frac{100}{\%T} \quad \text{Eq. 05}$$

Portanto, se um feixe radiante passa por uma substância e não ocasiona nenhuma absorção, obtém-se transmitância percentual igual a cem. Caso toda a radiação seja absorvida, a transmitância percentual corresponderá a zero e a absorvância é infinita (CECHI, 2003).

Tanto a leitura da absorvância como da transmitância são realizadas pelo espectrofotômetro que possui a capacidade de ler simultaneamente, múltiplos comprimentos de ondas, cujas intensidades de absorção geram bandas que configuram um espectro (BARBOSA, 2007).

Segundo Pavia *et al.* (2015), a unidade tecnicamente utilizada para expressar a intensidade das radiações eletromagnéticas corresponde ao comprimento de ondas, expresso em micrômetro (μm) ou micron (μ). Entretanto, a maior parte dos químicos refere-se à radiação na região vibracional do infravermelho em termos de uma unidade denominada por número de ondas que é expressa em centímetros inversos ou recíprocos (cm^{-1}). O número de ondas corresponde ao número de vezes que uma onda vibra em uma determinada unidade de distância, representando o inverso da frequência. O valor dessa medida é diretamente proporcional à energia vibracional de estiramento e de deformações angulares da substância analisada (BARBOSA, 2007).

A região do espectro eletromagnético do infravermelho se subdivide em três regiões correspondentes às diferentes frequências, comprimentos de ondas e números

de ondas denominadas como infravermelho próximo, médio e infravermelho distante. As três regiões do infravermelho apresentam as seguintes características: o infravermelho distante, com aproximadamente 400-10 cm^{-1} /números de ondas, apresenta baixa energia, sendo usado principalmente para a espectroscopia rotacional; o infravermelho médio (MIR), com número de ondas de aproximadamente 4000-400 cm^{-1} , podendo ser usado para estudar as vibrações fundamentais e está associado ao comportamento rotacional-vibracional das moléculas; por último, o infravermelho próximo (NIR), com número de ondas entre 14.000-4000 cm^{-1} , possui maior energia, sendo capaz de proporcionar vibrações em níveis mais elevados (FRANÇA E OLIVEIRA, 2011).

Na Tabela 5, apresentam-se os valores para frequência, número e comprimento de ondas nas diferentes regiões do infravermelho.

TABELA 5: Divisões da região do infravermelho

Região	Intervalo de número de ondas: cm^{-1}	Região em comprimento de onda: μm
Próximo (NIR)	12800 – 4000	0,780 – 2,500
Médio (MIR)	4000 - 400	2,500 – 25
Distante (FIR)	400 - 10	25– 1000

Fonte: Adaptado de FRANÇA E OLIVEIRA, 2011.

Tanto o NIR como o MIR são utilizados em análises de alimentos. Entretanto, o MIR é mais comumente utilizado para identificação estrutural de compostos orgânicos, devido às vibrações fundamentais promovidas nos grupos funcionais que proporcionam boa absorção e bandas melhor definidas que no NIR (KAROUI *et al.*, 2006; RODRIGUEZ-SAONA E ALLENDORF, 2011; SINELLI *et al.*, 2007).

Os espectros no infravermelho podem ser obtidos a partir de amostras: líquida, sólida e gasosa. As amostras líquidas são geralmente examinadas diretamente, enquanto as amostras sólidas ou gasosas, podem ou não, serem preparadas previamente (BARBOSA, 2007). Convencionalmente, a representação gráfica de um espectro no infravermelho médio expressa no eixo das abscissas o número de ondas (cm^{-1}) e no eixo das ordenadas, a absorvância de diferentes grupos funcionais químicos.

As regiões mais importantes do infravermelho médio, que auxiliam na identificação de substâncias estão relacionadas com o início e com o final do espectro, correspondendo às faixas de 4000-1300 cm^{-1} e 900-690 cm^{-1} . A região inicial, de alta energia, é a região de absorção dos grupamentos funcionais: hidroxila de álcool, ácido carboxílico, fenol, enol, vibrações de $-\text{NH}$ de aminas primárias e secundárias, grupo carbonila e outros. A região intermediária, que compreende a faixa de 1300-900 cm^{-1} , é conhecida como região de impressão digital, muito importante para a identificação de substâncias. Mesmo moléculas muito semelhantes apresentam diferenças significativas na distribuição dos picos de absorção nesta região do espectro. Na Tabela 6, são apresentados os grupos funcionais e suas respectivas faixas de absorção (cm^{-1} e μm) (BARBOSA, 2007).

TABELA 6: Faixas de absorção de diferentes grupos funcionais

Faixas de Absorção		
Grupo Funcional	Números de onda (cm^{-1})	Comprimento de onda (μm)
O-H	3600 - 3000	2,800 – 3,300
N-H ₂	3600 - 3100	2,800 – 3,200
C-H	3150 – 3000	3,200 – 3,300
C-H	3000 - 2850	3,300 – 3,500
C \equiv N	2400 – 2200	4,200 – 4,600
C \equiv C	2260 – 2100	4,400 – 4,800
COOR	1700 - 1750 -	5,700 – 5,900
COOH	1740 - 1670	5,700 – 6,000
C=O	1740 - 1660	5,700 – 6,000
CONH ₂	1720 - 1640	5,800– 6,100
C=C-	1670 - 1610	6,000 – 6,200

Fonte: SOARES *et al.*, 2006.

2.4.1 Equipamentos de infravermelho

Existem dois modelos de espectrofotômetros infravermelhos, um mais antigo, denominado de espectrofotômetro de dispersão e outro denominado de espectrofotômetro no infravermelho com transformada de Fourier (FTIR). Durante mais de quarenta anos espectrômetros dispersivos foram utilizados, porém, a baixa velocidade de varredura, a presença de componentes móveis sujeitos a desgastes, aquecimento da amostra dentre outras limitações, levaram ao aprimoramento do sistema analítico por meio de reflexão. (CIENFUEGOS E VAITSMAN, 2000).

A evolução da espectroscopia se deu, principalmente, a partir da combinação da transformada de Fourier (FTIR) e da nova geometria dos espectrofotômetros com a utilização do interferograma de Michelson. O funcionamento dessa ferramenta instrumental corresponde a um sistema interferométrico, que dá origem a um sinal complexo denominado de interferograma (BARBOSA, 2007). Dessa forma, um espectrômetro FTIR possui três componentes básicos: a fonte, o interferômetro de Michelson e o detector. Na Figura 2 é apresentado esquematicamente o modelo de um espectrômetro dotado de interferômetro de transformada de Fourier.

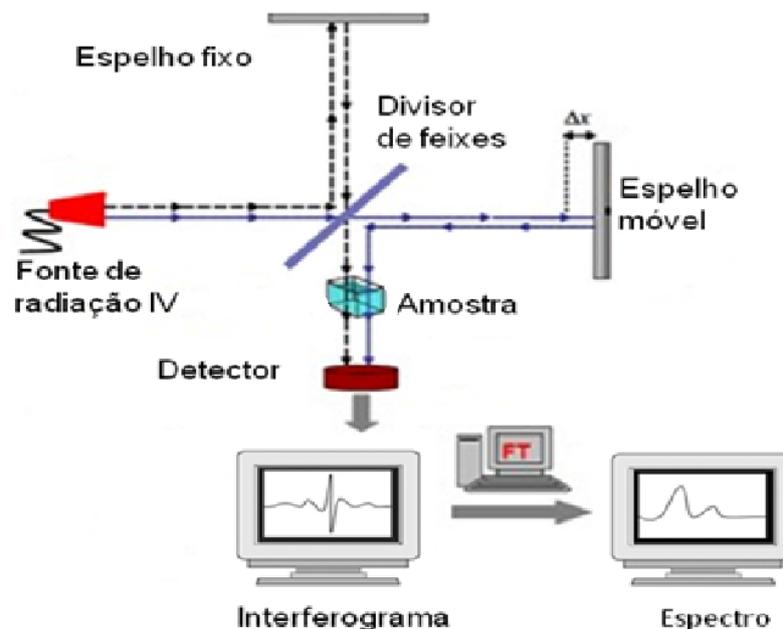


FIGURA 2: Diagrama esquemático de um espectrofotômetro com transformada de Fourier.

Fonte: Adaptado de HELFER *et al.*, 2006.

A radiação infravermelha no FTIR (fonte) corresponde à radiação promovida por elementos cerâmicos em altas temperaturas e *globars*, por meio de aquecimento elétrico. O feixe de radiação emitido pela fonte radiante é dividido por meio de um divisor de feixes constituído por um disco de brometo de potássio recoberto por um fino filme de germânio. Os dois feixes gerados após divisão (50% da radiação em cada feixe) são direcionados para um espelho fixo e outro móvel. Esses feixes percorrem diferentes caminhos e, posteriormente, são recombinados para gerar apenas um feixe simples que é captado por um detector composto por sulfato de triglicina. Esse feixe único resultante, passa através da amostra promovendo a absorção simultânea de todas as frequências (comprimentos de ondas) para configurar um interferograma (gráfico de sinais espectroscópicos). O interferograma contém informações sobre a quantidade de energia absorvida em cada comprimento de onda e transfere essas informações para um dispositivo computacional para aplicação da transformada de Fourier (algoritmo de Fourier).

Durante a trajetória dos dois feixes que corresponde à distância percorrida entre os espelhos fixo e móvel, um laser acoplado ao detector, monitora ininterruptamente o percurso. Ajustes na posição do espelho móvel fazem variar a distância percorrida por um dos feixes permitindo a obtenção de frequências de áudio correspondentes à região do infravermelho médio (diferentes extensões de ondas). Desta forma, dois interferômetros são gerados: um resultante da combinação dos dois feixes que passam pela amostra e outro resultante do monitoramento por meio do laser, das distâncias percorridas pelos espelhos. As medidas das frequências ocorrem simultaneamente e muito rapidamente, uma vez que, uma varredura (*scan*) em resolução padrão do espelho móvel dura cerca de um segundo (PAVIA *et al.*, 2015).

Para que a acumulação dos dados espectrais ocorra adequadamente, um ponto específico durante a varredura é definido para captação dos sinais que compõem o interferograma. Uma vez que o laser de referência monocromático proporciona obtenção de ondas cosseno que assumem um valor zero a cada período, cada passagem pelo zero do interferograma, sinaliza o computador para o registro de dados da amostra naquele ponto. Os pontos de captação de dados alinhados a cada varredura se acumulam para a configuração do espectro pelo interferômetro (BARBOSA, 2007; CIENFUEGOS E VAITSMAN, 2000; PAVIA *et al.*, 2015).

Entretanto, as informações geradas pelo interferômetro se configuram como um espectro no domínio temporal (intensidade x tempo) e necessitam ser transformadas

para um espectro no domínio da frequência (intensidade x frequência) para poderem ser interpretadas. Esse mecanismo é feito por meio do algoritmo de Fourier que é aplicado em posições sucessivas do espelho móvel para originar o espectro completo que expressa o valor da transmitância ou da absorvância de cada comprimento de onda. A transformada de Fourier funciona como um registrador, em que o espectro, é demonstrado em escalas de comprimento de onda (μm) ou número de onda (cm^{-1}) no eixo horizontal e absorvância (A) ou transmitância (T) em termos percentuais, no eixo vertical (AMORIM *et al.*, 2013).

A faixa de leitura (varredura) dos equipamentos comerciais disponíveis é realizada na região de número de ondas entre 5000 cm^{-1} a 400 cm^{-1} correspondente à absorção no infravermelho médio. A resolução do equipamento de infravermelho está relacionada ao caminho que o espelho móvel percorre levando em consideração o domínio do tempo e da frequência para captação de pontos que configuram os interferogramas. Quanto maior for a resolução do equipamento, maior será a distância a ser percorrida pelo espelho móvel. A resolução utilizada para aquisição dos espectros por meio da transformada de Fourier considera a densidade dos pontos como sendo a metade da resolução definida para o equipamento. Assim, uma resolução configurada para 4 cm^{-1} , terá a captação dos pontos (leitura) em 4000, 3998, 3996.....404, 402, 400 cm^{-1} (HAAK, 2010).

Os infravermelhos modernos, dotados da transformada de Fourier (FTIR), são extremamente rápidos e permitem obter diversos *scans* (leituras) de uma mesma amostra em questão de segundos. O FTIR possibilitou a digitalização dos dados e a habilitação de métodos estatísticos para melhor expressão das informações obtidas nas análises (HÉRBERGER, 2008; AMENDOLA, 2007; SITORIUS, 2010).

Outra característica instrumental do infravermelho, diz respeito às reflexões da radiação nos espectrofotômetros. Essas reflexões se diferenciam em: especular, difusa, interna e total atenuada (ATR). A ATR é uma técnica não destrutiva utilizada para amostras que absorvem fortemente na região do espectro infravermelho, como as amostras dissolvidas em água, muito absorventes (RODRIGUEZ-SAONA E ALLENDORF, 2011). Em amostras fortemente absorventes, as bandas características da água podem sobrepor bandas relacionadas com outra substância de interesse, além de ocasionar a saturação do sinal. A ATR minimiza esse problema e permite a obtenção de espectros de diferentes tipos de amostras de forma direta, requerendo que pequena quantidade da amostra seja introduzida em um dispositivo contendo cristal de

seleneto de zinco (ZnSe). Com o incremento da ATR (acessório para Reflexão Atenuada) ao FTIR, houve aumento do potencial de identificação de amostras líquidas e de adulterantes potencialmente prejudiciais (BARBOSA, 2007).

Na Figura 3, é representado o mecanismo da refletância total atenuada, em que I é o interferômetro, D o detector e L e M são respectivamente as lentes e os espelhos. As retas correspondem ao percurso do feixe do infravermelho (ETZION *et al.*, 2004).

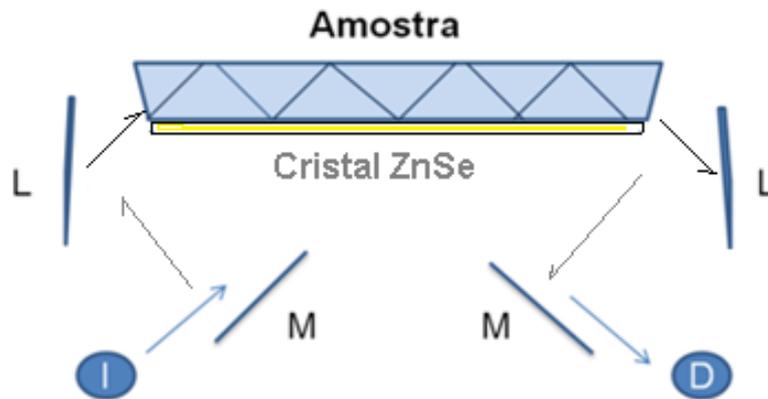


FIGURA 3: Esquema ilustrativo para dispositivo ATR

Fonte: Adaptado de ETZION *et al.*, 2004.

Os perfis espectrais gerados pela espectroscopia no infravermelho apresentam um grande conjunto de dados que não evidenciam prontamente todas as informações possíveis de serem obtidas. Para melhor aproveitamento dos dados, técnicas quimiométricas multivariadas são utilizadas como mecanismo capaz de identificar e quantificar os constituintes que estão envolvidos nas interações complexas dos componentes das amostras (PINTO, 2010).

2.4.2 Aplicação da espectroscopia do infravermelho médio em análises de alimentos e de leite

Desde a década de 1940, a espectroscopia no infravermelho é utilizada em estudos da composição de alimentos, mas, com o passar do tempo, inovações foram ocorrendo nos equipamentos e a aplicação ampliou significativamente (HOLLER, *et al.*, 2009). A intensificação da utilização do infravermelho na área de alimentos ocorreu na década de noventa, quando foram apresentadas as potencialidades da espectroscopia por reflexão no infravermelho médio (reflexão total atenuada, detecção fotoacústica e

reflexão difusa) para análise de alimentos e sua utilização no controle industrial (WILSON E TAPP 1999).

Inúmeras metodologias analíticas envolvendo aplicação de radiação infravermelha demonstram o potencial dessa técnica para determinação do teor de umidade, carboidratos e lipídeos em diferentes tipos de alimentos (CASSOLI *et al.*, 2011; KARTHEEK *et al.*, 2011; MORGANO *et al.*, 2005). Segundo França e Oliveira (2011), a espectroscopia no infravermelho corresponde a uma análise rápida e não destrutiva capaz de fornecer uma identificação precisa de cada diferente tipo de substância presente no alimento. Por meio de ampla revisão sobre a aplicação do FTIR em análises de alimentos, os pesquisadores constataram o uso da técnica em bebidas alcoólicas e não alcoólicas, em óleos comestíveis, em mel, em frutos e legumes, em café, em leite, em queijos, em carnes e seus derivados.

A técnica espectroscópica juntamente com modelos quimiométricos pode ser utilizada em práticas analíticas rotineiras, permitindo a identificação e quantificação de compostos conhecidos e desconhecidos, mesmo em baixas concentrações, que apenas poderiam ser detectadas por outras técnicas mais complexas e onerosas (HELFER *et al.*, 2006; INÁCIO, 2010; NICOLAOU *et al.*, 2010; RODRIGUEZ-SAONA E ALLENDORF, 2011; SANTOS *et al.*, 2013; SOUZA *et al.*, 2011).

Em artigo sobre uso do FTIR para rápida autenticação e detecção de adulterações em alimentos, Rodriguez-Saona e Allendorf (2011) apresentaram uma série de estudos relacionados à utilização do MIR e do NIR. Segundo os autores, o FTIR-MIR é comumente utilizado para identificação estrutural de compostos orgânicos, devido às vibrações fundamentais promovidas nos grupos funcionais que proporcionam boa absorção e bandas melhor definidas que no FTIR-NIR. Apesar das diferenças nos processos de absorção nas regiões do MIR e do NIR, ambos têm sido utilizados para determinação de padrões de alimentos de origem vegetal, sucos de frutas industrializados, produtos agroindustriais, óleos refinados, gorduras e em vários outros produtos alimentícios. Enfatizaram a aplicação do FTIR na detecção de grande número de adulterações, a partir de análise que demanda pequena quantidade de amostra, que dispensa preparo prévio da mesma e que proporciona resultados rápidos. Concluíram que os métodos baseados em análises de espectros vibracionais estão emergindo com rapidez, apresentando técnicas modernas e eficientes para monitoramento de adulterações e contaminações.

A obtenção de um espectro corresponde à base de um estudo no infravermelho e a visualização espectral do leite permite conhecer a intensidade de absorção de radiação em diferentes comprimentos de ondas. As bandas de absorção nas diferentes regiões do infravermelho correspondem aos grupos funcionais ou moléculas químicas presentes na amostra (PAVIA *et al.*, 2015).

O leite corresponde a uma emulsão composta por água, proteínas, lactose, lipídeos e sais minerais. Cada um desses componentes possui diferentes estruturas químicas que apresentam-se livres ou sob algum tipo de interação. Os componentes básicos do leite, de acordo Lynch *et al.* (2006), apresentam as seguintes características espectrais: a gordura apresenta os grupos carbonilas (C=O) das ligações ester dos triglicerídeos dos glóbulos de gordura e os grupos carbono-hidrogênio (C-H) que absorvem radiação no comprimento de onda de 5,723 μm e 3,480 μm , respectivamente. Os grupos amida (CONH) das ligações peptídicas das proteínas absorvem radiação em 6,465 μm e os grupos hidroxila (OH) da lactose em 9,610 μm . Também Wolfschoon (1977), analisando leite por espectroscopia no infravermelho, obteve resultados expressos por meio do número de ondas (cm^{-1}) ao invés do comprimento de ondas (μm) e constatou a presença de bandas características para gordura, proteína e lactose nas regiões de absorção correspondentes a 1735, 1535 e 1065 cm^{-1} , respectivamente.

Vários estudos já foram realizados para caracterização e identificação dos diferentes componentes do leite, principalmente, das proteínas de origem láctea. Segundo Kong e Yu (2007) as amidas I e II são as que possuem as bandas de vibração mais proeminentes para as proteínas. A região espectral mais sensível para os componentes estruturais da proteína, corresponde às vibrações C = O da amida I (80% da energia potencial vibracional), com banda de absorção de 1700-1600 cm^{-1} . As amidas II têm bandas de absorção derivadas das vibrações dos grupos NH (40 a 60% da energia potencial) e de ligações de carbono e nitrogênio (18 a 40% energia potencial). Outras amidas possuem bandas vibracionais muito complexas que variam de acordo com o campo de força, a natureza das cadeias laterais e com as ligações de hidrogênio.

Barbosa (2007) mencionou que quando a amida possui apenas um átomo de hidrogênio ligado ao nitrogênio, as bandas de absorção ocorrem entre 1700-1630 cm^{-1} em compostos não associados, em 1680-1630 cm^{-1} em compostos associados e em

amidas que não apresentam hidrogênio ligado ao nitrogênio (amidas III) a banda de estiramento C=O foi observada na faixa de absorção de 1670-1630 cm^{-1} .

Outra pesquisa realizada por Etzion *et al.* (2004) para avaliar proteínas lácteas por espectroscopia no infravermelho apresentou resultados semelhantes aos citados anteriormente, contemplando duas regiões de absorção das amidas I e II nas faixas entre 1600 cm^{-1} e 1700 cm^{-1} (6,3 μm a 5,9 μm) entre 1550 cm^{-1} e 1570 cm^{-1} (6,5 μm a 6,4 μm), respectivamente. Esses mesmos pesquisadores, para aprimorar o estudo sobre a identificação das proteínas lácteas, reduziram a concentração de água no meio. Conforme os autores, o modelo de calibração multivariada após concentração das proteínas possibilitou detecção mais nítida das bandas de absorção nas seguintes regiões: 1558-1705 cm^{-1} (amida I), 1480 a 1613 cm^{-1} (amida II), 1200-1280 cm^{-1} (amida III) e 1060-1100 cm^{-1} (estiramento O=P-O da caseína).

Em estudo realizado por Carvalho (2013) para caracterização por meio do FTIR-MIR de leite em pó integral instantâneo, obteve-se faixas de absorção capazes de serem atribuídas aos grupos funcionais característicos do leite. As regiões de absorção obtidas por meio do estudo e expressas em número de ondas (cm^{-1}) são apresentadas na Tabela 7.

TABELA 7: Regiões de absorção do leite de vaca no infravermelho médio.

Número de onda (cm^{-1})	Grupo	Atribuição	Modo vibracional
3438	ν_s (OH)	Gordura	Estiramento simétrico
2927-2925	ν_{as} (CH ₂)	Gordura	Estiramento assimétrico
2855-2854	ν_s (CH ₂)	Gordura	Estiramento simétrico
2350	ν_s (CH ₂)	Gordura	Estiramento simétrico
1747	ν (C=O)	Gordura	Estiramento simétrico
1660-1650	ν (C=O)	Amida I	Estiramento
1547-1541	δ (N-H)	Amida II	Deformação angular no plano (DAP)
1251-1241	ν (C-N)	Amida III	Estiramento
1160	ν (C-O)	Gordura	Estiramento
1150-1030	ν (C-O); ν (C-C); ν (C-O-O)	Carboidratos	Estiramento
800-1000	Anel de vibração	Carboidratos	(DAP)

Fonte: CARVALHO, 2013.

A faixa de absorção da estrutura lipídica do leite, devido aos diferentes constituintes do glóbulo de gordura, foi estudada por Soyeurt *et al.* (2006) por meio da comparação do espectro do leite com o de ácidos graxos específicos. A pesquisa constatou que no MIR, as principais regiões que implicam uma estimativa do perfil de ácidos graxos se localizam entre 1736 e 1805 cm^{-1} e entre 2823 e 3016 cm^{-1} . Também Santos (2013), em pesquisa sobre adulterações em leite utilizando infravermelho médio, identificou duas bandas de estiramento C-H de lipídeos na região de 3000 – 2800 e de 1852 cm^{-1} .

Amostras complexas como as de leites, possuem espectros que podem apresentar pequenas variações de absorção em função das diferentes composições influenciadas pela espécie, raça, estação do ano, período de lactação. Em estudo para avaliar detecção de fraude de leite de vaca com leite de cabra e de ovelha por meio do FTIR-MIR, os autores observaram variações na absorção das regiões correspondentes aos ácidos graxos, amidas I e II e lactose. Os modelos quimiométricos utilizados no processamento dos dados espectrais da pesquisa foram capazes de detectar a presença de leite de outras espécies em concentrações acima de 5 e 8% confirmando que o infravermelho médio, juntamente com a análise multivariada, pode ser utilizado para detectar a ocorrência de fraudes que implicam em alterações sutis em relação a amostra referência (NICOLAOU, *et al.*, 2010).

Cassoli *et al.* (2011), pesquisaram a aplicabilidade da técnica FTIR para detecção de adulterações em leite cru com bicarbonato de sódio, citrato de sódio e soro de queijo não ácido. Trabalharam com três concentrações de adulteração para cada tipo de adulterantes: 0,05%, 0,10% e 0,25% para bicarbonato; 0,025%, 0,050% e 0,75% para citrato e 5%, 10% e 20% para soro. Modelos quimiométricos de PCA foram construídos para cada grupo de adulterante. A sensibilidade da detecção por meio da técnica aplicada foi avaliada para cada uma das componentes principais (fatores) obtidas pelo modelo. Para adulteração com soro, cuja origem não foi mencionada pelos autores, os resultados apresentaram baixa sensibilidade de detecção, inclusive, para a maior concentração utilizada. Atribuíram esse resultado à limitação do equipamento MilkScan de realizar leitura em uma faixa espectral entre 1480 e 1490 cm^{-1} . Entretanto, para os demais adulterantes, consideraram os resultados satisfatórios.

Aplicação de infravermelho médio com reflectância atenuada foi testada em pesquisa desenvolvida por Santos *et al.* (2013). Os autores realizaram adulterações artificiais com soro de leite, uréia, peróxido de hidrogênio e leite sintético em leites

comerciais. Dois modelos quimiométricos foram aplicados aos dados espectrais: *Pattern Recognition Analysis by Soft Independent Modeling of Class Analogy* (SIMCA) e *Partial Least Squares Regression* (PLSR). Constataram que os dois modelos possibilitaram a identificação das adulterações artificiais promovidas no leite. A pesquisa comprovou que a técnica espectroscópica no FTIR-MIR dotado de reflectância total atenuada permite a detecção e a quantificação de adulterantes em leite em níveis satisfatórios. A Figura 5 apresenta os espectros da amostra de leite puro (amostra controle) e das demais amostras adulteradas artificialmente.

Esses estudos, comprovam que o FTIR-MIR permite a identificação rápida de proteínas e de outras substâncias presentes em leite e derivados, assim como adulterantes e contaminantes. Entretanto, técnicas de subtração de espectros podem tornar ainda mais eficiente a detecção de substâncias adulterantes ou contaminantes em leite fluido. Utilizando-se como espectro de fundo (espectro referência) a amostra de leite puro e, posteriormente, obtendo-se o espectro da amostra do leite adulterado, pode-se realizar a subtração do espectro de referência para se obter o espectro da substância adulterante. Mesmo procedimento pode ser adotado quando ocorrer sobreposição de bandas na mesma região de absorção que dificultam a identificação e caracterização das substâncias presentes nas amostras analisadas (AMORIM *et al.*, 2013; ETZION *et al.*, 2004; PINTO *et al.*, 2011).

2.5 Análises Multivariadas

O surgimento de inúmeras técnicas instrumentais de análises químicas e o crescente desenvolvimento computacional proporcionaram a geração de grande número de dados complexos que passaram a requerer a utilização de métodos matemáticos capazes de obter o máximo aproveitamento desses dados gerados. A aplicação de métodos matemáticos a um conjunto de dados de natureza multivariada permite uma simplificação do mesmo, no sentido de comprimir o espaço dimensional para melhor interpretação e visualização das informações. As análises multivariadas possuem dois objetivos básicos: fornecer métodos alternativos quando a análise estatística unidimensional é incapaz de tratar a grande quantidade de dados e possibilitar ao analista ou pesquisador e levaria à tomada de decisões melhores no contexto em que é dada a informação para o conjunto de dados analisados. À variáveis

multivariadas, métodos (modelos) quimiométricos podem ser aplicados com propósitos qualitativos (análise exploratória e reconhecimento de padrões) e quantitativos (calibração) (FERREIRA *et al.*, 1999; HAIR *et al.*, 2009).

2.5.1. Quimiometria

O termo “quimiometria” foi proposto no final dos anos 70 para descrever as técnicas e operações associadas ao tratamento matemático e à interpretação de dados químicos. Surgiu em função da necessidade de se trabalhar a vasta quantidade de dados gerados por instrumentos analíticos modernos. A quimiometria se divide em subáreas que envolvem diferentes métodos que visam a otimização e validação de metodologias analíticas, planejamento experimental, estimativa de parâmetros, processamento de sinais, análise de fatores e calibração multivariada (INACIO, 2010).

A quimiometria envolve a aplicação de métodos matemáticos estatísticos e computacionais para investigar, interpretar, classificar e fazer previsão de conjunto de dados de interesse químico (SOUZA E POPPI, 2012). Pode-se optar pela aplicação de um ou mais métodos conforme o objetivo da pesquisa (GONZÁLEZ, 2012). Esses métodos quimiométricos, também denominados de modelos, requerem, inicialmente, técnicas de pré-tratamento dos dados espectrais.

O objetivo do pré-tratamento é remover matematicamente fontes de variações indesejáveis que não são eliminadas durante a análise dos dados e que podem influenciar os resultados finais. Esses pré-tratamentos se dividem em tratamentos aplicados às amostras e aplicados às variáveis. Como técnica para pré-tratamento das variáveis a centragem dos dados na média é indicada para todos os modelo que contenham componentes principais ou variáveis latentes. Centrar os dados na média consiste em calcular a média das intensidades para cada comprimento de onda e subtrair cada uma das intensidades do valor médio (FONSECA, 2011; INÁCIO, 2010; PANERO, 2007; PARISOTTO *et al.*, 2007). A técnica para uma dada coluna j é representada pelas Equações 06.

$$\bar{X}_j = \sum_{i=1}^n \frac{x_{ij}}{n} \quad \text{e} \quad X_{ij(c)} = X_{ij} - \bar{X}_j \quad \text{Eq. 06}$$

Em que $X_{ij(c)}$ é o valor centrado na média para a variável j na amostra i ; X_{ij} é o valor da variável j na amostra \bar{X}_j

Como técnica para remoção de ruído aleatório nas amostras, utiliza-se o algoritmo móvel *Savitzky-Golay*. O método permite recalcular o ponto central de intervalos que contém determinado número de variáveis (janelas), auxiliando no alinhamento ou suavização (*smoothing*) dos sinais. O algoritmo requer a seleção do tamanho da janela (largura filtro) assim como da ordem polinomial (LIMA, 2007). Esse procedimento matemático é representado pela Equação 07.

$$x_i = \frac{1}{2m+1} \sum_{j=-m}^m w_j x_{i+j} \quad \text{Eq. 07}$$

Em que, $2m+1$ a janela móvel do algoritmo, m é um número inteiro igual ou maior que a ordem do polinômio e w_j os pesos específicos para a solução polinomial.

A correção da linha de base, outro pré-tratamento para espectros, evita que ocorram valores negativos no espectro, uma vez que, não existem valores negativos de absorvância. Para cada amostra o valor do menor ponto do espectro é subtraído de todas as variáveis. Esse procedimento é representado pela Equação 08.

$$f(x) = X - \min(X) \quad \text{Eq. 08}$$

Em que x é uma variável e X são todas as variáveis selecionadas para a amostra.

A definição da faixa espectral de um modelo também deve ser realizada antes da modelagem. Requer que sejam eliminadas as bandas que contenham ruídos visivelmente perceptíveis. Esses ruídos geralmente encontram-se localizados nas extremidades dos espectros. Entretanto, o redimensionamento da faixa espectral pode ser realizado após a obtenção do modelo visando à melhoria dos resultados. Por meio da observação dos gráficos dos escores (*scores*) e pesos (*loadings*) pode-se avaliar as regiões de absorvância de maior e menor relevância para o modelo. Os modelos nos quais são eliminadas regiões de baixa significância tendem a obter melhores resultados (BOTELHO, 2014; ETZION *et al.*, 2004; YU E IRUDAYARAJ, 2004).

Em pesquisa desenvolvida por Lyra *et al.* (2010), os autores optaram pela centralização dos dados pela média e, posteriormente, normalizaram a matriz de dados pelo desvio padrão, dividindo-se o valor de cada elemento centrado na média pelo desvio padrão da variável. Segundo os pesquisadores, o propósito dessa transformação foi permitir que todas as variáveis pudessem exercer influências equitativas nos resultados, tornando-as adimensionais. Entretanto, para cada método

existe um pré-tratamento mais adequado e simulações podem ser realizadas para verificar quais procedimentos proporcionam melhores resultados.

Após o pré-processamento dos dados, diferentes modelos de calibração podem ser aplicados aos dados espectrais. A calibração permite encontrar um algoritmo matemático capaz de converter os resultados instrumentais em informações que possam ser compreendidas e sustentadas por princípios matemáticos estatísticos (VALDERRAMA *et al.*, 2009). A Figura 4 representa, esquematicamente, passos que envolvem as análises quimiométricas multivariadas. Na calibração multivariada, busca-se estabelecer um modelo que relacione medidas químicas ou espectrais com uma determinada propriedade, como por exemplo, a concentração de um analito qualquer.

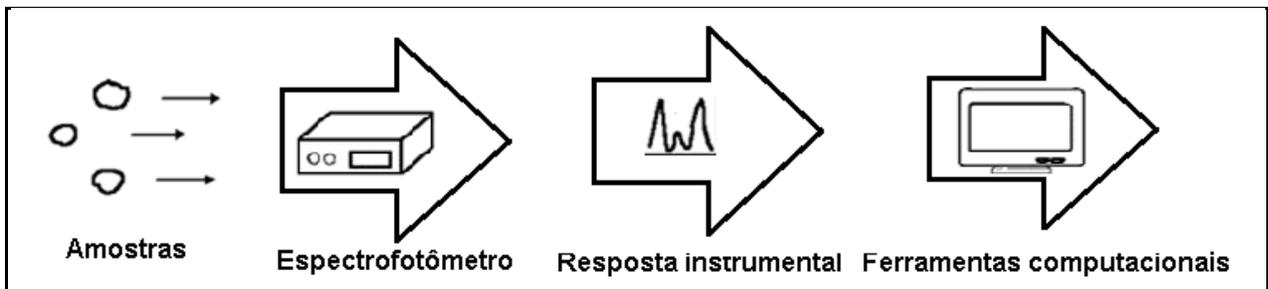


FIGURA 4: Diagrama simplificado de um processo de análise quimiométrica

Fonte: Adaptado de FERREIRA *et al.*, 1999.

A calibração multivariada consiste basicamente de duas fases: a calibração e a previsão. Na fase inicial, vários espectros são obtidos com diferentes comprimentos de onda, formando uma matriz \mathbf{X} , com n linhas e p colunas. Na Figura 5, é apresentada a construção da matriz de dados para calibração multivariada. A composição da matriz \mathbf{X} dispõe nas linhas as amostras e nas colunas as variáveis que corresponderam aos valores de absorvância (ROCHA, 2007; SIMÕES, 2008; VALDERRAMA *et al.*, 2014).

Os métodos ou modelos de calibração existentes podem ser divididos de acordo com a complexidade dos dados em: de ordem zero, de primeira ordem e de segunda ordem. Os arranjos de primeira e segunda ordem pertencem ao domínio multivariado e são processados por adequados procedimento quimiométricos (TEÓFILO, 2013; VALDERRAMA *et al.*, 2009).

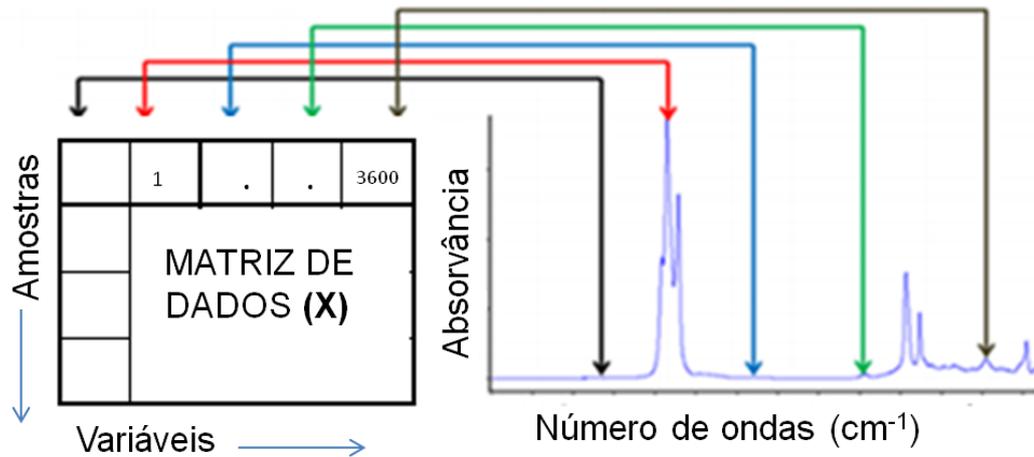


FIGURA 5: Organização dos dados em uma matriz

Fonte: Adaptado de SANTOS JUNIOR, 2011.

As análises multivariadas são capazes de processar grande quantidade de dados ao mesmo tempo, por isso, são muito utilizadas em análises de alimentos que apresentam grande complexidade na matriz (amostra). Nos métodos clássicos, univariados, somente uma absorvância atribuída a uma frequência de uma variável pode ser avaliada. Atualmente, a espectroscopia no FTIR permite manipular dados de absorvância de mais de uma frequência ao mesmo tempo. A digitalização dos dados a partir do FTIR, na presença de um interferômetro que gera os sinais referentes às absorvâncias em cada região do infravermelho, permitiu o uso sincronizado de métodos estatísticos de forma automática. Além do interferômetro, a transformada de Fourier oportunizou relevantes informações que são geradas simultaneamente a partir de diferentes dados espectrais e em certas situações, até mesmo de dados sobrepostos, relacionados aos vários componentes de uma amostra (HELFER *et al*, 2006).

Após a transformação (digitalização) dos dados espectrais para dados numéricos, estes são tratados por meio de análises multivariadas com programas computacionais independentes que apresentam interfaces gráficas amigáveis e de fácil utilização, como: Unscrambler (*Camo Software, Inc.*), o Pirouette (*Infometrix, Inc.*) e o PLS Toolbox (*Eigenvector Research, Inc.*), sendo esse último, pertencente ao ambiente Matlab (*The MathWorks, Inc.*) (BUSTAMANTE E HIERREZUELO, 2008).

Uma calibração multivariada envolve ferramentas da álgebra linear e estatística multivariada e requer um número mínimo de amostras para realização das etapas de

calibração, validação e previsão. A ASTM (*American Society for Testing and Materials*) sugere procedimentos para o desenvolvimento de análises quantitativas usando espectroscopia. Especificamente, a ASTM E1655-05 estabelece recomendações para definição do número de amostras para os conjuntos de calibração e validação interna, parâmetros estatísticos para avaliar, otimizar e recalibrar modelos de calibração, entre outras (SIMÕES, 2008). As especificações da norma, quando seguidas, propiciam bom desempenho instrumental e afetam positivamente os resultados dos modelos de calibração.

O número de amostras que devem compor o conjunto de calibração de um modelo multivariado depende da complexidade das amostras. Amostras que apresentam apenas variação na concentração requerem pequeno número de amostras para calibração (envolve pequeno número de variáveis) e amostras que apresentam maior número de componentes a serem investigados, requerem maior número de amostras para calibração (maior número de variáveis). Para se conhecer o número ideal de amostras para o conjunto de calibração, faz-se necessário rodar o modelo para que se tenha uma estimativa do número de componentes principais ou variáveis latentes requeridas (HAIR *et al.*, 2009).

A definição do número de amostras do conjunto de calibração pode ser feita também em função de k , que corresponde ao número de componentes principais para modelos PCA (*Principal Component Analysis*) e número de variáveis latentes para modelo PLS (*Partial Least Squares*) e PLS-DA (*Partial Least Square Discriminant Analysis*). Se um modelo requerer até no máximo três variáveis latentes $k < 3$, para viabilizar resultados confiáveis, o conjunto de amostras de calibração deverá ser composto por no mínimo, 24 amostras, após subtração das amostras anômalas. Se um modelo multivariado é desenvolvido usando $k > 3$ variáveis, o conjunto de calibração deverá conter um mínimo de $6k$ amostras, após a eliminação das amostras anômalas. Se o modelo é centrado na média, um mínimo de $6(k+1)$ amostras devem restar, excluindo-se, também, as amostras anômalas. Recomenda-se, ainda, fazer replicatas das amostras sempre em número maior que três (ASTM, 2012).

Uma das exigências da ASTM, normativa E 1655 05 para a construção do conjunto de calibração multivariada é que ele não possua amostras anômalas (*outliers*). Amostras anômalas são amostras que possuem alto resíduo e que, conseqüentemente, podem comprometer os resultados de um modelo. Essas amostras necessitam ser removidas e diferentes procedimentos podem ser adotados com essa

finalidade (ZHICHAO *et al.*, 2008). Usualmente, utilizam-se duas grandezas complementares para identificação de amostras anômalas: *leverage* e resíduos de Student.

O *leverage* que representa a medida da influência de uma amostra no modelo é interpretado como sendo à distância de uma amostra ao centróide do conjunto de dados. Se uma amostra tem uma medida de influência no modelo (*leverage*), maior que o valor limite ($h > 3k/nc$, onde k é o número de variáveis latentes do modelo PLS e nc o número de amostras de calibração), ela deve ser considerada uma amostra anômala (FERREIRA *et al.*, 1999; ROSSO, 2010).

Para calcular a influência de cada amostra do conjunto de calibração, utiliza-se o resíduo de Student representado pelas Equações 09 e 10 (FERREIRA *et al.*, 1999).

$$L_{rcsci} = \sqrt{\frac{(c - ci)^2}{(n - 1)(\hat{1} - hi)}} \quad \text{Eq. 09}$$

$$\text{Resíduo de Student} = \frac{(c - ci)^2}{L_{resci}\sqrt{\hat{1} - hi}} \quad \text{Eq. 10}$$

Em que L_{rcsci} é o resíduo da concentração da amostra i corrigido pela *leverage*. Sendo os resíduos de Student definidos em unidades de desvio padrão do valor médio, os valores acima de $\pm 2,5$ são considerados altos sob as condições usuais da estatística (FERREIRA *et al.*, 1999).

Matematicamente, os modelos PCA e PLS correspondem a uma regressão linear múltipla, que a partir de uma variável independente \mathbf{X} (variáveis das amostras), visa definir uma variável dependente \mathbf{Y} de n variáveis $X_1, X_2, X_3, \dots, X_n$. A função matemática que representa a relação existente entre a resposta instrumental de \mathbf{X} e a propriedade de interesse \mathbf{Y} por meio de modelo linear corresponde a Equação 11.

$$Y_{(np)} = X_{nj} B_{jp} + E_{(np)} \quad \text{Eq. 11}$$

Em que a matriz \mathbf{Y} contém os valores das p propriedades de interesse das n amostras, \mathbf{X} contém os valores de absorvância medidos ao longo dos j comprimentos de onda e \mathbf{B} , a matriz do coeficiente de regressão.

Após a etapa de calibração pela qual se obtém o modelo, procede-se à etapa de validação. Nessa etapa, é realizado um estudo quantitativo da aplicação dos resultados do modelo em amostras, com o objetivo de averiguar se o modelo desenvolvido está adequado. As amostras do conjunto de validação são retiradas aleatoriamente do

conjunto de calibração (encontram-se interpoladas no modelo) e são submetidas à mesma faixa espectral para a qual o modelo foi desenvolvido. As amostras para validação necessitam ser em número suficiente para serem distribuídas uniformemente ao longo da faixa espectral do modelo.

O número de amostras que deve compor o conjunto de validação pode ser definido em função de k . Se $k > 5$ variáveis latentes do modelo, então um mínimo de 4 k corresponde ao número de amostras de interpolação para a validação. Para modelos centrados na média, é recomendado um número de amostras para validação igual a 4 $(k+1)$ (ASTM, 2012). Não se deve utilizar menos que 10 amostras para validação do modelo, pois isso poderá acarretar elevado percentual de erro. Esse tipo de validação, que utiliza amostras do conjunto treinamento, é denominado de validação interna ou validação cruzada (*cross-validation*). Nesse procedimento de validação cada amostra do conjunto de treinamento (pelo menos uma de cada classe) é removida aleatoriamente e avaliada individualmente (TEÓFILO, 2013). Os modelos quimiométricos também podem ser validados por meio de amostras externas (avaliação externa) ou mesmo nem ser validado, caso a intenção seja uma avaliação superficial de um determinado problema.

Alguns modelos quimiométricos depois de validados podem ser utilizados para prever novas amostras. Na fase de previsão, os modelos matemáticos são aplicados para prever as concentrações ou propriedades de amostras desconhecidas, com base nos dados já conhecidos, responsáveis pela configuração do modelo.

Na Figura 6, são apresentadas as etapas do processo de calibração multivariada. No esquema, j corresponde ao conjunto dos diferentes comprimentos de ondas das amostras i (calibração) e k (validação). Esses conjuntos de dados gera uma matriz ***Xcal*** ($i*j$) para a fase de calibração e outra matriz ***Xval*** ($k*j$) para a fase de validação. Outras duas matrizes com q colunas, relacionadas aos diferentes componentes presentes nas amostras, ***Ycal*** ($i*q$) e ***Yval*** ($k*q$) também são formadas.

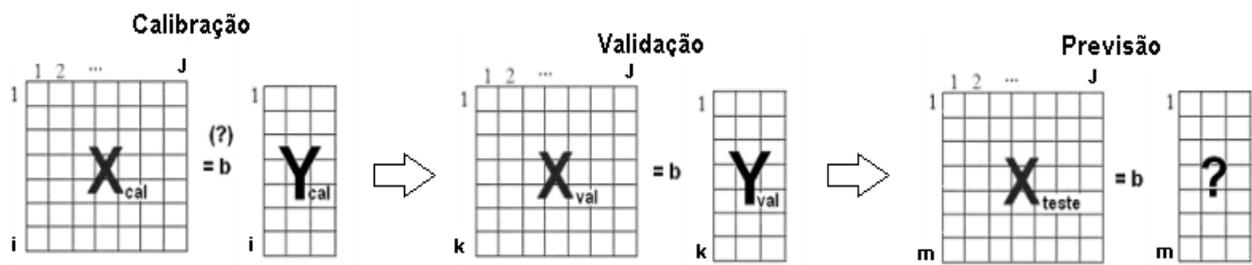


FIGURA 6: Esquema simplificado da calibração multivariada

Fonte: Adaptado de SIMÕES, 2008.

Diferentes modelos quimiométricos encontram-se disponíveis para serem utilizados em análises multivariadas, sendo classificados em dois grupos: métodos não supervisionados e métodos supervisionados. Os métodos não supervisionados são utilizados em análise exploratória cujo objetivo é extrair informações e detectar tendências nos dados, com base nas medidas multivariadas das amostras. São exemplos de métodos não supervisionados: PCA (*Principal Component Analysis*) e o HCA (*Hierarchical Cluster Analysis*).

Já os métodos supervisionados são aqueles para os quais se faz necessária a existência de informação inicial sobre as amostras para configuração das classes que servirão para reconhecimento de novas amostras. São exemplos de métodos supervisionados: LDA, (*Discriminant Analysis*), PLS (*Partial Least Squares*), KNN (*K-Nearest Neighbor*), PLS-DA (*Partial Least Square Discriminant Analysis*) e SIMCA (*Soft Independent Modeling of Class Analogy*) (SOUZA e POPPI, 2012).

2.5.2. Métodos de classificação não supervisionados

Análise exploratória de dados são técnicas utilizadas para organizar e evidenciar as informações contidas numa matriz de dados multidimensional. Técnicas de exploração de dados mais robustas são utilizadas para definir estruturas em espaços com várias dimensões, para identificar tendências, estudar as causas que influenciam as variâncias e obter variáveis latentes fundamentais ao sistema (BUSTAMANTE E HIERREZUELO, 2008).

Os métodos de análise exploratória são aqueles que não detêm prévias informações sobre as amostras ou variáveis, e por isso, são denominados métodos não supervisionados. Métodos como Análise de *Cluster*, Máquina de suporte de vetores e

Redes Neurais são considerados métodos não supervisionados. Contudo, o método de análise exploratória mais utilizado em análises de alimentos corresponde à PCA (PATAÇA, 2006).

A PCA é matematicamente definida como uma transformação linear ortogonal que transfere os dados para um novo sistema de coordenadas, de modo que a maior variância, por qualquer projeção dos dados, fica ao longo da primeira coordenada (PC1), a segunda maior variância fica ao longo da segunda coordenada (PC2) e assim por diante. O número de componentes principais a ser utilizado no modelo PCA é determinado pela percentagem da variância explicada (SOUZA E POPPI, 2012).

Normalmente, é utilizada para visualizar a estrutura dos dados, encontrar similaridades, detectar amostras anômalas e reduzir a dimensionalidade dos dados. A análise de componentes principais permite também identificar a relação entre as características extraídas dos dados, principalmente quando essas características possuem muitas dimensões (INÁCIO, 2010; SOUZA E POPPI, 2012) .

As PCs são construídas a partir de variáveis realmente representativas que desconsideram informações redundantes ou de caráter ruidoso. Correspondem a uma transformação linear onde as variáveis discretas são transformadas em coeficientes correlacionados. Esses coeficientes correlacionados se organizam no espaço configurando um ou mais componentes principais ortogonais objetivando explicar as variâncias dos dados. Para uma matriz X com n linhas e m colunas, com cada variável iniciando uma coluna e cada amostra uma linha, a matriz de covariância de X é dada pela Equação 12 (PANERO, 2007). .

$$Cov(X)_{n-1} = X^t X \quad \text{Eq. 12}$$

As colunas X depois de normalizadas são centradas na média e decompostas na soma do produto dos vetores t_i e p_i , mais uma matriz de erro E , conforme Equação 13.

$$X = t_1 p_1^t + t_2 p_2^t + \dots t_k p_k^t + E \quad \text{Eq. 13}$$

Outra forma de representar a decomposição da matriz X corresponde a Equação 14.

$$X = T P^T \quad \text{Eq. 14}$$

A PCA, portanto, decompõem uma matriz de dados X em duas novas matrizes menores: T correspondente a matriz de escores (*scores*) e P à matriz de pesos (*loadings*). Os escores correspondem à composição das PCs em relação às amostras e

os pesos correspondem às variáveis dessas amostras que definem a medida da contribuição de cada variedade original. A Figura 7 representa graficamente a decomposição da matriz de dados X em escores e pesos.

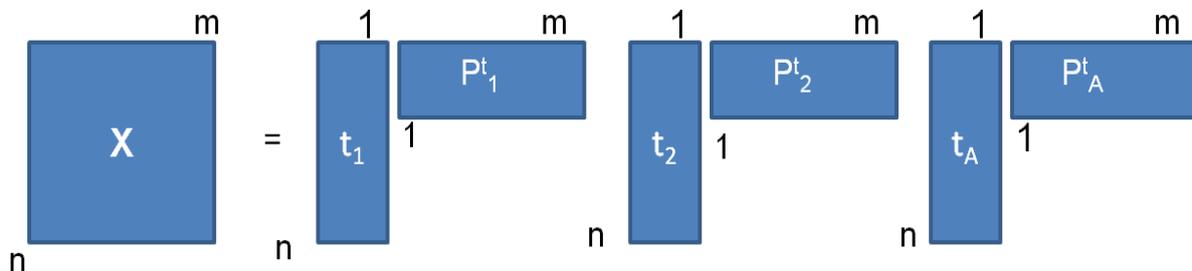


FIGURA 7: Representação da matriz X em uma PCA

Fonte: Adaptado de FONSECA, 2011.

De acordo com a representação gráfica, o número de colunas na matriz dos escores é igual ao número de linhas da matriz dos pesos e corresponde à quantidade de componentes principais (PCs) necessários para descrever as informações mais relevantes do conjunto de dados.

A construção dos eixos ortogonais: PC1, PC2, PC3, PC n , é realizada a partir da combinação linear do conjunto de variações originais. Os novos eixos (PCs) representam as direções de maior variabilidade na ordem da maior para a menor variância de PC1 até PC n , ou seja, sucessivos PCs descrevem quantidades decrescentes de variâncias. As variâncias (espalhamento dos dados ao redor do seu valor médio para uma única variável) e covariâncias (distribuição dos dados multivariados e suas relações) são capazes de explicar os dados do modelo. As informações mais relevantes sempre estão contidas nas primeiras componentes principais, e por essa razão, poucos PCs são necessários para que todas ou a maior parte da variância dos dados relevantes seja explicada (FREITAS, 2006; MORGANO, 2005; NETO, 2008).

A Figura 8 corresponde a uma representação geométrica de duas componentes principais em um espaço multidimensional delineado por 3 variáveis. As setas sinalizam a projeção das amostras em relação às PCs e os ângulos que permitem estimar a influência de cada variável em cada amostra.

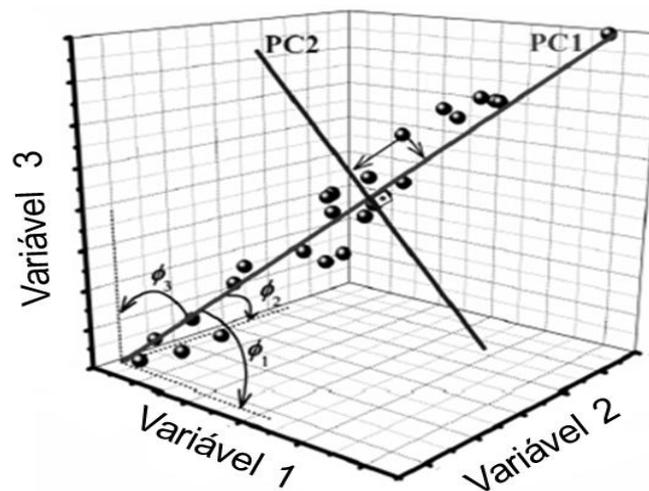


FIGURA 8: Representação geométrica de duas PCs em espaço multidimensional com três variáveis (3D)

Fonte: Adaptada de TEÓFILO, 2013.

Outro método exploratório muito utilizado corresponde ao HCA. Constitui uma ferramenta para análise preliminar de dados, sendo útil para determinar semelhanças entre amostras. O método agrupa amostras semelhantes, com base na medida da distância entre as amostras. O resultado dessas medidas ocasiona agrupamentos por similaridade, que são visualizados por diagramas conhecidos como dendrogramas. Os resultados apresentados por um dendrograma permitem a visualização das variáveis ou amostras em um espaço com duas dimensões (bidimensional), diferentemente do PCA cujos resultados podem se configurar em mais dimensões (CORREIA E FERREIRA, 2007).

Na construção do modelo HCA, um processo hierárquico vai agrupando gradativamente as amostras em classes, com base na similaridade dos participantes de uma mesma classe e nas diferenças entre os membros de outras classes. Matematicamente, o modelo vai agrupando em pares os pontos que estão mais próximos (CORREIA E FERREIRA, 2007). Para proceder à análise de agrupamentos, várias medidas de distâncias (algoritmos) podem ser utilizadas, porém, a mais usual, é a euclidiana (MORGANO, 2005; NASCIMENTO *et. al.*, 2010).

Em um espaço de m variáveis o cálculo da distância entre duas amostras a e b (D_{ab}) é representado pela Equação 15.

$$D_{ab} = \left[\sum_{j=1}^p (x_{aj} - X_{bj})^2 \right]^{1/2} \quad \text{Eq. 15}$$

Em que $p = 1, 2, 3, \dots, j$; X_{aj} é o valor da variável j para o indivíduo a e X_{bj} o valor da variável j para o indivíduo b .

Matricialmente, a distância Euclidiana é dada pela Equação 16.

$$d_{ab} = [(x_a - x_b)' * (x_a - x_b)]^{1/2} \quad \text{Eq. 16}$$

Em que: $x_a = [x_{a1}, x_{a2}, \dots, x_{ap}]'$ corresponde ao vetor de características do indivíduo a e ; $x_b = [x_{b1}, x_{b2}, \dots, x_{bp}]'$ corresponde ao vetor de características do indivíduo b .

No método de análise por agrupamento hierárquico as distâncias entre todas as amostras são calculadas e comparadas. Posteriormente a composição dos agrupamentos (*clusters*) é realizada pela definição de uma escala de similaridade. As técnicas mais utilizadas para esse fim são: ligação por vizinho mais próximo, por vizinho mais distante e pela distância entre os centróide de amostras similares (FREITAS, 2006).

A ligação por vizinho mais próximo e mais distante são calculadas pelas Equações 17 e 18.

$$S_{ab} = 1 \{D_{ab} | D_{min}\} \quad \text{Eq. 17}$$

$$S_{ab} = 1 \{D_{ab} | D_{max}\} \quad \text{Eq. 18}$$

Em que S_{ab} é a similaridade entre as amostras a e b , D_{ab} é a distância entre as mesmas, D_{min} , é a menor distância e D_{max} é a maior distância entre todas as amostras.

A representação gráfica de um dendrograma obtido por meio de um modelo HCA ilustra o arranjo de agrupamentos (*clusters*) (DANTAS E LIMA, 2010). Esses agrupamentos se assemelham aos ramos de uma árvore que vão se dividindo em outros de forma sucessiva.

Visando comprovar que os métodos quimiométricos exploratórios, PCA e HCA são ferramentas que podem ser utilizadas pelas autoridades de agências reguladoras para otimização dos resultados de análises investigativas, Souza *et al.* (2011), utilizaram esses modelos para averiguar ocorrências de adulterações em leite UHT (amido, cloro, formol, peróxido de hidrogênio e urina) produzidos em diferentes unidades industriais brasileiras.

Para avaliar com precisão concentrações e outras características quantitativas

em amostras por meio de métodos quimiométricos, o modelo de Regressão por Mínimos Quadrados Parciais (PLS) é o mais utilizado a partir de dados de primeira ordem. Por ser multivariado, mesmo que as amostras tenham interferentes, é capaz de fazer previsão da concentração do analito, desde que, os interferentes estejam presentes no conjunto de calibração.

O PLS é um método de regressão no qual o vetor da função-resposta \mathbf{y} e a matriz \mathbf{X} são decompostos em componentes principais, também denominadas variáveis latentes (LVs) (BADDINI *et al.*, 2014). O PLS é baseado na PCA e utiliza as componentes principais para modelar os diferentes analitos e interferências nos dados de \mathbf{X} . As informações das variáveis que compõem a matriz \mathbf{X} são reduzidas por meio de PCs, fazendo com que os dados se tornem mais robustos e mais fáceis de serem interpretados. No PLS, os ruídos espectrais são mantidos fora do modelo, na forma de resíduos (MARTENS E NAES, 1990).

A decomposição da matriz \mathbf{X} e do vetor \mathbf{y} é similar à do modelo PCA representada pela Equação 13. As relações entre as absorções dos espectros das amostras (matriz \mathbf{X}) com as respectivas concentrações (vetor \mathbf{y}) resultam na obtenção das matrizes de escores \mathbf{T} e \mathbf{U} e nas matrizes de *loadings* (pesos) \mathbf{P} e \mathbf{Q} . Além dessas matrizes, obtém-se também, \mathbf{E} e \mathbf{F} que correspondem aos resíduos (erros) das matrizes \mathbf{X} e \mathbf{Y} , conforme representam as equações 19 e 20 (FONSECA, 2011).

$$\mathbf{X} = \mathbf{T}\mathbf{P}^t + \mathbf{E} \quad \text{Eq. 19}$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^t + \mathbf{F} \quad \text{Eq. 20}$$

A decomposição do modelo PLS também pode ser representada graficamente pela Figura 9.

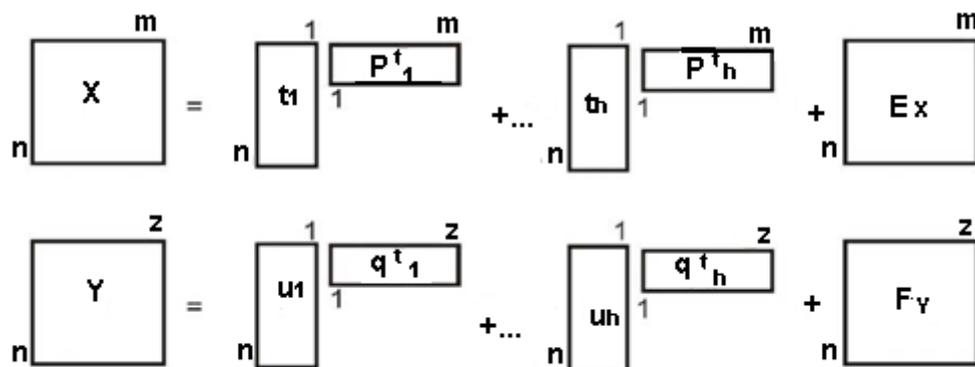


FIGURA 9: Decomposição das matrizes X e Y no modelo PLS

Fonte: Adaptada de FONSECA, 2011.

Após a decomposição das matrizes \mathbf{X} e \mathbf{y} , o modelo PLS estabelece uma relação linear entre os escores de \mathbf{X} e os escores de \mathbf{y} para cada variável latente (VL) por meio de um algoritmo. Atualmente há vários algoritmos para executar a regressão PLS e todos eles fornecem uma mesma previsão para uma única variável dependente. São eles: o algoritmo NIPALS (*Non-Linear Iterative Partial Least Squares*); o algoritmo baseado na decomposição bidiagonal (PLSBdg); o algoritmo SIMPLS, e o algoritmo Kernel, proposto por Lindgren *et al.* Entre estes, o NIPALS é o mais utilizado (TEÓFILO, 2007). A relação de linearidade é representada matematicamente pelas Equações 21 e 22.

$$U_h = b_h t_h \quad \text{Eq. 21}$$

$$Y = TBQ^t + F \quad \text{Eq. 22}$$

Em que \mathbf{b} é o vetor contendo os coeficientes de regressão linear da regressão do modelo para cada variável latente e o índice h indica o número de variáveis latentes do modelo. Depois de serem calculados os h valores (VLs), estes, passam a compor uma matriz B diagonal que contém os coeficientes de regressão dos escores das matrizes \mathbf{U} com \mathbf{y} e \mathbf{T} com \mathbf{X} . Para encontrar o melhor ajuste possível entre \mathbf{X} e \mathbf{Y} , pequenas rotações nas variáveis latentes são realizadas.

Segundo Ferreira *et al.* (1999), pode-se resumir o processo geral de calibração do PLS em duas etapas: modelagem aplicada ao conjunto de calibração e a validação do modelo de calibração. A validação permite escolher com maior precisão o número de componentes do modelo (VLs), detectar amostras anômalas e verificar preliminarmente sua robustez e capacidade de previsão. A validação é realizada com amostras que são retiradas do modelo de calibração para serem testadas uma a uma, no próprio modelo. Uma vez concluída a calibração e sua validação, o modelo também pode ser validado para amostras externas, não pertencentes ao conjunto de calibração. Esta etapa é conhecida como validação externa e objetiva garantir previsões confiáveis para concentrações/propriedades de amostras desconhecidas (PARISOTTO *et al.*, 2007).

A Figura 10 apresenta a representação geométrica da calibração do modelo PLS, em que se observa a correlação das variâncias das duas matrizes na configuração de nova variável latente (componente principal).

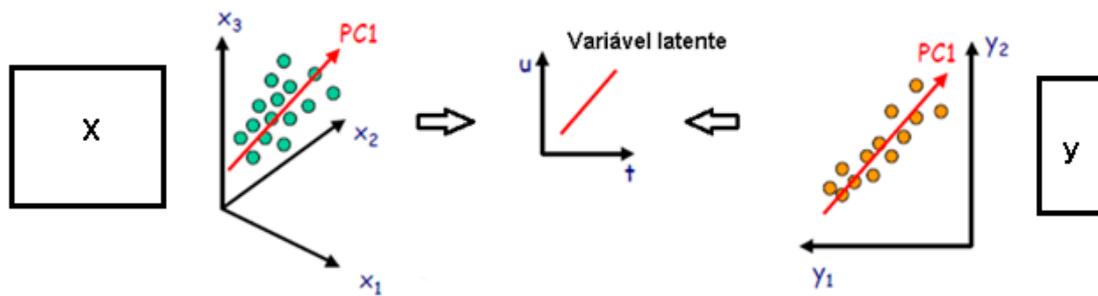


FIGURA 10: Representação geométrica do PLS

Fonte: Adaptada de ROCHA, 2007.

Os resultados do modelo PLS são expressos por meio dos valores da raiz do erro quadrático médio de calibração (RMSEC), pela raiz do erro quadrático médio de validação cruzada (RMSECV) e pela raiz do erro quadrático médio de previsão (RMSEP) (BADDINI *et al.*, 2014; INÁCIO, 2010; INON, 2004) . O cálculo de RMSEC e de RMSEP é obtido por meio das Equações 23 e 24, respectivamente.

$$RMSEC = \sqrt{\frac{\sum_{i=1}^n (y_i - y'_i)^2}{n - 1}} \quad \text{Eq.23}$$

$$RMSEP = \sqrt{\frac{\sum_i (y_i - y'_i)^2}{n}} \quad \text{Eq. 24}$$

Em que, y_i representa o valor previsto para a amostra i pelo modelo; y'_i representa o valor previsto pelo modelo e n corresponde ao número de amostras de validação.

Na etapa de validação externa, a capacidade de previsão do modelo é obtida por meio do erro quadrático médio de previsão (RMSEP) (TEÓFILO, 2013). O RMSEP é considerado um parâmetro robusto, uma vez que, é estimado por meio de um conjunto externo de amostras de validação. O valor de RMSEP (previsão) não pode ultrapassar três vezes o valor de RMSEC (calibração) (ASTM, 2012).

A validação do modelo, também pode ser expressa por meio da soma dos quadrados do erro de previsão (PRESS), representado pela Equação 25.

$$PRESS = \sum_i (y_i - y'_i)^2 \quad \text{Eq.25}$$

Além dos parâmetros de erros, a análise do coeficiente de determinação referente à correlação entre os valores previstos e valores reais permite avaliar a capacidade de previsão dos modelos PLS. O coeficiente de determinação linear R^2 explica, percentualmente, em qual proporção a variável independente X é capaz de explicar a variável dependente Y . Este coeficiente pode assumir valor na faixa $0 < R^2 < 1$. O valor de R^2 (que mede o ajuste do modelo) deve ser o mais próximo de 1 e os valores de RMSEC e RMSECV devem ser, o mais baixo possível, além de conter alta variância explicada (próxima a 100%) (BADDINI *et al.*, 2014).

Soyeurt *et al.* (2006) aplicaram modelo PLS em pesquisa para estimar o teor de ácidos graxos em leite cru por meio de espectroscopia do infravermelho médio. Segundo os autores, o PLS apresenta vantagens em relação aos modelos de regressão linear e de componentes principais porque é capaz de utilizar todas as informações espectrais, de comprimir dados e de maximizar a variância da variável dependente.

Almeida *et al.* (2012), utilizaram espectroscopia RAMAN e modelo PLS para detecção e quantificação de adulteração de leite em pó com soro de queijo a uma concentração correspondente a 2,32% (p/p). Compararam os resultados obtidos pelo método de regressão PLS e o teste da ninidrina ácida e constaram que não houve diferença significativa com o nível de confiança de 95% entre os valores encontrados. Concluíram que o modelo quantitativo PLS, mesmo com restrição de detecção de concentrações inferiores à que foi testada, podem ser usados como métodos de rastreio para análise de rotina.

2.5.3. Métodos de classificação supervisionados

Os métodos para classificação supervisionada correspondem a métodos utilizados para reconhecimento de padrões. De acordo com Berrueta *et al.*, (2007), os métodos para reconhecimento de padrão são: LDA, KNN, *Classification and Regression trees* (CART), PLS-DA, SIMCA e *Unequal Dispersed Classes* (UNEQ). Esses autores avaliaram diferentes métodos quimiométricos para reconhecimento de padrões em análises de alimentos e enfatizaram a grande importância do processamento adequado dos sinais instrumentais para obtenção de resultados rápidos e confiáveis. Concluíram que essas duas ferramentas, instrumentos analíticos e

métodos quimiométricos, apresentaram para o reconhecimento de padrões, grande potencial para análise de identidade e qualidade de alimentos.

Possuem também a capacidade de avaliar, interpretar e extrair o máximo de informações dos dados analíticos procedentes dos espectros. Os métodos supervisionados podem ser classificados como paramétricos ou não paramétricos, e como determinísticos ou probabilísticos. Os modelos paramétricos medem a proporção exata de variabilidade total dos resultados, enquanto os não paramétricos apresentam os resultados numa ordem de grandeza, medindo a variabilidade dos dados de uma forma indireta. Nos métodos LDA, PLS-DA, SIMCA e UNEQ, os dados assumem uma distribuição normal característica da estatística paramétrica (BUSTAMANT E HIERREZUELO, 2008).

Todo modelo de reconhecimento de padrões (classificação supervisionada) requer um conjunto de dados composto por conjunto de treinamento e conjunto teste. Esses conjuntos de dados são utilizados para modelagem, aprimoramento dos dados (quando necessário) e aplicação do modelo em novas amostras. O objetivo dos métodos de reconhecimento de padrões é identificar similaridades físicas e químicas, podendo, ainda, detectar a presença de amostras anômalas.

Entre os métodos de reconhecimento de padrões, destaca-se o PLS-DA, que permite construir modelos de classificação capazes de prever novas amostras. A concepção do PLS-DA foi baseada em características do método LDA e do PLS, que utiliza um algoritmo simples e de excelente poder de previsão (BERRUETA 2013). O PLS-DA, assim como o modelo PLS, pode ser aplicado em uma ou mais regiões do espectro que oportunize a obtenção de resultados mais precisos (PIZARRO *et al.*, 2013).

O PLS-DA é utilizado para a construção de modelos preditivos de classificação visando obter uma relação direta entre a resposta instrumental e a propriedade de interesse (qualitativa), utilizando os rótulos das classes como um vetor ou matriz dependente Y (ALMEIDA *et al.*, 2013; PEERBHAY *et al.*, 2013; PONTES *et al.*, 2011).

O PLS-DA configura-se a partir da existência de dois blocos de informações. As duas matrizes de dados do modelo são: matriz X ($n*j$) que representa os dados espectrais (n é o número de amostras e j o número de variáveis) e matriz Y ($n*m$) correspondente à propriedade de interesse (m número de propriedades). As matrizes são decompostas em variáveis latentes com a finalidade de diminuir o tamanho dos dados, assim como ocorre no PCA (Equações 19 e 20). Em que, T e U são as matrizes

de escores; P e Q as matrizes dos *loadings* (pesos) e E e F são os resíduos (erros) das matrizes X e Y , respectivamente.

No PLS-DA assim como no PLS existe um compromisso entre a explicação da variância em X e sua correlação com Y . Os escores de T são ortogonais e a combinação linear estimada em relação às variáveis originais são obtidas por meio do coeficiente de ponderação representado pela Equação 26 (ALMEIDA *et al.*, 2013).

$$X = XW^* \quad \text{Eq. 26}$$

Os escores de T são bons preditores de X , uma vez que, Y e X são modelados usando a mesma variável latente, conforme a Equação 27 (ALMEIDA *et al.*, 2013).

$$Y = TQ^t + F \quad \text{Eq. 27}$$

Em que F corresponde aos erros (desvios) da modelagem, representados nas Equações de correlação 28 e 29 (ALMEIDA *et al.*, 2013).

$$Y = XW^*Q^t + F \quad \text{Eq. 28}$$

$$Y = X\beta + F \quad \text{Eq. 29}$$

O coeficiente de regressão, β , pode ser descrito pelas Equações 30 e 31 (ALMEIDA *et al.*, 2013)..

$$\beta = W^*Q^t \quad \text{Eq. 30}$$

$$W^* = W(p^t w)^{-1} \quad \text{Eq. 31}$$

Em que W é definido como um conjunto de *loadings* que maximizam a covariância entre X e Y .

A previsão de novas amostras pelo PLS-DA pode ser realizada por meio do PLS-DA1 (algoritmo PLS1), quando se tem duas classes para as quais cada uma das amostras é testada para reconhecimento, ou PLS-DA2 (algoritmo PLS2) quando se tem três ou mais classes. Nesse segundo caso, o número de variáveis de resposta é correspondente ao número de classes envolvidas, obtendo-se não um vetor de correlação y (PLS1), mas sim, uma matriz de correlação Y (PLS2) (PATACA, 2006; WESTERHUIS, 2008).

O modelo PLS-DA difere do modelo PLS por fornecer dados qualitativos e não quantitativos. A classificação das amostras no PLS-DA corresponde a uma classificação binária que requer a atribuição dos valores 0 ou 1 para cada uma das

amostras. Entretanto, a classificação pelo modelo estima um valor para Y que pode ou não ser exatamente igual a 0 ou 1. Essa possibilidade requer a definição de um limite de decisão (*threshold*) que permite a separação das classes e a minimização dos números de falso positivos/ negativos. O valor do limite que separa as classes pode ser arbitrariamente fixado em 0,5 ou estimado pela teoria bayesiana. De acordo com o teorema de Bayes o valor do *threshold* corresponde ao encontro das curvas de probabilidade *a posteriori*. Valores de Y acima deste limite calculado ou fixado arbitrariamente indicam que a amostra pertence à respectiva classe, enquanto valores abaixo indicam que a amostra não pertence à classe (BOTELHO, 2014).

O estimador bayesiano é obtido a partir da ponderação probabilística que permite inferir sobre um determinado valor a partir de um conjunto de observações. O cálculo do limite bayesiano assume que a variância predita de Y segue uma distribuição similar à que será observada nas novas amostras (preditas). As duas distribuições de probabilidade dos modelos de treinamento (calibração) e teste (previsão) de uma amostra pertencer ou não a uma classe, se cruzam em um único ponto que é definido como o limite de decisão (NUNES, 2015).

O resultado da classificação binária é dado pela pontuação F (*F-score*). O valor de F corresponde ao produto dos valores das amostras verdadeiro positivas e falso negativas e dos valores das amostras falso positivo e verdadeiro negativo.

Para avaliar a capacidade classificatória do modelo PLS-DA, Reche e Franco (2009) diferenciaram quimicamente dois tipos de cachaça (alambiques de cobre e de aço inoxidável) e tipificaram esse destilado analisando 115 amostras que foram coletadas imediatamente após destilação. Realizaram a determinação quantitativa de vários compostos por cromatografia e espectrometria de emissão óptica com plasma. Posteriormente, aplicaram os métodos quimiométricos: análise de fatores (*FA-Factor Analysis*); análise discriminante com calibração multivariada por quadrados mínimos parciais (PLS-DA) e análise discriminante quadrática (*QDA - Quadratic Discriminant Analysis*). A análise de fatores conduziu a melhores resultados de explicação da variância do conjunto de dados original (66,0%) em relação à PLS-DA (54,2%). Entretanto, analisando os gráficos de escores apresentados pelas duas técnicas, observou-se que a separação se tornou mais evidente quando se utilizou PLS-DA, tendo sido constatada a ausência de *outliers*.

Botelho *et al.* (2014) propuseram um método para a detecção simultânea de cinco adulterantes em leite cru bovino utilizando reflectância total atenuada (ATR)

juntamente com espectroscopia no infravermelho médio e classificação multivariada supervisionada (PLS-DA). O método foi capaz de detectar a presença dos seguintes adulterantes na faixa de 0,5 a 10% (p/v): água, amido, citrato de sódio, formaldeído e sacarose no leite. Os autores concluíram que o método é adequado para o controle de leite cru em uma indústria de laticínios ou para o controle de qualidade.

Em determinadas situações, onde a definição da direção multidimensional de X não explica adequadamente a direção da variância multidimensional de Y e quando ocorre divergência significativa na variação entre duas classes no modelo PLS-DA, um algoritmo pode ser utilizado durante o pré-processamento para remover variação ortogonal sistemática. A correção ortogonal de sinais (OPLS-DA) visa a redução da variação na matriz X não correlacionada com a matriz Y , em que, a matriz X é corrigida por meio da subtração da variação que é ortogonal à matriz Y (PINTO *et al.*, 2012; WESTERHUIS *et al.* 2010).

Em 2005, o OPLS deu origem ao OPLS-DA (*Orthogonal Partial Least Squares Analysis Discriminant*) proporcionando uma maior capacidade discriminante para classificação de mostras por meio da redução da complexidade sua interrelação (BYLESJO *et al.*, 2006; PIOVEZAN, 2014; SADEGHI-BAZARGANI *et. al.*, 2011; WESTERHUIS *et. al.* 2010). Quando resultado do PLS-DA é bom, o do OPLS-DA tende a ser ruim e quando o resultado do PLS-DA é ruim, o do OPLS-DA tende a ser bom. (MINGOTI, 2013). A Figura 11 apresenta uma representação geométrica da diferença entre os modelos PLS-DA e OPLS-DA. Pode-se observar a característica rotacional do modelo OPLS-DA, onde a variação entre as classes (diferenças entre triângulos e círculos) encontra-se na componente de previsão tp , e a variação dentro das classes encontra-se na primeira componente ortogonal Y , para componente $t0$.

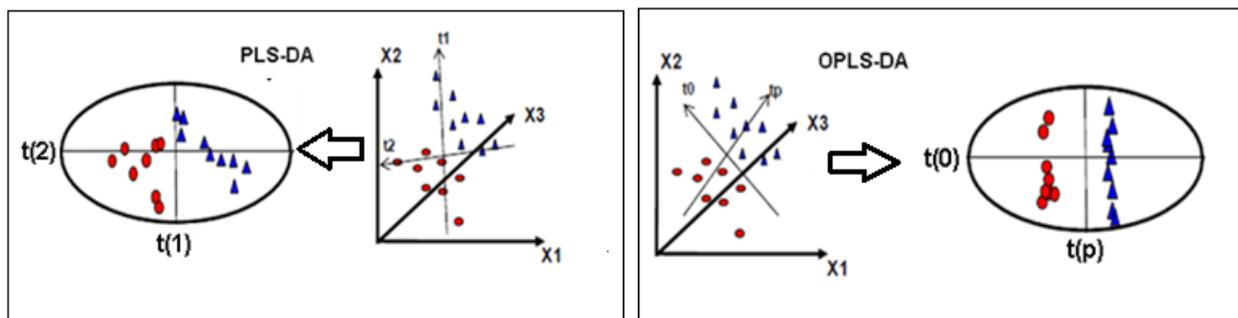


FIGURA 11: Representação geométrica dos modelos PLS-DA e OPLS-DA

Fonte: Adaptado de PIOVEZAN, 2014.

De acordo com González (2012) a decisão de se aplicar determinado método quimiométrico para a resolução de um problema, não corresponde a uma tarefa simples e irá depender da natureza particular do problema. Inúmeros métodos matemáticos estatísticos de natureza multivariada encontram-se disponíveis para uso, e, em certos casos, apresentam muita similaridade, dificultando assim a escolha. Na definição do método, deve-se levar em consideração critérios quantitativos, tais como, a capacidade de previsão e percentual de erros de classificação.

2.5.4. Validação de modelos quimiométricos

A validação de um método analítico corresponde ao uso de parâmetros matemáticos para evidenciar de forma objetiva, que os requisitos específicos para o uso pretendido do método analítico proposto são atendidos. Os parâmetros de validação de métodos analíticos envolvem seletividade, função resposta, intervalo de trabalho linearidade, sensibilidade, veracidade, precisão (repetitividade, precisão intermediária e reprodutividade), limite de detecção, limite de quantificação e robustez (ASTM 2012; BRITO *et al.*, 2003; GONDIM *et al.* 2011; IUPAC, 2012; INMETRO, 2011).;

Os parâmetros que certificam que o modelo proposto é confiável são denominados figuras de mérito (FOM). As figuras de mérito utilizadas para validação de modelos multivariados permitem avaliar a presença de erros sistemáticos nos modelos multivariados aplicados à espectroscopia. Os parâmetros a serem determinados durante a validação dependem da aplicação do método, do propósito da validação e da exigibilidade por órgãos de fiscalização (VALDERRAMA, BRAGA E POPPI, 2009).

Segundo Hatanaka (2015), a validação de métodos univariado e multivariados diferem em relação a alguns parâmetros. O cálculo de algumas figuras de mérito em modelos de calibração multivariada como veracidade, precisão, ajuste e bias são estimados de maneira similar aos métodos de calibração univariada. Entretanto, essa similaridade não é observada para a estimativa dos parâmetros de linearidade, sensibilidade, razão sinal/ruído, seletividade e intervalos de abrangência.

A *Pure and Applied Chemistry* (IUPAC, 2012) e a *American Society for Testing and Materials* (ASTM) possuem protocolos de recomendações para validação de métodos quimiométricos com base na qualidade do sinal analítico (NAS) e outras figuras de mérito (FOM), expostas a seguir:

- Sinal analítico líquido: em um modelo de calibração multivariada, o conceito de Sinal Analítico Líquido (NAS), corresponde à parte do sinal analítico que é exclusivamente relacionado com as propriedades de interesse e ortogonal a possíveis sinais interferentes (VALDERRAMA, BRAGA E POPPI, 2009). Objetiva definir adequadamente a relação entre a concentração e a resposta instrumental (BRITO, 2003). No cálculo do NAS, a matriz \mathbf{X} (dados instrumentais das amostras) é reconstruída para gerar a matriz $\hat{\mathbf{X}}_A$ e posteriormente a matriz $\hat{\mathbf{X}}_{A-k}$, que possui todas as propriedades das amostras, exceto, a propriedade de interesse k . O cálculo é representado pela Equação 32.

$$\hat{\mathbf{X}}_{A-k} = [I - \hat{\mathbf{y}}_{A,k} \hat{\mathbf{y}}_{A,k}^+] \hat{\mathbf{X}}_A \quad \text{Eq. 32}$$

Em que $\hat{\mathbf{y}}_{A,k}$ é o vetor de concentrações da espécie de interesse k estimado com A variáveis latentes e $\hat{\mathbf{y}}_{A,k}^+$ corresponde à pseudo inversa do vetor.

A matriz $\hat{\mathbf{X}}_{A-k}$, é livre do interferente k e a partir dela, se calcula o vetor NAS correspondente a $\hat{\mathbf{x}}_{A,k}$ também livre de interferentes, representado pelo escalar $nas = \|\hat{\mathbf{x}}_{A,k}^{nas}\|$. Com o valor escalar obtido é possível construir uma nova forma de calibração multivariada, em que o modelo pode ser representado em uma forma pseudounivariada que possibilita avaliar a porção do sinal que participa eficientemente do modelo. Por meio do NAS se calcula os parâmetros de seletividade e razão sinal/ruído.

- Veracidade: expressa o grau de concordância entre o valor estimado ou medido e o valor tido como verdadeiro ou de referência. Comumente, em calibração multivariada, os valores da raiz quadrada dos erros médios quadráticos de calibração (RMSEC, do inglês *Root Mean Square Error of Calibration*), calibração cruzada (RMSECV) e previsão (RMSEP- do inglês *Root Mean Square Error of Prediction*) podem ser referidos como estimativas da veracidade em trabalhos com calibração multivariada (GONDIM, *et. al.*, 2011; HATANAKA, 2015; VIM, 2012). As Equações 23 e 24 correspondem aos cálculos para RMSEC e RMSEP. Em geral, a veracidade é expressa pelo valor de RMSEP (VALDERRAMA, BRAGA E POPPI, 2009; ROCHA 2007).

- Sensibilidade: é definida como fração de sinal responsável pelo acréscimo de uma unidade de concentração da propriedade de interesse. Por meio do NAS, a sensibilidade é determinada pela Equação 33 (VALDERRAMA, BRAGA E POPPI, 2009).

$$\hat{S}_k^{nas} = \frac{\hat{X}_{A,k}^{nas}}{Y} \quad \text{Eq. 33}$$

- Sensibilidade Analítica: este parâmetro é aplicado em métodos analíticos que apresentam ruído da linha de base. É definido como a razão entre a sensibilidade e uma estimativa do desvio padrão para a flutuação do sinal analítico (ruído instrumental) e é representado por δx . A Equação 34 é utilizada para o cálculo da sensibilidade analítica (HATANAKA, 2015; VALDERRAMA, BRAGA E POPPI, 2009).

$$y = \frac{S\hat{E}N}{\|\delta x\|} \quad \text{Eq. 34}$$

- Seletividade: a seletividade mede a pureza do sinal analítico em relação ao sinal total. Em outras palavras, o grau de sobreposição entre o sinal da espécie de interesse e os interferentes presentes na amostra e indica a parte do sinal que é perdida por essa sobreposição. Para modelos de calibração multivariada esse parâmetro é definido pela Equação 35 (HATANAKA, 2015; BRAGA E POPPI, 2004).

$$SELi = \frac{n\hat{a}s_i}{\|x_i\|} \quad \text{Eq. 35}$$

- T_{Bias} : a norma ASTM E1655-05 propõe a investigação deste parâmetro por meio de um teste-t para as amostras de validação no nível de 95% de confiança. O valor de Bias é atribuído à presença de erros sistemáticos e é calculado pela Equação 36.

$$bias = \sum_{i=1}^n \frac{(y_i^{ref} - \hat{y}_i)}{n_v} \quad \text{Eq.36}$$

Em que y_i^{ref} e \hat{y}_i são os valores de referência e preditos, respectivamente, e n_v é o número de amostras do conjunto de validação.

Posteriormente, o desvio padrão dos erros de validação é estimado para avaliar o grau de concordância entre os resultados de cada teste, quando aplicado repetidamente às várias amostragens de uma mesma amostra (VALDERRAMA, BRAGA E POPPI, 2009). O desvio padrão dos erros de validação (SDV) é obtido pela Equação 37.

$$SVD = \frac{\sqrt{\sum[(y_i - \hat{y}_i) - bias]^2}}{n_v - 1} \quad \text{Eq.37}$$

Finalmente, o valor de *tbias* é calculado para determinar se a estimativa da validação apresenta tendências (*bias*) estatisticamente significativas. É obtido por meio da equação 38.

$$tbias = \frac{|bias|\sqrt{n}}{SDV} \quad \text{Eq.38}$$

O valor de *t* calculado é comparado ao valor de *t tabelado* para o número de grau de liberdade adequado (número de amostras de previsão). Quando o valor de *t* calculado é menor que o valor de *t tabelado*, significa que as amostras previstas pelo modelo multivariado apresentam o mesmo resultado médio obtido pelo modelo de treinamento (interpolação adequada do modelo). Se o valor de *t* calculado é maior que o valor de *t tabelado*, existe 95% de probabilidade que a estimativa do modelo multivariado não apresente a mesma média de resultados do método de referência (BRAGA E POPPI, 2004; HATANAKA, 2015; VALDERRAMA, BRAGA E POPPI, 2009).

A Linearidade (*R*) é utilizada para avaliar a dispersão da reta de regressão entre as variáveis independentes **X** e as dependentes **Y**, ou seja, mede a qualidade do ajuste de uma regressão linear. Seu valor pertence a um intervalo correspondente a $-1 \leq R \leq 1$. Para interpretação do valor de *R*, deve-se levar em consideração o número de amostras do modelo. A linearidade é calculada pela Equação 39.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^n (x_i - \bar{x})^2] [\sum_{i=1}^n (y_i - \bar{y})^2]}} \quad \text{Eq.39}$$

Em que, $(x_1, y_1), \dots, (x_n, y_n)$ são os *n* pares observados.

Especificamente para modelo qualitativo como o PLS-DA no qual o resultado não envolve uma regressão linear numérica, mas sim uma classificação binária, a validação da acórdância ou concordância se faz necessária. Esses dois termos são utilizados para definir a precisão do modelo e estão relacionados com a probabilidade de se obter para duas amostras idênticas, o mesmo resultado. O termo acórdância é utilizado quando essas duas amostras são testadas sob as mesmas condições, no mesmo laboratório. Concordância é o termo utilizado quando essas duas amostras são testadas em laboratórios diferentes (GONDIM, *et al.*, 2011).

A sensibilidade e seletividade são obtidas a partir de uma matriz que contém informações sobre as correlações entre as amostras classificadas. São medidas que podem ser realizadas tanto para conjuntos de calibração (treinamento) como para

conjunto de previsão (teste), mas que, usualmente são aplicadas no conjunto teste para validação do modelo.

Uma forma de analisar graficamente os resultados obtidos pelo modelo classificatório envolve a utilização da curva de características operativas do receptor (ROC, *Receiver Operating Characteristic*). O gráfico das curvas ROC corresponde a uma técnica de visualização gráfica que envolve a organização e seleção de classificadores com base no seu desempenho. O gráfico é constituído por duas curvas: uma representada por amostras não ligadas (correlação interclasse) e outra correspondente a amostras ligadas (correlações intraclasse), correspondentes à sensibilidade e seletividade. A correlação entre pares de amostras é dada pela função cosseno (ρ), correspondente ao cosseno quadrado do ângulo entre duas amostras, conforme representa a Figura 12.

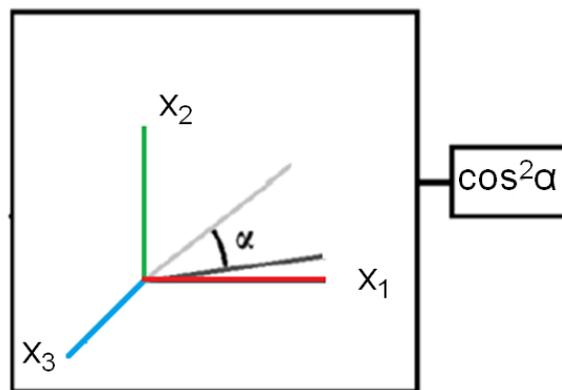


FIGURA 12: Gráfico da função cosseno de três amostras em um espaço tridimensional

Fonte: Adaptada de GROBÉRIO, 2012.

As amostras submetidas à avaliação da especificidade e sensibilidade por meio das curvas ROC podem ter os resultados classificados como Verdadeiro Positivo (VP) ou Falso Positivo (FP) quando apresentarem alto cosseno. Por outro lado, quando o valor do cosseno for baixo, podem ser classificadas como Verdadeiro Negativo (VN) ou Falso Positivo (FN). A curva ROC é um gráfico de sensibilidade (ou taxa de verdadeiros positivos) *versus* taxa de falsos positivos (seletividade), em que, os parâmetros envolvidos no cálculo da sensibilidade $[VP/(VP+FN)]$. e da seletividade $[VN/(VN+FP)]$ são: VP, que corresponde a resultado Verdadeiro Positivo; FP, que corresponde a resultado Falso Positivo; VN, que corresponde a resultado Verdadeiro Negativo e FN, que corresponde a resultado Falso Negativo (GROBÉRIO, 2012).

Na Tabela 8 são apresentados os critérios de classificação dos resultados de um modelo qualitativo. Quanto menor o número de resultados falso positivo e falso negativo melhor será a classificação das amostras.

TABELA 8: Possíveis resultados para classificação de amostras

Cosseno Quadrado	Amostras Ligadas (Parâmetros de sensibilidade)	Amostras não ligadas (parâmetros de seletividade)
Alto	VP	FP
Baixo	FN	VN
Resultados	VP + FN = 1	FP +VN = 1

Fonte: GROBÉRIO, 2012. Adaptada.

Os valores atribuídos pelas curvas ROC a um modelo, têm origem na área sob as curvas (AUC, do inglês, *área under the curve*) de amostras ligadas e não ligadas. Enquanto as curvas ROC expressam o desempenho do modelo, a AUC é utilizada como estatística de diagnóstico do modelo. Quanto maior o valor de AUC, menor é a área de sobreposição das curvas e melhor é a classificação do modelo. O valor máximo de confiabilidade obtido por meio da AUC de um modelo, é 1 (um) e o valor mínimo é 0,500 (SZYMANSKA *et al.*, 2012) .

A Tabela 9 apresenta as equações utilizadas para cálculo das figuras de mérito de modelos qualitativos, assim como, a definição sobre cada uma dessas FOM. É importante ressaltar que o *F-score* corresponde à ponderação dos parâmetros de sensibilidade e seletividade pode também ser obtida por meio da Equação 37.

$$F_{\beta} = (1 + \beta^2) \frac{\text{seletividade} + \text{sensibilidade}}{(\beta^2 * \text{seletividade}) + \text{sensibilidade}} \quad \text{Eq.37}$$

Em que β é a probabilidade de resultados falso-negativos e o seu valor, usualmente, corresponde a 5%. Essa probabilidade é também denominada como taxa de erro tipo II, sendo mais convencionalmente utilizada. Na avaliação do erro tipo II, o limite de detecção do modelo coincide com o limite superior da região de perda da confiabilidade (taxa de sensibilidade de 95%). No erro tipo I (α), que considera a probabilidade de ocorrência de resultados falso-positivos o limite de detecção do modelo coincide com o menor limite da região de perda de confiabilidade. A utilização de uma ou outra probabilidade de erro depende do objetivo da pesquisa (GONDIM, *et al.* 2011). A interpretação do *F-score* é realizada em função do valor de β (erro tipo II). Para $\beta=1$, modelo obteve ótima ponderação entre seletividade e sensibilidade; $\beta >1$, o

modelo apresenta maior característica de sensibilidade e se $\beta < 1$ modelo apresenta maior característica de seletividade (CAMPELO, 2015).

TABELA 9: Figuras de mérito para modelos qualitativos

Medida de desempenho	Cálculo	Definição
Sensibilidade (Taxa de VP ou Recall)	$\frac{VP}{VP + FN}$	Percentagem de amostras positivas classificadas corretamente sobre o total de amostras positivas
Seletividade	$\frac{VN}{VN+FP} = 1-TFP$	Percentagem de amostras negativas identificadas corretamente sobre o total de amostras negativas
Acurácia	$\frac{VP + VN}{VP + VN + FP + FN}$	Capacidade de predizer a classe de uma nova amostra
Precisão	$\frac{VP}{VP + FP}$	Percentagem de amostras positivas classificadas corretamente sobre o total de amostras classificadas como positivas
F-Score	$\frac{2 * (precisão * recall)}{(precisão + recall)}$	Média ponderada da precisão e do recall
Taxa de FP (TFP - Erro I)	$\frac{FP}{FP + TN} * 100$	Possibilidade de erroneamente detectar a presença do analito (admitir um resultado falso-positivo)
Taxa de FN (TFN- Erro II)	$\frac{FN}{FN + TP} * 100$	Possibilidade de erroneamente detectar a ausência do analito (admitir um resultado falso-negativo - erro tipo II)

Fonte: CAMPELO, 2015; DAVIS E GOADRICH, 2006. Adaptada.

3 METODOLOGIA

3.1 Testes preliminares

Algumas análises preliminares foram realizadas para dar sustentação aos propósitos da pesquisa: análises de crioscopia para verificar se a técnica é capaz de detectar adulterações em diferentes níveis com soro em leite; análise da qualidade do leite utilizado para simulação de adulterações artificiais (pH, densidade, avaliação das características sensoriais); análise exploratória espectroscópica para discriminação entre soros de origem láctea (modelos PCA e HCA); análise espectroscópica para classificação de soros de manteiga (leitelho) (modelo PLS-DA) e análise espectroscópica para classificação de soros de queijos (modelo PLS-DA). Os resultados dos testes preliminares foram apresentados no capítulo IV.

3.1.1 Análise de Crioscopia

Foram preparadas 13 amostras contendo: leite UHT (*Ultra High Temperature*) integral não adulterado (puro), soro de queijo, soro de manteiga (leitelho), leite UHT integral adulterado artificialmente com 1, 5, 15, 25 e 35% de soro de queijo e amostras de leite UHT integral adulterado artificialmente com 1, 5, 15, 25 e 35% de leite proveniente do processamento de manteiga (parcialmente diluído). Foram realizadas três leituras de cada amostra no crioscópio eletrônico. Concentrações acima de 25% de soro no leite fluido são perceptíveis sensorialmente e, por isso, tendem a não serem utilizados em fraudes (CORTEZ *et al.*, 2010).

Leites UHT integral (2 litros, mesma marca) foram adquiridos em supermercado da cidade de Uberlândia, MG, dentro do prazo de validade e com características expressas em rótulo que se enquadraram dentro dos padrões legais. Esses leites foram utilizados nas adulterações artificiais com dois diferentes tipos de soros lácteos. Os soros lácteos provenientes do processamento de queijo Minas Padrão e de manteiga de Primeira Qualidade foram cedidos por laticínios localizados em regiões

próximas à cidade de Goiânia, GO. Foram coletados um litro de cada tipo de soro em dois diferentes estabelecimentos (laticínios). Essas amostras foram congeladas e transportadas para cidade de Uberlândia.

Para confecção das amostras adulteradas artificialmente, os soros foram descongelados e adicionados ao leite UHT em diferentes concentrações. As amostras de leite adulteradas foram mantidas sob refrigeração em temperatura inferior a 5° C até serem encaminhadas em recipiente isotérmico para o laboratório LAMAM (Laboratório de Ensaio em Alimentos e Meio Ambiente) do SENAI, localizado na cidade de Uberlândia. Esse laboratório é acreditado pela Divisão de Acreditação de Laboratórios (DICLA) de acordo com os requisitos da Associação Brasileira de Normas Técnicas (ABNT) NBR ISO/IEC 17025. As análises crioscópicas foram realizadas seguindo os procedimentos analíticos do laboratório Nacional Agropecuário (LANAGRO), MET POA/10/01/01, em triplicata e em crioscópio eletrônico PZL 7000.

A análise de crioscopia utilizada principalmente para detectar fraude econômica por adição de água em leite fluido (cru, pasteurizado e UHT) foi testada para detecção de fraude econômica com soro de queijo e leiteiro. Operacionalmente, a análise requer que a amostra de leite seja colocada em pequeno tubo que é inserido em um cabeçote que abaixa com o acionamento do botão de operação para que a amostra possa ser submetida a um super resfriamento (-3°C). A amostra é mantida por um determinado tempo dentro da câmara de refrigeração sob agitação de uma haste que compõe um sistema vibratório. A agitação da amostra permite que ocorra a liberação do calor de fusão até atingir um *plateau* que permita a leitura do ponto de congelamento (PEREIRA *et al.*, 2001).

O equipamento forneceu o índice crioscópico na escala Hortvet (°H) e a equivalência entre as escalas Hortvet (°H) e Celsius (°C) puderam ser obtidas por meio do seguinte cálculo (Eq.41).

$$T(^{\circ}C) = 0,9656 * T(^{\circ}H) \text{ ou } T(^{\circ}H) = 1,0356 * T(^{\circ}C) \quad \text{Eq.41}$$

O valor da crioscopia do leite para ser considerada normal deve estar entre -0,512 e -0,531 °C (-0,530 °H e -0,550 °H) (BRASIL, 2006^a). Por meio da Figura 13 pode-se observar o equipamento utilizado para análise crioscópica do leite no laboratório do SENAI, Uberlândia, MG.



FIGURA 13: Crioscópio eletrônico PZL 7000.

Fonte: Própria autora

3.1.2 Avaliação da qualidade dos leites UHT utilizados nas adulterações com soro

Foram adquiridos leites UHT integral de diferentes lotes de três fabricantes (3 litros de cada) em supermercados de Goiânia, GO (amostras A, B e C). Os leites foram mantidos à temperatura ambiente até serem abertos. As análises de pH, densidade e avaliação das características sensoriais foram realizadas logo após a abertura das embalagens dos lotes de cada fabricante de leite UHT integral.

A medida do pH foi realizada usando-se pHmetro da marca PHTEK, modelo composto por um eletrodo de medição sensível ao íon presente na amostra e por um potenciômetro devidamente calibrado usado como referência. Antes da realização da análise, o pHmetro foi calibrado com as soluções tampão pH 4 e tampão pH 7. A leitura do pH foi feita à temperatura de 25°C.

Em leite normal o pH varia entre 6,6 e 6,8 com média de 6,7 a 20° C ou 6,6 a 25° C (PEREIRA *et al.*, 2001; BRASIL, 2006^a). Na Figura 14 é apresentado o pHmetro utilizado.



FIGURA 14: PHmetro modelo 100 da PHTEK

Fonte: Própria autora

A densidade, outra análise para avaliar os leites utilizados nas adulterações simuladas, foi realizada por meio de um termolactodensímetro calibrado da marca Incoterm de massa constante. O aparelho é dotado de duas escalas de medidas: uma para temperatura e outra para densidade. A técnica requereu que o termolactodensímetro fosse imerso lentamente, para evitar a formação de espuma, em uma proveta contendo 500 mL de leite. A massa do lactodensímetro promoveu o deslocamento de volume do leite na mesma proporção da massa do lactodensímetro. Esse deslocamento fez com que o líquido alcançasse um valor na escala graduada em graus densitométricos e outro valor na escala de temperatura. Para obtenção da densidade real, fez-se a conversão do valor encontrado para densidade à temperatura ambiente (densidade lida) para densidade a 15°C. A conversão foi realizada por meio de tabela que acompanha o aparelho. Essa conversão se fez necessária pelo fato do termolactodensímetro ser aferido a 15°C.

De acordo com os requisitos físicos-químicos para leite fluido, a densidade mínima a 15°C deve corresponder a 1,028 e a máxima a 1,034 g/mL (LERAYER *et al.*, 2002; BRASIL, 2006^a). A Figura 15 apresenta o termolactodensímetro utilizado nas análises de leite.



FIGURA 15: Termolactodensímetro

Fonte: Própria autora

Também foi realizada avaliação das características sensoriais dos lotes de cada fabricante a fim de detectar alguma anormalidade aparente para exclusão do lote. As observações sensoriais dos leites utilizados nas adulterações no laboratório foram realizadas pelo próprio pesquisador e correspondeu à observação da cor, do odor, do sabor, da ausência coágulos e de detritos.

Após a retirada de quantidade suficiente de leite para realização dessas análises iniciais para avaliação da qualidade, o restante dos leites foi mantido nas próprias embalagens e armazenados em temperatura inferior a 5°C, por no máximo 24 horas.

Além dessas análises, foram obtidos espectros de todos os lotes a fim de verificar evidências de alguma anormalidade com base no perfil espectral. Os resultados das análises para avaliação da qualidade do leite utilizado nas adulterações simuladas foram apresentados no capítulo IV.

3.2 Análise espectroscópica

3.2.1 FTIR-MIR

Neste trabalho foram utilizados espectrofotômetros no infravermelho médio da marca *Perkin Elmer* com transformada de Fourier, operado por microcomputadores contendo *software* específico, denominado de *Spectrum 10*. O *software* foi projetado para espectrofotômetros da marca *Perkin Elmer* dotados de transformada de Fourier e ferramentas computacionais capazes de realizar a coleta de dados espectrais, processamento e geração de resultados.

Para realização das análises no infravermelho, a fonte de luz incidente ininterruptamente foi dividida em dois feixes, cujos caminhos relativos variaram em função da presença de um elemento óptico móvel (espelho). Depois de percorridos os diferentes caminhos os feixes foram recombinados para gerar um feixe único que incidiu sobre a amostra e que, concomitantemente, produziu um padrão de interferência contendo todos os comprimentos de ondas da faixa de absorção no infravermelho médio, para qual o equipamento foi calibrado.

Ao incidir sobre a amostra, esse feixe de radiação infravermelha promoveu absorção eletromagnética e, conseqüentemente, movimentos vibracionais nas moléculas. Essa movimentação molecular ocasionou a liberação de energia (frequência) que foi captada pelo detector. O detector, ligado a uma interface computacional foi capaz de registrar os dados (número de ondas expressos em cm^{-1}) a cada intervalo de tempo. O equipamento computacional registrou os sinais eletromagnéticos emitidos em intervalos definidos pela configuração do equipamento.

Esses sinais, tanto de referência da fonte irradiante como da amostra, foram convertidos em dados matemáticos pela transformada de Fourier. O tratamento matemático promovido pelo algoritmo de Fourier oportunizou o processamento

simultâneo de dados e a eliminação de interferências geradas pelo próprio equipamento que poderiam comprometer a qualidade dos dados das amostras.

No presente trabalho a faixa de absorção utilizada inicialmente foi de 4000 a 600 cm^{-1} , porém, essa faixa sofreu alterações em determinados modelos para promoção de ajustes que levaram a melhores resultados. A resolução do equipamento foi de 4 cm^{-1} correspondente ao padrão para os equipamentos FTI-MIR. Essa resolução proporcionou uma varredura em aproximadamente um segundo.

Na Figura 16 está apresentado o conjunto de equipamentos que foram utilizados para realização das análises espectrofotométricas: infravermelhos *Spectrum* da *Perkin Elmer* e computadores de mesa acoplados, nos quais se encontram instalados *software Spectrum 10* responsável pelos comandos dos equipamentos.



Modelo *Spectrum Two* (*Perkin Elmer*)



Modelo *Spectrum 400* (*Perkin Elmer*)

FIGURA 16: Conjunto de espectrofotômetros da Perkin

Fonte: Própria autora

Os equipamentos apresentados na Figura 21 pertencem ao laboratório de biocombustíveis (LABIC-UFU), Uberlândia, MG, e ao laboratório Central de Análise Multiusuário (CAM) da UFG, em Goiânia, onde foram realizadas as análises espectroscópicas. Os equipamentos de infravermelho (IV) da marca *Perkin Elmer* possuem as mesmas configurações e o conjunto de espectros obtidos em cada equipamento foi utilizado em modelos específicos

3.2.2 Acessório de ATR

O espectrofotômetro da *Perkin Elmer* é dotado de acessório de reflectância total atenuada (*ATR-Attenuated Total Reflection*) que envolve uma técnica de amostragem simples e rápida que não requer preparo prévio da amostra, bastando que seja diretamente colocada no compartimento.

Na Figura 17 encontra-se apresentado o acessório de ATR com o compartimento onde foram colocadas as amostras. Esse acessório é comum aos dos equipamentos IV utilizados.



FIGURA 17: Acessório de ATR

Fonte: Própria autora

A cada troca de amostra o compartimento da ATR foi rigorosamente limpo com álcool isopropílico. Esse acessório também pertence ao laboratório de biocombustíveis (LABIC-UFU), Uberlândia, MG.

3.2.3 Obtenção dos espectros

Para obtenção dos espectros a amostra foi diretamente colocada no compartimento do acessório de reflectância atenuada em quantidade suficiente para cobrir todo o cristal de ZnSe. As amostras de leite puro cru e UHT integral, assim como, as amostras adulteradas com soros lácteos não foram submetidas a nenhum tratamento prévio. Os dados sobre os parâmetros da análise foram informados ao *software* do equipamento antes de iniciar as análises. Esses parâmetros corresponderam à identificação da amostra e aos dados referentes à configuração

necessárias para obtenção dos dados (modo de aquisição, faixa de leitura do espectro, número de varreduras, resolução e unidade para configuração dos espectros).

A definição do número de varreduras tem efeito sobre a relação sinal/ ruído captado pelo interferograma. As normas da *American Society for Testing and Materials* (ASTM) recomendam o uso de no mínimo 32 varreduras (*scans*) para obtenção de espectros isentos de ruídos. Visando atender a normativa, foram realizadas no mínimo três leituras por amostra e 16 varreduras para cada leitura, obtendo-se um total de 48 *scans* por amostra (mínimo). Esse procedimento garantiu, durante a modelagem, o tratamento de quantidade suficiente de sinais para geração de informações confiáveis (ASTM, 2012).

As unidades definidas para obtenção dos espectros correspondeu à absorvância (eixo das ordenadas) e a faixa de leitura no infravermelho foi obtida por meio do número de onda em centímetro inverso (cm^{-1}) (eixo das abscissas).

3.2.4 Interpretação dos espectros

A interpretação do perfil espectral das amostras foi realizada para promoção dos ajustes durante o pré-tratamento, identificação dos principais grupos funcionais presentes na amostra e para melhor compreensão dos resultados dos modelos.

Os grupos funcionais utilizados para interpretação da análise espectral (espectroscópica) foram: C=O (1800 a 1650 cm^{-1}), OH (3500 a 3200 cm^{-1}), banda larga para OH de ácidos (3400 a 2700 cm^{-1}), C-O (1300 a 100 cm^{-1}), N-H (3400 a 3200 cm^{-1}), C-H alifáticos saturados (3000 a 2800 cm^{-1}), $\equiv\text{C-H}$ insaturados (3100 a 3000 cm^{-1}), $\text{C}\equiv\text{C}$ aromáticos (1650 a 1450 cm^{-1}), entre outros (PAVIA *et al.*, 2015).

De acordo com Santos *et al.* (2013) as bandas de absorção que caracterizam o espectro do leite são: O-H das hidroxilas, N-H na forma associada, C-H das carbonilas, N-H das amidas I e II, C-O dos ésteres de lipídeos, O=P-O dos fosfatos ligados às proteínas e o grupo C-O-C da lactose

Para identificar os componentes presentes no leite e nas substâncias adulterantes, utilizou-se também os dados da Tabela 07. A tabela apresenta as regiões de absorvância dos grupos funcionais químicos presentes no leite bovino (CARVALHO, 2013).

3.2.5 Obtenção do espectro de fundo

Antes de iniciar o procedimento de obtenção dos espectros das amostras é necessário fazer um espectro de fundo (*background*). Esse espectro foi obtido sem que nenhuma amostra fosse colocada no compartimento de ATR. Esse procedimento permitiu que fossem detectados sinais referentes à umidade ou gás carbônico presentes no caminho óptico e também de outras interferências provenientes do funcionamento do equipamento.

O espectro de fundo foi automaticamente subtraído pelo *software* do equipamento dos espectros das amostras analisadas para garantir que os sinais finais obtidos fossem isentos de interferências. O modo de aquisição do espectro de fundo foi realizado por meio do comando “*modo single beam*” e o de obtenção do espectro da amostra pelo comando “*modo ratio*”, por meio do *software* do equipamento.

3.3 Modelagem

O MATLAB assim como o *PLS-Toobox* foram utilizados como ferramentas para construção dos modelos quimiométricos. Duas versões diferentes foram utilizadas na confecção dos modelos:

- MATLAB, versão 6.1 (*The MathWorks*, Natick, MA, USA) utilizando-se pacote de ferramentas contidas no *PLS-Toolbox*, versão 3.5 (*Eigenvector Technologies*, Manson, WA, USA);

- MATLAB, versão 7.1 (*The MathWorks*, Natick, MA, USA) utilizando-se pacote de ferramentas contidas no *PLS-Toolbox*, versão 6.1 (*Eigenvector Technologies*, Manson, WA, USA).

As diferentes versões do MATLAB não afetaram os resultados da pesquisa, uma vez que, os modelos de cada experimento, foram obtidos com o mesmo *software*. Os dados espectrais obtidos por meio do *software* do equipamento infravermelho puderam e podem ser utilizados em diferentes programas de modelagem multivariada, inclusive de diferentes versões. Cada modelo construído por meio desse trabalho independe um do outro.

3.3.1 Pré-tratamentos dos espectros para construção dos modelos

Os pré-tratamentos utilizados nos modelos neste trabalho foram:

- Centralização dos dados na média: para centralizar os dados na média o *software* MATLAB calculou automaticamente a média das intensidades de absorvância para cada número de onda (cm^{-1}) e subtraiu cada uma das intensidades do valor médio obtido para o grupo de espectros. Com a centralização dos dados, cada variável passou a ter média zero, ou seja, as coordenadas foram movidas para o centro dos dados permitindo que diferenças nas intensidades relativas das variáveis fossem mais facilmente percebidas. A centralização na média permitiu ainda que efeitos de variações de intensidade devido a sinais instrumentais fossem removidos. A técnica foi aplicada com base na Equação 6. A centralização pela média foi aplicada nos modelos preliminares PCA e HCA e nos modelos PLS.

- Suavização por *Savitzky-Golay*: o método permitiu recalcular o ponto central de intervalos que contém determinado número de variáveis (janelas), auxiliando no alinhamento ou suavização (*smoothing*) dos sinais. O algoritmo requereu a seleção do tamanho da janela (largura filtro) assim como da ordem polinomial (GODINHO, 2014). Foram utilizados polinômio de segundo grau e janela de 21 pontos. Esse procedimento matemático é representado pela Equação 07 e foi aplicado nos modelos preliminares PLS-DA das adulterações com leite e soro de queijos.

- Correção das linhas de base: para cada amostra o valor do menor ponto do espectro foi subtraído de todas as variáveis, fazendo com que, o menor valor de absorvância do conjunto final fosse zero e os demais valores positivos. Esse procedimento denominado de *Baseline* consistiu na aplicação de um algoritmo polinomial que removeu os sinais negativos não pertencentes à amostra (Equação 08) (SANTANA, 2015). A correção da linha de base foi aplicada em todos os modelos preliminares (PCA, HCA e PLS-DA das adulterações com leite e soro de queijos) e também nos modelos PLS e PLS-DA de todos os soros lácteos.

- Correção multiplicativa de sinal: os espectros foram corrigidos por meio da subtração dos valores originais dos coeficientes lineares das respectivas retas de regressão e divididos pelos coeficientes angulares dessas mesmas retas. A correção multiplicativa de sinal possibilitou a simplificação do modelo reduzindo o número de dados e proporcionou melhor linearidade ao modelo (HELLAND *et al.*, 1995; HONORATO *et al.*, 2007).

- Correção ortogonal de sinal: foi aplicada à matriz X (matriz de respostas instrumentais) visando reduzir a variação em X que não estava relacionada a Y (matriz dos parâmetros a serem preditos). Esse pré-tratamento evitou a perda de informações durante o pré-processamento dos espectros por meio da remoção de vetores ortogonais a Y e comuns aos equipamentos envolvidos (BYLESJO *et al.*, 2006; PIOVEZAN, 2014).

Especificamente nos modelos PLS-DA aplicou-se como pré-tratamento a correção ortogonal do sinal após terem sido comparados os resultados dos modelos, com e sem, a utilização da correção ortogonal de sinal (WESTERHUIS *et al.*, 2010).

3.3.2 Obtenção dos modelos

Os dados numéricos que expressavam as características de cada espectro foram obtidos por meio do *software* do equipamento de infravermelho e expostos sob forma de duas colunas: uma contendo o número de ondas (cm^{-1}) e outra contendo os valores de absorvância (%). Operacionalmente, a matriz de dados para calibração no MATLAB foi confeccionada apenas com os valores de absorvância. Uma vez que os valores de absorvância (variáveis) de cada amostra foram exportadas do *software* do equipamento sob forma de coluna (posição vertical), fez-se necessária a transposição dos dados para que estes passassem a ser dispostos na horizontal. Tal procedimento permitiu que os valores de absorvâncias fossem correlacionados no eixo do x , a um número de onda ou (comprimento de onda). Esse procedimento foi realizado no programa EXCEL e transferido para o MATLAB.

Posteriormente, acionou-se na plataforma operacional do MATLAB um gerenciador de dados denominado de *dataset* por meio do qual, se teve acesso ao ícone de caracterização das amostras ou parâmetros desejados (*row labels*). Em seguida, outras ferramentas do *PLS-Toobox* foram acionadas para realizar o pré-tratamento dos espectros visando a eliminação de variações sistemáticas e aleatórias.

De acordo com a norma ASTM 1655 (2012), 30 a 50 amostras são suficientes para fornecer dados sobre a propriedade de interesse por meio de uma análise multivariada. Todos os modelos de calibração e de treinamento do presente trabalho foram construídos com número de amostras superior a 40. Três a cinco leituras de

cada amostra compuseram os modelos de calibração e de treinamento, mas não foram contabilizadas como amostras.

Após a calibração os modelos foram submetidos à etapa de validação interna. A validação interna do modelo foi realizada acionando a ferramenta no *software cross-validation*. Cada amostra do conjunto de treinamento foi removida aleatoriamente e avaliada individualmente. Durante esse procedimento, novos modelos foram sendo construídos e os erros de validação interna foram armazenados para obtenção do erro total (RMSECV).

Os modelos preditivos PLS e PLS-DA requereram também conjunto de amostras para avaliar a capacidade preditiva do modelo (amostras de predição ou amostras teste). Para predição dos modelos, foram utilizadas amostras que não compuseram o conjunto de calibração (treinamento), e procedimento semelhante ao de validação cruzada foi realizado automaticamente pelo *software*. Para configuração do conjunto teste (previsão) adotou-se a proporção de um terço em relação ao número de amostras do conjunto de calibração. Em todas as etapas do modelo (calibração, validação e previsão) o número de amostras de leite fluido adulteradas com diferentes tipos de soros de origem láctea foram definidas conforme recomendações da ASTM.

3.3.3 Modelo PCA

O modelo exploratório de Análise de Componente Principal (PCA) foi obtido utilizando-se o *software* MATLAB e ferramentas do pacote *PLS-Toobox* para definição de uma regressão linear múltipla, cujo objetivo foi o de reduzir o espaço multidimensional para melhor visualização e compreensão do conjunto de dados. A redução do número de variáveis e obtenção das componentes principais basearam-se no cálculo da covariância dos dados por meio da Equação 12, capítulo II. A decomposição da matriz \mathbf{X} e a obtenção de uma matriz de resíduos \mathbf{E} foi obtida pela Equação 14, também demonstradas no capítulo II.

Gráficos contendo espectros das amostras dos diferentes tipos de soro lácteos utilizados nas adulterações de leite UHT foram gerados. A visualização dos espectros, sobretudo, das regiões de absorvâncias mais relevantes, possibilitaram a realização de análise visual interpretativa das características químicas de cada amostra.

Outros gráficos que permitiram compreender e avaliar os modelos foram gerados durante a modelagem: gráfico dos *scores* das amostras *versus* PCs e gráfico da variância explicada *versus* PCs.

3.3.4 Modelo HCA

O modelo de análise exploratória e classificatório HCA foi obtido utilizando-se o *software* MATLAB e ferramentas do pacote *PLS-Toobox*. A medida utilizada para medir a distância entre duas amostras *a* e *b* (D_{ab}) por meio das ferramentas do *PLS-Toobox*, no presente trabalho, correspondeu à Euclidiana, determinada analiticamente, pela Equação 16. Após definidas as distâncias entre todas as amostras foi obtida a escala de similaridade por meio do cálculo do vizinho mais próximo conforme Equação 17, capítulo II.

A modelagem da HCA resultou na obtenção de um histograma (dendrograma) que permitiu a identificação dos diferentes soros lácteos utilizados nas adulterações simuladas.

3.3.5 Modelo PLS

O modelo preditivo PLS foi obtido utilizando-se o *software* MATLAB e ferramentas do pacote *PLS-Toobox*, assim como os demais modelos. O modelo PLS é um modelo de regressão linear no qual o vetor da função resposta \mathbf{y} (dados sobre as concentrações) e a matriz \mathbf{X} (dados espectrais das variáveis das amostras) são decompostos em componentes principais (PC), que ao sofrerem leve rotação, passam a ser denominadas de variáveis latentes (VLs). Deste modo, as informações das medidas espectrais e das concentrações foram utilizadas simultaneamente na fase da calibração do modelo PLS. A Tabela com os valores de concentração do modelo PLS, encontram-se no Apêndice A.

Os passos para obtenção do modelo envolveram a decomposição da matriz \mathbf{X} e do vetor de concentração \mathbf{y} , conforme cálculo representado pela Equação 13 do modelo PCA. Posteriormente, algoritmo NILPALS (PLS1) foi aplicado nas matrizes decompostas, para definição de uma relação linear entre elas. Esse procedimento foi

realizado automaticamente pelas ferramentas do *PLS-Toobox* e foi representado matematicamente por meio das Equações 19 e 20 (capítulo II). Para encontrar o melhor ajuste possível entre \mathbf{X} e \mathbf{y} , pequenas rotações nas variáveis latentes foram automaticamente realizadas por meio do próprio *software PLS-Toobox*. Após o ajuste do modelo, os escores de \mathbf{y} (matriz \mathbf{U}) foram utilizados na previsão de novas amostras (conjunto de previsão do modelo PLS) a partir de \mathbf{X} (matriz \mathbf{T}).

Gráficos que auxiliaram na compreensão do modelo e expressaram sua eficiência, foram também obtidos: gráfico do número de VLs *versus* erros quadráticos médios; gráfico da variância explicada *versus* número de VLs; gráfico dos resíduos (erros) para cada amostra *versus* influência sobre o modelo e gráfico do ajuste do modelo de regressão linear.

Gráfico contendo o perfil espectral das amostras adulteradas com soro de queijo Minas Frescal nas diferentes concentrações foi obtido para possibilitar a interpretação das características químicas dessas amostras (PAVIA *et al.*, 2015). Também foi obtido perfil espectral do leite cru utilizado nas adulterações simuladas.

3.3.6 Modelo PLS-DA

Tanto os modelos PLS-DA preliminares referente às adulterações com soros de queijos e com leiteiros, assim como o PLS-DA que contemplou a junção de todos os tipos de soros lácteos utilizados nos testes preliminares, foram obtidos utilizando-se o *software* MATLAB e as ferramentas do pacote *PLS-Toobox*. Os modelos PLS-DA configuraram-se a partir das duas matrizes de dados do modelo necessárias para a construção do modelo: matriz \mathbf{X} referente aos dados espectrais e matriz \mathbf{Y} que correspondente às propriedades de interesse (classes).

Para obtenção dos modelos PLS-DA, inicialmente, as matrizes \mathbf{X} e \mathbf{Y} foram decompostas por meio das Equações 19 e 20 demonstradas. Nos modelos PLS-DA aplicou-se à matriz \mathbf{X} a correção ortogonal se sinal. Posteriormente, foram calculados automaticamente pelo *software* as variáveis latentes e as variâncias acumuladas em \mathbf{X} e em \mathbf{Y} . Equações 26 a 31 corresponderam aos procedimentos matemáticos por meio dos quais se obtiveram os dados básicos do modelo. Posteriormente aplicou-se PLS2 para classificação das amostras. Os resultados do modelo também foram expressos

por gráficos que auxiliaram na interpretação dos resultados: gráfico de classificação das amostras e gráfico dos *Vip-scores*.

Gráficos apresentando o perfil espectral de todas as amostras envolvidas no desenvolvimento dos modelos (leite UHT integral, leite adulterado com diferentes tipos de soros lácteos de queijos e manteigas) foram obtidos e analisados em relação à presença de determinados grupos funcionais químicos (SANTOS *et al.*, 2015).

3.3.7 Detecção de amostras anômalas

Com base nas Equações 9 e 10 o *software* calculou os resíduos de Student e a alavanca (influência sobre o modelo) de cada amostra e expressou esses valores por meio dos gráficos de *leverage*. A análise do gráfico dos resíduos de Student versus *leverage* para cada amostra correspondeu, à maneira pela qual, foram determinadas as amostras anômalas (FERREIRA *et al.*, 1999).

3.3.8 Interpretação dos resultados

Os resultados dos modelos foram obtidos por meio de diferentes parâmetros gerados pelo *software* MATLAB e das ferramentas do *PLS-Toobox*. Para cada diferente modelo construído, aplicou-se uma ou mais formas de avaliação dos resultados, conforme o objetivo da investigação.

Nos modelos PCA e HCA os resultados foram expressos pela definição do número de componentes principais em função da maior variância explicada pelo modelo e pela representação gráfica do dendrograma.

Os resultados dos testes preliminares do modelo PLS-DA para o conjunto de amostras adulteradas com soros de queijos e para o conjunto de amostras adulteradas com leitelhos foram obtidos pela definição dos números de variáveis latentes *versus* variâncias explicadas pelos modelos; pela análise dos gráficos de classificação das amostras e pelos valores de sensibilidade e seletividade definidos para cada classe dos modelos.

Para o modelo PLS os resultados envolveram a obtenção do número de variáveis latentes *versus* variância explicada pelo modelo e dos erros RMSEC, RMSECV e RMSEP por meio das Equações 23 e 24. Além dos parâmetros de erros, o valor do coeficiente de correlação R também foi obtido por meio do gráfico de correlação linear do modelo. Outra informação importante correspondeu à presença ou ausência de amostras anômalas. A presença de amostras anômalas foi analisada por meio do gráfico de resíduos *versus leverage*.

Os modelos PLS-DAs contendo amostras dos diferentes tipos de soros lácteos testados como adulterantes de leite UHT integral tiveram os resultados expressos pela definição do número de variáveis latentes *versus* a variância explicada pelo modelo; pelos gráficos de classificação de todas as amostras; pelos gráficos VIP-scores e pelas taxas de sensibilidade (percentagem de amostras positivas classificadas corretamente sobre o total de amostras positivas) e seletividade (percentagem de amostras negativas classificadas corretamente sobre o total de amostras negativas) (BOTELHO *et al.*, 2015).

3.3.9 Validação dos modelos

As figuras de mérito utilizadas no modelo PLS neste trabalho foram: veracidade, linearidade, seletividade, sensibilidade, erro sistemático (desvio padrão, grau de liberdade, $t_{calculado}$ e $t_{crítico}$). As ferramentas do *PLS-Toobox* oportunizaram a obtenção dos resultados de forma simples e rápida, uma vez que, calcularam automaticamente as figuras de mérito por meio do acionamento de comandos específicos. As figuras de mérito foram calculadas com base nas Equações 32 (sinal analítico), 23 e 24 (veracidade), 33 (sensibilidade), 34 (sensibilidade analítica), 35 (seletividade), 36 (Bias), 37 (desvio padrão), 38 (t Bias) e 39 (linearidade).

A validação analítica do método proposto foi realizada por meio do cálculo de parâmetros denominados de "figuras de mérito" usando sinal analítico líquido (NAS) (FERRÉ *et al.*, 1997; LORBER *et al.*, 1997).

A veracidade que corresponde ao grau de concordância entre o valor de referência e o valor previsto utilizando o modelo PLS foi determinada por meio do cálculo dos erros RMSEC, RMSECV e pelo erro quadrático médio de previsão (RMSEP) (SOUZA *et al.*, 2014; VALDERRAMA *et al.*, 2009; VIM, 2012).

A linearidade corresponde à relação entre os valores de concentração do soro de queijo Minas Frescal previstos pelo modelo PLS e os valores reais de concentração do soro para cada amostra. A linearidade do modelo que envolveu regressão linear (PLS) foi obtida por meio da correlação entre os valores reais e previstos, a partir da predição (SOUZA *et al.*, 2014; GONTIJO *et al.*, 2014).

O valor do NAS foi aplicado ao cálculo da seletividade. Esse parâmetro correspondeu à determinação da quantidade de sinal instrumental que foi utilizado pelo modelo para a determinação da concentração do soro. A sensibilidade, definida como a fração de sinal responsável para o aumento de uma unidade de concentração de soro também foi obtida por meio da Equação 33 (VALDERRAMA, BRAGA E POPPI, 2009).

A sensibilidade analítica calculada por meio da Equação 34 correspondeu à relação entre a sensibilidade e o desvio padrão estimada do ruído instrumental considerado como a única fonte de erro. A sensibilidade analítica foi calculada e obtida por meio de nove espectros do leite cru idênticos, isento de soro de queijo Minas Frescal (THOMPSON *et al.*, 2006). A seletividade (Eq. 34) propiciou a definição da faixa de concentração possível de ser detectada pelo modelo

O valor de "bias" utilizado para investigar a presença de erros sistemáticos no modelo PLS foi calculado a partir da diferença entre os valores médios e de referência, correspondendo aos erros que não são aleatórios. O desvio padrão (SDV) de validação que caracteriza a dispersão dos valores previstos de adulteração pelo conjunto de validação foi calculado pela Equação 37, correspondente ao desvio padrão dos erros de validação (SILVA *et al.*, 2012; SOUZA *et al.*, 2014).

A investigação dos erros sistemáticos foi realizada utilizando um teste t (Equação 38) com amostras de validação, com 95% de confiança e graus de liberdade igual ao número de amostras de previsão (24).

Especificamente para o modelo PLS-DA construído com todos os espectros dos modelos PLS-DA preliminares, no qual a variável dependente Y assume valores binários, a validação do modelo discriminante foi realizada por meio dos cálculos dos parâmetros: sensibilidade (*Recall*), seletividade, acórdância e F-score (BOTELHO, 2014; CAMPELO, 2015; DAVIS E GOADRICH, 2006).

Todos os parâmetros foram calculados por meio do acionamento das ferramentas do *PLS-Toobox*. As Equações matemáticas por meio das quais foram calculadas essas figuras de mérito encontram-se na Tabela 9.

3.4 Modelos Preliminares

3.4.1 Teste para avaliar existência de diferenças entre os soros de origem láctea por meio de modelos exploratórios: PCA e HCA

Para verificar, inicialmente, a existência de diferenças significativas entre soros de origem láctea, análises espectroscópicas foram realizadas no laboratório de biocombustíveis da Universidade Federal de Uberlândia (LABIC-UFU) em infravermelho médio da marca *Perkin Elmer* dotado de transformada de Fourier acessório de reflectância atenuada.

Amostras contendo um litro de cada tipo de soro de origem láctea foram coletadas e utilizadas em adulterações artificiais de leite UHT integral. Os diferentes soros utilizados na pesquisa foram: soro de queijo confeccionado artesanalmente com leite cru; soro de queijo elaborado com leite pasteurizado (ambas cedidas pela agroindústria experimental de leite e derivados do Instituto Federal de Ciência e Tecnologia do Triângulo Mineiro - IFTM); soro de manteiga de primeira qualidade (leitelho), soro de queijo Minas Frescal, soro de queijo Ricota e soro de Muçarela (as quatro últimas amostras foram doadas por laticínios localizado em Uberlândia, MG).

Leite integral UHT (único fabricante) foi adquirido em supermercado da cidade de Uberlândia para composição das amostras adulteradas artificialmente com os soros lácteos testados.

As amostras dos diferentes soros lácteos foram estocadas à temperatura inferior a 5,0°C por um período de 10 dias e analisadas em intervalos de tempo de 48 horas (cada amostra). As repetidas leituras espectroscópicas das amostras armazenadas visaram verificar se as características que distinguiram cada soro lácteo seriam alteradas em função da estocagem. As seis amostras de soros foram analisadas 5 vezes ao longo de 10 dias, totalizando 30 amostras. Em cada análise realizada, 5% (p/p) (0,25 g) de soro foram adicionados em 5 g de leite cru. Análises espectroscópicas do leite puro e de cada soro, também foram realizadas com o objetivo de se conhecer o perfil espectral das duas substâncias que estavam sendo misturadas.

Obeve-se inicialmente, o espectro de fundo e posteriormente, os espectros das 30 amostras. Cada amostra foi colocada sem preparo prévio diretamente no compartimento de ATR. Foram realizadas três leituras por amostra com 16 varreduras

para cada leitura. A faixa de absorção utilizada para varredura das amostras foi de 4000 a 600 cm^{-1} e a resolução do equipamento foi definida para 4 cm^{-1} . Posteriormente, os dados espectrais foram utilizados para confecção dos modelos exploratórios PCA e HCA.

Os modelos PCA e HCA foram obtidos por meio do MATLAB, versão 6.1 (*The MathWorks*, Natick, MA, USA) utilizando pacote de ferramentas contidas no *PLS-Toolbox*, versão 3.5 (*Eigenvector Technologies*, Manson, WA, USA). Para obtenção dos modelos acionou-se as ferramentas do *PLS-Toolbox*, após ter sido gerado o *dataset*. Os espectros originais foram pré-tratados com a remoção de ruídos visíveis na região de absorção próxima a 600 cm^{-1} , correção da linha de base (*baseline*) e com os dados centrados na média.

Os resultados dos modelos exploratórios PCA e HCA foram também apresentados no capítulo IV.

3.4.2 Teste para avaliar a possibilidade de classificação de soros de queijos em adulterações de leite por meio do modelo preditivo PLS-DA

Nas adulterações simuladas para detecção e classificação de amostras de soro de queijos que podem ser utilizadas em fraudes de leites fluidos, foram adquiridos em supermercados da cidade de Goiânia, GO., leites UHT integral provenientes de três diferentes fábricas de laticínios. Lotes de cada fabricante foram analisados previamente para confirmação da qualidade do leite (pré-teste das amostras A, B e C). Após essa avaliação, o leite UHT integral foi adulterado com soros de queijos Muçarela (M), Prato (K), Minas Frescal (MF), de Coalho (C) e Minas Padrão (MP). As amostras dos diferentes soros lácteos foram cedidas por laticínio localizado em Bela Vista, GO.

As concentrações de soro de queijo utilizadas nas adulterações simuladas do leite UHT integral corresponderam a 5, 15, 25 e 35%. Utilizou-se balança analítica para realização das adulterações (p/p), tendo sido as amostras adulteradas, colocadas em frascos âmbar, devidamente tampadas e resfriados (5°C) até a realização da análise no FTIR-MIR. A definição das concentrações de soro de queijos para adulteração artificial do leite UHT teve por base os estudos desenvolvidos por Cortez *et al.* (2010) e no fato desse problema estar relacionado a uma fraude econômica. Quantidades muito baixas de soro no leite pouco afetariam o acréscimo no volume para proporcionar

rentabilidade, por outro lado, quantidades muito grandes comprometeriam as características sensoriais do leite.

Foram realizadas cinco leituras por amostra em espectrofotômetro FTIR-MIR da marca *Perkin Elmer*, dotado de um acessório de reflexão total atenuada horizontal (ATR) de Seleneto de Zinco (ZnSe) com as seguintes configurações: varredura 4000 a 600 cm^{-1} , resolução 4 cm^{-1} e 16 varreduras por espectro. Foram analisadas as regiões de absorção mais relevantes para modelo e a faixa definida para obtenção de resultados mais significativos foi de 1550 a 1800 cm^{-1} . O equipamento utilizado pertence laboratório Central de Análise Multiusuário (CAM) da UFG, localizado em Goiânia.

Na Tabela 10 são apresentados os valores percentuais das concentrações utilizadas nas adulterações de leite UHT dos diferentes fabricantes para construção de modelo quimiométrico classificatório (PLS-DA).

TABELA 10: Concentrações de soro de queijo em leite UHT(amostras A, B e C)

Soros queijos	Fábrica A (%)	Fábrica B (%)	Fábrica C (%)
M	5 - 15 - 25 - 35	5 - 15 - 25 - 35	5 - 15 - 25 - 35
K	5 - 15 - 25 - 35	5 - 15 - 25 - 35	5 - 15 - 25 - 35
MF	5 - 15 - 25 - 35	5 - 15 - 25 - 35	5 - 15 - 25 - 35
C	5 - 15 - 25 - 35	5 - 15 - 25 - 35	5 - 15 - 25 - 35
MP	5 - 15 - 25 - 35	5 - 15 - 25 - 35	5 - 15 - 25 - 35

Obteve-se o espectro de fundo e, posteriormente, os espectros das amostras. Do total de 60 amostras, 40 foram destinadas para o conjunto de treinamento e 20 para o conjunto teste. Os espectros obtidos correspondentes às cinco leituras de cada amostra, não foram contabilizados como amostras (300 espectros). O número de espectros utilizados para rodar o modelo contribuiriam para a pureza do sinal e precisão dos resultados.

Realizou-se o pré-tratamento dos dados espectrais por meio da suavização Savitzky-Golay, correção da linha de base e correção ortogonal de sinal. O tratamento quimiométrico para obtenção do modelo PLS-DA foi realizado utilizando *software* MATLAB (*The Mathworks, Natick, USA*), versão 7.1 e ferramentas do *PLS-Toolbox* (*Eigenvector Research Inc*) versão 6.0.1.

Aplicou-se algoritmo PLS2 para classificação das amostras similares que compuseram cada classe. Os resultados do modelo de classificação do leite adulterado com diferentes soros de queijos foram apresentados no capítulo IV, juntamente com os resultados dos demais testes preliminares.

3.4.3 Teste para avaliar a possibilidade de classificação dos soros de manteiga em leite fluido por meio do modelo preditivo PLS-DA

Para detecção e classificação de amostras adulteradas com soros de manteigas (leitelho), foram utilizadas três amostras de leite UHT integral de diferentes lotes e fabricantes adquiridas em supermercados de Goiânia, GO, analisadas previamente para confirmação da qualidade (teste preliminar das amostras A, B e C).

Após essa avaliação, as amostras de leite UHT integral foram adulteradas com leitelhos provenientes de manteiga extra, elaborada com creme diluído com leite desnatado (MA); de manteiga extra elaborada com creme diluído com água (MB); de manteiga comum elaborada com creme de maior acidez diluído com água (MC), e de manteiga comum elaborada com creme de menor acidez (MD). Um litro de cada tipo de leitelho foram cedidas por dois laticínios situados em região próxima à cidade de Goiânia, GO.

As concentrações de leitelho utilizadas nas adulterações do leite UHT integral corresponderam a 5, 15, 25 e 35%. Utilizou-se balança analítica para realização das adulterações (p/p). As devidas proporções de leite e o leitelho foram colocados em pequenos frascos âmbar, devidamente tampados e resfriados (5°C) até a realização da análise no equipamento de infravermelho médio. Assim como para as adulterações com soros de queijos, a definição das concentrações de leitelho no leite teve por base, os estudos desenvolvidos por Cortez *et al.* (2010) e o comprometimento sensorial do leite decorrente da adição de percentual elevado de soro.

A Tabela 11 apresenta valores percentuais das concentrações utilizadas nas adulterações das diferentes fábricas de leite utilizadas para construção do modelo quimiométrico classificatório.

TABELA 11: Concentrações de leitelhos em leite UHT (amostras A, B e C)

Leitelho	Fábrica A (%)	Fábrica B (%)	Fábrica C (%)
MA	5 - 15 - 25 - 35	5 - 15 - 25 - 35	5 - 15 - 25 - 35
MB	5 - 15 - 25 - 35	5 - 15 - 25 - 35	5 - 15 - 25 - 35
MC	5 - 15 - 25 - 35	5 - 15 - 25 - 35	5 - 15 - 25 - 35
MD	5 - 15 - 25 - 35	5 - 15 - 25 - 35	5 - 15 - 25 - 35

Obteve-se o espectro de fundo e posteriormente, obteve-se cinco leituras por amostra em espectrofotômetro FTIR-MIR da marca *Perkin Elmer*, dotado de um acessório de reflexão total atenuada horizontal (ATR) de Seleneto de Zinco (ZnSe) com as seguintes configurações: varredura 4000 a 600 cm^{-1} , resolução 4 cm^{-1} e 16 varreduras por espectro. As análises foram realizadas no laboratório Central de Análise Multiusuário (CAM) da UFG, em Goiânia. As amostras foram colocadas diretamente no compartimento do acessório de ATR, uma a uma, sem terem sido submetidas a nenhum tipo de tratamento prévio.

Os espectros das amostras foram pré-tratados usando suavização por Savitzky-Golay, correção da linha de base e correção ortogonal de sinal. O tratamento quimiométrico para obtenção do modelo PLS-DA foi realizado por meio do *software* MATLAB (*The Mathworks, Natick, USA*), versão 7.1 e *software* PLS-Toolbox (*Eigenvector Research Inc*) versão 6.1.

Foram separadas oito amostras de cada classe para o conjunto de treinamento, somando um total de 32 amostras (dois terços do total de amostras) e para o conjunto teste foram selecionadas 4 amostras de cada classe, somando 16 amostras (um terço do total de amostras). Foram obtidas 5 leituras de cada amostra (240 espectros).

Inicialmente, obteve-se o espectro de fundo e realizou-se a calibração do equipamento utilizando dados de absorção provenientes de todo o espectro (4000-600 cm^{-1}). Posteriormente, o modelo foi testado utilizando faixa de absorção de 1550 cm^{-1} a 1800 cm^{-1} . Esta última faixa, apresentou melhores resultados para modelagem.

Modelos exploratórios de Análise de Componentes Principais (PCA) foram obtidos para faixa espectral mais significativa. Posteriormente, foram rodados dois modelos preliminares PLS-DA para amostras adulteradas com leitelho: um contendo apenas amostras de leitelho e outro contendo amostras de leitelho e de leite puro. No segundo modelo, três amostras de leite UHT integral, sem adulterante, passaram a compor o conjunto de calibração do modelo PLS-DA.

Aplicou-se algoritmo PLS2 para classificação das amostras de leite UHT adulteradas com leiteiro. Os resultados do modelo foram apresentados no capítulo IV.

3.5 Modelo preditivo PLS para detecção de adulteração de leite cru com soro em leite fluido

Para detecção e quantificação de soro de queijo em leite cru, amostras de leite provenientes de rebanho da região do em torno de Uberlândia (sudeste brasileiro) foram coletadas e cedidas pelo setor de produção do Instituto Federal do Triângulo Mineiro (IFTM), durante o período de duas semanas. A cada dois dias foram coletados drante a ordenha da manhã leite de diferentes animais que compunham o rebanho.

A obtenção do leite de animais saudáveis e a coleta realizada diretamente no local da ordenha corresponderam a procedimentos que visaram garantir a integridade (pureza) do leite utilizado na pesquisa.

Amostra de soro de queijo Minas Frescal elaborado artesanalmente, também foi obtida junto à agroindústria experimental do IFTM. Optou-se pela utilização do soro de queijo Minas Frescal para adulteração do leite cru, por se tratar de soro doce, cujo pH se assemelha ao do leite e em cujo processo de elaboração, não foram adicionadas outras substâncias, além do coalho.

Um total de 67 amostras de leite coletadas ao longo de duas semanas foi adulterada com concentrações entre 6,0 e 28% (p/p) de soro de queijo fresco utilizando balança de precisão da marca SARTORIUS. As amostras de leite cru adulteradas com soro de queijo Minas Frescal, após terem sido colocadas em frascos âmbar e tampadas, foram homogeneizados e mantidas sob refrigeração a 5,0 ° C por um período máximo de 36 horas (ISO 707, IDF 50 2008). A Figura 18 corresponde ao conjunto de amostras utilizadas.

As amostras foram colocadas, uma a uma, no compartimento do acessório de reflectância atenuada sem preparo prévio, após obtenção do espectro de fundo. Dados das concentrações de todas as amostras foram devidamente registrados para obtenção do modelo qualitativo. Para cada amostra adulterada com soro de queijo Minas Frescal, foram realizadas três leituras com 16 varreduras para cada leitura.



FIGURA 18: Amostras de leite cru adulteradas com soro de queijo Minas Frescal

Fonte: Própria autora

Os espectros foram obtidos no FTIR-MIR da marca *Perkin Elmer* com as seguintes configurações: varredura 4000 a 600 cm^{-1} , resolução 4 cm^{-1} e 16 varreduras por espectro. As análises foram realizadas no laboratório de biocombustíveis (LABIC-UFU), na cidade de Uberlândia, MG.

Os dados espectrais foram pré-tratados pelo *software* MATLAB aplicando-se correção da linha de base nas regiões 3750–3100 cm^{-1} e 2600–1900 cm^{-1} . Regiões de ruído (sinais não significativos) foram também eliminadas nas duas extremidades dos espectros, resultando numa região para modelagem, correspondente a 3750 cm^{-1} - 760 cm^{-1} .

Para calibração, matriz com dados de absorção de todas as amostras e suas triplicadas foi construída no programa EXCEL do Windows 2008 e salva no *software* MATLAB (*The Mathworks, Natick, USA*), versão 6.1, para gerar o *dataset*. Obtido o gerenciador de dados, acionou-se o comando das ferramentas do pacote *PLS-Toolbox* (Eigenvector Research Inc) versão 3.5 para obtenção do modelo PLS.

O número de amostras dos conjuntos de calibração e validação utilizados para construção do modelo PLS foi determinado de acordo com ASTM E1655-05, que especifica ser o número de variáveis latentes (k) o indicador do número mínimo de amostras para o conjunto de treinamento e de teste. De acordo com a norma o número de amostras do conjunto de calibração deve ser igual a 6 ($k+1$) para dados centrados na média e o número de amostras para o conjunto teste deve ser igual a 4 k (ASTM E1655-05, 2012). Do total de 67 amostras, 43 foram destinadas ao conjunto de calibração e 24 ao conjunto de validação. Durante modelagem do PLS, após acionamento dos comandos para aplicação dos pré-tratamentos, dados referentes às

concentrações das adulterações com soro de queijo Minas Frescal, foram digitalizados em janela específica do *software* para configuração da matriz **Y** do modelo.

O número de variáveis latentes foi definido por meio da relação linear que melhor explicou a correlação entre as variâncias das matrizes **X** (absorção) e **Y** (concentração adulterante) levando ainda em consideração, o valor correspondente ao erro quadrático médio de validação cruzada (RMSECV).

A presença de outlier foi avaliada por meio da observação dos valores de alavancagem e resíduos Q a 95% de confiança (SILVA *et al.*, 2012; VALDERRAMA *et al.*, 2009). Os resultados para o modelo PLS foram apresentados no capítulo V.

3.6 Modelo preditivo PLS-DA para classificação de adulterações com soro de queijos e leitelhos

Para aumentar a possibilidade de classificação de amostras adulteradas com maior diversidade de soros de origem láctea, as amostras de leite adulteradas com diferentes leitelhos e as amostras de leite adulteradas com diferentes soros de queijos, foram utilizadas para compor um único modelo de classificação. Uma grande matriz de dados foi construída no MATLAB contendo 120 amostras (48 amostras adulteradas com leitelho, 60 amostras adulteradas com soros de queijos e 12 amostras de leite puro). Foram utilizados para confecção do modelo, três espectros de cada amostra, totalizando 360 espectros.

Todos os espectros dessas amostras foram obtidos por meio de leitura direta em espectrofotômetro FTIR-MIR da marca *Perkin Elmer*, dotado de um acessório de reflexão total atenuada horizontal (HATR) de Seleneto de Zinco (ZnSe) com as seguintes configurações: varredura 4000 a 600 cm^{-1} , resolução 4 cm^{-1} e 16 varreduras por espectro, no laboratório Central de Análise Multiusuário (CAM) da UFG, localizado em Goiânia, GO.

O conjunto de treinamento foi composto por 80 amostras e conjunto teste por 40 amostras. Os espectros utilizados nesse modelo foram os mesmos utilizados nos modelos PLS-DA preliminares. As alterações relacionadas à modelagem conjuntamente de todas as amostras de leite UHT integral adulteradas com os diferentes soros lácteos, correspondeu à não aplicação do pré-tratamento de dados por

Savitzky-Golay e a ampliação da faixa de absorção. Utilizou-se inicialmente neste modelo toda a faixa espectral do infravermelho médio, visando capturar o máximo de informações. Após observação dos resultados iniciais do modelo, optou-se pela delimitação da faixa de absorção para 4000 cm^{-1} a 1000 cm^{-1} devido à necessidade da remoção de sinais caracterizados como ruídos. Os espectros das amostras foram pré-tratados usando correção multiplicativa de sinal e correção ortogonal de sinal.

Aplicou-se algoritmo PLS2 para classificação das amostras de leite UHT adulteradas com os diferentes soros lácteos.

O tratamento quimiométrico para obtenção do modelo PLS-DA foi realizado por meio do *software* MATLAB (*The Mathworks, Natick, USA*), versão 7.1 e *software* PLS-Toolbox (*Eigenvector Research Inc*) versão 6.1.

Os resultados do modelo foram apresentados no capítulo IV.

4 RESULTADOS E DISCUSSÃO DOS TESTES E MODELOS PRELIMINARES

4.1 Teste de crioscopia

Foram realizadas análises crioscópicas em leite UHT, adulterados com soros lácteos, com o objetivo de avaliar se a técnica seria capaz de detectar este tipo de fraude. Os resultados dos valores médios de crioscopia para diferentes teores de adulteração com soro de queijo e de manteiga foram apresentados na Tabela 12.

TABELA 12: Crioscopia (H°) das diferentes concentrações de adulterações de leite UHT integral

Teor de adulteração (%)	Valores de crioscopia (H°)	
	Soro de queijo	Soro de manteiga
0%	-0,544	-0,543
1%	-0,543	-0,539
5%	-0,546	-0,514
15%	-0,548	-0,483
25%	-0,541	-0,450
35%	-0,539	-0,422
Desvio Padrão	0,003	0,049

O valor da crioscopia do leite puro utilizado nas adulterações com soro de queijo foi de $-0,544^{\circ}\text{H}$. Este valor se encontra dentro da faixa de referência para leite fluido de $-0,530$ a $-0,550^{\circ}\text{H}$ (BRASIL, 2006^a).

Os resultados obtidos na análise de crioscopia para leite UHT adulterado com diferentes teores com soro de queijo também apresentaram valores dentro da faixa de referência para leite fluido não adulterado. Constatou-se que a análise de crioscopia não foi eficiente para detecção de fraudes com soro de queijo na faixa das adulterações testadas.

O valor da crioscopia do leite puro utilizado nas adulterações com soro de manteiga (leitelho) correspondeu a $-0,543^{\circ}\text{H}$. O valor médio de crioscopia obtido para

as adulterações artificiais com leiteiro foi de 0,492 com desvio padrão igual a 0,049. A adição de 1% de leiteiro promoveu pequeno decréscimo no valor da crioscopia em relação ao leite puro, porém, o valor encontrado não ultrapassou os parâmetros normativos para leites isentos de adulteração. Portanto, para adulterações de leite com adição de 1% de leiteiro a análise de crioscopia seria insuficiente para detectar a fraude. Para os demais teores de adulterações foram obtidos valores de crioscopia decrescentes e fora da faixa de normalidade.

É importante salientar que, o leiteiro oriundo do processo de fabricação da manteiga não corresponde ao leiteiro puro, cujo percentual de água na sua composição seria em torno de 90,4% (GILLE, 2011). No processo de elaboração de manteigas, tanto comum como extra, o creme é diluído com aproximadamente 38% de água. Essa água adicionada é eliminada juntamente com o leiteiro ocasionando aumento significativo no teor de água no leiteiro industrial. Essa característica justifica a maior alteração nos valores da crioscopia do leite adulterado com leiteiro em relação às pequenas variações proporcionadas pelas adulterações com soro de queijo.

Observou-se que, para fraudes em leite decorrente da adição de leiteiros obtidos por meio da bateção de cremes diluídos, a análise de crioscopia seria eficiente para detectar a presença do adulterante em concentrações de 5 a 35%, porém, a técnica não seria capaz de detectar fraudes com soro de queijo em nenhuma das concentrações testadas.

4.2 Teste da qualidade do leite

4.2.1 Resultados de densidade e acidez

Os resultados médios das análises de densidade e potencial hidrogeniônico (pH), obtidos para a avaliação prévia da qualidade dos leites UHT integral utilizados nas adulterações com diferentes tipos de soros lácteos, constam na Tabela 13.

A densidade média dos lotes dos três fabricantes analisados apresentaram o mesmo valor de 1,029 g/mL. O valor médio encontrado para densidade dos lotes utilizados na pesquisa se enquadra dentro dos limites normativos de referência para densidade corrigida do leite (1,023–1,040 g/mL) (BRASIL, 2006^a).

TABELA 13: Densidade e pH amostras A,B e C de UHT integral

Amostras	*Densidade (g/mL)	pH	Temperatura (°C)
A	1,029	6,7	25
B	1,029	6,5	25
C	1,029	6,6	25
Desvio padrão	-	0,1	-

Os valores médios de pH dos lotes das fábricas A (6,7), B (6,5) e C (6,6) também encontraram-se dentro dos limites normativos de referência correspondente ao pH 6,6 à uma temperatura de 25°C (BRASIL, 2006^a).

4.2.1 Resultados referentes ao perfil espectral

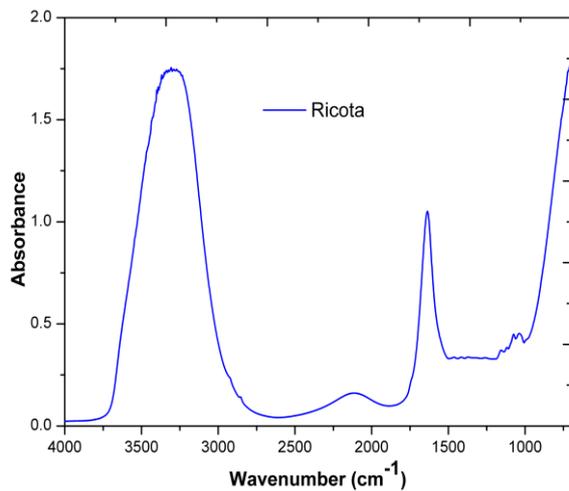
Foram obtidos espectros para os três lotes de cada um dos três fabricantes de leite utilizados nas adulterações em laboratório. Os dados para cada amostra foram compostos por 3351 valores de absorção na região do infravermelho. O cálculo da média das absorvâncias das nove amostras foi de 0,492 e o desvio padrão dos dados das amostras, correspondeu a 1,226.

Observou-se mesmo perfil espectral para todos os lotes de leite dos diferentes fabricante (sobreposição dos espectros). Consequentemente, inferiu-se que as amostras de leite UHT integral adquiridas, por apresentarem o mesmo perfil espectral e possuírem características físico-químicas dentro dos padrões, encontravam-se aptas para serem utilizadas nas adulterações artificiais da pesquisa.

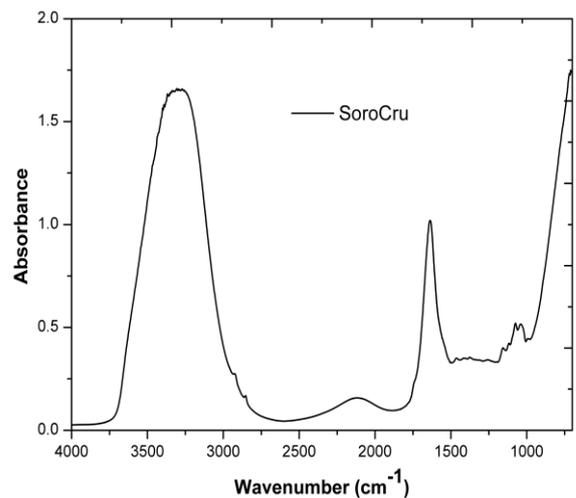
4.3 Modelo para avaliar a existência de diferenças entre soros de origem láctea: modelo exploratório PCA e HCA

4.3.1 Resultados referentes ao perfil espectral

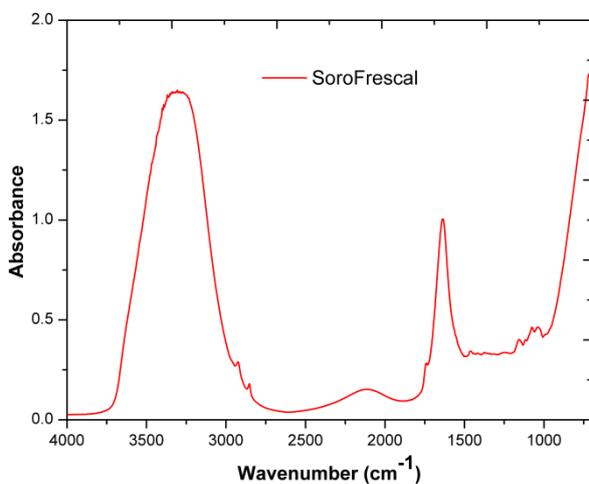
Importantes informações sobre as regiões de absorção promovidas pelos grupos funcionais presentes nas amostras auxiliaram a diferenciação e caracterização dos soros de origem láctea por meio da observação dos perfis espectrais das amostras. O perfil espectral de cada uma das amostras de leite adulteradas com os soros lácteos testados foi apresentado na Figura 19.



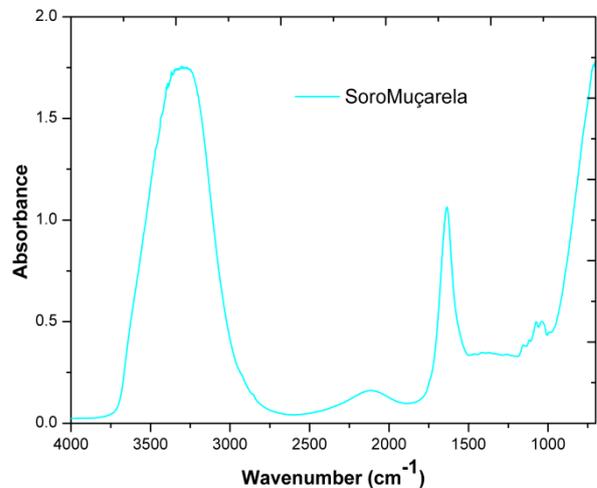
Soro de Ricota



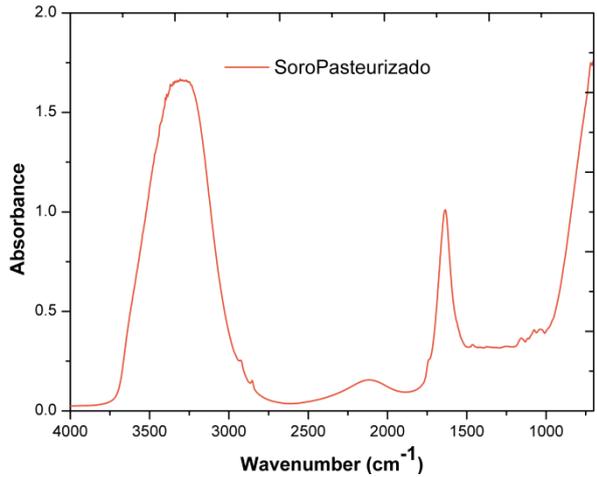
Soro de queijo elaborado com leite cru



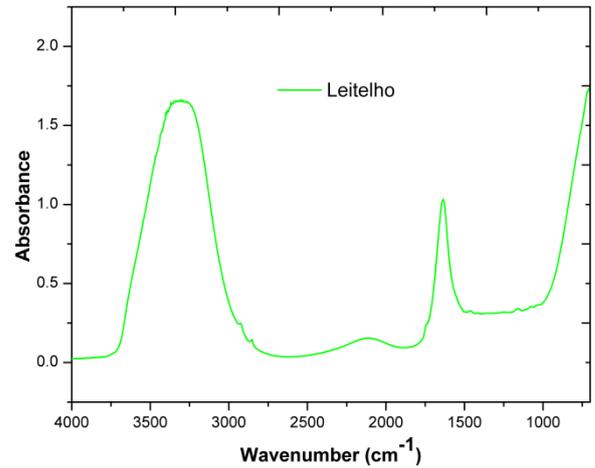
Soro de queijo Minas Frescal



Soro de queijo Muçarela



Soro de queijo elaborado com leite
pasteurizado



Leitelho

FIGURA 19: Espectros das amostras adulteradas com cada tipo de soro lácteo utilizados na construção do modelo PCA e HCA

Por meio da Figura 20 pôde-se observar o conjunto de todos os espectros (triplicatas das trinta amostras adulteradas artificialmente) dos diferentes soros lácteos utilizados em adulterações de leite UHT integral. Determinadas regiões de absorção do espectro sofreram alterações mais significativas em detrimento a outras regiões.

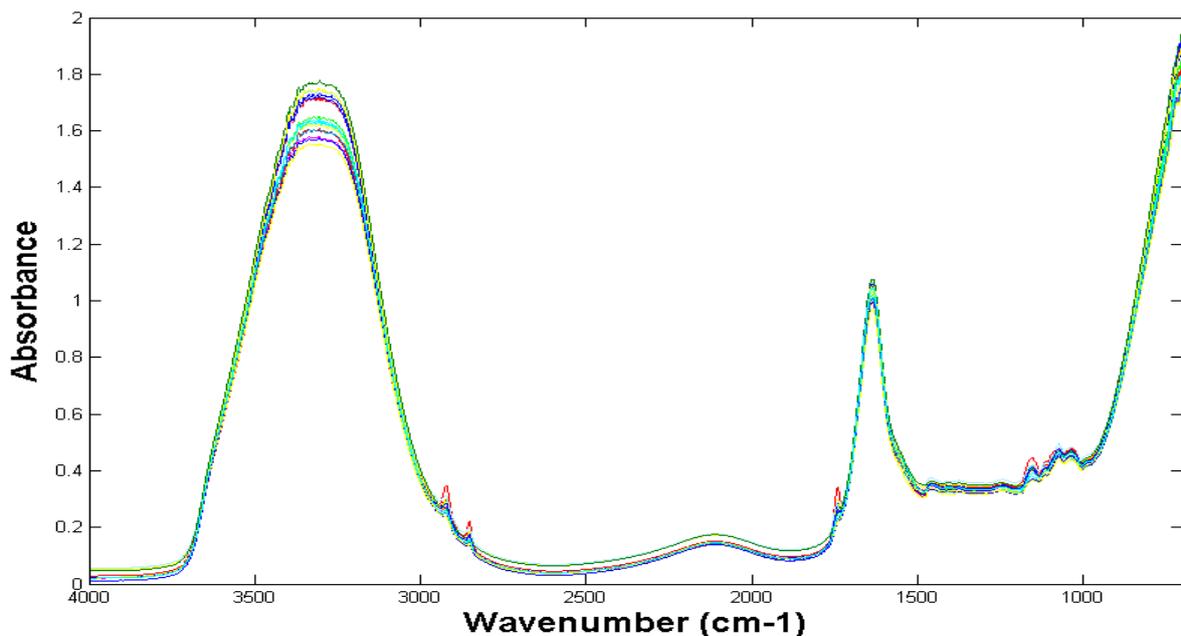


FIGURA 20: Espectros das amostras de leite adulteradas artificialmente com soros lácteos do modelo PCA e HCA

Foi confeccionado gráfico contendo um espectro de cada tipo de soro utilizado nas adulterações realizadas em leite UHT, a fim de facilitar a interpretação e visualização das características espectrais de cada tipo de soro utilizado nas adulterações artificiais (Figura 21).

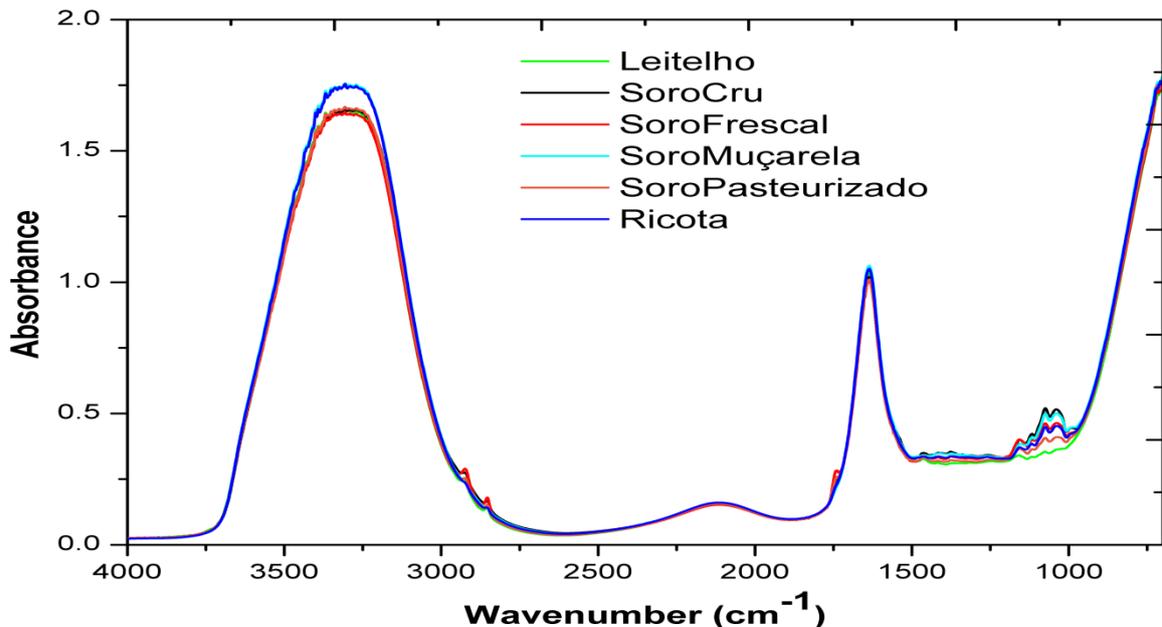


FIGURA 21: Conjunto contendo um espectro de cada tipo de soro adulterante no modelo PCA e HCA

4.3.1.1 Características dos espectros na região de 920 a 1500 cm^{-1}

A faixa de absorção de 920 a 1180 cm^{-1} apresentada na Figura 22 está relacionada à região de absorção de íons orgânicos dotados de ligações covalentes e também da lactose. Segundo Stuart (2005), os seguintes íons inorgânicos apresentam bandas de absorção no infravermelho: SO_4^{2-} (1130-1080 cm^{-1} , 880-800 cm^{-1}); PO_4^{3-} (1100-950 cm^{-1}); SiO_4^{2-} (1100-900 cm^{-1}) e NH_4^+ (1485-1390 cm^{-1}). Infere-se que pequenas bandas de absorção atribuídas aos sais citados Stuart (2005), estavam presentes nas amostras analisadas devido estarem associados naturalmente no leite a estruturas protéicas e lipoprotéicas (ORDONEZ, 2005).

Barbosa (2007) também faz referência à alguns íons inorgânicos dotados de ligações covalentes que apresentam absorção no infravermelho: H_2PO_4^- (forte absorção em 1080 – 1020 cm^{-1}), HPO_4^- (forte absorção em 1080 – 1020 cm^{-1}), SO_4^{2-} (forte absorção em 1150 – 1060 cm^{-1}) e MnO_4^- (forte absorção em 920 – 880 cm^{-1}).

Segundo Etzion *et al.* (2004), a região de absorção entre 1060 e 1100 cm^{-1} corresponde às vibrações das ligações covalentes dos fosfatos das micelas de caseína. Durante o processo de elaboração de queijos fatores ligados ao processamento, tais como a presença de enzimas específicas (coalho), alterações no pH e na temperatura, promovem modificações na estrutura das proteínas (caseínas), sobretudo, em relação aos sais originalmente presentes em sua estrutura. As concentrações de sais minerais variam de acordo com o pH do meio, conforme apresentado na Tabela 14.

TABELA 14: Composição dos principais minerais do soro doce e do soro ácido

Minerais	Soro doce (mg/100g)	Soro ácido (mg/100g)
Cálcio	36,500	92,800
Magnésio	6,500	9,000
Sódio	45,500	39,800
Potássio	123,000	153,000
Fósforo	43,000	58,000

Fonte: 1988.

Os maiores valores de absorvância atribuídos às regiões de absorção dos sais foram obtidos para o soro de Muçarela e para o soro proveniente da fabricação de queijo com leite cru. Esses dois soros, em geral, apresentam menores valores de pH (maior acidez) em relação aos demais. Durante o processamento de Muçarela adiciona-se fermento láctico para reduzir o pH do meio (valor próximo a 5,4) para obtenção da elasticidade na massa necessária para o processo de filagem. O leite cru normalmente possui uma elevada carga microbiana responsável pela produção de ácido láctico e aumento da acidez. Alterações no pH aumentam a solubilidade do fosfato de cálcio, permitindo a migração de elementos minerais, sobretudo de fosfatos, para o meio aquoso (ORDONEZ, 2005).

Pôde-se observar que a amostra de leitelho apresentou perfil espectral diferente das demais devido à menor absorvância nessa região. Na composição do leitelho e dos soros de queijos estão presentes teores semelhantes de gordura, lactose e sais minerais, sendo que a maior variação na composição ocorre em função do teor e do tipo de proteínas presentes. As proteínas presentes no soro de queijo correspondem às proteínas solúveis, enquanto as proteínas do leitelho correspondem às proteínas insolúveis em meio aquoso (caseínas). A composição protéica diferenciada do leitelho

e o tipo de sais associados a ela justificaram a diferenciação em relação aos demais soros oriundos de queijos e não de manteigas (ORDONEZ, 2005).

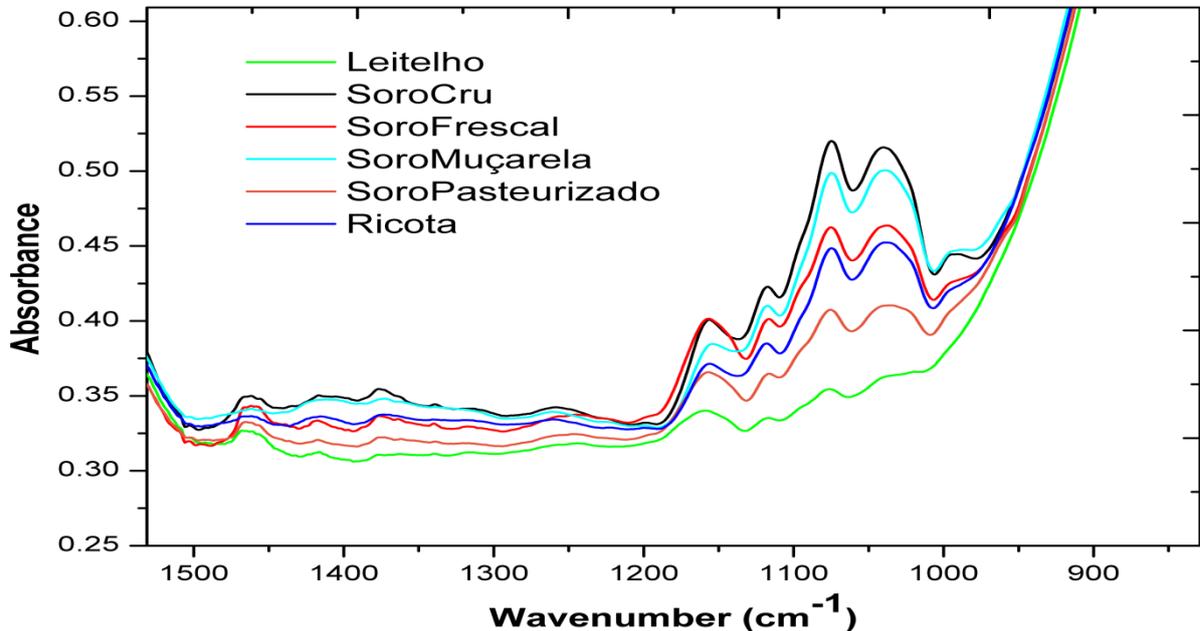


FIGURA 22: Absorção das amostras na região de 900 a 1500 cm^{-1}

A faixa de absorção da Figura 22 também apresentou bandas relacionadas à presença da lactose. A absorção da lactose ocorre entre 1030 a 1150 cm^{-1} devido à vibrações de estiramento das ligações C-O, C-C e C-O-C, responsáveis pela formação de carboidratos (AERNOOTS *et al.*, 2011; WU *et al.*, 2008). Nicolaou *et al.* (2010) pesquisando adulterações em leite bovino com outros tipos de leite constatou absorvância atribuída à lactose na região entre 1076 e 1159 cm^{-1} , valor este, semelhante ao encontrado na presente pesquisa.

Por meio dos perfis espectrais apresentados, observou-se que o leiteiro apresentou baixa absorção em relação à lactose. Por outro lado, os soros de queijo apresentaram maior absorção que o leiteiro e menor que o leite. Durante a fabricação de queijos, pequena quantidade de lactose é retida na massa, sendo, a maior parte liberada no soro. Essa característica do processamento de queijos justifica o comportamento das amostras apresentado na Figura 22 (FURTADO, 1991).

Segundo Etzion *et al.* (2004) a região de absorção das amidas III corresponde à 1200-1280 cm^{-1} . Estudo para caracterização da composição de leite em pó de cabra e de leite bovino por meio do MIR, o autor também atribuiu a região de absorção de 1241 a 1251 cm^{-1} à presença de amidas III (ligação C-N) (CARVALHO, 2013)..

4.3.1.2 Características dos espectros na região entre 1600 a 1680 cm^{-1}

Faixas de absorção também relacionadas à absorção de proteínas foram constatadas por meio da Figura 23. De acordo com Yu e Irudayaraj (2005) a região de 1637–1695 cm^{-1} é atribuída à absorção da amida I. As bandas da amida I compreendem 80% do estiramento do grupo carbonila (C=O das ligações peptídicas), 10% de deformação angular –CNH e 10% de estiramento –CN. As amidas saturadas quando se encontram envolvidas em ligações de hidrogênio apresentam absorção na região correspondente a 1620-1650 cm^{-1} conforme consta na figura apresentada abaixo.

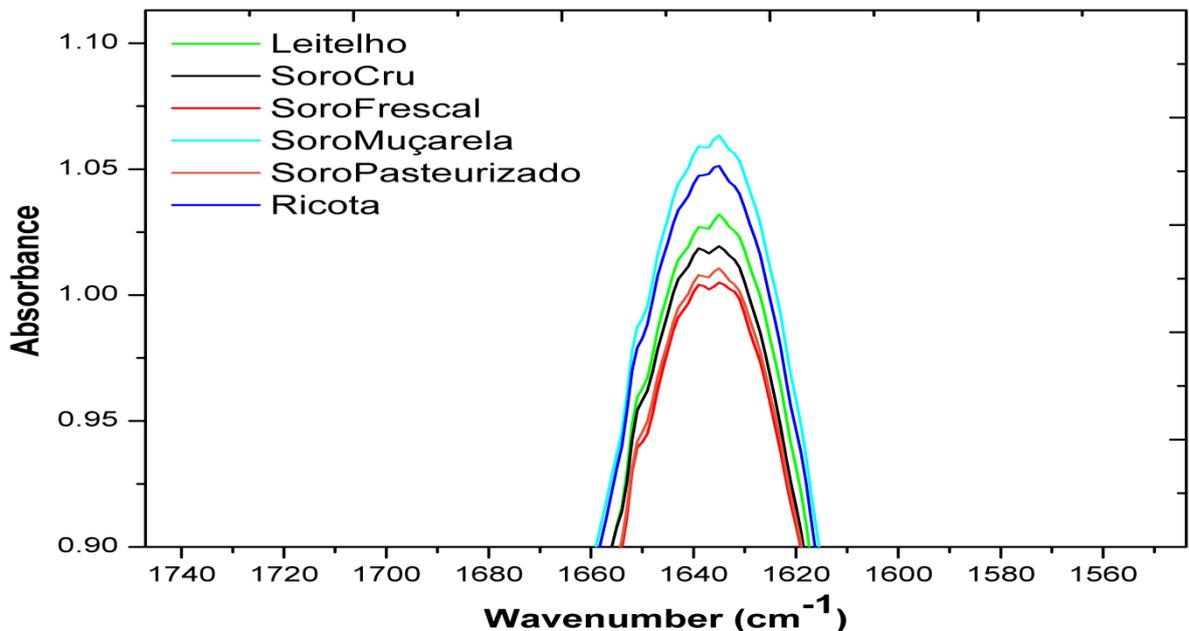


FIGURA 23: Absorção das amostras na região entre 1620 a 1650 cm^{-1}

Observou-se que todas as amostras apresentaram absorção nessa região, porém, algumas mais pronunciadas que outras. A região compreendida entre 1620 e 1640 cm^{-1} também corresponde à fortes bandas de dobramento N-H das amidas I em estado sólido (PAVIA *et al.*, 2015).

Os perfis espectrais dos soros nessa região demonstraram que a concentração de amidas I foram maiores no soro de Muçarela e no soro de Ricota, ambos submetidos a processo de aquecimento mais drástico.

4.3.1.3 Características dos espectros na região de 1720 a 1760 cm^{-1} e 2850 a 2940 cm^{-1}

A faixa de absorção correspondente a 1720-1760 cm^{-1} e as regiões de 2850 cm^{-1} e 2940 cm^{-1} são atribuídas à porção lipídica do leite composta pelos grupos carbonilas (C=O) das ligações éster dos triacilgliceróis e pelos grupos carbono-hidrogênio (C-H) (WOLFSCHOON, 1977). Segundo Barbosa (2007), a faixa de absorção de 1735 a 1750 cm^{-1} são características dos ésteres de ácidos graxos, atribuídas ao estiramento do C=O.

Os glóbulos de gordura possuem uma estrutura química complexa composta por diferentes ácidos graxos e diversas substâncias insaponificáveis. A fração lipídica predominante no leite corresponde aos triacilgliceróis (99%), cujas propriedades dos ácidos graxos promovem absorção em locais diferenciados na faixa do infravermelho (ARAUJO, 2011). Os principais ácidos graxos presentes na gordura do leite correspondem ao palmítico, oléico e esteárico. Entre estes, o ácido palmítico é o que está presente em maior quantidade e se caracteriza por possuir duas bandas nas regiões de absorção correspondentes a 2852 e 2920 cm^{-1} devido ao estiramento simétrico e assimétrico do C-H (CARVALHO, 2013).

Por meio das Figuras 24 e 25 pôde-se observar que, algumas amostras apresentaram menor absorvância nessas regiões, em relação às outras.

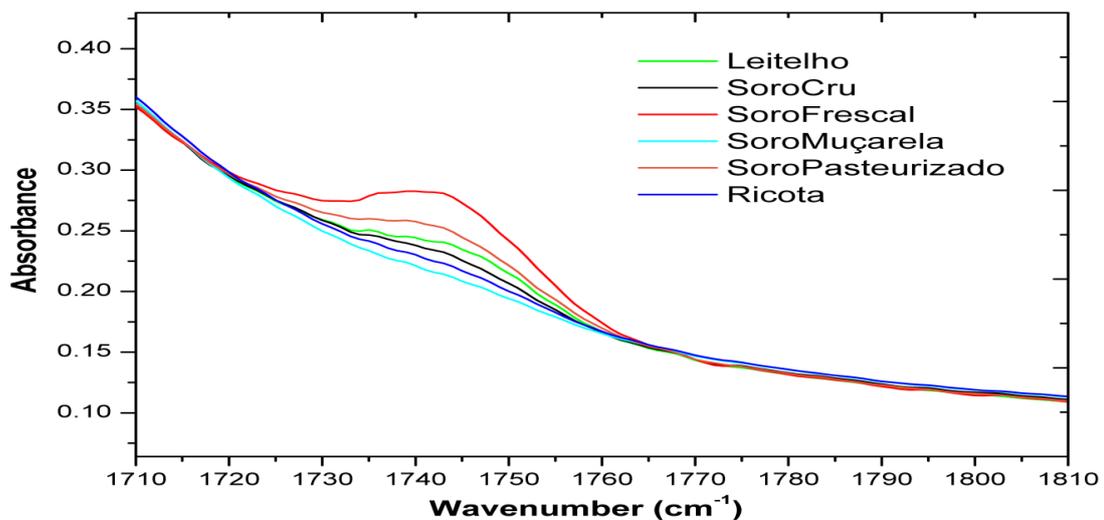


FIGURA 24: Absorção das amostras na região de 1720 a 1760 cm^{-1}

O leite possui diferentes triacilglicerídeos (ésteres de glicerol), além de outros tipos de lipídeos que compõem a gordura. O processamento de queijos, assim como de manteigas, podem ocasionar alterações nas estruturas lipídicas ocasionando liberação de diferentes compostos químicos, tanto de maior peso molecular como de menor peso molecular (componentes do glóbulo de gordura) que ocasionam diferentes absorções no infravermelho (ORDONEZ, 2005).

Os soros de Muçarela e de Ricota apresentaram bandas de absorção pouco pronunciadas nas regiões características de absorção promovidas pelos lipídeos. Essa característica indica baixa concentração de lipídeos nos soros desses queijos decorrente do processo de elaboração.

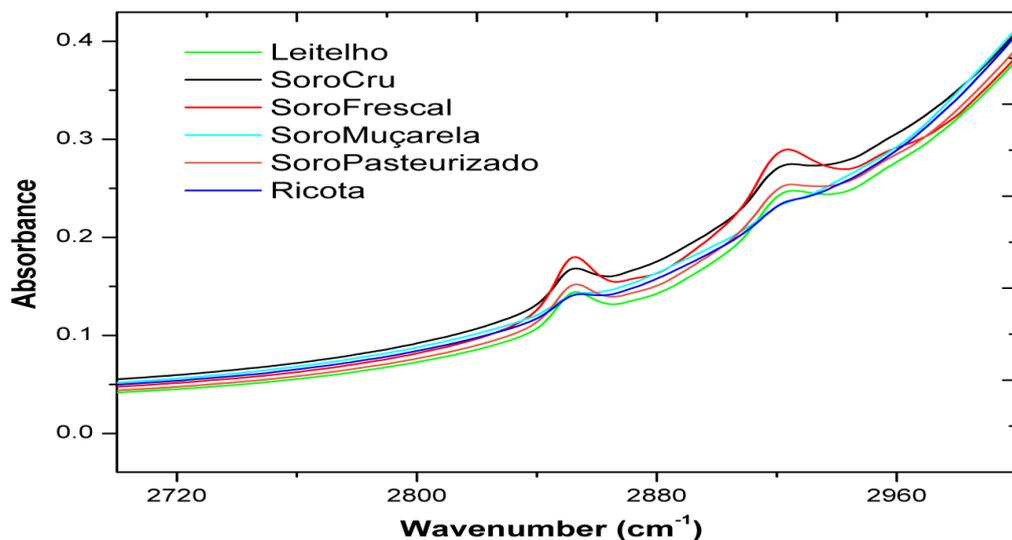


FIGURA 25: Absorção das amostras na região de 2850 a 2920 cm^{-1}

4.3.1.4 Características dos espectros na região de 3120 a 3460 cm^{-1}

Essa região correspondente à forte absorção da água no infravermelho e é apresentada na Figura 26.

Observou-se que houve uma diferença significativa na absorvância da amostra do soro de Muçarela e também do soro de Ricota nessa região de absorção da hidroxila (OH). Inferiu-se que as amostras que apresentaram maior absorção nessa região corresponderam às que continham maior teor de água.

A elaboração do queijo Muçarela envolve a adição de água para auxiliar no cozimento da massa durante o período de mexedura, fato que justifica a maior

absorção nessa região. Por outro lado, a fabricação da Ricota objetiva extrair do soro utilizado como matéria-prima proteínas solúveis, sobretudo das β -lactoglobulinas e α -lactoalbuminas, de grande parte da lactose, de minerais (cálcio, magnésio, fósforo, potássio, ferro e zinco), de lipídeo e de vitaminas hidrossolúveis (FURTADO E NETO, 1994; ORDONEZ, 2005). Conseqüentemente, o soro resultante do processamento da Ricota, possui baixo extrato seco e alto teor de água.

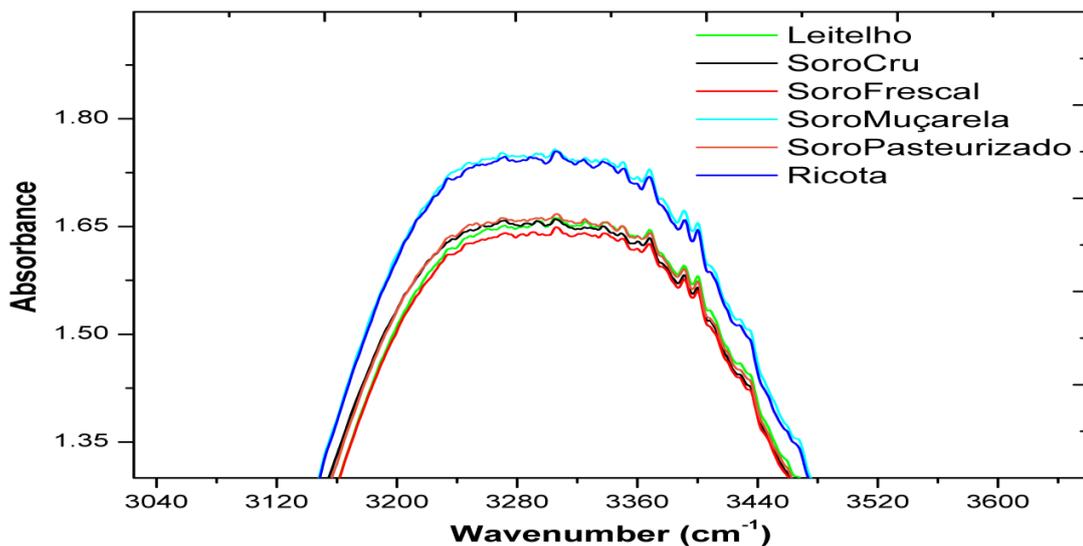


FIGURA 26: Absorção das amostras na região de 3160 a3500 cm^{-1}

Barbosa (2007) relata que absorções atribuídas à amida alifática saturada nas regiões de 3270-3370 cm^{-1} e 3440-3330 cm^{-1} são decorrentes respectivamente das formas *s-trans* e *s-cis* das vibrações do grupo NH. Também menciona absorção devido à vibração simétrica do NH_2 na forma não associada na região de 3420-3380 cm^{-1} e na forma associada na região de 3360-3320 cm^{-1} .

Os alcoóis associados por ligações de hidrogênio apresentam absorvância entre 3400 e 3200 cm^{-1} e podem, assim como os outros grupos citados anteriormente, serem responsáveis pelas pequenas bandas de absorção visualizadas nesta região.

4.3.2 Resultados referentes aos modelos PCA e HCA

Os seis grupos correspondentes a cada diferente tipo de soro lácteo foram adequadamente identificados pelo modelo exploratório conforme demonstra a Figura 27. Os grupos compostos pelas cinco amostras de cada tipo de soro lácteo,

configurados no espaço tridimensional, corresponderam a: amostras leitelho (1-5), amostras de soro de queijo cru (6-10), amostras de soro de Ricota (11-15), amostras de soro de leite pasteurizado (16-20), amostras de soro de queijo Minas Frescal (21-25) e amostras de soro de Muçarela (26-30).

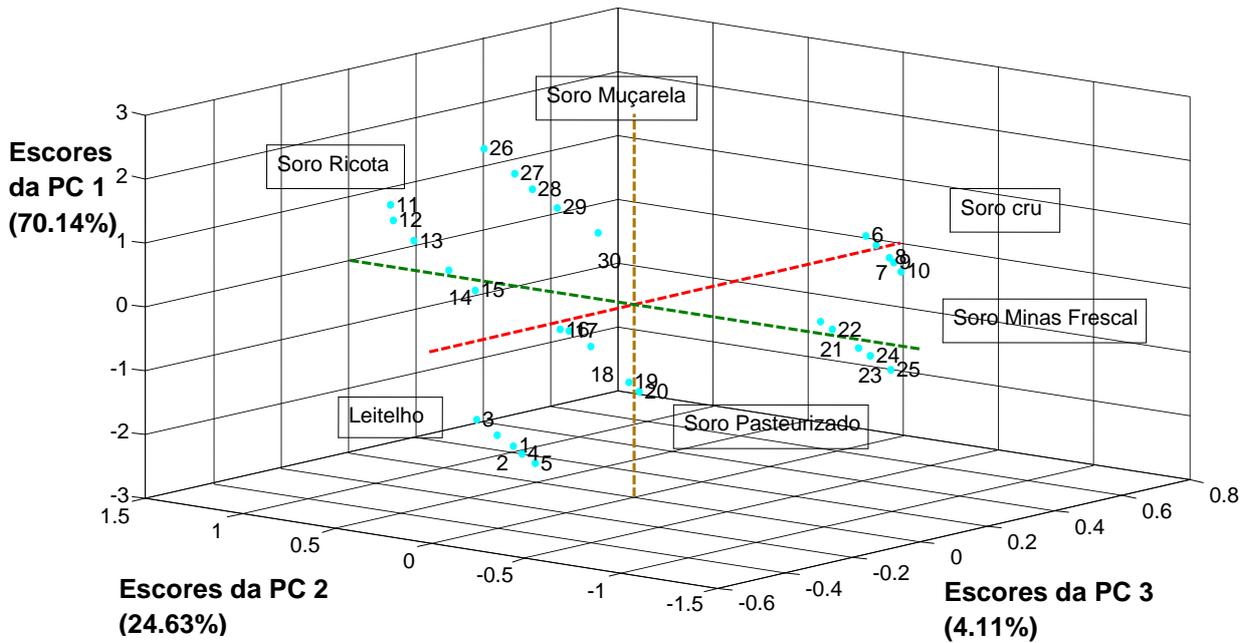


FIGURA 27: PCAs definidas pelo modelo e escores das amostras

O modelo PCA foi obtido por meio de três componentes principais que foram capazes de explicar 98,88% da variância dos dados. Na Figura 28, é apresentado o percentual de variância capturada acumulada em todas as PCs do modelo.

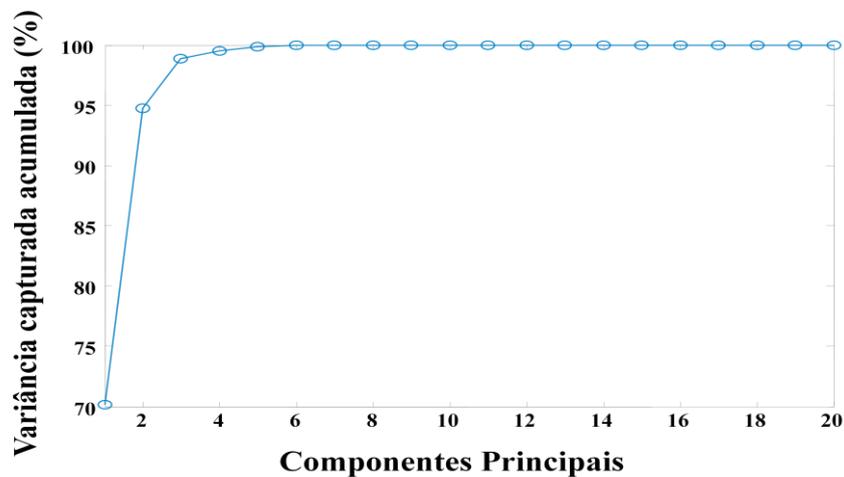


FIGURA 28: Variância acumulada em todos os PC (%)

A Figura 29 apresenta a variância percentual explicada pelas PCs: PC1, PC2 e PC3 correspondentes a 70,14%, 24,63% e 4,11% da variância total, respectivamente.

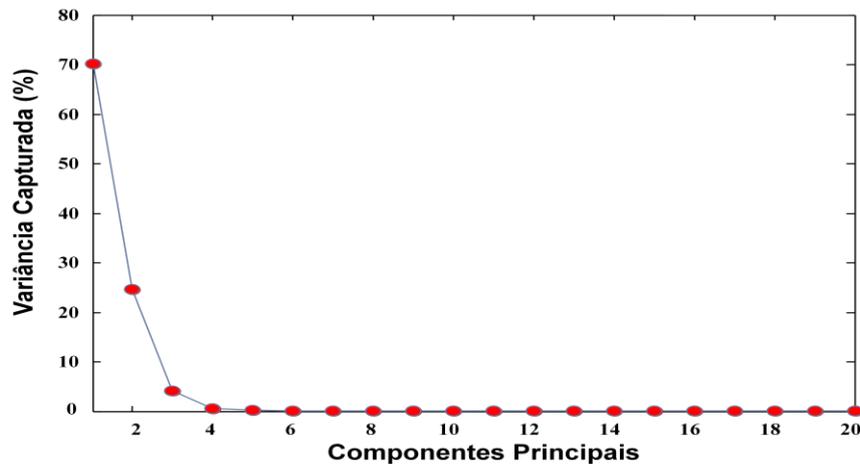


FIGURA 29: Variância explicada por todos os PC (%)

Após constatar diferença significativa entre os grupos de soros lácteos em um espaço tridimensional por meio da análise de componentes principais, aplicou-se modelo para análise de agrupamento hierárquico (HCA). O objetivo da modelagem com o HCA foi o de confirmar ou não a existência de diferenças significativas entre as amostras de soro.

A aplicação do modelo HCA resultou em um dendrograma no qual os grupos de soros lácteos foram agrupados em função da similaridade. O dendrograma apresentou os seis grupos distintos, assim como o PCA, conforme pode ser constatado na Figura 30.

Pôde-se observar que os grupos de amostras do leiteiro e do soro de queijo elaborado com leite cru se diferenciaram totalmente dos demais. Já os soros de queijo Minas Frescal e o soro de queijo elaborado com leite pasteurizado, apresentaram certa semelhança entre eles. Por meio da observação do dendrograma e das bandas de absorção nas Figuras 25 e 26, comprovou-se que os espectros do soro de Muçarela e de Ricota apresentaram características muito semelhantes.

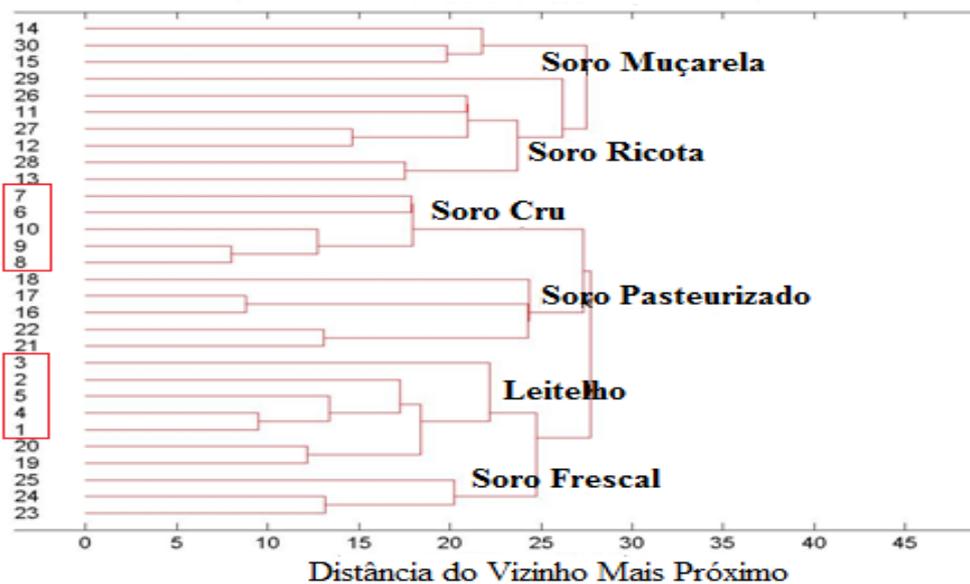


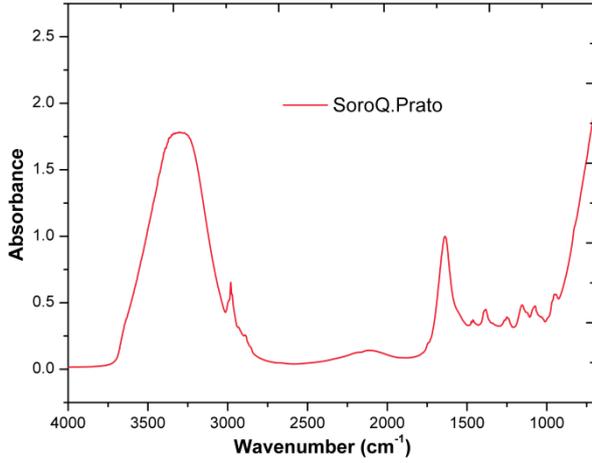
FIGURA 30: Dendrograma dos diferentes soros lácteos

Tanto o PCA como o HCA comprovaram poder ser utilizados como ferramentas para obtenção de informações sobre as características de amostras desconhecidas e a aplicação de ambos encontra-se bastante em pesquisas para reconhecimento de padrões. Os resultados desses dois modelos quimiométricos exploratórios demonstraram ser possível diferenciar soros provenientes do processamentos de diferentes derivados lácteos. Contatada a possibilidade de diferenciação entre os soros, modelo mais robusto para classificação como PLS-DA, foi testado.

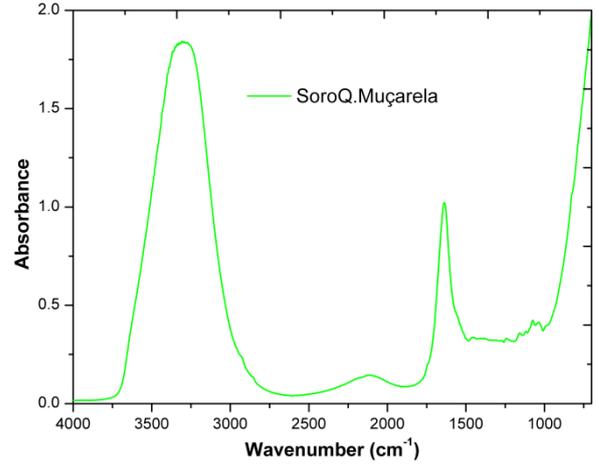
4.4 Modelo para avaliar a possibilidade de classificar soros de queijos em adulterações de leite UHT por meio do FTIR/MIR e modelo PLS-DA

4.4.1 Resultados referentes ao perfil espectral

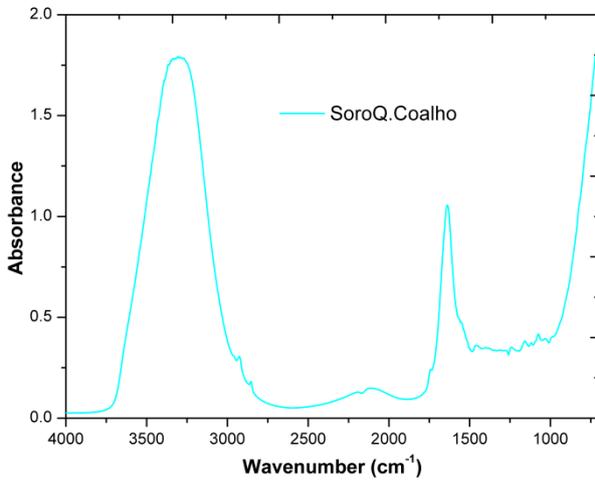
Espectros puros de diferentes tipos de soros de queijos foram obtidos e apresentados por meio da Figura 31 (gráficos A, B, C, E e F). A visualização do perfil espectral possibilitou verificar as diferenças nas composições dos soros de queijos utilizados em adulterações de leites fluidos. Essas diferenças foram posteriormente constatadas por meio das análises multivariadas.



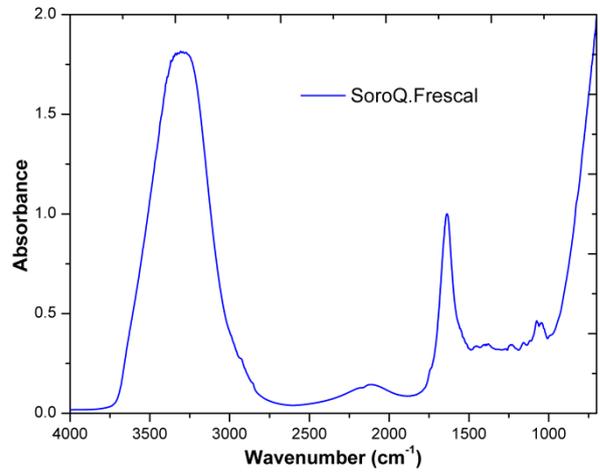
A - Soro de queijo Prato



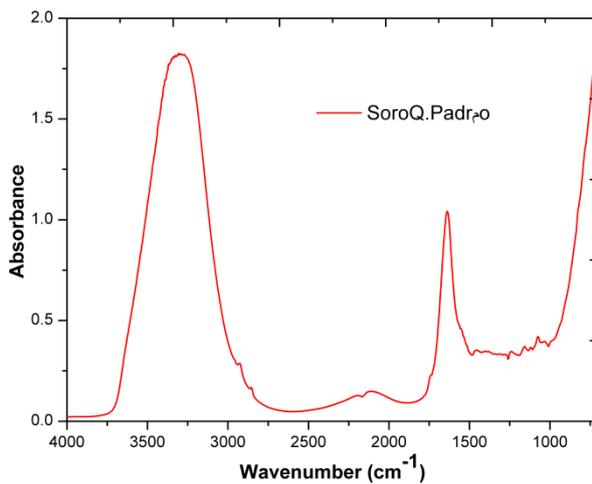
B - Soro de queijo Muçarela



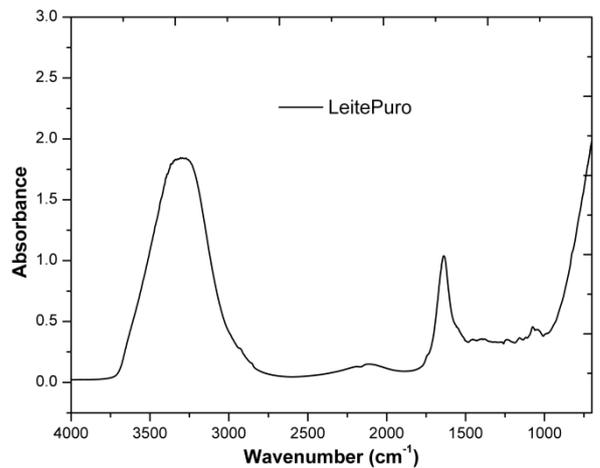
C - Soro de queijo Coalho



D - Soro de queijo Minas Frescal



E - Soro de queijo Minas Padrão



F - Leite puro

FIGURA 31: Espectros de amostras de soros de queijos e de leite (puros)

Entre as amostras de soros de queijos, a que mais se diferenciou das demais foi a amostra de queijo Prato.

Gráfico contendo os espectros de amostras de leite UHT adulteradas com 35% de cada um dos tipos de soro de queijos foi confeccionado para possibilitar visualização e comparação entre o perfil espectral de cada soro de queijo (Figura 32).

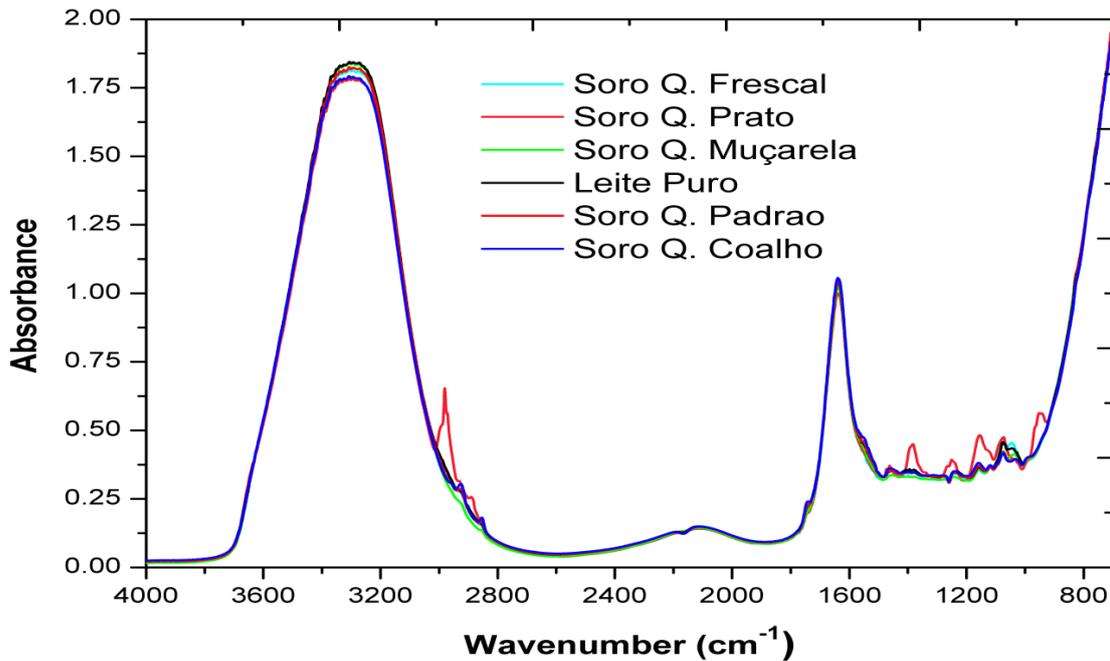


FIGURA 32: Conjunto de espectros das amostras de leite adulteradas com diferentes soros de queijos

Ampliação da região de absorvância correspondente a 900-1600 cm^{-1} foi realizada e apresentada na Figura 33. A observação dessa região espectral possibilitou identificar banda de absorção na região próxima a 900 cm^{-1} específica para o queijo Prato, sugerindo ser atribuída à adição um insumo utilizado apenas nesse tipo de queijo (corante) ou a uma peculiaridade do processamento (maturação).

A adição de corante de urucum no queijo Prato, assim como nas manteigas, é autorizada pela legislação brasileira e corresponde a uma prática convencional durante a fabricação desses derivados (LERAYER *et al.*, 2002; FURTADO E LOURENÇO NETO, 1994; MULVANY, 1970).

Todas as amostras apresentaram bandas de absorção devido à presença da lactose (1030-1150 cm^{-1}) correspondentes a vibrações C-O, C-C e C-O-O dos carboidratos. O teor de lactose que permanece na massa do queijo é inferior a 1%, por isso, a presença da lactose é significativa nos soros (CARVALHO *et al.*, 2007;

ORDONEZ, 2005). A amostra de soro de queijo Minas Frescal foi a que apresentou maior absorção nessa região indicando maior teor de lactose em relação às demais amostras.

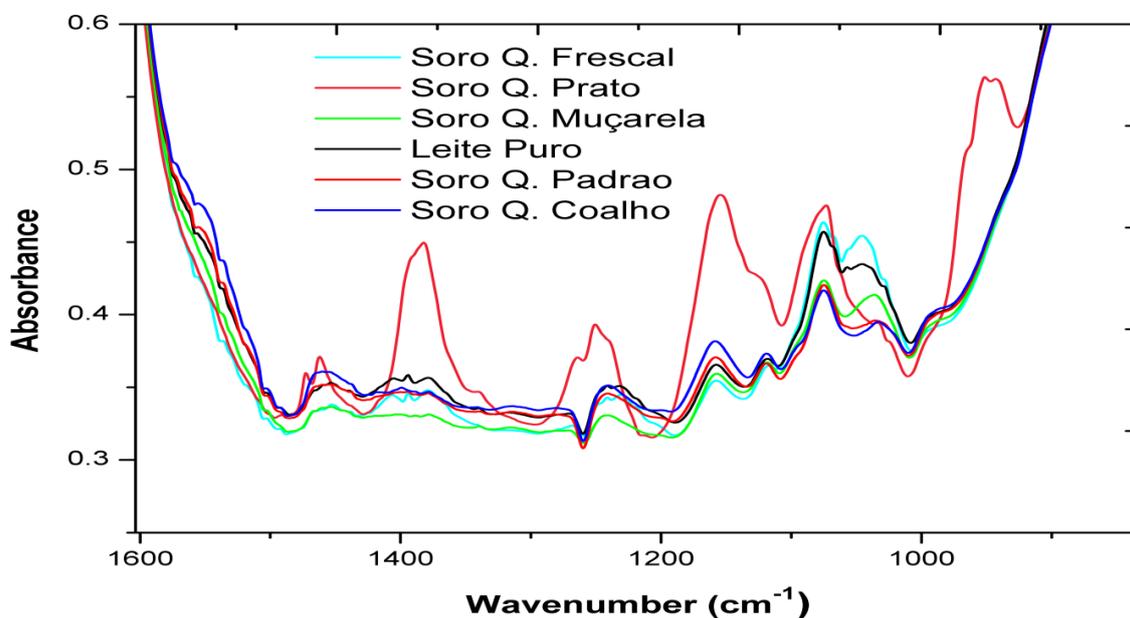


FIGURA 33: Bandas de absorção das amostras de leite adulteradas com soros de queijos na região de 900-1600 cm⁻¹

A banda de absorção atribuída à presença de grupos fosfatos (1060-1100 cm⁻¹) foram semelhantes para todos os soros lácteos e também para o leite. Infere-se que todo o fosfato presente no leite é carregado para o soro tanto de queijos como de manteiga.

As demais bandas que diferenciaram a amostra de queijo Prato foram atribuídas à alterações no anel benzeno nas ligações C=C (1630-1430 cm⁻¹) e C-C (1275-1000 cm⁻¹ e 900-690 cm⁻¹) e são consideradas promotoras de características aromáticas (STUART, 2005).

Por meio da Figura 34 observou-se pequena banda de absorvância na região de 1740 a 1735 cm⁻¹ relacionada com a absorção de lipídeos, mais especificamente, à absorção dos ésteres de cadeia aberta saturado (BARBOSA 2007). As amostras adulteradas com soro de queijo Minas Frescal e Minas Padrão apresentaram mesma absorção que o leite. O soro de queijo Coalho foi o que apresentou absorção mais pronunciada nessa região.

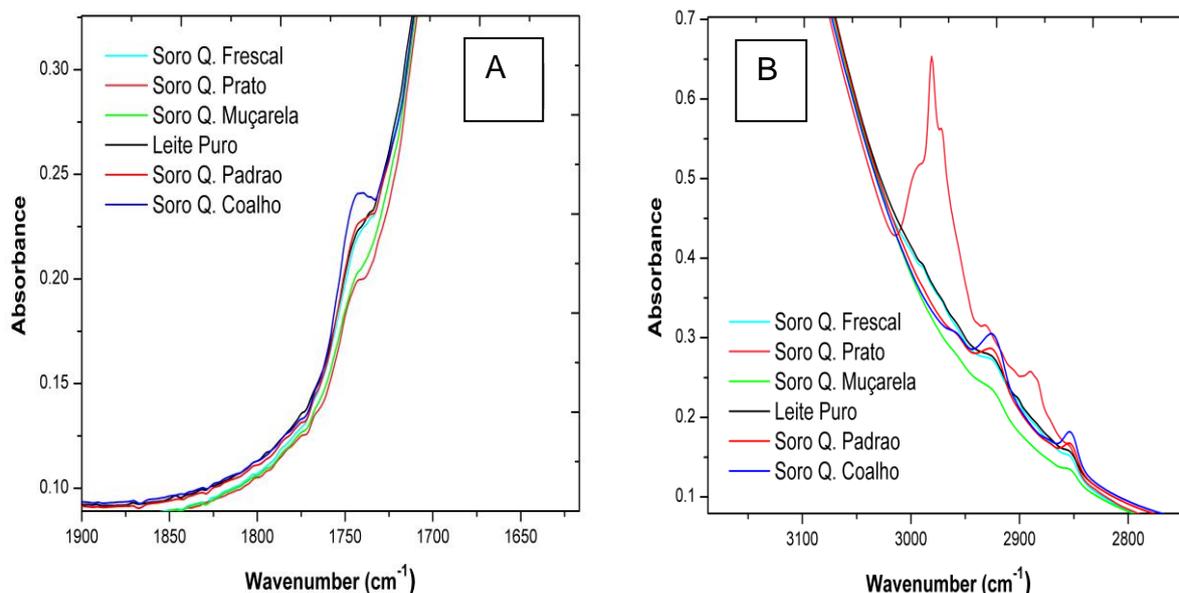


FIGURA 34: Bandas de absorção das amostras de leite adulteradas com soros de queijos nas regiões de 1700 a 1820 cm^{-1} (A) e 2800 a 3000 cm^{-1} (B).

Absorções na região de 2855 a 2854 cm^{-1} e 2927 a 2925 cm^{-1} relacionadas às vibrações de estiramento simétrico e assimétrico do CH_2 presentes nos ácidos graxos dos lipídeos, foram semelhantes para as amostras de leite puro e para as adulteradas com soro de queijo Minas Frescal e Minas Padrão.

Pequenas variações de absorção nas regiões características de absorção das frações lipídicas apresentadas na Figura 34 se justificaram pela existência de diferentes tipos de ácidos graxos e pelo grau de degradação dos mesmos, em detrimento ao processamento por meio do qual, os soros foram obtidos (CONWAY *et al.*, 2014; NICOLAOU *et al.*, 2010; PINTO *et al.*, 2011; SOYEURT *et al.*, 2006; INON, 2004).

O soro de queijo Prato apresentou uma banda de absorção totalmente diferenciada das demais amostras podendo ser atribuída à maior presença de ácido láctico. Segundo Stuart (2005), o ácido láctico possui banda de absorção na região de 3000 a 3100 cm^{-1} . O processo de fabricação do queijo Prato envolve a utilização de fermento láctico e maturação, fatores estes, que acarretam a produção de ácido láctico (FURTADO, 1991; ORDONEZ, 2005; PINTO, 2010).

4.4.2 Resultados referentes ao modelo preliminar PLS-DA contendo amostras de soros de queijos

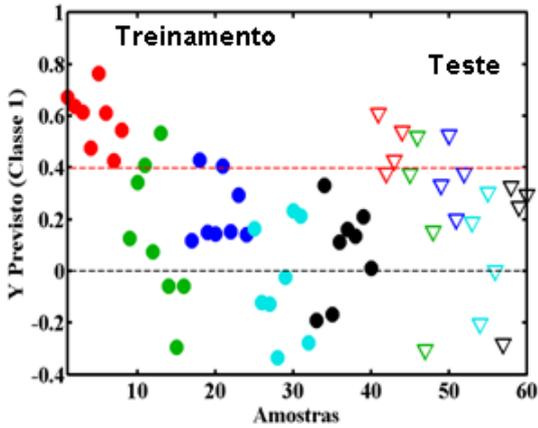
As classes para o modelo de treinamento (calibração) e de teste (previsão) foram identificadas da seguinte forma: classe 1 para soro de queijo Muçarela (M), classe 2 para soro de queijo Prato (K), classe 3 para soro de queijo Minas Frescal (MF), classe 4 para soro de queijo Coalho (C) e classe 5 para soro de queijo Minas padrão (MP).

A correção ortogonal utilizada como pré-tratamento dos sinais espectrais permitiu evitar perda de informações durante o pré-processamento dos espectros, reduzindo as variações em X (matriz das respostas instrumentais) não relacionadas a Y (matriz dos parâmetros a serem preditos pelo modelo).

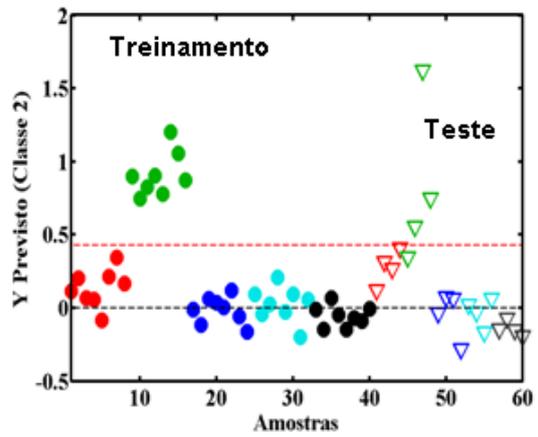
O melhor modelo PLS-DA após aplicação da correção ortogonal de sinal foi obtido com 5 VLs. Os dados foram centrados na média e as 5 VLs foram capazes de explicar 98,68% da variância de X e 55,94% da variância de Y .

O modelo gerou gráficos de classificação para cada uma das classes tanto para treinamento como para teste. Nos gráficos, cada classe composta por amostras similares foi apresentada na região acima da linha tracejada em vermelho. Este tracejado corresponde ao limite de distinção (limite de corte) definido pelo modelo para cada classe (*threshold*).

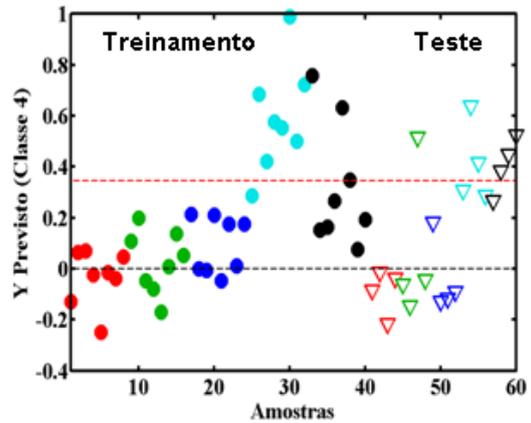
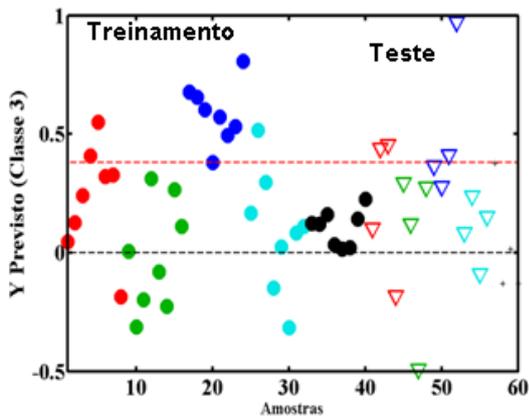
Na Figura 35 (gráficos A, B, C, D e E), podem ser visualizados os gráficos de classificação para as etapas de treinamento e teste do modelo para cada classe. Na etapa de calibração foram utilizadas todas as amostras experimentais e na etapa de teste foram utilizadas algumas amostras extraídas aleatoriamente do conjunto de treinamento para testar a eficiência do modelo. A eficiência do modelo correspondeu à constatação da real capacidade de reconhecer e classificar as amostras do modelo de teste. As amostras do modelo de treinamento foram representadas por círculos e as de teste por triângulos.



A -Classe 1: Leite com soro de queijo Muçarela



B- Classe 2: Leite com soro de queijo Prato



Legenda:

Vermelho: Classe 1: Soro de queijo Muçarela

Verde: Classe 2: Soro de queijo Prato

Azul escuro: Classe 3: Soro de queijo Minas Frescal

Azul claro: Classe 4: Soro de queijo de Coalho

Preto: Classe 5: Soro de queijo Minas Padrão

E - Classe 5: Leite com soro de queijo Padrão

FIGURA 35: Gráficos de classificação das amostras de leite adulteradas com diferentes soros de queijos

Os resultados do modelo PLS-DA, foram apresentados por meio das taxas de sensibilidade e seletividade (especificidade) cujos limites máximo e mínimo correspondem respectivamente a 1,0 e 0,5 tanto para o modelo de treinamento como para o modelo teste (previsão). A confiabilidade dos modelos PLS-DA se limitaram a um valor mínimo de 0,5 para os dois parâmetros de validação, entretanto, o ideal é que obtenha valores o mais próximo possível de 1,0 (alta confiabilidade) (PALERMO, PIRAINO E ZUCHT, 2009; SZYMANSKA *et al.*, 2012).

Os resultados do modelo de classificação foram expressos na Tabela 15.

TABELA 15: Sensibilidade e seletividade do modelo PLS-DA para soros de queijos

Modelos PLS-DA	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5
Sensibilidade (treinamento)	1,000	1,000	0,875	0,875	0,750
Seletividade (treinamento)	0,875	1,000	0,906	0,906	0,813
Sensibilidade (teste)	0,750	0,750	0,505	0,500	1,000
Seletividade (teste)	0,875	1,000	0,875	0,750	0,688

*Classe 1: leite com soro de Muçarela; classe 2: leite com soro de queijo Prato; classe 3: leite com soro de queijo Minas Frescal; classe 4: leite com soro de queijo Coalho; classe 5: leite com soro de queijo Minas padrão.

Observou-se que, o PLS-DA para a classe, 1 apresentou máxima sensibilidade (relação ente resultados comprovadamente positivos e o total de amostras positivas) para modelo de treinamento e índice de taxa de 0,750 no modelo teste. Já o valor de seletividade (relação ente resultados comprovadamente negativos e o total de amostras negativas) do modelo de treinamento e no de teste foi de 0,875.

A classe 2 referente ao soro de queijo Prato apresentou alta sensibilidade e seletividade para o modelo de treinamento e teste. O modelo de treinamento e de teste para a classe 2 mostrou alta similaridade entre amostras do mesmo grupo e alta diferença em relação às demais classes. O perfil espectral do soro de queijo Prato foi o que mais diferenciou dos demais. Na elaboração desse tipo de queijo é diferenciada pela adição de corante e pelo processo de maturação que propicia a formação de compostos aromáticos e ácido láctico (ORDONEZ, 2005).

A Classe 3 obteve valores acima de 0,875 tanto para sensibilidade como para seletividade no modelo de treinamento. Para o modelo teste, obteve-se taxa de

sensibilidade de 0,500 e de seletividade de 0,875. O valor de sensibilidade de 0,500 é considerado como mínimo para a confiabilidade do modelo.

A classificação das amostras de queijo Coalho (classe 4) demonstrou similaridade entre as amostras do mesmo grupo e também certa similaridade com duas amostras de queijo Minas Padrão, tanto no modelo de treinamento como no de teste. A classe 4, assim como a 3, apresentou taxa de 0,500 de sensibilidade para o modelo teste e 0,875 para o modelo de treinamento. Resultados acima de 0,750 foram obtidos para seletividade nos modelos de treinamento e teste.

A classe 5 apresentou máxima sensibilidade para o modelo teste (1,000) e seletividade de 0,688. O conjunto treinamento obteve valores considerados satisfatórios tanto para sensibilidade (0,750) como para seletividade (0,813). Essa classe apresentou alta similaridade entre amostras do mesmo grupo e certa similaridade com amostras de soro de queijo coalho.

De um modo geral, os resultados para os modelos de treinamento foram satisfatórios para todas as classes, uma vez que, apresentaram taxas de 0,750 tanto para sensibilidade como para seletividade. Os modelos teste de novas amostras, atenderam o limite mínimo requerido para classificação (erro tipo II máximo de 5%) para as classes 3 e 4. Segundo Szymanska *et al.* (2012) 0,5 corresponde ao valor mínimo para discriminação entre classes, ou seja, sensibilidade mínima de classificação de 95% e máximo de erro tipo II de 0,5%.

Modelos PLS-DA capazes de classificar adequadamente o tipo de soro de queijo utilizado na adulteração de leite permitem compreender a causa de complicações nos organismos de pessoas que ingerirem o leite adulterado, como por exemplo, o aparecimento de sintomas alérgicos a corantes.

Informações geradas por meio da classificação dos possibilitam, ainda, o rastreamento da fonte fornecedora do adulterante para que esse tipo de prática fraudulenta seja contida.

4.5 Modelo para avaliar a possibilidade de classificar soros de manteiga (leitelho) em adulterações de leite UHT por meio do modelo PLS-DA

4.5.1 Resultados referentes ao perfil espectral

Por meio da Figura 36 pôde-se observar o perfil espectral do leite UHT integral utilizado nas adulterações simuladas com diferentes tipos de leitelhos.

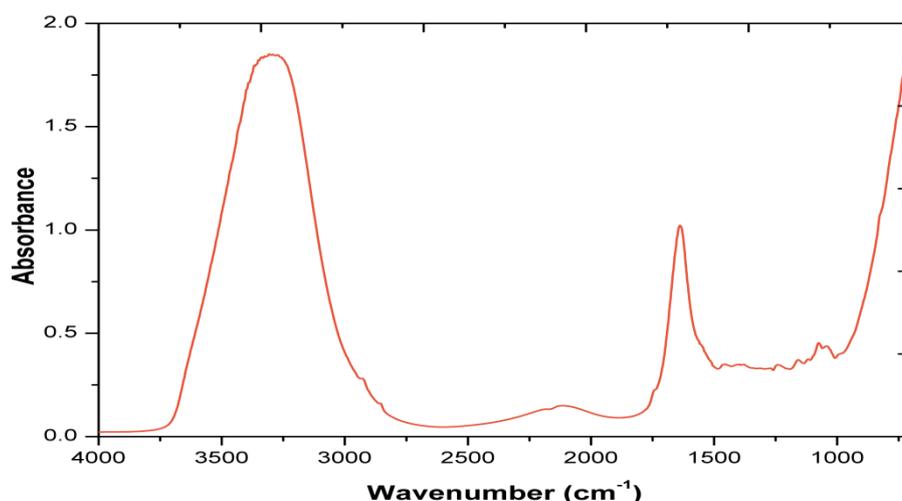
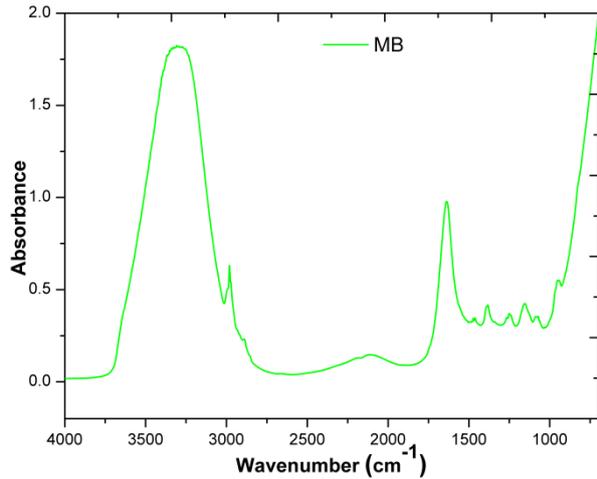


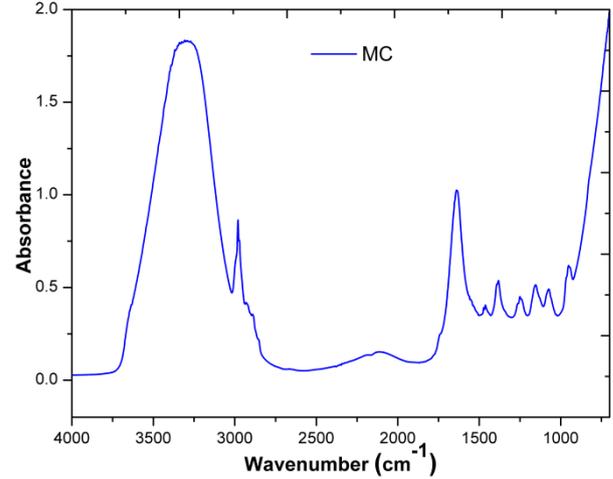
FIGURA 36: Espectro do leite puro utilizado nas adulterações com leitelho

Para ampliar o estudo sobre adulteração de leites fluidos por soros lácteos, espectros puros de diferentes tipos de leitelhos foram obtidos e apresentados por meio da Figura 37 (gráficos A, B, C e D).

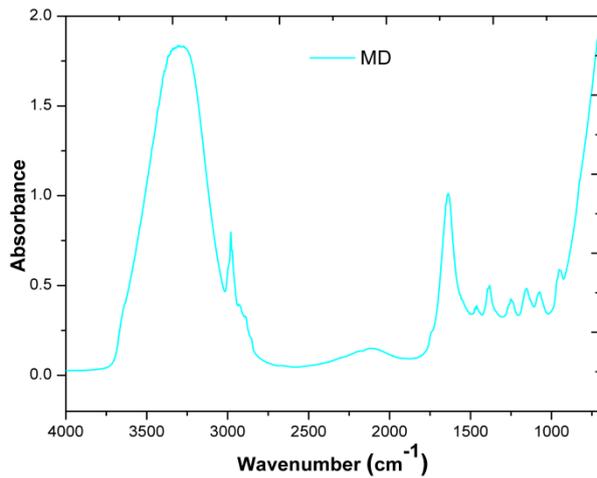
Os gráficos permitiram visualizar as características das bandas de absorção típicas de cada tipo de leitelho utilizado nas adulterações artificiais do leite UHT integral. Os leitelhos utilizados foram: leitelho de manteiga extra com creme diluído com leite desnatado (MA); leitelho de manteiga extra com creme diluído com água (MB); leitelho de manteiga comum com menor acidez (MC) e leitelho de manteiga comum com maior acidez (MD).



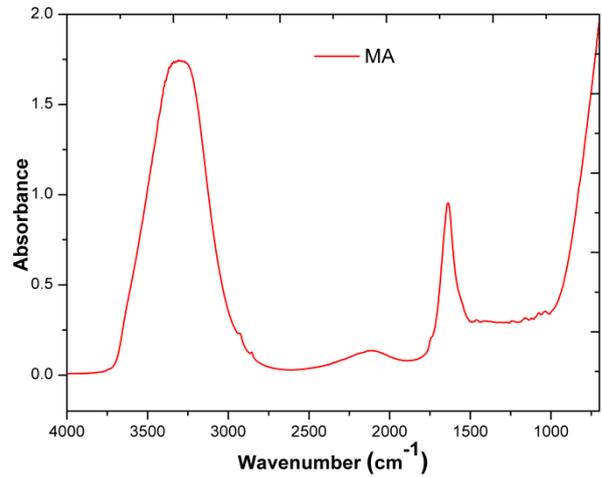
A - Leiteiro de mateiga extra diluído com água



B - Leiteiro de manteiga comum de maior acidez



C - Leiteiro de manteiga comum com menor acidez



D - Leiteiro de manteiga extra com creme diluído com leite desnatado

FIGURA 37: Espectros puros dos diferentes leiteiros utilizados nas adulterações

Gráfico contendo os espectros de cada tipo de leiteiro foi confeccionado para facilitar a observação e comparação das absorvâncias (Figura 38).

Observou-se que o soro MA possui um perfil espectral semelhante ao do leite, correspondendo ao leiteiro que mais se diferenciou dos demais. O leiteiro MA foi obtido pelo processamento de manteiga extra em que o creme foi diluído com leite desnatado. Esse fato justifica sua maior semelhança com o leite. Os demais leiteiros foram obtidos de manteiga elaborada com creme diluído com água.

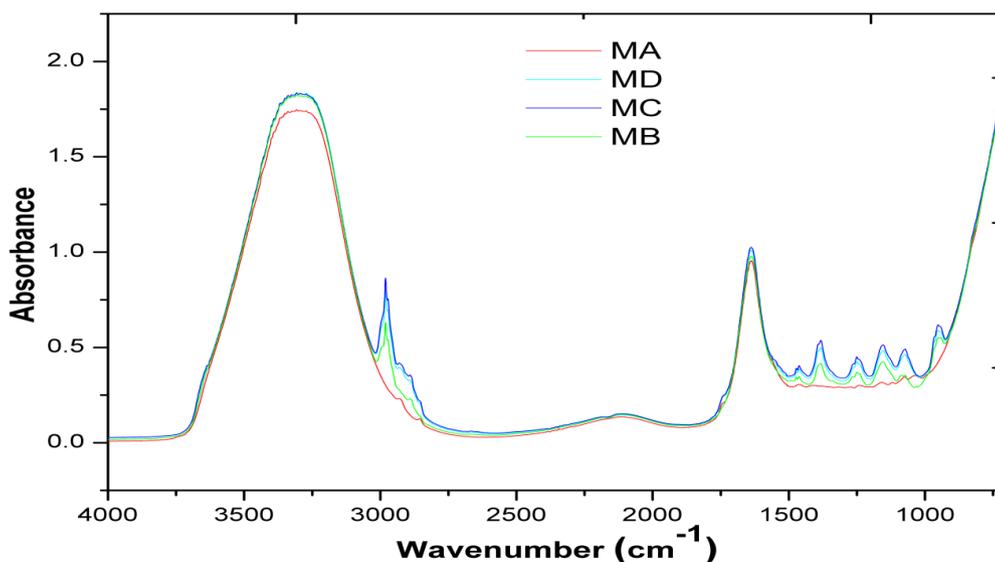


FIGURA 38: Espectros de cada tipo de leite adulterante do modelo preliminar PLS-DA

Absorções atribuídas à presença da lactose na região de $1030\text{-}1150\text{ cm}^{-1}$ (vibrações C-O, C-C e C-O-O dos carboidratos) e à presença de grupos fosfatos ($1060\text{-}1100\text{ cm}^{-1}$) ocorreram em todas as amostras. Absorções nas regiões correspondentes a $1251\text{-}1241\text{ cm}^{-1}$ atribuída às amidas III e a região de $1547\text{-}1541\text{ cm}^{-1}$ atribuída às amidas II foram similares para os leitelhos MC, MD e MB. Durante o processamento da manteiga componentes proteicos não são retidos na manteiga e são liberados no leite. A membrana dos glóbulos de gordura, rompida durante a bateção da manteiga, é rica em lipoproteínas (SANTOS, 2013).

Nas amostras dos leitelhos MC, MD e MB, absorções na região de $1020\text{-}1200\text{ cm}^{-1}$ (álcoois) e $1200\text{-}1250\text{ cm}^{-1}$ (grupos fenólicos) foram observadas, assim como, absorções que podem ser atribuídas a compostos aromáticos (alterações no anel benzeno nas ligações C=C em $1630\text{-}1430\text{ cm}^{-1}$ e C-C em $1275\text{-}1000\text{ cm}^{-1}$ e $900\text{-}690\text{ cm}^{-1}$) (STUART, 2005). Absorções similares ocorreram nas amostras adulteradas com soro de queijo Prato.

Absorção relacionadas aos lipídeos nas amostras de leite foram atribuídas às regiões próximas à 1747 cm^{-1} (estiramento vibracional C=O), e também nas regiões de $2855\text{-}2854\text{ cm}^{-1}$ e $2927\text{-}2925\text{ cm}^{-1}$ relativas às vibrações de estiramento simétrico e assimétrico do CH_2 . Essas bandas de absorção de lipídeos foram também observadas por Carvalho (2013) em estudo para levantamento de diferenças entre leites bovino e de cabra.

Região de absorção entre $2850\text{ e }3000\text{ cm}^{-1}$ foi apresentada pela Figura 39.

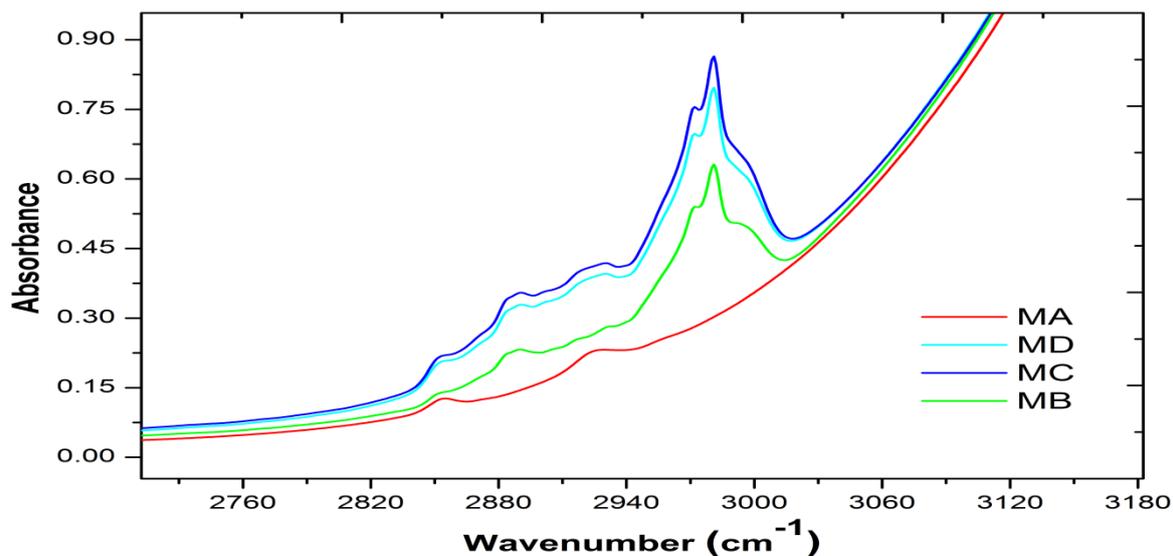


FIGURA 39: Absorção na região de 2820 a 3000 cm^{-1}

O perfil espectral da região da amostra de leite adulterada com os leiteiros MC, MD e MB apresentaram semelhança facilmente perceptível devido a uma banda de absorção pronunciada próxima à região de absorção de 3000 cm^{-1} . Segundo Stuart (2005), essas bandas de absorção estariam relacionadas à presença de compostos aromáticos (grupo alifático) que absorve na região de 3000 a 3100 cm^{-1} .

Outras regiões no espectro do infravermelho médio atribuídas à absorções da ligação O-H de lipídeos presentes no leite foram mencionadas por Aernouts *et al.* (2011) na região de 3000 - 3650 cm^{-1} e por Carvalho (2013) na região de 3438 cm^{-1} . em suas pesquisas.

O estudo das características dos espectros de cada tipo de leiteiro utilizado como adulterante contribuiu significativamente para a compreensão dos resultados obtidos por meio da modelagem.

4.5.2 Resultados referentes ao modelo contendo amostras de leiteiro

Um modelo exploratório usando Análise de Componentes Principais foi realizado na faixa espectral considerada mais significativa correspondente à região de absorção das proteínas (1550 a 1800 cm^{-1}), após algumas simulações de modelagem.

O melhor modelo PCA foi obtido com os dados centrados na média e com 2 CPs que foram capazes de explicar 95,58 % da variância dos dados.

No gráfico da Figura 40 podem ser observados os escores para as 2 primeiras CPs. As amostras MA e MB foram classificadas em grupamento localizado à direita do gráfico (menor acidez) da PC1 enquanto as amostras MC e MD foram classificadas à esquerda da PC1 (maior acidez).

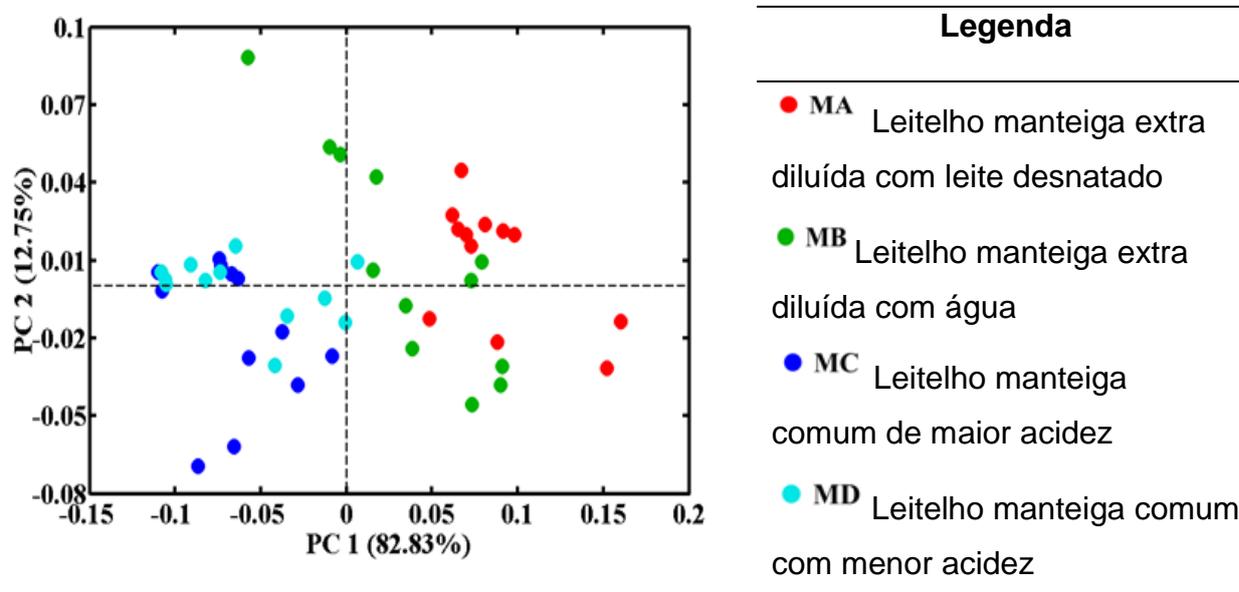


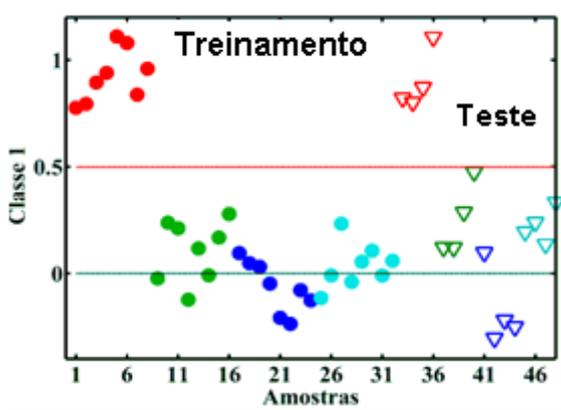
FIGURA 40: Escores das duas primeiras CPs para modelo PCA

Após a construção do modelo exploratório foi construído modelo PLS-DA utilizando a correção ortogonal de sinal para melhor captura de sinais. Para a construção do modelo PLS-DA foi criada uma matriz Y contendo valores binários para cada uma das classes.

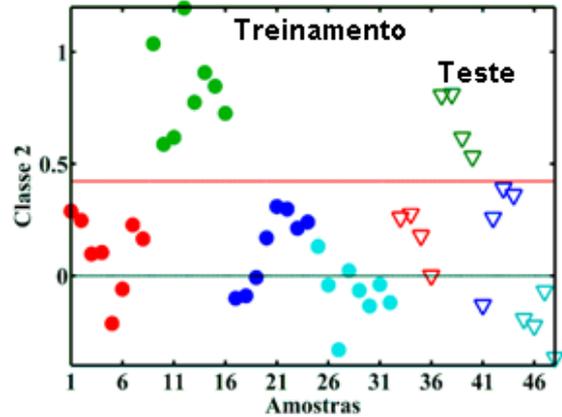
O melhor modelo PLS-DA foi obtido com 5 VLs e com os dados centrados na média explicando 98,90% da variância de X e 70,73% da variância de Y .

A Figura 41 apresenta a classificação dos leitelhos: classe 1 (MA), classe 2 (MB), classe 3 (MC) e classe 4 (MD), correspondentes às classes originadas pelos modelos de treinamento e teste. As amostras de treinamento foram representadas por círculos e as amostras de previsão por triângulos, para cada classe obtida pelo modelo PLS-DA.

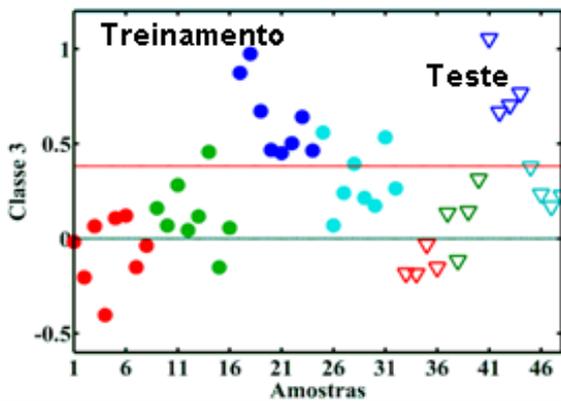
Pôde-se observar que as amostras de leite UHT contendo leite manteiga extra diluído com leite desnatado (classe1-MA) apresentaram alta similaridade entre as amostras de mesmo grupo tanto para o modelo de treinamento como para o de teste. O modelo foi capaz de prever todas as do conjunto teste.



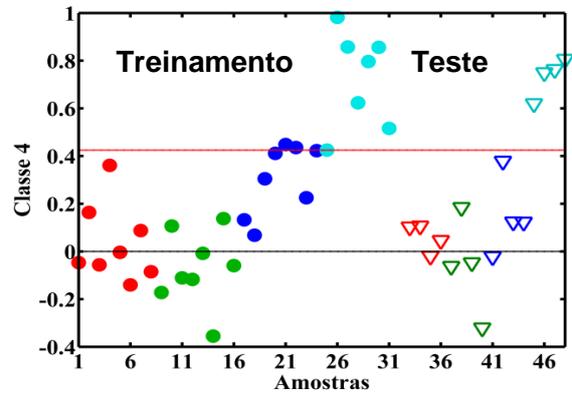
Classe 1 - Leiteinho mant. extra dil. leite



Classe 2 - Leiteinho mant. extra dil. água



Classe 3 - Leiteinho mant. comum maior acidez



Classe 4 - Leiteinho mant. comum menor acidez

Legenda

Vermelho: Leiteinho manteiga extra diluída com leite desnatado

Verde: Leiteinho manteiga extra diluída com água

Azul escuro: Leiteinho manteiga comum de maior acidez

Azul claro: Leiteinho manteiga comum com menor acidez

FIGURA 41: Modelo PLS-DA para classificação das amostras de leite UHT adulteradas com diferentes leiteinhos.

A classe 2 (MB) correspondente ao leiteinho de manteiga extra diluída com água, também apresentou alta similaridade entre as amostras do grupo, tendo sido adequadamente previstas, todas as amostras do conjunto teste.

As amostras de leite UHT contendo leiteinho de manteiga comum de menor acidez (MC), correspondente à classe 3, apresentou alta similaridade entre as amostras do mesmo grupo no modelo de treinamento. Observou-se também certa

similaridade das amostras dessa classe com três amostras da classe 4 e com uma amostra da classe 2. Contudo, o modelo teste apresentou alta capacidade de classificação para as amostras previstas.

A classe 4 referente às amostras de leite UHT contendo leite de manteiga comum com maior acidez (MD), apresentou alta similaridade entre amostras do mesmo grupo, tanto para o conjunto de treinamento como para o de teste.

Por meio da observação dos gráficos de classificação constatou-se que o modelo foi capaz de classificar as amostras de leite UHT adulteradas com diferentes leitelhos. Todas as amostras do modelo foram adequadamente classificadas em relação à similaridade inter e intragrupo.

A Tabela 16 apresenta os valores de sensibilidade e seletividade (especificidade), tendo sido considerada como faixa de confiabilidade para o modelo PLS-DA valores entre 0,5 e 1,0 (máxima capacidade de classificação).

TABELA 16: Sensibilidade e seletividade do modelo PLS-DA para leitelhos

Modelos PLS-DA	Classe 1	Classe 2	Classe 3	Classe 4
Sensibilidade (treinamento)	1,000	1,000	1,000	1,000
seletividade (treinamento)	1,000	1,000	0,833	0,917
Sensibilidade (teste)	1,000	1,000	1,000	1,000
Seletividade (teste)	1,000	1,000	1,000	1,000

Legenda: Classe1: manteiga extra diluída com leite desnatado; classe 2: manteiga extra diluída com água; classe 3: manteiga comum com maior acidez; classe 4: manteiga comum com menor acidez.

Os resultados demonstraram a eficiência de classificação do modelo PLS-DA, uma vez que, apresentou valores elevados (próximos a 1,0) para sensibilidade e seletividade para os modelos (etapas) de treinamento e teste. O modelo foi capaz de classificar 100% das amostras verdadeiramente positivas em relação ao total de amostras positivas (elevada sensibilidade). O modelo também apresentou alta seletividade para as classes nos modelos de treinamento e máxima capacidade de classificar amostras verdadeiramente negativas dentre as amostras supostamente negativas nos modelos de teste.

4.5.3 Resultados referentes ao modelo contendo amostras de leite e amostra de leite puro

Outro modelo PLS-DA com os mesmos espectros dos leitelhos utilizados no modelo anterior foi também obtido com a inclusão de amostra de leite UHT isento de adulteração. Esse novo conjunto de amostras foi utilizado para verificar se a modelagem seria capaz de classificar leite não adulterado.

Inicialmente, análise exploratória por meio da PCA, foi realizada. O melhor modelo PCA foi obtido com os dados centrados na média e com 2 CPs explicando 95,3 % da variância dos dados. No gráfico da Figura 42 podem ser observados os escores para as 2 primeiras CPs com grupamentos de amostras (em função da similaridade) à direita (menor acidez e maior similaridade com o leite puro) e à esquerda do gráfico (maior acidez).

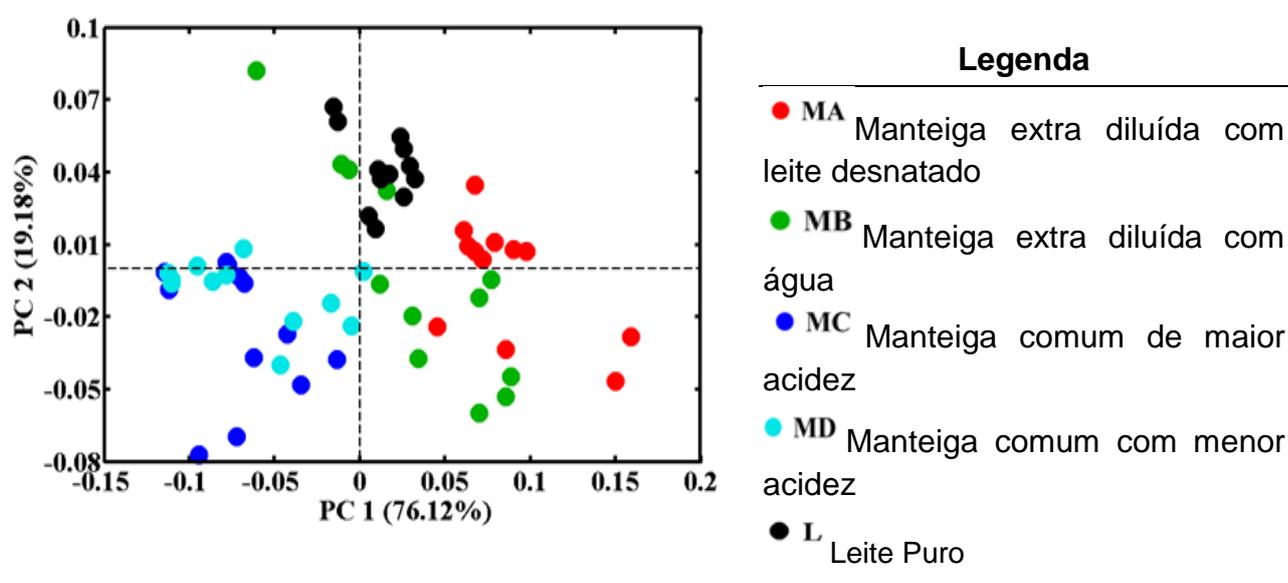
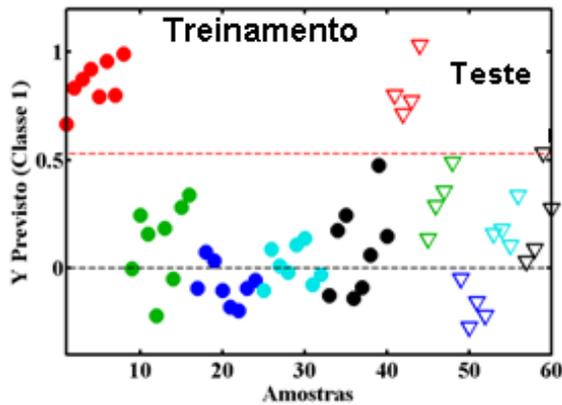
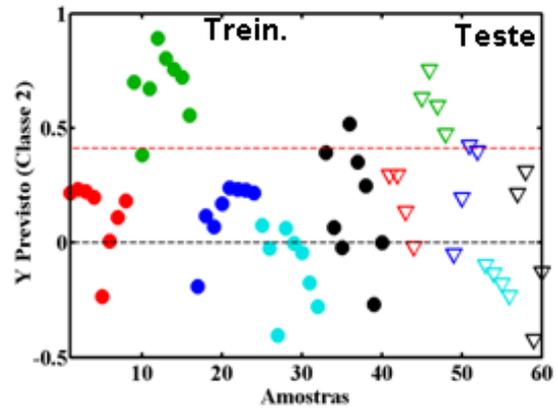


FIGURA 42: Escores das duas primeiras PCs para modelo com leite puro

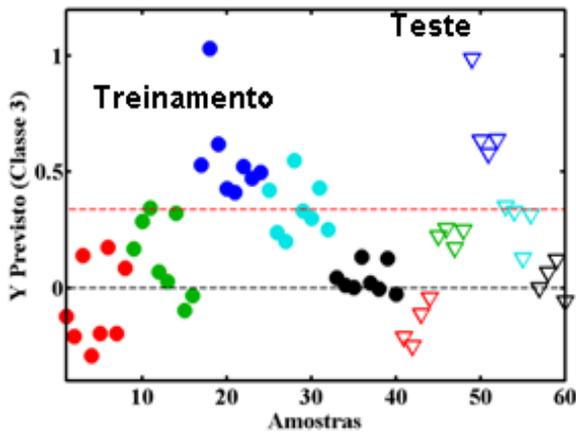
A separação das classes do modelo PLS-DA após inclusão da amostra de leite puro foi apresentada na Figura 43.



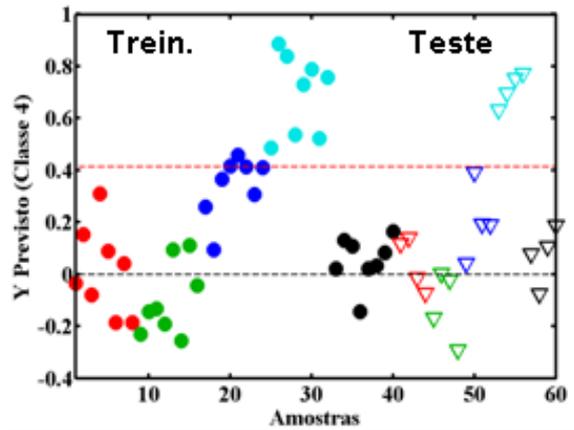
Classe 1 - Leiteinho mant. extra diluída com leite desnatado



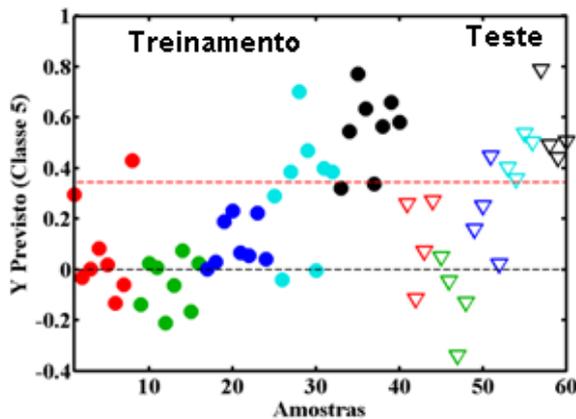
Classe 2 - Leiteinho mant. extra diluída com água



Classe 3 –Leiteinho manteiga comum maior acidez



Classe 4 – Leiteinho manteiga comum menor acidez



Classe 5 – Leite Puro

Legenda

Vermelho: Leiteinho manteiga extra diluída com leite desnatado

Verde: Leiteinho manteiga extra diluida com água

Azul escuro: Leiteinho manteiga comum de maior acidez

Azul claro: Leiteinho manteiga comum com menor acidez

Preto: Leite Puro

FIGURA 43: Classificação das amostras adulteradas com leiteinho

O melhor modelo PLS-DA foi construído com 5 VLs e com os dados centrados na média explicando 98,54% da variância de X e 61,33% da variância de y . A inclusão da amostra de leite não alterou de forma significativa o modelo PLS-DA, que obteve melhores resultados com o mesmo número de variáveis latentes do modelo anterior (isento do leite puro).

Observou-se que, assim como no modelo anterior, que as amostras de leite de manteiga extra diluído com leite desnatado (MA) referente à classe 1, apresentaram alta similaridade dentro do grupo e alta distinção em relação às demais classes, tanto no modelo de treinamento como no de teste.

Todas as amostras do conjunto teste foram adequadamente classificadas. As amostras da classe 2 (leitelho de manteiga extra diluída com água) apresentaram alta similaridade no grupo para os modelos de treinamento e teste. Todas as amostras de previstas para a classe 2, foram adequadamente classificadas.

A classe 3 referente ao leite de manteiga comum de menor acidez, apresentou no modelo de treinamento a formação de um grupo distinto, porém, pôde-se observar certa similaridade com três amostras da classe 4. Já no modelo de teste, o grupo apresentou alta similaridade entre as amostras e alta discriminação em relação às demais classes.

As amostras da classe 4 (leitelho de manteiga comum com maior acidez) apresentaram tanto para o modelo de treinamento como para o de teste, alta similaridade dentro do próprio grupo e diferença significativa em relação às amostras dos outros grupos.

A classe 5 correspondente às amostras de leite UHT utilizado para realização das adulterações, apresentou similaridade entre as amostras do próprio grupo e também certa similaridade em relação às amostras da classe 4, no modelo de treinamento.

É importante ressaltar que os resultados do modelo em relação à classificação das amostras foram obtidos para faixa de absorção específica e não para todo o espectro. Os resultados dos valores de sensibilidade e seletividade gerados para o modelo foram apresentados na Tabela 17.

TABELA 17: Sensibilidade e seletividade para modelo PLS-DA (leitelhos e leite puro)

Modelos PLS-DA	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5
Sensibilidade (treinamento)	1,000	0,875	1,000	1,000	1,000
Seletividade (treinamento)	1,000	0,969	0,875	0,938	0,969
Sensibilidade (teste)	1,000	1,000	1,00	1,000	1,000
Seletividade (teste)	0,938	0,938	0,938	1,000	1,000

Legenda: Classe1: manteiga extra diluída com leite desnatado; classe 2: manteiga extra diluída com água; classe 3: manteiga comum com maior acidez; classe 4: manteiga comum com menor acidez; classe 5: leite puro.

Os resultados da Tabela 17 demonstraram que o PLS-DA para a classe 1, apresentou alta probabilidade de classificar amostras verdadeiramente positivas e verdadeiramente negativas em relação ao total de amostras positivas e negativas, tanto no modelo de treinamento como no de teste.

A classe 2 apresentou valores acima de 0,875 tanto para sensibilidade como para seletividade no modelo de treinamento e apresentou taxa de 0,938 para seletividade e máxima sensibilidade para o modelo teste.

A classe 3 apresentou máxima sensibilidade (constatação das amostras verdadeiramente positivas) para os modelo de treinamento e teste e apresentou valores acima de 0,875 para seletividade nos modelos de treinamento e teste.

A classe 4 também apresentou máxima sensibilidade (treinamento e teste) e resultados para seletividade muito próximos do ideal (1,0) nos modelos de treinamento e teste.

A classe 5 obteve resultados semelhantes à classe 4.

Concluiu-se que o modelo foi capaz de identificar (classificar) as amostras adulteradas com leite e as amostras isentas de adulteração. Na prática, novas amostras submetidas a esse modelo seriam reconhecidas como adulteradas com leite, caso fossem classificadas como pertencentes às classes 1, 2, 3 ou 4. Seriam consideradas como isentas de adulteração com leite se fossem reconhecidas como pertencente à classe 5.

A adição de uma classe contendo amostras de leite puro proporcionou ao modelo a capacidade de constatar a ocorrência ou não da adulteração com soro de queijo (presença ou ausência do adulterante testado).

Essa informação seria suficiente para propor a aceitação ou rejeição do leite, uma vez que a legislação proíbe a adição de soro no leite. Entretanto, conhecer o tipo de soro lácteo, corresponde a uma informação que pode contribuir para localização da fonte fornecedora do insumo adulterante, auxiliando os órgãos fiscalizadores no combate a essa prática criminosa.

5 RESULTADOS E DISCUSSÃO REFERENTES À TÉCNICA FTIR/MIR E MODELO PREDITIVO PLS PARA DETECÇÃO DE ADULTERAÇÃO DE LEITE CRU COM SORO DE QUEIJO MINAS FRESCAL

5.1 Resultados referentes ao perfil espectral

O modelo PLS foi construído utilizando todo o espectro a fim de avaliar as alterações nas bandas de absorção do leite cru decorrentes da adição do soro de queijo Minas Frescal. O soro de Minas Frescal foi escolhido para o desenvolvimento desse modelo qualitativo em virtude de suas características sensoriais relacionadas ao sabor suave e à baixa acidez.

O perfil espectral do leite segundo Wolfschoon (1977) possui bandas de absorção na região de 3500 cm^{-1} referente às moléculas de água; banda na região de 1735 cm^{-1} referente ao estiramento C=O das ligações ésteres dos triglicerídeos; banda na região de 1535 cm^{-1} referente a absorção da amida II (N-H) das proteínas e banda de absorção na região de 1065 cm^{-1} referente às vibrações C-CH das moléculas de álcool presentes na lactose. As constatações deste estudo de absorvância podem ser confirmadas por meio da Figura 44 que apresenta o espectro do leite cru utilizado nas adulterações.

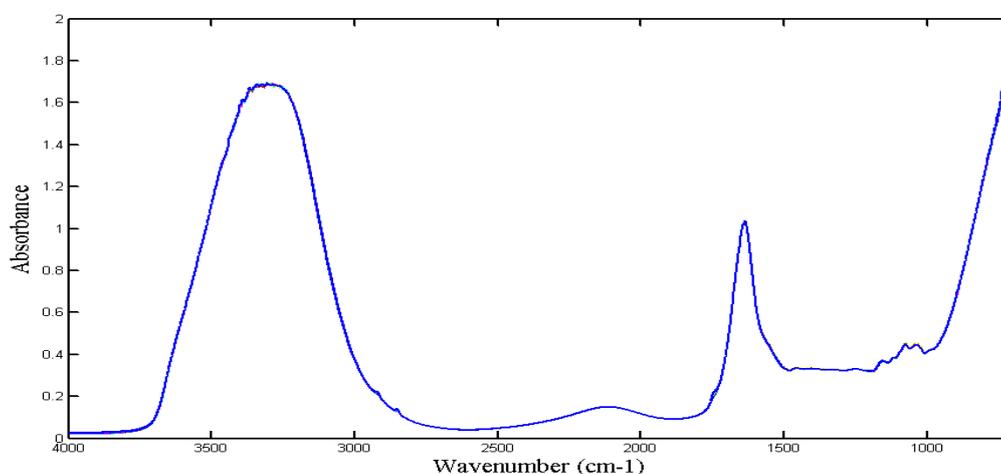


FIGURA 44: Espectro do leite cru utilizado nas adulterações com soro Minas Frescal

O conjunto de espectros das amostras de leite contendo concentrações de soro de queijo Minas Frescal na faixa de 6 a 28% constam na Figura 45.

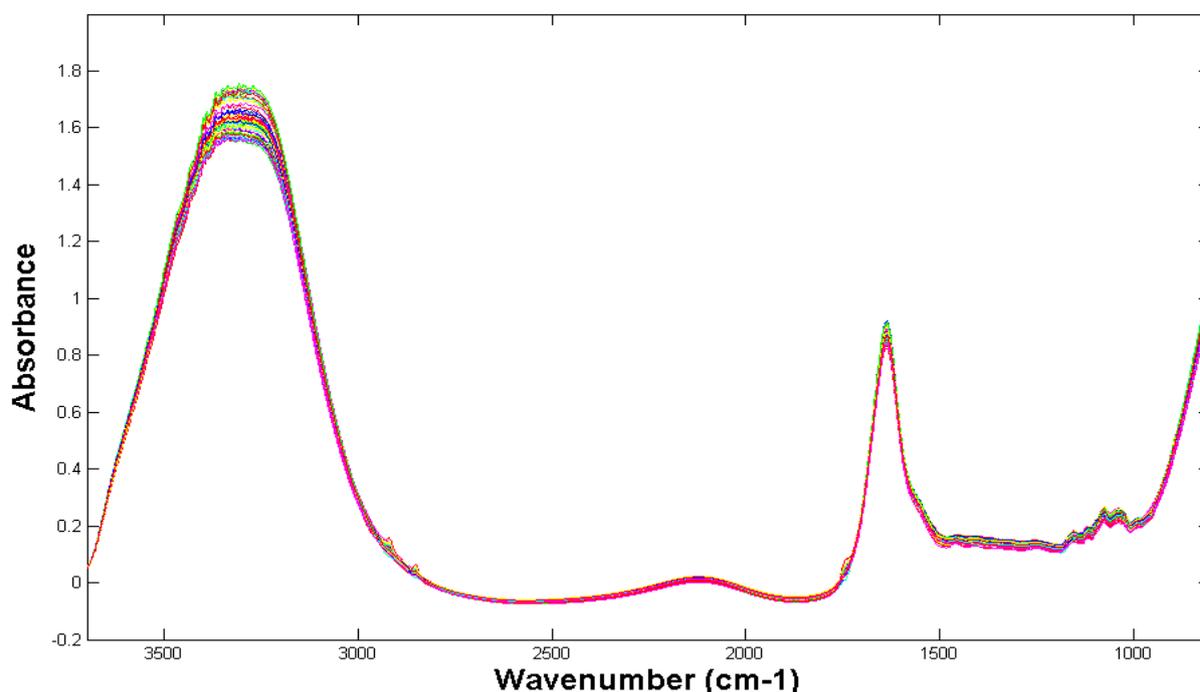


FIGURA 45: Conjunto espectros de leite cru adulterado com diferentes concentrações de soro Minas Frescal

Os perfis espectrais das amostras adulteradas com soro fresco demonstraram pequenas variações em determinadas bandas de absorção em relação ao espectro da amostra de leite cru, isenta de adulteração. Em pesquisa realizada por Santos, Pereira-Filho e Rodriguez-Saona (2013) para desenvolvimento de método rápido para detecção e quantificação de adulterações em leite utilizando FTIRMIR e técnicas quimiométricas, os autores também detectaram alterações no espectro do leite após adulterações com diferentes substâncias, inclusive soro de queijo. Alterações nas regiões de absorção entre 1500 e 1700 cm^{-1} foram atribuídos à presença do soro. Nessa região a absorção é atribuída ao estiramento das amidas I e II. Bandas mais proeminentes de absorção foram observadas em 1620 cm^{-1} e correlacionadas pelos autores à absorção das estruturas β das proteínas presentes nas β lactoglobulinas que se encontram dispersas no soro.

Comparando o perfil espectral do leite puro (cru) com os espectros do leite adulterado, observou-se em algumas regiões que as bandas de absorção aumentaram significativamente e que outras três pequenas regiões de absorção passaram a existir.

Por meio do recorte e ampliação dos espectros das amostras adulteradas artificialmente (Figura 46) constatou-se absorção na região de 1020 a 1140 cm^{-1} atribuída ao estiramento dos grupos C-O, C-C e C-O-O presentes na molécula de lactose. Absorção da lactose na região de 1030 a 1150 cm^{-1} foi também constatada em estudo para análise de leite em pó bovino e de cabra por meio de infravermelho médio (CARVALHO, 2013).

Nicolaou *et al.* (2010) analisando leites de diferentes espécies utilizados na adulteração de leite bovino, atribuiu a região de 1159 - 1076 cm^{-1} à absorção de carboidratos, enfatizando ser este, um importante fator de diferenciação entre leites de diferentes espécies.

Por meio da Figura 46 também observou-se absorvância de grupo O=P-O da caseína na região próxima a 1000 cm^{-1} e do grupo amida na região próxima a 1500 cm^{-1} .

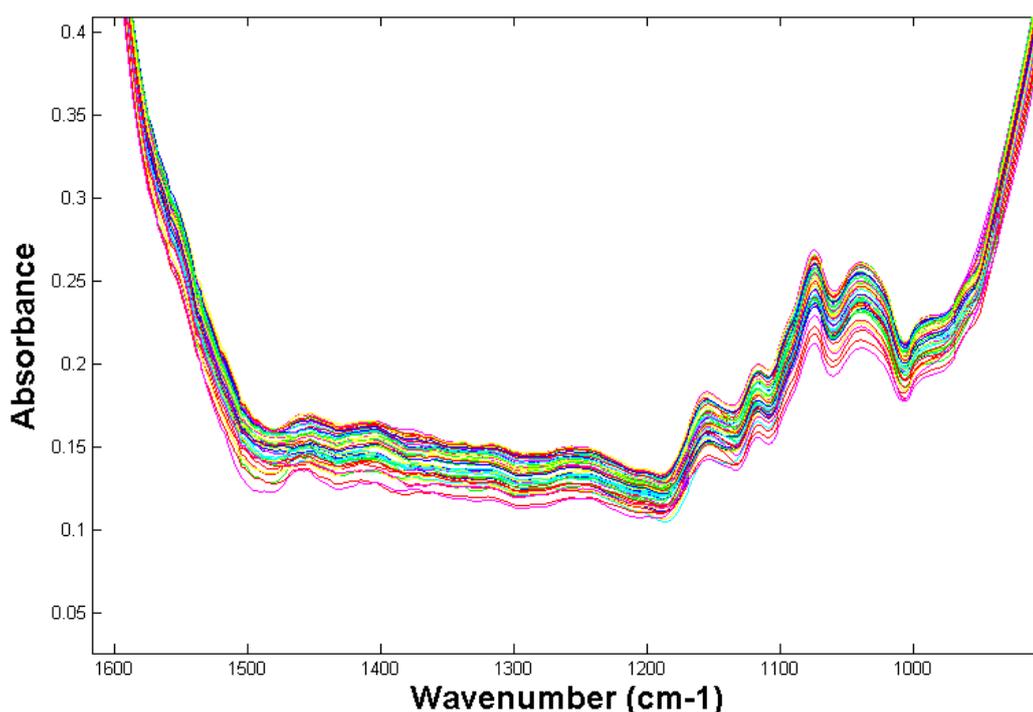


FIGURA 46: Região de absorção de 950 a 1450 cm^{-1}

Etzion *et al.* (2004), em estudo sobre a concentração de proteínas lácteas observou, por meio de infravermelho médio, que a absorvância das proteínas do leite se dava em duas bandas de absorção: uma na região de comprimento de onda 1500-1700 cm^{-1} e outra na região de absorção de comprimento de onda de 1060 a 1100 cm^{-1} relacionada aos grupos de fosfatos covalentemente ligados à micela de caseína. Essa

região apareceu mais pronunciada nas amostras de soro de queijo do que na amostra de leite. Isso ocorreu devido à liberação dessas minerais quando a micela de caseína foi rompida pela enzima quimosina utilizada para coagulação do leite nos processos de elaboração de queijos (ORDONEZ, 2005).

Suaves bandas de absorção correspondentes à absorção das amidas III e II foram observadas entre 1200 cm^{-1} e 1500 cm^{-1} .

Variações nas absorções da região situada entre 1620 a 1660 cm^{-1} atribuídas ao grupamento C=O acoplados às flexões N-H dos aminoácidos existentes nas ligações peptídicas das amidas I podem ser observadas por meio da Figura 47.

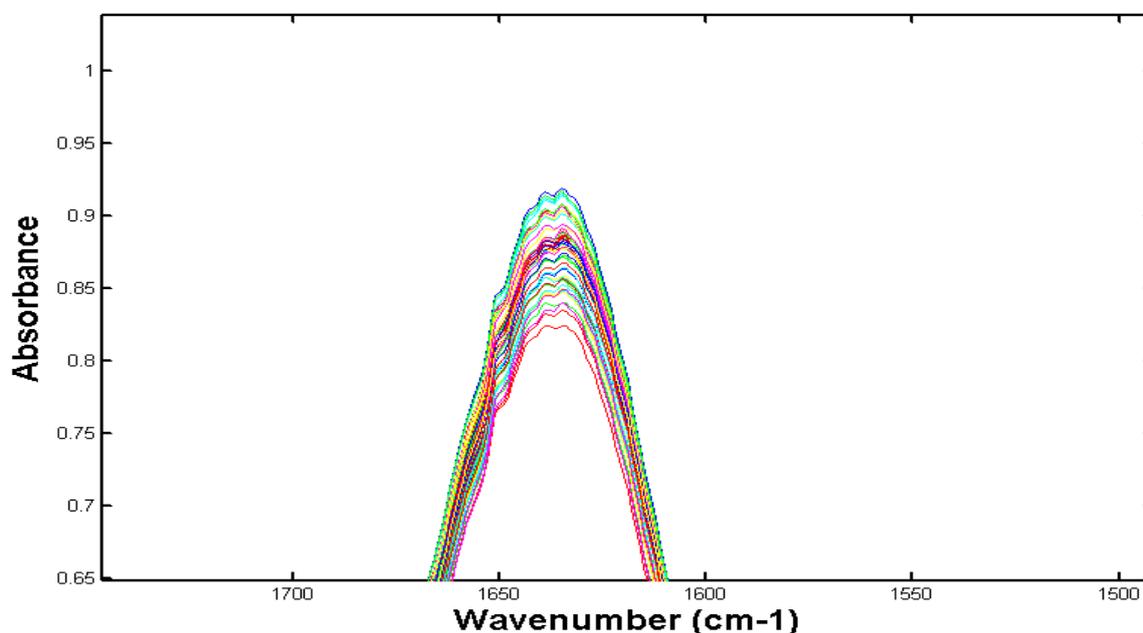


FIGURA 47: Região de absorção de 1600 a 1650 cm^{-1}

A forma e a intensidade das bandas de absorção existentes nessa região foram influenciadas pelas diferentes estruturas secundárias das proteínas, em particular os sinais atribuídos a configurações em α -hélices (1645 - 1662 cm^{-1}), estruturas β -pregueadas (1613 - 1637 cm^{-1}), turns (1662 - 1682 cm^{-1}) e estruturas randômicas (1637 - 1645 cm^{-1}). Essas estruturas correspondem à conformação geométrica da cadeia peptídica que se apresentam na forma de hélice (α), folha (β), dobras (β) ou em mais de um tipo de conformação na mesma estrutura secundária da proteína (PAVIA *et al.*, 2015).

Estudos sobre sensibilidade de indivíduos à proteínas do leite atribuem à beta-lactoglobulina, a alfa-lactoalbumina e a fragmentos de caseína, a capacidade de provocar alergia em certos indivíduos (CUCU *et al.*, 2012; LAM *et al.*, 2008). A região de absorvância entre 1613 a 1682 cm^{-1} pode sinalizar, portanto, a presença de compostos responsáveis por processos alergênicos. Esse tipo de informação apresenta relevância para determinado grupo de pessoas, cuja sensibilidade às proteínas do leite, pode aumentar com o consumo de leite fluido adulterado com soro de queijo.

Absorções características do soro Minas Frescal (não presentes no espectro do leite cru) corresponderam às faixas de 1720-1760 cm^{-1} , 2840- 2870 cm^{-1} e 2950-2940 cm^{-1} (Figura 48).

Segundo Barbosa (2007), as regiões de 1730 a 1755 cm^{-1} são referentes a absorções do grupamento C=O das ligações ésteres dos triglicerídeos. Santos *et al.* (2013) em pesquisa envolvendo adulteração em leite, atribuíram bandas de absorção próximas a 3000 cm^{-1} ao estiramento da ligação CH de carbonilas de ácidos graxos. Bandas semelhantes às citadas pelos pesquisadores foram observadas nas amostras adulteradas com soro de queijo Minas Frescal, entretanto, não foram observadas no leite cru.

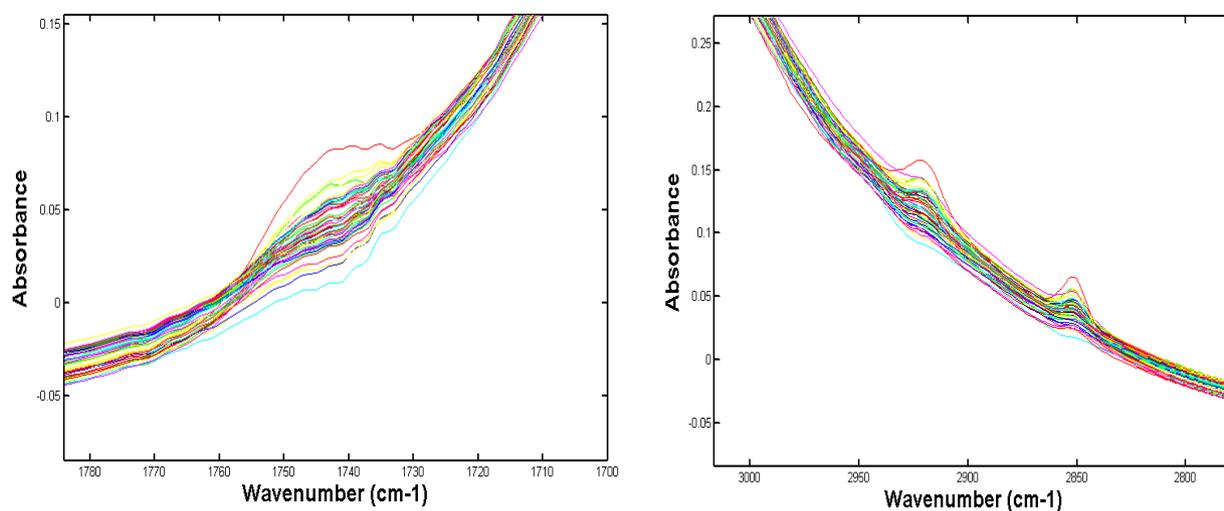


FIGURA 48: Região de absorção 1720-1750 cm^{-1} ; e de 2840 a 2940 cm^{-1}

Interessante observar que, diferentes autores que utilizaram a espectroscopia no infravermelho médio para análise de leite UHT e em pó, mencionaram a presença dessas bandas de absorção de lipídeos, independentemente, da presença do soro no leite. O fato dessas bandas não estarem presentes no espectro do leite cru permitiu

inferir que o aparecimento dessas regiões de absorção esteja relacionado com o tratamento térmico dispensado ao leite.

Faixa de absorção de diferentes ácidos graxos presentes no leite, foi estudada por Soyeurt *et al.* (2006). A pesquisa constatou que no MIR, as principais regiões que implicam uma estimativa do perfil de ácidos graxos se localizam entre 1736 e 1.805 cm^{-1} e entre 2823 e 3016 cm^{-1} . Fagan *et al.* (2009) ao avaliarem qualidade de queijos detectaram absorções no MIR referente a presença de lipídeos nas regiões 1712-1751 cm^{-1} , 1740-1747 cm^{-1} e 2854-2927 cm^{-1} . Esses autores também detectaram bandas de absorção nas regiões de 2854 e 2924 cm^{-1} , atribuídas aos lípidos.

Constatou-se que a maior variação na absorvância do leite promovida pela adição do soro Minas Frescal ocorreu na região de 3100-3500 cm^{-1} (Figura 49). Essa região corresponde predominantemente à forte absorção promovida pelo estiramento do grupo O-H presente na água. Também Fagan *et al.* (2009) observou em sua pesquisa, bandas de absorção de 3047-3703 cm^{-1} atribuídas à água.

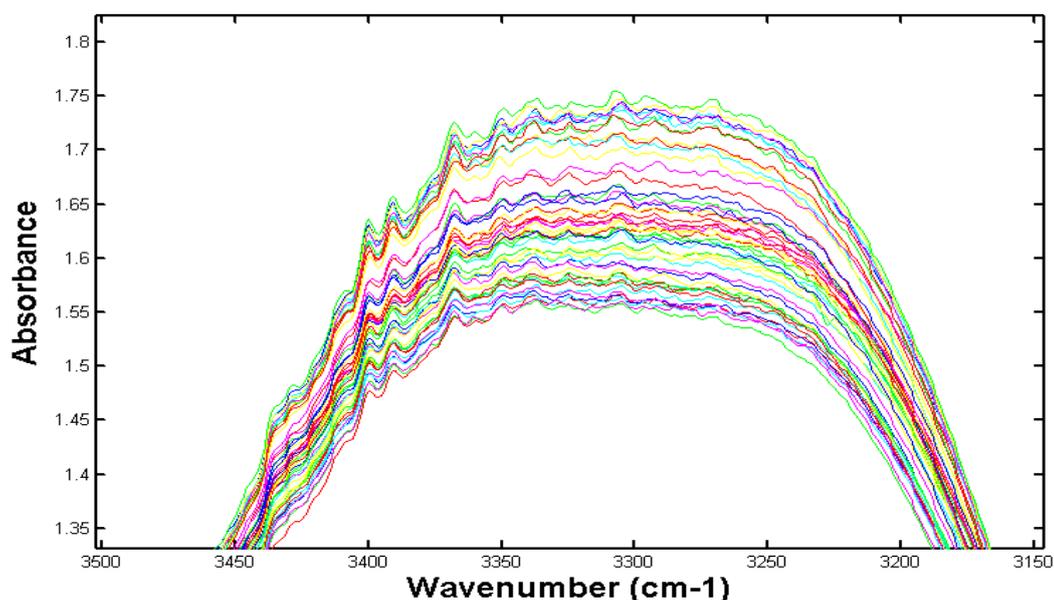


FIGURA 49: Região de absorção de 3150-3500 cm^{-1}

A absorção predominante das amostras adulteradas nessa região está relacionada com o alto teor de água presente no soro de queijo. O soro utilizado nas adulterações artificiais tem a função de carrear água para o leite, fato que caracteriza essa prática, como fraude econômica. O valor da absorvância do leite puro nessa região em relação ao valor da absorvância das amostras adulteradas com soro de

queijo demonstra que essa região pode ser considerada como referência para confirmação de adulterações do leite com água ou soros de queijos.

Para verificar se a fraude econômica é devida ao soro ou a água, faz-se necessário analisar as demais regiões do espectro. Nessa mesma região, onde ocorre a absorção da água, observou-se a presença de várias pequenas bandas. Com base na literatura, atribuiu-se essas bandas a grupos NH em solução diluída nas formas *cis* e *trans* e ao estiramento simétrico e assimétrico do NH₂ (BARBOSA, 2007).

5.2 Resultados referentes ao modelo PLS

Vários estudos comprovam que a adição de diferentes tipos de adulterantes em leite pode ser detectada por diferentes métodos espectroscópicos. Porém, o preparo das amostras, a sua conservação, o tipo de adulterante utilizado, o pré-tratamento dos espectros, a escolha da região espectral para análise e a escolha do método quimiométrico, proporcionam diferentes capacidades de detecção dessas adulterações.

Os níveis de sensibilidade para a detecção de diferentes substâncias fraudadoras podem, portanto, sofrer variações significativas de acordo com o modelo escolhido e o delineamento do experimento (ZHANG *et al.*, 2014; CASSOLI *et al.*, 2011; SANTOS E PEREIRA-FILHO, 2013). O fato da adição do soro no leite promover alterações ao mesmo tempo em diferentes regiões de absorção ao longo do espectro (picos), justificou a necessidade de se tratar quimiometricamente, todo o espectro.

Os resultados do modelo PLS foram obtidos por meio de 6 variáveis latentes, definidas pelos valores de RMSEC e RMSECV, conforme Figura 50. A capacidade de recuperação ou captura dos dados do modelo é que define o número de variáveis latentes necessárias para explicar a variâncias dos dados do modelo.

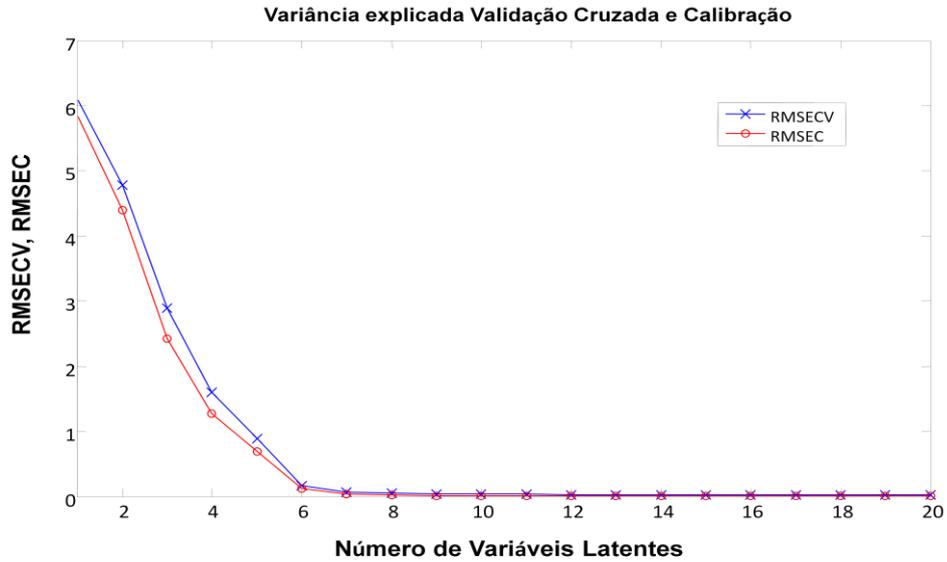


FIGURA 50: Variância explicada pelo modelo definida pelos valores de RMSEC E RMSECV

À medida que se aumenta o número de variáveis latentes, maior é a capacidade de explicar a variância dos dados e menor tende a ser o erro médio. Por meio da Figura 51, pode-se constatar que as 6 VL foram capazes de explicar (capturar), cerca de 99% da variância dos dados do modelo.

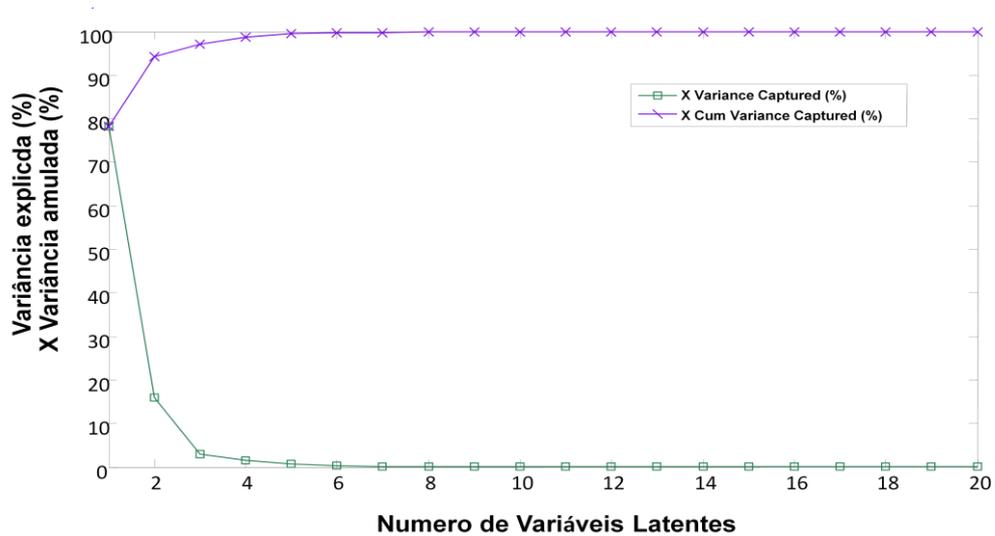


FIGURA 51: Variância explicada pelo modelo em função do número de variáveis latentes

Em função do conjunto de amostra ter sido adequadamente manuseado em laboratório, ou seja, mediante condições controladas, não foram detectadas amostras anômalas (*outliers*).

A Figura 52 apresenta os valores de *leverage* e Q residual das amostras do grupo de calibração. Pôde-se observar que nenhuma das amostras foi considerada como *outlier* na região de absorção no infravermelho definida para o modelo, com limites de confiança de 95%.

Observou-se que apesar de algumas amostras (5, 41 e 43) terem apresentado maior Q residual, não apresentaram alto *leverage*, e por isso não se configuraram como *outliers*.

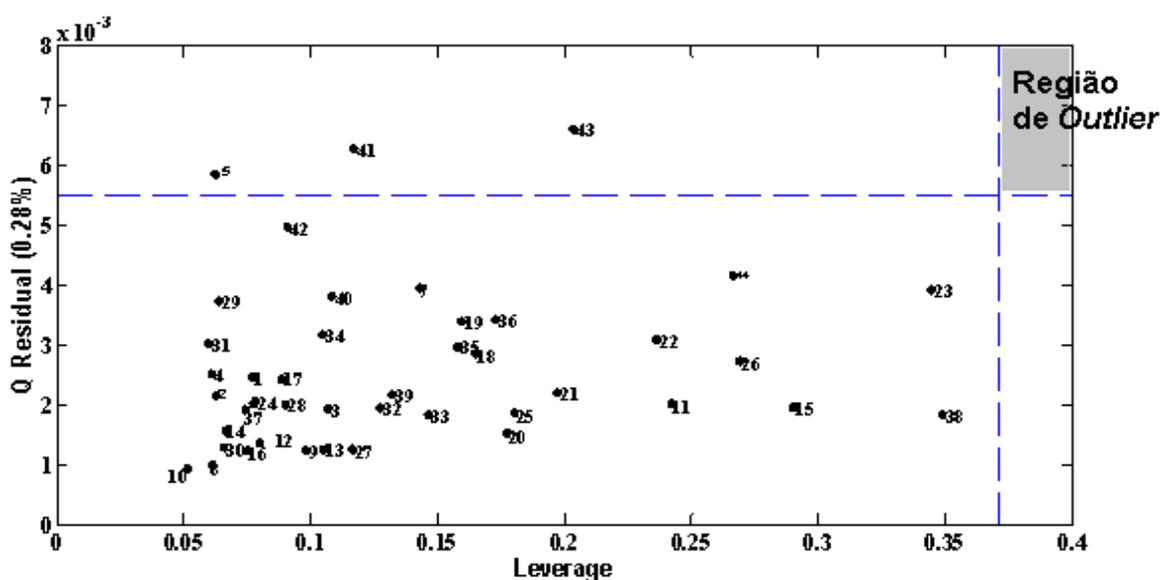


FIGURA 52: Valores de *leverage* e Q residual das amostras do grupo de calibração com intervalo de confiança de 95%.

A eficiência dos modelos de calibração multivariada é avaliada pelo cálculo dos valores do erro quadrático médio (RMSE) e erro relativo (E). Erros quadráticos médios são calculados a partir das amostras de calibração (RMSEC), das amostras validação cruzados (RMSECV) e das amostras de previsão (RMSEP). Normalmente, o erro quadrático médio é expresso em termos percentuais e demonstra os resultados para cada etapa do modelo (calibração, validação e previsão). Em cada etapa, definida com o mesmo número de variáveis latentes, um valor percentual de erro é obtido.

Os resultados dos cálculos dos erros padrão para calibração, validação e previsão do modelo obtido com 6 variáveis latentes foram: RMSEC (%) = 0,116; RMSECV(%) = 0,161 e RMSEP (%) = 0,184.

O modelo PLS exibiu baixos valores percentuais para os erros padrão quadráticos. A diferença entre os valores dos erros do grupo de calibração (RMSEC) e do erro do grupo de *cross* validação (RMSECV) foi pequena, conforme exigido na norma ASTM E1655-05. O valor do RMSEP foi maior que os de RMSEC, não ultrapassando 3 vezes o valor de RMSEC. O RMSEP é considerado um parâmetro robusto, uma vez que é estimado a partir de um conjunto externo de amostras de validação. Esses resultados demonstram um excelente ajuste do modelo, tanto para o conjunto de calibração como de validação.

Por meio da Figura 53 pôde-se observar o que o valor do erro de previsão decresce em função do número de variáveis latentes, ou seja, quanto menor o erro, maior o número de dados recuperados pelo modelo. A correlação entre o menor erro e a maior variância explicada, é que define o número de variáveis latentes ideal para o modelo.

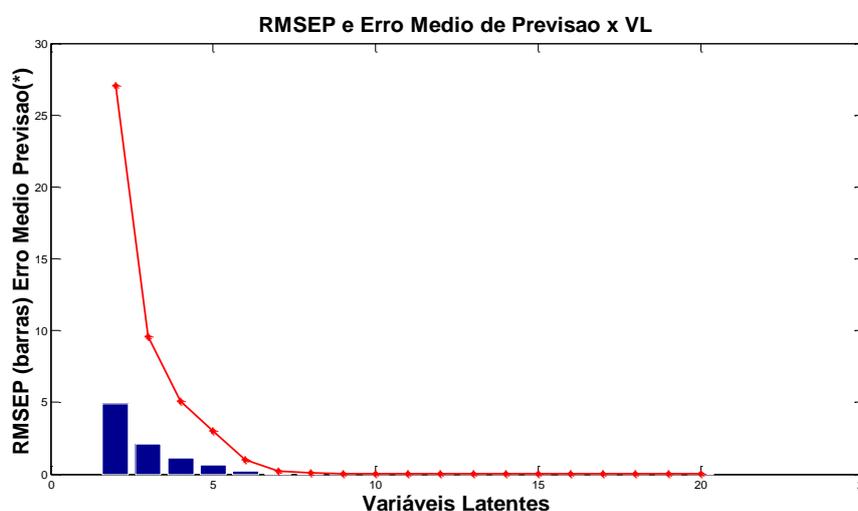


FIGURA 53: Relação entre o erro médio de Previsão e o número de Variáveis Latentes

O modelo PLS construído com seis variáveis latentes obteve um bom ajuste entre os valores previstos e medidos, com o valor de $R=0,998$. O coeficiente de correlação simbolizado pela letra R corresponde à avaliação da dispersão entre as variáveis X e Y e o valor encontrado, deve ser inferido no intervalo: $-1,0 \leq R \leq 1,0$. Quanto mais próximo de 1 for o valor obtido, maior a linearidade da correlação entre as variáveis e melhor é o modelo. Para que se considere um modelo como robusto, o

valor de R (coeficiente de correlação) deve ser próximo a 1, os valores de RMSEC e RMSECV devem ser os mais baixos possíveis e deve ter uma alta variância explicada (próximo a 100 %) (BADDINI *et al.*, 2014).

O gráfico da Figura 54 apresenta os valores de R que correlacionaram os valores reais de concentração das adulterações com os valores previstos pelo modelo tanto para calibração (R= 0,998) como para previsão (R=0,996). O algoritmo PLS do modelo foi capaz de correlacionar os valores de absorvância obtidos pelo FTIR- MIR para cada amostra, com as proporções das concentrações de soro de queijo Minas Frescal nas adulterações artificiais possibilitando a construção de uma curva de calibração que foi capaz de prever a proporção de soro nas amostras, e principalmente, em novas amostras.

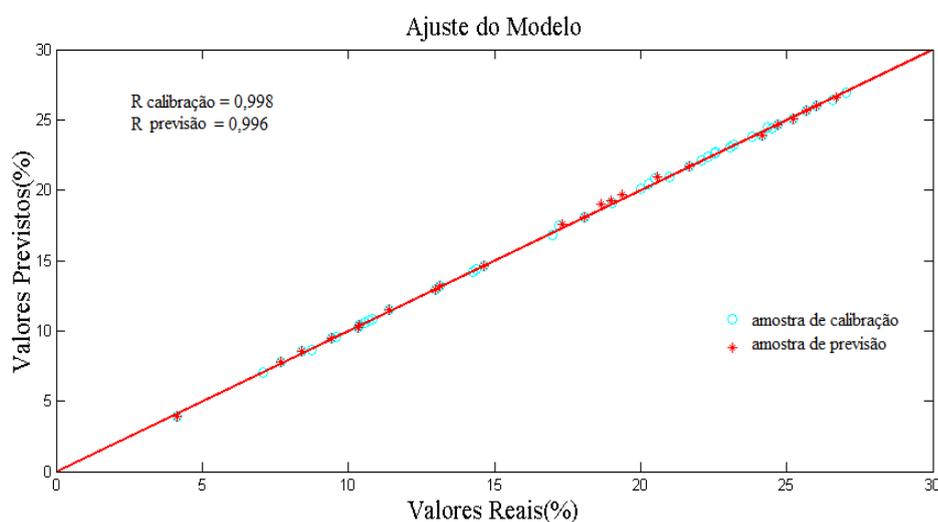


FIGURA 54: Correlação linear do modelo PLS para amostras de calibração e de previsão

Pesquisa desenvolvida por Santos *et al.*, (2013) visando rápida detecção de adulteração em leite por meio do FTIR-MIR, envolveu a realização de diferentes tipos de adulterações artificiais em lotes de leites adquiridos em supermercados. Adulterações com soro foram realizadas numa faixa de concentração correspondente a 1,87 e 30% (p/v). Diferentemente da metodologia utilizada para desenvolvimento do modelo PLS para detecção de adulteração com soro de queijo Minas Frescal em leite cru, os autores realizaram pré-tratamento das amostras para remoção da gordura do leite a fim de minimizar interferentes. Apesar desse procedimento, os resultados obtidos pelos autores foram semelhantes aos apresentados neste trabalho. Santos *et al.*, (2013) obtiveram modelo PLS com 6 VLs explicando 97% dos dados, erro de

validação cruzada de 0,191% e de previsão de 0,233%. Os coeficientes de correlação apresentados pelos autores foram de 0,96 para o modelo de calibração e também de 0,96 para o modelo de validação. A comparação dos resultados apresentados pela pesquisa dos autores citados e pela desenvolvida nesse trabalho demonstrou a acurácia do modelo PLS desenvolvido para detecção de soro de queijo Minas Frescal em leite cru.

Os resultados obtidos para as figuras de mérito confirmaram a veracidade e a linearidade do modelo. A veracidade do modelo pôde ser constatada por meio dos baixos valores percentuais dos erros obtidos para os modelos de calibração (RMSEC), validação cruzada (RMSECV) e previsão (RMSEP) correspondentes a 0,116; 0,161 e 0,184 respectivamente. Esses valores são considerados aceitáveis e eficientes para a detecção de adulterações de leite cru com soro de queijo Minas Frescal em concentrações entre 6,00-28,00% (p/p).

A linearidade correspondeu ao ajuste do modelo PLS e avaliou a correlação entre os valores reais (adulterações artificiais das amostras) e os valores previstos (calculados pelo modelo) a partir da predição. A linearidade de 0,99 obtida pelo modelo alcançou o valor muito próximo a 1,0 (máxima linearidade).

Os parâmetros relacionados à seletividade, sensibilidade, sensibilidade analítica e erro sistemático (bias) foram calculados em função da qualidade do sinal analítico. O baixo valor para seletividade (0,01) correspondeu à fração do sinal original que foi aproveitado pelo modelo. Esse resultado demonstrou que não houve perda de sinal analítico decorrente da sobreposição de sinais da propriedade de interesse e de interferentes (BRAGA E POPPI, 2004).

A sensibilidade de 0,11 correspondeu à fração do sinal incrementada a cada unidade de concentração. Por outro lado, a sensibilidade analítica de 38,67 correspondeu à relação entre a sensibilidade e o desvio padrão estimado do ruído instrumental obtido por meio de nove espectros do leite cru (idênticos) isentos do soro de queijo Minas Frescal. O valor inverso deste parâmetro (0,025) expressou a diferença em concentração mínima que pode ser discernível pelo modelo, considerando o ruído instrumental, como a única fonte de erro (BRAGA E POPPI, 2004; SOUZA *et al.*, 2014).

O valor de "bias" usado para investigar a presença de erros sistemáticos (não aleatórios) no modelo PLS foi de 0,04. O valor do desvio padrão (SDV) que caracterizou a dispersão dos valores preditos de adulteração pelo conjunto de

validação composto por 24 amostras foi de 0,18. O SDV demonstrou nível aceitável de dispersão e boa concordância do modelo de previsão.

Em conformidade com a norma ASTM E1655-05, os erros sistemáticos do modelo foram avaliados usando teste t para amostras de validação com 95% de confiança e graus de liberdade igual ao número de amostras de predição (ASTM, 2012). Após o cálculo de t_{bias} foi constatada a baixa influência dos erros sistemáticos no modelo, uma vez que, o $t_{calc} (1,33) < t_{critc} (2,06)$ para o grau de liberdade igual a 24.

O delineamento experimental e os procedimentos analíticos contribuíram significativamente para a obtenção de dados consistentes. O número de amostras do modelo para calibração e validação foram superiores ao que é determinado pela ASTM – E 1655 de 2012. A conferência da calibração do equipamento assim como a limpeza do compartimento da ATR para leitura de cada amostra proporcionaram a obtenção de sinais livres de interferentes. A Tabela 18 apresenta uma síntese dos resultados dos parâmetros utilizados para validação do modelo.

TABELA 18: Valores das Figuras de Mérito (FOM) para modelo PLS

Figuras de mérito	Parâmetros	Valor
Veracidade	RMSEC (%p/p)	0,110
	RMSECV (%p/p)	0,160
	RMSEP (%p/p)	0,180
Linearidade	-	0,99
Seletividade	-	0,01
Sensibilidade	-	0,11
Sensibilidade analítica	-	38,67- 0,025
Erro sistemático (Bias)	-	0,04
Desvio Padrão		0,18
Grau de liberdade		24
t_{bias} calc.		1,33
t_{bias} critico-		2,06

*O modelo PLS foi desenvolvido para adulterações com soro emleite na faixa de 6,0 a 28% (p/p).

6 RESULTADOS E DISCUSSÃO REFERENTES À TÉCNICA FTIR/MIR E MODELO CLASSIFICATÓRIO PLS-DA PARA DETECÇÃO DE ADULTERAÇÃO DE LEITE UHT COM SOROS DE MANTEIGA E DE QUEIJOS

Com base nos resultados dos modelos exploratórios PCA e HCA que constataram por meio do FTIR-MIR a existência de diferenças entre os soros lácteos que podem ser utilizados em adulterações de leite e também nos modelos PLS-DA preliminares desenvolvidos para classificação de soros de queijos e de leiteiro separadamente, modelo PLS-DA mais amplo foi desenvolvido, com todos os espectros das amostras adulteradas com os diferentes soros lácteos.

6.1 Resultados referentes ao perfil espectral

A Figura 55 apresentou o conjunto de 360 espectros (120 amostras em triplicata) de leite adulteradas artificialmente (concentrações de 5, 15, 25 e 35%) com os diferentes tipos de soro de queijo (60 amostras) e de leiteiro (48 amostras). No gráfico do perfil espectral das amostras também foram incluídas 12 amostras de leite UHT integral (puro) provenientes de diferentes lotes.

A composição desses soros lácteos utilizados nas adulterações artificiais de leite UHT integral foram configuradas por meio do perfil espectral gerado pelo FTIR-MIR. Observou-se ao longo do perfil espectral do conjunto de amostras a existência de algumas regiões de absorção mais pronunciadas em relação às demais em virtude das diferentes características dos soros lácteos utilizados nas adulterações do leite UHT integral. Diferentes formas de processamento do leite acarretam soros residuais que possuem componentes não incorporados ao produto processado.

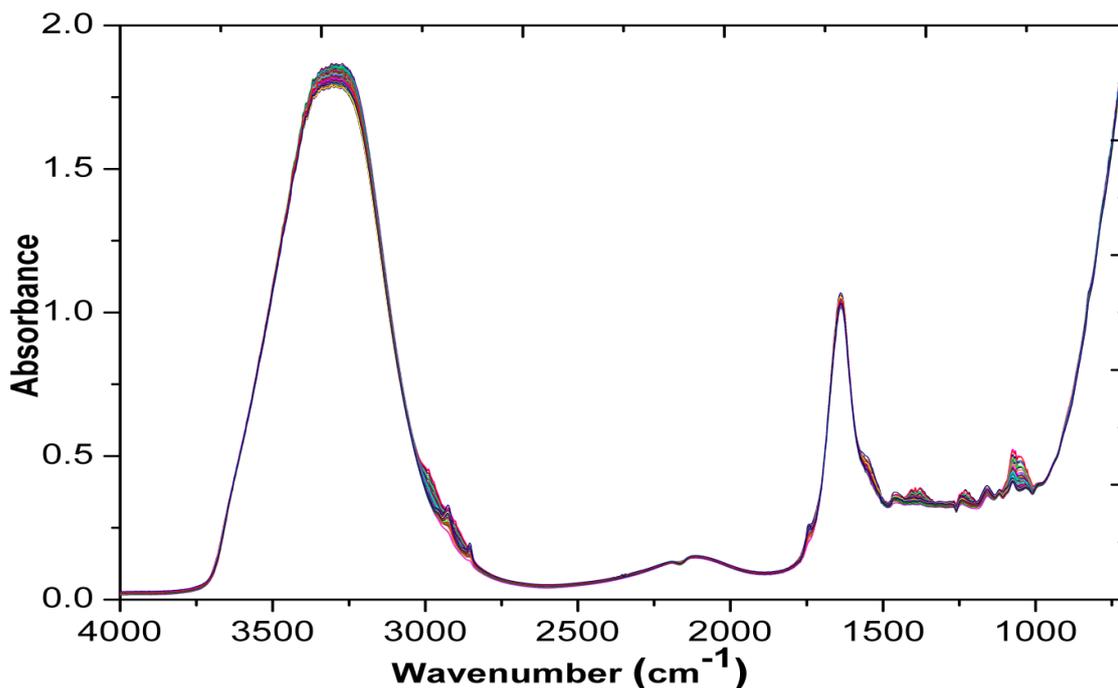


FIGURA 55: Conjunto dos espectros das amostras do modelo PLS-DA contendo adulterações de soros de queijos e leiteiros.

A Figura 56 representou a região de absorção no infravermelho correspondente à faixa de 950 a 1580 cm^{-1} . Na região localizada à esquerda do perfil espectral das amostras observou-se diferentes bandas de absorção atribuídas à presença de: sais livres ou que se encontravam fazendo parte de complexos iônicos ou não ionizados, lactose, amidas III e II.

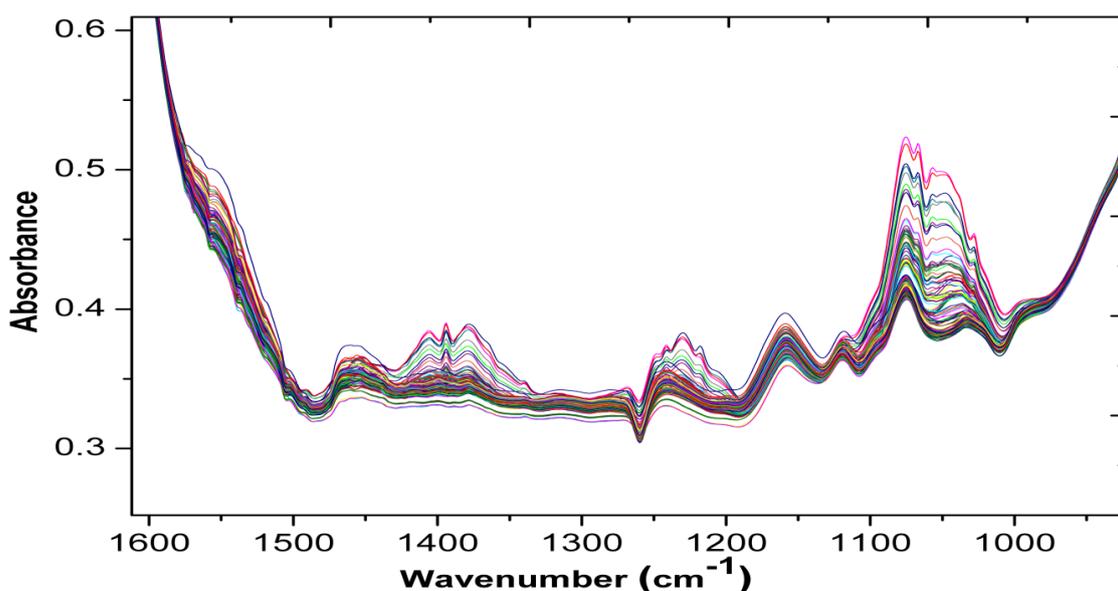


FIGURA 56: Região de absorção referente à faixa de 950 a 1580 cm^{-1}

Os fosfatos e sulfatos, apresentaram absorções nas regiões de 960 a 1150 cm^{-1} (BARBOSA, 2007). Sobrepondo a absorção desses grupos minerais presentes no leite e nos diferentes soros de origem láctea, observou-se absorções promovidas pelas vibrações de estiramento das ligações C-O, C-C e C-O-C da lactose na região de 1030 e 1150 cm^{-1} (AERNOUTS *et al.*, 2011, CARVALHO, 2013; WU *et al.*, 2008).

Os soros de queijos frescos, apresentam teor mais elevado de lactose em relação ao leite, justificando a intensidade dessas bandas em certas amostras, isentas de fermentos ou microrganismos que degradam a lactose (CARVALHO *et al.*, 2007). Os soros utilizados nas adulterações artificiais foram coletados junto às indústrias logo após serem obtidos e utilizados em prazo inferior a 24 horas.

Absorções na região de 1200 a 1250 cm^{-1} foram atribuídas às amidas III (AERNOUTS *et al.*, 2011; CARVALHO, 2013; ETZION *et al.*, 2004; SANTOS, 2013). Absorções em 1020-1200 cm^{-1} atribuídas à presença de álcoois e absorções em 1200-1250 cm^{-1} , decorrentes da presença de grupos fenólicos, apresentaram evidenciadas em algumas amostras.

A presença de citratos naturalmente presentes no leite (absorção em 1354, 1392 e 1437 cm^{-1}) e de outros ions inorgânicos já mencionados nos estudos dos modelos preliminares PLS-DA foi também observada (BARBOSA 2007; CANNATA, 2008; FURTADO, 1991; STUART, 2005).

Bandas de absorção relacionadas às amidas II foram identificadas na faixa correspondente a 1540 a 1580 cm^{-1} . Esses valores foram semelhantes aos encontrados por Etzion *et al.* (2004) e Aernouts *et al.*, (2011)

Outra região do perfil espectral das amostras que compuseram o modelo PLS-DA para adulterações com soros de queijos e leitelhos correspondeu às absorções situadas entre 1620 e 1650 cm^{-1} .

Nessa faixa de absorção o perfil espectral foi comum a todas as amostras. Ela foi relacionada à absorção das amidas I saturadas em estado sólido e ligadas a hidrogênios (estiramento vibracional C=O) (ETZION *et al.*, 2004; INON *et al.*, 2004; PAVIA *et al.*, 2015; YU E IRUDAYARAJ, 2005). Barbosa (2007) também menciona que na faixa de absorvância correspondente a 1630 a 1680 cm^{-1} , as bandas de estiramento das amidas I sobrepõem as bandas de deformação angular das amidas II por apresentarem menor intensidade de absorção.

Em pesquisa realizada para avaliar alterações promovidas pelo processo de homogeneização do leite, os autores constataram presença de ligações das hidroxilas

da água em intervalo de absorvância de 1600 a 1680 cm^{-1} . Essa observação foi possível após análise do leite com e sem a presença de água (AERNOOTS *et al.*, 2011).

Por meio da Figura 57 pôde-se observar o perfil espectral de absorção na faixa entre 1700 e 1780 cm^{-1} . Constatou-se bandas de absorção em praticamente todas as amostras na região de 1735 e 1750 cm^{-1} atribuída ao estiramento do C=O dos ésteres saturados de cadeia aberta que compõem a porção lipídica do leite e dos soros de origem láctea (BARBOSA, 2007).

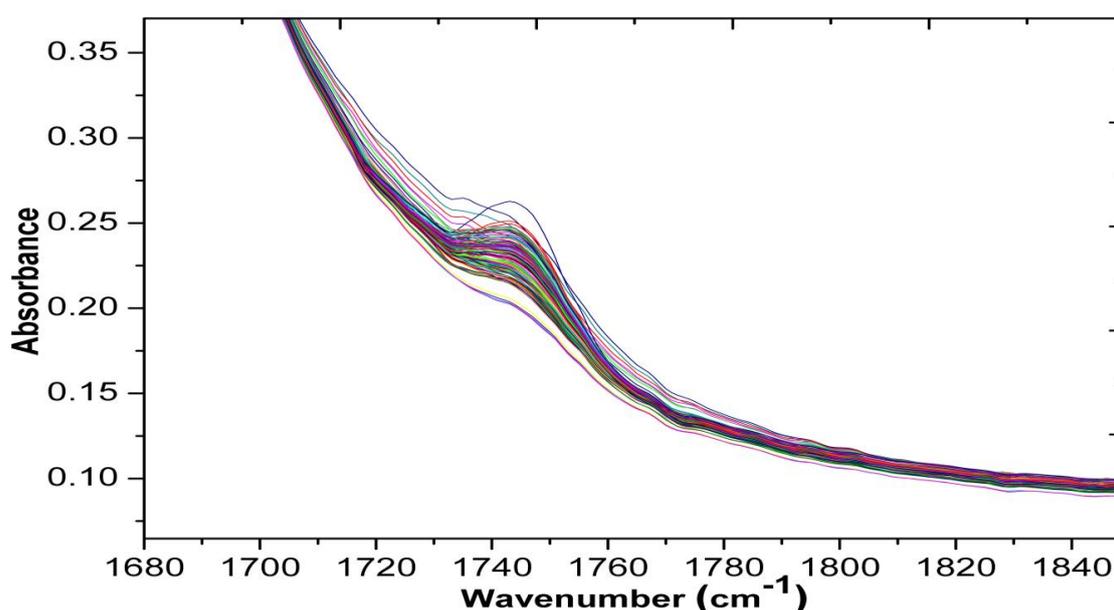


FIGURA 57: Região de absorção referente à faixa de 1700 a 1780 cm^{-1}

Em algumas amostras observou-se absorção ocorrendo a partir de 1700 cm^{-1} . De acordo com Barbosa (2007), ligações C=O dos ácidos carboxílicos apresentam absorção na faixa de 1700 a 1725 cm^{-1} , assim como os grupos cetonas.

A região do perfil espectral com bandas de absorção entre 2850 e 3000 cm^{-1} foram visualizadas na Figura 58. Nessa região observou-se duas bandas mais pronunciadas nos espectros do conjunto de amostras nas regiões 2855 e 2925 cm^{-1} correspondentes às vibrações de estiramento simétrico e assimétrico do CH_2 dos ácidos graxos dos lipídeos.

Pôde-se constatar a presença de bandas largas que se estenderam ao longo dessa região de absorção (2850 e 3000 cm^{-1}). Segundo Barbosa (2007) bandas muito largas na região entre 2500 e 3300 cm^{-1} podem ser atribuídas à ligação O-H dos ácidos carboxílicos e também à presença de grupos fenólicos. Observou-se que o número de

espectros que apresentaram bandas largas, mais especificamente na região entre 2850 e 3000 cm^{-1} foram restritos a algumas amostras. Inferiu-se com base nas considerações de Barbosa (2007) que essas bandas de absorção estariam relacionadas com a composição de corantes presentes nessas amostras.

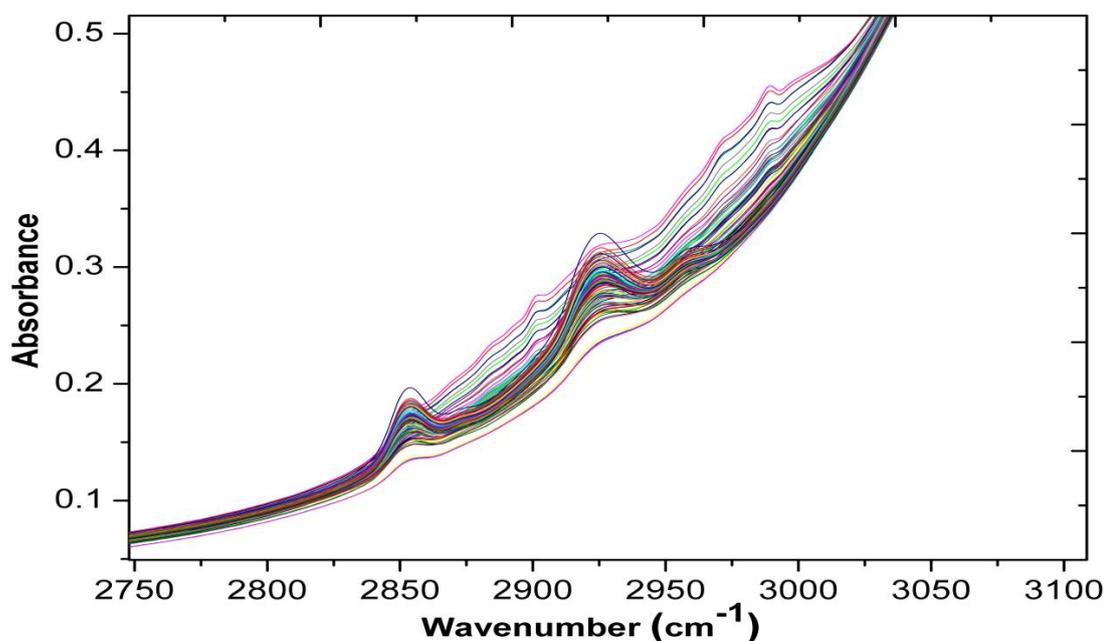


FIGURA 58: Região de absorção referente à faixa de 2850 a 3000 cm^{-1}

Na região de 3000-3100 cm^{-1} dos modelos PLS-DA preliminares foram observados picos pronunciados para três amostras de leiteiros e uma de soro de queijo Prato atribuídas à possível presença de ácido láctico e compostos aromáticos. Entretanto, no conjunto espectral contendo todos os soros lácteos, esses picos não foram evidenciados. Inferiu-se que as absorções de grande quantidade de amostras simultaneamente tratadas dificultaram a visualização de determinadas peculiaridades.

Outra região de absorção cujo perfil espectral foi comum a todas as amostras correspondeu à faixa entre 3200 e 3400 cm^{-1} representada por meio da Figura 59.

Essa região é caracterizada predominantemente pelas vibrações das ligações O-H da água e ligações O-H presentes nas moléculas da gordura do leite (glicerol ligado a ácidos graxos ou na forma livre).

Absorções promovidas pelas vibrações de estiramento das formas *trans* (3270-3370 cm^{-1}) e *cis* (3330-3440 cm^{-1}) das ligações N-H, assim como, das vibrações simétricas NH_2 na forma não associada (3380 – 3420 cm^{-1}) e associada (3320-3360 cm^{-1}) são características dessa faixa espectral.

Esses diferentes tipos de absorções no infravermelho nessa região, justificaram os vários pequenos picos observados no perfil espectral das amostras (BARBOSA, 2007; CARVALHO, 2013; LYNCH *et al.*, 2006; WOLFSCHOON, 1977).

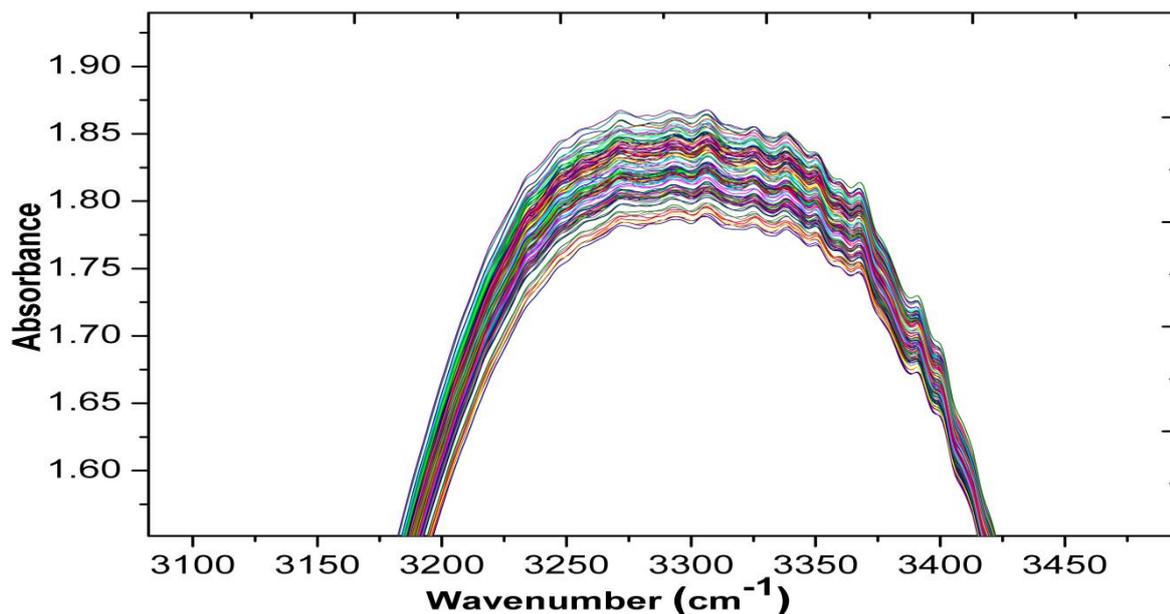


FIGURA 59: Região de absorção referente à faixa de 3200 e 3400 cm^{-1}

6.2 Resultados referentes ao modelo PLS-DA

O modelo PLS-DA obtido com todos os tipos de soros lácteos foi construído com faixa espectral mais ampla do que a utilizada nos modelos PLS-DA preliminares que classificaram separadamente soros de queijos e de manteigas. Expandiu-se a faixa espectral de modelagem para 1000 a 4000 cm^{-1} a fim de proporcionar captação de maior quantidade de dados (sinais).

No modelo PLS-DA mais abrangente utilizou-se como pré-tratamento dos espectros a correção multiplicativa de sinal e a correção ortogonal de sinal. Essa decisão foi tomada após verificar, de forma comparativa, que os resultados do modelo tornaram mais satisfatórios ao serem aplicados esses pré-tratamentos.

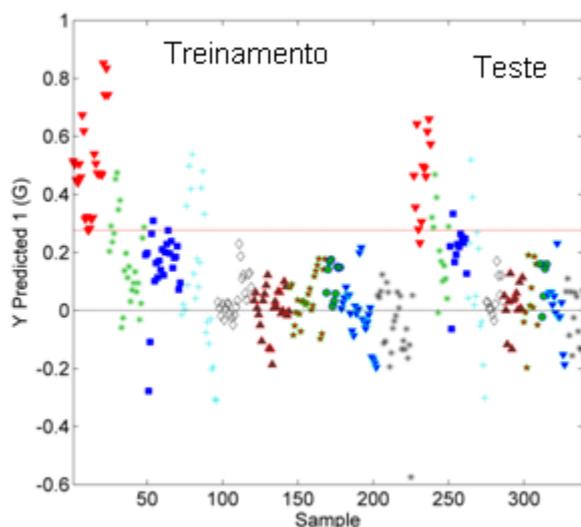
No modelo PLS-DA, assim como no PLS, a explicação da variância em X sofre uma correlação em Y , dessa forma, as variáveis multidimensionais de X (independentes) foram explicadas por meio das variáveis multidimensionais de Y .

O modelo PLS-DA foi obtido com 11 VLs que explicaram 99,94% da variância dos dados em X e 53,97 em Y .

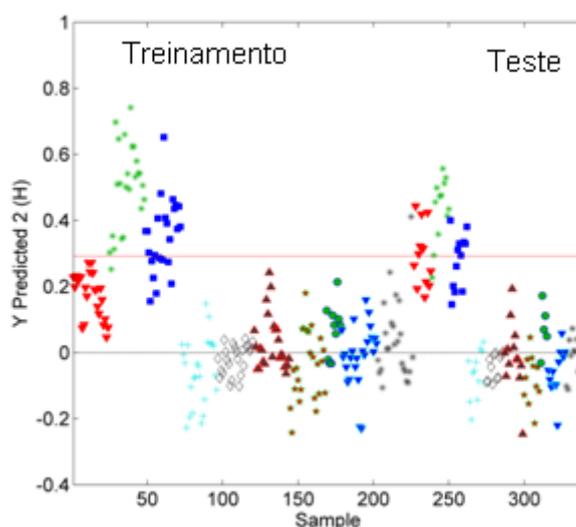
A Figura 60 apresentou as dez classes reconhecidas pelo modelo PLS-DA composto por amostras de leite UHT integral adulteradas artificialmente com cinco tipos de soros de queijos e quatro tipos de soros de manteiga (leitelho). Cento e vinte amostras (360 espectros) foram utilizadas na confecção do modelo. Cada classe foi identificada por letras maiúsculas (G, H, I, J, K, L, N, R, Q e P) representadas por figuras geométricas de cores diferentes.

Nos gráficos apresentados a seguir, as amostras reconhecidas como participantes de cada classe foram dispostas na região superior do gráfico, acima da linha tracejada em vermelho (*threshold*). Abaixo desse limite que define as amostras que compõem cada classe identificada, encontraram-se as demais amostras do modelo.

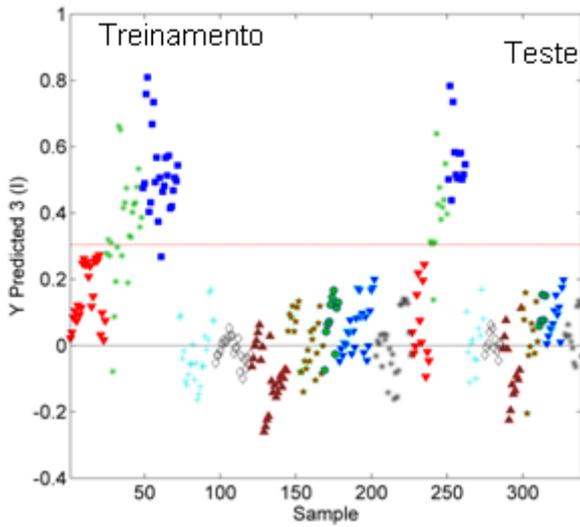
Do lado esquerdo da região superior dos gráficos de classificação, foram situadas as amostras pertencentes ao conjunto de treinamento e do lado direito, as do conjunto teste (predição).



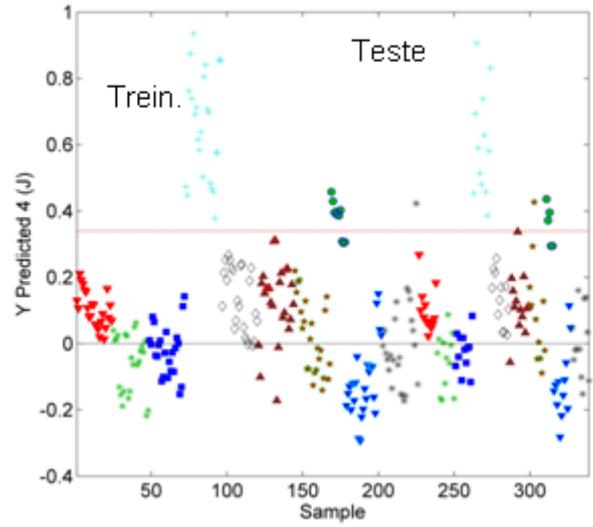
▼ Classe G: Amostras de leite UHT com soro de queijo Minas Frescal



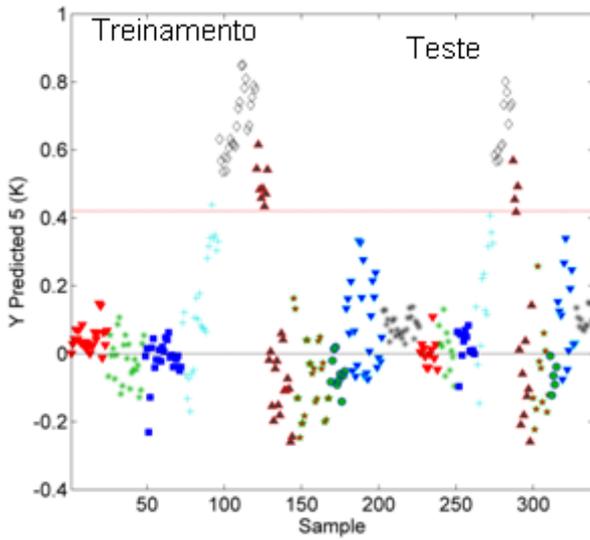
* Classe H: Amostras de leite UHT com soro de queijo Coalho



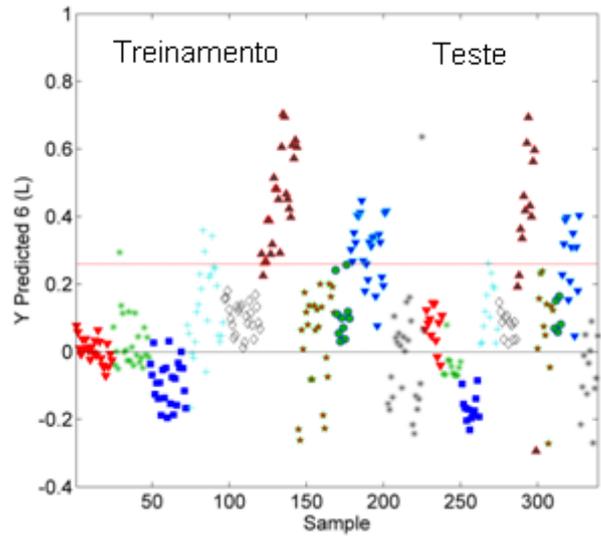
■ Classe I: Amostras de leite UHT com soro de queijo Minas Padrão



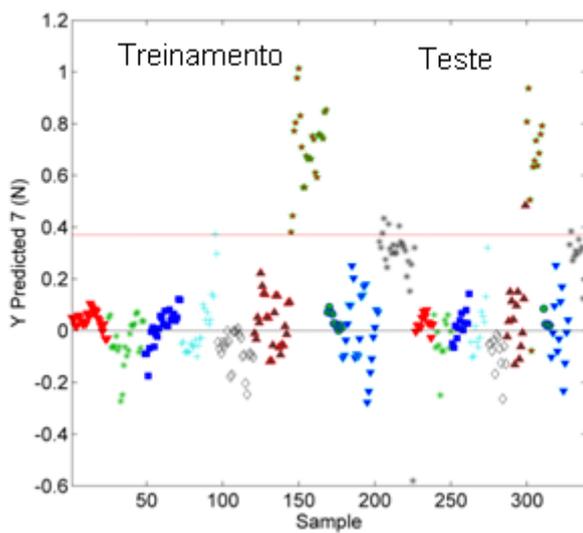
+ Classe J: Amostras de leite UHT com soro de queijo Muçarela



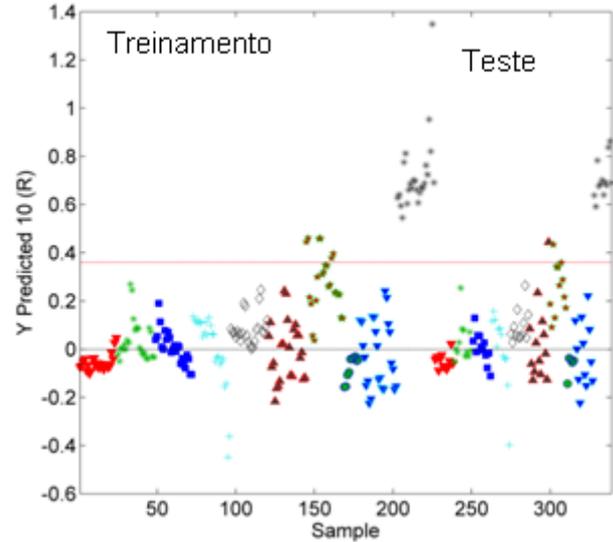
◇ Classe K: Amostras de leite UHT com soro de queijo Prato



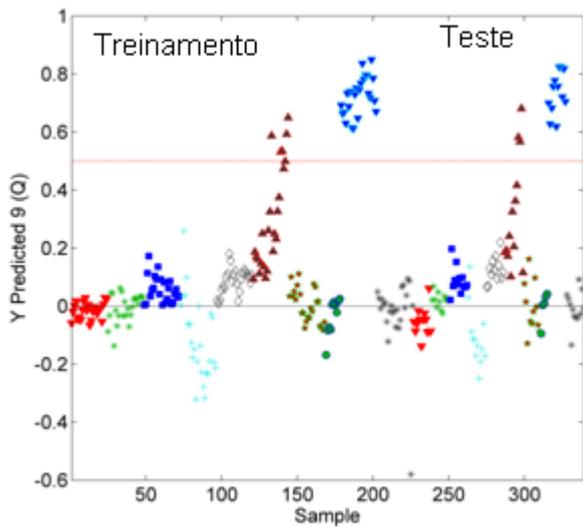
▲ Classe L: Amostras de leite UHT com leite lho mant. comum maior acidez



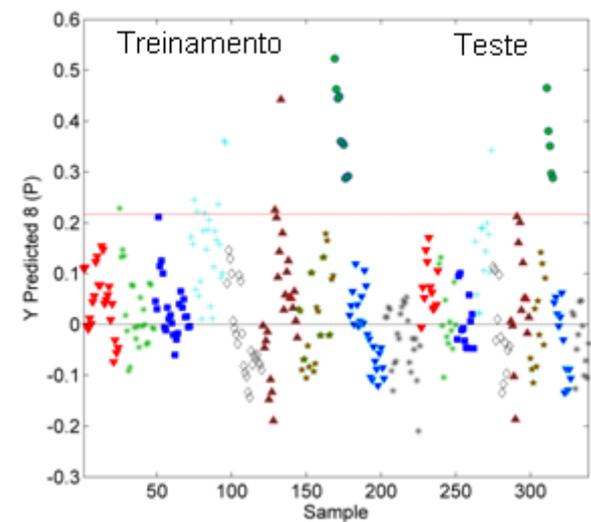
★ Classe N: Amostras de leite UHT com leiteinho mant. extra diluída com água



✱ Classe R: Amostras de Leite UHT com leiteinho mant. extra diluída com leite desnatado



▼ Classe Q: Amostras de Leite UHT com leiteinho mant. comum menor acidez



● Classe P: Amostras de leite UHT puro (isento de adulteração)

FIGURA 60: Gráficos das classes reconhecidas pelo modelo PLS-DA para adulterações de leite com soros de queijos e leiteinhos

Todas as amostras de leite adulteradas com soros lácteos foram adequadamente classificadas pelo modelo PLS-DA. O modelo também demonstrou robustez ao prever novas amostras.

Por meio da observação dos gráficos das classes constatou-se a existência de similaridade entre as amostras adulteradas com soro de queijo Minas Padrão e soro de queijo de Coalho.

Observou-se também semelhança entre amostras adulteradas com soro de queijo Prato, leite de manteiga comum com maior acidez e leite de manteiga comum com menor acidez. A semelhança entre essas amostras foi justificada pela presença de compostos aromáticos proveniente de processos de maturação e pela presença de corante de urucum que comumente é acidificado ao queijo Prato e também à manteigas, sobretudo, as classificadas como comus.

Em pesquisa desenvolvida por Costa e Chaves (2005) para extração de pigmentos de sementes de urucum por meio do infravermelho médio verificou-se que o espectro do extrato de sementes de urucum apresentou banda larga e forte na região de 3500-2500 cm^{-1} . Nessa região, segundo os autores, o estiramento O-H de álcool (geraniol) sobrepõe a banda de estiramento O-H de ácido. A banda de absorção do O-H do ácido juntamente com a banda forte de estiramento de carbonila de éster α,β -insaturado (1717 cm^{-1}) assemelharam-se a um “ombro” sugerindo a presença de outra carbonila, que caracterizam a presença de bixina. Conforme os autores, o espectro do extrato de urucum apresenta ainda, entre outras, bandas de estiramento C=C (1608, 1600 e 1563 cm^{-1}) e C-H (2856, 2919, 2969 cm^{-1} e 3031 cm^{-1}), C-O de álcool primário (1161 cm^{-1}) e deformação C-H de olefinas (1006 e 964 cm^{-1}).

As amostras de leite UHT integral isentas de adulteração (leite puro), também apresentaram características específicas em relação às demais classes reconhecidas pelo modelo. Essa classe definida pelo modelo é relevante para a predição de novas amostras de leite não adulteradas.

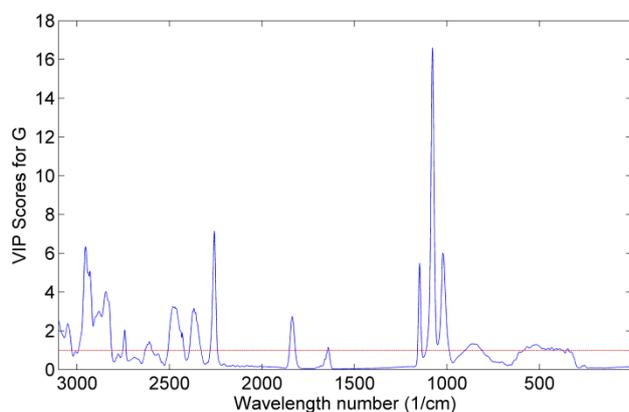
Gráficos da Importância das Variáveis na Projeção dos escores (VIP-scores, *Variable Importance in Projection*) para os modelos de classificação foram também obtidos para cada classe identificada pelo modelo PLS-DA. Os valores de VIPs foram estimados com base na covariância entre as variáveis dependentes e independentes. A inclusão dos pesos (scores) na estimativa do cálculo dos VIPs permitiu não apenas constatar o quanto a variável dependente é descrita, mas também, como as variáveis independentes são importantes para o modelo (BOTELHO *et al.*, 2014; NUNES, 2015).

Uma variável com um valor de VIP próximo ou superior a 1 (um) pode ser considerada importante, enquanto, variáveis com valores de VIP menores do que 1 são menos importantes e podem ser excluídas do modelo (SANTANA, 2015).

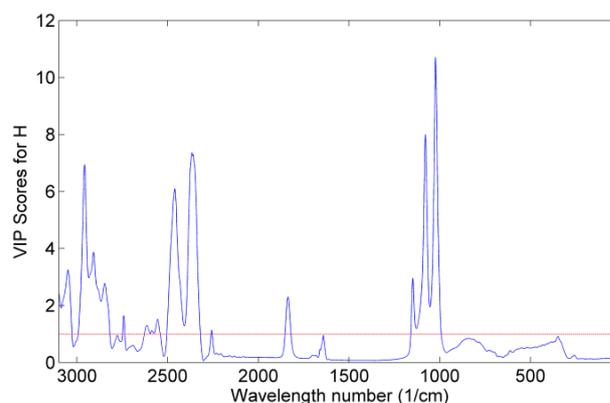
Dez gráficos VIP-scores, apresentados por meio da Figura 61, foram gerados pelo modelo e permitiram identificar quais variáveis dos espectros MIR foram mais relevantes para o reconhecimento e classificação das amostras. Por meio desses gráficos pôde-se constatar quatro regiões ao longo do espectro cujo valor de influência para o modelo foram superiores a um.

Os parâmetros VIPs foram analisados em função dos grupos funcionais responsáveis pela discriminação das amostras em cada classe.

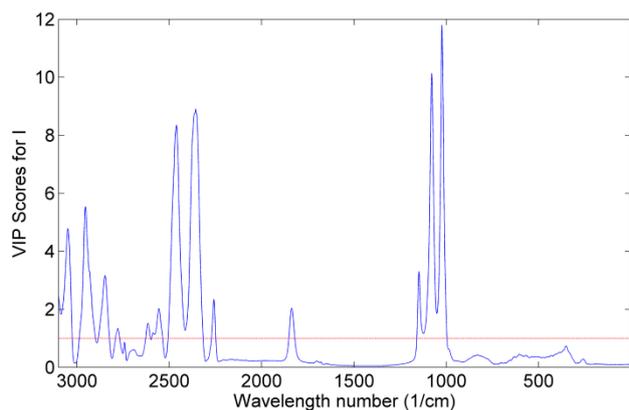
Sete das dez classes identificadas pelo modelo apresentaram valores VIP-scores significativos na região compreendida entre 3000 e 2800 cm^{-1} e nove em região próxima a 1000 cm^{-1} .



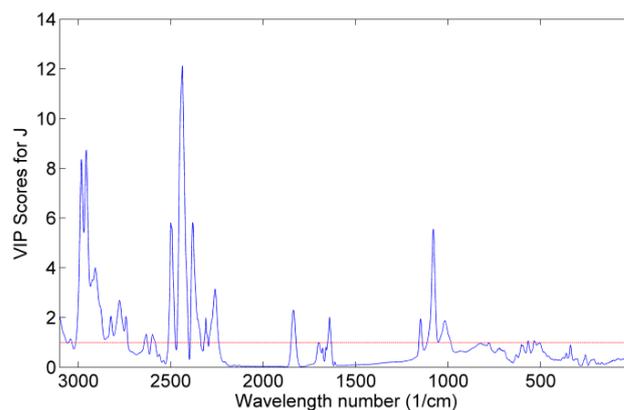
Classe G: Amostras de leite UHT com soro de queijo Minas Frescal



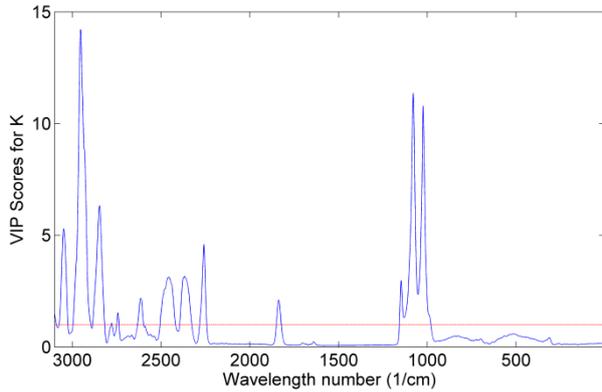
Classe H: Amostras de leite UHT com soro de queijo Coalho



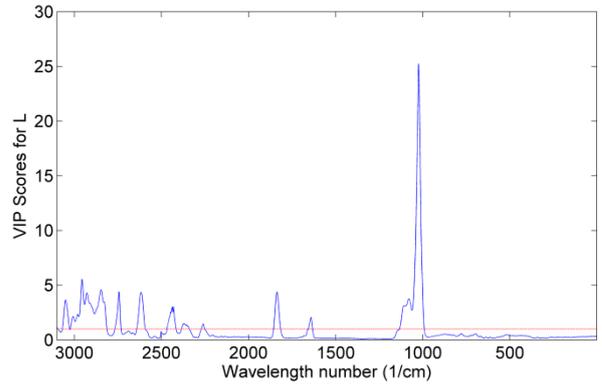
Classe I: Amostras de leite UHT com soro de queijo Minas Padrão



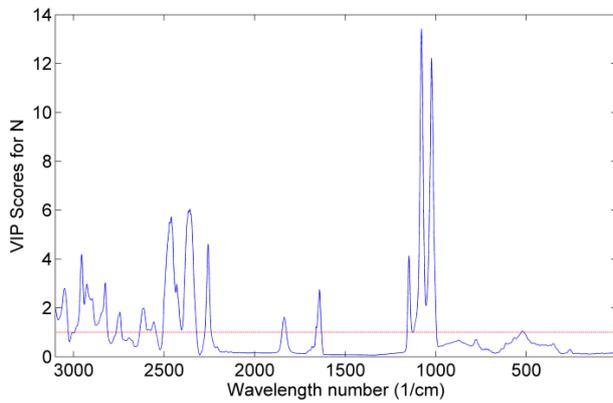
Classe J: Amostras de leite UHT com soro de queijo Muçarela



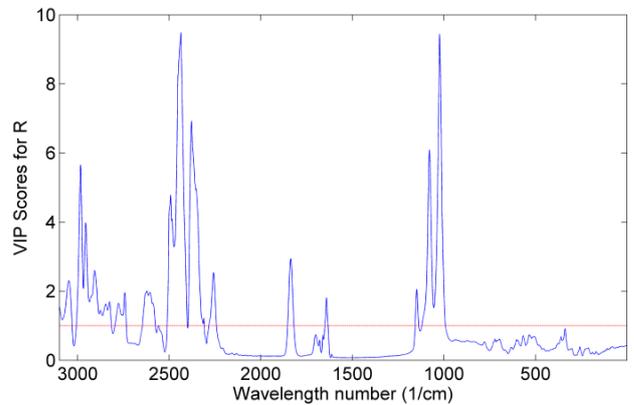
Classe K: Amostras de leite UHT com soro de queijo Prato



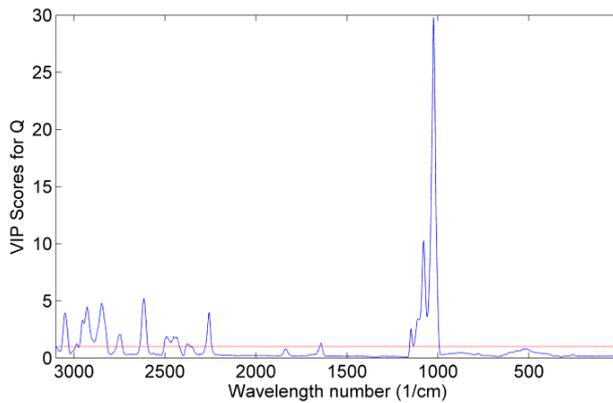
Classe L: Amostras de leite UHT com leiteinho mant. comum maior acidez



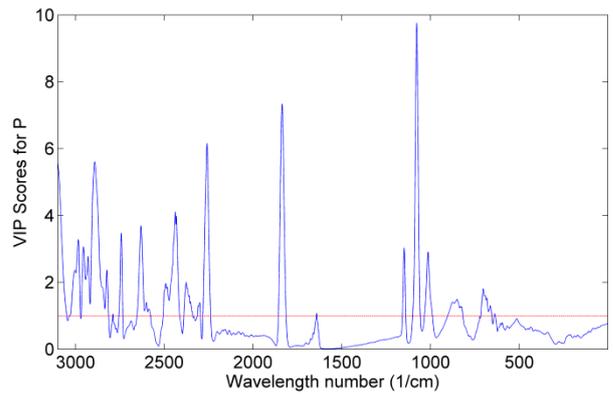
Classe N: Amostras de leite UHT com leiteinho mant. extra diluída com água



Classe R: Amostras de Leite UHT com leiteinho mant. extra diluída com leite desnatado



Classe Q: Amostras de Leite UHT com leiteinho mant. comum menor acidez



Classe P: Amostras de leite UHT puro (isento de adulteração)

FIGURA 61: Modelo de gráfico VIP-scores para cada classe do modelo PLS-DA(B)

A absorvância na região de 3000 e 2800 cm^{-1} foi relacionada às vibrações de estiramento simétrico e assimétrico do CH_2 presentes nos ácidos graxos dos lipídeos (CARVALHO, 2013; BARBOSA, 2007; OLIVEIRA, 2010).

Essa região contribuiu significativamente com grande quantidade de dados diferenciados (sinais espectroscópicos) para a classificação das amostras.

O leite possui diferentes ácidos graxos e diversas substâncias insaponificáveis em sua composição. Os ácidos graxos que compõem os triacilgliceróis (predominância de 99%) possuem diferentes configurações e propriedades físico-químicas. Cada um desses lipídeos (triacilglicerol) apresentam bandas de absorção que se diferenciam e que, mediante fatores de processamento do leite, podem ocasionar a formação de variados compostos químicos.

Outra região de elevada influência demonstrada por meio dos gráficos VIP-scores correspondeu à região de absorção entre 1000 e 1200 cm^{-1} . A configuração das bandas de maior influência nos modelos, nessa região, foi atribuída especialmente à presença de ácido láctico e também à absorção promovida por carboidratos (lactose). A absorção do ácido láctico promove bandas nas regiões entre 1000 e 1600 cm^{-1} , em 1800 cm^{-1} , em 3000 cm^{-1} (pequena banda) e em 3600 cm^{-1} (STUART, 2005):

A Tabela 19 apresentou os principais grupos funcionais discriminantes das classes.

TABELA 19: Grupos funcionais que maior influencia exerceram sob o modelo para definição das classes de acordo com os gráficos VIP-escores

Regiões de maior influência no modelo (cm^{-1})	Grupos funcionais	Atribuições
1150 - 1030	ν (C-O); ν (C-C); ν (C-O-)	Carboidratos (Lactose)
1547 – 1541	δ (N-H)	Amida II (Proteínas)
1660-1650	ν (C=O)	Amida I (Proteínas)
1747	ν (C=O)	Lipídeos (Gordura)
2855 – 2854	ν_s (CH_2)	Lipídeos (Gordura)
2927-2925	ν_{as} (CH_2);	Lipídeos (Gordura)
3250 - 3500	ν_s (OH)	Lipídeos (Gordura)

A absorção relacionada à presença da lactose foi justificada pelo fato dos soros de queijos, assim como, os leitelhos, possuírem quantidade significativa de lactose. O

queijo e a manteiga durante o processamento retêm pequena quantidade de lactose propiciando a liberação desse componente no soro e no leiteiro.

Constatou-se que, apesar da existência de regiões de absorção comuns às classes, os gráficos VIP-scores demonstraram a existência de grande diversidade de sinais específicos que justificaram a alta capacidade de classificação do modelo.

Além da interpretação espectral e gráfica, parâmetros numéricos foram utilizados para análise do modelo e validação por meio de figuras de mérito. A Tabela 20 apresenta esses resultados para o modelo de treinamento.

TABELA 20 - Parâmetros de mérito para validação do modelo PLS-DA contendo amostras do conjunto de treinamento

Classe	TFP	TFN	Acordância	Recall	F-scores
G	0,000	0,000	1,000	1,000	1,000
H	0,019	0,208	0,852	0,821	0,836
I	0,019	0,167	0,857	0,857	0,857
J	0,000	0,000	1,000	1,000	1,000
K	0,033	0,000	0,816	1,000	0,899
L	0,000	0,417	1,000	0,583	0,737
N	0,000	0,083	1,000	0,917	0,957
P	0,009	0,000	0,857	1,000	0,923
Q	0,010	0,000	0,929	1,000	0,963
R	0,010	0,000	0,929	1,000	0,963

Legenda: TFP – Taxa de falso Positivo; TFN – Taxa de falso Negativo.

Os valores ideais para acordância, *recall* e F-score correspondem a 1,0 e o valor mínimo que garante a confiabilidade do modelo em relação a esses parâmetros, corresponde a 0,5. Por outro lado, as taxas de classificação falso positiva e falso negativa que configuram um bom modelo de classificação, devem possuir valores próximos de zero (BOTELHO *et al.*, 2014).

Os valores de F-scores, que representam a razão harmônica entre a acordância e o *recall*, demonstraram a boa qualidade do modelo de treinamento. Valores de F-scores inferiores a 0,5 são considerados insatisfatórios e demonstram a incapacidade do modelo de classificar adequadamente as amostras. No modelo de treinamento o valor de F-score de cada uma das classes ficou acima de 0,5. O menor valor obtido foi

para a classe L (leite adulterado com leite de manteiga comum com maior acidez) igual a 0,737. A obtenção desse valor foi justificada por meio da observação dos gráficos das classes onde se constatou a existência de certa similaridade entre essa amostra e as amostras K (leite adulterado com soro de queijo Prato) e Q (leite adulterado com leite de manteiga comum com menor acidez).

A acordância para cada classe do modelo, relacionada à precisão em termos da capacidade de classificação por meio da repetitividade (percentual de amostras positivas classificadas corretamente sobre o total de amostras classificadas como positivas) também apresentou valores próximos ao ideal (0,1). As taxas de classificações falso positivas e falso negativas foram baixas. Apenas a classe L apresentou valor mais elevado para a taxa de classificações falso negativa, correspondente a 0,417.

Após a construção do modelo de treinamento o modelo PLS-DA foi testado com amostras externas de predição. Essas amostras de leite UHT integral contendo diferentes percentuais de adulterações artificiais com soros lácteos compuseram o conjunto denominado de “teste”. Essa etapa visou avaliar se o modelo seria, de fato, capaz de classificar (reconhecer) novas amostras. Os resultados do modelo de predição e das figuras de mérito calculadas para validação da capacidade preditiva do PLS-DA foram apresentados na Tabela 21.

TABELA 21: Parâmetros de mérito para validação do modelo PLS-DA contendo amostras do conjunto teste (predição)

Classe	TFP	TFN	Acordância	Recall	F-score
G	0,078	0,036	0,704	0,950	0,809
H	0,127	0,026	0,649	0,960	0,774
I	0,099	0,000	0,688	1,000	0,815
J	0,040	0,000	0,800	1,000	0,889
K	0,030	0,000	0,833	1,000	0,909
L	0,078	0,103	0,692	0,857	0,766
N	0,020	0,063	0,867	0,929	0,897
P	0,009	0,000	0,857	1,000	0,923
Q	0,030	0,000	0,833	1,000	0,909
R	0,020	0,000	0,875	1,000	0,933

Legenda: TFP – Taxa de falso Positivo; TFN – Taxa de falso Negativo

Constatou-se que os valores de *F-score* para todas as classes do conjunto teste também obtiveram valores superiores a 0,5 comprovando a capacidade do modelo para classificar novas amostras.

Nesse modelo preditivo, a amostra H apresentou, assim como a amostra L, valores menores para *F-score*. O valor encontrado para a amostra H pôde ser justificado por meio da observação dos gráficos de classificação das amostras. Constatou-se que a amostra H (leite adulterado com soro de queijo Coalho), apresentou certa similaridade com amostras adulteradas com soro de queijo Minas Padrão (amostra I). Os gráficos *VIP-scores* das amostras H e I, também apresentaram bandas de absorção em regiões e intensidades semelhantes. Essas semelhanças ocasionaram baixos valores de acordância (taxa de amostras positivas classificadas corretamente sobre o total de amostras classificadas como positivas) para essas duas classes.

As taxas de classificação falso positivas e falso negativas apresentaram baixos valores para o modelo teste demonstrando a confiabilidade na classificação das amostras.

A adequada classificação das amostras de leite adulteradas com soros lácteos apresentadas pelo modelo com base nos valores de *VIP-scores* e nos parâmetros de mérito apurados (conjunto de treinamento e de predição) contribuíram para comprovar a qualidade e a robustez do modelo PLS-DA desenvolvido pelo presente trabalho. A metodologia proposta, que objetivou desenvolver uma ferramenta para detecção e identificação de diferentes tipos de soros lácteos em adulterações de leite UHT integral por meio do reconhecimento de padrões, demonstrou ser eficiente e aplicável.

7 CONCLUSÕES

7.1 Conclusões gerais

A técnica convencional de crioscopia simples e rápida utilizada para detecção de fraudes econômicas demonstrou ser inadequada para soros de queijos, indicando a necessidade do desenvolvimento de outra técnica também rápida, eficiente e que não necessite de tratamento prévio da amostra.

Os modelos exploratórios HCA e PCA utilizados inicialmente para investigar a existência de diferenças entre os soros de origem láctea que podem ser utilizados para adulterar leites fluidos foram capazes de constatar características próprias em cada soro, ou seja, uma composição química diferenciada. As leituras espectroscópicas das amostras adulteradas e refrigeradas ao longo de dez dias comprovaram que o armazenamento não comprometeu as características químicas específicas de cada tipo de soro. As informações iniciais obtidas pelos modelos PCA e HCA nortearam o desenvolvimento de modelos quimiométricos preditivos com maior capacidade de cruzamento de dados.

Os modelos preditivos preliminares PLS-DA construídos para classificação de amostras de leite UHT integral adulteradas artificialmente com soros de queijos e leitelhos (faixa de concentração de 5 a 35% p/p) foram capazes de definir classes para cada tipo de soro de queijo e de leitelho, separadamente, com níveis de confiabilidade satisfatórios. Durante a etapa de definição dos pré-tratamentos a comparação dos resultados dos modelos, com e sem, a ortogonalização dos sinais espectroscópicos, constatou a viabilidade da aplicação dessa ferramenta para o aprimoramento dos resultados nos modelos PLS-DA.

O modelo PLS, testado para avaliar adulteração em leite com apenas um tipo de soro foi capaz de identificar e quantificar a presença do soro de queijo Minas Frescal na faixa de concentração testada (6 a 28% p/p). Os testes estatísticos multivariados aplicados indicaram que diferentes concentrações de soro podem ser determinadas em amostras de leite cru com ótima veracidade. O modelo PLS desenvolvido foi adequadamente validado e apresentou resultados satisfatórios para todos os valores estimados de mérito. A viabilidade da utilização do modelo desenvolvido para detecção e quantificação de adulterações em leite cru com soro fresco foi confirmada. Outros

modelos semelhantes a este podem vir a ser desenvolvidos para quantificação de soro em outras matrizes, como por exemplo, em bebidas lácteas cujo percentual máximo de soro a ser adicionado é definido pela legislação e necessita ser monitorado.

O modelo PLS-DA construído para classificação conjunta de soros de manteigas e de queijos presentes em adulterações de leite UHT integral (faixa de concentração entre 5 e 35% p/p) também demonstrou excelente capacidade classificatória e preditiva comprovada pela validação por meio de parâmetros qualitativos. Para aceitação ou rejeição de amostras de leite adulteradas com soro, bastariam duas classes; uma contendo amostras isenta de adulterações e outra contendo amostras suspeitas. Porém, identificar o tipo de soro utilizado na fraude pode vir a contribuir com o rastreamento da fonte fornecedora do adulterante para que intervenção seja realizada. Possibilita ainda, prever sintomas que podem ocorrer em pessoas sensíveis a determinados componentes dos soros utilizados nas adulterações.

Uma revelante constatação correspondeu ao fato dos modelos PLS e PLS-DA terem permitido a detecção da adulteração, independentemente, da presença do glicomacropéptido utilizado convencionalmente como marcador para detecção de adulterações de leite com soro.

7.2 Conclusões finais

Os experimentos realizados comprovaram que o FTIR-MIR juntamente com modelos quimiométricos são capazes de fornecer importantes resultados que se traduzem em relevantes informações norteadoras de pesquisas e de metodologias para detecção de adulterações. As análises espectrais das amostras que compuseram os modelos permitiram conhecer as regiões de absorvância do leite cru e UHT integral, bem como as regiões relacionadas às alterações decorrentes de adulterações com soros lácteos (leitinhos e soros de queijos). Os métodos propostos apresentaram como importante diferencial a simplicidade nos procedimentos que não requereram reagentes ou solventes que acarretam resíduos químicos para o meio ambiente. Essa característica corresponde a uma nova tendência analítica que deve ser implementada por laboratórios de órgãos fiscalizadores, instituições de pesquisa e empresas de laticínios.

REFERÊNCIAS BIBLIOGRÁFICAS

AERNOOTS B.; POLSHIN E.; SAEYS W.; LAMMERTYN J. Mid-infrared spectrometry of milk for dairy metabolomics: A comparison of two sampling techniques and effect of homogenization. **Analytica Chimica Acta** v. 705, p. 88– 97, 2011.

AIMUTIS W.R. Bioactive properties of milk proteins with particular focus on anticariogenesis. **J Nutr.** 134 (4), p. 989-995, 2004.

ALMEIDA M. R; OLIVEIRA K. S; STEPHANI R.; OLIVEIRA L.F. Application of FT-Raman Spectroscopy and Chemometric Analysis for Determination of Adulteration in Milk Powder. **Analytical Letters**, v. 45 (17), p. 2589-2602, 2012.

ALMEIDA, M.R.; CORREA, D.N.; ROCHA, W.F.C.; SCAFI, F.J.O.; POPPI, R.J. Discrimination between authentic and counterfeit banknotes using Raman spectroscopy and PLS-DA with uncertainty estimation. **Microchemical Journal**, v. 109, p.170- 177, 2013.

AMENDOLA M. C. **Algoritmo genético e espectroscopia no infravermelho – algumas aplicações na indústria cosmética.** São Paulo: Instituto de Química da USP. 2007. 125 p. (Dissertação, Mestrado em Química).

AMORIM S. R.; KLIER A. H.; ANGELIS L. H. Quality control in the pharmaceutical industry: Identification of infrared spectroscopy. **Rev. Bras. Farm.** v. 94 (3), p. 234-242, 2013.

ARAÚJO J. M. **Química dos alimentos: teoria e prática.** Viçosa: Ed. UFV, 5° Ed., 2011. 601 p.

ASTM. Standards Practices for Infrared, multivariate, quantitative analysis, e1655-05: **Anual book of ASTM Standards.** West Conshohocken, 2012.

BADDINI, A. L. Q.; NETTO, A. D. P.; TEIXEIRA, V. L. Metabolômica baseada em RMN e no Método dos Mínimos Quadrados Parciais Discriminante. **Rev. virtual quim.**, v. 6 (2), p. 396-415, 2014.

BARBOSA L. C. A. **Espectroscopia no infravermelho na caracterização de compostos orgânicos.** Viçosa: Ed. UFV, 2007. 189 p.

BERRUETA L. A.; ALONSO-SALCES R. M.; HEBERGER K. Supervised pattern recognition in food analysis–Review. **Journal of Chromatography** n.1158, p. 196–214, 2007.

BOTELHO B. G. **Desenvolvimento e validação de métodos para análise direta de alimentos usando ferramentas quimiométricas, espectroscopia no infravermelho**

e imagens digitais. Belo Horizonte: Departamento de Química da UFMG, 2014. 136 p. (Tese, Doutorado Ciências–Química).

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

BRAGA J. W. B; POPPI R.J. Validação de modelos de calibração multivariada: uma aplicação na determinação de pureza polimórfica de carbamazepina por espectroscopia no infravermelho próximo. **Química Nova**, v. 27 (.6), 2004.

BRANDÃO S.C.C. **Fraude do leite**, In: BARBOSA S.B.P., BATISTA A.M.V.; MONARDES H. III Congresso Brasileiro de qualidade do leite. Recife: CCS Gráfica e Editora, 2008, v.1. p. 193-199.

BRASIL. ANVISA, RDC n. 348 de 2003. **Regulamento Técnico de identidade e Qualidade de soro de queijo, p.1-10. Disponível em:** <<http://www.terraViva.com.br/cliQue/minuta.html>> **Acesso em: 13 de maio de 2014.**

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Portaria nº. 124 de 23 set. 1991. Aprova métodos analíticos qualitativo e quantitativo de detecção de soro em leite. **Diário Oficial da União, Brasília, 20 nov. 1991**, p.26245-26246. 7.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Portaria nº 53 , 11 abril 2013. **Submete a consulta pública o projeto de Instrução normativa e seu anexo que estabelece aos padrões de identidade e qualidade de soro de leite.** Disponível em: <<http://.lex.com.br/legis24319917portarian53de10deabrilde2013>> Acesso em 10 fev 2016.

BRASIL – **Projeções do Agronegócio: Brasil 2014-2015 a 2024-2025.** Ministério da Agricultura, Pecuária e Abastecimento. Brasília, julho de 2015, p. 1-133.

BRASIL^a. Ministério da Agricultura Pecuária e Abastecimento. Instrução Normativa nº 68 de 12 de dezembro de 2006. Oficializa os Métodos Analíticos Oficiais Físico-Químicos, para Controle de Leite e Produtos Lácteos, em conformidade com o anexo desta Instrução Normativa, determinando que sejam utilizados nos Laboratórios Nacionais Agropecuários. **Diário Oficial da União, Brasília, 14 dez. 2006.** p.1-74.

BRASIL. Ministério da Agricultura Pecuária e Abastecimento. Instrução Normativa nº 62 de 29 de dezembro de 2011. Aprova o Regulamento Técnico de Produção, Identidade e Qualidade do Leite tipo A, o Regulamento Técnico de Identidade e Qualidade de Leite Cru Refrigerado, o Regulamento Técnico de Identidade e Qualidade de Leite Pasteurizado e o Regulamento Técnico da Coleta de Leite Cru Refrigerado e seu Transporte a Granel. **Diário Oficial da União, Brasília, 30 dez. 2011**, p.1-24.

BRASIL^b. Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa nº 69, 13 dez. 2006. **Institui critério de avaliação da qualidade do leite in natura, concentrado e em pó, reconstituídos, com base no método analítico oficial físico-**

químico denominado “Índice CMP”. Disponível em: <<http://extranet.agricultura.gov.br/sislegis-consulta/consultarLegislacao.do?>> Acesso em: 03 set. 2014.

BRASIL. Lei nº 8.078, de setembro de 1990. **Código de Defesa do Consumidor**. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/L8078.htm>. Acesso em 8 dez 2016.

BRASIL, Decreto 30.691/52. Regulamento de Inspeção Industrial e Sanitária de Produtos de Origem Animal (RIISPOA). Brasília: Ministério da Agricultura e do Abastecimento. **Diário Oficial da União de 07/07/1952 , Seção 1 , Página 10785**.

BRITO N. M.; AMARANTE O.P.J.; POLESE L.;RIBEIRO M.L. Validação de métodos analíticos: estratégia e discussão. **Pesticidas: R. Ecotoxicol. E Meio Ambiente**, v.13, p.129-146. 2003.

BUSTAMANTE, I. T.; HIERREZUELO, J. L. R. **Estado del Arte del Reconocimiento de Patrones en la Quimiometría**. Havana: CENATAV. 2008. 54 p.

BÜTTEL B.; FUCHS M.; HOLZ B., Freezing point osmometry of milk to determine the additional water content – an issue in general quality control and German food regulation. **Chemistry Central Journal**, v.2 (6), p.1-7, 2008.

BYLESJO M. B.*et. al.* OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. **J. Chemometrics** v. 20, p.341–351, 2006.

CAMPELO R. J. G. B. **Classificação IV: Avaliação de classificadores**, 2015. Disponível em <<http://docslide.com.br/documents/avaliacao-de-classificadores.html> > Acesso em 15 agosto de 2016.

CANNATA M. G. **Síntese e caracterização do complexo formado através da reação entre o dicloreto de difeniletano e ácido cítrico e sua atividade fungicida**. Lavras: Departamento de Química da UFLA. 2008. 51 p. (Dissertação, Mestrado em Agroquímica).

CARVALHO A. M. B. **Análises do leite em pó de vaca e cabra por métodos físicos**. Juiz de Fora: Faculdade de Farmácia da UFJF. 2013. 95 p. (Dissertação, Mestrado em Ciências e Tecnologia de Leite e Derivados).

CARVALHO, B. M. A.; CARVALHO, L. M.; AYRA, L.; ALCÂNTRA, P.; BONOMO, R. C. F. Métodos de detecção de fraude em leite por adição de soro de queijo. **Revista eletrônica de Veterinária**. v. 8 (6), p. 1695-7504, 2007.

CASSOLI L. D.; SARTORI B.; ZAMPAR A.; MACHADO P. F. An assessment of Fourier Transform Infrared spectroscopy to identify adulterated raw milk in Brazil. **International Journal of Dairy Technology**, v. 64 (n. 4), p. 480-485, 2011.

CECCHI H. M. Fundamentos teóricos e práticos em análise de alimentos. 2º ed. rev. Campinas, S.P.: Editora da UNICAMP, 2003 p.

CHAVERON M.; SILVER J. J.; DUPERREX H. Desmineralization of whey. **Societe D'assistance Technique Pour Produits Nestlé S.A.** v.4,(138, 501), p. 1-10, 1979.

CIENFUEGOS F.; VAITSMAN D. S. **Análise instrumental**. Rio de Janeiro: Interciência, 2000. 606 p.

CODEX ALIMENTARIUS. Official Standards. CAC/GL 69. **Guide for the Validation of Food Safety Control Measures**, 2008. Disponível em: < <http://fao.org/fao-who-codexalimentarius/en/>> .Acesso em: 12 dez. 2013.

CODEX ALIMENTARIUS. Official Standards. CAC/RCP 57: **Code of Hygienic Practice for Milk and Milk Products**, 2004. Disponível em: < <http://fao.org/fao-who-codexalimentarius/en/>> .Acesso em: 12 dez. 2013.

CONWAY V.; GAUTHIER S. F.; POULIOT Y. Buttermilk: Much more than a source of milk phospholipids. **Animal Frontiers**, v. 4 (2), p. 44-51, 2014.

CORREIA P. R. M.; FERREIRA M. M. C. Reconhecimento de padrões por métodos não supervisionados: explorando procedimentos quimiométricos para tratamento de dados analíticos. **Quim. Nova**, v. 30 (2), p. 481-487, 2007.

CORTEZ M. A. S., DIAS V.G.; MAIA R. G.; COSTA C.C.A. Physicochemical characteristics and sensorial evaluation of pasteurized milk added with water, cheese whey, 0.9% sodium chlorite solution and 5.0% dextrose solution. **Rev. Inst. Latic. Cândido Tostes**, v.376 (65), p. 18-25, 2010.

COSTA C. L. S.; CHAVES M. H. Extraction of pigments from seeds of *Bixa orellana* L.: an alternative for experimental courses in organic chemistry. **Quim. Nova**, v. 28 (1), p. 149-152. 2005.

CLOSS E. ; SOUZA C. F. V. SOUZA. Avaliação do teor de caseínomacropéptido (CMP) nos leites cru e UAT ao longo do tempo de armazenamento. **Revista Brasileira de Pesquisa em Alimentos**, v.2 (.2), p.111-119. 2011.

CUCU T.; MEULENAER B.; KERKAERT B.; VANDENBERGHE I; DEVREESE B. MALDI based identification of whey protein derived tryptic marker peptides that resist protein glycation. **Food Research International**, v. 47, p. 23-30. 2012.

DAVIS J.; GOADRICH M. **The Relationship Between Precision-Recall and ROC Curves**. Proc. 23rd Int. Conf. On Machine Learning. 2006.

DANTAS H. V.; LIMA M. B. Classificação quimiométrica de vinagres usando espectros UV-Vis. **IFPB**, 1º Congresso Químico do Brasil. João Pessoa. 2010. P 1-5.

ETZION, Y.; LINKER, R.; COGAN, U.; SHMULEVICH, I. Determination of protein concentration in raw milk by mid-infrared fourier transform infrared/attenuated total reflectance spectroscopy. **Journal of Dairy Science**, v. 87 (9), p. 2779-88. 2004.

FAGAN C. C.; CASTILLO M. ; O'CALLAGHAN D. J.; PAYNE F. A.; O'DONNELL C. P. Visible-near infrared spectroscopy sensor for predicting curd

and whey composition during cheese processing. *Sens. & Instrumen. Food Qual.* v.3, p.62–69. 2009.

FERRÉ J., BOQUÉ R., FERNÁNDEZ-BAND B., LARRECHI M.S., RIUS F. Figures of merit in multivariate calibration. Determination of four pesticides in water by flow injection analysis and spectrophotometric detection. *Analytica Chimica Acta.* v. 348, p. 167-175. 1997.

FERREIRA, M. M. C.; ANTUNES, A. M.; MELGO, M. S.; VOLPE, P. L. O. Quimiometria I: calibração multivariada, um tutorial. *Química Nova*, v. 22 (5), p. 724-731. 1999.

FONSECA F. S. **Espectrometria no infravermelho médio e calibração multivariada por PLS para quantificação de biodiesel metílico e etílico de soja, milho e girasol em misturas com diesel.** Uberlândia: Instituto de química, 2011, 77 p. (Dissertação, Mestrado em Química).

FRANCA, A.S.; OLIVEIRA, L. S. Potential Uses of Fourier Transform Infrared Spectroscopy (FT-IR). *Food Processing and Engineering*, cap. 4, p. 211-257, 2011.

FREITAS S. K. B. **Uma Metodologia para Screening Analysis de sucos cítricos utilizando um analisador automático em fluxo-batelada, espectrometria UV-VIS e técnicas quimiométricas.** João Pessoa: Centro de ciências Exatas e da natureza da UFPB. 2006. 55 p. (Dissertação, Mestrado em Química).

FUKUDA S. P. **Estudo de metodologia quantitativa para determinação espectrofotométrica de ácido siálico em leite.** Campinas: Faculdade de Engenharia de Alimentos da UNICAMP, 1994. 142 p. (Dissertação, Mestrado em Tecnologia de Alimentos).

FURTADO M. A. M. **Fraudes em leite de consumo.** In. I Simpósio de Qualidade do Leite e Derivados, UFRRJ – Seropédica, RJ –Ago de 2010.

FURTADO M. M.; LOURENÇO NETO J. P. M. **Tecnologia de queijos.** São Paulo: Dipemar. 1994, 118 p.

FURTADO M. M. **A arte e a ciência do queijo.** São Paulo: Globo, 1991, 297p.

GILLE D. The health aspects of buttermilk components: a review. *ALP Science* n. 540 |, p. 1-22, 2011.

GIOMBELLI C. J. *Avaliação da qualidade microbiológica, físico-química e dos parâmetros enzimáticos de leite pasteurizado e leite tipo B, produzidos no Paraná.* **Semina: Ciências Agrárias**, Londrina, v. 32 (4), p. 1539-1546. 2011.

GONDIM C.S, JUNQUEIRA R.G, SOUZA S.V.C. Tendências em validação de métodos de ensaios qualitativos. *Ver. Inst. Adolfo Lutz* v 70(4), p.433-447. 2011.

GONTIJO, L. C.; GUIMARÃES, E.; MITSUTAKE, H.; SANTANA, F. B.; SANTOS, D. Q.; BORGES NETO, W. Development and Validation of PLS Models for Quantification of

Biodiesels Content from Waste Frying Oil in Diesel by HATR-MIR Rev. **Virtual Quim.**, v. 6 (5), p. 1517-1528. 2014.

GONZALES, L.S.; BEZERRA, V.N.R.I.; RICO, N.; RAYMUNDO, S.N. Elaboração de pães com adição de soro de manteiga. **Ambiência: Revista de Ciências Agrárias e Ambientais**, v. 5 (3), p.391-399. 2009.

GONZÁLEZ, A. G. Critical Aspects of Supervised Pattern Recognition Methods for Interpreting Compositional Data. In: VARMUZA, K. **Chemometrics in Practical Applications**, p. 21-40, 2012.

GROBÉRIO T. S. **Desenvolvimento de uma metodologia para comparação de amostras de sal de cocaína pela determinação de solventes residuais e análises quimiométricas**. Brasília: Instituto de Química da UnB. 2012. 129 p. (Dissertação, Mestrado em Química).

HAIR JR., J. F. *et al.* **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009. 688 p.

HATANAKA R. R. **Desenvolvimento, validação, avaliação da incerteza de medição e análise de custos de método baseado em espectroscopia no infravermelho e análise multivariada para previsão de parâmetros de qualidade de amostras de E85**. Araraquara: Instituto de Química, 2015, 150 p. (Tese, doutorado em Química).

HELLAND I. S.; NAES T.; ISAKSSON T. Related versions of the multiplicative scatter correction method for preprocessing spectroscopic data. **Chemometrics and Intelligent Laboratory Systems**, v. 29 (2), P. 233-241. 1995.

HELPER, G. A.; FERRÃO, M. F.; FERREIRA, C. V.; HERMES, N. Aplicação de métodos de análise multivariada no controle qualitativo de essências alimentícias empregando espectroscopia no infravermelho médio. **Ciência e Tecnologia de Alimentos** v. 26 (4), p. 779-786. 2006.

HÉRBERGER, K. Chemoinformatics- multivariate mathematical – statistical methods for data evaluation. **Medical Applications of Mass Spectrometry**, v. 7, p. 141-169. 2008.

HOLLER, F. J.; SKOOG, D. A.; CROUCH, S. R. **Princípios de análise instrumental**. 6ª edição. São Paulo: Bookman. 2009, 1056 p.

HONORATO F. A.; NETO B. B.; MARTINS M. N.; GALVÃO R. K. H.; PIMENTEL M. F. Calibration transfer in multivariate methods. **Quím. Nova**, v. 30 (5). 2007.

INMETRO. - Instituto Nacional de Metrologia, Normalização e Qualidade Industrial - **Orientações sobre validação de métodos de ensaios químicos**. DOQ-CGCRE-008, Rio de Janeiro, 2011. Revisão 04, 20p.

INÁCIO, M. R. C. **Ferramentas quimiométricas aplicadas à classificações de amostras de leite em pó e quantificação de proteínas.** Natal: Centro de Ciências Exatas e da Terra da UFRN, 2010. 119 p. (Dissertação, Mestrado em Química).

INON, F.A.; S. GARRIGUES AND M. DE LA GUARDIA. Nutritional parameters of commercially available milk samples by FTIR and chemometric techniques. **Analytica Chimica Acta**, v. 513 (2), p. 401-412. 2004.

IUPAC. **Certificado de Acreditação** - Anexo Técnico de Acreditação N.º L0115 - 1, 14 ed., Caparica, 2012.

JOHNSON R. Food Fraud and “Economically Motivated Adulteration” of Food and Food Ingredients. In: **Congressional Research Service**. p.1-45. 2014.

KAROUI R. *et al.* A comparison and joint use of NIR and MIR spectroscopic methods for the determination of some parameters in European Emmental cheese. **Eur Food Res Technol**. v. 223, p. 44–50. 2006.

KARTHEEK M. A.; SMITH A.; MUTHU A. K.; MANAVALAN R. Determination of adulterants in food: A review. **J. Chem. Pharm. Res.** v. 3 (2), p. 629-636. 2011.

KENNEDY, D. G. *et al.* Use of liquid chromatography–mass spectrometry in the analysis of residues of antibiotics in meat and milk. **Journal of Chromatography A**, v. 812 (1–2, 3), p. 77–98. 1998.

KONG J.; YU SHAONING. Fourier Transform Infrared Spectroscopic Analysis of Protein Secondary Structures. **Acta Biochimica et Biophysica Sinica** v. 39(8): p. 549–559. 2007.

LAM H.Y. *et al.* Cow's milk allergy in adults is rare but severe: both casein and whey proteins are involved. **Clin Exp Allergy** v. 38(6), p.995-1002. 2008.

LERAYER, A. L. S. *et al.* Nova legislação comentada de produtos lácteos – revisada e ampliada. São Paulo: Varela, 2002. 327 p.

LIMA I. C. A. **Aplicação de métodos de classificação e calibração multivariada acoplados com técnicas espectroscópicas em amostras ambientais e alimentos.** Rio de Janeiro: Instituto de Química da UERJ. 2011.105 p. (Dissertação de Mestrado em Química).

LORBER A., FABER K., KOWALSKI B.R. Net Analyte Signal Calculation in Multivariate Calibration. **Anal Chem**. v. 69, p. 1620 – 1626. 1997.

LYNCH, J. M. *et al.* Precalibration Evaluation Procedures for Mid-Infrared Milk Analyzers. **Journal of Dairy Science**. v. 89 (7), 2006.

LYRA W. S. *et al.* Classificação periódica: um exemplo didático para ensinar análise de componentes principais. **Quim. Nova**, v. 33 (7), p. 1594-1597. 2010.

MCDERMOTT A, VISENTIN G.; MARCHI M.; BERRY P. D.; , FENELON M. A.; O'CONNOR P. M.; KENNY O. A.; MCPARLAND S. Prediction of individual milk proteins including free amino acids in bovine milk using mid-infrared spectroscopy and their correlations with milk processing characteristics. **J. Dairy Sci.** v. 99, p.3171–3182, 2015.

MAGALHÃES, M.A. **Determinação de fraude de leite com soro de leite pela análise de CMP e pseudo-CMP por cromatografia líquida de alta eficiência em fase reversa com detecção por espectrometria de massa.** Viçosa: Faculdade de Eng. De Alimentos da UFV. 2008. 74 p. (Dissertação, Mestrado em Ciência e Tecnologia de Alimentos).

MAPA (Ministério da Agricultura, Pecuária e Abastecimento). **Operação Leite adulterado I e II.** Brasília, 2014. Disponível em:<<http://www.agricultura.gov.br/animal/noticias/2014/08/operacao-leite-adulterado-i-e-ii>> Acesso em 20 mar. 2015.

MAPA (Ministério da Agricultura, Pecuária e Abastecimento). **Projeção do Agronegócio.** Brasília, 2016. Disponível em: <http://sapc.embrapa.br/arquivos/consorcio/informe_estatistico/Proj_Agronegocio2016_2026_cafe.pdf> Acesso set. 2014.

MAPA (Ministério da Agricultura, Pecuária e Abastecimento). **Programa Nacional de Combate à Fraude no Leite.** Brasília, 2013. Disponível em: <<http://agricultura.gov.br/animal/noticias/2013/05/ministerio-intensificacoes-de-combate-a-irregularidade-no-leite>> Acesso em: 18 nov. 2013.

MARTENS, H.; NAES, T.: Multivariate Calibration. **Spectrochimica Acta Part: A Molecular Spectroscopy** v. 46(10), p. 419-436, 1990.

MARTIN, J. G. P. Resíduos de antimicrobianos em leite – uma revisão. **Revista Segurança Alimentar e Nutricional**, v. 2 (18), p. 80-87. 2011.

MENDES, C. G.; SAKAMOTO, S. M.; SILVA, J. B. A.; JÁCOME, C. G. M.; LEITE, A. I. Análises físico-químicas e de fraude do leite informal comercializado no município de Mossoró-RN. **Ciência Animal Brasileira**. v. 11(2), p. 349-356. 2010.

MINGOTI S. A. Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada. Belo Horizonte: Editora UFMG, 2005. 297 p.

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 (4), p. 118-126. 2012.

MORGANO, M. A. **Aplicação de métodos quimiométricos em análise de alimentos.** Campinas: Instituto de Química da UNICAMP. 2005. 127 p. (Dissertação de mestrado em Química).

MORGANO, M. A.; FARIA, C. G.; FERRÃO, M. F.; BRAGAGNOLO, N.; FERREIRA, M. M. C. Determinação de proteína em café cru por espectroscopia NIR e regressão PLS. **Food Science and Technology**, v. 25 (1), p. 25-31. 2005.

MPF (Ministério Público Federal–Procuradoria da Republica em Minas Gerais) **Operação ouro branco**. Belo Horizonte, 2007. Disponível em: < <http://prmg.mpf.mp.br/imprensa/noticias/criminal/operacao-ouro-branco-desarticula-quadrilhas-que-adulteravam-leite> >. Acesso em: 10 jun 2015.

MULVANY J. L. Indústria da manteiga. Juiz de Fora: ILCT, 1970. 219 p.

NASCIMENTO, J. A.; DIONÍSIO, A. G. G.; NASCIMENTO, E. C. L.; FREITAS, S. K. B.; ARAÚJO, M. C. U. Análise screening de vinhos empregando um analisador fluxo-batelada, espectroscopia UV-VIS e quimiometria. **Química Nova**, v. 33 (2), p. 351-357. 2010.

NEELIMA; SHARMA R.; MANN B. Chemical and functional properties of glycomacropeptide. **Dairy Sci. & Technol**, v. 93, p. 21–43. 2013.

NETO F. F. G. **Classificação de óleos vegetais utilizando voltametria de onda quadrada e métodos quimiométricos**. João Pessoa: Centro de ciências Exatas e da Terra da UFPB. 2008. 117 p.(Dissertação de Mestrado em Química).

NETO W. B. **Parâmetros de qualidade de lubrificantes e óleo de oliva através de espectroscopia vibracional, calibração multivariada e seleção de variáveis**. Campinas: Instituto de Química da UNICAMP, 2005. 142 p. (Tese, Doutorado em Química).

NICOLAOU, N.; XU, Y.; GOODACRE, R. Fourier transform infrared spectroscopy and multivariate analysis for the detection and quantification of different milk species. **Journal of Dairy Science**. v. 93 (12), p. 5651-5660. 2010.

NUNES P. G. A. **Uma nova técnica para seleção de variáveis em calibração multivariada aplicada às espectrometrias UV-VIS e NIR**. João Pessoa: Centro de Ciências Exatas e da Terra da UFPB. 2008. 121 p. (Tese, doutorado em Química).

NURRULHIDAYAH, A.F.; CHE MAN, Y.B.; ROHMAN, A.; AMIN, I.; SHUHAIMI, M.; KHATIB, A. Authentication analysis of butter from beef fat using Fourier Transform Infrared (FTIR) spectroscopy coupled with chemometrics. **International Food Research Journal**. v. 20 (3), p. 1383-1388. 2013.

OANCEA S. Identification of glycomacropeptide as indicator of milk and dairy drinks adulteration with whey by immunochromatographic assay. **Society of Biological Sciences**, v. 14 (1), p. 4146-4151. 2009.

OLIVEIRA, M. C. P. P. **Espectrofotometria no Infravermelho – por metodologia FTIR (Fourier transform infrared): validação da análise do teor de uréia e de outros parâmetros de qualidade do leite**. Belo horizonte: Escola de Veterinária da UFMG. 2011. 70 p. (Dissertação, Mestrado em Tecn. e Insp. de alimentos de origem animal).

OLIVEIRA, R. R. **Aplicação da espectroscopia de infravermelho próximo para a determinação do caseinomacropéptido em leite UAT**. Goiânia: Escola de Veterinária e Zootecnia da UFG, 2010. 77 p. (Dissertação, Mestrado em Ciência Animal).

ORDONEZ, J. A. **Tecnologia de Alimentos**. Porto Alegre: Artmed, 2005. 279 p.

PALERMO G.; PIRAINO P.; ZUCH H. Performance of PLS regression coefficients in selecting variables for each response of a multivariate PLS for omics-type data. **Advances and Applications in Bioinformatics and Chemistry** v.2, p. 57–70. 2009.

PANERO J.S. **Determinação de proteína, óleo e umidade por espectroscopia NIR em grãos de soja do estado de Roraima**. Boa Vista: Departamento de Química da UFRO, 2007. 77 p. (Dissertação, Mestrado em Química).

PARISOTTO G.; FERRÃO M. F.; FURTADO J. C.; MOLZ R. F. Determination of amoxicillin content in powdered pharmaceutical formulations using DRIFTS and PLS. **Rev. Bras. Cienc. Farm.** vol.43(1). 2007.

PASQUINI C. Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications, **J. Braz Chem Soc.**, v.14, 198-219. 2003.

PATACA L. C. M. **Análises de mel e própolis utilizando métodos Quimiométricos de Classificação e Calibração**. Campinas: Departamento de Química Analítica da UNICAMP, 2006. 102 p. (Tese, Doutorado em Química).

PAVIA D; KRIZ G. S.; LAMPMAN G. M. **Introdução a espectroscopia**. Trad. Pedro Barros. – Sao Paulo: Cengage Learning, 2015. 692 p.

PEERBHAY K. Y.; MUTANGA O.; ISMAIL R. Commercial tree species discrimination using airborne AISA Eagle hyperspectral imagery and partial least squares discriminant analysis (PLS-DA). **Journal of Photogrammetry and Remote Sensing** v. 79, p.19–28. 2013.

PEREIRA, D. B. C.; SILVA, P. H. F.da.; COSTA JÚNIOR, L. C. G.; OLIVEIRA, L. L. de. **Físico-química do leite e derivados: métodos analíticos**. 2. ed. Juiz de Fora: EPAMIG, 2001. 234 p.

PERRY K. S. P. Queijos: aspectos químicos, bioquímicos e microbiológicos. **Quim. Nova**, v. 27 (2), p. 293-300. 2004.

PINTO C. R.; TRYGG J.; GOTTFRIES J. Advantages of orthogonal inspection in chemometrics. **J. Chemometrics** v. 26, p. 231–235. 2012.

PINTO F. A. **Metodologia de espectroscopia no infravermelho para análises de soros provenientes da fabricação de queijos minas padrão e prato**. Belo Horizonte: Escola de veterinária da UFMG, 2010. 45 p. (Dissertação de Mestrado em Tecnologia e inspeção de produtos de origem animal).

PINTO F. A.; Fonseca L. M.; L.R. Abreu L. R.; Souza M. R.; Oliveira D. L. S.; Clementino L. A.; Andrade E. H. P. Infrared spectroscopy method to measure the whey composition in "minas padrão" and "prato" cheeses. **Arq. Bras. Med. Vet. Zootec.** v. 63 (4), p. 1039 – 1042. 2011.

PIOVEZAN M. **Otimização e caracterização química de extratos de amora preta (Rubus fruticosus var. Lochness e seu efeito sobre a dieta hiperlipídica em ratos mediante análise metabólica).** Florianópolis: Departamento de Química, UFSC, 2014. 212 p. (Tese de doutorado em Química).

PIZARRO C.; RODRÍGUEZ-TECEDOR S.; PÉREZ-DEL-NOTARIO N.; ESTEBAN-DÍEZ I.; GONZÁLEZ-SÁIZ J. M. Classification of Spanish extra virgin olive oils by data fusion of visible spectroscopic fingerprints and chemical descriptors. **Food Chemistry**, v 183, p. 915–922. 2013.

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

RAMOS J. M. T. A.. **Estudo espectroscópico vibracional de complexos de Ni(II) com os aminoácidos serina, glicina e ácido guanidoacético.** Niterói: Instituto de Química da UFF, 2006. 168 p. (Dissertação de Mestrado em Química Inorgânica).

RECHE, R. V.; FRANCO, D. W. Distinção entre cachaças destiladas em alambiques e em colunas usando quimiometria. **Química Nova**, v. 32 (2), p. 332-336. 2009.

RIGUEIRA, J. C. S. **Desenvolvimento de metodologia analítica para detecção de adulteração pela adição de leite em pó e leite fluido.** Viçosa: Departamento de Tecnologia de alimentos da UFV, 2006. 58 p. (Dissertação, Mestrado em Ciências e Tecnologia de Alimentos).

ROCHA W. F. C. **Utilização do sinal analítico líquido para validação de modelos de calibração multivariada através do cálculo de figuras de mérito e de cartas de controle.** Campinas: Departamento de Química Analítica da UNICAMP, 2007. 99 p. (Dissertação, Mestrado Química).

RODRIGUES, N. P. A.; GIVISIEZ, P. E. N.; QUEIROGA, R. C. R. E.; AZEVEDO, P. S.; GEBREYES, W. A.; OLIVEIRA, C. J. B. 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, p. 2749–2752. 2012.

RODRIGUEZ-SAONA, L. E.; ALLENDORF, M. E. Use of FTIR for Rapid Authentication and Detection of Adulteration of Food. **Annual Review of Food Science and Technology.** v. 2, p. 467–83. 2011.

ROSSO S. **Predição das propriedades de Eucalyptus grandis Maiden por espectroscopia no infravermelho próximo.** Curitiba: Setor de Ciências Agrárias da UFPR.2010, 126 p.(Tese de Doutorado em Engenharia Florestal).

SADEGHI-BAZARGANI *et al.*. Compared application of the new OPLS-DA statistical model versus partial least squares regression to manage large numbers of variables in an injury case-control study. **Scientific Research and Essays**, v. 6 (20), p. 4369-4377. 2011.

SANTOS JUNIOR A. A. Determinação de parâmetros de óleos lubrificantes de motores de ciclo Otto e ciclo Diesel através da espectroscopia no infravermelho, métodos multivariados e cartas de controle. Brasília: Instituto de Química da UnB. 2011, 134 p.(Dissertação, Mestrado em Química).

SANTOS P. M.; PEREIRA-FILHO E. R. Digital image analysis – an alternative tool for monitoring milk authenticity. **Journal: The Royal Society of Chemistry**. Anal.Methods, v.5, p. 3669–3674. 2013.

SANTOS P. M. **Aplicação de imagens digitais e técnicas espectroanalíticas combinadas com quimiometria para detecção e quantificação de adulteração em leite bovino**. São Carlos: Departamento de Química da UFScar. 2013, 79 p. (Tese, Doutorado em Química Analítica).

SANTOS, P. M.; PEREIRA-FILHO, E. R.; RODRIGUEZ-SAONA, L. E. Rapid detection and quantification of milk adulteration using infrared microspectroscopy and chemometrics analysis. **Food Chemistry**, London, v. 138 (1), p. 19-24. 2013.

SILVA M. A.M.; FERREIRA M. H.; BRAGA J. W.B.; SENA M. M. Development and analytical validation of a multivariate calibration method for determination of amoxicillin in suspension formulations by near infrared spectroscopy. **Talanta** v. 89, p. 342-35. 2012.

SIMÕES S. S. **Desenvolvimento de métodos validados para a determinação de captopril usando espectrometria NIR e calibração multivariada**. João Pessoa: Departamento de Química da UFPB, 2008. 97 p. (Tese, Doutorado em Química).

SINELLI N.; COSIO M.S.; GIGLIOTTI C.; CASIRAGHI E. Preliminary study on application of mid infrared spectroscopy for the evaluation of the virgin olive oil “freshness”. **Analytica Chimica Acta**. v. 598, p. 128-134. 2007.

SISO G. M. I. The biotechnological utilization of cheese whey: A review. **Bioresource Technology**, v. 57, p. 1-11. 1996.

SITORIUS E. D. **Development of Fourier Transform Mid-Infrared Spectroscopy as a Metabolomic Technique for Characterizing the Protective Properties of Grain Sorghum Against Oxidation**. Lincoln, Nebraska: Food Science and echnology Department University of Nebraska at Lincoln, 2010, 94 p.

SJAHFIRDI L.; MAYANGSARI; NASIKIN M. Protein identification using transform infrared (FTIR). **IJRRAS** v. 10 (3), p.418 – 421. 2012.

SOARES A., GIL E. S.; REZENDE K. R. Análises preliminares de substâncias viscosas em espectrômetro infravermelho utilizando suporte alternativo. **Revista Eletrônica de Farmácia** v.3 (2), p.63-67. 2006.

- SODINI L.; MORIN P.; OLABI A.; JIMÉNEZ-FLORES R. Compositional and Functional Properties of Buttermilk: A Comparison Between Sweet, Sour, and Whey Buttermilk. **J. Dairy Sci.** v.89, p. 525–536. 2006.
- 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 (1), p. 223-229. 2012.
- SOUZA L.M., MITSUTAKE H, GONTIJO L.C, BORGES NETO W. Quantification of residual automotive lubricant oil as an adulterant in Brazilian S-10 diesel using MIR spectroscopy and PLS. **Fuel**. v.130, p. 257 – 262. 2014.
- SOUZA, S. S., *et al.* Monitoring the authenticity of Brazilian UHT milk: A chemometric approach. **Food Chemistry**, v.124 (2), p. 692-695. 2011.
- SOYEURT, H. *et al.* Estimating Fatty Acid Content in Cow Milk Using Mid-Infrared Spectrometry. **J. Dairy Sci.** v.89, p.3690–3695. 2006.
- STUART B. H. Infrared Spectroscopy: **Fundamental and Applications**. 2005. Disponível em: < <http://onlinelibrary.wiley.com/book/10.1002/0470011149>>. Acesso em dez 2015.
- SZYMANSKA E. *et al.* Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. **Metabolomic**. v. 8, p. 3-16. 2012.
- TEIXEIRA L. V.; FONSECA L. M. Perfil físico-químico do soro de queijos mozzarella e minas padrão produzidos em várias regiões do estado de Minas Gerais. **Arq.Bras. Med. Vet. Zootec.**, v. 60 (1), p. 243-250. 2008.
- TEÓFILO, R. F; **Métodos Quimiométricos: Uma Visão Geral – Conceitos básicos de quimiometria**. Universidade Federal de Viçosa, Viçosa, v. 1, 2013. 118 p.
- THOMPSON M, ELLISON S.L.R, WOOD R. The International Harmonized Protocol for the proficiency testing of analytical chemistry laboratories (IUPAC Technical Report). **Pure Appl. Chem.** v.78 (1), p. 145 – 196. 2006.
- EU. EUROPEAN COMMUNITY. **Regulamento 2002/178/EC**. Disponível em: <<http://data.europa.eu/eli/reg/2002/178/oj>> Acesso em 8 dez 2016.
- USDEC - U. S. DAIRY EXPORT COUNCIL. **Ingredientes lácteos para uma alimentação saudável**. USDEC News, v. 2 (4), p.1-3. 2002.
- VALDERRAMA L.; GONÇALVES R. P.; MARÇO P. H.; VALDERRAMA P. Espectroscopia Uv-Vis e Método Quimiométrico na Avaliação de Adulterações e Fraudes em Azeite de Oliva Extra Virgem. *Revista Brasileira de Pesquisa em Alimentos* v. 5 (2), p. 32–40. 2014.
- VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. Estado da arte de figuras de mérito em calibração multivariada. **Química Nova**, v. 32 (5), p. 1278-1287. 2009.

VIM. International Vocabulary of Metrology – Basic and general concepts and associated terms – **JCGM** 200:2012.

WESTERHUIS J. A. *et. al.* Multivariate paired data analysis: multilevel PLSDA versus OPLSDA. **Metabolomics** v. 6, p. 119-128. 2010.

WESTERHUIS J. A. *et. Al.* Assessment of PLSDA cross validation. **Metabolomics**, v. 4, p. 81–89. 2008.

WILSON, R. H.; TAPP, H. S. Mid-infrared spectroscopy for food analysis: Recent new applications and relevant developments in sample presentation methods. **Trends Anal. Chem.**, Amsterdam, v. 18 (2), p. 85-93. 1999.

WOLFSCHOON A. F. Milk Analysis by Infrared Spectroscopy. **Rev. do ILCT**, v.32 p. 3-10. 1977.

WONG, N.P. Chemistry of processing. **Fundamentals of dairy chemistry**. 3 ed., Van Nostrand Reinhold: N.Y., 1988, p. 739-766.

WU D.; HE Y.; SUN D. Study on infrared spectroscopy technique for fast measurement of protein content in milk powder base. **Journal of Food Engineering** v. 84, p.124-131. 2008.

YU C.; IRUDAYARAJ J. Spectroscopic haracterization of Microorganisms by Fourier Transform Infrared. **MicrospectroscopyBiopolymers**, v. 77, p. 368–377. 2005.

ZADOW, J.G., Modern dairy technology: advances in milk processing. London: **Elsevier** v. 2. 1997.

ZHANG L.G; ZHANG X.; NI L.J.; XUE Z. B.; GU X.; HUANG S. X. Rapid identification of adulterated cow milk by non-linear pattern recognition methods based on near infrared spectroscopy. **Food Chemistry**, v. 145, p. 342-348. 2014.

ZHICHAO L.; WENSHENG C.; XUEGUANG S. Outlier detection in near-infrared spectroscopic analysis by using Monte Carlo cross-validation. **Science in China Series B: Chemistry**, v. 51(8), p. 751-759. 2008.